

ABSTRACT

Application of Chemometric Analysis to UV-Visible and Diffuse Near-Infrared Reflectance Spectra

Christopher Brent Davis, Ph.D.

Mentors: Kenneth W. Busch, Ph.D. and Marianna A. Busch, Ph.D.

Multivariate analysis of spectroscopic data has become more common place in analytical investigations due to several factors, including diode-array spectrometers, computer-assisted data acquisition systems, and chemometric modeling software. Chemometric regression modeling as well as classification studies were conducted on spectral data obtained with chili peppers and fabrics samples.

Multivariate regression models known as partial least squares (PLS-1) were developed from the spectral data of alcoholic extracts of Habanero peppers. The developed regression models were used to predict the total capsaicinoids concentration of a set of unknown samples. The ability of the regression models to correctly predict the total capsaicinoids concentration of unknown samples was evaluated in terms of the root mean square error or prediction (RMSEP). The prediction ability of the models produced was found to be robust and stable over time and in the face of instrumental modifications.

A near-infrared spectral database was developed from over 800 textile samples. Principal components analysis (PCA) was performed on the diffuse near-infrared

reflectance spectra from these commercially available textiles. The PCA models were combined together into a soft independent modeling of class analogy (SIMCA) in order to classify the samples according to fiber type. The samples in the study had no pretreatments. The discriminating power of these models was tested by creating validation sets within a given fiber type as well as attempting to classify samples into a category that they do not belong to.

The apparent sub-class groupings within the same fiber class were investigated as to whether or not they were caused by chemical processing residues, multipurpose finishes, or dyes.

Application of Chemometric Analysis to UV-Visible
and Diffuse Near-Infrared Reflectance Spectra

by

Christopher Brent Davis, B.S.

A Dissertation

Approved by the Department of Chemistry and Biochemistry

David E. Pennington, Ph.D., Interim Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Kenneth W. Busch, Ph.D., Chairperson

Marianna A. Busch, Ph.D., Chairperson

C. Kevin Chambliss, Ph.D.

Stephen L. Gipson, Ph.D.

Darrin J. Bellert, Ph.D.

Stephen I. Dworkin, Ph.D.

Accepted by the Graduate School
May 2007

J. Larry Lyon, Ph.D., Dean

Copyright © 2007 by Christopher Brent Davis

All rights reserved

TABLE OF CONTENTS

List of Figures	vi
List of Tables	xiv
Acknowledgments.....	xvi
CHAPTER ONE – Introduction	1
Chemometrics	1
Multivariate Data	2
Pattern Recognition Methods in Chemometrics	4
Unsupervised Pattern Recognition.....	5
Supervised Pattern Recognition.....	11
Multivariate Regression Techniques in Chemometrics	17
Partial Least Squares Regression Modeling	23
Overview and Research Objective.....	24
CHAPTER TWO – Determination of Capsaicinoids in Habanero Peppers by Chemometric Regression Analysis of UV Spectral Data	26
Introduction.....	26
Background.....	29
Ultraviolet-Visible Spectrometry.....	35
Materials and Methods.....	47
Experimental	47
Results and Discussion	50
Univariate Determination of Capsaicinoids.....	53

Multivariate Determination of Capsaicinoids	55
Independent Validation Study	67
Determination of Individual Capsaicinoids	71
Conclusion	76
CHAPTER THREE – Introduction of Textiles	78
Background	78
Fiber and Textile Taxonomy	79
History and Background of Natural Fiber	87
History and Background of Manufactured Fibers	102
Determination of Textile Fibers	116
CHAPTER FOUR – Soft Independent Modeling of Class Analogy for Textile Classification by Diffuse Near-Infrared Reflectance Spectroscopy	127
Introduction	127
Background	129
Near-Infrared Spectroscopy	129
Materials and Methods	140
Results and Discussion	143
Preliminary Study	143
Expanded Sample Population Study	158
Conclusion	176
CHAPTER FIVE – Identification of Chemically Treated Textiles by Soft Independent Modeling of Class Analogy Analysis of Diffuse Near-Infrared Reflectance Spectra	178
Introduction	178
Background	178

Softening Finishes.....	179
Flame-retardant Finishes.....	180
Chemical Dyes	184
Methods and Materials.....	189
Experimental	189
Results and Discussion	191
Finishes and Groupings.....	194
Water Content Modeling.....	203
Conclusion	205
CHAPTER SIX – Conclusions	207
APPENDICES	210
REFERENCES	225

LIST OF FIGURES

Figure	Page
1.1 Decision tree for pattern recognition	5
1.2 Result of unsupervised pattern recognition analysis multiple clusters present and data present in one cluster	6
1.3 Result of unsupervised pattern recognition analysis clusters conform to known class memberships and clustering not related to known information.....	6
1.4 Dendrogram derived from data and single link clustering HCA	7
1.5 Row plot of mean-centered data, with one PC, with two PCs, and with three PCs	8
1.6 Residual unexplained variance in a data set drops to zero.....	9
1.7 Soft Independent Modeling of Class Analogy.....	16
1.8 Decision tree for multivariate calibration	19
1.9 Two X-variables with high collinear nature	22
2.1 Energy differential between ground state and excited state of an electron....	36
2.2 Vibrational and rotational transitions superimposed on electronic transitions.....	37
2.3 Electronic transitions from (a) $\sigma \rightarrow \sigma^*$, (b) $n \rightarrow \sigma^*$, and (c) $\pi \rightarrow \pi^*$	38
2.4 Plot of absorbance as a function of concentration	42
2.5 Chemical deviations from Beer's law for unbuffered solutions of indicator HIn	43
2.6 Apparent deviation from Beer's law caused by varying amounts of stray radiation	45
2.7 Multichannel diode array spectrometer	46

2.8	Alcoholic extracts from Habanero peppers.....	48
2.9	HPLC chromatogram of natural capsaicin in ethanol.....	51
2.10	Calibration curve of capsaicinoids from HPLC peak area.....	52
2.11	HPLC chromatogram of Habanero pepper extract in ethanol	53
2.12	Plot of UV absorbance of pepper extracts at 230 and 280 nm versus concentration of capsaicinoids as determined by HPLC	55
2.13	UV absorption spectra of 31 Habanero peppers extracted with ethanol from 190-600 nm	57
2.14	Expanded (215-300 nm) UV absorption spectra of 31 Habanero peppers extracted with ethanol	58
2.15	Expanded (215-300 nm), mean-centered UV absorption spectra of 31 Habanero peppers extracted with ethanol	59
2.16	Summary of regression results for total capsaicinoids concentration.....	61
2.17	Comparison of RMSEP of 9 test sets with 2 and 4 PLS components.....	63
2.18	RMSEP for nine test sets at various wavelength windows.....	66
2.19	UV absorbance spectra for 12 Habanero pepper extracts used for Independent validation of the regression models prepared five months Earlier.....	67
2.20	Summary of regression results for total capsaicinoids concentration in independent validation series.....	68
2.21	Three-dimensional scores plot of independent validation set at 2 and 27 Weeks.....	71
2.22	Calibration curve for estimated capsaicin concentration	73
2.23	Summary of regression results for estimated capsaicin concentration.....	74
2.24	Calibration curve for estimated dihydrocapsaicin concentration.....	74
2.25	Summary of regression results for estimated dihydrocapsaicin concentration.....	75

2.26	Root mean square error of prediction results for individual capsaicinoids in nine test sets	75
3.1	Construction breakdown for a textile.....	80
3.2	General taxonomy for fiber classes.....	80
3.3	Cellulose molecule comprised of two β -d-glucose residues.....	81
3.4	Detailed taxonomy for natural fibers	84
3.5	Detailed taxonomy for manufactured fibers	86
3.6	Cross-section of mature cotton fiber.....	88
3.7	Cross-section and longitudinal view of unmeccerized and mercerized cotton fibers	89
3.8	Cross-sectional view of a flax fiber	90
3.9	Photomicrographs of flax fiber	91
3.10	Physical structure of a wool fiber	95
3.11	Chemical formula of wool molecule.....	96
3.12	α -helix geometry of wool	98
3.13	β -sheet structure of silk fibroin.....	100
3.14	The Viscose rayon process.....	103
3.15	Synthesis of rayon from cellulose.....	103
3.16	Orbital interactions in cuprammonium rayon.....	104
3.17	Photomicrographs of viscose rayon and cuprammonium rayon.....	105
3.18	Chemical structures for cellulose esters acetate and triacetate	106
3.19	Chemical structures for common nylon-6,6 and nylon-6	108
3.20	Chemical structures for common aramid fibers Novex and Kevlar	109
3.21	Chemical structure of polyamide polymer fiber Qiana.....	111

3.22	Chemical structure of polyethylene terephthalate (PET) polyester	112
3.23	Chemical structure of poly-1,4-cyclohexylenedimethylene terephthalate (PCM) polyester.....	113
3.24	Chemical structure of acrylic fibers.....	114
3.25	Chemical structure of modacrylic fibers.....	115
3.26	Comparison of photomicrographs for textile fibers.....	118
3.27	Near-infrared reflectance spectra of mercerized and unmercerized Cotton.....	125
4.1	Energy diagram of vibrational mode calculated as an ideal diatomic or as an actual anharmonic oscillator	131
4.2	Schematic representation of a sample for which Kubelka-Munk equation was derived	138
4.3	Schematic of dual-channel near-infrared spectrometer	142
4.4	Diffuse near-infrared spectra for 114 textile samples from 1100-2200 nm	143
4.5	Diffuse near-infrared spectra for 114 textile samples from 1334-1906 nm	145
4.6	Summary of principal components analysis for cotton and polyester samples with a wavelength window from 1334-1906 nm	146
4.7	Three-dimensional scores plot from principal components analysis of cotton and polyester fabric samples.....	147
4.8	Summary of principal components analysis for cotton and silk samples with a wavelength window from 1334-1906 nm	148
4.9	Three-dimensional scores plot from principal components analysis of cotton and silk fabric samples.....	149
4.10	Three-dimensional scores plot from principal components analysis of fibers from natural sources	149
4.11	Three-dimensional scores plot from principal components analysis of fibers from synthetic sources	151

4.12	Three-dimensional scores plot from principal components analysis of both natural and synthetic fibers	151
4.13	Residual x-variance plot from principal components analysis of cotton Samples	152
4.14	Comparison of cotton, polyester, and silk models classification abilities with fiber types not included in the calibration phase	155
4.15	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of cotton, polyester, and silk textiles from 1334-1906 nm.....	156
4.16	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 61 acetate textiles from 1334-1906 nm.....	159
4.17	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of acetate samples with a wavelength window from 1334-1906 nm	160
4.18	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 274 cotton textiles from 1334-1906 nm.....	162
4.19	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of acetate samples with a wavelength window from 1334-1906 nm	162
4.20	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 109 polyester textiles from 1334-1906 nm	164
4.21	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of polyester samples with a wavelength window from 1334-1906 nm	164
4.22	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 77 rayon textiles from 1334-1906 nm.....	166
4.23	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of rayon samples with a wavelength window from 1334-1906 nm	166
4.24	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 45 silk textiles from 1334-1906 nm.....	168
4.25	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of silk samples with a wavelength window from 1334-1906 nm	168

4.26	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 192 wool textiles from 1334-1906 nm.....	170
4.27	Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of wool samples with a wavelength window from 1334-1906 nm	170
4.28	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of misclassified acetate sample with representative acetate and polyester fibers	173
4.29	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of representative rayon and silk fiber samples	173
4.30	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of representative samples and the suspected “fake” silk textile.....	174
4.31	Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of nylon and acetate samples from 1334-1906 nm with insert from 1674-1846 nm	175
5.1	Schematic orientation of cationic and anionic softeners.....	180
5.2	Combustion cycle of fibers	181
5.3	Dehydration of cellulose by strong acids.....	182
5.4	Thermal degradation of cellulose to levoglucosan	183
5.5	Polyester flame retardants.....	184
5.6	Chemical structure for anionic dye Acid Blue 78.....	185
5.7	Chemical structure for anionic dye Direct Red 185.....	186
5.8	Chemical structure for anionic dye Mordant Brown 35	186
5.9	Chemical structure for anionic dye Reactive Red 1.....	187
5.10	Chemical structure for cationic dye Basic Blue 1.....	187
5.11	Chemical structure for vat dye Blue 1 (Indigo)	188
5.12	Chemical structure for sulfur dye Green 6.....	189
5.13	Chemical structures for coupling components of and azoic dye	189

5.14	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for wool samples with a wavelength window of 1334-1906 nm	192
5.15	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples with a wavelength window of 1334-1906 nm	193
5.16	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for rayon samples with a wavelength window of 1334-1906 nm	193
5.17	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for polyester samples with a wavelength window of 1334-1906 nm	194
5.18	Diffuse near-infrared reflectance spectra for cotton, linen, 65:35 polyester/cotton blend, and polyester	195
5.19	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton, linen, 65:35 polyester/cotton blend, and polyester	196
5.20	Diffuse near-infrared reflectance spectra for cotton, linen, 65:35 polyester/cotton blend, and polyester	197
5.21	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton, linen, 65:35 polyester/cotton blend, and polyester	198
5.22	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples	199
5.23	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples	200
5.24	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples	200
5.25	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples	201
5.26	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton samples	202

5.27	Diffuse near-infrared reflectance spectra for cotton organdy, Greige cotton duck, cotton fine filtercloth, and polyester coated with various fabric treatments from 1400 to 1900 nm.....	204
5.28	Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared spectra for cotton fine filtercloth, cotton organdy, Greige cotton duck, and polyester samples treated with various finishes for a wavelength region of 1334-1906 nm	205
A.1	Principal components analysis X-loadings plots for acetate at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components	212
A.2	Principal components analysis X-loadings plots for cotton at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components	213
A.3	Principal components analysis X-loadings plots for polyester at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components	214
A.4	Principal components analysis X-loadings plots for rayon at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components	215
A.5	Principal components analysis X-loadings plots for wool at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components	216

LIST OF TABLES

Table	Page
1.1 Typical response matrix for a set of the absorbance and concentration of sample solutions	21
2.1 Naturally occurring capsaicinoids.....	30
2.2 Scoville Heat Unit scale.....	34
2.3 Colors of visible radiation and their compliments.....	36
2.4 Typical absorptions of simple isolated chromophores.....	39
2.5 Absorption characteristics of some common chromophores	40
2.6 Solvent gradient for HPLC analysis	49
2.7 Physical data of natural capsaicinoids	54
2.8 Prediction results for the ten samples in test set 1	64
2.9 Summary of slope, offset, correlation coefficient, and RMSEP for 9 test sets	65
2.10 Prediction results for the 12 samples	69
2.11 Prediction of 12 samples after light source replacement	70
2.12 Estimated capsaicin and dihydrocapsaicin concentrations in commercial capsaicin.....	72
3.1 Protein structures	83
3.2 Composition of typical cotton fibers.....	89
3.3 Selected plant fibers of commercial interest.....	94
3.4 Amino acid contents of wool keratin fibers.....	97
3.5 Amino acid contents in silk fibroin of the <i>Bombyx mori</i> larva	99

3.6	Identification of textile fibers by burn testing.....	120
3.7	Solubility tests for fibers.....	121
4.1	Near-infrared spectral bands.....	133
4.2	Near-infrared bands for trichloroethane.....	134
4.3	Assumptions of the Kubelka-Munk equation.....	137
4.4	Principal bands in NIR spectra of cotton and polyester Terephthalate.....	144
4.5	Classification results of preliminary study for cotton, polyester, and silk.....	154
4.6	Classification results of preliminary study for cotton, polyester, and silk with Savitzky-Golay smoothing of raw diffuse near-infrared reflectance spectra.....	157
4.7	Textile fiber catalog for expanded study.....	159
4.8	SIMCA classification results for acetate fibers in expanded study.....	161
4.9	SIMCA classification results for cotton fibers in expanded study.....	163
4.10	SIMCA classification results for polyester fibers in expanded study.....	165
4.11	SIMCA classification results for rayon fibers in expanded study.....	167
4.12	SIMCA classification results for silk fibers in expanded study.....	169
4.13	SIMCA classification results for wool fibers in expanded study.....	171
5.1	Chemical structures of typical cationic and anionic softeners.....	181

ACKNOWLEDGMENTS

I would like to thank my advisors Dr. Kenneth W. Busch and Dr. Marianna A. Busch for their supervision, advice, encouragement, and patience with me during my graduate studies at Baylor. Without the support of my friends and colleagues from the Busch group, Jody Harvey, Patricia Diamond, Carolyn Markey, Dr. Jemima Ingle, Dr. Dennis Rabbe, Carlos Calleja, Selorm Modzabi, and Dr. Karen Humphrey, I would not have completed this monumental undertaking, and to my coworkers I say, thank you.

Additional thanks must also go to my committee members Dr. Kevin Chambliss, Dr. Stephen Gipson, Dr. Darrin Bellert, and Dr. Stephen Dworkin for their time and effort to read my dissertation in order to improve it.

Special thanks must be extended to Dr. Judith Lusk. Her knowledge of textiles and assistance in appropriating samples is what made my fabric classification studies possible. I would also like to thank Dr. Alton Hassell for his assistance for being my initial intermediary between Dr. Lusk and myself.

My appreciation also goes to the Department of Chemistry and Biochemistry, as well as Baylor University for allowing me to study here and for the financial support given to me.

My parents and family have always given me the utmost, unwavering support that I needed to complete this endeavor. I cannot express how much their love and encouragement has meant to me.

CHAPTER ONE

Introduction

Chemometrics

The term chemometrics was first coined by the Swedish physical organic chemist Svante Wold in June 1972.^{1a} Wold discussed the application of splines to fit data to a statistical model in the journal *Kemisk Tidskr.*² The origins of chemometrics can be traced as far back as 1969 when a series of papers authored by Jurs, Kowalski, and Isenhour was published in *Analytical Chemistry* on the use of a linear learning machine to classify a collection of low resolution mass spectra.³⁻⁵ In 1974, Kowalski and Wold founded the Chemometrics Society. The society defined chemometrics as a chemical discipline that combines mathematical and statistical methodologies to design or select optimal measurement procedures and experiments in order to afford maximum chemical information by analyzing the chemical data available.^{1a}

The basis of any chemometric analysis is the development of a statistical model that will relate the experimentally obtained variables to one another and show how these variables can be used to make qualitative and quantitative determinations about the samples. Pattern recognition or classification analyses are the preference for making qualitative assessments of chemical data. Multivariate regression modeling (MRM) is chosen when quantitative results concerning a dependent variable are desired. The multivariate nature of chemometric modeling allows the examiner to employ empirical modeling techniques, such as PCA and PLS, that can visually depict unexpected patterns

in the data because the joint effect of all the variables in the model are taken into account.⁶

Multivariate Data

Nature itself is multivariate. The weather patterns on the planet Earth are influenced by a number of factors, including wind, air pressure, temperature, and dew point among others. The health of a human being relies on a combination of variables, like genes, social position, stress, and eating habits. The multivariate character of nature does not stop at a laboratory door. In many scientific disciplines, the underlying casual relationships between components of a system manifest themselves in the form of observable data. The data is rarely dependent on one and only one variable.⁷

The question could be raised as to why chemists and technologists would concern themselves with using statistics when they spend large amounts of time and energy to calibrate their analytical instruments. The answer would be that statistics has useful concepts that are frequently lacking in other sciences, like chemistry. The quality of analytical data can be improved by applying multivariate analyses.⁸ Most chemical measurements are inherently multivariate, because more than one measurement can be made on a single sample. In a spectroscopic analysis, a spectrum can be recorded at hundreds of wavelengths on a single sample. The traditional methods of analysis in chemistry are univariate, where only one wavelength, or other measurement, is used per sample. Though tried and true, univariate techniques ignore the large amounts of information contained in the complete spectrum.⁹

Multivariate data analysis is used for a number of distinct and different purposes. The three main groups of objectives are data description, discrimination and

classification, and regression and prediction.⁷ Often times, multivariate data analysis is merely concerned with looking at the data. The characterization of the data is usually in the form of written summaries or displaying the intrinsic data graphically in plots.

Discrimination of the data deals with the separation of groups of data. The possibility exists to derive a quantitative data model in order to discriminate between two or more groups. Classification is related to discrimination, in this case however, typically, the relevant groups of the samples in the data set and the reasons behind those groupings are known. Regression is an approach for relating two sets of variables to each other. The process derives a relationship between one or more Y-variables on the basis of a well-chosen set of X-variables. Prediction is a means of determining Y-values for new sets of X-values based on a previously calibrated X-Y model.

Practical applications of statistical methods often have one or more observations showing a departure from the bulk of the data. These observations are called outliers or abnormal observations.⁸ Outliers can be seen in both calibration and predictions and occur for many different reasons. A system should be in place to identify possible outliers in a data set as well as a system to determine whether or not a sample should be removed from a classification or regression model. The sheer fact that a sample does not match up and cluster with others that are “supposedly” the same, does not mean a mistake has been made. Outliers may sometimes be highly informative and not an erroneous sample. The sample marked as an outlier may simply span a certain type of important variability in the X-data.

Pattern Recognition Methods in Chemometrics

In analytical chemistry, it is often necessary to observe similarities and differences in chemical samples based on the results of whatever analysis has been conducted. The determination of trends within the data set is called pattern recognition. Even as children, we are taught to discriminate between different shapes and colors while playing with blocks, games, and other toys. Television programs, such as Sesame Street, actively involve toddlers and young children in learning simple pattern recognition skills. This early indoctrination into pattern recognition spills over into the adult lives of most chemists. Spectroscopic data is usually presented as a continuous curve as opposed to a table of numbers, due to our ability to perceive shapes. When an absorbance or reflectance spectrum is examined, more often than not, it can be clearly seen that certain peaks are present, or absent.¹⁰

As human beings, our pattern recognition abilities are finite. Correlating and examining a collection of twenty different spectra can be a daunting task. When the collection of spectra grows to one hundred or more, the determination of class membership can be nearly impossible due to the sheer amount of data. In this case, minute variations may not be readily apparent to the human eye. Chemometric pattern recognition utilizes computer-assisted statistical analysis to make up for the shortcomings of simple visual inspection. Pattern recognition methods are used for elucidating or confirming groups of samples in multivariate row space. There are two discrete types of pattern recognition: unsupervised methods, also known as cluster analysis or numeric taxonomy, and supervised methods, also known by the blanket term of classification.

Figure 1.1 is a decision tree in order for an examiner to determine which type of classification analysis would be most appropriate for their analysis.

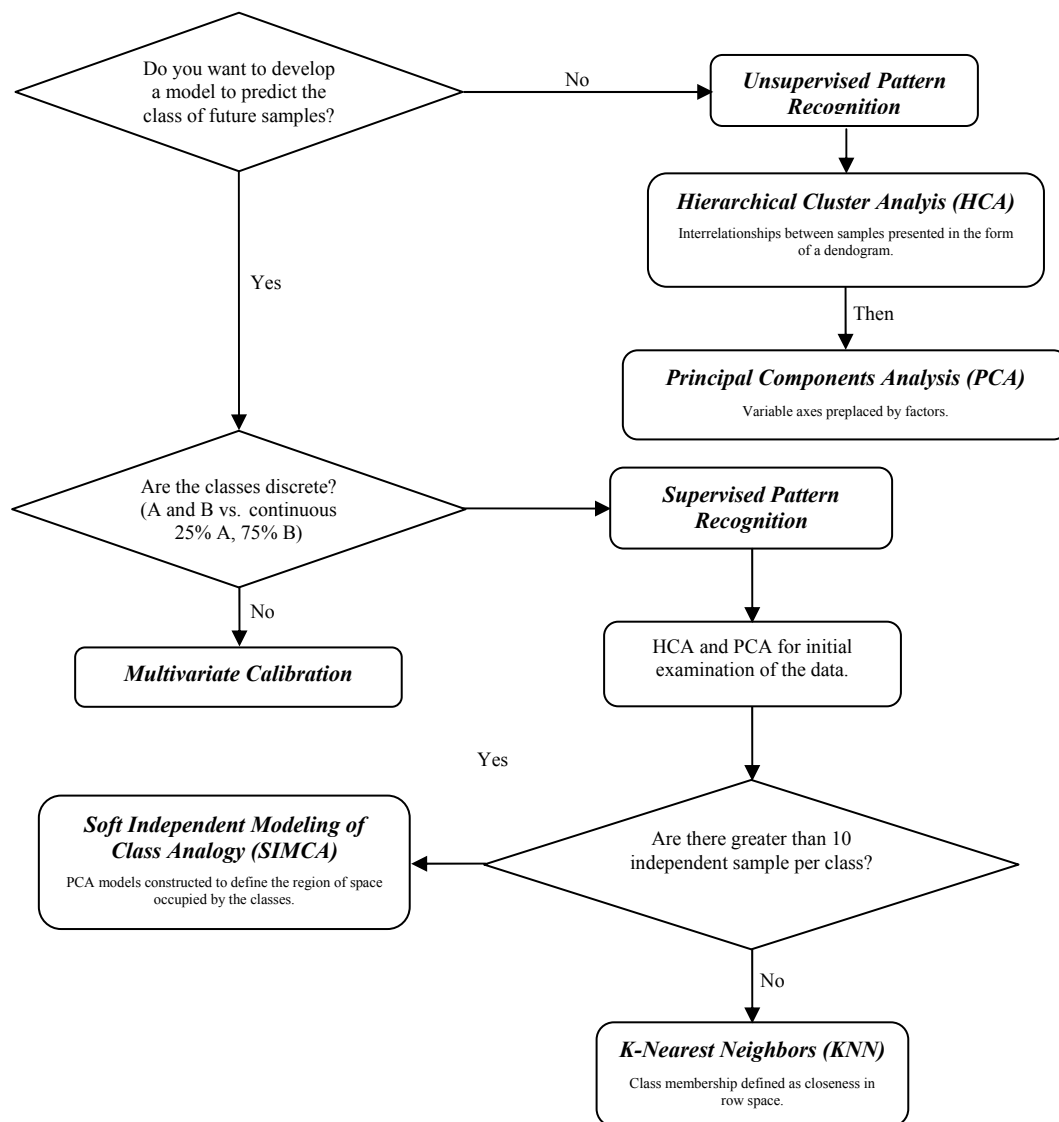


Figure 1.1. Decision tree for pattern recognition (Adapted from reference 10. Copyright 1998 John Wiley & Sons, Inc.)

Unsupervised Pattern Recognition

The primary goal of an unsupervised pattern recognition analysis is to determine whether or not clustering exists in a data set without the use of the class membership in

the actual calculation. Class membership of the samples in the data set is either not known or simply ignored, as can be seen in Figure 1.2. The groupings observed in the examination of row space are defined by the type of measurements taken from the samples. An examiner may simply wish to see if there are groupings within the set of samples or if outliers are present. If the class memberships are known, the examiner may choose to display or identify any groupings in the data swarm (Figure 1.3). This process is accomplished without imposing a class membership on the samples. The examination of natural clusters of the samples can lead to an increased understanding of the data set. Hierarchical Cluster Analysis (HCA) and Principal Components Analysis (PCA) are two readily used methods of unsupervised pattern recognition in chemometrics.

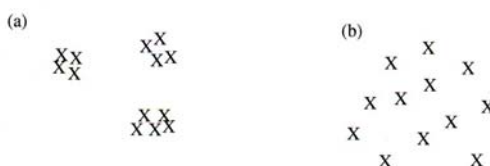


Figure 1.2. Result of unsupervised pattern recognition analysis (a) multiple clusters present and (b) data present in one cluster (Modified from reference 10. Copyright 1998 John Wiley & Sons, Inc.)



Figure 1.3. Result of unsupervised pattern recognition analysis (a) clusters conform to known class memberships (b) clustering not related to known information (Modified from reference 10. Copyright 1998 John Wiley & Sons, Inc.)

1. Hierarchical Cluster Analysis. HCA is an unsupervised technique that examines the interpoint distances between all of the samples and represents that

information in the form of a two-dimensional plot known as a dendrogram (Figure 1.4a). This depiction brings higher-dimensional row space into a simpler format by which human pattern recognition skills can deduce natural clusters. Dendrograms used in HCA methods form clusters based on the samples' nearness in row space. Initially, every sample is treated as a cluster and the closest clusters are joined together. This process is repeated over and over again until only one cluster remains. Depending on what variation of HCA is used, the measurement of distances between clusters can be calculated by single versus centroid linkage or Euclidean versus Mahalanobis or statistical distance. Samples joined together at small distances are similar based on the measurement system. However, samples joined at large distances can indicate the presence of outliers in the data matrix.

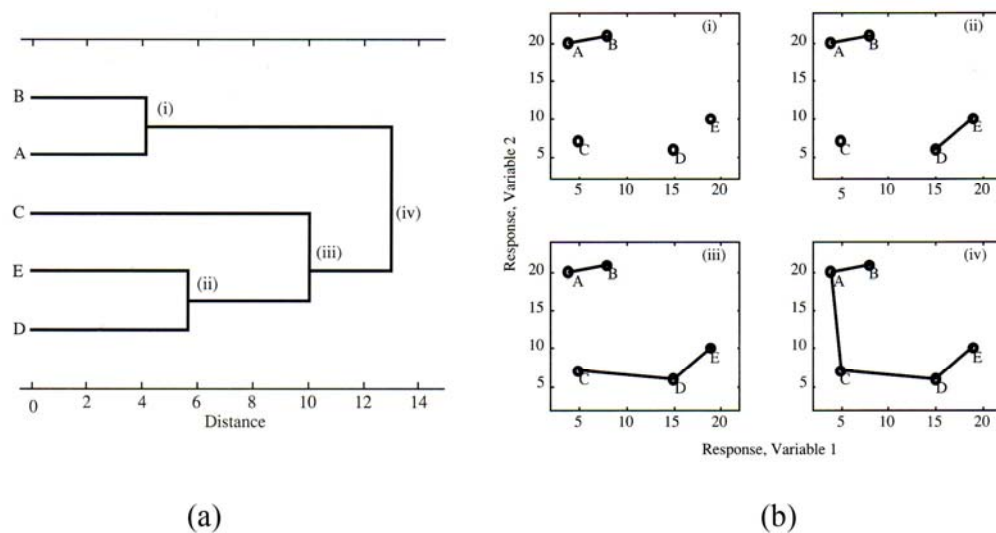


Figure 1.4. (a) Dendrogram derived from data (b) Single-link clustering HCA (Modified from reference 10. Copyright 1998 John Wiley & Sons, Inc.)

2. Principal Components Analysis. Another unsupervised method of pattern recognition is principal components analysis. The aim of PCA is to represent the

variation present in many variables using a reduced number of factors by mathematically manipulating the data matrix. These factors redefine an axis system in a new row space as opposed to the original measurement variables. The new axes, also referred to as principal components (PCs), allow the matrix variables to be viewed in a true multivariate fashion in a relatively small number of dimensions. A successful PCA describes the interpoint distances using as few axes, or dimensions, as possible.

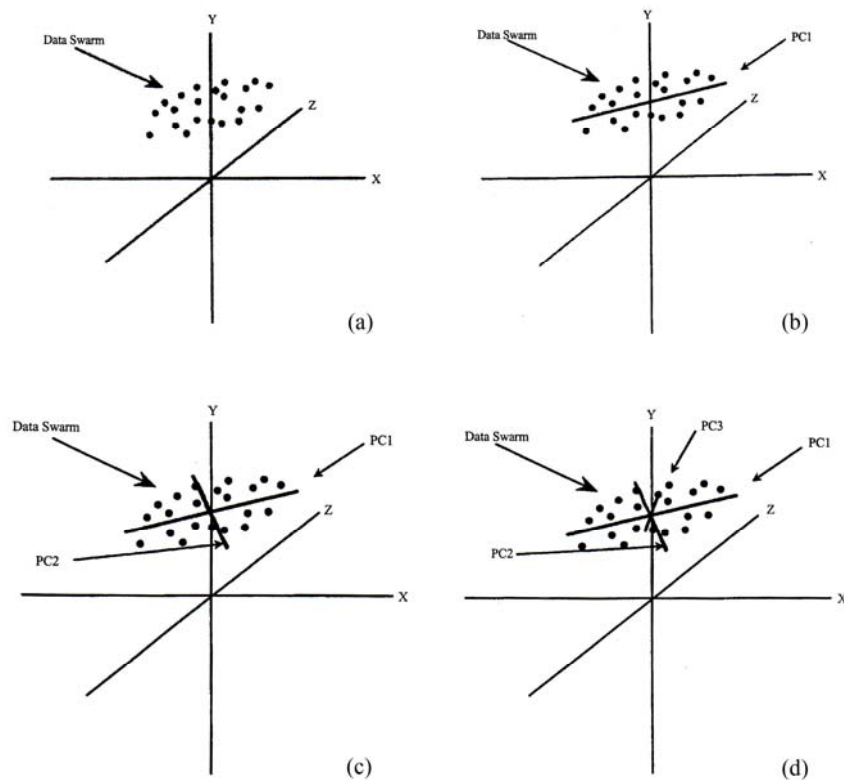


Figure 1.5. A row plot of (a) mean-centered data (b) with one PC (c) with two PCs and (d) with three PCs (Reproduced with permission from reference 11. Copyright 2004 Sayo Fakayode)

The maximum amount of variation possible in the data set (Figure 1.5.a) in one direction is explained by the first principal component calculated (Figure 1.5.b). Each successive PC will describe ever decreasing amounts of variation in the data set. This

fact can be helpful when determining the appropriate number of principal components to include in order to describe the full variation in the sample set (Figures 1.5c-d). Any determinations based on a PC plot should be gauged by the amount of variation that is explained by the factors in the plot. When an examiner removes non-significant principal components, noise from the data set can be filtered out. Every principal component contains some degree of noise; however, the noise is spread throughout all of them. The signal-to-noise ratio is greatest for the first PC and decreases for the PCs subsequently calculated as is seen in Figure 1.6. The subsequent PCs generated describe more and more of the noise in the data matrix, as opposed to the relevant chemical information. Principal components, as used in chemometrics, are orthogonal to one another. The maximum number of principal components capable of being calculated is always less than the number of samples or variables.

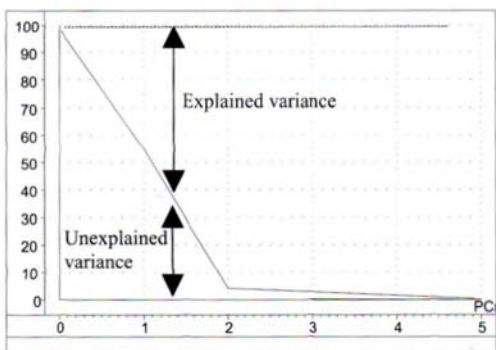


Figure 1.6. Residual unexplained variance in a data set drops to zero (Reproduced with permission from reference 7. Copyright 2004 Camo AS)

PCA is most useful when the dimensionality of the measurement space is large, but the samples reside in a small dimensional space, also referred to as having small inherent dimensionality or rank. When discussing PCA, inherent dimensionality is

defined as the number of principal components needed to describe a data set down to the noise level. Striking the balance of relevant PCs and excluding the pure noise components is difficult. The inherent dimensionality of the data matrix can be constrained by a number of factors, including: number of measurements, number of samples analyzed, and the number of chemical components present in a sample.

Both hierarchical cluster and principal components analyses have their individual strengths and weaknesses.¹⁰ When using HCA, the analysis does not identify which variable or variables contribute to the overall clustering found in the dendrogram. The individual discriminating power of the variables used in the model cannot be determined. Any poorly resolved clusters found are often difficult to interpret using HCA, due to the fact that some of the geometric information is not retained. Dendrograms are fixed representations of the variation in the sample matrix. There is not a method of manipulation that can transform the dendrogram into a form that would be more accessible to human pattern recognition abilities.

There are some advantages to using HCA. For instance, all of the variation in the original data set is depicted in the dendrogram. Principal components analysis displays different fractions of variation in a succession of scores plots. HCA also negates the need to determine the rank of a data matrix.

PCA requires that the rank of a data set be determined. Rank determination is not always a straightforward procedure. The goal in creating a PCA is to explain the variation in a sample set with the least number of principal components possible. Even with decreased dimensionality, it is still possible to have a large number of plots explaining the variation in the sample set. The larger the inherent dimensionality of a

data matrix becomes, the less that PCA will aid in deriving meaningful information from it.

PCA does provide information about the measurement variables in the data set. A loadings plot will indicate the degree of variation contained in each individual measurement variable. Another added feature of PCA is the use of interactive tools that can allow the examiner to explore the analysis results in different ways. It is possible to rotate the PC axes in three dimensions to afford different views which may have obscured various clusters or outliers behind another.

The unsupervised methods of hierarchical cluster analysis and principal components analysis are often times done in conjunction with one another. They compute similar information, but in dramatically different forms. Investigation of both analyses allows for a broader examination of the data by HCA, and PCA will allow for examination of the samples and clusters that are identified by HCA. Both methods have proven their validity in preliminary data analysis.

Supervised Pattern Recognition

Supervised pattern recognition analyses do include class membership information in calculations of models. The prediction of class membership for future samples is the primary goal in the construction of models from different analytical measurements. In order to properly calibrate the models generated, both the class shape and location are taken into account.

1. K-Nearest Neighbor. The supervised pattern recognition technique of K-Nearest Neighbor (KNN) predicts the class of an unknown sample by calculating the

nearest sample to the unknown in multidimensional space.¹⁰ This distance can be Euclidean, which is calculated by equation (1.1):

$$Distance = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1.1)$$

where x_i and y_i are the coordinates of samples x and y in the i^{th} dimension of the row space, where i ranges from 1 to n . A training, or calibration, set, with a known class membership, is used to classify an unknown by calculating the Euclidean distance between it and the training set. The closest samples (K) to the unknown are then used to make the determination of class membership. In order to determine the optimal K for class determination, the original data set with known class identities undergoes a cross-validation to determine the degree of separation between different groups. A K-value of 1 indicates that the closest neighbor to a given sample has a high likelihood of being a member of the same class.

The K-values offer a more qualitative measure of class membership for the unknown sample. One quantitative measure of validating the prediction is “goodness value”, G .¹⁰ This approach involves the comparison of the distance from the unknown sample to the suspect class, relative to an expected distance for known members. The goodness value is obtained by first calculating the distance (d_u) from the unknown to the closest member of the suspect class. Next, the interpoint nearest-neighbor distances for each of the samples in the calibration set for the class being considered are determined. Finally, the mean and standard deviation for these values are calculated. These three values, the unknown distance, the standard deviation and mean for the interpoint nearest-neighbor distances are placed into equation 1.2 to calculate the goodness value as can be seen below:

$$G = \frac{d_u - \bar{d}_c}{sd(d_c)} \quad (1.2)$$

where, d_u is the distance of the unknown from the suspect class, \bar{d}_c is the mean of the interpoint distances, and $sd(d_c)$ is the standard deviation of those values.

The goodness value, G , acts similarly to the t-value often used in traditional univariate statistics which indicates the number of standard deviation units the distance of the unknown is from the average class distance. With this in mind, it is logical to see that the smaller the G -value is, the more confidence one will have in describing the unknown as a class member. It is possible for the G -value of a data set to be negative if the unknown sample is closer to the interpoint distance than all of the samples used in the calibration set. The larger the G -value becomes, the less certainty an examiner will have in declaring the unknown sample to be a member of the class.

The traditional KNN technique does not detect the presence of outliers in the sample data set. The unknowns placed into a KNN classification will always be classified into one of the member groups in the training sets provided. In addition, K-Nearest Neighbor also does not take advantage of the class shape information that is available. KNN does have an added value in the fact that it is a simple methodology to implement. KNN is applicable when the number of samples per class is relatively small.

The conceptually simple approach of KNN works well for a variety of situations, but its limitations must be understood. The numbers in each class should be approximately equal, otherwise, the votes will be biased towards the class with the most representatives.⁹ Also, for the simplest implementations, each variable in the data matrix assumes equal significance. In spectroscopy, an examiner may record hundreds of wavelengths. A large number of those variables might not be diagnostically important, or

worse, be highly correlated in a way that does not offer a classification model that accurately depicts the status of the sample population. A way of circumventing this problem is to either use some form of variable selection or to use a distance calculation other than Euclidean, such as the Mahalanobis distance. The Mahalanobis distance is a distance measure based on correlations between variables by which different patterns can be identified and analyzed. This measure is useful for determining the similarity of an unknown sample set to a known one. The primary difference between Euclidean and Mahalanobis distances is that Mahalanobis takes correlations in the data into account when calculations are made and is scale-invariant, not dependent on the scale of the measurements. Another problem is the presence of ambiguous or outlying samples in the data set. Also, KNN does not take into account the spread or variance within a class. If an examiner was trying to determine whether a forensic sample was a forgery, it is likely that the class of forgeries has a much higher variance compared with the class of non-forged samples.

2. Soft Independent Modeling of Class Analogy. SIMCA was first introduced by Svante Wold in 1977.¹² The idea behind soft modeling is that two classes can overlap. In soft modeling, there is no problem with belonging to both of the classes or neither. The added value of SIMCA is that the analysis will not automatically place a sample into a category solely because that sample was placed into the analysis. The philosophy of soft modeling is that, in many situations in chemistry, it is entirely legitimate for an object to fit into one or more classes simultaneously. For example, a compound could contain an ester and an alkene group. When the compound is analyzed by a spectroscopic method, it will show characteristics of both functional groups. A

classification method that assumes the compound belongs to only one or the other is unrealistic. When hard modeling is employed, like in other areas of statistics, an object is required to belong to a discrete class. K-Nearest Neighbor is an example of a hard-modeling technique.

In a SIMCA pattern recognition analysis, a separate principal component analysis is conducted for each class in the data set.⁹ Each class is represented by an individual PCA that takes the form of equation 1.3:

$$\chi_i = 1X_i + T_i \cdot P_i + E_i \quad (1.3)$$

where χ_i is the $n \times p$ data matrix of class i , $1X_i$ is the mean vector matrix of class i , T_i is the $n \times F$ scores matrix, P_i is the $F \times P$ loadings matrix, and E_i is the residual $n \times P$ matrix. The scores matrix provides the coordinates for each sample in the principal component space. The loadings matrix contains the necessary information for transforming the original variables into the principal components. The F -value is determined by the number of principal components necessary to accurately model the data within each class. The residual matrix indicates the extent to which the arrived upon model explains the data within the original matrix.¹³ The residual variance for a particular class to determine the quality of the cluster can be calculated with equation 14:

$$s_o^2 = \sum_{i=1}^N \sum_{j=1}^P \frac{(e_{ij})^2}{(P - F)(N - F - 1)} \quad (1.4)$$

where e_{ij} are the components of the residual matrix E_i , P is the dimensionality of the data, F is the number of principal components necessary to define the class, and N is the number of samples in the given class. The residual variance of fit indicates the degree to which a sample fits into a class model, and is given by equation 1.5:

$$s_i^2 = \sum_{j=1}^P \frac{(e_{ij})^2}{(P-F)} \quad (1.5)$$

The comparison of the s_o^2 and s_i^2 values will show whether or not sample i is representative of the class in question. When the two values are equivalent to one another, sample i is said to be representative of the class. If s_i^2 is significantly larger than s_o^2 , the sample is not representative of the class. The difference in the two benchmarks is evaluated using an F -test to see if s_i^2 is significantly larger than s_o^2 .

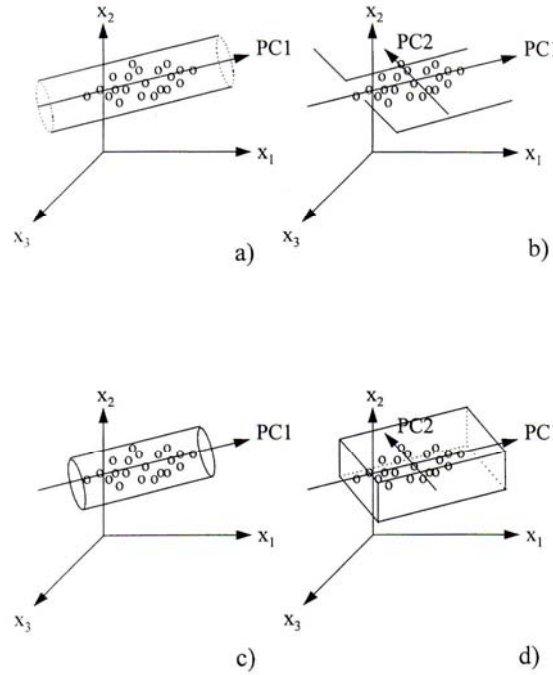


Figure 1.7. Soft Independent Modeling of Class Analogy (Reproduced with permission from reference 1b. Copyright 1998 Elsevier Science)

The modeling power of each variable for each separate class being modeled is defined by:

$$M_j = 1 - \left(\frac{s_{jresid}}{s_{jraw}} \right) \quad (1.6)$$

where $s_{j\text{raw}}$ is the standard deviation of the variable in the raw data and $s_{j\text{resid}}$ is the standard deviation of the variable in the residuals given by equation 1.4, which provides the difference between the observed data and the PC model data for the class. The modeling power (M) varies between 1 for excellent discriminating power and 0 for no discrimination. Variables with an M -value of 0.5 or lower are of little use unless they provide minute amounts of information that could be added to a model to achieve a classification for a difficult sample.

Multivariate Regression Techniques in Chemometrics

The ability to obtain ideal measurements in chemical analysis is often difficult. More often than not, measurements that are only selective for the constituents of interest are rarely found. In addition to the presence of usually more or less random measurement noise, the data may be affected by chemical and physical interferences due to phenomena in the samples themselves. Experimental interferences arising in the measurement process also contribute. Non-linearities can cause additional problems, because instruments seldom respond linearly to changes in the constituent concentrations and to changes in levels of interferences. Interferences can often be removed by means of separation procedures like filtration or some type of chromatography in order to ensure both selectivity and linearity. Often times, the purification of the samples of interest is prohibitively expensive or even physically impossible. The interferences and individual non-linearities represent less of a problem when the samples are subjected to multivariate calibration.⁸

Chemical interferences are described as systematic errors in the quantitative determination of a certain analyte when these errors are caused by other chemical

constituents or by chemically-induced variations in the analyte's own instrument response. Many chemical samples, especially those of a biological type, are mixtures consisting of several chemical constituents. Chemical similarity of those constituents can be to the degree that it affects the measurement. Physical interferences in a sample occur when systematic errors in the quantitative determination of a chemical constituent are caused by physical effects and not chemical ones. Irrelevant physical phenomena can affect the measured signal quite intensely. A common physical interference, seen in absorbance spectroscopy, is light scattering due to changing turbidity in a sample. Another physical effect is temperature. The water absorption peaks in the near-infrared spectrum, at approximately 1940 nm, are greatly affected by variations in temperature. If the temperature of the samples is not kept constant or compensated for in some way, the prediction of chemical composition for things, such as protein, lipid, carbohydrate, and water content, in foods and feeds using NIR reflectance measurements will be inaccurate.

In regression calibration, a regression model is developed to describe the experimental data by forming a mathematical relationship between factors or dependent variables, such as spectral data, and a response or dependent variable, like the concentration of an analyte(s). Regression calibration can either be univariate linear regression or multivariate regression modeling.

Univariate linear models are developed from a single dependent variable and single independent variable. The common example of this univariate approach is when a calibration curve is constructed by plotting the absorbance or emission intensity, at a particular wavelength, as a function of concentration of the absorbing or emitting species.

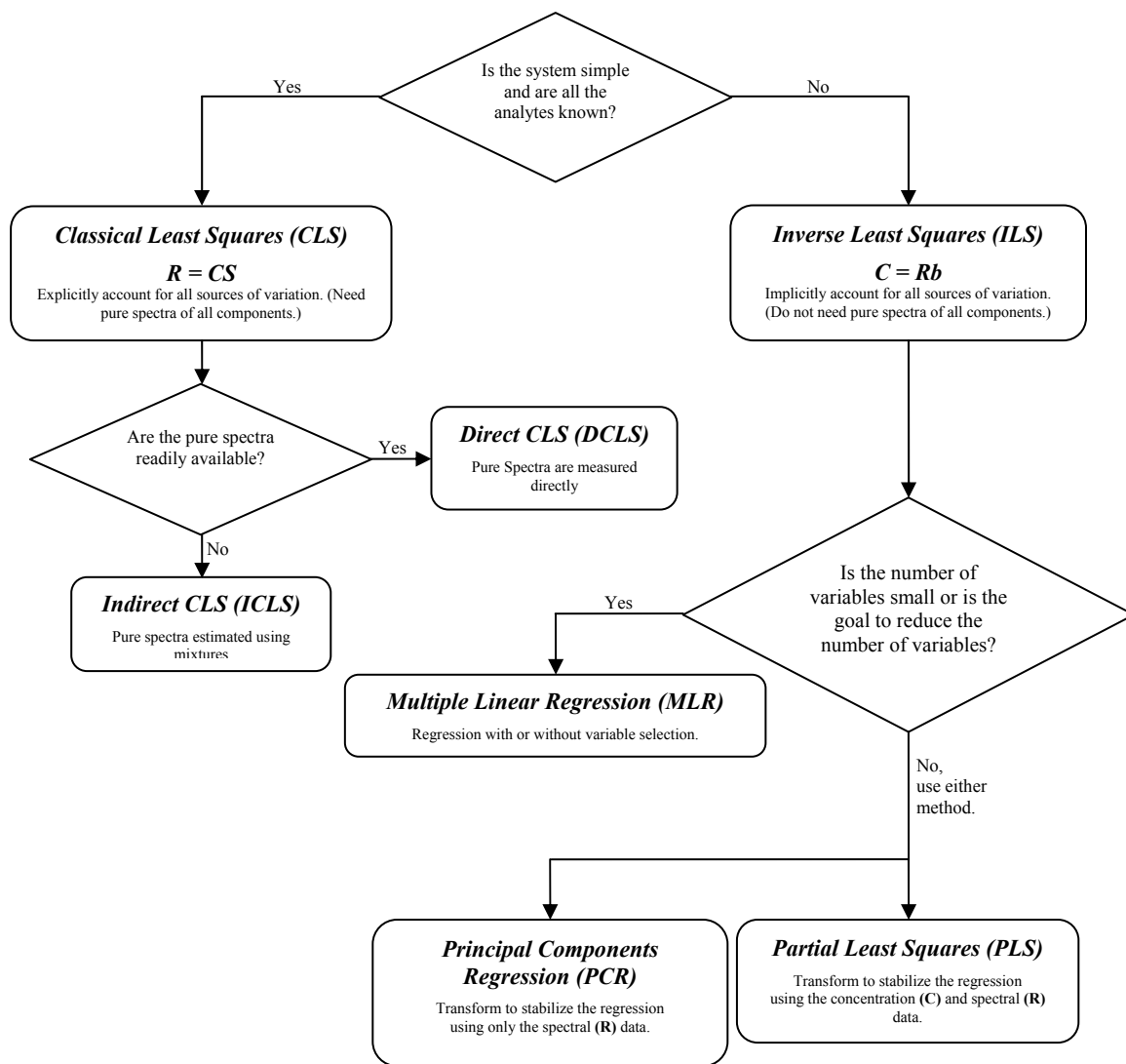


Figure 1.8. Decision tree for multivariate calibration (Adapted from reference 10. Copyright 1998 John Wiley & Sons, Inc.)

The relationship is described as follows:

$$y = b_0 + b_1x \quad (1.7)$$

where y is the dependent concentration variable, x is the independent absorbance or emission variable, b_0 is the intercept, and b_1 is the slope. While the determination of a single analyte constitutes a critical area of analytical science, multi-component analysis and multivariate data analysis are becoming more important. Chemometric approaches

like multiple linear regression (MLR), principal components regression (PCR), and partial-least-squares regression (PLSR) are often employed to carry out multivariate regression analyses. A decision tree to determine the appropriate type of regression modeling to be used by an examiner can be seen in Figure 1.8.

Consider a situation where the spectral data of sample solutions are recorded over a certain wavelength range. The spectral data can be seen in Table 1.1. The relationship between the concentration of the absorbing species and the spectral data can be described by the following equation:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_mX_m \quad (1.8)$$

where y is the concentration, b_i are the regression coefficients, and X_i are the measured absorbances at different wavelengths where $i = 1 \dots m$. The relationship between the response, the coefficients, and the independent variable can be expressed in matrix form by:

$$y = Xb \quad (1.9)$$

where the vector y contains the concentration, b is the vector containing the coefficients of the regression model, and X is the matrix of the response or spectral data.

The coefficient of the regression can be calculated as follows:

$$b = (X^T X)^{-1} X^T y \quad (1.10)$$

where X^T is the transpose of X . A variety of useful information can be obtained from the regression coefficients, including the significance of those coefficients in the regression model. The contribution of individual wavelengths to a given model can be seen in a plot of the regression coefficients as a function of wavelength. The future predictions of new samples is also dependent on the coefficients. The new spectral data (X_{fs}) introduced will

be weighted by the regression coefficients to predict concentrations (y) by the following equation:

$$y = X_{fs}b \quad (1.11)$$

In many analytical chemistry experiments, the $X^T X$ is not always invertible. There can be extensive collinearity between the variables involving large numbers of independent variables or columns of the sample matrix. Collinearity adds redundancy to regression model since more variables may be included in the model than are necessary

Table 1.1. Typical response matrix for a set of the absorbance and concentration of sample solutions

Sample No.	Absorbance at different wavelength (λ) (X-Data)								Concentration (Y-Data)
1	$X_{1\lambda 1}$	$X_{1\lambda 2}$	$X_{1\lambda 3}$	X_{1--}	X_{1--}	X_{1--}	X_{1--}	$X_{1\lambda m}$	y_1
2	$X_{2\lambda 1}$	$X_{2\lambda 2}$	$X_{2\lambda 3}$	X_{2--}	X_{2--}	X_{2--}	X_{2--}	$X_{2\lambda m}$	y_2
3	$X_{3\lambda 1}$	$X_{3\lambda 2}$	$X_{3\lambda 3}$	-----	-----	-----	-----	-----	y_3
4	$X_{4\lambda 1}$	$X_{4\lambda 2}$	$X_{4\lambda 3}$	-----	-----	-----	-----	-----	y_4
5	$X_{5\lambda 1}$	$X_{5\lambda 2}$	$X_{5\lambda 3}$	-----	-----	-----	-----	-----	y_5
-----	-----	-----	-----	-----	-----	-----	-----	-----	y_{--}
-----	-----	-----	-----	-----	-----	-----	-----	-----	y_{--}
-----	-----	-----	-----	-----	-----	-----	-----	-----	y_{--}
-----	-----	-----	-----	-----	-----	-----	-----	-----	y_{--}
-----	-----	-----	-----	-----	-----	-----	-----	-----	y_{--}
n-1	-----	-----	-----	-----	-----	-----	-----	$X_{1\lambda 1}$	y_{n-1}
n	$X_{n\lambda 1}$	$X_{n\lambda 2}$	$X_{n\lambda 3}$	$X_{n\lambda 4}$	X_{n--}	X_{n--}	X_{n--}	$X_{n\lambda m}$	y_n

for adequate description of the spectral variation and predictive performance. In mathematical terms, the X matrix is called collinear, or multi-collinear, if the columns in

X are approximately or exactly linearly dependent. In other words, X is collinear if at least one of the X -variables can be written as an approximate or exact linear combination of the others. If an examiner chooses to use multiple linear regression modeling for a data set, any collinearity in the X matrix could have a detrimental effect on the stability of the coefficients of b and render them useless for the casual interpretation. An example of collinear data can be seen in Figure 1.9. The high degree of correlation in the data can be used as basis the estimation of the first principal component or partial-least-squares component of a PCR or PLS analysis, respectively.

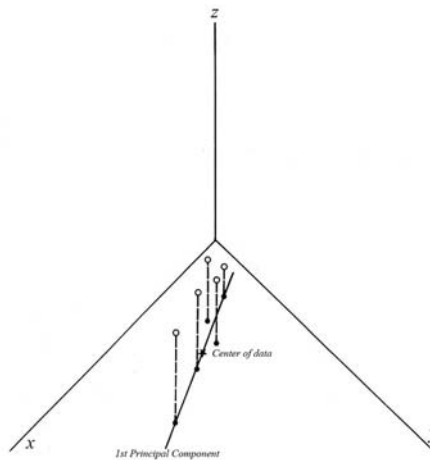


Figure 1.9. Two X -variables with high collinear nature (Modified from reference 8. Copyright 1989 John Wiley & Sons, Ltd.)

To avoid problems with collinearity in spectral data, all multivariate regression techniques require an orthogonal basis set or coordinate system on which to represent the data. Bilinear modeling techniques, such as principal components and partial-least-squares regression, employ projection methods to obtain a series of variance-scaled eigenvectors that can serve as a new coordinate system. A new matrix, U , with columns

that are linear combinations of the original columns from X is generated. The relationship is then given as the following:

$$y = Ub. \quad (1.12)$$

The inversion, $U^T U$, is now possible and eliminates the problem of collinearity among the spectral data. The reduction of the colinearity in the data matrix is also desirable because it reduces the dimensionality of the data.

Partial Least Squares Regression Modeling

Partial-least-squares regression (PLSR) is a powerful multivariate regression technique for reducing the colinearity of spectral data. PLS is often regarded as the major regression technique to use when dealing with multivariate data. The PLS method proceeds by successive linear combinations of the predictors in the X -matrix. Partial least squares balances the objectives of seeking factors that explain both response and prediction variation. PLSR differs from principal components regression by including the dependent variable in the data compression and decomposition operations. This means that both the X and Y data are actively involved in the construction of the new basis set made up of PLS components.¹¹ The information in the data that is not of importance to the model's predictive abilities is not weighted as heavily as that which does directly correlate to the parameter of interest.

PLS also takes into account the errors in both the concentration estimates and the spectra. Methods such as PCR assume that the concentration estimates are free of error. In chemistry, there are often errors involved in sample preparations, including weighing of solids and making dilutions. PLS assumes that the error in the experimentation is spread throughout the spectral and concentration data.

Overview and Research Objective

The research studies reported in this dissertation address some of the shortcomings of existing analytical methods for determining chili pepper pungency as well as identifying fiber composition of textiles. A new spectrophotometric method for the determination of total capsaicinoid concentration (capsaicin and dihydrocapsaicin) was developed by combining UV-visible absorption spectroscopy and partial-least-squares regression modeling. Typical pepper pungency analysis involves the use of liquid chromatography to separate the individual capsaicinoids before they can be detected and quantified. The new method developed does not require the separation of the capsaicinoids prior to the collection of the UV spectrum. The results of this study can be seen in Chapter 2.

The analysis of textiles by diffuse near-infrared reflectance spectroscopy is investigated in Chapter 4. Soft Independent Modeling of Class Analogy (SIMCA) classification was used to process the near-infrared spectra and make predictive principal components models that were used to classify commercially available textiles. The possibilities of classifying fibers on the basis of their origin (natural cellulose, mineral, synthetic, etc.) as well as by their fiber type (cotton, flax, polyester, etc.) were investigated. A preliminary study was carried out as a test to see whether the modeling of the spectral data would be able to distinguish samples based on fiber content. An expanded study, which comprised a database of over 800 individual fabric samples, was completed in order to increase the accuracy of the predictive classifications by allowing for more variability within a fiber class.

The results the classification studies yielded different substructures in the scores plots for the principal components analyses (PCA) of some individual fiber classes. The reasons behind the phenomena were investigated. A collection of untreated fabric samples was coated with a variety of different textile finishes and dyes. The resulting scores plots from the PCAs were examined to see whether or not the finishing procedures common to textile manufacturing could cause stratification within a particular fiber grouping. This study is reported in Chapter 5.

CHAPTER TWO

Determination of Capsaicinoids in Habanero Peppers by Chemometric Analysis of UV Spectral Data

Introduction

While UV-visible spectrophotometry is often highly sensitive – particularly for analytes with high molar absorptivities – selectivity can frequently be a problem with complex samples because contaminants in the sample can produce absorption bands that overlap those of the analyte. As a result, ordinary UV-visible spectrophotometry at a single wavelength generally requires that the analyte be separated from the other absorbing constituents in the sample matrix before the absorbance measurements are made. This prior separation step can be problematic when the sample matrix is complex. In the case of chili pepper samples, capsaicinoids are generally extracted from the flesh of the pepper with solvents like ethyl alcohol. While extracting capsaicinoids from a chili pepper is straightforward, other absorbing compounds – particularly carotenoids – are simultaneously co-extracted from the pepper. As a result, traditional spectrophotometric methods require sample purification before the absorbance spectrum is collected.¹⁴

Though techniques for the chemical purification of pepper extracts have been developed for spectrophotometric procedures, sample cleanup can be expensive and time-consuming because it adds an extra step to the procedure and may generate chemical waste that requires proper disposal. In analytical situations like this, chemometric techniques can be beneficial.

Chemometric techniques have gained acceptance in the past several years as an effective tool in extracting information out of seemingly chaotic uncontrolled systems by use of different statistical algorithms. In chemistry, for example, multivariate regression modeling has been used as a means of correlating spectral data with known compositional changes.¹⁵⁻¹⁷ Techniques like partial-least-squares (PLS-1) regression modeling have been shown to be especially effective in removing the influence of unwanted variables that are impractical or impossible to control in the laboratory.⁶ The overall goal of this approach is to attempt to compensate for the interferences and other uncontrolled variables mathematically rather than resorting to chemical purification strategies.

Research into the determination of chili pepper pungency in our laboratory has been an ongoing process. The earliest work was a feasibility study to determine the validity of using multivariate modeling and near-infrared transmittance spectroscopy to determine pepper hotness.¹⁸ The results of the study demonstrated that there is correlation, as would be expected, between the predicted concentration of capsaicinoids and the actual values determined by high-performance liquid chromatography. No independent validation sets were created and analyzed during this study. The study was limited in regard to the number of samples included in the modeling process and used leverage correction. Also, the sample preparation required 70 g of chopped peppers to be used in the extraction of capsaicinoids.

In a further study, NIR transmittance was used to analyze chili pepper extracts using partial-least-squares modeling. The study also investigated the use of NIR reflectance measurements on pepper extracts adsorbed onto the solid support

Chromosorb.¹⁹ The results of this study showed that adsorption of the pepper extract onto the solid support improved the overall predictions from the regression models produced. However, the error of prediction for the validation samples reached as high as 43 ppm. The sample preparation was also cumbersome due to the 15 g of chopped peppers needed to complete the capsaicinoid extraction procedure, as well as preparing the samples adsorbed on Chromosorb.

The most recent work done on determining chili pepper hotness, prior to the work reported later in this Chapter, involved UV spectroscopy, as opposed to NIR.²⁰ This study attempted to build a universal partial-least-squares regression model that would predict capsaicinoid concentrations in a wide variety of peppers, from the mild jalapeno, to the pungent habanero. The sample preparation still utilized a large mass of chopped peppers from which the capsaicinoids were extracted.

PLS-1 is used to develop a mathematical model that correlates two sets of data so that the independent X-variables (spectral data in this case) can be used to predict the dependent Y-variable (capsaicinoids concentration in this case). PLS-1 regression modeling is a two-step multivariate process that makes use of actual real samples (rather than laboratory-prepared standards) in calibration. By using actual real samples for calibration, all possible interferences are ideally present in the spectral data used to develop the regression model. The procedure is multivariate because it makes use of spectral ranges rather than a single wavelength.

This chapter will report on the application of multivariate statistical analyses, particularly PLS-1, to UV spectral data of habanero pepper extracts to develop regression models aimed at predicting the combined capsaicin and dihydrocapsaicin content of

pepper extracts from their UV-visible absorption spectra. This technique has the potential to permit direct spectrophotometric determination of pepper pungency in pepper extracts without prior separation of the matrix.

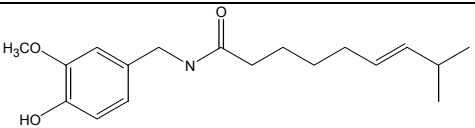
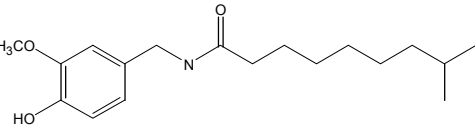
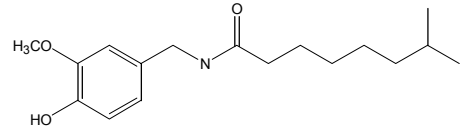
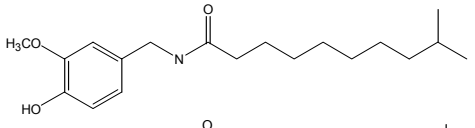
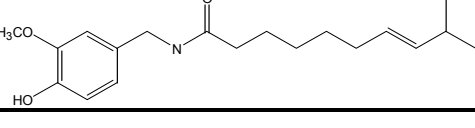
Background

Capsaicin (*N*-[(4-hydroxy-3-methoxy-phenyl)methyl]-8-methyl-non-6-enamide) and dihydrocapsaicin (*N*-[(4-hydroxy-3-methoxy-phenyl)methyl]-8-methyl-nonanamide) are two members of a family of naturally occurring capsaicinoids (Table 2.1), which make up the pungent components of *Capsicum* fruits. Capsaicinoids are all *N*-vanillylamides of fatty acids, which differ by the length of their aliphatic side chain, the presence or absence of a double bond, the branching point, and their relative pungency. The first true data concerning the chemical structure of capsaicin was not seen until 1919. Nelson developed a proposed structure; however, it was not proven to be correct until the following year. Lapworth and Royle were able to obtain pure capsaicin, with a melting point of 64-65° C, but they disagreed with Nelson on his theory that capsaicin was a condensation product of vanillylamine and decenoic acid.

Until the discovery of five naturally occurring capsaicinoids by Bennett and Kirby in 1968, it was believed that the pungency of peppers was due solely to capsaicin. When evaluating the “hotness” of a pepper, capsaicin and dihydrocapsaicin constitute approximately 80-90% of the total capsaicinoids present.^{14, 21} The remaining capsaicinoids are nordihydrocapsaicin, homocapsaicin, and homodihydrocapsaicin, which are vanillylamides that can be derived from 7-methyloctanoic acid, 9-methyldec-trans-7-enoic acid, and 9-methyldecanoic acid, respectively. The relative concentrations of the naturally occurring capsaicinoids were determined to be 69% capsaicin, 22%

dihydrocapsaicin, 7% nordihydrocapsaicin, 1% homocapsaicin, and 1% homodihydrocapsaicin.²² The concentrations of capsaicin and dihydrocapsaicin in a chili pepper vary, depending on the species of the pepper, its growing conditions, and the time when the pepper was harvested.^{23, 24} The capsaicinoids are not evenly distributed throughout the pepper, and substantial amounts of the capsaicinoids exist in the pericarp and placenta of the individual fruits.²² The total concentration of capsaicinoids found in different varieties of peppers can be highly variable. Woodbury and colleagues measured the capsaicinoids in a variety of peppers in a study conducted in 1980. Paprikas were found to have between 0-30 ppm, 30-600 ppm in chili peppers, and a staggering 13,000 ppm in red peppers.

Table 2.1. Naturally occurring capsaicinoids in chili peppers

Capsaicinoid	Relative abundance	Scoville heat units	Chemical structure
Capsaicin	69%	15,000,000	
Dihydrocapsaicin	22%	15,000,000	
Nordihydrocapsaicin	7%	9,100,000	
Homodihydrocapsaicin	1%	8,600,000	
Homocapsaicin	1%	8,600,000	

Capsicums were first cultivated by Native American Indians that lived in tropical climates. Archeological evidence shows that *Capsicum* fruits appeared as early as 5000 to 7000 BC in Mexico and Central and South America.²⁵ The remains of dried pepper pods were found at the burials sites in Peru that dated back more than 2,000 years. The name, *Capsicum*, was originally ascribed to these plants by herbalists and later by taxonomists. Though the origin of the term is not known, the word may come from the Latin root *capsa*, meaning box, referring to the box-like shape some pepper pods have. Another possible derivation may stem from the Greek work *kapto*, meaning bite referring to the intense nature of the pungent principles contained within the *Capsicum* fruit. Taxonomists generally agree that there are five different cultivated species of peppers (genus *Capsicum*) that include *C. annum*, *C. frutescens*, *C. pendulum*, *C. pubescens*, and *C. chinense*. Investigations into the domestication and dispersal of chili pepper have identified samples from the genus *Capsicum* by analyzing microfossils present on artifacts.^{26, 27}

Habanero peppers belong to the *Capsicum chinense* variety. This species is typically grown in tropical lowland areas of Central and South America, specifically the Amazon Basin. The peppers are also grown in the Yucatan Peninsula of Mexico, Belize, Costa Rica, as well as in the states of Texas, Idaho, and California. Other peppers are grown in India, China, and throughout Southeast Asia.²⁵ On average, these pepper plants have large leaves and have fruits that are round and quite pungent. The mature fruits are commonly red, orange, or yellow in color.

Capsicum fruits have long been sought for their use in food preparation,²⁸ medicinal applications,^{29, 30} and personal protection aerosols.²³ Chile peppers and spices,

such as paprika, with different capsaicin contents are used routinely in salsas and other food products.²² Of all spices, Capsicums are the most popular and widely cultivated.³¹ Pungent sauces found Mexican and Asian cuisine often contain pungent peppers like the habanero.³² Capsaicin has been used in the treatment of rheumatoid arthritis, osteoarthritis, and other peripheral neuropathic disorders as a therapeutic pain reliever.¹⁹

Since capsaicin does not cause blistering or other surface damage to the skin, as such, it has been used as an ointment, liniment, or plaster applied externally, providing some pain relief by having an irritating effect on the sensory nerve endings. In addition, there are no deleterious effects on the capillaries, arteries, or veins in the patients' skin. Capsaicin is also used to gauge a patient's cough threshold after ingestion of an antitussive agent.²⁹ Recent studies have used capsaicin in other applications, such as neurobiological research,³⁰ weight management,^{33, 34} local/topical analgesia,³⁵ and antimicrobial defense.^{36, 37} Capsaicin has also been investigated as a possible treatment for prostate^{38, 39} and lung cancer.^{40, 41} Personal self-defense aerosols rely on the lachrymator-like effects caused when capsaicin is introduced to the mucous membranes.^{23, 42} The painful, burning sensations caused by capsaicin is due to its interaction with sensory neurons. Capsaicin binds with the vanilloid receptor subtype one (VR1), which is an ion-channel receptor. VR1 can be stimulated by heat and physical abrasion that allows for cations to pass through the cell membrane and into the cell when activated. When the receptor is bound to capsaicin, the molecule produces the same effect that excessive heat or abrasion damage would cause. The vanilloid receptor subtype one has become included in a super family of TRP ion channels. It is now commonly referred to as TRPV1.

In the past, organoleptic tests, like the Scoville Heat Test, introduced in 1912 by Wilbur Scoville, were used to determine pepper pungency.⁴³ The Scoville analysis required that a given amount of sample be macerated with alcohol and sweetened with 5-10 % aqueous sucrose solution. Panels of five judges then consume the mixtures and vote by committee as to the appropriate Scoville Heat Unit (SHU) to assign the sample in question. The basis of the test is dependent upon whether the judges can still detect the “bite” of the capsaicin even after many dilutions. A breakdown of the Scoville Heat Unit Scale can be seen in Table 2.2. While the fact remains that this analysis is simple and cost effective, it is clearly obvious that it has significant drawbacks. Problems include fatiguing of the judges, poor reproducibility, and no true standard panel. As a result, these methods have given way to modern instrumental analyses.^{14, 44}

Analytical methods employed for the determination of capsaicinoids have included colorimetry,^{39, 45} paper chromatography, thin layer chromatography, gas chromatography,⁴⁶⁻⁴⁹ liquid chromatography,^{32, 42, 50-60} GC/LC-MS,^{23, 61} NMR-flow probe analysis,⁶² spectrophotometry,⁶³⁻⁷⁰ amperometric titration,⁷¹ micellar electrokinetic capillary chromatography,⁷² and sensory methods such as an electronic nose.⁷³

Thin layer and paper chromatography techniques have been found to be an efficient, fast, and economical way to purify crude capsicum extracts. Colorimetric methods involve the reaction of the phenolic group on the capsaicin with a chromogenic reagent to induce a color change, or frequency shift, in the sample. This process can be quite tedious due to the fact that all pigments, fats, and other interferences must be removed in order to assure accuracy. The use of solvent extractions and chromatographic columns to purify capsaicin can add large amounts of time to the analysis as well as

Table 2.2. Scoville Heat Unit Scale

Type of pepper	Scoville heat units	Type of pepper	Scoville heat units
Capsaicin	1.5×10^7 - 1.6×10^7	Serrano Pepper	1.0×10^4 - 2.3×10^4
Nordihydrocapsaicin	9.1×10^6	Habanero Tabasco	7.0×10^3 - 8.0×10^3
US Grade pepper spray	2.0×10^6 - 5.3×10^6	Wax Pepper	5.0×10^3 - 1.0×10^4
Naga Jolokia	8.6×10^5 - 1.0×10^6	Jalapeno Pepper	2.5×10^3 - 8.0×10^3
Dorset Naga	8.7×10^5 - 9.7×10^5	Tabasco sauce	2.5×10^3 - 5.0×10^3
Red Savina	3.5×10^5 - 5.8×10^5	Rocotillo Pepper	1.5×10^3 - 2.5×10^3
Habanero		Poblano Pepper	1.0×10^3 - 1.5×10^3
Habanero chili	1.0×10^5 - 3.5×10^5	Green Pepper	6.0×10^2 - 8.0×10^2
Scotch Bonnet	1.0×10^5 - 2.0×10^5	Tabasco	
Jamaican Hot Pepper	1.0×10^5 - 2.0×10^5	New Mexico Pepper	5.0×10^2 - 1.0×10^3
Thai Pepper	5.0×10^4 - 1.0×10^5	Pimento	1.0×10^2 - 5.0×10^2
Cayenne Pepper	3.0×10^4 - 5.0×10^4	Bell Pepper	0

possibly losing a large portion of the analyte. UV spectroscopy is a reliable method for the determination of capsaicin; however, it also requires the removal of all interfering substances from the sample matrix before an analysis is conducted.

Of all these methods, the most widely used is high-performance liquid chromatography, which offers sufficient accuracy and precision. Lee and coworkers were the first to report on the clean separation of capsaicin and dihydrocapsaicin using reverse-phase HPLC.²⁵ The standard analytical method, as described by the American Spice Trade Association, called the Gillett method, employs HPLC for testing the pungency of capsicums and their oleoresins.⁷⁴ In this method, the pump flow rate is set

between 0.6-1.8 ml/min, and the UV detector is set at 280 nm for samples greater than 700 ppm. For samples less than 700 ppm, fluorescence detection is employed with excitation at 288 nm and, emission recorded at 320 nm. The analytical column is a 4.6 x 250 mm C-18 column (10 μ m packing) with a mobile phase consisting of a mixture of acetonitrile, dioxane, water, methanol, and perchloric acid depending on the sample concentration. Significant drawbacks to chromatographic methods include: expense of columns and solvents, analysis time, and production of chemical wastes that require proper disposal.

Ultraviolet-Visible Spectrometry

When one examines the visible region of the electromagnetic spectrum, it is found to cover a wide range of different wavelengths of light from approximately 400 to 750 nm. A breakdown of the individual components of the visible spectrum can be seen in Table 2.3. Beyond the visible region, at shorter wavelengths from about 400 nm down to 190 nm, is the ultraviolet region.

1. Types of Electronic Transitions. The energies of the photons between 200-800 nm correspond to the energies of excitation of outer valence electrons and inner shell, *d-d* transitions with related vibrational levels. An electron will be promoted from an occupied orbital of low energy, the ground state, to an orbital of higher energy, known as the excited state. This can be seen in Figure 2.1. The most probable transition is that of the electron moving from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO).

The magnitude of energy that an electron can absorb is specific due to the fact that the energies of the orbitals involved in the electronic transitions are quantized.¹¹ This implies that absorptions in the UV-visible spectrum should be sharp, well-defined

Table 2.3. Colors of visible radiation and their compliments

Wavelength range (nm)	Color of reflected light	Color of absorbed light
400-465	Violet	Yellow-green
465-482	Blue	Yellow
482-487	Greenish-blue	Orange
487-493	Blue-green	Red-orange
493-498	Bluish-green	Red
498-530	Green	Red-purple
530-559	Yellowish-green	Reddish-purple
591-571	Yellow-green	Purple
571-576	Greenish-yellow	Violet
576-580	Yellow	Blue
580-587	Yellowish-orange	Blue
587-597	Orange	Greenish-blue
597-617	Reddish-orange	Blue-green
617-780	Red	Blue-green

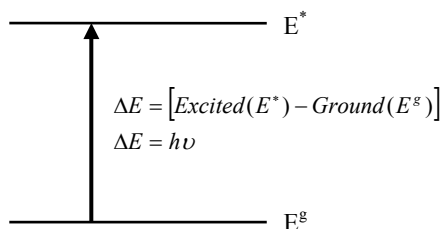


Figure 2.1. Energy differential between the ground state and excited state of an electron (Adapted from reference 79. Copyright 1996 Harcourt Brace & Company)

peaks. As can be seen in Figure 2.2, the broad bands in the spectrum are due to the superposition of vibrational and rotational levels on the electronic transitions. This phenomenon is the reason that the absorption spectrum is a broad continuum instead of the anticipated sharp peaks.

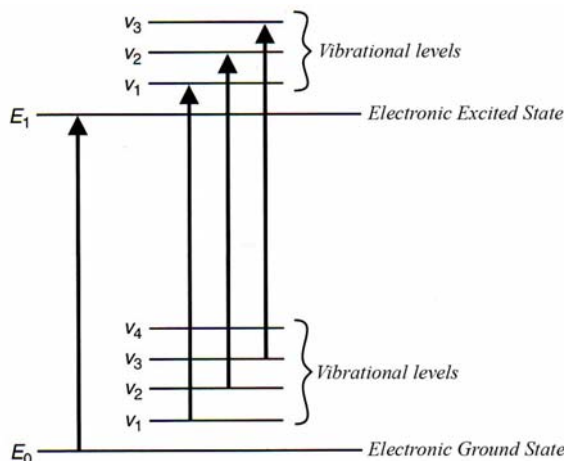


Figure 2.2. Vibrational and rotational transitions superimposed on electronic transitions (Modified from reference 79. Copyright 1996 Harcourt Brace & Company)

There are three basic types of electronic transitions involving: (1) sigma(σ), pi (π), and non-bonding (n) electrons; (2) charge transfer electrons; and (3) $d-d$ transitions. The $\sigma \rightarrow \sigma^*$ and $\pi \rightarrow \pi^*$ transitions involve the excitation of sigma and pi bonding orbital electrons to their corresponding antibonding orbitals. Molecules that contain only single bonds and lack atoms with unshared electron pairs only exhibit transitions of the $\sigma \rightarrow \sigma^*$ type. Because these transitions are of sufficiently high energy, the ultraviolet energy is absorbed at a very short wavelength. Due to instrumental shortcomings, the $\sigma \rightarrow \sigma^*$ transitions are not normally observed in the UV. This particular type of transition is shown in Figure 2.3a. The $n \rightarrow \sigma^*$ transition (Figure 2.3b) becomes important in saturated

molecules, such as alcohols, ethers, amines, and sulfur compounds, containing atoms bearing nonbonding pairs of electrons. Unsaturated molecules, like alkenes and

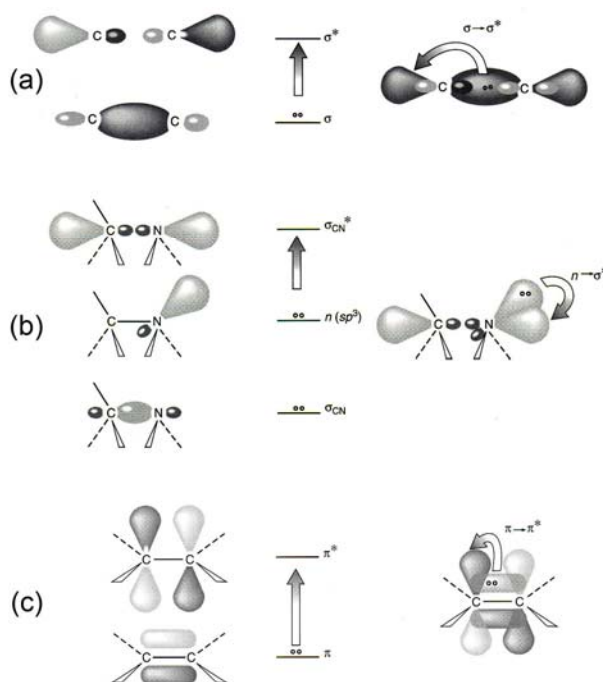


Figure 2.3. Electronic transitions from (a) $\sigma \rightarrow \sigma^*$, (b) $n \rightarrow \sigma^*$, and (c) $\pi \rightarrow \pi^*$ ⁷⁹

alkynes, that have multiple bonds and pi electrons can undergo a $\pi \rightarrow \pi^*$ transition (Figure 2.3c). The $n \rightarrow \pi^*$ transition is forbidden by selection rules, and, as a result, has a low intensity in the UV. Unsaturated molecules that contain heteroatoms such as oxygen or nitrogen, particularly in carbonyls, sometimes undergo $n \rightarrow \pi^*$ transitions. A listing of typical absorptions for some isolated chromophores can be seen in Table 2.4.

The environment around a chromophore has an effect on which wavelength will be absorbed by the chromophore. Substituent groups in place of hydrogen on a simple chromophore unit change the position and intensity of the absorption band.⁷⁹ Even if the substituents themselves do not absorb in the ultraviolet region of the spectrum, their presence will alter the absorption of the principal chromophore (Table 2.5). When a

substituent increases the intensity of the absorption, it is called an auxochrome. Common auxochromes include methyl, hydroxyl, alkoxy, halogen, and amino groups. The increase in absorption intensity is called a hyperchromic effect. The converse of this is a hypochromic effect, where the absorption intensity decreases. A bathochromic shift, also known as a red shift, is when the absorption band is moved to a longer wavelength. When the shift is to a shorter wavelength, it is termed a hypsochromic, or blue, shift.

Table 2.4. Typical absorptions of simple isolated chromophores⁷⁹

Class	Transition	λ_{max} (nm)	$\log \epsilon$
R-OH	$n \rightarrow \sigma^*$	180	2.5
R-O-R	$n \rightarrow \sigma^*$	180	3.5
R-NH ₂	$n \rightarrow \sigma^*$	190	3.5
R-SH	$n \rightarrow \sigma^*$	210	3.0
R ₂ C=CR ₂	$\pi \rightarrow \pi^*$	175	3.0
R-NO ₂	$n \rightarrow \pi^*$	271	<1.0
R-CHO	$\pi \rightarrow \pi^*$	190	2.0
	$n \rightarrow \pi^*$	290	1.0
R ₂ CO	$\pi \rightarrow \pi^*$	180	3.0
	$n \rightarrow \pi^*$	280	1.5
RCOOH	$n \rightarrow \pi^*$	205	1.5
RCOOR'	$n \rightarrow \pi^*$	205	1.5
RCONH ₂	$n \rightarrow \pi^*$	210	1.5

Table 2.5. Absorption Characteristics of Some Common Chromophores⁷⁹

Chromophore	Example	Solvent	λ_{\max} (nm)	ϵ_{\max}	Type of Transition
Alkene	$\text{C}_6\text{H}_{13}\text{CH}=\text{CH}_2$	<i>n</i> -Heptane	177	13,000	$\pi \rightarrow \pi^*$
Alkyne	$\text{C}_5\text{H}_{11}\text{C}\equiv\text{C}-\text{CH}_3$	<i>n</i> -Heptane	178	10,000	$\pi \rightarrow \pi^*$
			196	2,000	---
			225	160	---
Carbonyl	$\text{CH}_3\overset{\text{O}}{\parallel}\text{CCH}_3$	<i>n</i> -Hexane	186	1,000	$n \rightarrow \sigma^*$
			280	16	$n \rightarrow \pi^*$
	$\text{CH}_3\overset{\text{O}}{\parallel}\text{CCH}$	<i>n</i> -Hexane	180	large	$n \rightarrow \sigma^*$
			293	12	$n \rightarrow \pi^*$
Carboxyl	$\text{CH}_3\overset{\text{O}}{\parallel}\text{COH}$	Ethanol	204	41	$n \rightarrow \pi^*$
Amido	$\text{CH}_3\overset{\text{O}}{\parallel}\text{CNH}_2$	Water	214	60	$n \rightarrow \pi^*$
Azo	$\text{NH}_3\text{C}=\text{NCH}_3$	Ethanol	339	5	$n \rightarrow \pi^*$
Nitro	CH_3NO_2	Isooctane	280	22	$n \rightarrow \pi^*$
Nitroso	$\text{C}_4\text{H}_9\text{NO}$	Ethyl ether	300	100	---
			665	20	$n \rightarrow \pi^*$
Nitrate	$\text{C}_2\text{H}_5\text{ONO}_2$	Dioxane	270	12	$n \rightarrow \pi^*$

2. *Beer-Lambert Law.* The phenomenon of light being absorbed as it passed through a given sample was first published in the paper *Essai d'optique sur la gradation de la lumière* by the French mathematician and astronomer Pierre Bouguer in 1729. His work examined the quantity of light that was lost by passing it through a given extent of atmosphere. This study led to the work of both Johann Heinrich Lambert and August Beer. Lambert is credited with the discovery that the amount of light absorbed is related

to the thickness of the sample, and Beer was able to deduce that the concentration of the sample in question was also directly related to the absorbance. Their contributions resulted in the Beer-Lambert Law, also known simply as Beer's law that can be seen in equation 2.1:

$$A = \log\left(\frac{I_0}{I}\right) = \epsilon cl \quad (2.1)$$

where A is the absorbance of the sample, I_0 is the incident intensity of the light upon the sample, I is the intensity of the light leaving the sample, c is the molar concentration of the solute, l is the pathlength of the sample cell, and ϵ is the molar absorptivity of the solute. At a specific wavelength, there is a linear relationship between the absorbance, concentration of the absorbing molecules, and the path length of the cell. The concentration of unknown samples is routinely determined by generating a calibration curve of various known standards. The concentrations of the unknowns can be determined by comparing their absorbance intensities to those in the calibration curve.

The Beer-Lambert law is obeyed for samples that are of low concentrations or in dilute solutions. Deviations from the Beer-Lambert law can be seen in Figure 2.4. These deviations arise from real limitations of the law itself, chemical changes associated with concentration, known as chemical deviations, as well as the manner in which the absorbance measurements are made, instrumental deviations.⁷⁵

Because Beer's law is used to describe the absorption behavior of dilute solutions only, it can be thought of as a limiting law. At concentrations above 0.01 M, the average distance between the species that are responsible for absorption is decreased to the extent each molecule's neighbors will affect the charge distribution on the central molecule and

vice versa. As a result, this interaction can alter a species' ability to absorb a given wavelength of light. The extent of this phenomenon is concentration dependent, which

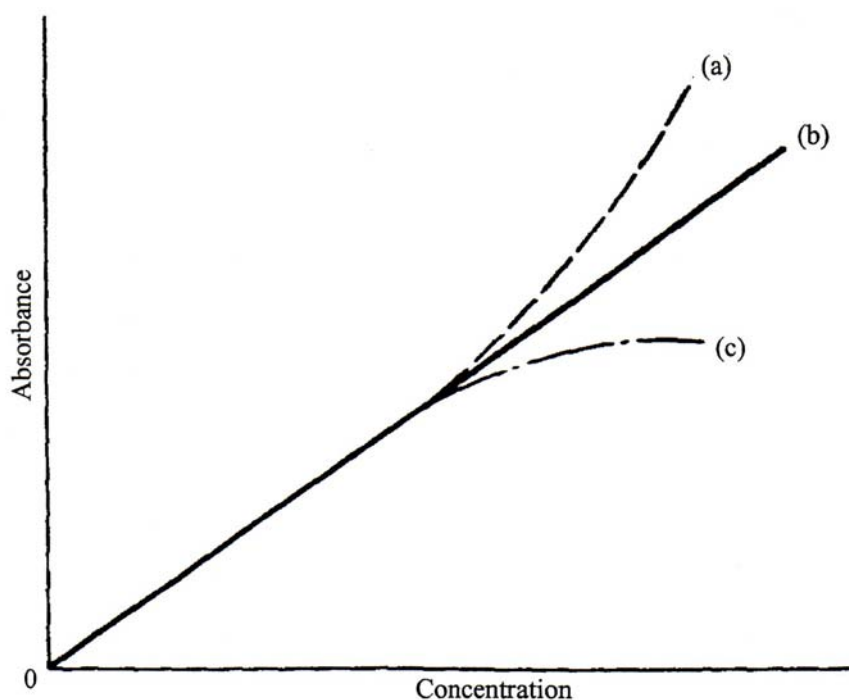


Figure 2.4. Plots of absorbance as a function of concentration for (a) a sample which exhibits a positive deviation from Beer's law, (b) a sample which obeys Beer's law, (c) a sample which exhibits a negative deviation from Beer's law (Reproduced with permission from reference 11. Copyright 2004 Sayo Fakayode)

will result in deviations from the linear relationship between absorbance and concentration. Another situation that can occur is if a sample contains a low concentration of an absorbing species and a high concentration of another species, such as an electrolyte. Electrostatic interactions between the absorber and the non-absorbing species can produce a change in the molar absorptivity of the former. The molar absorptivity is also dependent upon the refractive index of the solution the solute is dissolved in.⁷⁶ Concentration changes can lead to significant alterations in the refractive

index of a solution. It is possible to correct for this error by including the refractive index, n , in the Beer's law calculations. The corrected molar absorptivity constant can be seen in equation 2.2:

$$\varepsilon' = \frac{\varepsilon n}{(n^2 + 2)^2} \quad (2.2)$$

where ε is the standard molar absorptivity and n is the refractive index of the solvent.

Generally, correction is never very large and is rarely significant at concentrations below 0.01 M.

Deviations from Beer's law can also be seen when the analytes in a sample dissociate, associate, or react with a solvent to form a new complex that has a different absorption spectrum than the analyte itself. An example of this situation can be seen in the behavior of acid/base indicators, where a color change arises from a shift in

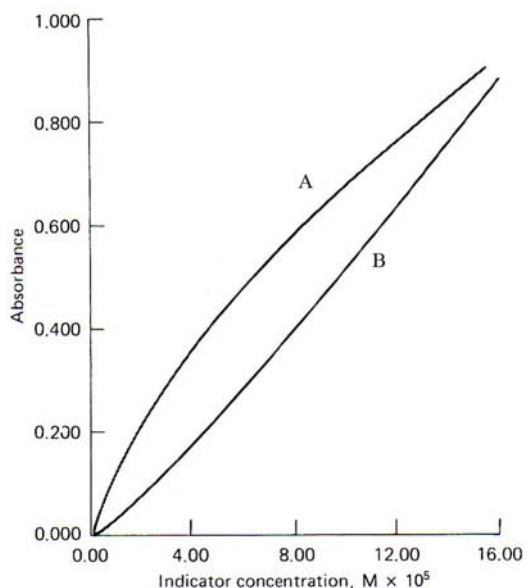
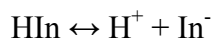


Figure 2.5. Chemical deviations from Beer's law for unbuffered solutions of indicator HIn at 430 nm (A) and 570 nm (B). (Modified from reference 75. Copyright 1992 Harcourt Brace Jovanovich)

equilibrium. This is represented by the following reaction:



where In is the indicator. The protonated indicator species has one color while the deprotonated anionic form of the indicator has a different color. The deviations from Beer's law for this scenario can be seen in Figure 2.5.

Instrumental deviations from Beer's Law can arise from polychromatic radiation and stray radiation. The use of radiation that is constrained to a single wavelength is hardly ever practical because devices that isolate portions of the output of a continuous source produce a symmetric band of wavelengths around the desired one. Stray radiation, that often times contaminates radiation exiting the monochromator, usually differs from the wavelength of the principal radiation. When stray light is present in the monochromatic radiation, the observed absorbance is given by:

$$A' = \log \frac{P_0 + P_s}{P + P_s} \quad (2.3)$$

where A' is the apparent absorbance in the presence of stray radiation, P is the incident power of the radiation, P_0 is the power of the radiation leaving the sample (without the stray radiation), P_s is the power of nonabsorbed stray radiation. In Figure 2.6, the absorbance as a function of concentration for a given sample is plotted. There is marked deviation in the linear relationship between these two factors as the relative percentage of stray radiation is increased from 0% to 5%. It should also be mentioned that at high concentrations and at longer path lengths, stray radiation can also cause a deviation in the linear relationship between absorbance and path length.⁷⁷ Both polychromatic light and stray radiation always lead to negative absorbance errors.⁷⁸

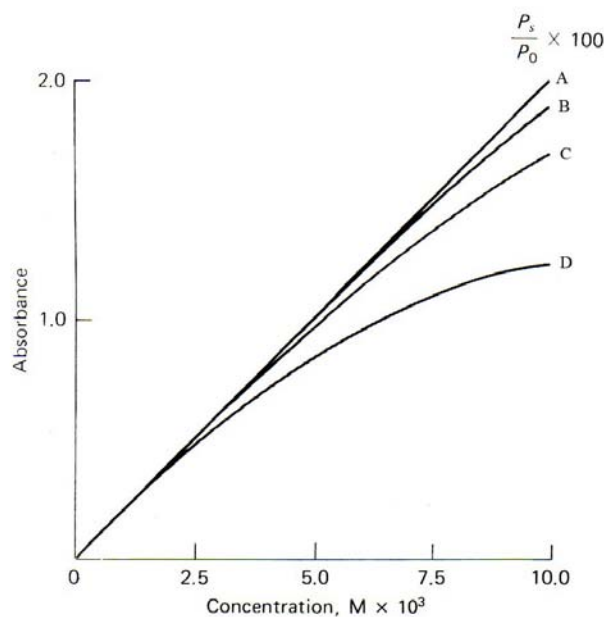


Figure 2.6. Apparent deviation from Beer's law caused by varying amounts of stray radiation, 0.0% (A), 0.2% (B), 1% (C), 5% (D) (Modified from reference 75. Copyright 1992 Harcourt Brace Jovanovich)

3. UV-Visible Spectrometer: Instrumentation. A common UV-Visible

spectrometer is comprised of a source of electromagnetic radiation, a wavelength

selector, sample containers, radiation detector, a signal processor, and read-out device.⁷⁵

Typical sources include deuterium or hydrogen discharge lamps, tungsten filament bulbs, or xenon arc lamps. Deuterium lamps are used to produce radiation in the UV. A modern deuterium lamp is operated at a low-voltage that forms an arc between a heated, oxide-coated filament and a metal electrode. They produce a continuous spectrum from 160 to 375 nm. Xenon arc lamps produce intense radiation by passing a current through an atmosphere of xenon gas. The spectrum is continuous from the UV (250 nm) to the visible (600 nm).

The monochromator is used to select wavelengths in the spectrum. A

monochromator usually employs a diffraction grating that is used to spread out the beam

into its component parts.⁷⁹ An exit slit out of the monochromator allows the light to pass through the sample. Sample cells, commonly known as cuvettes, can be made of an assortment of materials. Glass and plastic are used when the measurements taken are in the visible region of the spectrum. Because these materials absorb in the UV, a UV transparent material, quartz, is used in the cuvette.

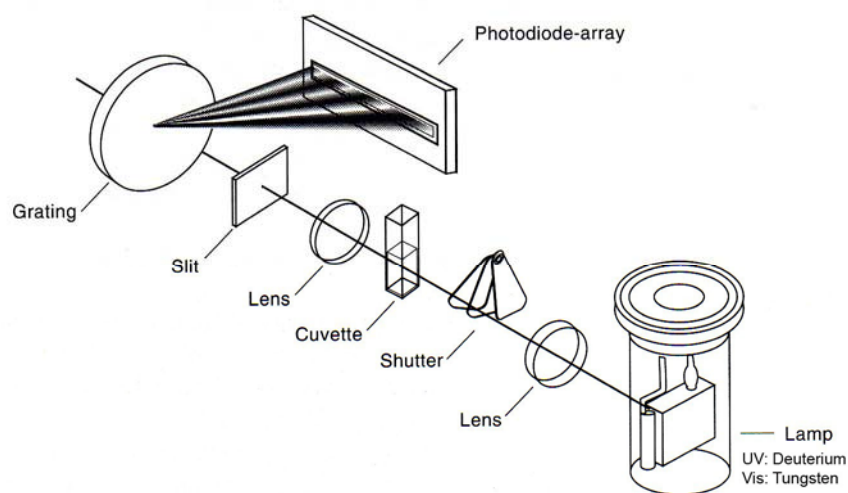


Figure 2.7. Multichannel diode array spectrometer (Modified from reference 75. Copyright 1992 Harcourt Brace Jovanovich)

The detector in a spectrometer primarily translates the electromagnetic radiation signal that has passed through the sample into an electrical signal. The intensity of the signal is proportional to the intensity of light that has reached the detector. In the past, barrier-layer cells, photo tubes, and photomultiplier tubes were used as detectors in the ultra-violet region of the spectrum. Diode array detectors are the most popular detector found in UV-Vis spectrometers today. The silicon diode detector is a reverse-biased *pn* junction on a silicon chip.⁷⁵ The reverse bias creates a depletion layer that reduces the conductance of the junction to almost zero. When radiation from the sample cell reaches

the chip, holes and electrons are formed in the depletion layer. This effect produces a current that is proportional to the radiant power of the light. Diode detectors have a spectral range from 190 to 1100 nm. Though vacuum phototubes have no moving parts and a spectrum can be collected quickly using them,^{11, 79} they are, however less sensitive than photomultiplier tubes.⁷⁵ Diode detectors can be combined into an array that will allow the entire spectrum to be analyzed simultaneously. The photodiodes are each designed to record a narrow band of the spectrum. This type of arrangement can be seen in Figure 2.7.

Diode array detectors have the advantage that they can have their output shunted directly to a computer for analysis. Other common types of recording devices include analog meters, digital meters, recorders, printers, and video displays.

Materials and Methods

Experimental

Natural capsaicin (~65% capsaicin and 35% dihydrocapsaicin), ethanol, HPLC acetonitrile, and HPLC water were purchased from Aldrich Chemical. Glacial acetic acid was purchased from Dupont. A series of calibration standards (20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 ppm) were prepared by diluting a 1025 ppm stock solution of the natural capsaicin in ethanol. These standards were subsequently used to prepare an HPLC calibration curve using the peak areas of the capsaicin and dihydrocapsaicin peaks. The calibration curve was then used to determine the amounts of the two capsaicinoids in the pepper extracts.

Habanero peppers were purchased from local area supermarkets in the Waco metropolitan area and pureed in a Black and Decker Handy Chopper Plus blender. Approximately 4.0 grams of the pepper mush was placed in a 150 ml beaker with 50 ml of 100% ethanol. The mixture was then stirred for 30 seconds. A watch glass was placed over top the beaker and the contents were heated to approximately 78 °C. If the solvent level dropped below 10 ml, more solvent was added. After boiling for 30 minutes, the samples were removed from the heat and allowed to cool to room temperature. The liquid extract was then separated from the solids using vacuum filtration through a nylon filter mat into a 250 ml filter flask. The filtered extract was then transferred to a 50 ml volumetric flask and diluted to the calibration mark. Extracts were then stored in two 25 ml glass vials (Figure 2.8).

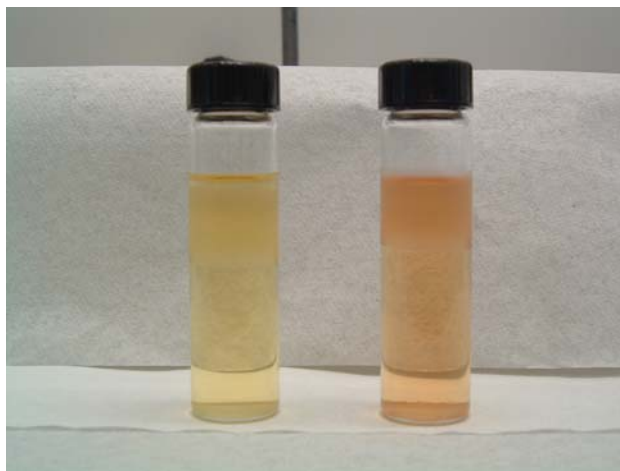


Figure 2.8. Alcoholic extracts from Habanero peppers

HPLC Analysis

The HPLC analysis was conducted using a Hewlett-Packard Liquid Chromatograph Model 1090. The chromatograph was controlled by a computer

interfaced via a Hewlett-Packard interface card (Model 82341C). Data analysis and instrument control were performed by HP Chemstation software. Sample injections were formed with the HP-1090 sample injection valve using a 5.0 μl sample loop. The column used was a Restek Corporation Pinnacle IITM (C-18) 250 x 4.6 mm with a 5 μm particle size. A 1.0 cm x 4.6 mm Hypersil guard column with 5 μm packing diameter was attached to the column. The column was maintained at a temperature of 40 °C. The signal from the Hewlett-Packard UV-Vis Diode Array Detector (Model 1050) was collected and processed by the computer using the Chemstation software. The UV detector was set to 284 nm. The flow rate through the column was 2.0 ml min^{-1} , and the total run time was 11 minutes. The solvents were degassed with helium and gradient elution was used (see Table 2.6). A period of 2-5 minutes was allowed to pass before a new sample was injected to the liquid chromatograph to allow the late-eluting homocapsaicin to come off the analytical column.

Table 2.6. Solvent Gradient for HPLC Analysis

Time (min)	Acetonitrile (%)	Acetic Acid ^a (%)
0	50	50
2	50	50
5	80	20
7	80	20
8	50	50
9	50	50

^aAcetic acid, 1% by volume in water

UV-Visible Spectroscopy

Absorbance spectra were collected with a Hewlett-Packard photodiode array UV-vis spectrometer (Model 8455) using a 2.0 mm path length quartz cell over the wavelength range from 190 to 1100 nm.

Chemometric Analysis

Partial least squares (PLS-1) regression modeling was used to develop the regression models. Because UV spectral data are relatively noise-free, no data pre-treatment was employed and the raw spectral data were input directly into the PLS software. The mean-centered spectral data were subjected to multivariate analysis using The Unscrambler© 9.1 software package (CAMO, Inc., Corvallis, OR). Partial least squares regression (PLS-1) was preformed on the spectral data using test set validation. Independently prepared validation samples were also prepared to further validate the models. The performance of the prediction models was evaluated using the root-mean-square error of prediction (RMSEP):

$$RMSEP = \sqrt{\frac{1}{I} \sum_{i=1}^I (y_i - \hat{y}_i)^2} \quad (2.4)$$

where I is the number of samples in the validation set, y is the predicted value from the regression model, and \hat{y} is the accepted value for the sample.

Results and Discussion

The ultimate goal of multivariate regression modeling is to correlate measured X-data with some desired Y-data to produce a regression model that can be used to predict future Y-values from their X-data. One reason for doing this is to attempt to substitute a simpler and perhaps faster technique (like absorption spectrophotometry) for a slower,

more expensive technique (like HPLC). In this case, the goal is to correlate the spectral data obtained for the pepper extracts with the known capsaicin concentrations determined independently by HPLC. By using actual pepper samples in the calibration phase of the modeling process, the goal is to develop a model that focuses in on those spectral components that correlate with the capsaicin content of the sample while ignoring those spectral components that arise from other matrix components like carotenoids that are not of interest. To accomplish this, it is important that the calibration set consists of real samples and that the members of the set exhibit all possible natural variations that occur in the peppers. To span all the possible variations that can occur, it is necessary to use as large a number of samples in the calibration phase as possible. It is also important to realize that a calibration set of samples cannot be prepared artificially by adding capsaicin to a pepper extract sample (such an attempt would not lead to the same types of natural variations that occur in real samples, and a model developed in this way would fail in the predictive stage).

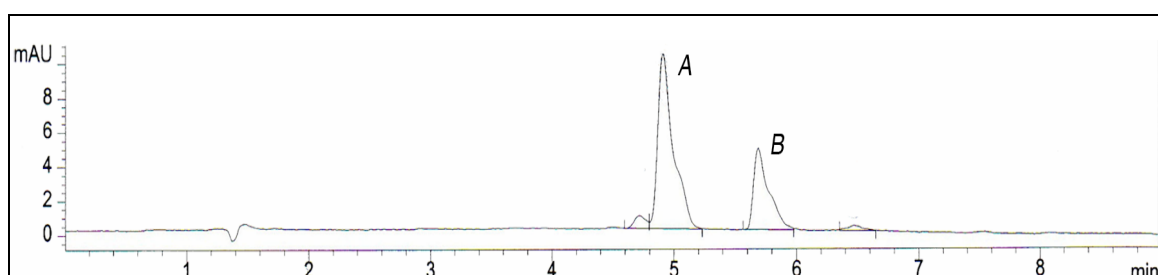


Figure 2.9. HPLC chromatogram of natural capsaicin in ethanol at 284 nm. The total concentration of capsaicin and dihydrocapsaicin for this sample was 160.0 ppm. A: capsaicin; B: dihydrocapsaicin.

The sample used to obtain the chromatogram in Figure 2.9 contained ~65% capsaicin and ~35% dihydrocapsaicin along with minor amounts of the other capsaicinoids. Standard solutions (0-200 ppm) of this commercially available capsaicin

were prepared and subsequently used to a prepare calibration curve of peak area versus the total concentration of the two major capsaicinoids (i.e., the sum of capsaicin and dihydrocapsaicin) in Figure 2.10.

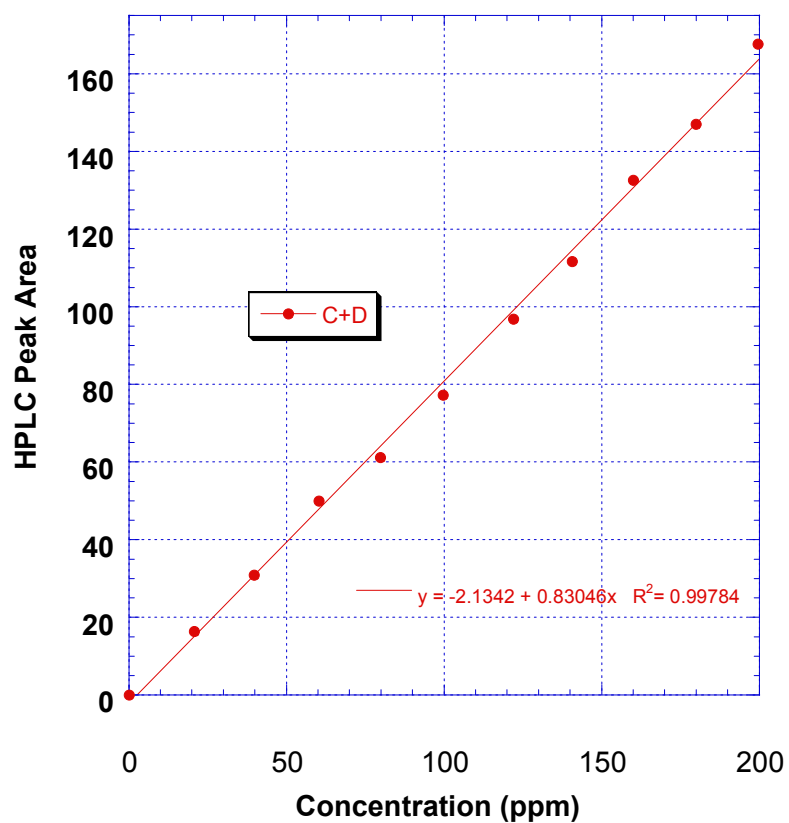


Figure 2.10. Calibration curve of total capsaicinoids, capsaicin (C) and dihydrocapsaicin (D) from HPLC peak area

This HPLC calibration curve was used subsequently to determine the reference concentrations needed to prepare the multivariate regression model relating the absorption spectral data with the capsaicin concentration of the pepper extracts.

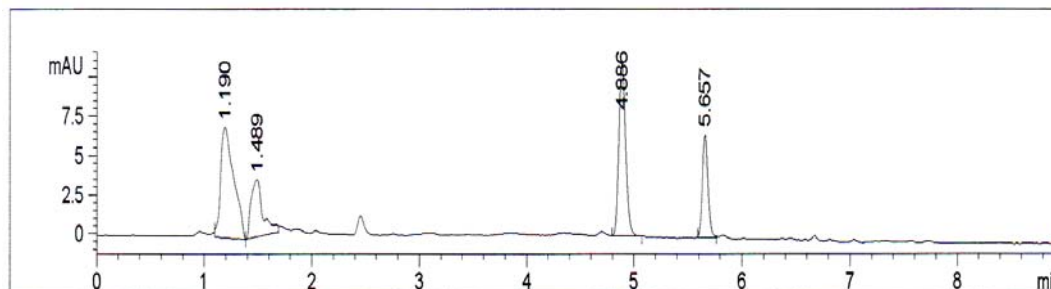


Figure 2.11. HPLC chromatogram of Habanero pepper extract in ethanol

Figure 2.11 shows a typical HPLC chromatogram obtained with a habanero pepper extract using absorbance detection at 284 nm. The peaks eluting prior to 4 minutes on this chromatogram are carotenoids and other plant materials in the sample matrix. Under these chromatographic conditions, both capsaicin and dihydrocapsaicin give strong, well-resolved peaks that can be used in conjunction with the calibration curve prepared with the commercially-available capsaicin to determine the concentration of the two major capsaicinoids in the pepper extracts (i.e., the sum of capsaicin and dihydrocapsaicin). A total of 31 habanero pepper extracts were analyzed by HPLC to determine their capsaicin levels.

Univariate Determination of Capsaicinoids

In a traditional analytical analysis using UV-visible spectroscopy, a univariate analysis would be performed where a single wavelength is selected and correlated to the concentration of the analyte of interest. This analysis is usually carried out with a pure sample having any possible interferences removed. The goal of this project was the quantitation of the pungent principles of a habanero chili pepper without any preparative separation. Capsaicin and dihydrocapsaicin are the two main components accounting for approximately 90% of the pepper's hotness.

Table. 2.7. Physical Data of Natural Capsaicinoids²²

	C ^a	DHC ^b	NDC ^c	HC ^d	HDC ^e
Formula	C ₁₉ H ₂₇ NO ₃	C ₁₈ H ₂₉ NO ₃	C ₁₇ H ₂₇ NO ₃	C ₁₉ H ₂₉ NO ₃	C ₁₉ H ₃₁ NO ₃
MW	305	307	293	319	321
MP (°C)	64.5-65.5	65.6-65.8	65.6	---	66.5
UV abs ^e	280.5 nm (ε2690)	281 nm (ε2630)	280 nm (ε3110)	282 nm (ε3180)	280.5 nm
	230.0 nm (ε8130)	228 nm (ε6390)	228 nm (ε7490)	231 nm (ε7720)	

^aCapsaicin, ^bDihydrocapsaicin, ^cNordihydrocapsaicin, ^dHomocapsaicin, ^eHomodihydrocapsaicin, ^eλ_{max} in ethanol

The peak absorbances for these components of the pepper extract are at 230 and 280 nm (Table 2.7). Figure 2.12 is a plot of absorbances at these wavelengths using a UV-visible spectrometer versus the total concentration of the two capsaicinoids, capsaicin and dihydrocapsaicin pepper extracts as determined by HPLC. In accordance with Beer's law, a linear relationship should exist between the absorbance and the concentration, assuming the path length is kept constant. By examining the data in Figure 2.12, there is a linear relationship between the absorbance and concentration of the capsaicinoids. However, the plot shows significant scattering of the data, not in keeping with linear best-fit trend lines. The correlation of the data when monitored at 230 nm had an R² value of 0.84255. The selection of 280 nm yielded a poorer correlation with an R² of 0.80586. These results demonstrate that the determination of capsaicinoids in future pepper extract samples using either a set of capsaicin standards in a traditional univariate calibration routine or real pepper extracts would be problematic and not produce reliable results. The application of multivariate regression modeling, particularly partial-least-squares (PLS-1), could provide a means to determine the total capsaicinoid concentration in a pepper extract, even in the presence of interferences such as carotenoids.

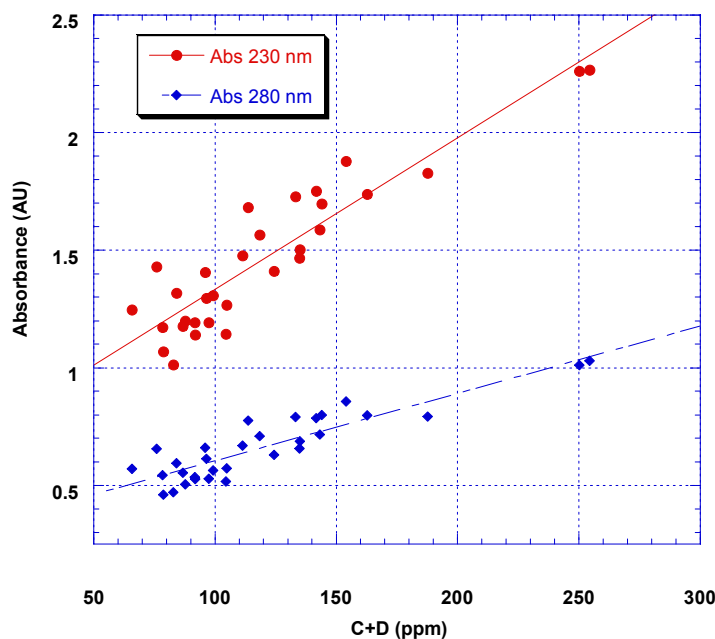


Figure 2.12. Plot of UV absorbance of pepper extracts at 230 and 280 nm versus concentration of capsaicinoids as determined by HPLC

Multivariate Determination of Capsaicinoids

In the calibration phase of regression modeling, spectral data are acquired for a calibration set of samples whose capsaicinoid concentrations are known independently from some reference analytical method (like HPLC). In this part of the process, the computer is trained to predict the capsaicinoid levels from spectral data obtained on real pepper samples that contain all the possible interfering matrix components. In selecting the calibration samples, it is important to have as much inherent variability in the samples as possible so that all the potential interferences are present in the calibration set. The more samples in the calibration set, the more likely the regression model will give good prediction results.

The PLS-1 algorithm is extremely powerful in dealing with interferences associated with the spectral data because it focuses those spectral features that correlate

with the parameter of interest while simultaneously minimizing the effect of spectral features that do not correlate with the parameter of interest⁸⁰⁻⁸². The mathematical model developed in this stage of the process takes the form of a regression vector made up of regression coefficients that are determined by the PLS-1 algorithm. This regression vector can be represented by:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (2.5)$$

where \hat{y} is the concentration of capsaicinoids predicted by the model for a given sample, the b -values are the coefficients determined by the PLS-1 algorithm during the calibration phase, and the x -values are the measured absorbances at the different wavelengths that make up the spectral range from 1 to n .

In the validation phase of the process, a second independent set of samples is collected and analyzed by the reference method to determine the analyte concentrations. The spectra of these samples are then input into Eqn. 2.5 and the concentrations of capsaicinoids are predicted. The predicted concentrations are then compared with those obtained by the reference analytical method. If the predicted concentrations agree with the values obtained by the reference analytical method, then Eqn. 1 can be used to determine the capsaicinoid levels of future samples simply from the absorption spectra of the samples¹⁵⁻¹⁷.

In this study, a total of nine regression models were prepared. These nine models were prepared by randomly selecting 21 pepper extracts from the 31 samples that were analyzed by HPLC. These 21 randomly selected samples were then used to prepare a particular regression model, and the remaining 10 samples were used to validate that

model. This process was repeated nine times with different samples selected randomly to prepare the model and validate it.

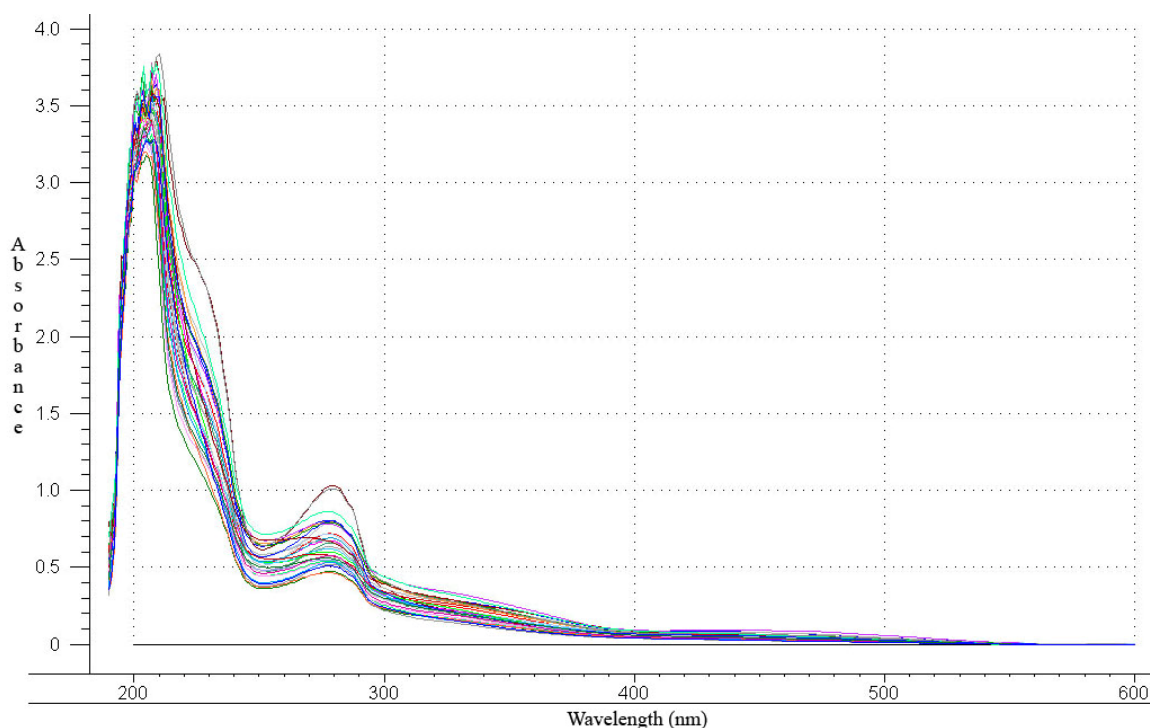


Figure 2.13. UV absorption spectra of 31 habanero peppers extracted with ethanol from 190-600 nm

The selection of wavelengths to include in a regression model to achieve the best predictive capabilities can be challenging. The complete UV spectra for the pepper extracts can be seen in Figure 2.13. As was discussed in the previous section, a classical univariate analysis of the spectral data for the determination of pepper pungency based on a single wavelength calibration curve will not provide a reliable spectrophotometric method for the determination of pepper pungency. As shown in Figure 2.13, the spectra for the extracts show two broad peaks at 230 nm and 280 nm characteristic of the absorbing components in pepper extracts. The question then becomes, how many individual wavelengths or wavelength windows are necessary to produce a model whose

ability to predict future samples is high. Since the absorbing species of interest, capsaicin and dihydrocapsaicin absorb at 230 and 280 nm, respectively, both of those wavelengths must be included in the model. The overall wavelength region initially chosen was 215-300 nm (Figure 2.14).

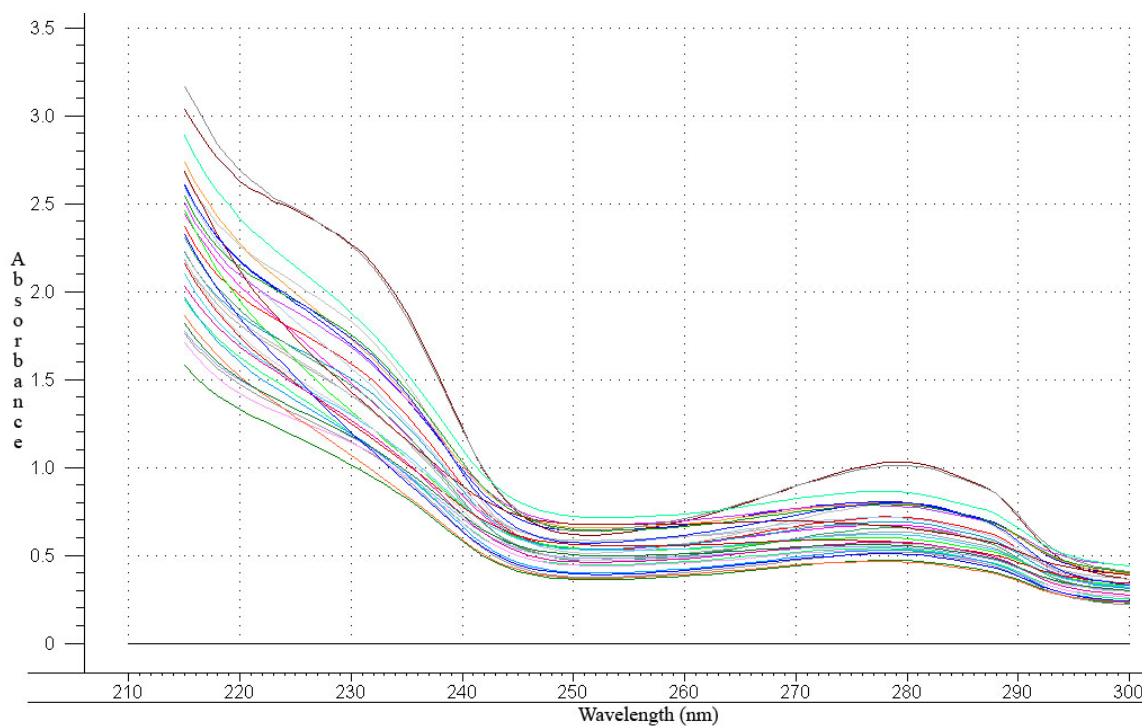


Figure 2.14. Expanded (215-300 nm) UV absorption spectra of 31 habanero peppers extracted with ethanol

A plot of the mean-centered spectral data over the wavelength range from 215-300 nm can be seen in Figure 2.15. The mean-centered plot was obtained by averaging the 31 individual spectra on a wavelength-by-wavelength basis (i.e., adding the absorbances of each spectrum on a wavelength-by-wavelength basis and dividing each sum by 31). The average spectrum was then subtracted from each individual spectrum on a wavelength-by-wavelength basis to give the mean-centered spectra. The mean-centered spectra show the regions of the pepper extract spectra that vary the most among the

different pepper samples. Because these spectral ranges where variation occurs are likely candidates for multivariate regression modeling studies, the cut off for the wavelength window should be where the variations drop back to zero.

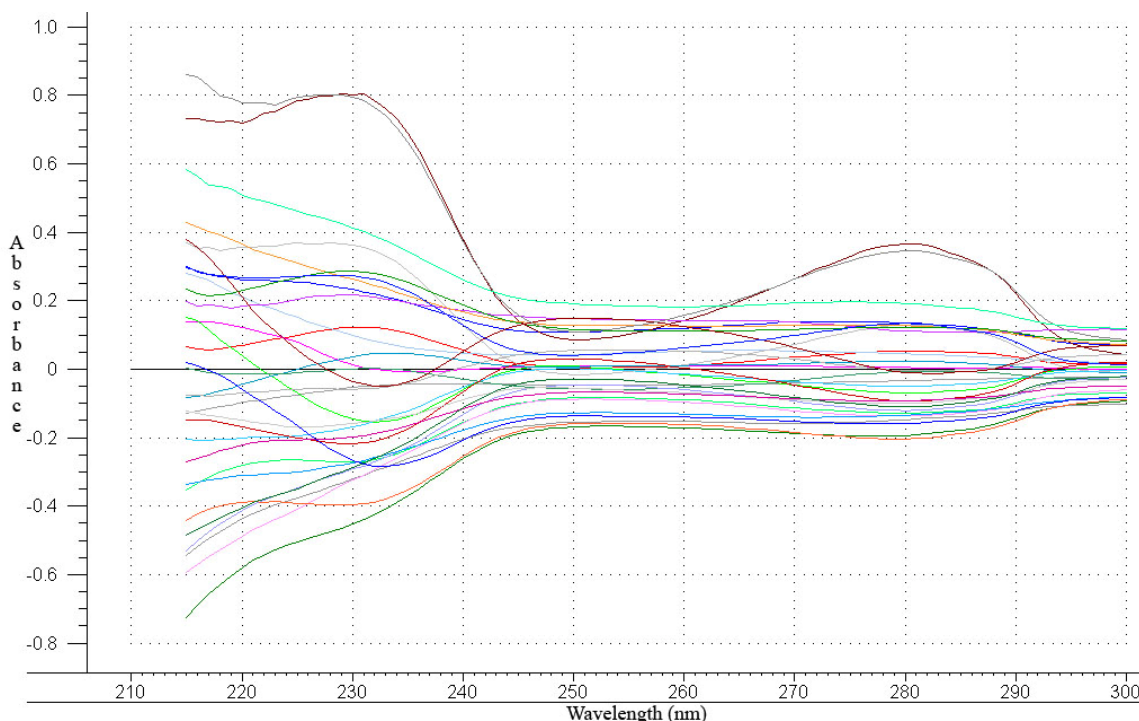


Figure 2.15. Expanded (215-300 nm), mean-centered UV absorption spectra of 31 habanero peppers extracted with ethanol

Figure 2.16 shows a summary of the results of the regression modeling for the 21 samples used to prepare model number 1. Figure 2.16a shows a scores plot for the first two PLS components in the model. From the scores plot, which shows the samples plotted on the new coordinate system made up of PLS components, there are no apparent groupings in the sample set and the first PLS component appears to be related primarily to the concentration of the capsaicinoids (it accounts for 76% of the variance in the concentration data). Figure 2.16b shows a plot of unexplained variance as a function of

the number of PLS components in the model. Figure 2.16c shows a plot of the regression coefficients that make up the model as a function of wavelength. This plot shows the values for the different b_i coefficients in Eqn. 2.5. The plot shows that over the wavelength range from 215 nm to 300 nm the coefficients vary sinusoidally with some being positive and others being negative. The maximum positive values in the plot occur at approximately 230 nm and 280 nm, which correspond to regions in the pepper extract spectra that show the most variability for the different samples (see Figures 2.16b and 2.16c). Wavelengths in these regions contribute significantly to the regression model in a positive sense. By contrast, the region around 250 nm contributes significantly to the regression model in a negative sense.

Figure 2.16d shows a plot of the capsaicin concentrations predicted by the model versus the values obtained by HPLC. For the calibration set, this linear plot has a slope of 0.994, an offset of 0.665, and correlation coefficient of 0.997. A perfect model would have a slope of 1, and offset of 0, and a correlation coefficient of 1. It should be stressed that Figure 2.16d is not a calibration curve.

While the above model parameters look quite good, the real test of any regression model is its ability to predict future samples correctly. This ability is evaluated in the validation phase of regression modeling. In this part of the study, the spectra of the 10 samples that were not used to develop a given model were input into the model and the

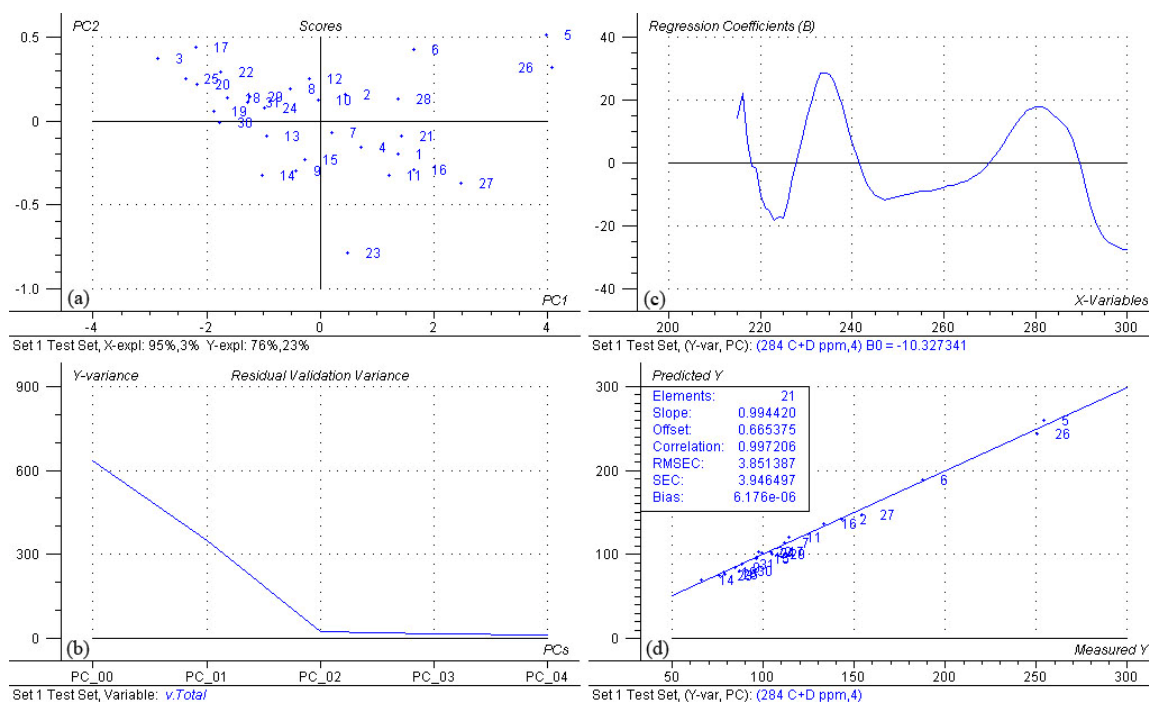


Figure 2.16. Summary of regression results for total capsaicinoids concentration: (a) scores plot; (b) residual variance as a function of the number of PCs; (c) regression coefficients as a function of wavelength; (d) plot of concentration of total capsaicinoids predicted by the model versus the known values

capsaicin levels were predicted using the model and the spectral data. Table 2.8 gives the results obtained for this phase of the study for model number 1. In this table, the total capsaicin concentration (capsaicin plus dihydrocapsaicin) predicted by the model is given along with the corresponding value determined by HPLC. From this data, the errors for the individual validation samples were calculated along with the RMSEP for all 10 samples.

In developing a regression model, selecting the proper number of PLS components to use is an important consideration. Using too few PLS components will result in sub-optimal predictive capability, while using too many PLS components will result in over fitting the model. By using too many PLS components, one risks incorporating noise into the model, which has no predictive capability. As a result, use of

too many PLS components will also reduce the predictive capability of the model. The maximum number of PLS components possible in a model is given by the smaller of: a) the number of variables in the model or b) the number of samples in the model minus one. In this study, there were 86 variables (wavelengths) and 21 samples. As a result, 20 PLS components would be possible with this set of data.

In examining Figure 2.16a, it can be seen that two PLS components account for 98% of the variance in the spectral data and 99% of the variance in the concentration data. Figure 2.16b also indicates that 2 PLS components account for essentially all the variance in the data. This means that 2 PLS components should be all that are needed in the model. However, it is necessary to validate the number of PLS components selected to explain the spectral data.

To determine the optimum number of PLS components to use in the model, the predictive abilities of models with two and four PLS components were studied with the 9 models developed as described above. Figure 2.17 shows the results of the study in terms of a bar graph. It is clear that in almost all cases, models with four PLS components had better predictive capabilities (lower RMSEP) than models with only two PLS components. Only for models 7 and 9 were the RMSEP values almost the same or better for two PLS components as opposed to four. Accordingly, four PLS components were used for the models in this study.

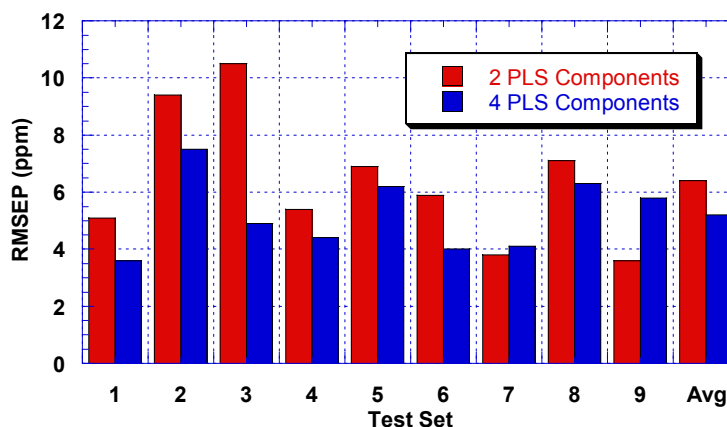


Figure 2.17. Comparison of RMSEP of 9 test sets with 2 and 4 PLS components

Table 2.9 summarizes the results obtained for the nine regression models prepared in this study. The RMSEP values range from 4 to 8 ppm. The average RMSEP for all nine models made with the pepper extracts was 5 ppm. This is quite good considering that only 21 samples were used in the calibration phase of the regression modeling. Smaller error values would be expected if a larger calibration set were employed.

With the optimal number of PLS components selected, the final step to optimize the model is to determine the proper wavelength window to be used in the modeling. As a logical starting point, the peak absorbances of the capsaicinoids will be used. Varying wavelength windows about the peak absorbances were selected, models were made, and the concentration of total capsaicinoids was predicted. The results of these predictions can be seen in Figure 2.18.

The average RMSEP for 228-232 and 278-282 nm window was 9 ppm. This is higher than the average of 5 ppm given by the 215-300 nm window used in the bulk of this study. The higher error value is to be expected. The univariate method described

Table 2.8. Prediction Results for the ten samples in Test Set 1^a

Total (ppm) ^b	Predicted (ppm) ^c	Error (ppm) ^d
144	138	-6
82.7	78.7	-4
118	124	6
124	121	-3
135	132	-3
135	135	0
91.8	88.4	-4
91.9	90.5	-1
142	144	2
163	167	4
		4 ^e

^aFour PLS components were used in this model, ^bTotal concentration of capsaicin and dihydrocapsaicin as determined by HPLC at 284 nm, ^cTotal concentration of capsaicin and dihydrocapsaicin as predicted by the regression model, ^dabsolute error, ^eRoot-mean-square error of prediction

Table 2.9. Summary of slope, offset, correlation coefficient, and RMSEP for 9 test sets

Model ^a	Cal/Val ^b	Slope	Offset	Correlation	RMSEP ^c
Set 1	Cal	0.994	0.665	0.997	4
	Val	1.05	-7.50	0.992	
Set 2	Cal	0.997	0.390	0.998	8
	Val	1.07	-4.73	0.981	
Set 3	Cal	0.994	0.733	0.997	5
	Val	1.03	-3.51	0.991	
Set 4	Cal	0.995	0.632	0.997	4
	Val	1.04	-4.56	0.990	
Set 5	Cal	0.995	0.636	0.997	6
	Val	0.930	6.16	0.995	
Set 6	Cal	0.994	0.777	0.997	4
	Val	1.03	-3.48	0.994	
Set 7	Cal	0.994	0.725	0.997	4
	Val	0.991	2.75	0.994	
Set 8	Cal	0.996	0.458	0.998	6
	Val	0.987	0.845	0.984	
Set 9	Cal	0.994	0.661	0.997	6
	Val	1.06	-7.44	0.995	

^aFour PLS components were used in all the models, ^bCalibration (Cal)/Validation (Val), ^cRoot-mean-square error of prediction (ppm)

previously shows the departure in linearity from Beer's Law. The reduced number of wavelengths in the multivariate analysis is removing information that is crucial to the prediction of total capsaicinoid concentration in the pepper samples. The RMSEP values show a decreasing trend as the model window increases in the number of wavelengths used. The error values for the window of 210-250 and 260-300 nm were comparable to those observed at using the 215-300 nm window. The RMSEP for the shorter wavelength range was slightly higher at 6 ppm. Though the difference is only 1 ppm, it demonstrates that even small changes in the window can cause deviations in the predictions from the models.

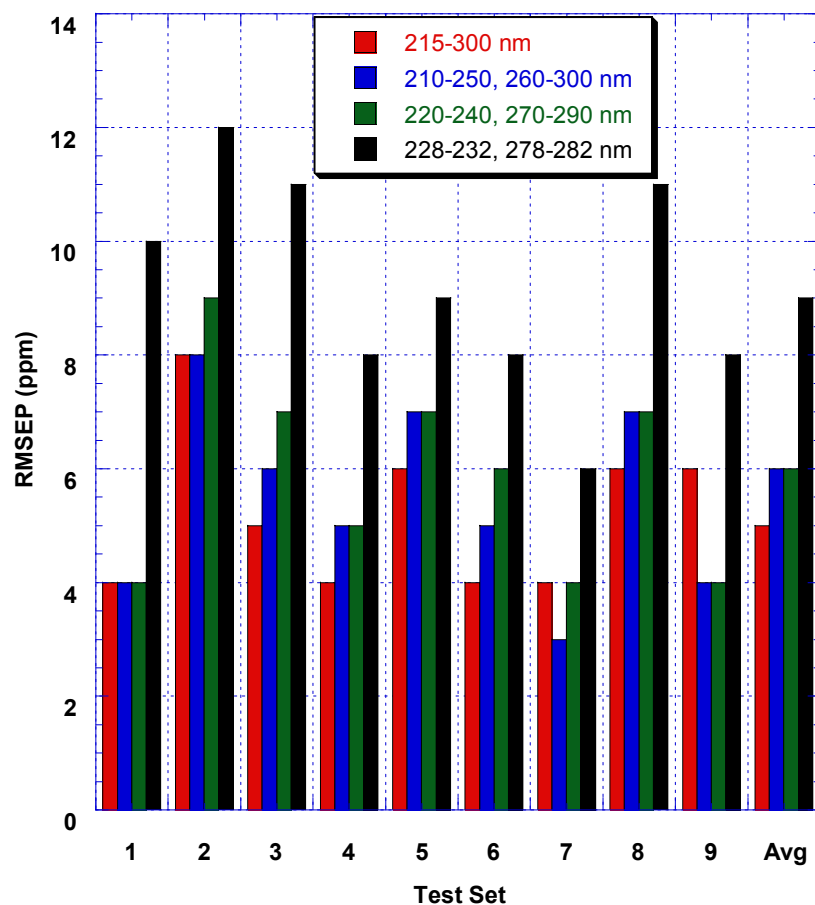


Figure 2.18. RMSEP for nine test sets at various wavelength windows

Independent Validation Study

To validate this optimized spectrophotometric procedure, a series of 12 fresh pepper samples was prepared and analyzed with a regression model that had been prepared over 5 months prior to this validation study. The UV spectra used in the calculations can be seen in Figure 2.19. Figure 2.20 shows the regression model, which used the original 31 pepper extracts to predict the 12 validation samples using 4 PLS components from 215-300 nm. Table 2.10 shows the prediction results obtained for these pepper samples analyzed after a two week equilibration period. The RMSEP obtained for these samples was 4 ppm, suggesting that the regression model prepared from the UV spectral data is quite stable over time.

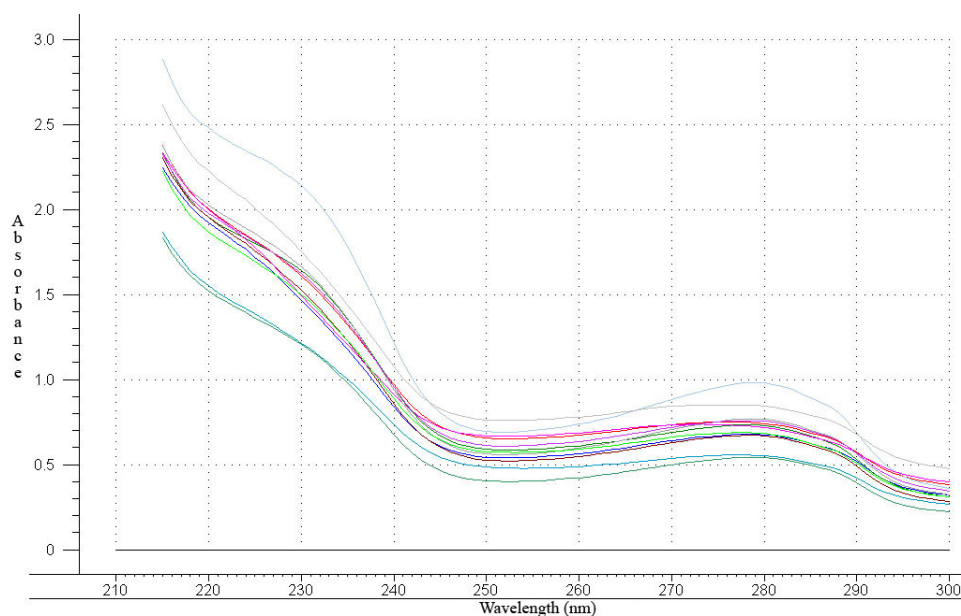


Figure 2.19. UV absorbance spectra for 12 pepper extracts used for independent validation of the regression model prepared five months earlier

The 12 pepper samples were then stored in the dark at room temperature for a period of 25 weeks. During this time, the UV lamp in the spectrophotometer failed and

was replaced. The 12 samples were taken out of storage and their UV spectra were taken. Table 2.11 shows the prediction results obtained for these samples. The RMSEP obtained was 5 ppm, which demonstrates that the model is stable over time, as well as when instrument adjustments occur.

A principal components analysis was conducted on the UV spectral data of the validation samples taken at two weeks and 27 weeks. The three dimensional scores plot seen in Figure 2.21 shows, graphically, that the samples show a high degree of correlation with one another even after a five month storage with no tendency to segregate into groups.

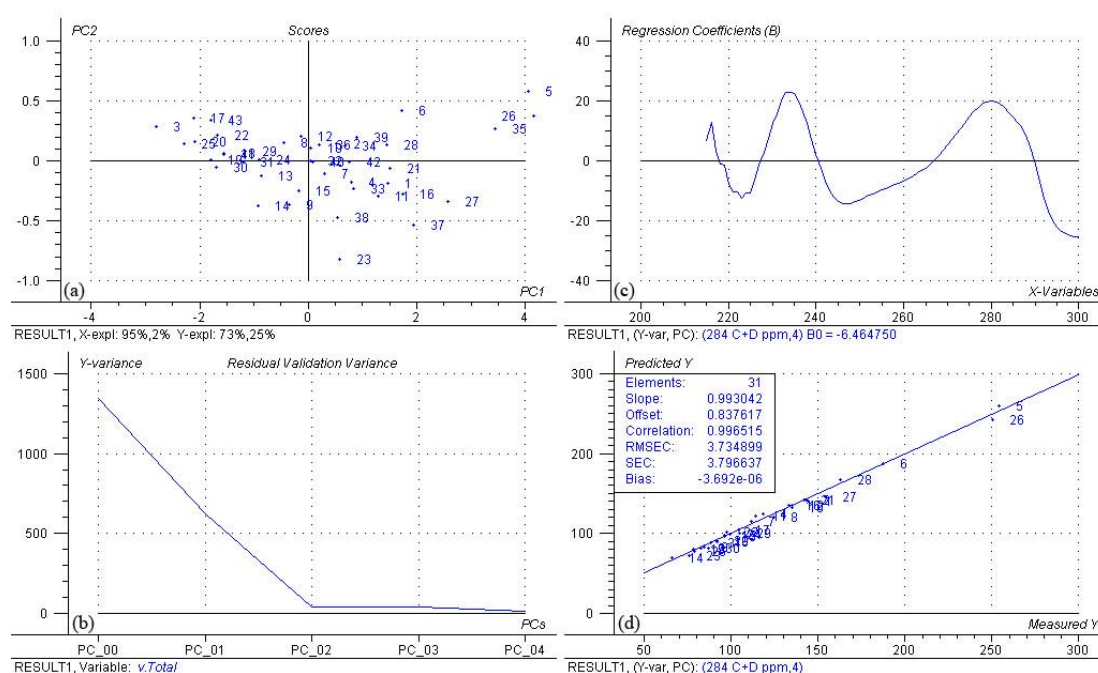


Figure 2.20. Summary of regression results for total capsaicinoids concentration in independent validation series: (a) scores plot; (b) residual variance as a function of the number of PCs; (c) regression coefficients as a function of wavelength; (d) plot of concentration of total capsaicinoids predicted by the model versus the known values

Table 2.10. Prediction Results for the 12 samples^{a,b}

Total (ppm) ^c	Predicted (ppm) ^d	Error (ppm) ^e
114	116	2
112	119	7
144	147	3
229	223	-6
130	131	1
115	116	1
89	90	1
152	156	4
119	123	4
90	92	2
146	138	-8
113	111	-2
		4 ^f

^aFour PLS components were used, ^bCalibration set used all 31 samples, ^cTotal concentration of capsaicin and dihydrocapsaicin as determined by HPLC at 284 nm,

^dTotal concentration of capsaicin and dihydrocapsaicin as predicted by the regression model, ^eabsolute error, ^fRoot-mean-square error of prediction

Table 2.11. Prediction of 12 samples after light source replacement^{a,b}

Total (ppm) ^c	Predicted (ppm) ^d	Error (ppm) ^e
118	115	-3
115	117	2
142	145	3
230	220	-10
138	129	-9
121	115	-6
88	88	0
154	154	0
123	120	-3
93	90	-3
137	136	-1
111	108	-3
		5 ^f

^aFour PLS components were used, ^bCalibration set used all 31 samples, ^cTotal concentration of capsaicin and dihydrocapsaicin as determined by HPLC at 284 nm, ^dTotal concentration of capsaicin and dihydrocapsaicin as predicted by the regression model, ^eabsolute error, ^fRoot-mean-square error of prediction

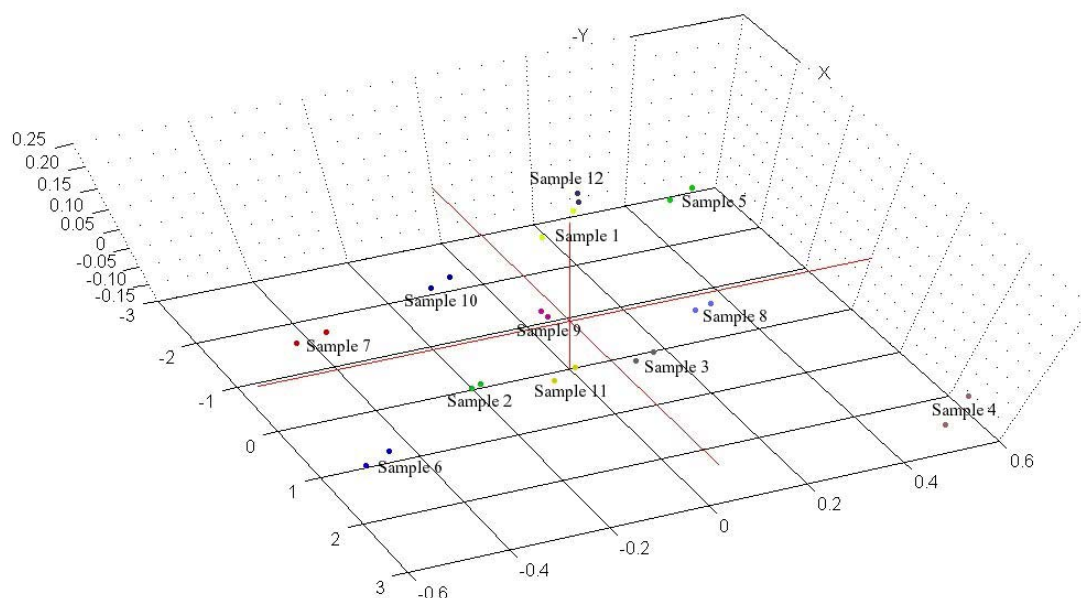


Figure 2.21. Three dimensional scores plot of independent validation set at 2 and 27 weeks

Determination of Individual Capsaicinoids

Commercially available capsaicin, available from retailers such as Aldrich, Alfa Aesar, and Fluka, contains approximately 65% capsaicin and 35% dihydrocapsaicin, along with several other naturally occurring capsaicinoids in miniscule quantities. As mentioned above, capsaicin and dihydrocapsaicin constitute 80-90% of the total pungency of a chili pepper. Throughout this study, the total capsaicin concentration was always taken as the sum of the capsaicin and dihydrocapsaicin. It is possible to determine an approximation of the two constituents from the total capsaicinoid data as well.

Since the total concentration for the two components is the concentration of the stock solutions prepared for the initial calibration curve, it is merely a matter of determining the relative percent of capsaicin and dihydrocapsaicin present by looking at the peak areas on the chromatograms for the calibration standards. For this study, the

residual capsaicinoids, though present in the samples, will be ignored, due to their negligible concentrations. The estimations for the concentrations of capsaicin and dihydrocapsaicin can be seen in Table 2.12. The average values for the percentages of capsaicin and dihydrocapsaicin are 69% and 31%, which are slightly different than the values of 65% and 35% reported previously by the manufacturer.

Table 2.12. Estimated capsaicin and dihydrocapsaicin concentrations in commercial capsaicin

C Peak Area	DHC Peak Area	C+DHC Peak Area	C (ppm)	DHC (ppm)	C+DHC (ppm)	%C	%DHC
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11.3	5.10	16.4	14.2	6.44	20.7	68.9	31.1
21.4	9.54	30.9	27.4	12.3	39.7	39.1	30.9
34.7	15.3	50.0	41.7	18.4	60.2	69.4	30.6
42.0	19.2	61.2	54.7	25.1	79.8	68.6	31.5
53.2	24.1	77.2	68.6	31.0	99.6	68.9	31.2
66.6	30.2	96.9	83.9	38.1	122	68.8	31.2
76.6	35.1	112	96.4	44.2	141	68.6	31.4
91.2	41.4	133	110	50.0	160	68.8	31.2
102	45.0	147	125	55.0	180	69.4	30.6
117	51.0	168	139	60.7	199	69.6	30.4
Average						69.0	31.0

With this concentration data, it was possible to build calibration curves for both capsaicin and dihydrocapsaicin in Figures 2.22 and 2.24. The concentrations of these two components in the alcoholic pepper extracts was then modeled based on the UV spectral data already collected. The same nine test set groupings used previously, with a wavelength window from 215-300 nm, were used as calibration and validation sets to

determine the ability of the model to quantify individual pungent principles instead of their total. Summaries of the regression models can be seen in Figure 2.23 and 2.25.

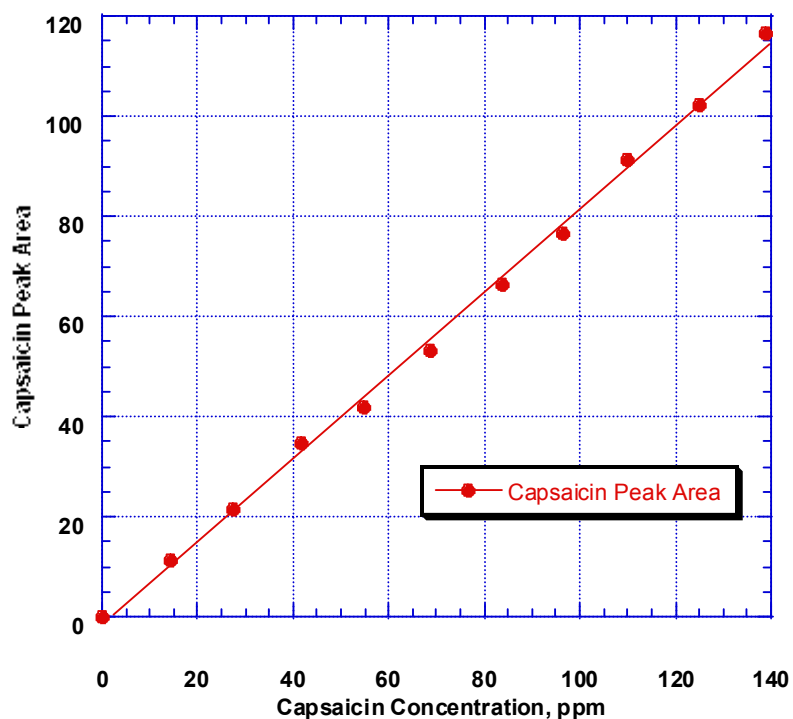


Figure 2.22. Calibration curve for estimated capsaicin concentration

The predicted concentrations for capsaicin had RMS error or prediction of 4 to 7 ppm and had an average value of 5 ppm. This is comparable to the values of total capsaicinoids determined earlier. The prediction of dihydrocapsaicin had RMSEP values of between 2 and 3 ppm for the test sets and an average value of 2 ppm. The full summary of the error values can be seen in Figure 2.26.

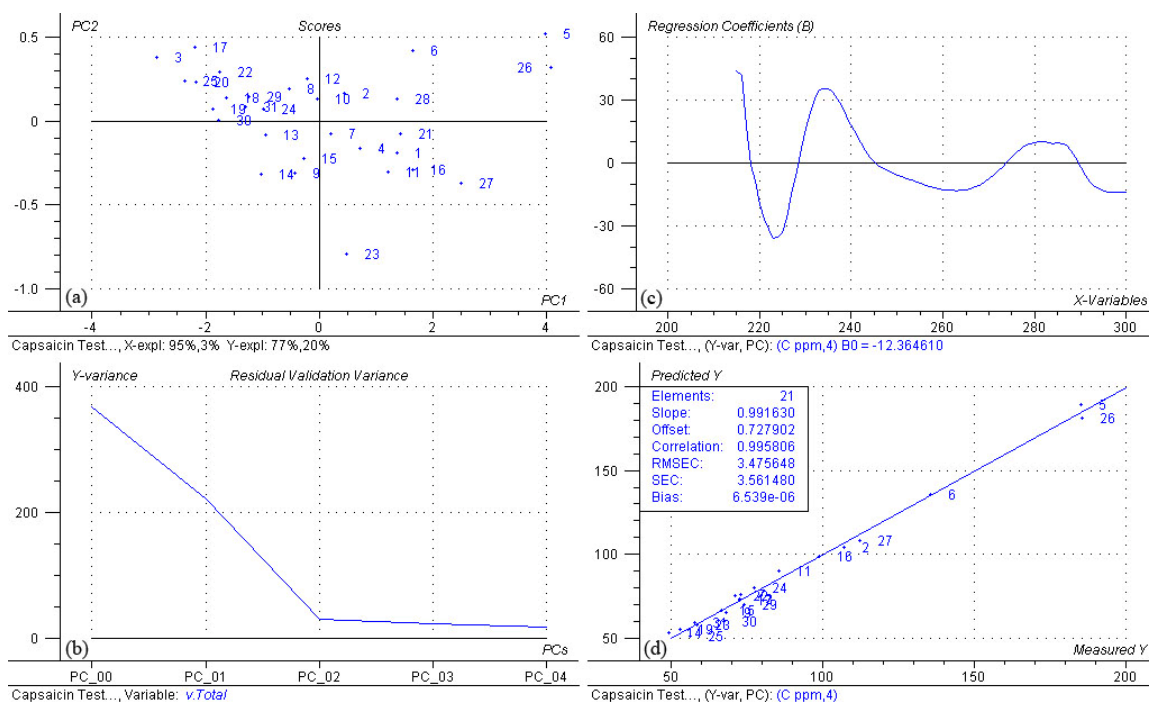


Figure 2.23. Summary of regression results for estimated capsaicin concentration: (a) scores plot; (b) residual variance as a function of the number of PCs; (c) regression coefficients as a function of wavelength; (d) plot of concentration of capsaicin predicted by the model versus the estimated values

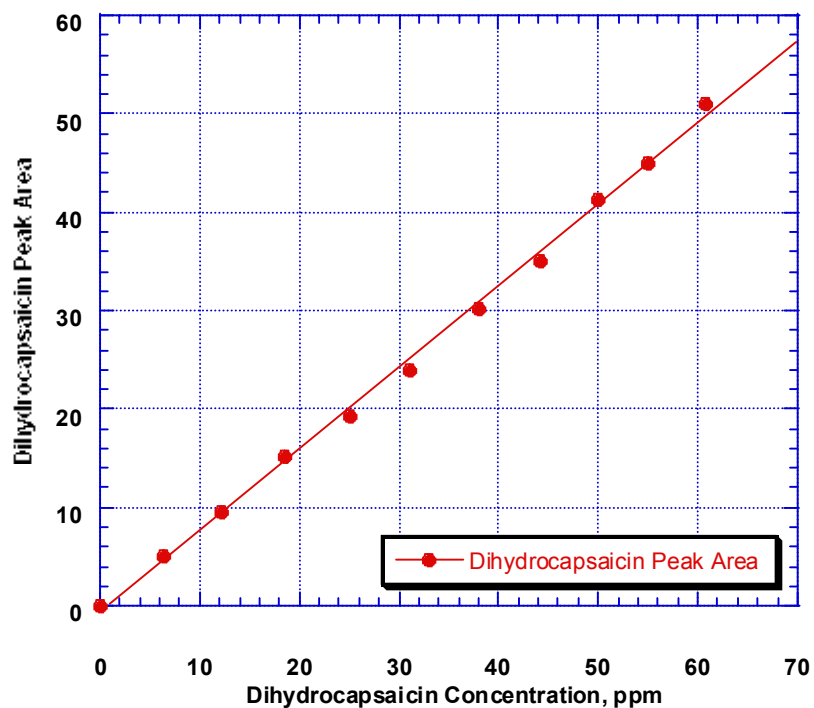


Figure 2.24. Calibration curve for estimated dihydrocapsaicin concentration

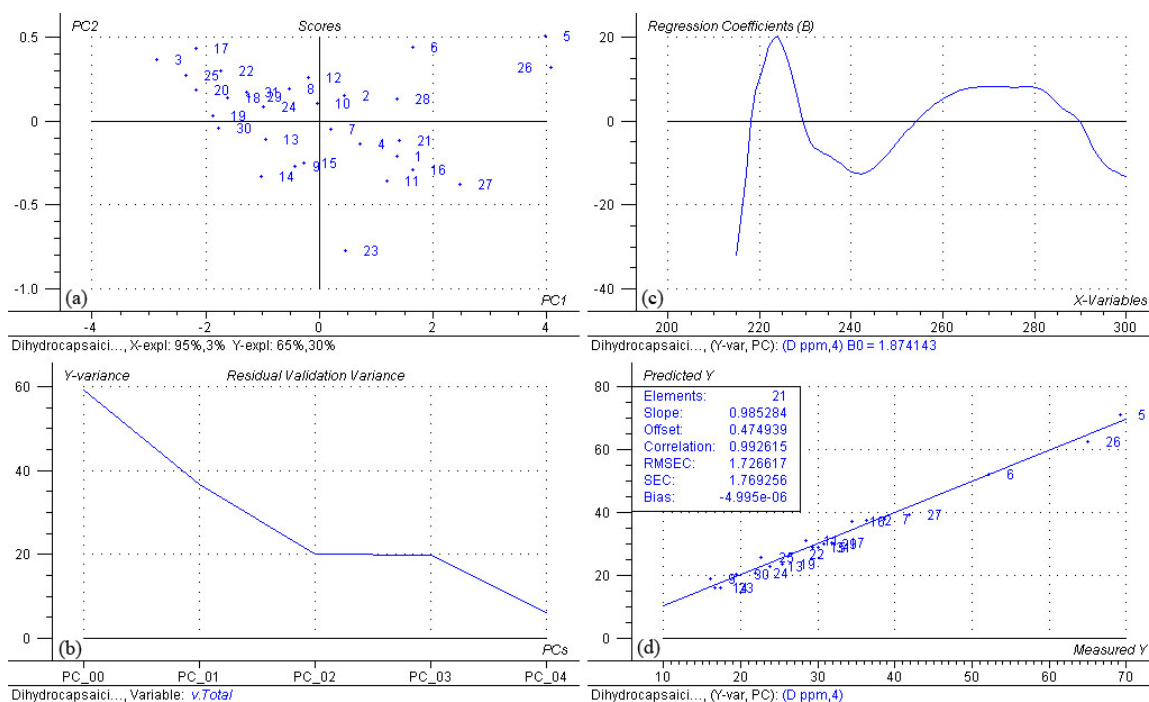


Figure 2.25. Summary of regression results for estimated dihydrocapsaicin concentration: (a) scores plot; (b) residual variance as a function of the number of PCs; (c) regression coefficients as a function of wavelength; (d) plot of concentration of dihydrocapsaicin predicted by the model versus the estimated values

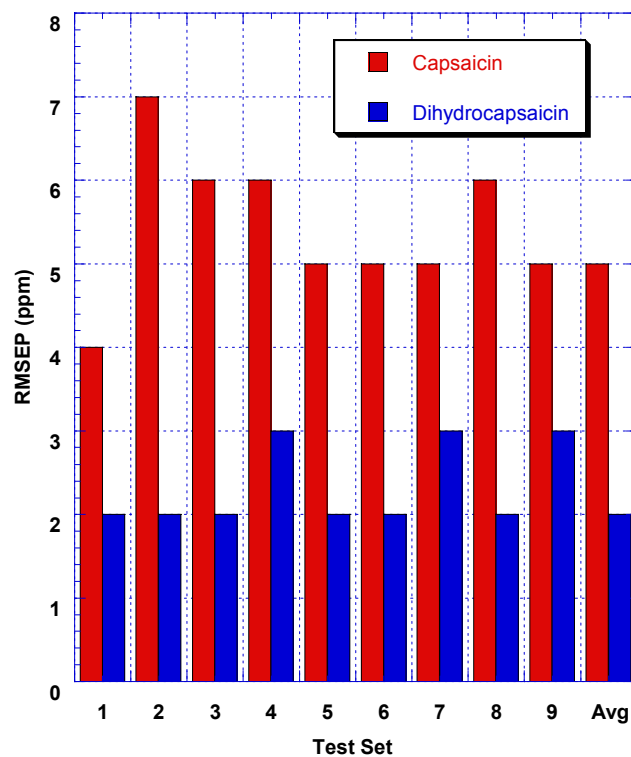


Figure 2.26. Root mean square error of prediction results for individual capsaicinoids in nine test sets

Conclusion

The motivation in conducting this study was to show that multivariate regression modeling of spectral data could be used as a means of developing regression models that could be used to predict pepper pungency from ordinary absorption spectrophotometric data. We chose to predict the sum of the capsaicin and dihydrocapsaicin levels because they are the major capsaicinoids present that primarily determine the pepper pungency. If one wanted, however, similar models could be made individually for capsaicin and dihydrocapsaicin separately.

We selected habanero peppers for this study because they are one of the hotter commonly available peppers with relatively high levels of capsaicinoids. Because they have high capsaicinoid concentrations, these pepper extracts were more compatible with our diode-array absorption detection systems than less pungent peppers that would require fluorescence detection. The smaller 4.0 g sample size was used due to the fact that the models developed were aimed at predicting the total capsaicinoid concentration of a single pepper and not the relative hotness of a certain type of pepper as was seen in previous research.²⁰

This study has shown that reasonable multivariate regression models can be prepared from ordinary absorption spectrophotometric data obtained on pepper extract samples using HPLC as a reference method to determine the actual capsaicinoid content of the pepper extracts. For the habanero peppers used in this study, an average predictive error of about ± 5 ppm was observed for regression models prepared with 21 samples. The average error for the NIR solid support studies previously conducted were between

41.71 and 41.91 ppm using the Kubelka-Munk function and $\log(I/R)$. The previous study conducted in the UV showed relative error values ranging from 18 to 34%.

It should be realized that regression modeling is an ongoing process and models can be updated periodically with new samples. As this is done, the predictive error will decrease.

CHAPTER THREE

Introduction to Textiles

Background

Textiles are among the most ubiquitous materials in society and have been important to mankind since before recorded history. They have served to protect people from the elements, used in the construction of homes and shelters, and provided for decorations like wall coverings and objects such as baskets and furniture. Other uses of textiles include tire reinforcement, tenting, filter media, conveyor belts, and insulation.⁸³ Even though the term textiles has become synonymous with household products and apparel, the use of industrial fibers is of ever increasing importance.

The term textile is derived from the Latin *textilis* meaning woven.⁸⁴ Over time, the word has taken on broad connotations, that include: (1) staple filaments and fibers for use in yarns or preparation of woven, knitted, tufted, or non-woven fabrics, (2) yarns made from natural or man-made fibers, (3) fabrics and other products made from fibers or yarns, and (4) apparel or other articles made from the above which retain the flexibility and drape of the original fabrics.⁸⁵

Little regard is given as to the many fabrics that are around us everyday and their origin. Though many natural materials are fibrous in nature, not all of them are suitable to be made into a fabric. Corn silk and wood slivers are prime examples of fibrous materials that do not have correct properties to be used in making textiles. Fibers must be

thin, flexible, long, cohesive, and strong.⁸⁴ Of course, each of these factors must be modified in various degrees in order to produce a fabric with the desired characteristics.

The remaining chapters in this dissertation will discuss the use of chemometric multivariate analysis of diffuse near-infrared spectra collected from a variety of commercially available textiles. With the countless types and variations of fibers available in today's market, it is necessary to provide some background outlining general information on textile fibers. The current chapter will present relative information pertaining to the taxonomy, historical background, chemical and physical characteristics of fibers, and various analytical methods of textile identification.

Fiber and Textile Taxonomy

Yarns and fibers are interlaced or entangled in a specific manner to produce the planar structure commonly referred to as a textile fabric. A diagram of this arrangement can be seen in Figure 3.1. The individual yarns are continuous strands made up of textile fibers. The structural arrangement and orientation of the molecules that make up the fibers will dictate the fibers' properties.

The broad spectrum of textile fibers can be classified into two broad categories based on the chemical makeup: natural fibers and man-made or manufactured fibers. A general flowchart for the divisions can be seen in Figure 3.2. Fibers in a textile may be either staple or filament. Staple fibers are relatively short, measuring only a maximum of a few inches in length. Filament fibers are relatively long and can reach up to yards in length.⁸⁶

The fibers that come from natural sources and do not require formation or reformation are classed as natural fibers.⁸⁵ Natural fibers that come from plant or

vegetable origins are composed of cellulose (Figure 3.3), which is a polymeric substance made from 1,4- β -anhydroglucose subunits, connected by β -ether linkages,⁸⁵ bound to lignin and associated with varying amounts of other natural materials.⁸³

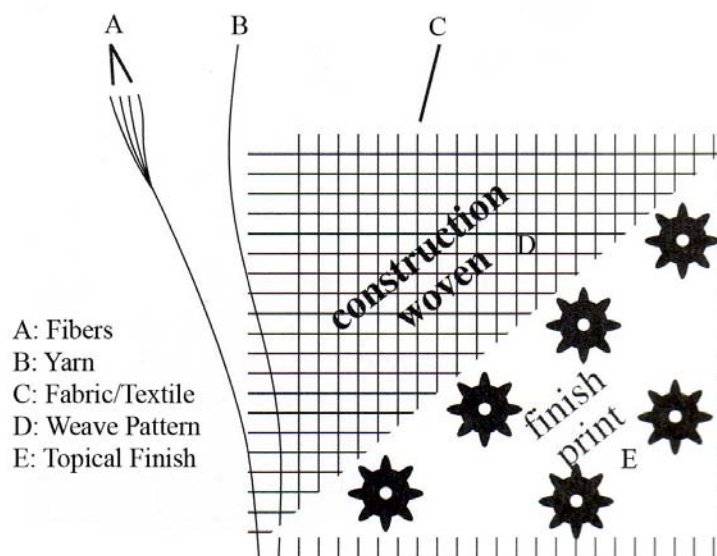


Figure 3.1. Construction breakdown for a textile (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

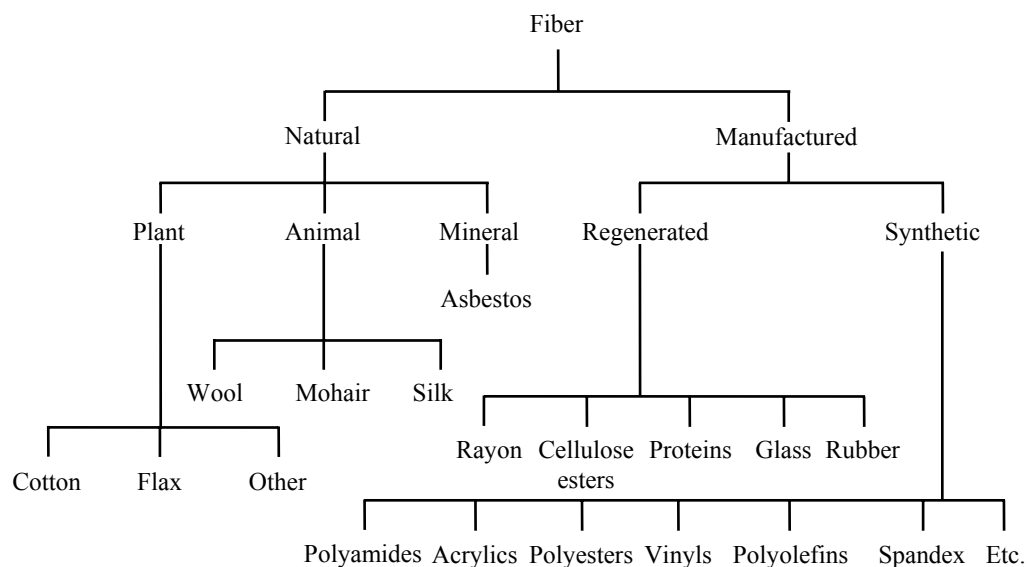


Figure 3.2. General taxonomy breakdown for fiber classes (Adapted from reference 85. Copyright 1986 Noyes publications)

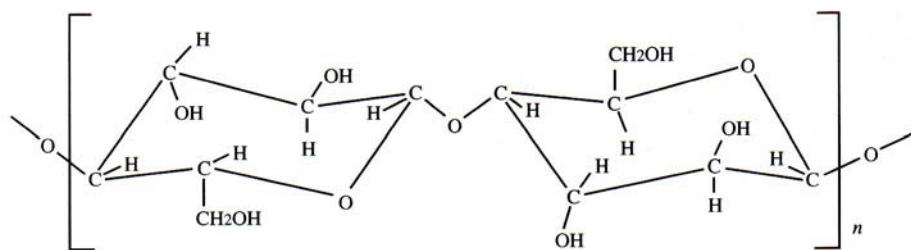


Figure 3.3. Cellulose molecule comprised of two β -D-glucose residues (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

The number of cellulosic repeating monomers can be less than 1,000 and up to as many as 18,000, depending on the fiber source. Being a hemiacetal, cellulose is hydrolyzed in dilute acid solutions to form the monosaccharide glucose. Primary and secondary hydroxyl functional groups are the predominant reactive sites in cellulose with the primary moieties being the more reactive.

Plant-based natural fibers, also called cellulosic fibers, are further segregated by which part of the plant the fiber is derived from.⁸³ Leaf fibers are obtained from the leaves of monocotyledonous plants, from the fibro-vascular systems, mostly from tropical regions. Examples of commercially important leaf-based fibers include abaca, sisal, and henequen, which are used for tasks such as cordage due to their hard and robust nature. Bast fibers are procured from the bast tissue or bark of a plant's stem. Common examples of bast fibers include flax, hemp, jute and ramie, also called soft-fibers that are converted into textiles, thread, yarn, and twine. The multi-celled fibers of this category are long and can readily be split into finer cells. Seed hair fibers, such as cotton, kapok, and the flosses, are obtained from seeds, seedpods, and the inner walls of the fruit. The fibers from this grouping are short and single celled. The final grouping of plant-based natural fibers is that of those from more obscure sources. They include fibers obtained

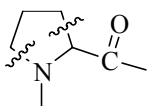
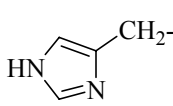
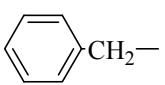
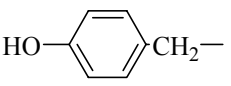
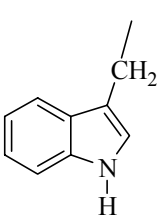
from sheathing-leaf stalks of palms, stem segments, and fibrous husks like piassava and coir. These fibers are often used as broom bristles, matting, and stuffing because of their straw-like, woody, and coarse nature.

Natural fibers can also be based on proteins and fall under the category of animal fibers. Protein fibers are formed naturally from the condensation of α -amino acids to form polyamide units with various substituents on the α -carbons.⁸⁵ The various protein units can be seen in Table 3.1. The sequence and type of amino acid residues in the condensed polymer chain will dictate the overall properties of the resultant fiber. Protein fibers can be grouped into two subclasses: (1) keratin, derived from hair or fur, and (2) secreted from insects. Keratin fibers are characterized by proteins that are highly cross-linked by disulfide bonds in the cystine residues in the protein chain. Secreted fibers tend not to be cross-linked and are derived from a smaller array of less complex amino acids. Keratin fibers exhibit helical structures periodically throughout; however, secreted fiber protein chains are arranged in a linear pleated-sheet structure with hydrogen bonding between amide groups in adjacent protein chains. On average, protein fibers can be regarded to have moderate strength, resiliency, and elasticity. Though they do not build up a static charge and are fairly resistant to acids, protein fibers are readily attacked by bases and other oxidizing agents. The two main protein fibers of commercial interest are wool and silk.

The final class of natural fibers is that of inorganic fibers, also known as mineral fibers. The common example of an inorganic fiber is asbestos, or “rock wool”.⁸⁴ Despite the fact that the material is formed inside of a rock, it can be spun and made into fabrics

just as its plant and animal derived cousins can. A complete classification scheme for natural fibers can be seen in Figure 3.4.

Table 3.1. Protein Structures⁸⁵

$\left[\begin{array}{c} \text{H} \quad \text{O} \quad \text{H} \\ \quad \quad \\ -\text{C}-\text{C}-\text{N}- \\ \\ \text{R} \end{array} \right]_n$			
Protein			
R	Amino Acid	R	Amino Acid
H-	Glycine	HOCH ₂ -	Serine
CH ₃ -	Alanine	$\text{H}_3\text{C}-\underset{\text{OH}}{\text{CH}}-$	Threonine
(CH ₃) ₂ CH-	Valine	-CH ₂ SSCH ₂ -	Cystine
(CH ₃) ₂ CHCH ₂ -	Leucine	CH ₃ SCH ₂ CH ₂ -	Methionine
$\text{H}_3\text{CH}_2\underset{\text{CH}_3}{\text{CH}}-$	Isoleucine	$\text{H}_2\text{N}-\underset{\text{NH}}{\underset{ }{\text{C}}}-\text{NH}(\text{CH}_2)_3-$	Arginine
	Proline		Histidine
	Phenylalanine	NH ₂ (CH ₂) ₄ -	Lysine
	Tyrosine	HOOCCH ₂ -	Aspartic Acid
	Tryptophan	HOOC(CH ₂) ₂ -	Glutamic Acid

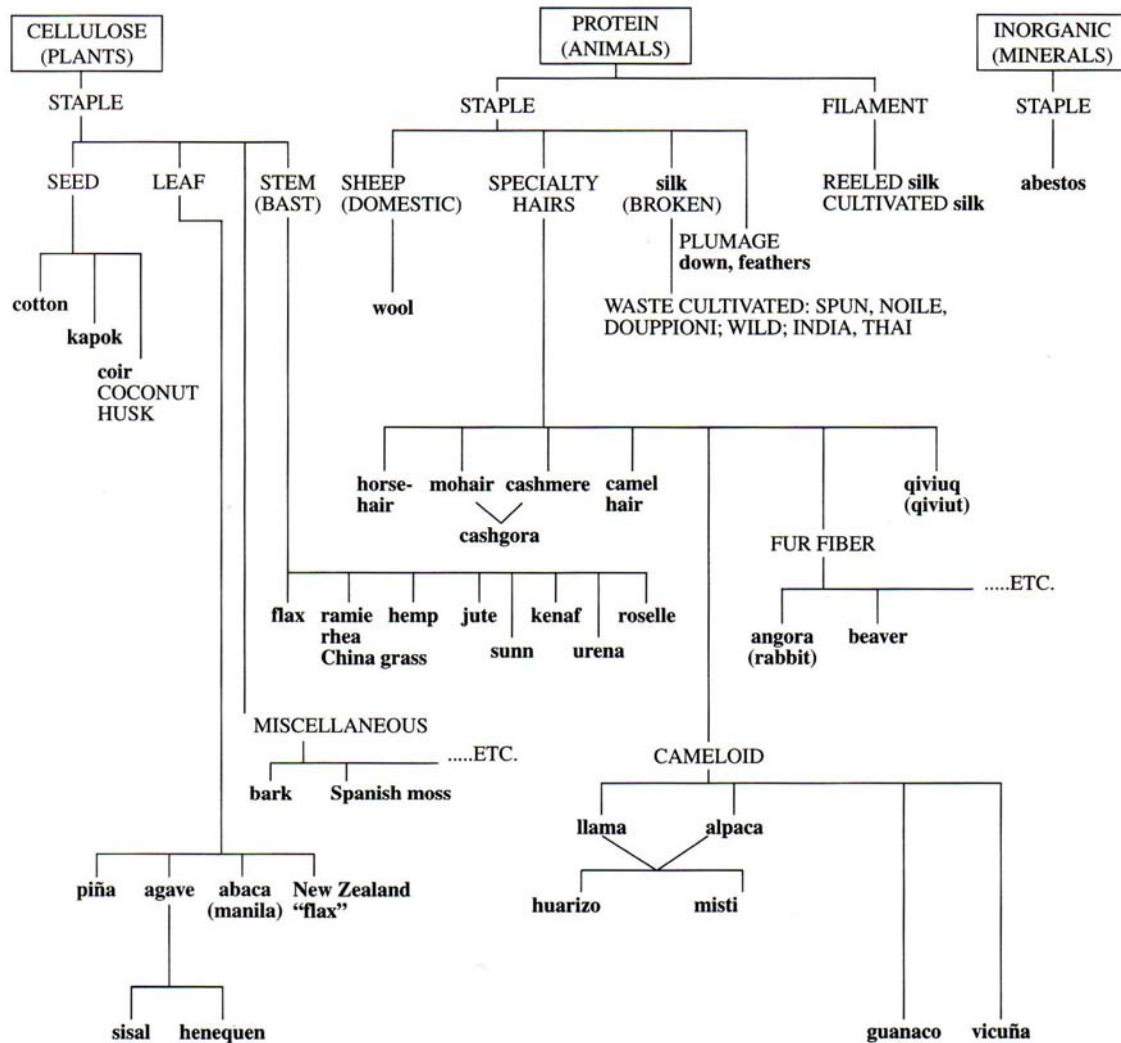


Figure 3.4. Detailed taxonomy for natural fibers (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

Man-made, synthetic, or manufactured fibers are those that are made by chemical synthesis followed by fiber formation. Manufactured fibers can also be produced by taking naturally occurring polymers, such as cellulose, dissolving and then regenerating them.⁸⁵ Two sub-classifications exist for manufactured fibers: natural polymer and synthetic polymer.⁸⁷ Natural polymer manufactured polymers are derived from sources, such as regenerated cellulose and proteins. The first recorded patent for a natural

polymer manufactured fiber is credited to Count Hilarie de Chardonnet of France in 1885 for the production of an artificial silk.⁸³ The fabric was, in fact, modern day rayon, which was synthesized by converting nitrated cellulose into a fiber and then chemically regenerating the cellulose. This was done because the flammability of cellulose nitrate was extremely hazardous. Chardonnet's discovery paved the way for more work to be done in the area of man-made fibers from the late 1880's to the 1930's. Cellulose remained the basis for most of the work accomplished during this period.

In the 1920's and 1930's, research conducted, especially that of Hermann Staudinger and W. H. Carothers positioned the foundation for future advances, particularly those relating to synthetic fibers. The 1953 Nobel prize winner, Staudinger, proved that polymers were molecules with a high molecular weight. Carothers was responsible for defining the two types of polymerization, addition and condensation, demonstrating that polymers of a high molecular weight could be synthesized, and that some of these polymers could be formed into filaments, resulting in oriented, strong fibers.⁸⁸ Carothers' work led in 1939 to two major breakthroughs. The first was the commercialization of nylon. Nylon was the first true synthetic polymer fiber. The second milestone was the use of the melt spinning technique to form the nylon fibers after they were synthesized. By 1977, the global production of manufactured fibers was approaching that of natural fibers.

Whether naturally occurring or synthesized, polymers must meet several conditions in order to be suitable for fiber production. To be of use in the textile industry, the fibers must be capable of being converted into fibrous form and be of sufficient molecular weight. The molecular weight will dictate the overall fiber

properties of the finished fiber. A complete categorical breakdown of the types of manufactured fibers can be seen in Figure 3.5.

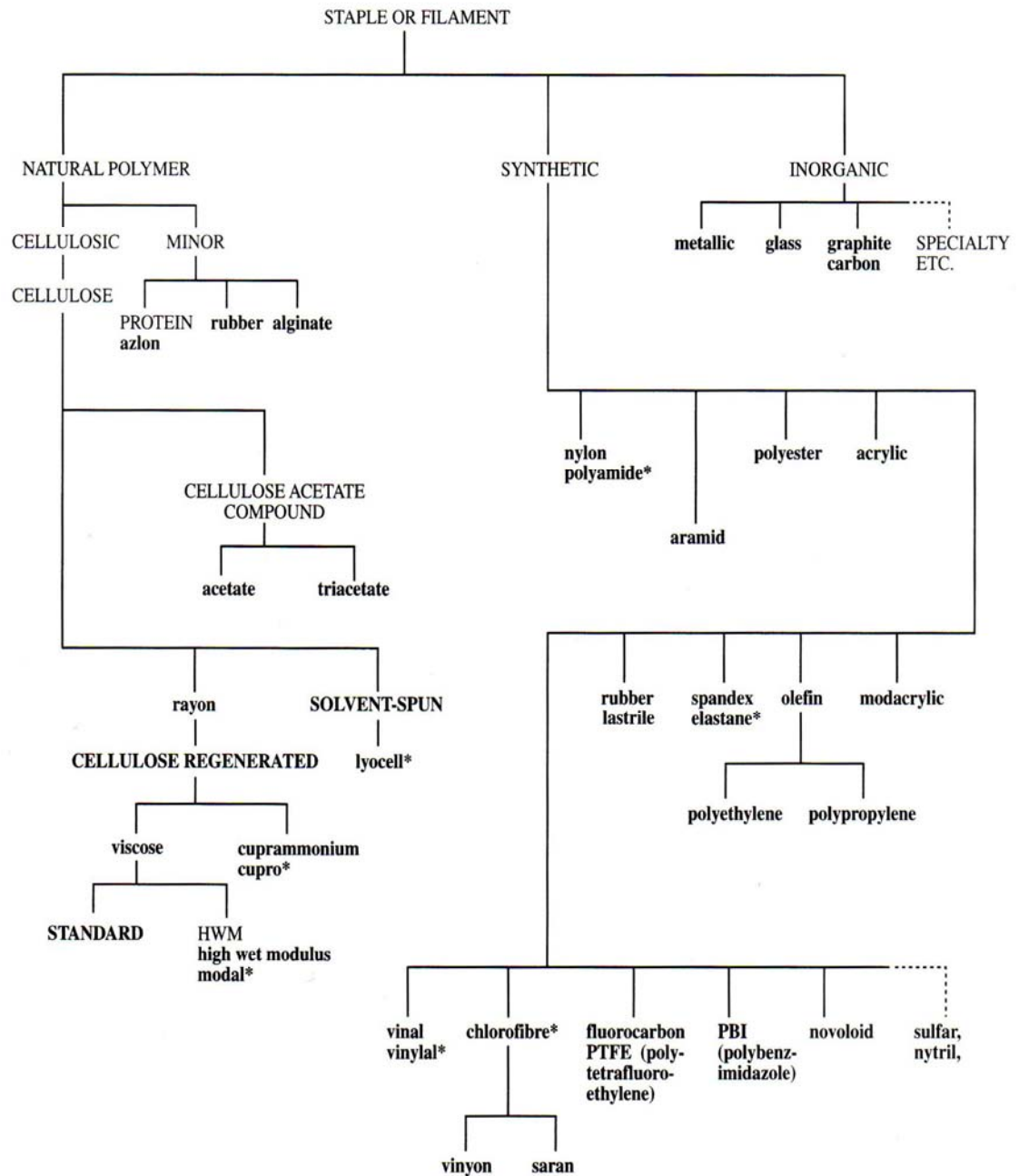


Figure 3.5. Detailed taxonomy for manufactured fibers (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

History and Background of Natural Fibers

1. Cotton. The first use of the cotton plant to make textiles remains unknown as its use predates recorded history. There is evidence for the existence of cotton plants in Egypt as early as 12,000 BC. Archeological findings prove that cotton was in use as of 3,000 BC. Records of the cultivation of cotton plants can be traced to 700 BC in the Indian subcontinent. Herodotus, a Greek historian and often dubbed the “father of history”,⁸⁹ described trees growing in the wild in India, bearing wool as soft and beautiful as that of the sheep in Greece.⁸³ This description was chronicled in his writings in the 5th century BC. The account may not be reliable due to the debate as to the authenticity of Herodotus’ work.⁹⁰ Alexander the Great was said to have introduced Indian cotton to Egypt in the 4th century BC, where it later spread to Greece, Italy, and Spain. The discovery of cotton in North America was made by Christopher Columbus on his historic voyage in 1492. It was not until 300 years later when the first cotton mill was introduced in Beverly, Massachusetts. Eli Whitney was granted a patent for the cotton gin in 1794.

Cotton is a member of the Malvaceae or mallow family. It is a plant of the genus *Gossypium* with important species being *hirsutum*, *barbadense*, *arboretum*, and *herbaceum*. *Gossypium hirsutum* originates from Central America and can be traced back to the time of the Mayan civilization. The many varieties of *Gossypium hirsutum* comprise those grown in the southern United States. Cotton is a single-cell fiber that originates in the epidermis of the seed coat at the time when the flower opens. The fibers of the cotton boll grow to a maximum of 2,500 times their overall width. The cells that make up the fibers are initially composed of a thin wall, known as the primary wall that is covered by a waxy pectinaceous material which encloses the cytoplasm. When the boll is

half mature, cellulose is deposited on the inside of the thin casing to become what is known as the secondary wall. The cellulose layers are continually built up until only a central cavity called the lumen is left clear (Figure 3.6). A microscope magnified view of cotton fibers can be seen in Figure 3.7.

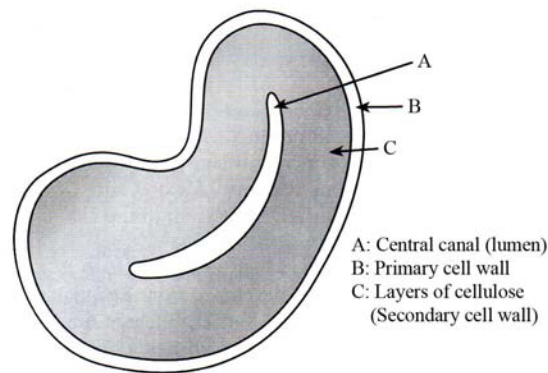


Figure 3.6. Cross-section of mature cotton fiber (Modified from reference 86. Copyright 2002 Prentice-Hall Inc.)

In 1999, cotton made up 54% of the fibers that went into apparel, making it the most important apparel fiber in the world, as well as being the cash crop of more than 80 countries. Major producers of cotton today include China, the United States, India, Eastern Europe, Pakistan, Turkey, and Brazil.⁸⁶ Because of the importance of cotton, recent advances in plant breeding have produced cotton plants that are resistant to insects, herbicides, and stress.

Since cotton is primarily cellulose, 95 wt%, it typically behaves like a cellulose polymer. Other components of the cotton fiber include waxes, pectinaceous substances, protein, organic acids, sugars, and nitrogenous matter.⁹¹ These secondary materials are located mostly in the primary cell wall. Residual protein is the component second only to

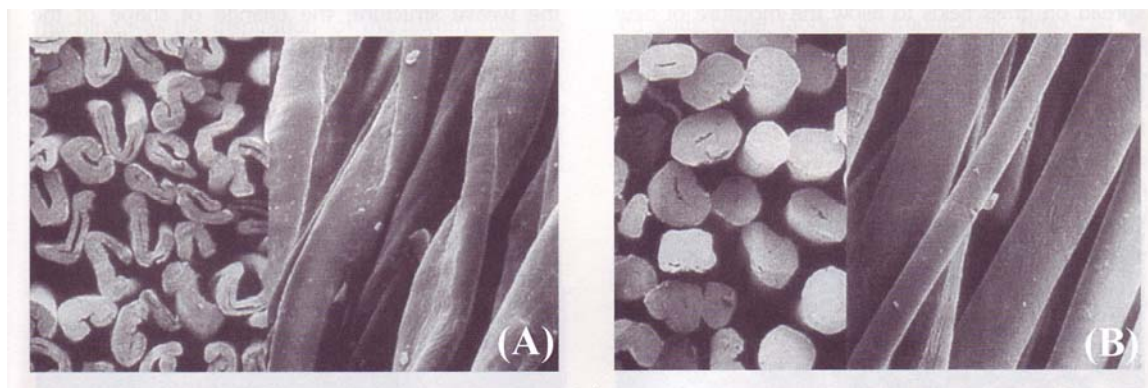


Figure 3.7. Cross-section and longitudinal view of (a) unmercerized and (b) mercerized cotton fibers (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

cellulose in the cotton fiber. The protein is what remains after the protoplasm of the cell has dried. The relative weight percentages for the component parts of a cotton fiber can be seen in Table 3.2.

Table 3.2. Composition of Typical Cotton Fibers⁸³

	Composition, % of dry weight	
	Typical	Range
cellulose	94.0	88.0-96.0
protein	1.3	1.1-1.9
petic substances	1.2	0.7-1.2
ash	1.2	0.7-1.6
wax	0.6	0.4-1.0
total sugars	0.3	
pigment	trace	
others	1.4	

2. *Flax*. Another cellulosic natural fiber is flax. The fibers are derived from the stem of the annual plant, *Linum usitatissimum* that grows in temperate and subtropical

areas of the world.⁸⁵ Flax is used to manufacture linen textiles and fabrics. The word linen is derived from the Celtic term *llin*.⁹² The history of flax fibers is also one of the oldest in human history.⁸⁶ Remnants of linen fabric have been found in prehistoric lake-side settlements in present day Switzerland. The mummified pharaohs of ancient Egypt were wrapped in fine linen cloth as far back as 3,000 years ago. Until the 18th century, the linen industry thrived in Europe. The invention of power spinning methods drove cotton to replace flax as the most used fabric.^{86, 92} Today, most flax is produced in western Europe, in Belgium, Italy, Ireland, and the United Kingdom. Flax is also cultivated in Belarus, Russia, as well as New Zealand. The use of flax has been greatly reduced, due to the extensive labor and cost of producing linen.

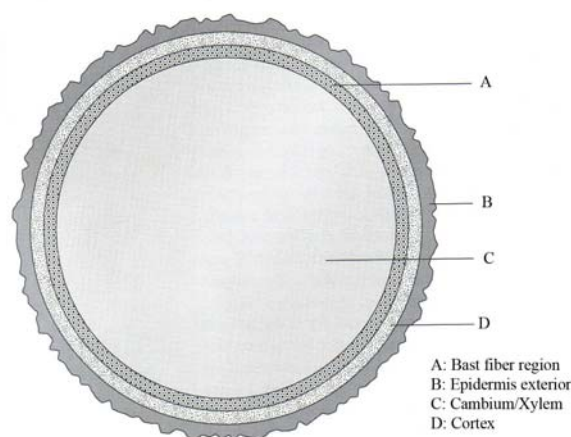


Figure 3.8. Cross-sectional view of a flax fiber (Modified from reference 86. Copyright 2002 Prentice-Hall inc.)

The flax fibers are cleaved from the plant using a fermentation process called retting. Fermentation accelerated by acids or bases have been used with success; however, the natural fermentation process works just as well and at a lesser cost. Two processes called scotching and hackling are done to remove all excess woody material

and orient the fibers for spinning respectively. The desired bast fiber is located between the central woody core of the plant and the outer two layers of the stem, the cortex and epidermis (Figure 3.8).

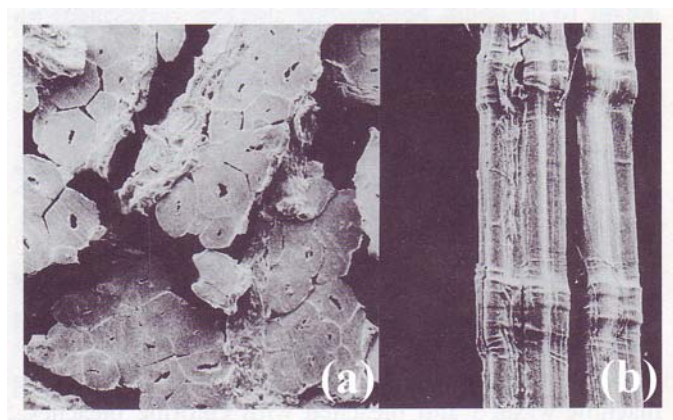


Figure 3.9. Photomicrographs of flax fiber (a) Cross-section and (b) longitudinal view (Modified from reference 84. Copyright 1996 Prentice-Hall Inc.)

As with cotton, flax is almost pure cellulose. The two fibers also share many of the same physical properties as a result. Flax fibers are comprised of many individual fiber cells, or fibrils. The average molecular weight for this cellulosic polymer chain is about 3 million. The cross-section of the flax fibers (Figure 3.9) is polygonal with a large lumen.⁸⁵ Flax is highly resistant to bases, and chemical bleaches.⁹² Flax is also resistant to moths and other insects. Exposure to acid and direct sunlight will cause damage flax fibers.

3. Other Cellulosic Fibers. Natural cellulose fibers are found in most localities throughout the world. There are a large variety of cellulosic fibers that are of importance to the world economy that are not cotton or flax. The term hard fiber is used to indicate

that the fiber in question was derived from the leaf of a plant, while a soft fiber is obtained from the base or stem of the plant.⁹²

Abaca is a hard fiber extracted from the *Musa testilis* plant that is native to the Philippines.⁸³ The fiber, also known as Manila hemp, was of prime importance due to its use as cordage in the Napoleonic and Crimean Wars.⁹² Finer grades of abaca can also be woven into fabrics for uses as tablecloths, placemats, and purses in addition to its use in ropes and twine. Another leaf fiber is sisal, which is derived from fibers taken from the *Agave sisalana* plant. Like abaca, sisal is composed of cellulose and lignin; however, sisal also contains other carbohydrates, hemicelluloses, pectins, and wax. Because it comes from the stiff, parallel fibers of the leaf, sisal has a high degree of tensile strength. It has been used in the manufacture of ropes, sacking, twine, sailcloth, and cable insulation. Other types of hard leaf fibers include: henequen, cantala, istle, Mauritius, phorium, snsevieria, caroa, pissava, piña, and broomroot.^{83, 86} All of these hard fibers have similar physical characteristics.

Of the soft-bast fibers, flax is the most popular. Another soft-bast fiber is hemp, which is derived from the annual plant *Cannabis sativa*.⁸³ Early reports of cultivation of hemp come from China in 2,800 BC.⁹² The introduction of the plant to Europe came prior to the Christian era. The use of hemp originally was, as is with most of the bast fibers, as cordage; however, abaca and sisal, and then later manufactured fibers, have supplanted the use of hemp in rope. Ramie is a prehistoric soft fiber that is also known as rhea in India and China grass in China. It was not introduced to Europe until the 18th century and America in 1850. Applications for ramie include its use as twine, fishnets, industrial sewing thread and packing, canvas, fire hose, filter cloths, rugs, upholstery,

clothing, and household fabrics. Jute is similar to hemp, and can be used as a replacement for hemp. As with most cellulose-based materials, jute is damaged by exposure to acid. Jute is also damaged by alkalis and damaged by bleaching. Other types of soft-bast fibers include: kenaf, roselle, urena, China jute, and sunn hemp.

A few other examples of cellulose fibers are found in use today. Two seed fibers are kapok and coir.⁸⁶ Kapok is taken from the seed pod of the tropical tree *Ceiba pentandra*. Because it is hollow, the fiber itself is buoyant. This characteristic makes it a perfect material to stuff life jackets; nevertheless, different foams and polyester fiberfill has largely superseded the use of kapok. Coir is a coarse, stiff, and strong fiber made from the husk of a coconut. It is often used as cordage and rough matting. Other seed fibers with minor amounts of use include East Indian basla, Indian kumbi, American milkweed floss, and cattail.⁸³ A more complete listing of plant fibers of commercial interest can be found in Table 3.3. Bark fiber cloth, known as Tapa, is produced by using the inner bark of the paper mulberry tree, *Broussonetia papyrifera*.⁸⁴ The bark is pounded into bark cloth in regions like the Pacific islands and is considered a natural non-woven fabric. Spanish moss, related to the pineapple, has been used in the southern United States as a stuffing material as an alternative to horsehair.

4. *Wool*. The highly crimped protein hair fiber wool is procured from sheep. The variety of the sheep will determine the fineness, structure, and properties of the fiber. The varieties of sheep that are sheared for their wool include the Merino, Lincoln, Leicester, Sussex, and Cheviot breeds.⁸⁵ The use of wool by man for clothing and other purposes dates back before written records were kept. There is evidence that wool was used as household items and clothing by the Egyptians, Babylonians, and Romans.⁹² It

Table 3.3. Selected Plant Fibers of Commercial Interest⁸³

Commercial name	Botanical name	Source	Use
Leaf fibers			
Abaca	<i>Musa textilis</i>	Borneo, Philippines	cordage
Cantala	<i>Agave cantala</i>	Philippines, Indonesia	cordage
Caroa	<i>Neoglaziovia vareigata</i>	Brazil	cordage, coarse textiles
henequen	<i>Agave fourcroydes</i>	Australia, Cuba, Mexico	cordage, coarse textiles
Istle	<i>Agave (various)</i>	Mexico	cordage, coarse textiles
Mauritius	<i>Fureraea gigantea</i>	Brazil, Venezuela, tropics	cordage, coarse textiles
phormium	<i>Phorium tenax</i>	Argentina, Chile	cordage
bowstring hemp	<i>Sansevieria</i>	Africa, Asia, S. America	cordage
Sisal	<i>Agave sisalana</i>	Haiti, Java, Mexico	cordage
Bast fibers			
China jute	<i>Abutilon theophrasti</i>	China	cordage, coarse textiles
Flax	<i>Linum usitatissimum</i>	N. S. temperate zones	textiles, threads
Hemp	<i>Cannabi sativa</i>	All temperate zones	cordage, oakum
Jute	<i>Corchorus capsularis</i>	India	cordage, coarse textiles
Kenaf	<i>Hibiscus cannabinus</i>	India, Iran, Russia	coarse textiles
Ramie	<i>Boehmeria nivea</i>	China, Japan, US	textiles
Roselle	<i>Hibiscus sabdarifa</i>	Brazil, Indonesia	cordage, coarse textiles
Sunn	<i>Crotalaria juncea</i>	India	cordage, coarse textiles
Cadillo	<i>Urena lobata</i>	Africa, Brazil	cordage, coarse textiles
Seed hair fibers			
Cotton	<i>Sossypium sp.</i>	US, Asia, Africa	textiles, cordage
Kapok	<i>Ceiba pentrandia</i>	tropics	stuffing
Miscellaneous fibers			
broom root	<i>Muhlenbergia macroura</i>	Mexico	brooms, brushes
Coir	<i>Cocos nucifera</i>	tropics	cordage, brushes
crin vegetal	<i>Champaerops humilis</i>	North Africa	stuffing
Piassava	<i>Attalea funifera</i>	Brazil	cordage, brushes

was not until the mid 1600's that wool was introduced to the colonies in North America, following its use in Italy, England, and Spain. Though the consumption of wool has declined in recent years due to the assortment of man-made fibers, which have similar physical properties, wool is still used in the United States.

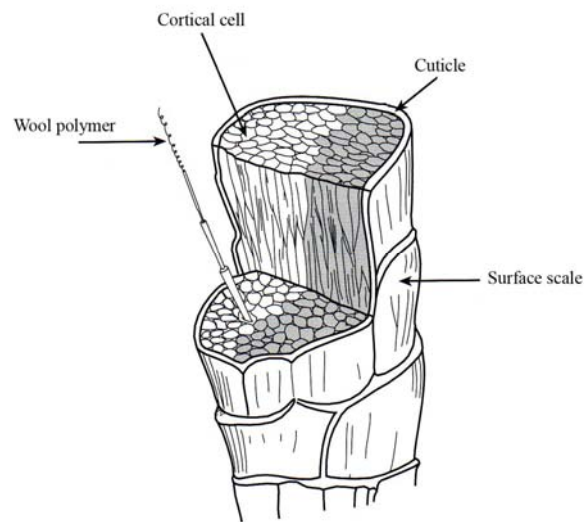


Figure 3.10. Physical structure of a wool fiber (Modified from reference 86. Copyright 2002 Prentice-Hall Inc.)

The exterior of the wool fiber is covered by a thin membrane called the epicuticle (Figure 3.10). This membrane protects the next layer which is known as the cuticle. These layers of protection insulate the cortex, which is made up of long, flattened, tapered cells. The grouping of cells on the different sides of the fiber, the ortho- and paracortex, respond differently to moisture and temperature leading to wool's natural crimp. The crimp is what helps the fibers cling to one another in a yarn, increasing its overall strength and resiliency. Within the cortical cells, are a collection of fibrils that delineate down until the wool biopolymer is seen.⁸⁶

Wool is a highly cross-linked keratin protein polymer comprised of over 17 amino acid residues (Figure 3.11). The protein chains in wool are connected periodically through the disulfide cross-linked cystine. The average amino acid contents of the major varieties of wool can be found in Table 3.4.

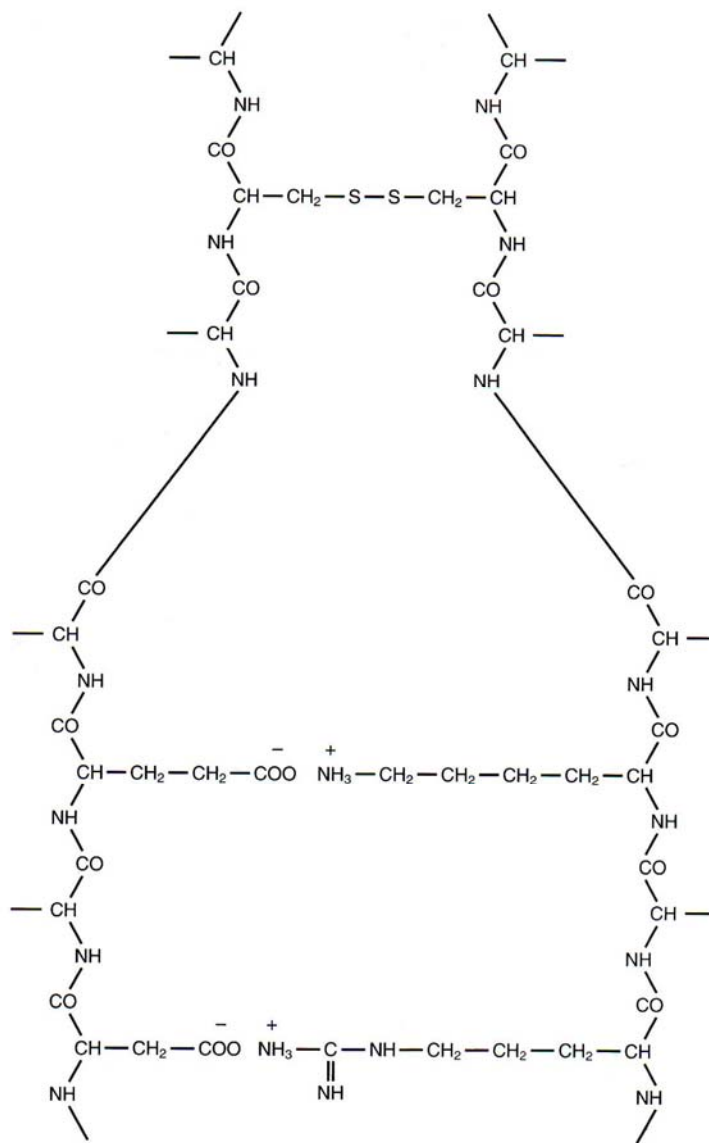


Figure 3.11. Chemical formula for wool molecule (Modified from reference 86. Copyright 2002 Prentice-Hall Inc.)

Table 3.4. Amino Acid Contents in Wool Keratin Fibers⁸⁵

Amino Acid	Content in Keratin ^a	Amino Acid	Content in Keratin ^a
Glycine	5-7	Alanine	3-5
Valine	5-6	Leucine	7-9
Isoleucine	3-5	Proline	5-9
Phenylalanine	3-5	Tyrosine	4-7
Tryptophan	1-3	Serine	7-10
Threonine	6-7	Cystine	10-15
Methionine	0-1	Arginine	8-11
Histidine	2-4	Lysine	0-2
Aspartic acid	6-8	Glutamic acid	12-17

^aContent g/100g Wool

Approximately 40% of the protein chains in wool are in the form of an α -helix (Figure 3.12), due to the chains spiraling onto themselves and being held together by internal hydrogen bonding. Close packing of the polymer chains is not possible at points where the periodic cystine crosslinks or where proline and other amino acids with bulky substituents exist. This results in a less regular non-helical geometry.

The 1998 domestic consumption of wool in the United States was 190 million pounds. This amount was approximately 1.0% of the all the fibers used in that same year.⁸⁶ The main use of wool is in the manufacture of business suits due to the fabric's performance and durability. Though the amount of wool used in furnishings is quite low, it has been established as the standard by which carpets are judged. Many fire-safety blankets used in chemical and other types of laboratories are made from wool. Industrial

felts that are used as insulators to cut down on machine noise as well as absorb oil and other spills are derived from wool. Landscapers often use wool mulch mats to aid in weed control.

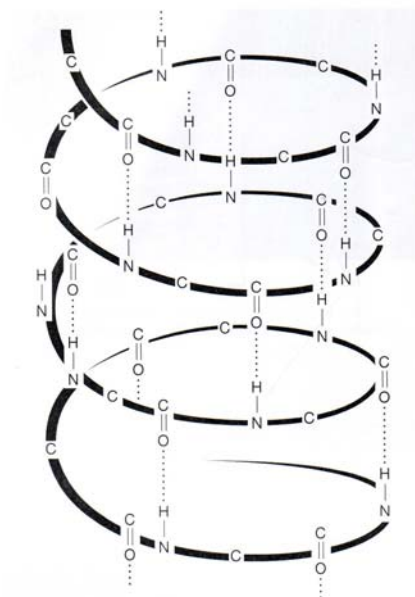


Figure 3.12. α -helix geometry of wool (Modified from reference 86. Copyright 2002 Prentice-Hall inc.)

5. *Silk*. The natural protein polymer known as silk is excreted by the moth larva *Bombyx mori*, which is more commonly known as a silkworm.⁸⁵ Silk is a fine continuous monofilament fiber which has high luster and strength. The fiber and textiles made from it are often considered prestigious or a status symbol due to the high cost of silk. There have been several business ventures attempted to commercialize silk from spiders, but all of the attempts met with failure. The amino acid content for the silk fibroin fibers can be seen in Table 3.5. The predominant amino acids found in fibroin are glycine, alanine, tyrosine, and serine.

Table 3.5. Amino Acid Contents in Silk Fibroin of the *Bombyx mori* larva)⁸⁵

Amino Acid	Content in Fibroin ^a	Amino Acid	Content in Fibroin ^a
Glycine	36-43	Alanine	29-35
Valine	2-4	Leucine	0-1
Isoleucine	0	Proline	0-1
Phenylalanine	1-2	Tyrosine	10-13
Tryptophan	0-1	Serine	13-17
Threonine	1-2	Cystine	0
Methionine	0	Arginine	0-2
Histidine	0-1	Lysine	0-1
Aspartic acid	1-3	Glutamic acid	1-2

^aContent g/100g Fibroin

Without the presence of cystine, silk is essentially an uncross-linked polymer and has a relatively simple amino acid composition compared to keratin fibers, such as wool. With no cross-linking agents and limited numbers of bulky substituents, fibroin molecules align themselves parallel to each other and hydrogen bond to form a pleated or beta sheet structure (Figure 3.13).

The liquid silk protein is secreted from two glands at the head of the larva. The fibers emerge from the spinneret and harden into a single strand by a water-soluble protective gum called sericin.^{85, 86} The cocoons of the larva are then soaked in hot water to loosen the sericin binding agent. Once the binding agent has released the fibers, they are then unwound and washed in warm detergent solutions to remove the remaining

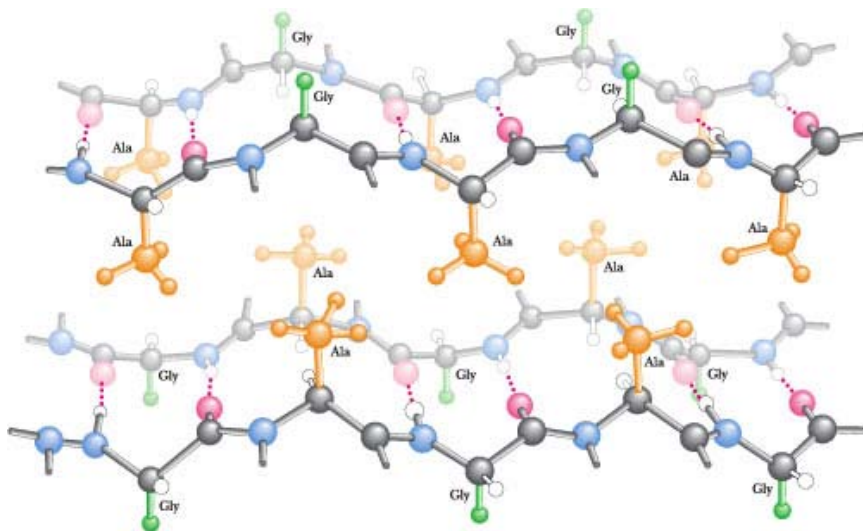


Figure 3.13. Beta sheet structure of silk fibroin

sericin. The finished fibers are strong with moderate degrees of recovery from deformation. Silk is soluble in aqueous lithium bromide, phosphoric acid, and cuprammonium solutions. Silk show resistance to damage by acids; however, with long exposure, the fibers will deteriorate. Biological agents have little effect on silks, but the fibers do yellow and lose strength when put in direct sunlight. Tin and other metal salts are used to weight silk fibers. This makes the silk more susceptible to light-induced oxidative attack.

6. Other Natural Protein Fibers. Other major keratin based fibers, sometimes called specialty wools,⁸⁶ include mohair, cashmere, llama, alpaca, and vicuña.⁸⁵ Mohair and cashmere are resilient fibers obtained from the angora and cashmere goats, respectively. Mohair posses a brilliant luster and is often blended with other fibers to impart this luster on the finished textile. There is a scratchy feel to mohair that is not

seen in cashmere. Cashmere fibers are extremely fine and soft and used in luxury applications.

Llama, alpaca, guanaco, and vicuña are all part of the camel family or cameloids. These animals are raised in South America, and the fibers derived from them are usually found in expensive luxury items like those made from cashmere.⁸⁵ Both guanaco and vicuña are now protected. As of mid-1999, it was illegal to bring any items into the United States that contained fibers from the vicuña because of the Endangered Species Act.⁸⁶

7. Inorganic Fibers. Asbestos has been known since the time of the ancient Greeks for its valuable property of being flame retardant. In fact, the term asbestos is derived from a Greek word $\alpha\sigma\beta\epsilon\sigma\tau\omicron\varsigma$, meaning “will not burn”. The material can melt, but only when the ambient temperature reaches 1450-1500 °C. Various types of asbestos exist, including hydrated silicates of magnesium and calcium, along with other minerals. The best fiber-quality asbestos, used for textiles, is chrysolite asbestos, $\text{Mg}_3(\text{Si}_2\text{O}_5)(\text{OH})_4$, was mined in the eastern townships of Quebec.⁸⁶ Chrysolite asbestos is also known as white asbestos. Amosite, $\text{Fe}_7\text{Si}_8\text{O}_{22}(\text{OH})_2$, or brown asbestos is another type of asbestos found in South Africa. Blue asbestos, Riebeckite, is found in Africa and Australia.⁹³

In the past, asbestos had been used as lamp wicks and cleaning cloths. Asbestos has also been used in sheetrock taping, vinyl flooring, plasters and stuccos, roofing tar, siding, countertops, acoustical ceilings, brake pads, as well as fireproof clothing and blankets. The United States Environmental Protection Agency (EPA) has made efforts to ban or restrict the use of asbestos in consumer products. Of all the types of asbestos, blue is considered to be the worst biohazard. Some conditions linked to asbestos include

asbestosis, mesothelioma, and cancer.⁹³ All of these diseases are a direct result of asbestos fibers entering and being lodged in the inner lining of the lungs.

History and Background of Manufactured Fibers

1. Rayon. It was observed by Frederick Schoenbein, in 1846 that cellulose pretreated with nitric acid would dissolve in a mixture of ether and alcohol. This process may have lead to mass production of rayon fibers; however, the resulting fibers were found to be highly explosive. The dangerous nitrated cellulose was converted back to ordinary cellulose by Count Hilarie de Chardonnet in 1889. Rayon fibers are regenerated cellulose produced from a solution of cellulose, usually obtained from wood pulp, cotton waste, or other sources, that is extruded through a spinneret and subsequently regenerated into fiber form.⁸⁵ The United States Federal Trade Commission defines rayon as a manufactured fiber composed of regenerated cellulose in which substituents have not replaced more than 15% of the hydroxyl hydrogens.⁸⁶ Rayon was the first regenerated manufactured fiber to be produced commercially. Viscose rayon, cuprammonium rayon, and saponified cellulose acetate are the three different forms of rayon produced, by various methods. Viscose rayon is the most important and least expensive to produce among these fiber types.

Viscose rayon production in the United States was begun in 1911. The fibers were originally dubbed artificial silk, as was de Chardonnet's original patented material, until the name rayon was adopted in 1924.⁸⁶ Viscose rayon fibers are produced by what is known as the viscose process (Figure 3.14), developed in 1891 by Cross, Bevan, and Beadle in England. The overall reaction can be seen in Figure 3.15. First, the

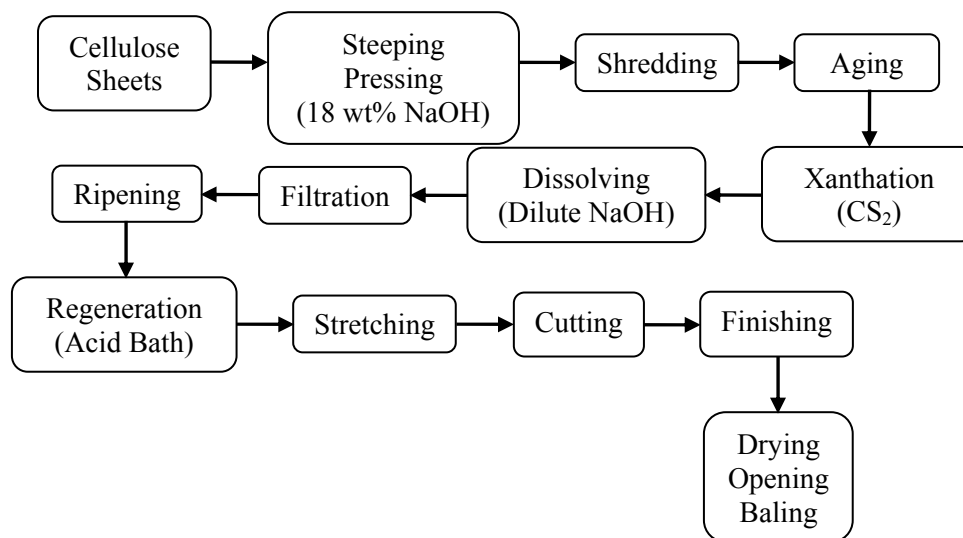


Figure 3.14. The viscose rayon production process (Adapted from reference 83. Copyright 1984 John Wiley & Sons)

cellulose sheet raw material is converted to alkali cellulose by placing the sheets in a steeping press and filling the press with a closely controlled solution of 18-20 wt% sodium hydroxide (± 0.2 wt%).⁸³ After the steeping process is complete, the cellulose is pressed under high pressure to remove as much excess sodium hydroxide. The resulting material is then shredded to allow for a more even spread of the caustic sodium

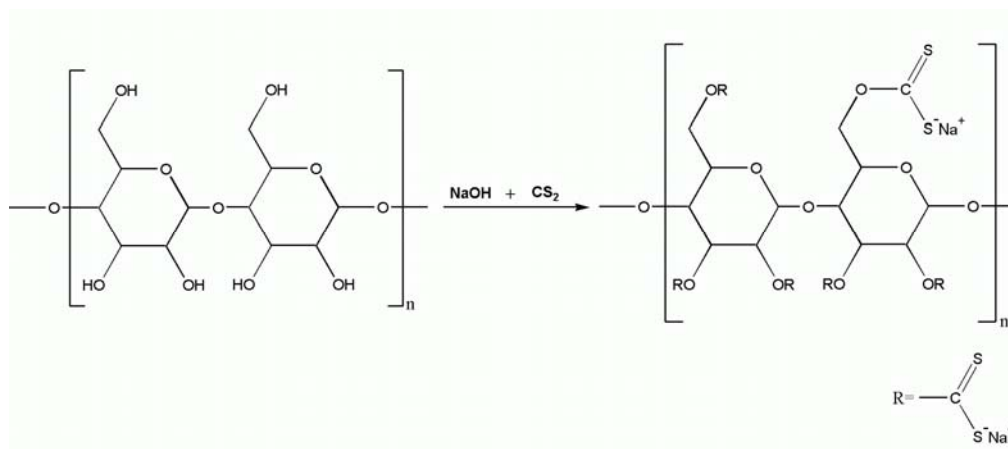


Figure 3.15. Synthesis of rayon from cellulose

hydroxide over the surface of the fibers. Next, the cellulose, often called cellulose cruds, is aged in covered containers. After aging, the crumbs are combined with carbon disulfide to form cellulose xanthate. The material is then placed in a dilute sodium hydroxide solution and stirred before it is allowed to age and ripen until the proper viscosity is achieved.

The cellulose solution is then forced through the spinneret into a dilute sulfuric acid solution. The acidic environment will decompose the xanthate and regenerate the cellulose in what is called wet-spinning. Various additives can be placed in the acid bath, such as sodium sulfate, zinc sulfate, and glucose. The additives can regulate the decomposition of the xanthate as well as the regeneration of the cellulose itself. Regular viscose rayon is characterized by lengthwise lines called striations (Figure 3.17a).

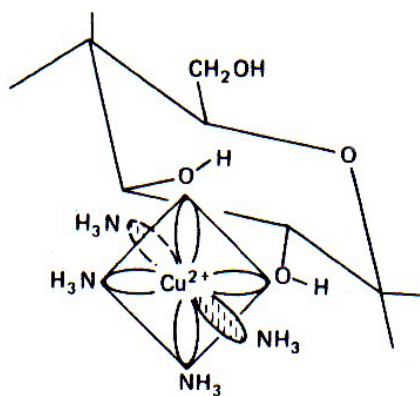


Figure 3.16. Orbital interactions in cuprammonium cellulose (Modified from reference 83. Copyright 1984 John Wiley & Sons)

Cuprammonium rayon (CAR) is produced by adding cellulosic materials to a solution of cuprammonium hydroxide at low temperature under nitrogen. Then, the solution is extruded through a spinneret into a water bath, followed by a sulfuric acid

rinse to decompose the cuprammonium complex formed (Figure 3.16) and to regenerate the cellulose. The process was discovered by Louis Despeissis in 1890. The commercialization of cuprammonium rayon was achieved in 1919 by J. P. Bemberg.⁸⁶

CAR is more silk-like than any of the other cellulose derived fibers.⁸⁵ This physical attribute comes with an increase in the cost of production as well.

Photomicrographs of both viscose and cuprammonium rayon can be seen in Figure 3.17b.

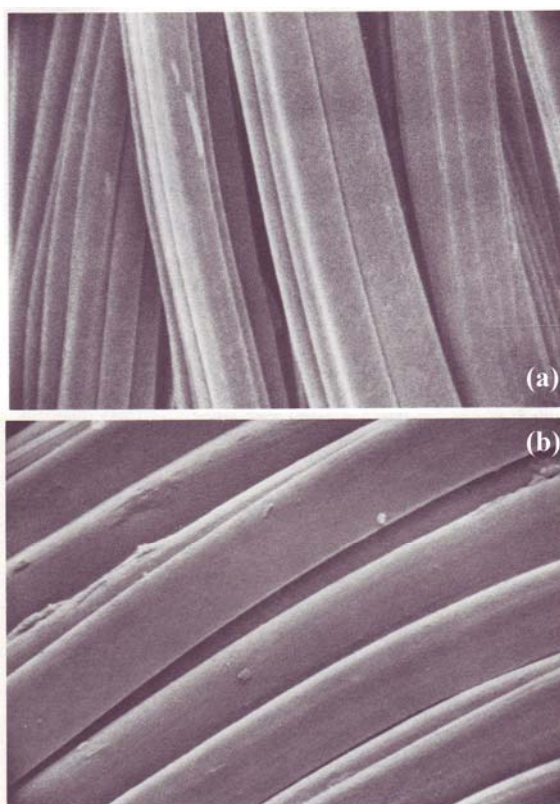


Figure 3.17. Photomicrographs of (a) viscose rayon x2000 and (b) cuprammonium rayon x1100 (Modified from reference 85. Copyright 1986 Noyes publications)

Since it is derived from cellulose, the chemical properties of rayon are similar to those of cotton. Whether in dilute or concentrated solution, bases at low temperatures do not significantly affect rayon fibers. However, dilute acids at elevated temperatures will

attack rayon at a faster rate than that of cotton. Oxidizing agents have little effect with the exception of hydrogen peroxide, which readily attacks rayon.⁸⁵ Rayon is resistant to most microorganisms and insects, with an exception being silverfish. Excessive sunlight exposure will cause a reduction in the fibers' strength and resiliency.⁸⁵

2. Cellulosic Esters. The cellulosic esters, or acetates for short, originated in Europe during World War I. The Dreyfus brothers experimented with acetate in Switzerland, before going to England. They managed to perfect an acetate dope used as a varnish for airplane wings.⁸⁶ The two major cellulosic esters are acetate, also known as secondary acetate, and triacetate. The Federal Trade Commission defines acetate fibers as manufactured fibers in which the fiber-forming substance is cellulose acetate.^{85, 86} When a cellulose fiber has more than 92% of its hydroxyl groups acetylated, it is called a triacetate fiber. Acetate fibers are categorized as those cellulose acetates that have less than 92% of their hydroxyl groups acetylated. Structures for both polymers can be seen in Figure 3.18.

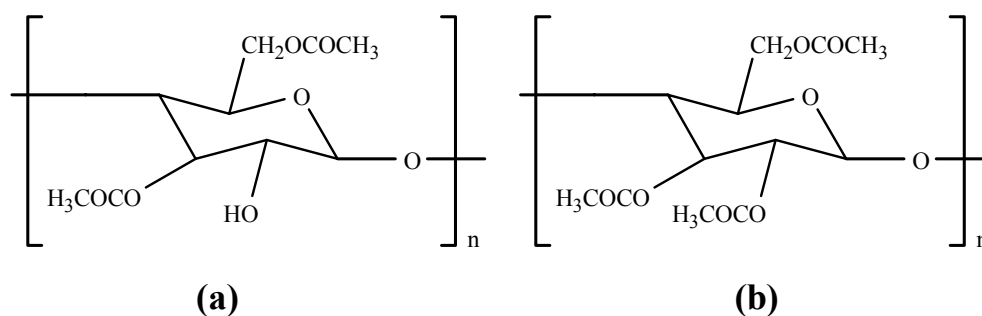


Figure 3.18. Chemical structures for the cellulose esters (a) acetate and (b) triacetate

The preparation of triacetate fibers begins with purified cellulose, usually wood pulp or cotton linters. They are mixed with glacial acetic acid, acetic anhydride, and a

sulfuric acid catalyst. The solution is left to age for 20 hours, which allows for partial hydrolysis to occur. The triacetate is precipitated as an acid-resin flake. These flakes are then dissolved in acetone, filtered, solution spun, and then dried with warm air.⁸⁶ Acetate is obtained by an acid catalyzed hydrolysis of triacetate to achieve an average degree of substitution of 2.4 acetyl groups on each glucose unit.⁸³

While both acetate and triacetate fibers are resistant to dilute acid solutions, both are readily attacked by concentrated acids that cause hydrolysis and removal of the acetate ester groups. Both cellulosic esters show resistance to alkali solutions; nonetheless, acetate is more susceptible to attack than triacetate.⁸⁵ Microorganisms and insects have very little effect on acetates. Both acetates exposed to direct sunlight perform admirably, but the greater stability was found to be with triacetate.

3. *Polyamides*. Polyamide fibers include the nylons as well as the aramid fibers, which are both formed from polymers of long-chain polyamides. Nylon fibers have less than 85% of their polyamide units attached directly to two aromatic rings, while aramid fibers have more than 85% of their amide groups directly attached to aromatic rings.⁸⁵

Nylon was the first synthetic manufactured fiber, and the first fiber developed in the United States. DuPont initiated a research effort to develop nylon fibers in 1928.⁸⁶ Wallace Carothers' team at DuPont synthesized a wide assortment of polymers. It was not until a team member discovered that the solutions they were making could be drawn out into a stable solid filament that Carothers and co-workers began to concentrate on textile fibers. Commercial release of nylon into the marketplace occurred after 1939⁹² with a successful pilot plant. Two common examples of nylons are nylon-6 and nylon-

6,6 (Figure 3.19). The name nylon was originally going to be No-Run, but DuPont felt that making an unjustified claim about the new polymer would not be practical.

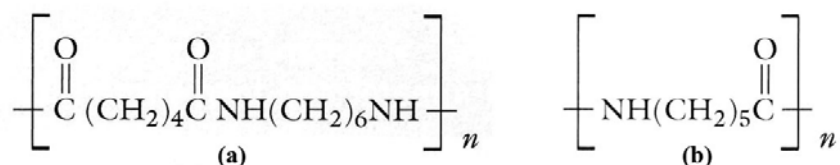


Figure 3.19. Chemical structures for common (a) nylon-6,6 and (b) nylon-6

Both nylons have similar properties. Nylon polymers are extremely strong and durable.^{85, 86} A ring-opening chain growth polymerization of caprolactam in the presence of water vapor and an acid catalyst is the method to synthesize nylon-6. The fibers are melt-spun at 250-260 °C after the water and acid have been removed. Nylon-6,6 is produced by a step-growth polymerization of hexamethylene diamine and adipic acid and melt spun at 280-290 °C. The polyamide chains, which vary in degrees of polymerization from 100 to 250 monomer units, lie parallel to one another in a pleated sheet geometry similar to that of natural silk protein polymers. The chains are held together by hydrogen bonding between the amide linkages on neighboring chains.

Nylon fibers are moderately hydrophilic with the moisture uptake into the fibers of up to 4-5% under standard conditions.⁸⁵ With a relative humidity of 100%, the moisture regain can be as high as 9%. Nylons are soluble a variety of solvents like phenols, 90% formic acid, and benzyl alcohol. Acids, bases, and oxidizing agents show little effect on nylon fibers. Only under extreme conditions are the fibers degraded by these reagents.

Nylon is one of the most widely used fibers in the United States.⁸⁶ End use applications of nylons include hosiery, lingerie, underwear, sweaters, and other knitted goods.⁸⁵ Their use has been extensive in garments that are light and sheer such as windbreakers, nightgowns, pajamas, and lightweight robes. Perhaps the most important use of nylon today is in carpets, due to its aesthetic, appearance, durability, and ability to be cleaned in place. Industrial applications vary to include tire cordage as well as automobile interiors.

Aramid polyamide fibers are produced from a long chain of synthetic polyamides in which at least 85% of the amide linkages are attached to aromatic rings.⁸⁶ These polyamides are characterized with high melting points and excellent property retention and durability even at extreme temperatures. The fibers are moisture resistant and inherently flame retardant. Aramids are the strongest of the man-made fibers, depending on polymer structure, spinning method, and degree of orientation of the fiber.⁸⁵ Nomex and Kevlar are two examples of common aramid fibers commercially available (Figure 3.20).

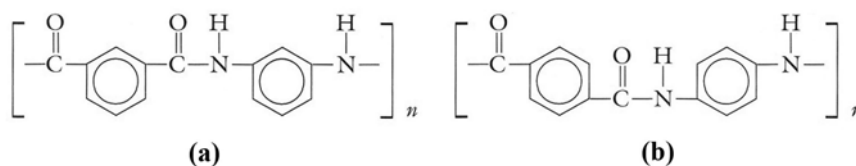


Figure 3.20. Chemical structures for common aramid fibers (a) Nomex and (b) Kevlar

The aramids are synthesized through a step-growth polymerization of aromatic diacids, like terephthalic or isophthalic acid, with aromatic diamines in a polar-aprotic solvent, such as *N,N*-dimethylformamide.⁸⁵ Like with nylon-6 and 6,6, the degree of

polymerization for these fibers is between 100-250 monomer units. The meta- (Nomex) and para- (Kevlar) substituted benzene dicarboxylic acid chlorides and diamines are routinely used in aramid fiber production. The resulting aramid fibers in suspension are passed through the spinneret and either dry spun in hot air or wet spun in a coagulating bath. Aramids have limited flexibility and mobility due to the aromatic units along the aramid polyamide chain. They constrain the polymeric chains and cause stiffness. Hydrogen bonding between the amide groups and adjacent chains as well as tremendously strong van der Waals interactions between aromatic rings planar to adjacent aromatic rings provides a tightly-packed, forcefully held molecular structure that accounts for the strength and thermal resistance of aramid fibers.

Polar aprotic solvents and strong acids can dissolve the aramid fibers. For the most part, aramids are resistant to biological attack from microorganisms and insects, as well as most chemicals. Sunlight exposure will cause discoloration because of an initial oxidative attack. The initial discoloration and slight loss of strength is the extent of the damage the sun can cause. Further exposure will have little or no effect. Aramid fibers have also been used as effective screens from high energy nuclear radiation. The fibers have the ability to trap and stabilize radical and ionic species induced by a nuclear blast.⁸⁵

The need for fibers with certain physical or chemical characteristics has given rise to a group of specialty polyamide fibers. Some specialty nylon fibers include Qiana, nylon-4, nylon-11, nylon-6,10, and biconstituent nylon-polyester.⁸⁵ Qiana (Figure 3.21) is the trade name for a luxury nylon fiber formed through step-growth polymerization of trans, trans-di(2-aminocyclohexyl)methanol and a dibasic acid having between 8-12

carbons. The properties of Qiana resemble those of nylon-6 and 6,6 as well as having a unique silk-like texture.

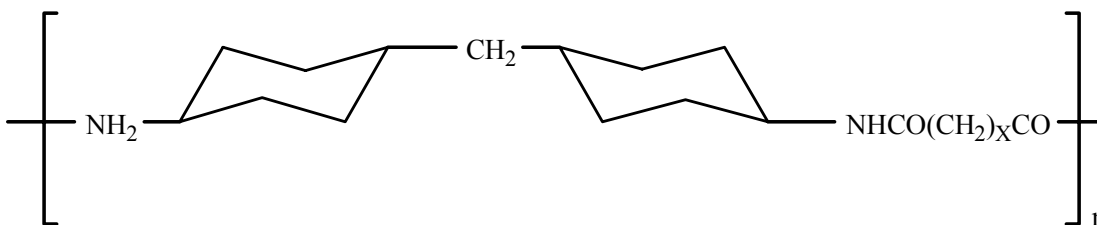


Figure 3.21. Chemical structure of polyamide polymer fiber Qiana

Nylon-4 was produced by a polymerization of pyrrolidone using carbon dioxide as a catalyst.⁸⁵ The material was then melt spun above the polymer's melting point of 273 °C. Due to the status of the textile market of the time, with the presence of such fibers as nylon-6 and 6,6, nylon-4 was never commercially produced in any great quantities. Nylon -11 is synthesized during a self-condensation polymerization reaction of 11-aminoundecanoic acid. It is also melt spun above its 189 °C melting point. This fiber is often used as an insulator in electrical devices and in common household applications such as brush bristles and clothing. A condensation reaction between hexamethylene diamine and sebacic acid produces nylon-6,10.

4. Polyesters. The first polyester fiber, known as Terylene, was produced in England.⁸⁶ DuPont acquired the patent for polyester in the United States in 1948, while Imperial Chemical Industries secured the patent rights for the United Kingdom and the rest of the world.⁸³ Commercially availability of polyester was not seen until 1953. Manufactured fibers containing at least 85% of a polymeric ester of a substituted

aromatic carboxylic acid including, but not restricted to, terephthalic acid and *p*-hydroxybenzoic acid are called polyesters.⁸⁵

The major polyester of commercial importance is polyethylene terephthalate (PET) polyester seen in Figure 3.22. PET is an ester formed from a step-growth polymerization of terephthalic acid and ethylene glycol in the presence of a catalyst. The degree of polymerization is between 100-250 monomer units. Generally, when someone says a material is polyester, it is a reference to this generic type. The finished polymer molecular chains are stiff and rigid due to the presence of periodic phenylene groups, which results in a tightly packed fiber, held together by van der Waals interactions.

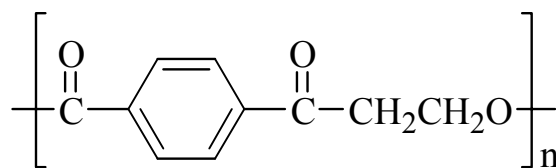


Figure 3.22. Chemical structure of polyethylene terephthalate (PET) polyester

PET is highly resistant to attack from acids, bases, oxidizers, and reducing agents. Significant damage is only seen when concentrated acids or bases at high temperatures are used. Biological agents have no deleterious effects on PET polyester. Oxidative damage, along with discoloration, is seen when the fibers are exposed to sunlight. The fibers are hydrophobic and are non-absorbent unless chemical modifications are made.

Poly-1,4-cyclohexylenedimethylene terephthalate (PCM) polyester (Figure 3.23) is formed through the step-growth polymerization of terephthalic acid and 1,4-cyclohexylenedimethanol.⁸⁵ Most of the properties of PCM resemble those of PET polyester. The addition of the cyclohexylene group in this fibers structure gives rise to

additional rigidity; however, the packing of the polymer chains in the fiber may not be as tight as PET due to its increased bulk. PCM is superior to PET when considering recovery from stretch as well as resistance to pilling. PCM is also more resilient and well suited to being blended with cellulose and wool fibers.

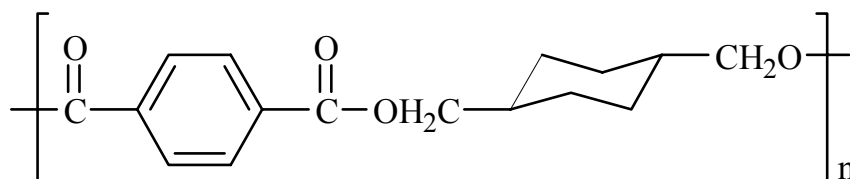


Figure 3.23. Chemical structure of poly-1,4-cyclohexylenedimethylene terephthalate (PCM) polyester

5. *Acrylics*. The research into synthetic fiber production based on polyacrylonitrile (PAN) started in Germany in the 1930's.⁸³ Parallel research was independently begun in the United States. There are two types of acrylics available in today's textile market: acrylic and modacrylic. An acrylic fiber is defined as a manufactured fiber in which the fiber-forming substance is any long-chain synthetic polymer composed of at least 85 wt% acrylonitrile units, which can be seen in Figure 3.24. Modacrylic fibers are defined as those composed of less than 85 but at least 35 wt% acrylonitrile units. The structure of modacrylic can be seen in Figure 3.25.

DuPont announced the first commercial acrylic fiber in 1949 under the trade name Orlon. The filament Orlon was first produced in 1950, with the staple length fiber reaching the market in 1952.⁸³ Because the acrylic fibers are 85 wt% polyacrylonitrile, many of their properties are determined by the chemical nature and physical behavior of long chain polyacrylonitrile molecules in an orientated structure. PAN molecules are described as stiff rods with a diameter of 0.6 nm, but are able to bend. A randomly

twisted and kinked polymer molecule view was given to PAN's by Bohn and coworkers⁹⁴ due to strong steric and dipolar repulsive forces between adjacent nitrile groups.⁹⁵

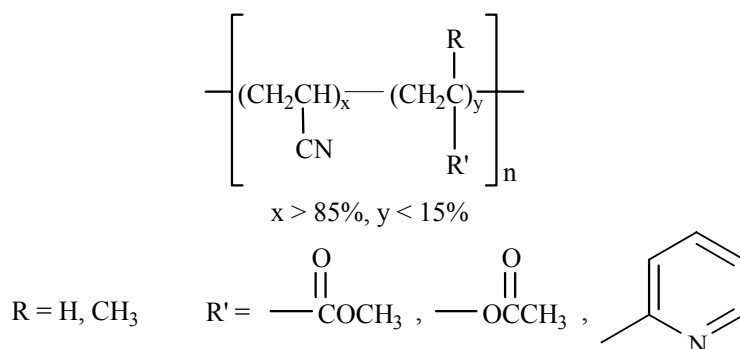


Figure 3.24. Chemical structure of acrylic fibers

Acrylic fibers are extremely resistant to direct sunlight, as compared to rayon, polypropylene, wool, cotton, and polyester. After a period of 19 months of direct sun exposure, the acrylic fibers had lost 50% of their overall strength. The same structural integrity loss in the other fibers occurred between two and a half and six and a half months. Acrylic fibers are also resistant to attack by insects and microorganisms found in soil. Most chemicals have little to no effect on acrylic fibers. Strong acids may cause fibers to swell; however, concentrated nitric or sulfuric acids will dissolve the fibers. Most organic solvents have a negligible effect on acrylics. Common oxidizers have little effect on these fibers.

The United States market uses a large amount of acrylics in home furnishing like carpets and rugs. It replaced wool because of increased resiliency and resistance to staining. Draperies, curtains, and awnings are also routinely made from acrylic fibers. The warmth, lightness, and ease of cleaning also make acrylics good for use in blankets.

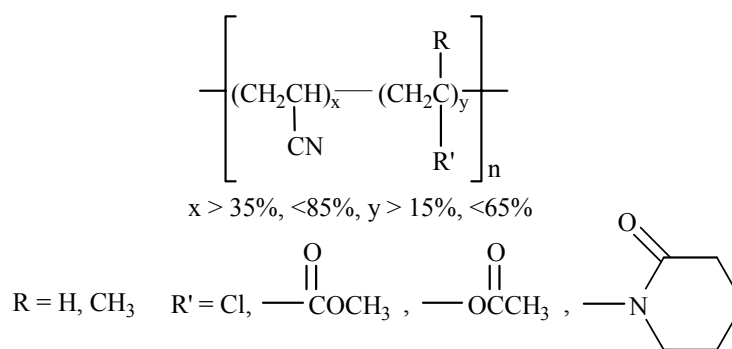


Figure 3.25. Chemical structure of modacrylic

The first modacrylic fiber was a result of the research efforts from Union Carbide on copolymers of vinyl chloride.⁸³ The continuous filament fiber was introduced in 1948, and the staple fiber was produced in 1950 by Dynel. Different trade names for modacrylics have included Eastman Chemical's Verel and Kanekalon from the Japanese company Kanegafuchi. Modacrylic fibers are similar in behavior to acrylic fibers. Due to the presence of halogen-containing monomers, the temperatures at which modacrylics are dimensionally stable are lower than those of acrylics. The halogen presence also changes the overall fiber properties including moisture regain, light stability, and resistance to degradation on heating.

Perhaps the most important property of modacrylics is their flame retardance. Consumer concern of the flammability of textiles has made this property of commercial significance. Modacrylic fibers, like acrylics, are highly resistant to biological agents such as insects and microorganisms. Though modacrylics are resistant to sunlight, they are not equal to the lightfastness of their acrylic cousins, due to the loss of hydrogen halides.⁸³ Because of their flame retardance, modacrylics are used in a wide range of applications. The largest amounts of modacrylics are used in apparel, such as sleepwear

for children and adults, and home furnishings. Wigs and industrial applications are also important areas of modacrylic use.

Determination of Textile Fibers

Forensic examinations of trace evidence are based upon Locard's Exchange Principle. The principle states that "every contact leaves a trace". This means, for instance, that if a person enters a room, they will either pick up something from the room, leave something behind in the room, or both. Problems can arise when the amount of material transferred is so minute it can not be accurately identified.⁸⁷

The methods generally used for the identification of fibers can be applied to both natural and manufactured fibers. However, some methods of analysis are more efficient and reliable when analyzing one class of fibers or the other.⁹² Microscopic examination can be used to readily identify most of the natural fibers. When identifying man-made or regenerated fibers, other testing methods must be used. Infrared spectroscopy is one of the preferred methods of analyzing manufactured fibers. Nevertheless, natural fibers are difficult to identify because more often than not, the basic cellulosic structure of the plant fibers and the amino acid polymers of animal fibers are too similar to offer definitive results.

A variety of testing methods can be performed on a given set of fibers or textiles, assuming there is enough of the sample. The combinations of different tests often provide enough information as to the generic class or specific type in a given sample. Examiners do not need to use every technique available to identify a sample. Positive identifications can be made without exhausting every resource available. An assortment of analysis methods frequently used in forensic textile analysis are highlighted below

with the first four being the routine methods.^{86, 96} All of the analytical methods described can be complimentary and need not be done alone.

1. Visual Inspection. The first method which most examiners would use to identify a textile would be visual inspection and hand analysis. The process involves looking at and feeling the sample. The length of the fibers is examined by untwisting the yarn and determining the overall length. All fibers can be made to be staple length; however, not all fibers can be filament length. Cotton and wool are both always staple length and never filament. The luster or lack thereof can be another factor observed through visual inspection. Manufactured fibers can have lusters ranging from matte, to dull, to shinny. The body, texture, and hand of a fabric can be related to the fiber size, surface contour, stiffness, and cross-sectional shape. Because manufactured fibers can resemble natural fibers, as well as other manufactured fibers, it is no longer possible to make a determination of fiber type solely on the basis of visual inspection.

2. Microscopy. The use of microscopy will be of use when the examiner has a working knowledge of the physical structure of the fibers. The physical characteristics of a fiber, such as longitudinal contours and cross-sectional shape, lead to the identification of an unknown fiber. The identification of natural fibers is best accomplished using microscopy.⁸⁶ Manufactured fibers are more difficult to identify because many of the man-made fibers look similar. Also, changes can occur during the manufacturing process to alter the shape and morphology of a fiber. Photomicrographs of an assortment of fibers can be seen in Figure 3.26.

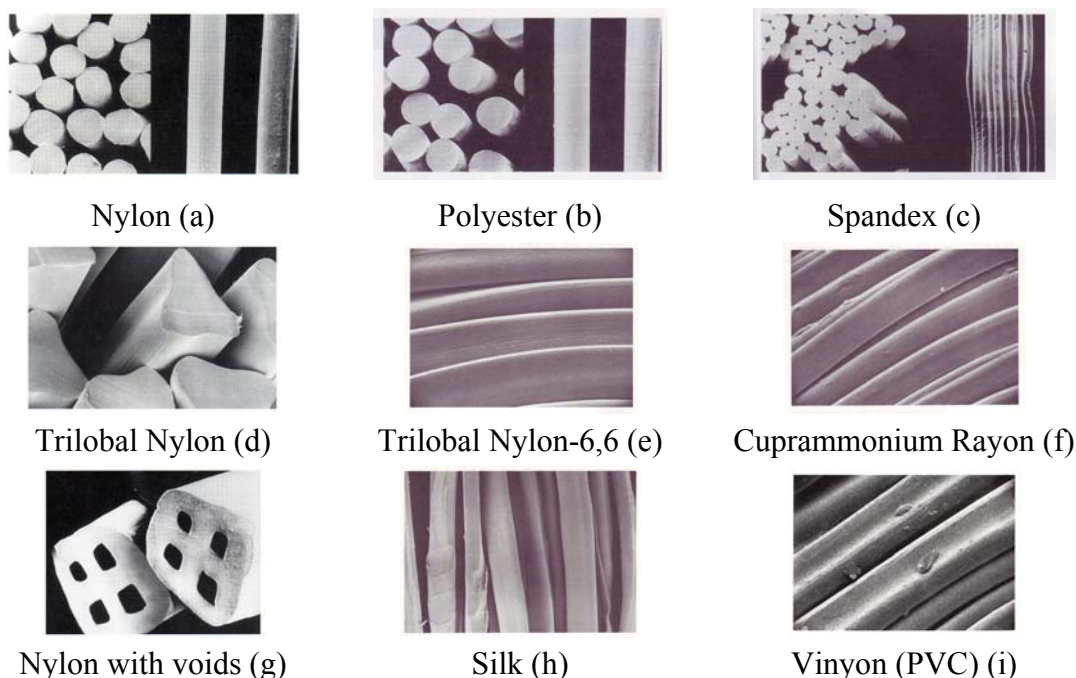


Figure 3.26. Comparison of photomicrographs for textile fibers⁸⁶

Figure 3.26a-c shows the cross-sectional and longitudinal photomicrographs of nylon, polyester, and spandex. All three fibers have a circular cross-section. The longitudinal scans show that nylon and polyester are smooth, yet spandex appears to be fluted. If the nylon and polyester were being analyzed, it would be impossible to show which was which, based solely on microscopy. The shape of the spinnerets used to make the individual fibers can be changed to produce a different type of fiber. Micrographs (a), (d), and (g) in Figure 3.26 show that nylon can be extruded to have at least three completely different cross-sections, making identification quite dubious. The remaining micrographs, (e), (f), (h), and (i) demonstrate that longitudinal scans also can lead to problems determining fiber type. The trilobal nylon-6,6 and cuprammonium rayon show similar morphology to one another. The natural fiber silk and the completely synthetic vinyon, also known as polyvinyl chloride, also resemble one another.

3. Burning Characteristics. The use of burn testing can give information as to the general chemical composition of a fiber (i.e. cellulose, protein, mineral, synthetic).⁸⁶ Textiles that are fiber blends cannot be identified by burn analysis.⁹² Visual inspection can be used in conjunction with burn testing. Burn testing requires a well ventilated area or hood vent to be used. To conduct the analysis, a collection of unknown fibers are held in a pair of tongs. The sample is then brought close to an open flame, usually from a Bunsen burner. The fibers' behavior as they approach the flame, while they are in the flame, and after they are removed from the flame is monitored. The characteristics of the smoke and ash produced, as well as the odor while burning is recorded. All of this data is then correlated and compared to known behaviors for different fibers.

As can be seen in Table 3.6, all of the cellulosic fibers, including the regenerated rayon, behave in an identical manner. The behaviors of all the fibers are the same before, while in, and out of the flame. Ash and odor are the same for all of the fibers. A similar trend can be observed for the other general classes of fibers, protein and mineral/synthetic. This demonstrates why the burn testing method is limited in scope and cannot be used for reliable identification on its own. Also, the method is inherently destructive as well as necessitating the need for proper facilities for performing the analysis.

4. Solubility Analysis. The use of chemical solvents to soften or dissolve different fibers as a means of identification has been done for many years.^{86, 92} Solubility testing is most useful for the identification manufactured fibers by their generic class and to

Table 3.6. Identification of textile fibers by burn testing⁸⁶

Fibers	Approach	In Flame	Remove	Ash	Odor
Cellulose					
Cotton	Does not fuse or shrink from flame	Burns with light gray smoke	Continues to burn, afterglow	Gray, feathery, smooth edge	Burning paper
Flax					
Lycocell					
Rayon					
Protein					
Silk	Curls away from flame	Burns slowly	May self-extinguish	Crushable black ash	Burning hair
Wool					
Mineral/Synthetic					
Acetate	Melts/pulls away from flame	Melts/burns	Continues to burn/melt	Brittle, black, hard bead	Acrid
Acrylic	Melts/pulls away from flame	Melts/burns	Continues to burn/melt	Brittle, black, hard bead	Chemical odor
Glass	No reaction	Does not burn	No reaction	Fiber remains	None
Modacrylic	Melts/pulls away from flame	Melts/burns	Self-extinguish, white smoke	Brittle, black, hard bead	Chemical odor
Nylon	Melts/pulls away from flame	Melts/burns	May self-extinguish	Hard gray/tan bead	Celerylike
Olefin	Melts/pulls away from flame	Melts/burns	May self-extinguish	Hard tan bead	Chemical odor
Polyester	Melts/pulls away from flame	Melts/burns	May self-extinguish	Hard black bead	Sweet odor
Saran	Melts/pulls away from flame	Melts/burns	May self-extinguish	Hard black bead	Chemical odor
	Melts/burns				
Spandex	Melts/does not pull away from flame	Melts/burns	Continues to burn/melt	Soft black ash	Chemical odor

confirm the identification of natural fibers. A fiber sample is placed into a beaker along with a particular solvent at a set temperature and stirred for 5 minutes. A variety of solvents are available in increasing strength. The fibers in question should be placed in the different solvents in order of increasing strength to be sure proper identifications are made. The list of solvents routinely used in solubility testing, in order of increasing strength can be seen in Table 3.7.

Table 3.7. Solubility tests for fibers⁸⁶

Solvent	Temp. (°C)	Fiber dissolved
Acetic acid, 100%	20	Acetate
Acetone, 100%	20	Acetate, modacrylic, vinyon
Hydrochloric acid, 20%	20	Nylon-6, nylon-6,6, vinal
Sodium hypochlorite, 5%	20	Silk, wool, azlon
<i>m</i> -Xylene, 100%	139	Olefin, saran, vinyon
Dimethyl formamide, 100%	90	Spandex, modacrylic, acrylic, acetate, vinyon
Sulfuric acid, 70%	38	Cotton, flax, rayon, nylon, acetate, silk
<i>m</i> -Cresol	139	Polyester, nylon, acetate

5. *Pyrolysis Gas Chromatography/Mass Spectrometry.* Gas chromatography has been used as another method of fiber analysis.⁸⁷ The sample is pyrolyzed by a thin wire suspended in the GC column. When an electric current is passed through the wire, the sample is heated and vaporized. The pyrolyzed products of the combustion are then carried onto the GC column to be separated and analyzed.^{97, 98} The resulting 'pyrogram' produced by this analysis is characteristic of a specific fiber type and can be used as a fingerprint in the identification of polymers and fibers.

6. *Thermal Analysis.* The physical and chemical changes in fibers are also investigated by the measurement of changes in selected properties as samples of fibers are heated at a steady rate over a given temperature range in an inert atmosphere, such as argon or nitrogen. The four thermal characterization methods used are: differential thermal analysis (DTA), differential scanning calorimetry (DSC), thermal gravimetric analysis (TGA), and thermal mechanical analysis (TMA).⁸⁵

Differential thermal analysis monitors physical and chemical alterations of matter. These changes can be detected by measuring temperature differences arising between a sample and a thermostable, inert reference material. The sample and reference materials are heated or cooled at a constant rate of change. The temperature range for these analyses can range anywhere from -150 °C to 1000 °C with a rate of change between 5-10 °C. Total analysis time to get a thermogram is dependent on the temperature range and rate of change. These methods all necessitate some degree of sample preparation before the analysis can be conducted. The chemical and physical events that are monitored in the samples include changes in crystallinity or crystal structure, loss of water, solvents, or other volatile materials, melting, or decomposition of the fiber.⁸⁵ Differential scanning calorimetry is similar to DTA; however, DSC measures changes in heat capacity (ΔH) rather than temperature (ΔT) as the fibers are heated. This analysis method provides quantitative insight on the thermodynamic processes involved in the fiber and the environment it resides in. Thermal gravimetric analysis monitors the changes in mass of the fibers (ΔM) as the temperature is increased at a constant rate. TGA offers information about the loss of volatile materials, the rate and mode of sample decomposition, and the effect of finishes on the fiber's decomposition. TMA measures

the change in a specific mechanical property, such as hardness and flow under stress, as the temperature of the fiber is raised at a constant rate.

7. Infrared Spectroscopy. The use of infrared spectroscopy is not usually needed when the fibers under investigation are natural fibers. The use of microscopy, chemical testing, and staining techniques can regularly make a valid determination. If two manufactured fibers have the same basic monomer unit and the same copolymer, it may not be possible to identify the fiber by use of IR spectroscopy.⁹² There are different approaches by which infrared spectroscopy can be used to analyze a fiber. The first is by pressing the fibers into a potassium bromide pellet. This will afford a qualitative identification even with minute samples. Another approach is the determination of the attenuated internal reflectance of the fiber samples.

8. Near-Infrared Spectroscopy. Raw materials as well as finished textile products can be characterized using near-infrared spectroscopy (NIRS). NIR methods continue to diversify and be applied to a variety of wide-ranging quantitative and qualitative analyses. The quantitative techniques allow for the rapid, accurate monitoring of chemical, physical, and morphological properties of a fiber, yarn, fabric, or chemical textile auxiliary.⁹¹ NIRS analyses are useful due to their speed, accuracy, and precision, without having to destroy the sample being analyzed. Diffuse reflectance measurements can be made since the samples in question are, for the most part, solids. This also removes the time consuming and often times expensive sample preparations necessary for textiles.

In the processing of cotton, reducing sugars present in the fibers, such as glucose, mannose, fructose, and pentoses, can accumulate on the surfaces of the equipment. Depending on the relative humidity and temperature, the sugars can become extremely sticky, leaving a coating on the finished fibers and in the machinery. Perkins and coworkers devised a method of determining sugars by means of a complex titration analysis known as Perkin's method.⁹⁹ An aqueous extract of cotton is reacted with potassium ferricyanide in the presence of sodium carbonate. The potassium ferricyanide oxidizes the reducing materials and is itself reduced to potassium ferrocyanide. The quantity of potassium ferrocyanide produced is determined by titrating with ceric sulfate in an acidic solution containing *o*-phenanthroline/ferrous sulfate complex (ferroin) as an indicator.

Though Perkin's method is accurate, it involves large amounts of sample preparation and actual analysis time. NIR techniques have been developed to determine the amounts of these compounds in cotton fibers, so that excessive levels of sugars, over 0.5%, can be removed or reduced before production begins.⁹¹ These methods are streamlined and user friendly. Fabric blends have also been analyzed by NIRS. Polyester and cotton are often combined in various ratios to afford a finished textile that has the soft characteristics of cotton, while at the same time, the durability and resiliency of polyester. Traditionally, the process of determining the ratio of polyester to cotton was by dissolving the fibers in 70% sulfuric acid. This analysis required 8 hours before the final results were known. Near-infrared reflectance analysis (NIRA) can determine the fiber blend composition within two minutes.

Mercerization is a process by which fabrics are treated with 20% sodium hydroxide under tension. The mercerization process is done to enhance the finished fabric's dye affinity, dimensional stability, tensile strength, and luster.^{100, 101} The degree of mercerization must remain constant to ensure the final product is of the proper quality. The most common problem, associated with changes in the degree of mercerization, is dye shade variability.⁹¹

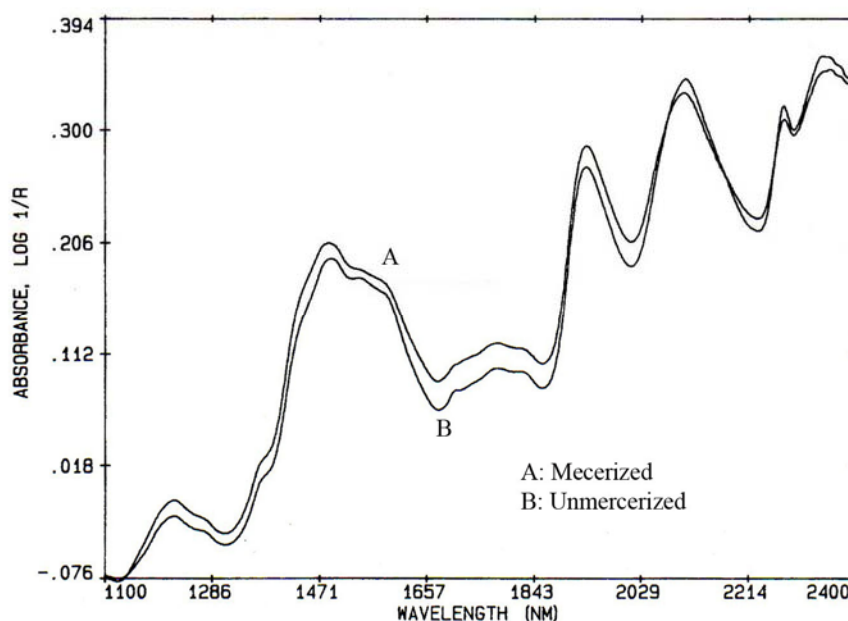


Figure 3.27. Near-infrared reflectance spectra of mercerized and unmercerized cotton samples (Modified from reference 91. Copyright 2001 Marcel Dekker Inc.)

The conventional method of mercerization testing is barium-activity number measurements. The process begins with the both mercerized and unmercerized fabrics being boiled in soap and soda ash. The fabrics are then washed until they are at a neutral pH. Next, both samples are treated with a barium hydroxide solution. The ratio of the amount of barium hydroxide absorbed by the two different samples is determined by

titration. The total time required for this analysis was approximately 6 hours. The NIRA for both mercerized and unmercerized cotton can be seen in Figure 3.27. The hydroxyl group overtone region near 1471 nm shows a more pronounced band in the mercerized sample.

9. Elemental and End Group Analysis. Both quantitative and qualitative information concerning chemical elements and end groups can be derived from fibers to elucidate the fabric class. The presence of process residues, dyes, and other finishes can affect the nature and elemental composition that could skew the results.⁸⁵ A variety of gravimetric and instrumental methods are available for both elemental and end group analysis. The determination of amino acids in protein fibers, amino acids in polyamides and proteins, and acid groups in polyamides and polyesters can aid in structure analyses, molecular characterization, and the identification of fibers.

CHAPTER FOUR

Soft Independent Modeling of Class Analogy for Textile Classification by Diffuse Near-Infrared Reflectance Spectroscopy

Introduction

In today's global business community, the textile and clothing industries make up a substantial portion of the domestic and international trade marketplace. The textile industry is expected to grow to a staggering \$320 billion in retail sales this year alone. With such a large investment of money in this industry, manufacturers and retailers demand assurances that the raw materials, textiles, and finished garments they are purchasing to provide to their customers are authentic. The Federal Trade Commission Act requires all clothing retailers and textile manufacturers to run identification tests on all fabrics sold in the United States. Prices for different fabrics can range from less than one dollar to several hundred dollars a yard. Cotton Inc. levies a tariff on all imported bales of foreign cotton entering the United States. This revenue is then used to promote the purchase of cotton materials in television and print media. The cotton entering the country must be screened to determine its authenticity. At international borders and customs checkpoints, travelers entering a country with textiles must also have those items screened to determine legitimacy and assess whatever import fees or taxes might apply. In the past, such screenings could take days or even weeks before someone would have their fabric released by customs officials. And perhaps of paramount importance, fiber classification and characterization is a crucial element in many criminal investigations.⁹⁶

The mission statement of the Baylor University Department of Family and Consumer Sciences is to prepare students to become professionals in business, education, and service careers that use the family and consumer sciences knowledge base and to assume individual, family, and community roles in an increasingly complex global society. The programs offered by Family and Consumer Sciences (FCS) department include fashion design and merchandising, as well as interior design, in which textile materials are used constantly. A fast, non-destructive method of fiber analysis would be quite beneficial to the FCS department.

The Baylor Center for Analytical Spectroscopy was approached and agreed to enter into a collaborative effort with FCS. Both entities provided expertise that was invaluable to the success of the project. Dr. Judith Lusk, professor of fashion merchandising and design, provided the textile samples, as well as technical information concerning the manufacturing and finishing techniques associated with textiles. The Center for Analytical Spectroscopy provided the instrumentation and technical knowledge of both near-infrared spectroscopy and chemometrics.

Chemometric classification analyses are more often used when the examiners wish to know if a certain compound is present in a given sample, without necessarily determining its concentration in the sample matrix. The identification of a textile fiber is an excellent example of where a classification analysis would be beneficial. A common classification technique is soft independent modeling of class analogy (SIMCA).

The purpose of both the preliminary and expanded studies reported here was to investigate the feasibility of using SIMCA on the diffuse near-infrared spectra of textile samples in order to classify future samples. The Baylor NIR textile database has grown

to over 800 samples and spectra. The creation and use of such a large and diverse NIR spectral library to perform classification analyses on textiles has not been reported

Background

Near-Infrared Spectroscopy

The discovery of the near-infrared (NIR) region of the electromagnetic spectrum has long been attributed to Sir William Herschel in 1800. The noted astronomer, formerly a musician, authored two papers detailing his work on the heating effect in the spectrum of solar radiation.¹⁰² Herschel used a large glass prism to disperse sunlight onto three thermometers that had carbon-blackened bulbs. A heating effect was observed in the study towards the red end of the spectrum. However, beyond the visible red light from the prism, where no visible light was seen, the temperature recorded by the thermometer was at its highest. The heating effect was termed radiant heating and the thermometrical spectrum by Herschel.

In 1835, Ampere utilized a thermocouple to show that near infrared has the same optical properties as visible light. He concluded that they both were the same phenomenon and gave birth to the concept of an extended spectrum, beyond the ordinary visible. Abney and Festing were able to record and interpret the first true NIR spectrum of organic liquids from 1-1.2 μm in 1881. Further headway into the understanding of the NIR was made when W.W. Coblentz was able to record the spectra of hundreds of compounds in the 1-15 μm wavelength region. Coblentz observed that no two chemical compounds had the exact same NIR spectrum even if they contained the same elements in the same proportions, such as with different isomers.

The development of methodologies for using NIR measurements is often more laborious and quite different from those in the ultraviolet/visible (UV-vis) and infrared (IR) due to hydrogen bond shifts that dominate in the NIR, as well as the fact that interactions and nonlinearities show nearly a total disregard for Beer's law.

1. Spectral Characteristics. The fundamental concept to remember with vibrational spectroscopic methods is that the atom-to-atom bonds within molecules vibrate at discrete frequencies (Figure 4.1). These vibrations can be described by the laws of physics, and the different resonant frequencies can be calculated mathematically. Most molecules, while at room temperatures, will only vibrate in their least energetic state that is dictated by quantum mechanics. Molecules in this state are said to be at rest.¹⁰² The fundamental frequency at which any two atoms connected by a chemical bond can be roughly calculated by making the assumption that the bond energies arise from the vibration of a diatomic harmonic oscillator and obey Hooke's Law. This can be seen in equation 4.1:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (4.1)$$

where ν is the vibrational frequency, k is the force constant, and μ is the reduced mass for the two atoms in the bond, which is given by:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (4.2)$$

where m_1 and m_2 are the respective masses for the two atoms connected by the chemical bond in question.

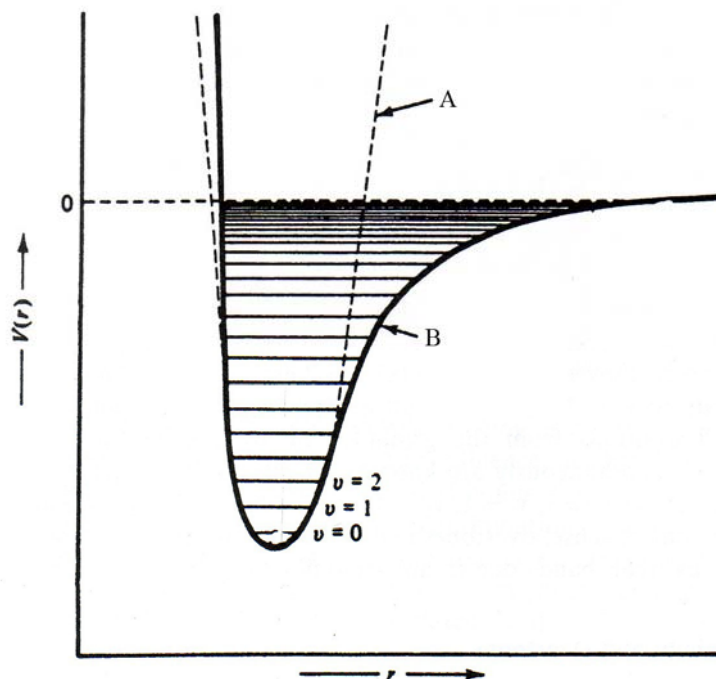


Figure 4.1. Energy diagram of vibrational mode calculated as (a) an ideal diatomic oscillator or (b) as an actual anharmonic diatomic oscillator (Modified from reference 102. Copyright 2001 Marcel Dekker, Inc.)

This approach works well for the fundamental vibrational frequency of simple diatomic molecules and is not too far divergent from the average value of a two-atom stretch within a polyatomic molecule. This approximation only gives the average or center frequency of the diatomic bond. Since the reduced masses of the C-H, O-H, and, N-H are 0.85, 0.89, and 0.87 respectively, one might expect the ideal frequencies of these pairs to be quite similar. These pairing also constitute the major absorption bands of the NIR spectrum. The electron withdrawing and donating properties of neighboring atoms and groups in real molecules greatly influence the bond strength and length, thus affecting the frequency of the X-H bonds.

The classical spring model for molecular vibrations has a continuum of energy levels; however, molecular vibrations have discrete energy levels as described by

quantum theory. The quantum mechanical treatment requires that the energy E , at any given time, be restricted to specific fields according to:

$$E_v = (v + \frac{1}{2})h\nu \quad (v = 0, 1, 2, \dots) \quad (4.3)$$

where v is the vibrational quantum number, h is Plank's constant, and ν is the frequency of vibration. When the molecule in question is a non-linear polyatomic molecule, the equation can be generalized as:

$$E(v_1, v_2, v_3 \dots) = \sum_{i=1}^{3N-6} (v_i + \frac{1}{2})h\nu_i \quad (4.4)$$

where $v_1, v_2, v_3 = 0, 1, 2, \dots$

When a vibrational transition occurs from $v=0$ to $v=1$ in any of the vibrational states ($v_1, v_2, v_3 \dots$), the transition is considered a fundamental transition, permitted by selection rules. A vibrational overtone occurs when the transition occurs from the ground state to a state $v=2, 3, 4, \dots$. The majority of the overtone peaks seen in the NIR spectrum arise from the X-H stretching modes of O-H, C-H, S-H, and N-H due to energy considerations. A selection of group frequencies in the near-infrared region can be seen in Table 4.1. Because these transitions are quantum mechanically forbidden, overtones are commonly seen to be 10 to 1000 times weaker than the fundamental vibrational bands.¹⁰² When a band arises from a bending or rotating atoms, which have lower energies than stretching vibrational modes, it would not be until the third or fourth overtone that it would be seen in the NIR spectrum. Large numbers of combination bands are prominent features of the NIR. Combination bands are those arising from movement from the ground state to a state for with $v_i = 1$ and $v_j = 1$. An example of combination and overtone bands for chloroform can be seen in Table 4.2.

Table 4.1. Near-infrared spectral bands

Wavelength (nm)	Bond Vibration	Structure
1450	O-H Stretch first overtone	Starch
1463	N-H Stretch first overtone	CONH ₂
1500	N-H Stretch first overtone	NH
1540	O-H Stretch first overtone	Starch
1620	C-H Stretch first overtone	=CH ₂
1695	C-H Stretch first overtone	-CH ₃
1705	C-H Stretch first overtone	-CH ₃
1900	C=O Stretch second overtone	-CO ₂ H
1940	O-H Bend second overtone	-H ₂ O
1990	N-H Stretch/bend combination	Urea
2180	N-H Bend second overtone	Protein
2310	C-H Bend second overtone	Oil
2380	C-H Stretch/C-C stretch combination	Oil
2470	C-H Combination	CH ₂

2. *Diffuse Reflectance Spectroscopy.* The use of near-infrared reflectance for the quantitative analysis of many products and commodities has gained wide acceptance in recent years. Different algorithms have been developed to achieve multi-component determinations from the diffuse reflection spectra of powdered samples. The strict adherence to a linear dependence of band intensity is not absolutely mandatory for an analytical result to be obtained. One algorithm that is frequently used is $\log 1/R$, where R is the reflectance of a sample. The use of $\log 1/R$ as a preferred ordinate is contrary to what most physical scientists would consider appropriate for a diffuse reflection measurement. The understanding of the basic theories of diffuse reflection and the

validity of the assumptions for each would be helpful in understanding the strengths and limitations of diffuse NIR.¹⁰²

Table 4.2. NIR bands of Trichloroethane¹⁰³

Assignment ^a	Wavenumber (cm ⁻¹)	Wavelength (nm)
$\nu_1 + 2\nu_4$	5373	1861
$2\nu_1$	5911	1692
$2\nu_1 + \nu_4$	7091	1410
$3\nu_1$	8680	1152

The phenomenon of diffuse reflection is observed in everyday life, such as when sunlight is reflected from a matte surface. This observable fact led Lambert to be the first to attempt a mathematical description of diffuse reflection. Lambert proposed that the radiation flux, I_r , in an area $f \text{ cm}^2$, and solid angle ω steradians (sr), is proportional to the cosine of the angle of incidence α and the angle of observation, ϑ , in the form:

$$\frac{dI_r / df}{d\omega} = \frac{CS_o}{\pi} \cos \alpha \cos \vartheta = B \cos \vartheta \quad (4.5)$$

where S_o is the irradiation intensity in W/cm^2 for normal incidence, B is the radiation density or surface brightness in $\text{W cm}^{-2} \text{ sr}^{-1}$, and the constant C is the fraction of the incident radiation flux that is remitted, reflected. C is less than 1 because some of the incident radiation is always absorbed. Eq. 4.5 is known as the Lambert cosine law and can be derived from the second law of thermodynamics. An ideal diffuse reflector has never been found in practice, therefore deviations, whether large or small, always occur from the Lambert cosine law. There have been many studies trying to prove or disprove

the Lambert cosine law.⁹¹ It was found, in general, that the law only holds when both the angle of incidence α and the angle of observation ϑ are relatively small.

Mie scattering is another widely accepted general theory concerning the scattering of light. The phenomenon is related primarily to the scattering of radiation by isolated particles. Mie's efforts led to a series of equations that describe the angular distribution of both the intensity and the polarization of scattered radiation for a plane wave scattered once by a particle. The particle in question was spherical and had no limitation imposed on its size. The primary equation developed by Mie is as follows:

$$\frac{I_{\theta s}}{I_0} = \frac{\lambda^2}{8\pi^2 R^2} (i_1 + i_2) \equiv q(\theta_s) \quad (4.6)$$

where $I_{\theta s}$ is the scattered intensity at a distance R from the center of the sphere, I_0 is the intensity of the incident radiation, and λ is the wavelength of the incident radiation. The values of i_1 and i_2 are functions of the angle of the scattered radiation. The Mie theory, although applicable for spherical particles of any size, is only valid for single scattering and therefore only applicable to chemical systems where particles are well separated.

Most theories attempt to describe diffuse reflection of radiation by means of a radiation transfer equation. In simplest terms, the equation can be stated as:

$$-dI = \kappa \rho I dS \quad (4.7)$$

where dI describes the change in intensity of a beam of radiation of a given wavelength in a sample whose density is ρ and for which the pathlength is dS . The attenuation coefficient, κ , is associated with the total radiation lost, whether that loss is due to scattering or absorption. The first attempt to find a simplified solution to the radiation transfer equation was found by Schuster in 1905. The solution was arrived at by first

making the assumption of using two oppositely directed radiation fluxes, I and J . The forward traveling radiation was designated as I . The following two differential equations were derived using this simplification:

$$\frac{-dI}{d\tau} = (k + s)I - sJ \quad (4.8)$$

$$\frac{dJ}{d\tau} = (k + s)J - sI \quad (4.9)$$

where

$$k = \frac{2\alpha}{\alpha + \sigma} \quad (4.10)$$

and

$$s = \frac{\sigma}{\alpha + \sigma} \quad (4.11)$$

In equations 4.10 and 4.11, α is the true absorption coefficient of single scattering and σ is the scattering coefficient for single scattering. The abbreviation s used by Schuster is identical to the albedo ω_0 for single scattering.⁹¹ When equations 4.8 and 4.9 are solved, they give the reflectance at “infinite depth”, which is defined as the depth at which a sample must be in order to have no further change in the measured reflectance. Schuster’s final equation is:

$$\frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} \quad (4.12)$$

where R_∞ is the diffuse reflectance.

Table 4.3. Assumptions of the Kubelka-Munk Equation¹⁰²

No.	Statement
1	The radiation flux (I and J) travels in two opposite directions
2	The sample is illuminated with monochromatic radiation of intensity I_0
3	The distribution of scattered radiation is isotropic so that all regular (specular) reflection is ignored
4	The particles in the sample layer (defined as the region between $x = 0$ and $x = d$) are randomly distributed
5	The particles are very much smaller than the thickness of the sample layer d
6	The sample layer is subject only to diffuse irradiation
7	Particles are much larger than the wavelength of irradiation (so that the scattering coefficient will be independent of wavelength), although if only one wavelength is to be used then this assumption is not relevant
8	The breadth of the macroscopic sample surface (in the yz plane) is great compared to the depth (d) of the sample and the diameter of the beam of incident radiation (to discriminate against edge effects)
9	The scattering of particles are distributed homogeneously throughout the entire sample

A similar solution to Schuster's was derived by Kubelka and Munk (K-M) in 1931.¹⁰² The primary difference was in the definition of the two constants k and s . Schuster defined these constants in terms of the absorption and scattering coefficients for single scattering. K-M theory simply defined k and s in their equations as the absorption and scattering coefficients for the densely packed sample as a whole. K-M theory made a series of assumptions in their theoretical considerations that can be seen in Table 4.3. A schematic representation of the Kubelka-Munk equation can be seen in Figure 4.2.

The general Kubelka-Munk equation is expressed as:

$$f(R_\infty) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} \quad (4.13)$$

where R_∞ is the absolute reflectance of the layer, k is the molar absorption coefficient, and s is the scattering coefficient.

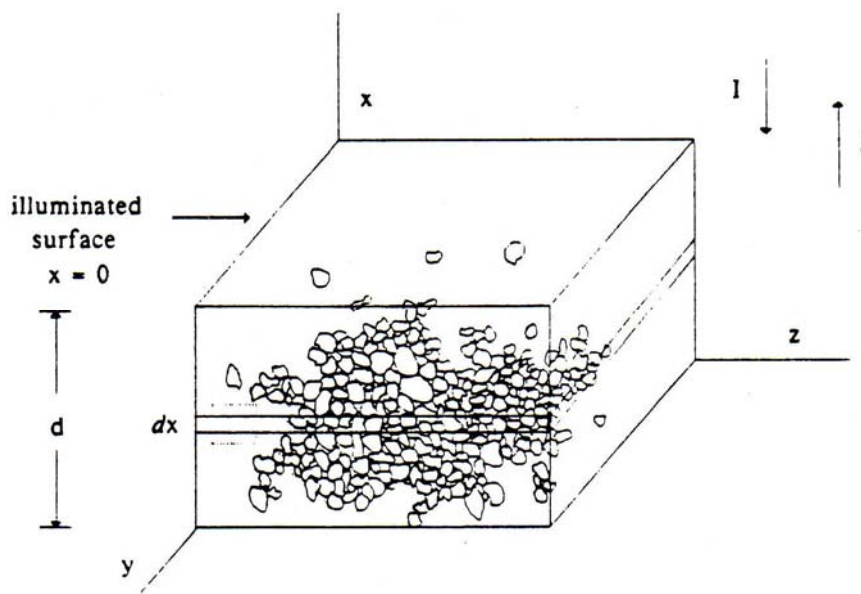


Figure 4.2. Schematic representation of a sample for which Kubelka-Munk equation was derived (Modified from reference 102. Copyright 2001 Marcel Dekker, Inc.)

Kubelka-Munk theory predicts that a linear relationship will exist between spectral data and sample concentration under conditions of a constant scattering coefficient and infinite sample dilution in a non-absorbing matrix. Therefore, the K-M theory can only be strictly applied under the stated conditions. The K-M equation only applies to an infinitely thick sample layer, which is approximately 3 mm for NIR spectroscopic determinations. The work of Humphrey and coworkers demonstrated that the comparison between the Kubelka-Munk and $\log I/R$ calculations of NIR reflectance showed no clear preference for one method over another.¹⁹

3. Near-Infrared Analysis of Textiles. Use of near-infrared analysis for textiles has been growing over the past two decades. It can be used to characterize both raw materials and finished textile products. Both qualitative and quantitative analyses have been conducted using NIR spectroscopy.⁹¹ In 1994, Jasper and coworkers pioneered the

idea of using neural networks and NIR spectrophotometry to identify 17 categories of fibers including both natural and man made.¹⁰⁴ The study showed that the neural network could, with a fair degree of certainty, categorize unknown fibers into their respective groups. Confidence levels for all but one category were a minimum of 94%. Cotton had only a 51% confidence level, and further examination of the data showed that cotton showed 32% chance of being said to contain linen. The misclassification can be traced to the inherent similarities of the structures (Chapter 3), which lead to similar spectra.

Cleve and colleagues used chemometric methods in conjunction with near-infrared spectroscopy in order to determine not only fabric type, but also such factors as moisture content and textile coatings.¹⁰⁵ The study used seven fabric types as well as a number of two component blends of those fibers. The blends' spectra were analyzed by use of the PLS-2 algorithm. Sample pretreatment in this research involved pulverizing the fibers after they had been frozen in liquid nitrogen.

A comparison of several vibrational spectroscopic methods including FT-Raman, mid-infrared attenuated total reflectance, mid-infrared diffuse reflectance, and near-infrared diffuse reflectance to determine the composition of wool/polyester blends was completed by Church et al.¹⁰⁶ When the diffuse near-infrared reflectance data were analyzed, it was found that a partial-least squares (PLS) approach was superior to the classical least squares approach familiar to most statisticians, because a PLS analysis examines spectral regions and not a singular wavelength. The PLS algorithm is an outgrowth of multiple linear regression. A regression as opposed to a classification is more appropriate for this research study due to the fact that the samples analyzed had varying compositions, and there was not a uniform fiber makeup between all samples. A

partial-least squares regression model takes into account different minute spectral variations and processes them to identify the important information to be retained. Interferences can come from many factors including dyes, finish resins, and weave patterns.

The use of near-infrared and chemometrics in the analysis of fiber blends continues. In August 2005, Sohn and coworkers reported on the use of FT-NIR spectroscopy for determining the linen/cotton content of different textiles.⁹⁶ As with the previous work done by Church, this study utilized a partial-least squares analysis of the near-infrared spectra. The models created and validated showed less than a 3% error of prediction.

Methods and Reagents

Preparation of Textile Samples

Commercially available textiles were attained from Dr. Judith Lusk of the Baylor University Department of Family and Consumer Science. Samples approximately one-inch square were cut from the textile swatches provided. The samples were then placed in acid-free polypropylene sleeves containing a 4 x 6 inch index card marked with the sample's fiber type and a numerical designation which indicated the samples position in the data matrix.

Near-Infrared Spectroscopy

NIR analysis of textile samples was carried out on a dual-channel spectrometer previously assembled and tested in our laboratory.¹⁸ The radiation source for the spectrometer was a quartz tungsten-halogen lamp. The radiation from the source lamp

was passed through a long-pass filter with a cut-off of 1400 nm before being modulated with a rotary chopper and focused onto the entrance slit of the monochromator. The exit slit of the monochromator was mounted onto a custom-built sample compartment. The radiation from the exit slit was collimated before being divided by a beam splitter into a reference and sample beams. The beam was reflected off a mirror within the sample compartment and entered an eight-inch diameter integrating sphere through the sample port. The diffuse reflectance signal was detected by a lead sulfide detector mounted 45° from the sample window, and passed into a lock-in amplifier that was referenced to the frequency of the rotary chopper. The signal from the lock-in amplifier, in analog form, was then digitized by a 16-bit analog-to-digital converter and passed to the instrument computer. The instrument control and data processing were carried out by a LabView software graphical user interface, which controlled the stepping of the monochromator and acquisition of the signal from the lock-in amplifiers. A diagram of the instrumental setup is shown in Figure 4.3.

The textile samples were placed in the sample holder at the sample window of the integrating sphere. NIR reflectance scans were run from 1100 to 2200 nm, with a sampling interval of 2.0 nm. Measurements were reported as $\log 1/R$, where R is equal to the reflectance. Data were written into ASCII tables by the LabView software and converted to Microsoft® Excel spreadsheets used in chemometric analysis.

Chemometric Analysis

The spectral data, contained in the Excel spreadsheet files, were imported into the Unscrambler© software package (versions 9.1 and 9.6 – CAMO, Inc., Woodbridge, NJ). The data were then transposed to list the wavelengths as the x -variables and the samples as the y -variables. The multivariate principal components analysis was performed on the data. Raw spectral data was subjected to a Savitzky-Golay smoothing routine, using a zero-degree polynomial with 5 averaging points on both ends.

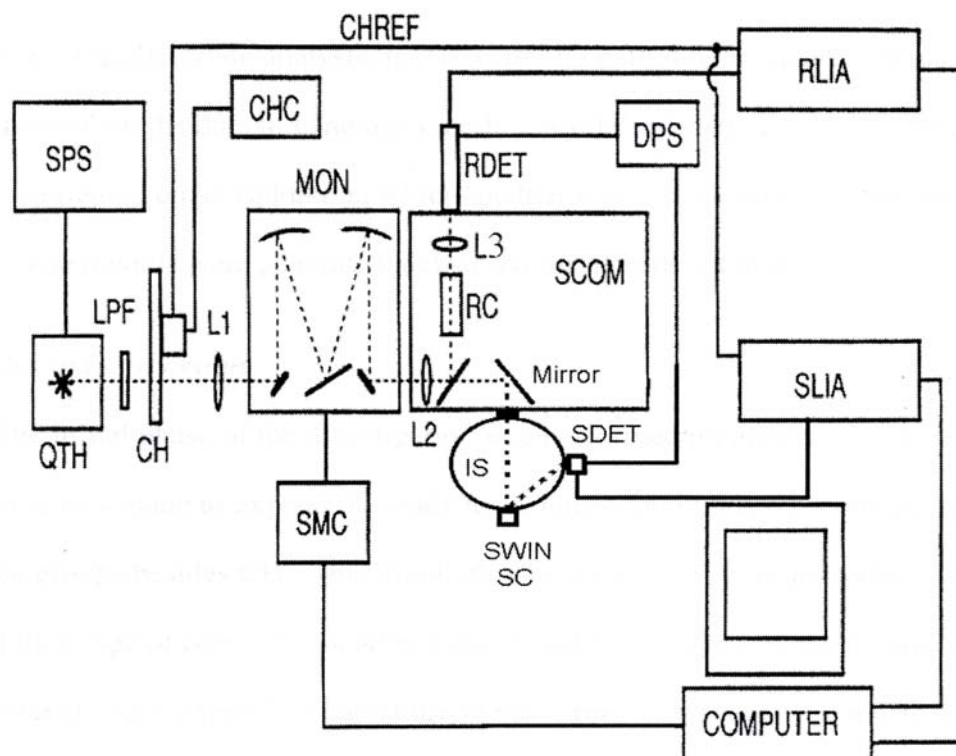


Figure 4.3. Schematic of dual-channel near-infrared spectrometer, SPS = spectral power supply, QTH = quartz tungsten halogen lamp, LPF = long pass filter, CH = chopper, CHC = chopper controller, MON = monochromator, CHREF = reference channel, RDET = reference detector, SDET = sample detector, DPS = diode power supply, SCOM = sample compartment, RC = reference channel, SC = sample cell, L1-L3 = lenses, RLIA = reference lock-in amplifier, SLIA = sample lock-in amplifier, IS = intergrating sphere, SWIN = sample window (Modified with permission from reference 18. Copyright 2000 Olusola Soyemi)

Results and Discussion

Preliminary Study

The complete spectra for the 114 samples in the preliminary study can be seen in Figure 4.4. The high degree of signal noise at the wavelength extremes is due to the photoconductive lead sulfide detector used by the spectrometer. The spectroscopic sensitivity of lead sulfide detectors drops off at their wavelength limits of 1.0 and 2.5 μm . For this reason, the wavelength region selected for these studies was from 1334 to 1906 nm, where the signal-to-noise ratio is substantially higher.

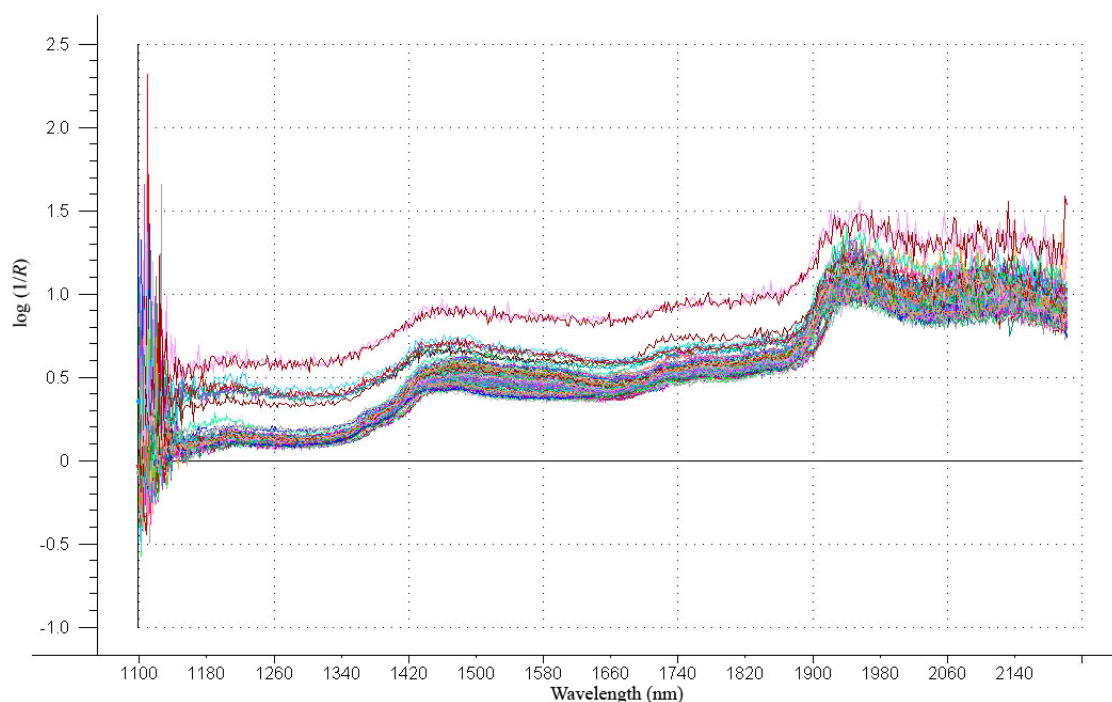


Figure 4.4. Diffuse near-infrared reflectance spectra for 114 textile samples from 1100 to 2200 nm

The inherent spectral characteristics of the fiber samples must also be considered when selecting the wavelength region to be used in the principal components analyses. For a classification analysis based on SIMCA to be plausible, the samples in the data sets

must segregate themselves in multivariate space to a degree that hyper-dimensional regions can be erected around the sample clusters. The principal NIR bands for cotton and polyethylene terephthalate, PET polyester, can be seen in Table 4.4. Many of the spectral characteristics for both fibers can be seen in the selected wavelength window of 1334 to 1906 nm. Cotton and polyester are among the two most important fibers in the world. As such, the characteristic spectral bands for these samples must be included in the principal components analysis.

Table 4.4. Principal bands in NIR spectra of cotton and polyester terephthalate¹⁰²

Cotton		Polyester Terephthalate		
λ_{\max} (nm)		λ_{\max} (nm)		
1216	1708	1128	1800	2156
1270	1776	1172	1828	2184
1372	1824	1368	1868	2256
1444	1930	1412	1908	2336
1490	2104	1616	1952	2396
1550	2276	1660	2088	
1590	2340	1716	2132	

A principal components analysis was conducted using the diffuse near-infrared reflectance spectra for the cotton and polyester samples in the data set. There were no sample pretreatments, as well as no manipulation of the raw spectral data collected on the spectrometer. The wavelength region of 1334 to 1906 nm was used (Figure 4.5). A total of six principal components were calculated using full cross-validation. The result of the principal components analysis can be seen in Figure 4.6. The two-dimensional scores plot (Figure 4.6a) shows that the samples have segregated into two distinct groups. The factor discriminating the samples in each group is the fiber type. The segregation can

also be seen in a three-dimensional scores plot (Figure 4.7) from the same analysis. A border has been superimposed on the samples in the different groups, as is done in a SIMCA classification. This result demonstrates that two fibers, one a natural and the other a manufactured, can be distinguished by using a principal components analysis. The separating power can also be attributed to the fact that the polymers involved, cellulose and substituted aromatics for cotton and polyester, respectively, are chemically dissimilar. With the naked eye, a prominent spectral band between 1668 and 1684 nm is seen, which is indicative of polyesters, but totally lacking in the spectra for cotton.

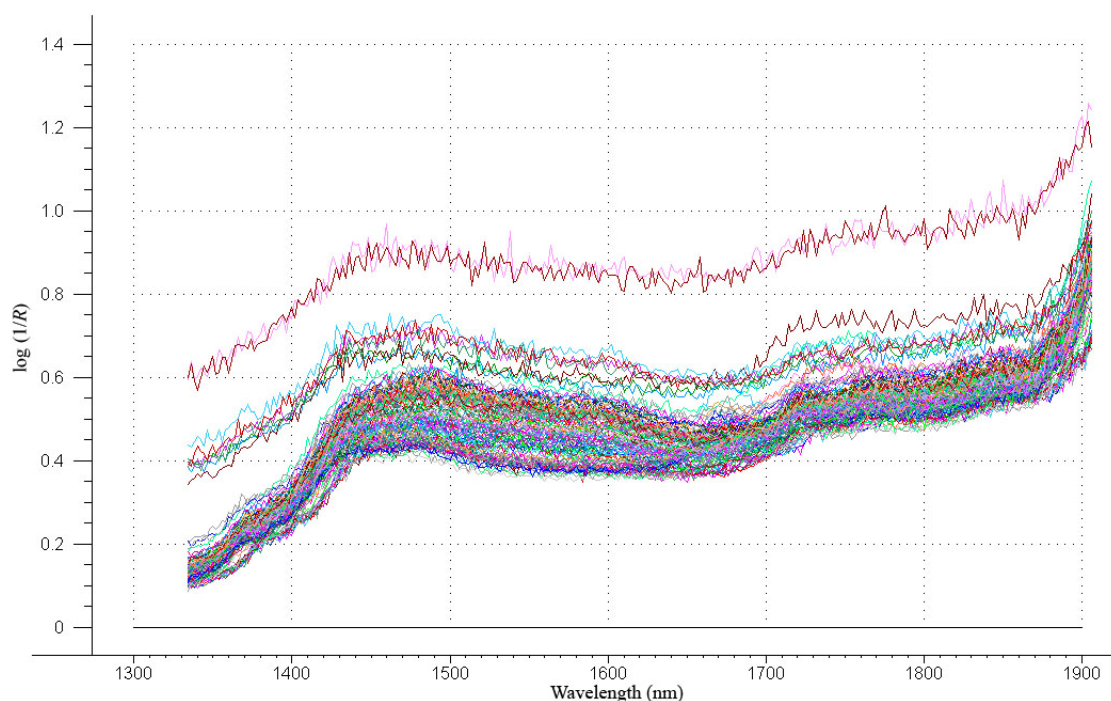


Figure 4.5. Diffuse near-infrared reflectance spectra for 114 textile samples depicting the wavelength window of 1334 to 1906 nm.

The next question to be answered is whether or not the principal components analysis can distinguish, not just between a natural and manufactured fiber, but also between two fibers that are both natural or both manufactured. Cotton and silk samples

were used as the two natural fiber candidates to be differentiated. A PCA was conducted on the spectral data for these fiber types using the same wavelength region of 1334 to 1906 nm. As before, a total of six principal components were used in the analysis. The results of the PCA can be seen in Figure 4.8. The scores plot (Figure 4.8a) shows the distribution of the samples in the new principal component coordinate system. At first glance, the samples appear to have formed a single cluster about the origin, with a small grouping of outliers near the extreme edge of the plot. However, it is again necessary to view the samples in higher dimensional space to get a true and accurate picture of the data.

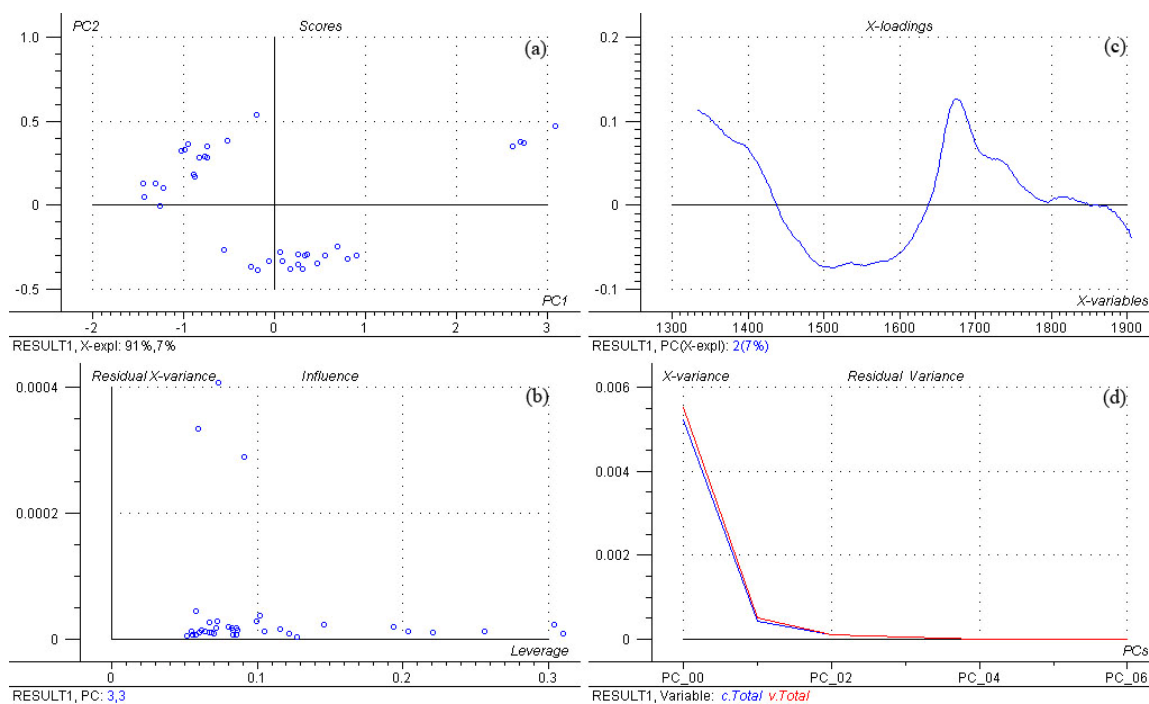


Figure 4.6. Summary of principal components analysis for cotton and polyester samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a of the number of PCs

In figure 4.9, the three-dimensional scores plot is shown with borders of the cotton and silk samples drawn in. As with the previous example, the textiles of the same fiber type clustered together, which formed two distinct groups. This experiment also shows the ability of the model generated to distinguish a natural cellulose fiber, cotton, and a natural protein fiber, silk.

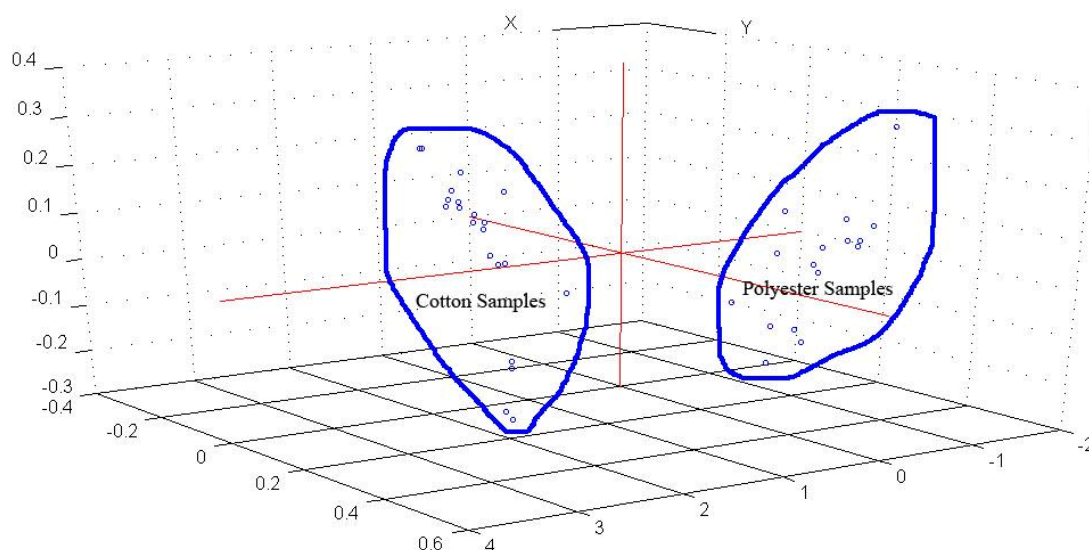


Figure 4.7. Three-dimensional scores plot from principal component analysis of cotton and polyester fabric samples

This pattern becomes further complicated when the addition of more fiber types are added to the PCA. There are six different natural fibers used in this preliminary study: cotton, linen, mohair, rayon, silk, and wool. When conducting a PCA on these fibers, one would expect that cotton and rayon would be difficult to separate due to the fact that rayon is simply chemically regenerated cellulose. The NIR spectra for these samples would be virtually identical, and thus, difficult to classify. Linen, derived from flax, is also a cellulose based fiber. Though the remaining fibers, mohair, silk, and wool

are all protein based, the relative amounts of the amino acid residues in the fibers is different. This will provide a means of separating these fibers. The three-dimensional scores plot for the principal components analysis run on all six natural fibers can be seen in Figure 4.10.

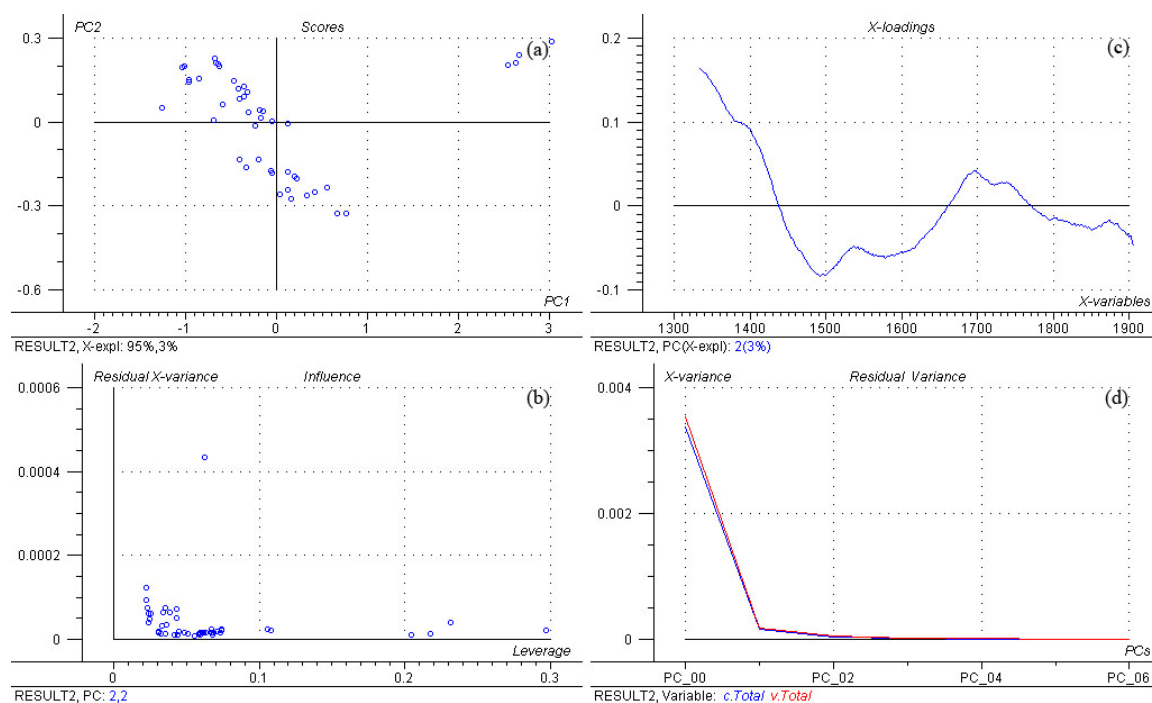


Figure 4.8. Summary of principal components analysis for cotton and silk samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a function of the number of PCs

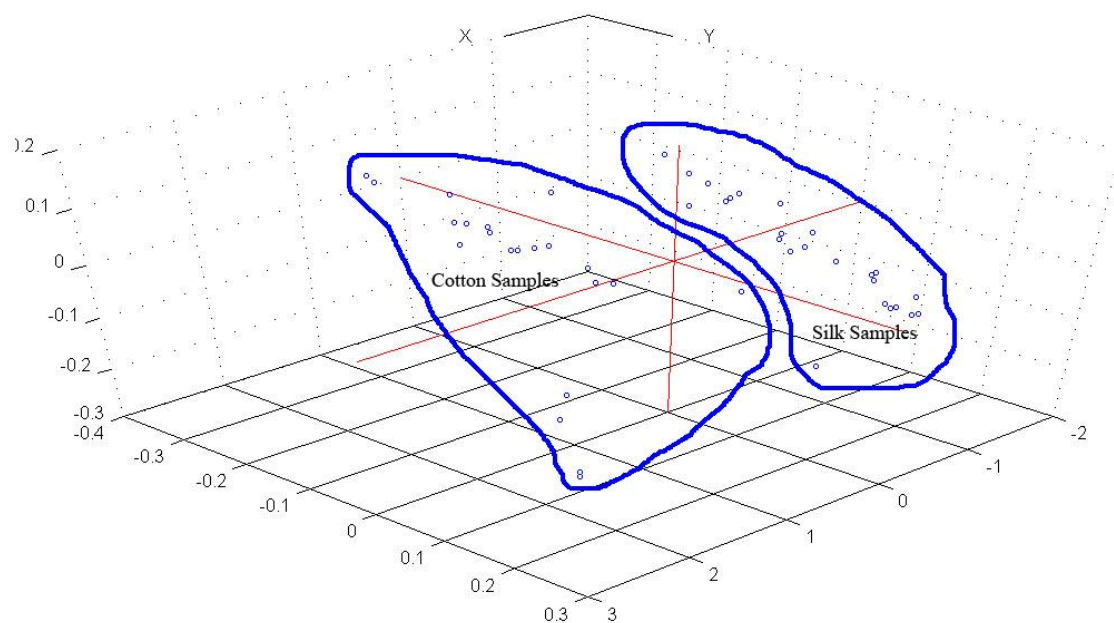


Figure 4.9. Three-dimensional scores plot from principal component analysis of cotton and silk fabric samples

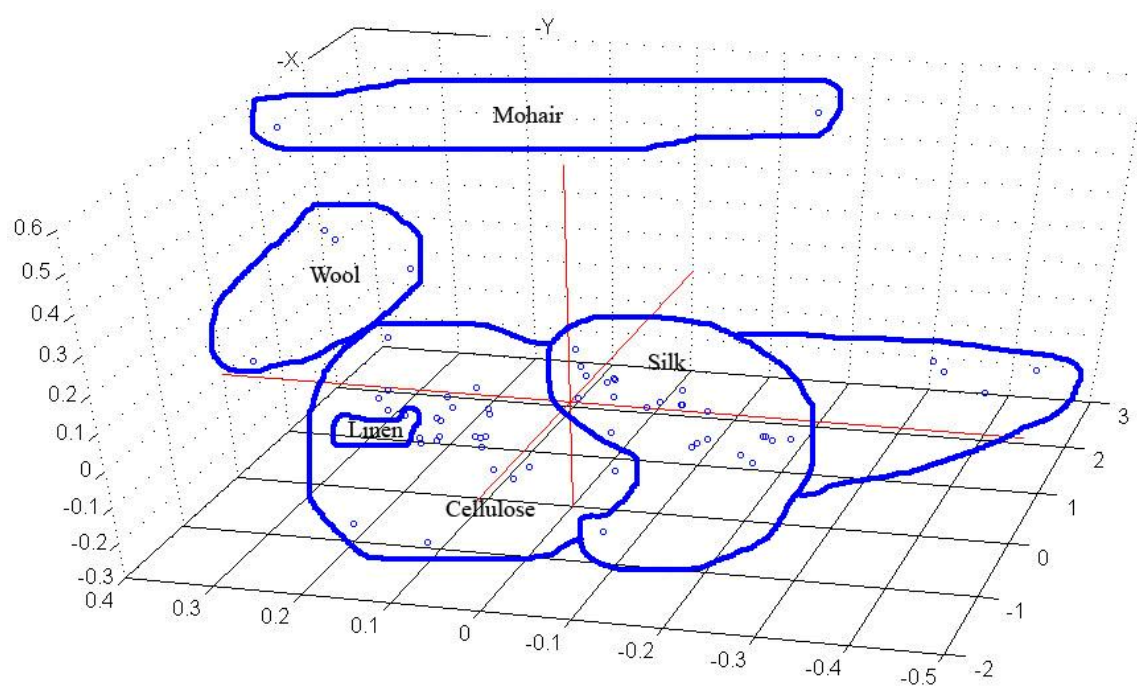


Figure 4.10. Three-dimensional scores plot from principal components analysis of fibers from natural based sources cellulose (cotton, rayon), linen, mohair, silk, and wool

As was expected, the cotton, rayon, and cellulose clustered together into a single group. It should be noted, however, that since there are very few samples of each type in the full data, a true accurate picture of the grouping behavior for these fiber types may not be seen. The protein fibers did cluster into their own groups as was anticipated based on the amino acid content of the keratin and secreted fibers.

The six types of manufactured fibers in the data set were examined next. A principal components analysis was run on spectral data collected from the acetate, acrylic, nylon, olefin, polyester, and polyvinyl chloride using the 1334 to 1906 nm wavelength window, six calculated principal components, and full cross-validation of the model. The resulting three-dimensional scores plot for the analysis can be seen in Figure 4.11. The degree of separation is quite prominent. The arbitrary size of the clusters is quite small, lending more credence to the possibility of using the diffuse near-infrared reflectance spectra of textiles in a SIMCA analysis to classify them based on fiber type. The final test of the discriminating power of the principal components analysis was to place all the pure fiber samples, excluding any blends, into a single model and investigate their placement in the resulting scores plot. Figure 4.12 shows the three-dimensional scores plot for this study. The natural and manufactured fibers, as two groups, clustered around one another. The cellulose based fibers, cotton, rayon, and linen, again showed an association with one another, and not decomposing into three distinct groups. Mohair, silk, and wool fibers did segregate into three groups, removed from the other natural fibers, as well as the manufactured. The manufactured fibers all formed distinct clusters, based on their chemical make up.

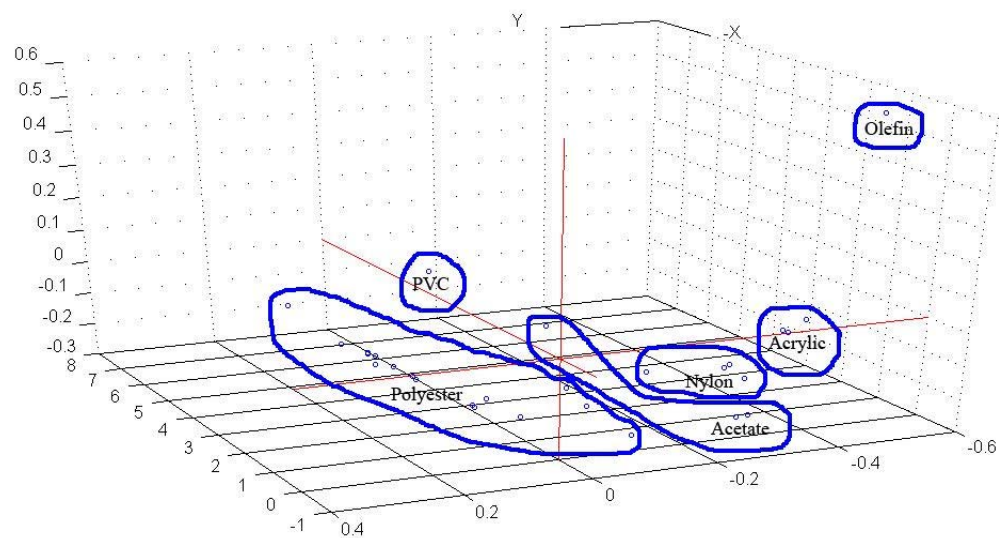


Figure 4.11. Three-dimensional scores plot from principal component analysis of synthetic fibers (acetate, acrylic, nylon, olefin, polyester, PVC)

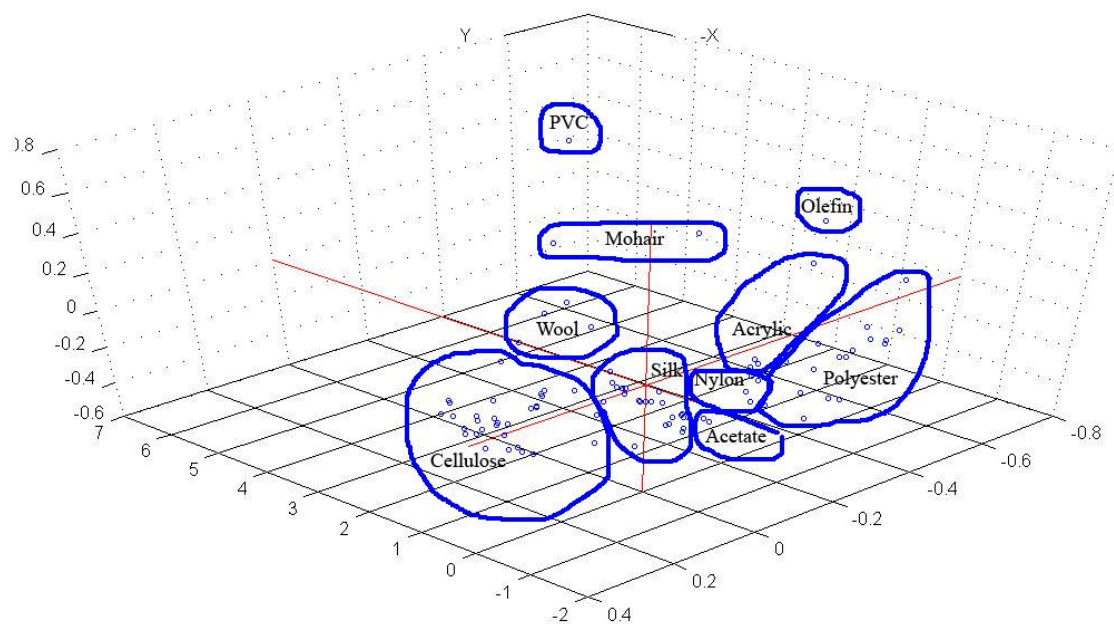


Figure 4.12. Three-dimensional scores plot from principal component analysis of both natural and synthetic fibers

With the results of the PCA models for different fiber types in hand, the near-infrared data was then used to create class based models to describe a particular fiber type. In order to have models based on an appropriate sample size and still have enough samples for validation, the three fibers selected for this portion of the study were cotton, silk, and polyester. The Unscrambler program automatically defaults to a set number of principal components to be used in a classification, even if the number of PCs computed in a model is higher. The suggested number of principal components is what the program deems necessary to explain the variation in the spectral data. But, when one looks at the residual x-variance plot for a model based on 16 cotton samples (Figure 4.13), it can be seen that the residual variance continues to drop, closer to zero, beyond the second principal component. The recommended number of PCs for the cotton models studied was always two. However, each model's residual x-variance plots showed further explanation of the spectral data when more PCs were used. For this reason, the modeling and validation done was computed using both the suggested and a predetermined number of 6 principal components.

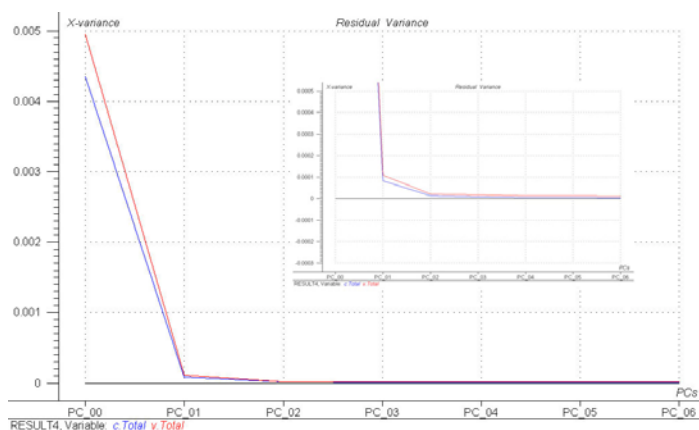


Figure 4.13. Residual x-variance plot for a principal components analysis of cotton samples

In Table 4.5, the classification results for five cotton sets can be seen. When the suggested number of principal components was used, the classification results showed two errors (102-cotton). One sample in test sets 2 and 4 was not identified as a cotton sample. However, when the number of principal components used was increased to 6, the two errors were not present, and all the samples were correctly identified as cotton. The same type of analysis was performed on the polyester samples. The suggested number of principal components for these analyses was 2. Both the suggested and predetermined number of principal components yielded the same results. All the polyester fiber samples were correctly classified (Table 4.5).

The classifications run on the silk samples were successful as well. The five sample sets selected provided correct classifications with one exception. A sample designated silk-13 was not classified in either model for which it was used as a validation sample, at neither the suggested or predetermined number of principal components. For this fiber type, the suggested number of PCs changed from 2 to 3, depending on which model was used. The results for the classification can be seen in Table 4.5.

To further optimize the method of classification, the PCA models developed for the cotton, silk, and polyester samples in the previous section were used in an attempt to classify fibers in the data set that were not included in the models (Figure 4.14). These fibers included: acetate, acrylic, linen, mohair, nylon, olefin, PVC, rayon, and wool. When using SIMCA methodology, a sample will not be classified as a member of a group, unless it is deemed to belong to that group. This is a major advantage to negating false positives when classifying. In the best case scenario, no samples from the other

fiber types would be flagged when being compared to the models for cotton, silk, and polyester.

Table 4.5. Classification results of preliminary study for cotton, polyester, and silk

Cotton				Polyester			Silk		
Test Set	Sample	Class ^a	Class ^b	Sample	Class ^c	Class ^b	Sample	Class ^d	Class ^b
1	22	X	X	15	X	X	13		
	36	X	X	49	X	X	44	X	X
	46	X	X	59	X	X	80	X	X
	91	X	X	72	X	X	84	X	X
	100	X	X				98	X	X
2							109	X	X
	36	X	X	6	X	X	13		
	46	X	X	48	X	X	40	X	X
	73	X	X	49	X	X	44	X	X
	100	X	X	72	X	X	69	X	X
3	102		X				80	X	X
							108	X	X
	18	X	X	6	X	X	24	X	X
	21	X	X	47	X	X	40	X	X
	102	X	X	49	X	X	42	X	X
4	105	X	X	66	X	X	58	X	X
	107	X	X				78	X	X
							99	X	X
	18	X	X	4	X	X	37	X	X
	23	X	X	59	X	X	43	X	X
5	100	X	X	65	X	X	44	X	X
	102		X	75	X	X	58	X	X
	104	X	X				80	X	X
							84	X	X
	21	X	X	15	X	X	24	X	X
	38	X	X	65	X	X	40	X	X
	91	X	X	71	X	X	69	X	X
	106	X	X	75	X	X	76	X	X
	111	X	X				84	X	X
							109	X	X

^aTwo PCs, ^bSix PCs, ^cThree PCs, ^dTwo PCs for sets 1, 2, and 4, three PCs for sets 3 and 5

In Figure 4.14, any asterisk within a cell is an indication of a false positive. There should be no positive hits in this figure. Whether or not the recommended number of

	Recommended PCs															Non-Recommended PCs															
Sample	C	C	C	C	C	P	P	P	P	P	S	S	S	S	S	C	C	C	C	C	P	P	P	P	P	S	S	S	S	S	
Acetate							*	*	*					*		*						*	*	*	*	*			*	*	*
Acrylic																															
Olefin																															
Acrylic																															
Acetate							*	*	*					*		*						*	*	*	*	*			*	*	*
Rayon																															
Rayon																*		*	*												
Nylon						*	*	*	*	*				*	*	*						*	*	*	*	*			*	*	*
Nylon						*	*	*	*	*				*	*	*						*	*	*	*	*			*	*	*
Nylon						*	*	*	*	*				*		*						*	*	*	*	*			*	*	*
Acrylic																															
Wool																															
Linen	*	*	*	*	*									*	*	*	*	*	*	*	*						*	*	*	*	*
Linen	*	*	*	*	*									*	*	*	*	*	*	*	*						*	*	*	*	*
Wool																													*	*	
Wool															*	*	*	*	*	*	*						*	*	*	*	*
Rayon	*	*	*	*	*						*	*	*	*	*	*	*	*	*	*	*						*	*	*	*	*
PVC																															
Wool																											*		*	*	
Acetate														*		*								*				*	*	*	
Mohair																															
Nylon						*	*	*	*	*	*	*	*	*	*	*						*	*	*	*	*	*	*	*	*	*
Rayon	*	*	*	*	*						*	*	*	*	*	*	*	*	*	*	*						*	*	*	*	*
Mohair																															
Acrylic																															
Rayon	*	*	*	*	*						*	*	*	*	*	*	*	*	*	*	*						*	*	*	*	*
Rayon	*	*	*	*	*												*	*	*	*	*						*				
Rayon	*	*	*	*	*												*	*	*	*	*						*	*	*		
Linen	*	*	*	*	*												*	*	*	*	*										

Figure 4.14. Comparison of cotton (C), polyester (P), and silk (S) models classification abilities with fiber types not included in the calibration phase

principal components was used, the misclassifications are in the same places. Rayon and linen fibers are being misidentified as cotton and silk. Wool is being marked as a silk fiber. Nylon and acetate fibers are being misclassified as polyesters as well as silks. To use SIMCA to classify fiber samples, these errors must be greatly reduced or removed. In an attempt to clear up some signal noise, the reflectance spectra from the data set were subjected to a Savitzky-Golay smoothing routine, fitted to a zero degree polynomial with

5 averaging points on the left and right side. The smoothed spectra for the cotton, polyester, and silk samples can be seen in Figure 4.15.

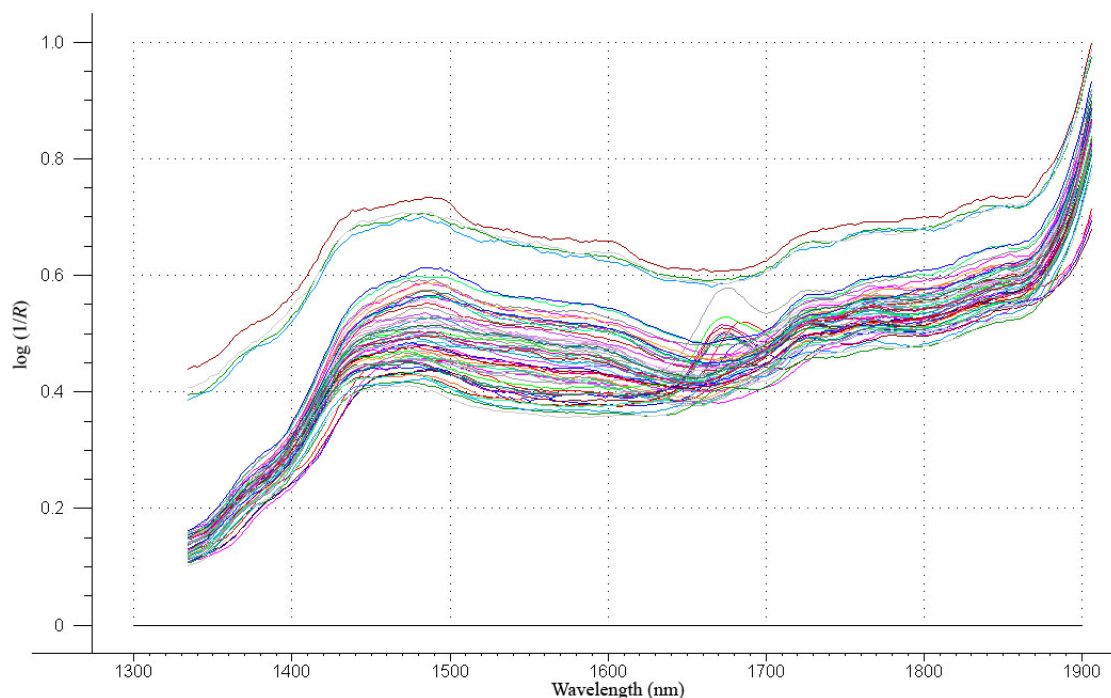


Figure 4.15. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of cotton, polyester, and silk textiles from 1334-1906 nm

The smoothed reflectance spectra for cotton, polyester, and silk were then used to construct models for those fiber types. These models were then validated as were the PCA models developed from the unsmoothed data. The results of the classification analysis can be seen in Table 4.6. There was no change in the classification abilities of the polyester model when compared to the unsmoothed model. All polyesters were correctly identified. There are two classification errors present in the cotton samples (46-cotton) when the suggested number of PCs is used. These errors are absent from the classification when 6 PCs are used; however, there is an additional error present in test

Table 4.6. Classification results of preliminary study for cotton, polyester, and silk

with Savitzky-Golay smoothing of raw diffuse near-infrared reflectance spectra

Cotton				Polyester			Silk		
Test Set	Sample	Class ^a	Class ^b	Sample	Class ^c	Class ^b	Sample	Class ^d	Class ^b
1	22	X	X	15	X	X	13		
	36	X	X	49	X	X	44	X	X
	46	X		59	X	X	80	X	X
	91	X	X	72	X	X	84	X	X
	100	X	X				98	X	X
2							109	X	X
	36	X	X	6	X	X	13		
	46	X	X	48	X	X	40	X	X
	73	X	X	49	X	X	44	X	X
	100	X	X	72	X	X	69	X	X
	102		X				80	X	X
3							108	X	X
	18	X	X	6	X	X	24	X	X
	21	X	X	47	X	X	40	X	X
	102	X	X	49	X	X	42	X	
	105	X	X	66	X	X	58	X	X
	107	X	X				78	X	X
4							99	X	X
	18	X	X	4	X	X	37	X	X
	23	X	X	59	X	X	43	X	X
	100	X	X	65	X	X	44	X	X
	102		X	75	X	X	58	X	X
	104	X	X				80	X	X
5							84	X	X
	21	X	X	15	X	X	24	X	X
	38	X	X	65	X	X	40	X	X
	91	X	X	71	X	X	69	X	X
	106	X	X	75	X	X	76	X	X
	111	X	X				84	X	X
							109	X	X

^aTwo PCs, ^bSix PCs, ^cThree PCs, ^dTwo PCs for sets 1 and 2, three PCs for sets 3 and 5, four PCs for set 4

set one that was not present the analysis for 2 PCs. The overall performance of the cotton model is still acceptable with only one error at 6 PCs. The silk models saw the same problem sample, designated silk-13, as the lone hold-out for both models in test sets 1 and 2. An additional error can be seen in test set 3 (42-silk).

On the surface, the models using smoothed spectra seem to provide the same classification abilities as their unsmoothed counterparts. The smoothed models are far superior however. As was done previously, the models from the test sets for the cotton, polyester, and silk fibers were used to classify the other sample fiber types in the data set. As seen in Figure 4.14, the unsmoothed spectra models showed rampant misclassifications. When the smoothed spectra models using the imposed 6 PCs were compared to the other fiber types, all the misclassifications from the polyester and silk models vanished. The cotton models are a different matter. The PCA models still classified rayon as a cotton fiber, but at a greatly reduced frequency. Linen fibers were also misclassified. The lack of selectivity of the cotton models for cotton alone could indicate that the sample population must be increased in order to better define the PCA models and in turn the classification.

Expanded Sample Population Study

The next progression in the textile study was an increase in the number of fabric samples within each individual fiber class. A total of 826 textile samples were cataloged and had their NIR reflectance spectra collected. The specific breakdown as to the types of fibers and the total number of each can be seen in Table 4.7. The raw spectra were smoothed, using the Savitzky-Golay moving average with 5 averaging points on the left and right sides, and a wavelength window from 1334 to 1906 nm was selected (Figures 4.16, 18, 20, 22, 24, 26). Principal components analyses were conducted (Figures 4.17, 19, 21, 23, 25, 27) on the fiber classes of acetate, cotton, polyester, rayon, silk, and wool. The total numbers of samples in the remaining categories were deemed too low to

achieve a PCA model well enough defined as to classify future samples. Results can be seen in Tables 4.8-4.13.

Table 4.7. Textile fiber catalog for expanded study

Fiber Type	No. of Samples	Fiber Type	No. of Samples
Acetate	61	Olefin	1
Acrylic	4	Polyester	109
Blends ^a	50	PVC	1
Cotton	274	Rayon	77
Linen	5	Silk	46
Mohair	2	Wool	192
Nylon	4		

^aBlends include a variety of fiber mixtures

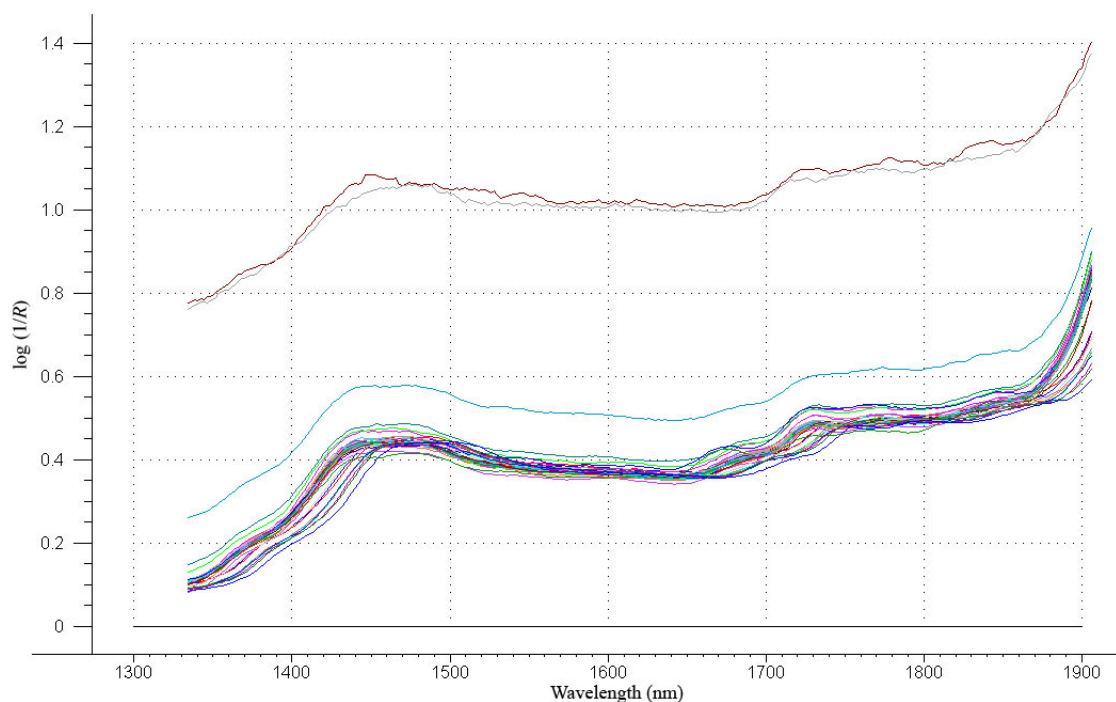


Figure 4.16. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 61 acetate textiles from 1334-1906 nm

Misclassifications and Counterfeit Determination of Silk Sample

There was no additional examination of the textile fibers acquired from the Department of Family and Consumer Science. The fabrics were labeled as to their fiber type, and as such, cataloged on that basis alone. The models created for the individual

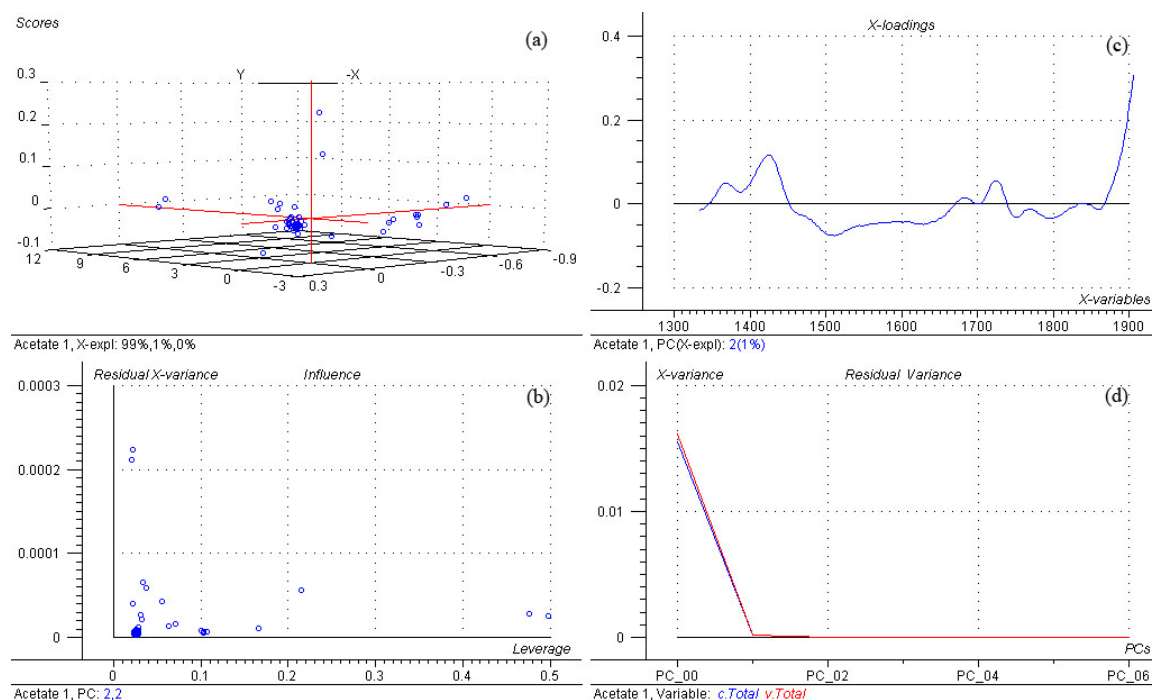


Figure 4.17. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of acetate samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a of the number of PCs

fiber categories were validated not only by classifying samples of the same fiber type, but also by using the analyses to attempt classification of samples of different fiber types.

False positives, as were seen in the preliminary study, could be a major obstacle in the implementation of this method as a routine textile analytical technique. The expanded study, with more inherent variability, was used to classify samples of each fiber type.

Acetate samples showed no false classifications when categorized using the cotton, rayon, and wool. Of the 10 polyester class models, 8 of them produced a false positive

on a single fiber of acetate. The spectrum for the suspect sample, designated Acetate-582, along with representative spectra of acetates and polyester can be seen in Figure 4.28. When a principal components analysis was conducted on the spectra present in Figure 4.28, the suspect sample did not cluster with the polyester samples.

Table 4.8. SIMCA classification results for acetate fibers in expanded study^a

<u>Test Set 1</u>		<u>Test Set 2</u>		<u>Test Set 3</u>		<u>Test Set 4</u>		<u>Test Set 5</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
9	X	5	X	5	X	1	X	10	X
13	X	7	X	24	X	3	X	15	X
15	X	19	X	28	X	7	X	18	X
27	X	26	X	29	X	12	X	20	X
33	X	41	X	34	X	14	X	26	X
35	X	42	X	38	X	19	X	33	X
40	X	45	X	41	X	30	X	39	X
43	X	51	X	50	X	32	X	43	X
45	X	56	X	53	X	36	X	44	X
59	X	59	X	55	X	50	X	53	X
<u>Test Set 6</u>		<u>Test Set 7</u>		<u>Test Set 8</u>		<u>Test Set 9</u>		<u>Test Set 10</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
3		5	X	4		4		3	X
5	X	9	X	8		6	X	15	X
14	X	10	X	12	X	9	X	22	X
17	X	11	X	18	X	11	X	33	X
24	X	16	X	30	X	20	X	37	X
26	X	20	X	31		31		40	X
31		27	X	39	X	43	X	41	X
34	X	42	X	45	X	45	X	46	X
60	X	48	X	49	X	49	X	52	X
61		58	X	61		52	X	53	X
Correct Classifications									91%

^aClassifications based on six principal components

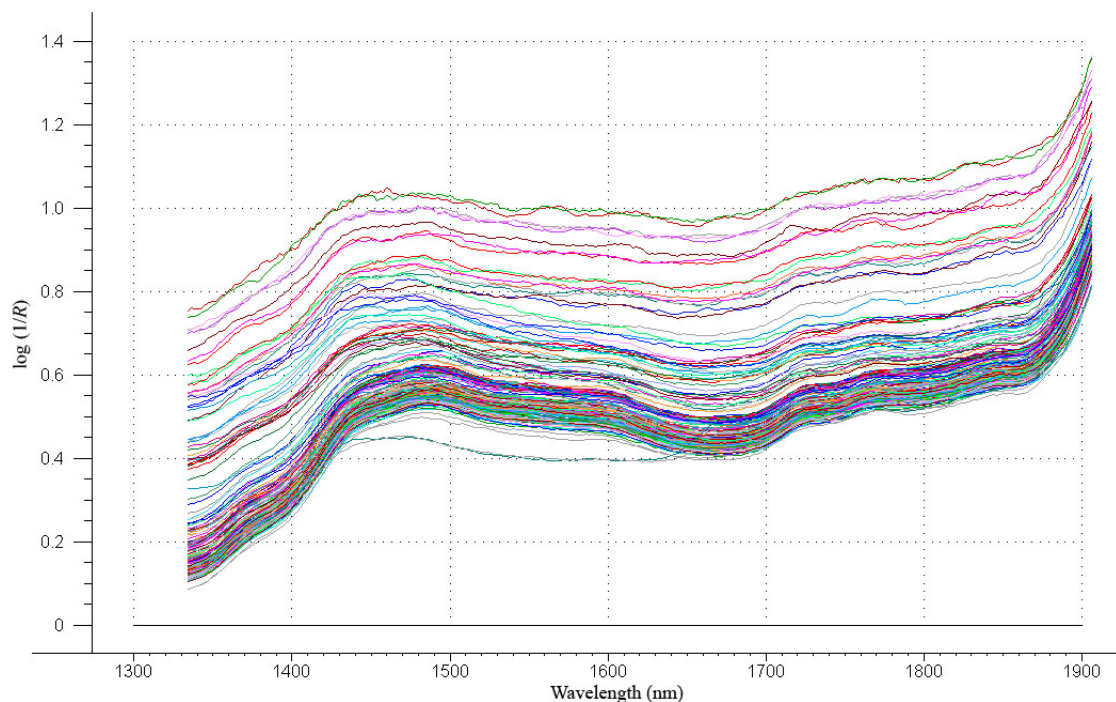


Figure 4.18. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 274 cotton textiles from 1334-1906 nm

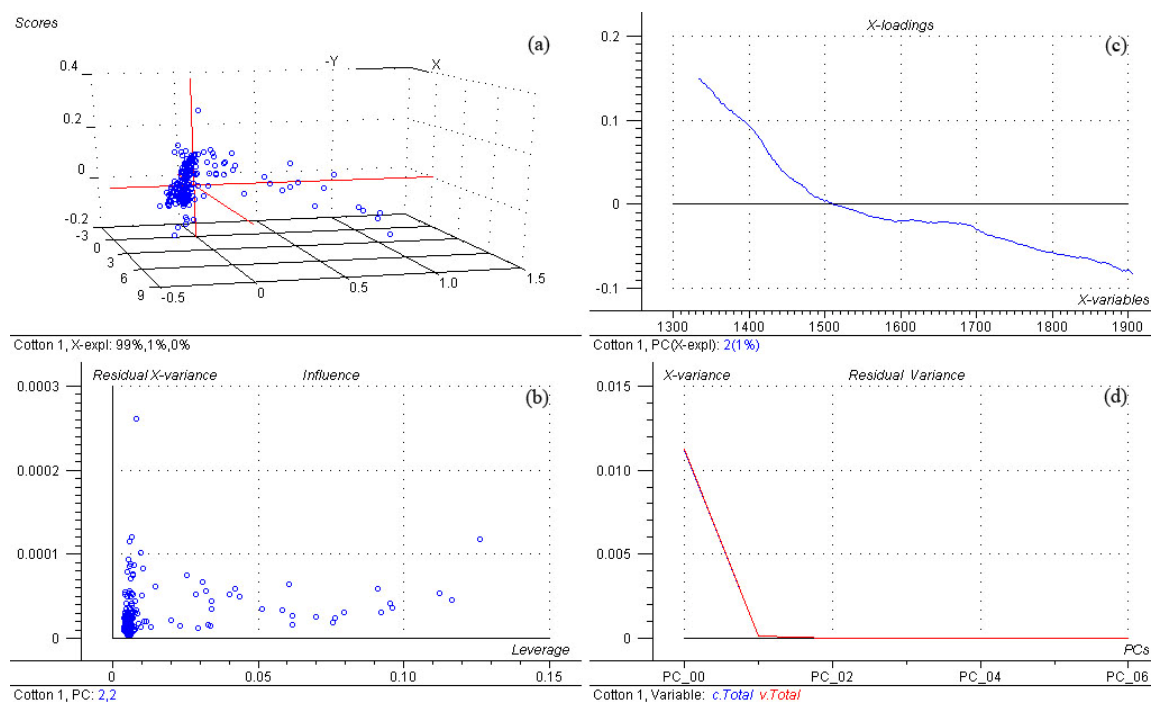


Figure 4.19. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of cotton samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a of the number of PCs

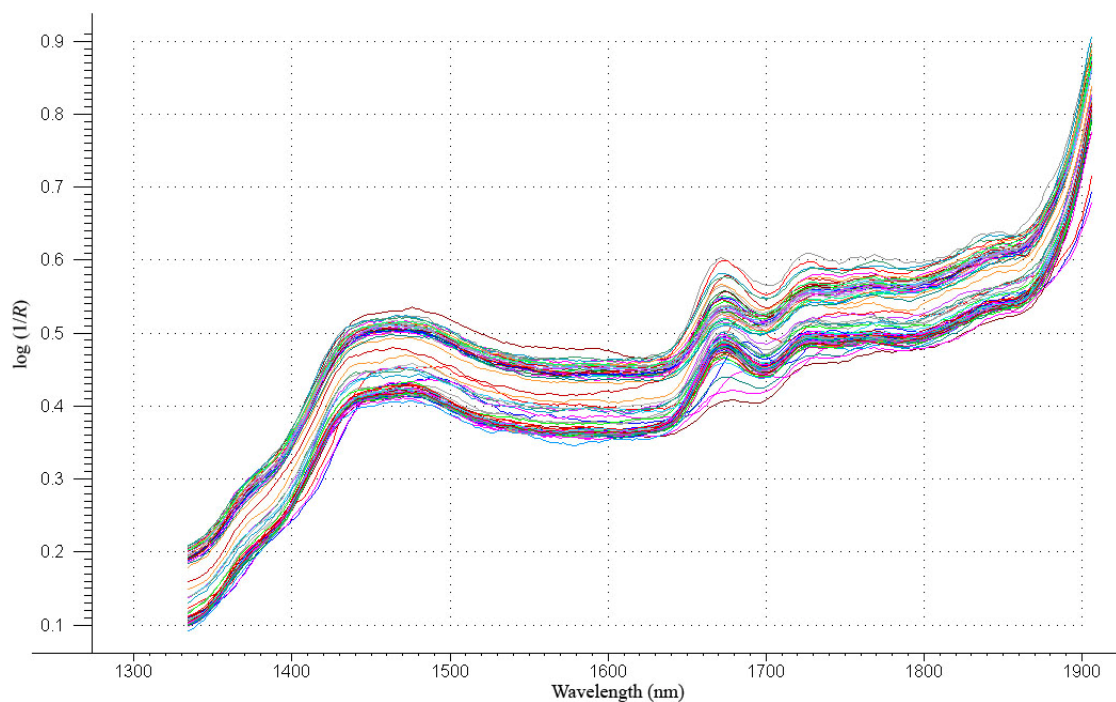


Figure 4.20. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 109 polyester textiles from 1334-1906 nm

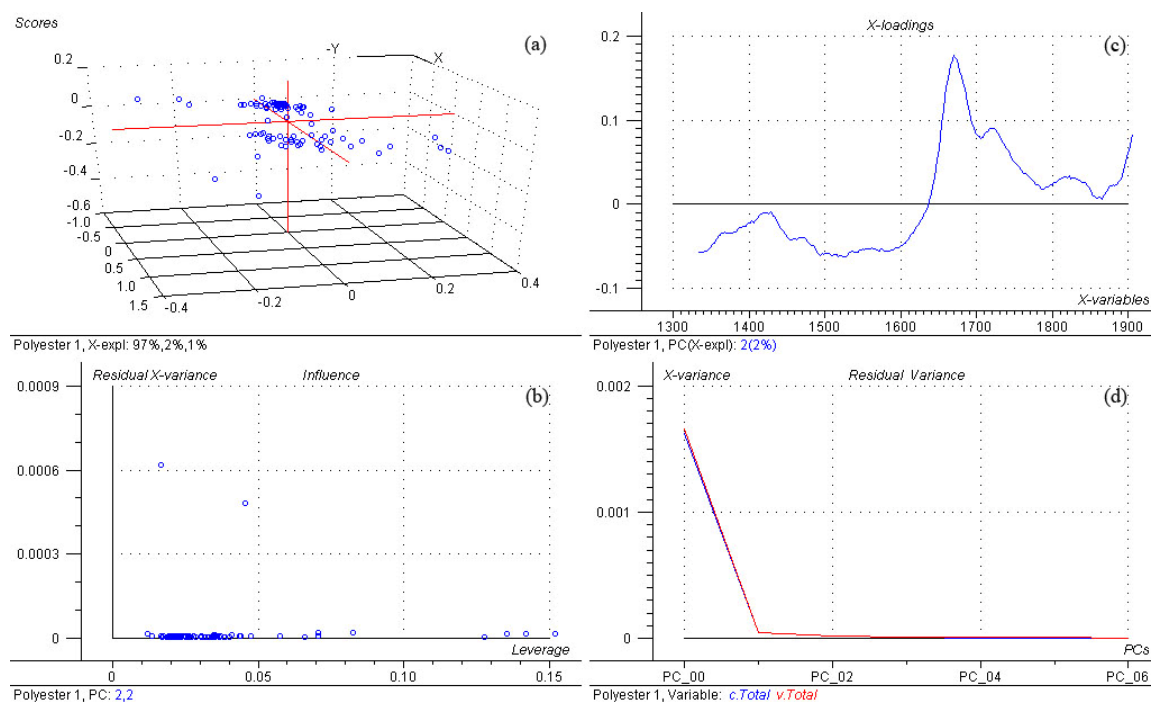


Figure 4.21. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of polyester samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a function of the number of PCs

Table 4.10. SIMCA classification results for polyester fibers in expanded study^a

Test Set 1		Test Set 2		Test Set 3		Test Set 4		Test Set 5	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
3		2		2		19	X	2	
13	X	3		3		25	X	4	
15	X	4		11	X	31	X	5	X
17	X	6	X	12	X	39	X	12	X
19	X	11	X	16	X	41	X	14	
28	X	19	X	18	X	42	X	29	X
34	X	41	X	27	X	44	X	31	X
40	X	45	X	36	X	53	X	33	X
41	X	48	X	44	X	54	X	35	X
43	X	51	X	46	X	55	X	43	X
48	X	55	X	49	X	56	X	46	X
49	X	58	X	50	X	61	X	50	X
54	X	60	X	56	X	62	X	51	X
57	X	63	X	58	X	63	X	56	X
58	X	71	X	69	X	69	X	71	X
61	X	72	X	70	X	79	X	88	X
66	X	87	X	74	X	83	X	92	X
67	X	89	X	75	X	91	X	96	X
72	X	107	X	96	X	96	X	104	X
76	X	108	X	99	X	100	X	108	X
Test Set 6		Test Set 7		Test Set 8		Test Set 9		Test Set 10	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
10	X	5	X	15	X	5	X	5	X
18	X	7	X	19	X	8	X	10	X
20	X	11	X	23	X	10	X	11	X
23	X	14		28	X	12	X	19	X
25	X	20	X	31	X	14		21	X
30	X	21	X	32	X	17	X	35	X
43	X	26	X	38	X	22	X	42	X
50	X	29	X	43	X	33	X	45	X
51	X	41	X	47	X	34	X	47	X
53	X	44	X	55	X	38	X	50	X
57	X	51	X	61	X	47	X	53	X
61	X	57	X	67	X	54	X	55	X
67	X	60	X	71	X	55	X	56	X
71	X	61	X	74	X	57	X	69	X
79	X	73	X	79	X	58	X	72	X
80	X	81	X	82	X	61	X	81	X
93	X	86	X	95	X	84	X	89	X
97	X	100	X	99	X	86	X	90	X
98	X	101	X	101	X	88	X	93	X
99	X	102	X	108	X	97	X	98	X
Correct Classifications								95%	

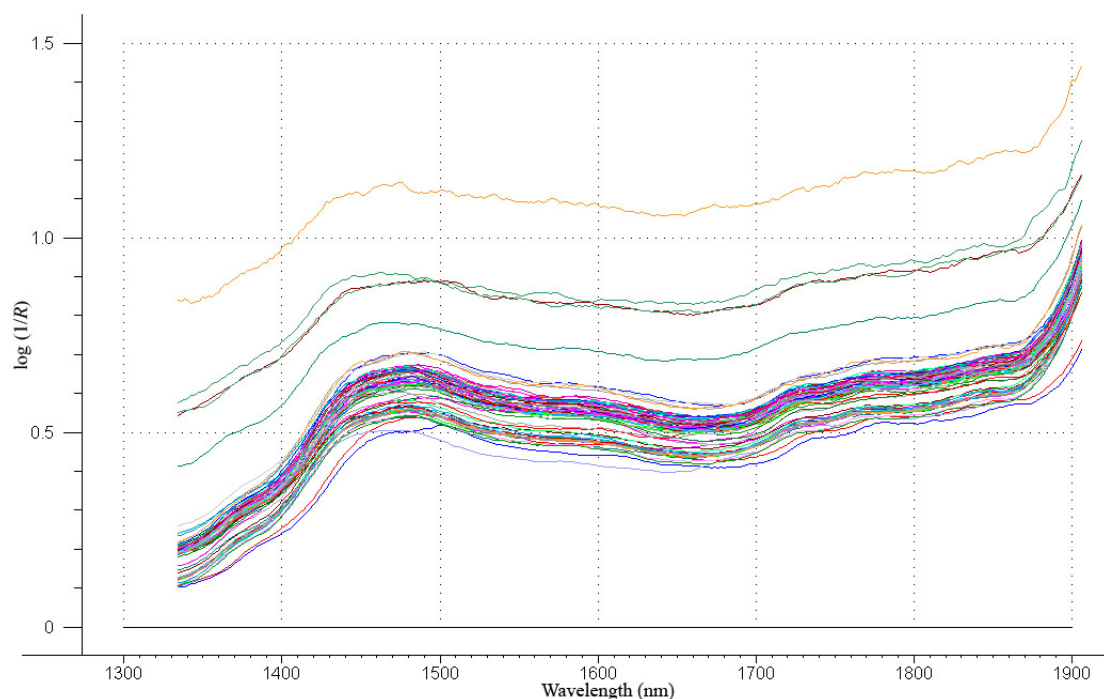


Figure 4.22. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 77 rayon textiles from 1334-1906 nm

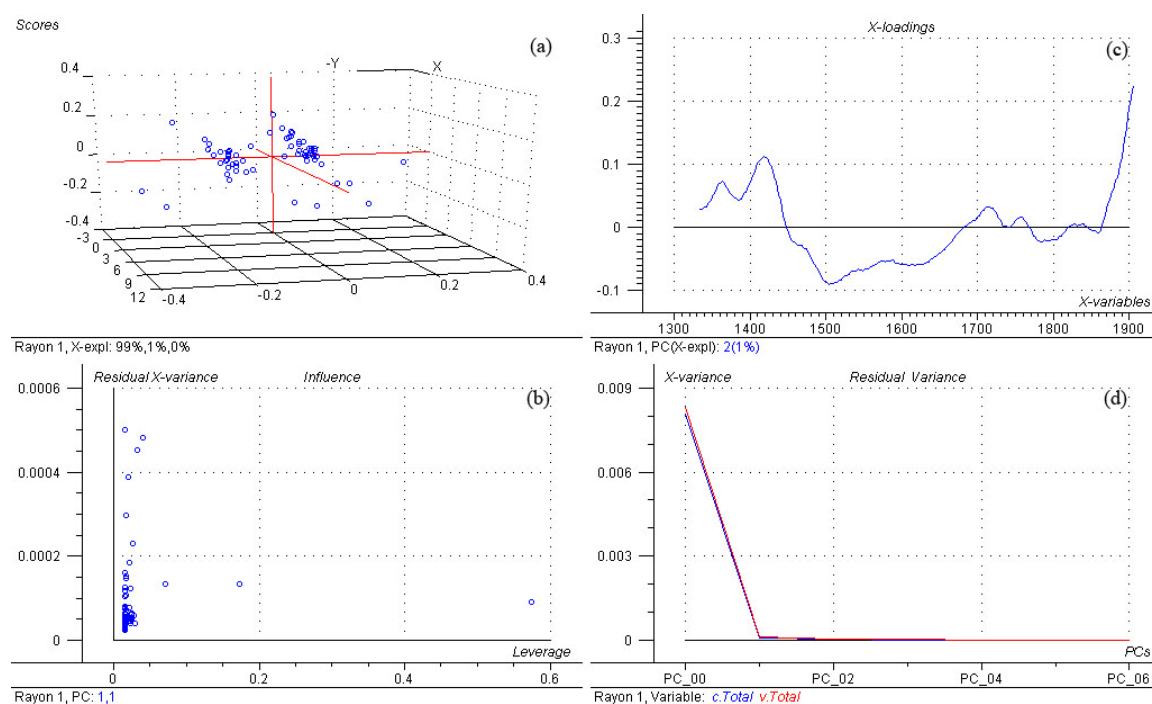


Figure 4.23. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of rayon samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a function of the number of PCs

Table 4.11. SIMCA classification results for rayon fibers in expanded study^a

<u>Test Set 1</u>		<u>Test Set 2</u>		<u>Test Set 3</u>		<u>Test Set 4</u>		<u>Test Set 5</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
16	X	2	X	11	X	7		5	X
19		5	X	13	X	9	X	19	
25	X	8	X	18	X	17	X	20	X
33	X	11	X	31	X	23	X	21	X
34	X	22	X	33	X	28	X	41	X
36	X	35	X	45	X	29	X	44	X
42	X	44	X	47	X	49	X	49	X
72		58	X	48	X	51	X	62	X
73	X	61	X	52	X	55	X	71	X
76	X	64	X	72		59	X	77	X
<u>Test Set 6</u>		<u>Test Set 7</u>		<u>Test Set 8</u>		<u>Test Set 9</u>		<u>Test Set 10</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
4	X	2	X	1		4	X	16	X
18	X	10	X	11	X	9	X	23	X
29	X	17	X	14	X	10	X	33	X
41	X	19		25	X	13	X	43	X
53	X	21	X	28	X	25	X	47	X
55	X	25	X	32		28	X	50	X
56	X	31	X	37	X	34	X	62	X
63	X	63	X	40	X	36	X	70	X
64	X	74	X	50	X	47	X	72	
69	X	75	X	73	X	64	X	75	X
Correct Classifications									92%

^aClassifications based on six principal components

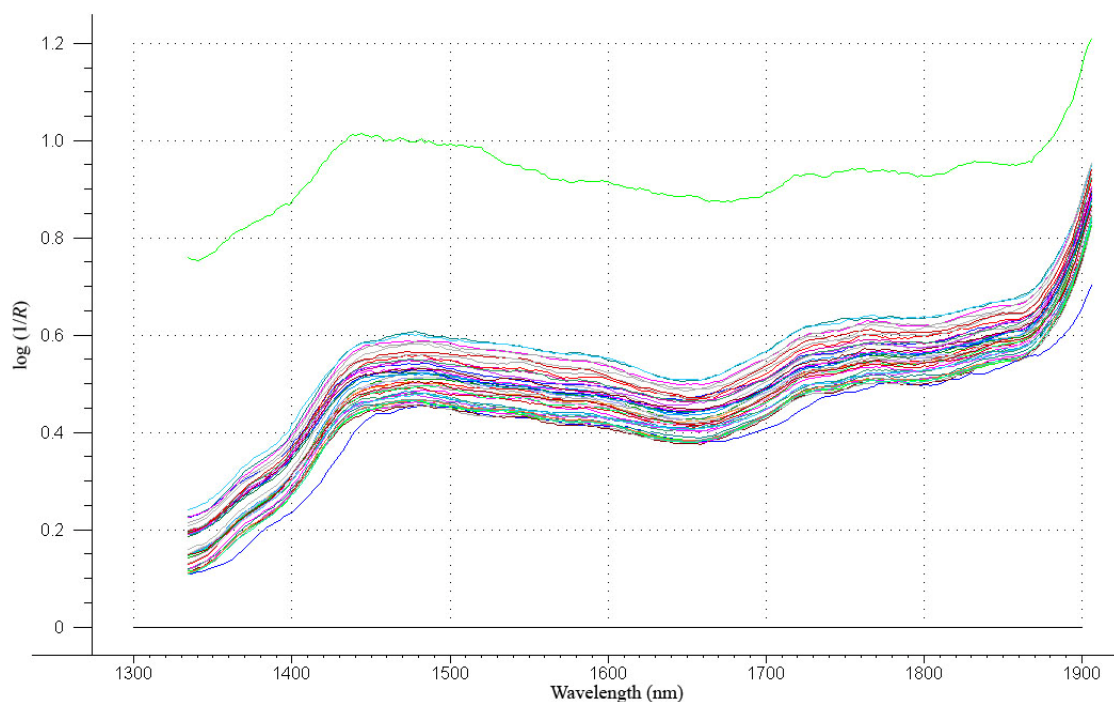


Figure 4.24. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 45 silk textiles from 1334-1906 nm

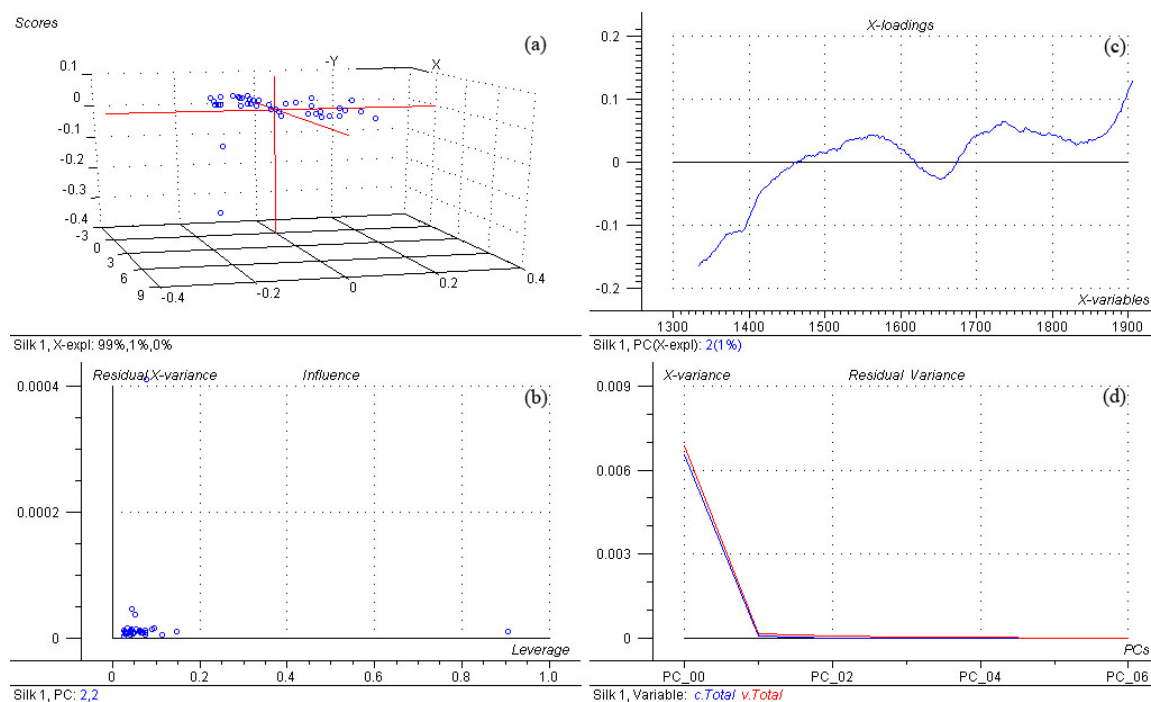


Figure 4.25. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of silk samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a function of the number of PCs

Table 4.12. SIMCA classification results for silk fibers in expanded study^a

<u>Test Set 1</u>		<u>Test Set 2</u>		<u>Test Set 3</u>		<u>Test Set 4</u>		<u>Test Set 5</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
5	X	11	X	1		2	X	5	X
24	X	13	X	13	X	11	X	10	X
37	X	23	X	23	X	16	X	28	X
42	X	28	X	42	X	31	X	35	X
44	X	30	X	45	X	34	X	37	X
<u>Test Set 6</u>		<u>Test Set 7</u>		<u>Test Set 8</u>		<u>Test Set 9</u>		<u>Test Set 10</u>	
Sample	Class	Sample	Class	Sample	Class	Sample	Class	Sample	Class
12	X	23	X	2	X	7	X	2	X
17	X	27	X	24	X	13	X	5	X
32	X	32	X	33	X	15	X	20	X
35	X	35	X	39		20	X	27	X
41	X	41	X	42	X	21	X	42	X
Correct Classifications									96%

^aClassifications based on six principal components

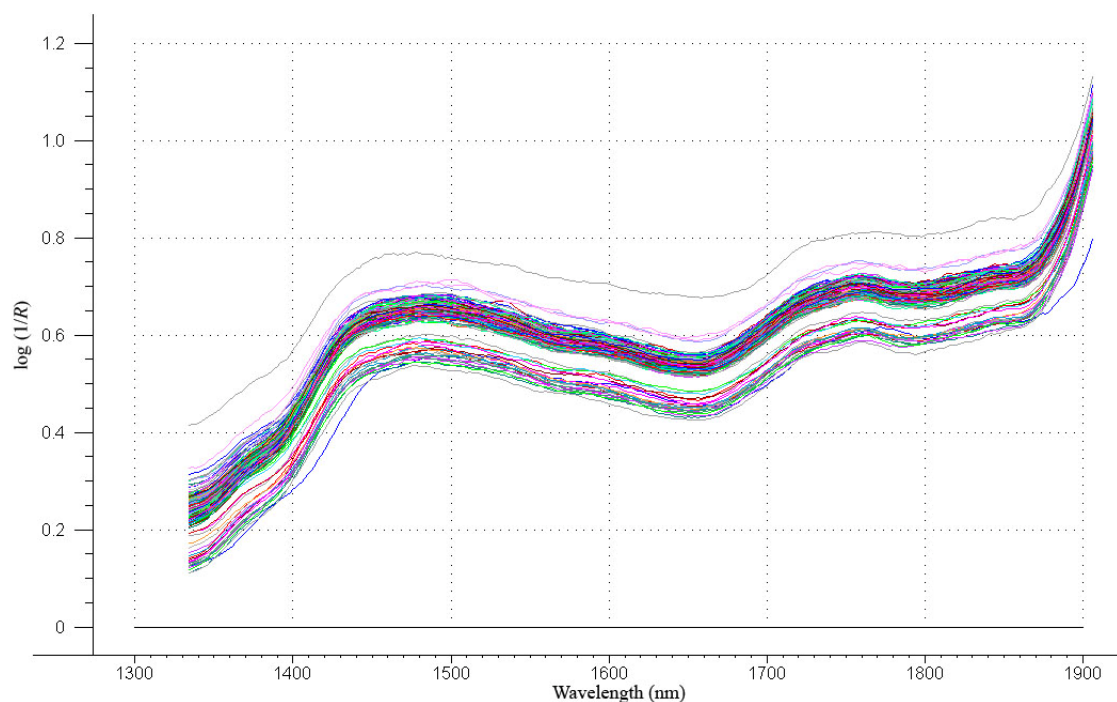


Figure 4.26. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 192 wool textiles from 1334-1906 nm

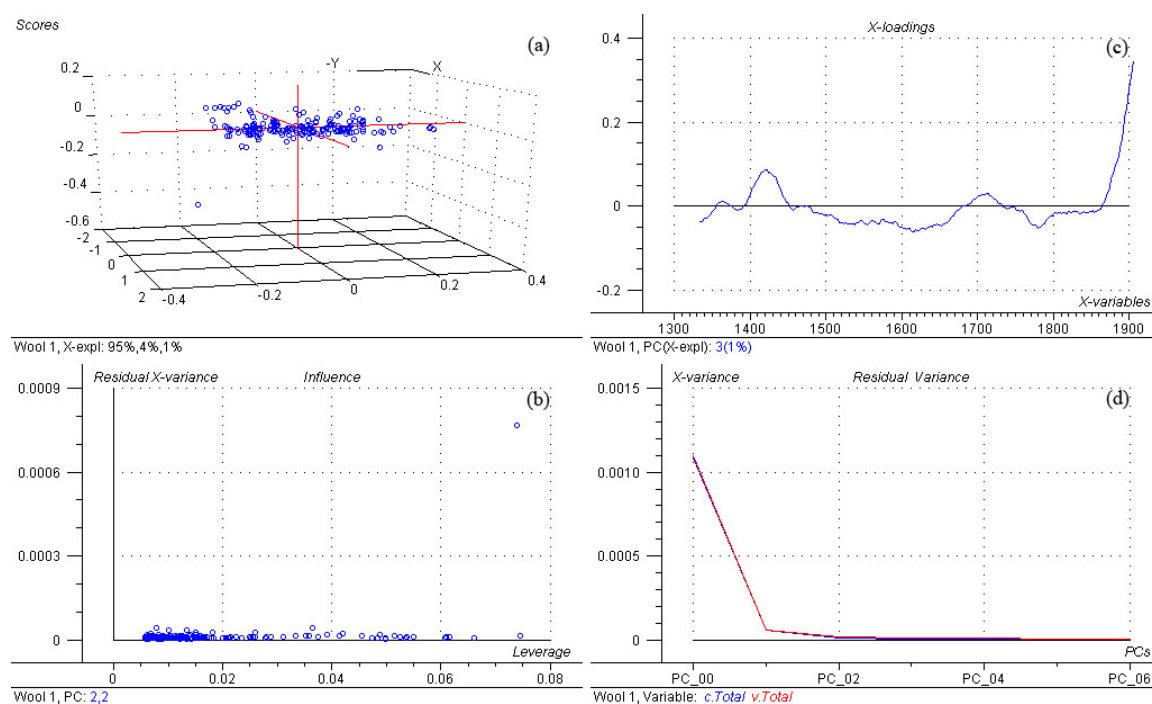


Figure 4.27. Summary of principal components analysis for Savitzky-Golay smoothed diffuse near-infrared spectra of wool samples with a wavelength window from 1334-1906 nm: (a) scores plot; (b) residual X-variance plot; (c) x-loadings as a function of wavelength; (d) residual variance as a function of the number of PCs

The peak at approximately 1670 nm, commonly seen in the spectra of polyesters, is present in the spectra for Acetate-582. It should be noted, however, that the peak is substantially weaker than in the example spectra for polyesters. The presence of this spectral feature could account for the erroneous classification.

When cotton samples were classified, no misclassifications were seen when the acetate, polyester, silk, and wool models were used. The ten rayon models produced an enormous amount of errors in the classification of the cottons. The sheer number of errors indicates that the 77 rayon samples available for modeling are still too small. The cotton samples are too chemically similar to rayon to have a distinction made with such a reduced sample population. When rayon samples are subjected to classification by the cotton models, there was a single false positive sample across the data sets. One acetate model flagged a single sample in the analysis. The polyester, silk, and wool models showed no erroneous classifications for the rayon samples.

Polyester samples showed no false positives when classified using the acetate, cotton, rayon, silk, or wool models. The models for acetate, cotton, polyester, and wool all showed excellent performances in regards to the classification of silk. The models for rayon, unfortunately yielded a large number of misclassifications. The sample size for the rayon models is not large enough. The spectra for the rayon and silk samples can be seen in Figure 4.29. Since the spectral features are so broad and ill-defined, the principal components analysis is not able to ascertain the differences between the two fabric types. A suspect silk textile was identified and can be seen in Figure 4.30. The sample has the characteristic peak at approximately 1678, indicative of polyester.

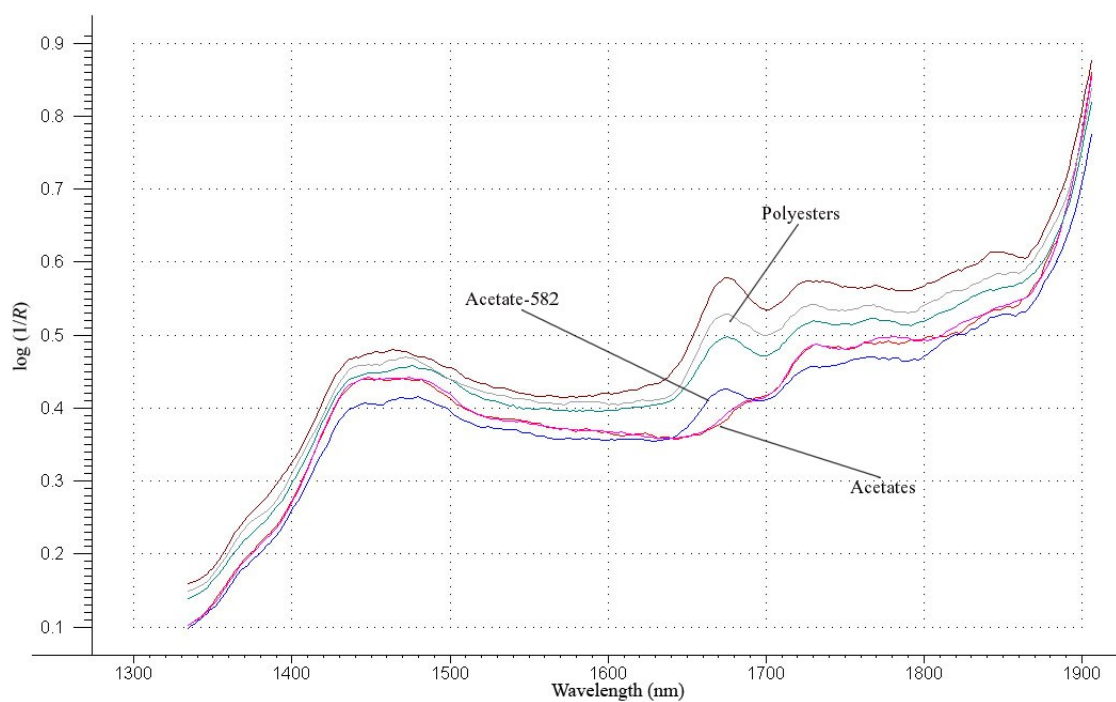


Figure 4.28. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of misclassified acetate sample with representative acetate and polyester fibers

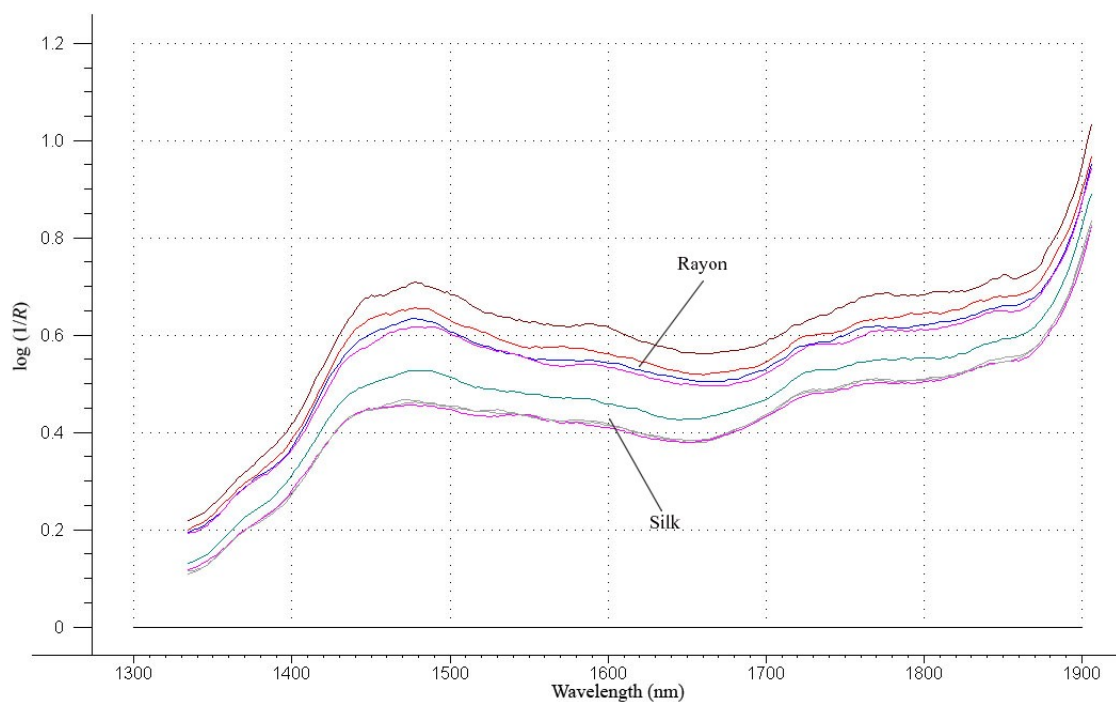


Figure 4.29. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of representative rayon and silk fiber samples

The models for acetate, cotton, polyester, and rayon show no misclassifications for wool fibers. There are, however, a large number of errors in classification when the silk models were used. Both silk and wool are protein based natural fibers. Though the proportions of amino acid residues they possess are different, there are some similarities in the minor constituents. Wool fibers contain, predominantly serine, cystine, arginine, and glutamic acid, while silk fibers contain no cystine and little to no arginine. Both fibers do contain an ample supply serine. Wool and silk contain 7-10 g and 13-17 g of serine per 100 g of fiber, respectively. The amounts of the amino acids valine, phenylalanine, tryptophan, histidine, and lysine are quite similar in both silk and wool. The overall serine content, as well as these other components, could give rise to the misclassifications found with these models.

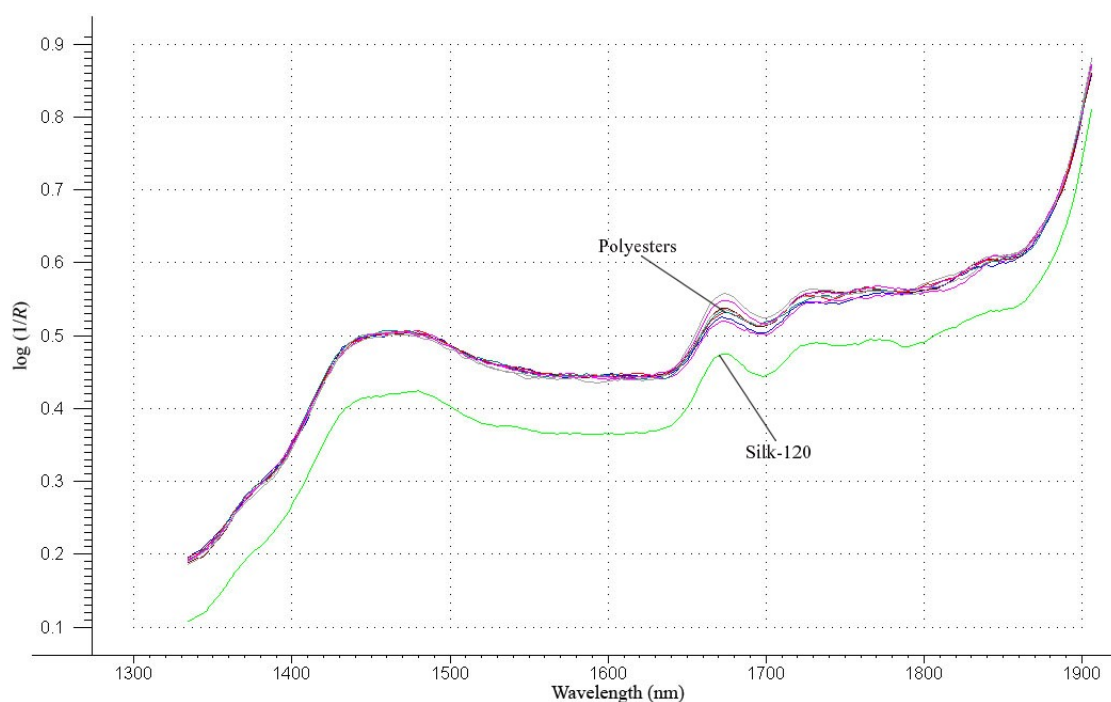


Figure 4.30. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of representative samples and the suspected “fake” silk textile

The final validation step was to use the model sets to attempt to classify the remaining fiber samples in the data set, which include: acrylic, linen, mohair, nylon, olefin, and PVC. The acrylic, mohair, and olefin fibers had no false classifications. The linen fibers were erroneously marked as cotton and rayon. Cotton and rayon are seed fibers, while linen is a bast fiber. This difference notwithstanding, they are all cellulosic natural fibers. The misclassification here is similar to the error of cotton fibers being misidentified as rayon by the rayon designed models. A number of nylon fiber samples were identified as acetates in the validation. The spectra for the samples can be seen in Figure 4.31. The spectra for the acetate and nylon samples overlap and cross, but still have distinctly different spectral bands at 1688 to 1818 nm.

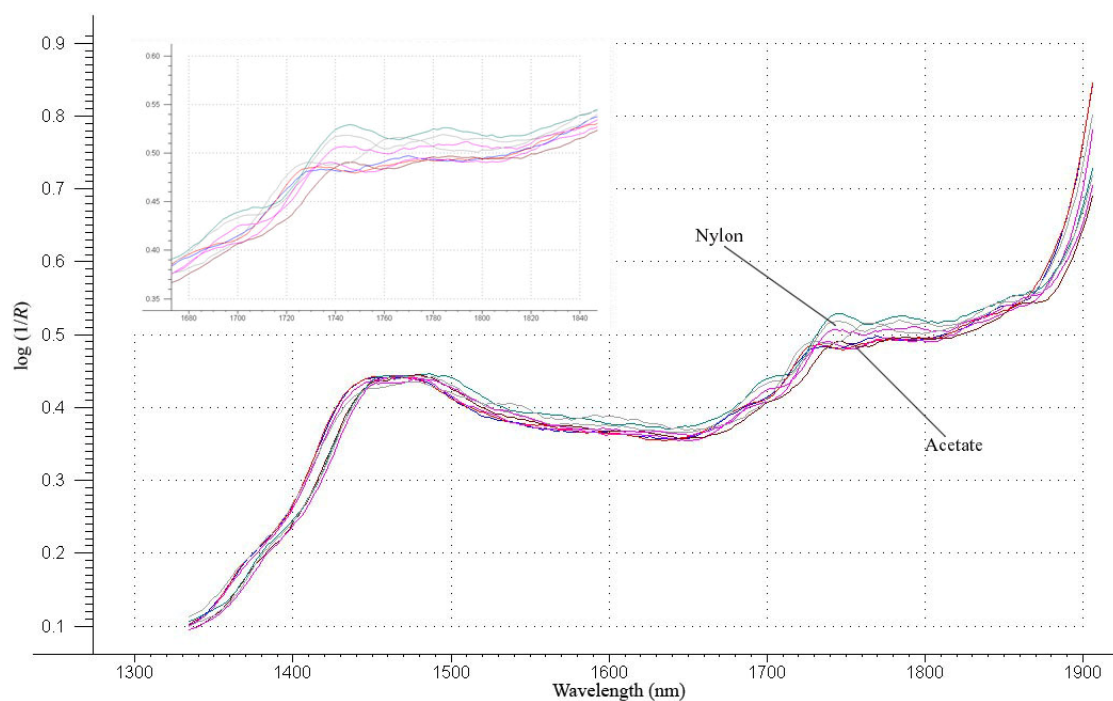


Figure 4.31. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of nylon and acetate samples from 1334-1906 nm with inset from 1674-1846 nm

Conclusion

The preliminary study into the use of soft independent modeling of class analogy (SIMCA) with the diffuse near-infrared reflectance spectra of textile samples as a means of classification was very promising. It was shown that a principal components model could differentiate a grouping of natural fibers from those that were manufactured. The principal component coordinate system plotted the samples along the various axes in order of decreasing variance in order to explain the most amount of variance in the sample matrix with the least number of factors. The PCA models were also shown to segregate samples of two different types of natural fibers. The example shown in this work was between collections of cotton and silk samples. The further investigations also showed the samples of each fiber type would sequester themselves into relatively compact clusters. This clustering behavior was the key to the construction of models to be used in a SIMCA classification.

Though there was no sample pretreatment in this study, it was deemed appropriate to use a Savitzky-Golay smoothing routine, using 5 averaging points at either end of the sample point being averaged. Model classifications improved after the use of the smoothing routine.

Determining the number of principal components in the models was also an obstacle. The residual variance for the different models usually dropped to below 1% at a point far before the calculation of the third principal component (PC). The recommended number of PCs usually was between 1 and 3, which depended on the model and the fiber that model was derived from. Samples of cotton, linen, and rayon are all cellulose

polymers. The subtle deviations in the spectra of these fibers, among others, necessitated the use of more PCs than was recommended by the chemometric software program.

The limited number of samples in the preliminary study was a hindrance; however, it was still possible to build PCA models that could be used in subsequent sample classifications. The expanded study showed that the increased number of samples of the different fiber types helped to better define the fiber types as a whole, spectroscopically. The inherent variations from sample to sample in a single fiber group allowed the PCA models to account for these changes in the classifications. It is what allowed the cotton and rayon samples to be distinguished from one another.

CHAPTER FIVE

Identification of Chemically Treated Textiles by Soft Independent Modeling of Class Analogy Analysis of Diffuse Near-Infrared Reflectance Spectra

Introduction

In the previous studies discussed in Chapter 4, various examples of secondary clustering could be observed within the scores plots of principal components analysis. The degree of variability in the sample population for the study was quite high, due to over 800 samples in the investigation. While the primary clusters of samples observed is mainly due to fiber type, manufacturing process residues, topical finishes, weave pattern, and dye content undoubtedly contribute to variability of fibers of a given type.

In this Chapter, the clustering patterns of laboratory-prepared samples will be investigated in an effort to determine what effect chemical finishing and dyeing has on the behavior of textile samples in a principal components analysis.

Background

Chemical finishing can be defined as the use of chemicals to achieve a desired fabric property. Chemical finishing has always been an important component of textile processing; however, in recent years, the trend of using more high-tech products has increased the interest in the use of chemical finishes. The use of high performance textiles has grown, and with it, the need for chemical finishes to provide the fabric properties required in special-use applications. The estimated amount of textile chemical auxiliaries sold and used globally was about one tenth the world's fiber production.¹⁰⁷

Chemical finishing, also referred to as 'wet' finishing, can change the chemical composition of a fabric as the finish is applied. By doing so, an elemental analysis of the finished fabric would be different than that of the same fabric done prior to the treatment. Chemical treatments can be durable, if they can undergo numerous launderings and dry cleanings without losing their effectiveness. They are non-durable, if they are temporary properties or if the fabric is not intended to be washed. The methods for application of textile finishes vary based on the particular chemical finish, the fiber to be treated, and the machinery available. The type of textile finishes include: softeners, hand builders, easy-care and durable press agents, repellents, soil-release agents, flame retardants, antistatics, anti-pillers, elastomerics, ultraviolet protectants, antimicrobials, as well as other novel finishes like anti-odor and fragrant finishes. A sampling of important chemical finishes and dyes are discussed below.

Softening Finishes

Softening finishes are among the most important chemical after-treatments. The purpose of softening finishes is to achieve an agreeable, soft hand, some smoothness, along with better drape and flexibility. The hand of a fabric is a subjective sensation felt by the skin when a textile fabric is touched with the finger tips and gently compressed. Softeners provide their main effects on the surface of the fibers.

Cationic softeners orient themselves with their positively charged ends toward the partially negatively charged fiber. This process creates a new surface of hydrophobic carbon chains that provide the characteristic excellent softening and lubricity. Anionic softeners mode of action is exactly the opposite of cationic softeners. The negatively charged heads of the softener are repelled away from the fiber. The result is a higher

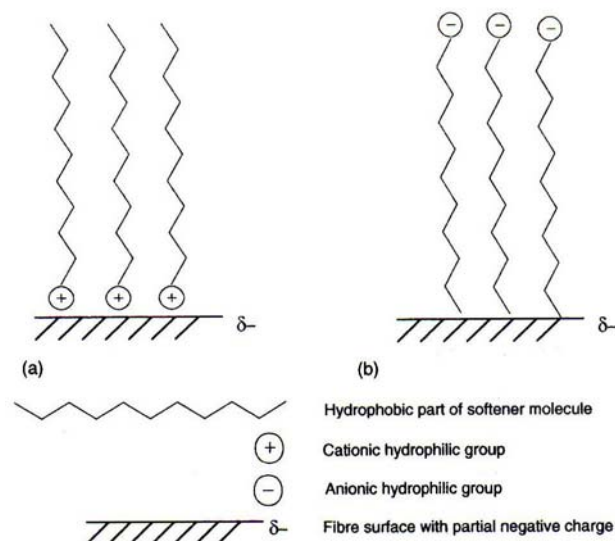


Figure 5.1. Schematic orientation of (a) cationic and (b) anionic softeners on fiber surfaces¹⁰⁷

hydrophilicity and less softening effects than seen with cationic softeners. The modes of action for both cationic and anionic fiber finishes can be seen in Figure 5.1. Non-ionic softeners are also used; however, the orientation of them is dependent on the nature of the fiber surface, where hydrophobic portions of the softener are attracted to hydrophobic portions of the fiber. Common examples of both cationic and anionic softeners can be seen in Table 5.1. Silicone based cationic and non-ionic softeners are also available that provide very high softness. A common example of a silicone softener is polydimethyl siloxane.

Flame-retardant Finishes

The protection of the consumer is the most important criteria for a company manufacturing a product for the general public. Flame-retardant finishes provide textiles with an important performance characteristic by inhibiting the decomposition of fibers when exposed to heat or an open flame. Fire fighters and other emergency personnel

Table 5.1. Chemical structures of typical cationic and anionic softeners

Cationic		
Quaternary ammonium salt	$\begin{array}{c} \text{CH}_2 \\ \\ \text{R}_1-\text{N}^+-\text{R}_2 \\ \\ \text{CH}_3 \end{array} \quad \text{X}^-$	$\text{X} = \text{HSO}_4^-, \text{Cl}^-$ $\text{R}_1 = (\text{CH}_2)_n\text{CH}_3, n = 11-17$ $\text{R}_2 = \text{CH}_3, (\text{CH}_2)_n\text{CH}_3, n = 11-17$
Amine salt	$\text{R}_1-\text{NH}_3^+ \quad \text{X}^-$	$\text{R}_1 = \text{long alkyl chain}$
Imidazoline	$\begin{array}{c} \text{R}_3 \\ \\ \text{CH}_3(\text{CH}_2)_{16}-\text{C} \begin{array}{l} \nearrow \text{N}-\text{CH}_2 \\ \searrow \text{N}-\text{CH}_2 \end{array} \end{array}$	$\text{R}_3 = \text{H}, \text{CH}_2\text{CH}_2\text{NH}_2$
Anionic		
Alkylsulfate salt	$\text{R}_1-\text{O}-\text{SO}_3^- \quad \text{Na}^+$	$\text{R}_1 = \text{long alkyl chain}$
Alkylsulfonate salt	$\text{R}_1-\text{SO}_3^- \quad \text{Na}^+$	

require flame resistant clothing and other items when they enter into a dangerous situation. The military and airline industry also have multiple needs for flame retardant textiles.

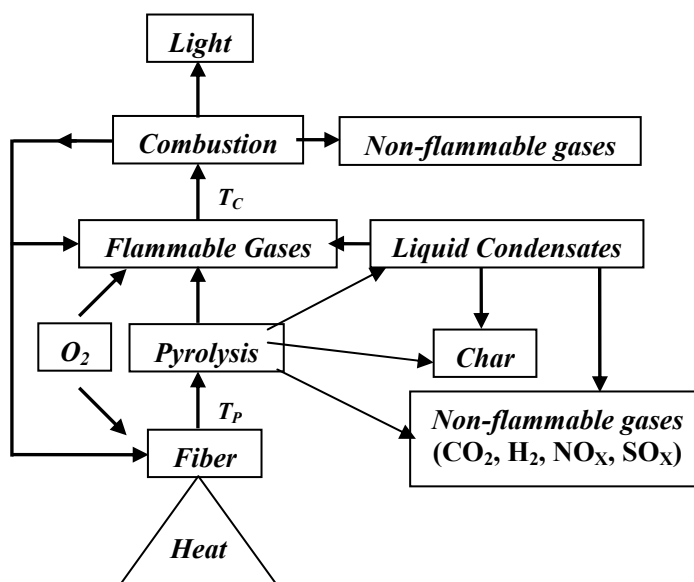


Figure 5.2. Combustion cycle of fibers (Adapted from reference 107. Copyright 2004 CRC Press, LLC)

When a fiber is heated, the fiber's temperature increases until it reaches the pyrolysis temperature, T_p . The fiber begins to decompose, producing carbon dioxide, water vapor, higher oxides of nitrogen and sulfur, carbonaceous char, tars, and flammable gases. If the temperature continues to rise, it will reach the temperature of combustion, T_c , where the flammable gases will then ignite and burn off. The free radical reactions occurring at the point of combustion are highly exothermic, producing large amounts of heat and light. The heat expelled perpetuates the reaction by providing thermal energy to further pyrolyze the fiber and produce more flammable gases.

A main goal of researchers has been to disrupt the combustion cycle. One attempt was the insertion of a heat sink on or in the fiber itself, using materials that thermally decompose through strongly endothermic reactions. If enough energy can be absorbed by the heat sink, the pyrolysis temperature of the fiber will not be reached and no combustion will take place. Aluminum oxide and calcium carbonate have been used as fillers in polymers and coatings to inhibit combustion. Another approach seeks to apply a material that will form an insulating layer around the fibers to be sure they do not reach their pyrolysis temperature. Flame retardancy can also be achieved through a dehydration reaction that influences the pyrolysis reaction to produce less flammable volatiles as well as more residual char. The condensed phase mechanism (Figure 5.3)

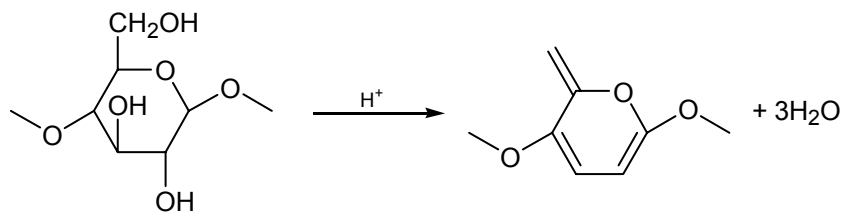


Figure 5.3. Dehydration of cellulose by strong acids

can be seen in the action of phosphorous-containing flame retardants, which, after producing phosphoric acid through thermal decomposition, crosslink hydroxyl containing polymers, thereby altering the pyrolysis to yield less flammable byproducts.

One thermal degradation mechanism for cellulose fibers, such as cotton, rayon, and linen, is the formation of the small depolymerization product levoglucosan. This material and its volatile pyrolysis products are extremely flammable and are the main contributors to cellulose combustion. Compounds that inhibit the formation of levoglucosan (Figure 5.4) are expected to act as flame retardants for cellulose-based fibers. The crosslinking of cellulose polymer chains, as mentioned above, reduces levoglucosan generation, catalyzes dehydration and carbonization. A variety of inorganic salts have long been known as flame retardants. Gay-Lussac proposed the use of borax and ammonium sulfate as a flame retardant for cotton as far back as 1820. The three commercially important flame retardants today are diammonium phosphate, ammonium sulfamate, and ammonium bromide.

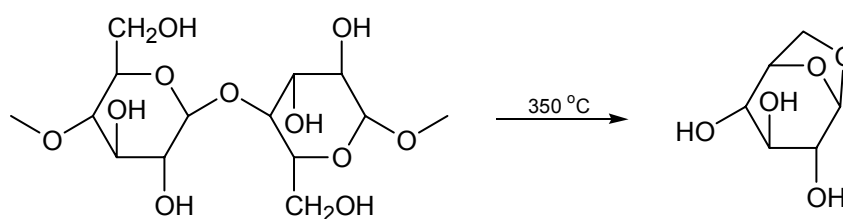


Figure 5.4. Thermal degradation of cellulose polymer to levoglucosan

The three approaches to flame-retard polyester have included additives to the polymer melt, flame-retardant copolymers, and topical finishes. Each of these methods has been used commercially to produce flame-retardant polyester textiles and employ phosphorus- or bromine-containing compounds. A common example is

trisbromopropylphosphate (Figure 5.5a), more commonly known as “tris”. This bromine containing phosphate ester is an extremely versatile and effective product. The

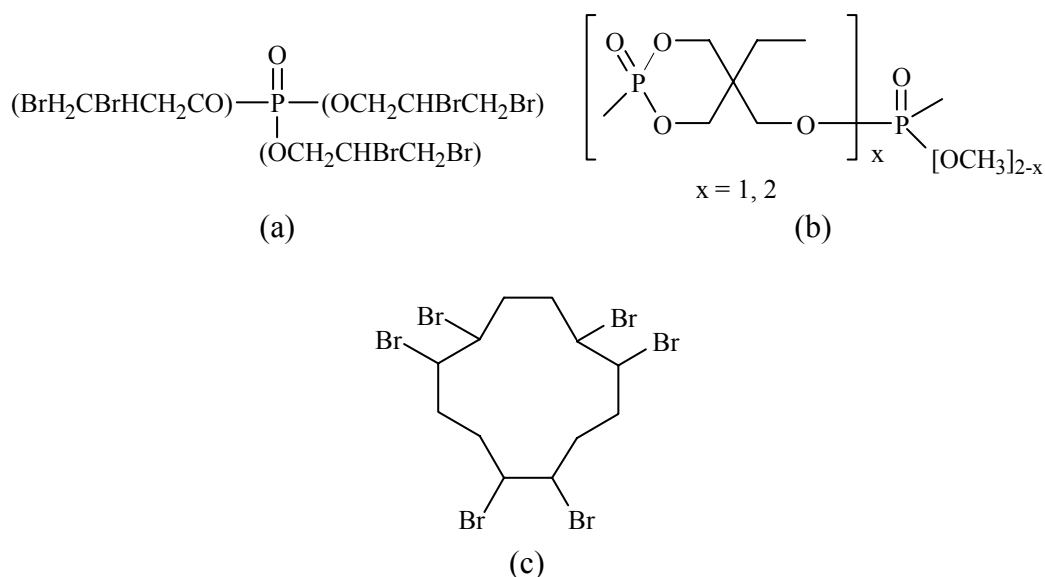


Figure 5.5. Polyester flame retardants (a) trisbromopropylphosphate, (b) cyclic phosphate/phosphonate, and (c) hexabromocyclodecane

use of tris was abandoned, however, after the material was found to be a carcinogen. The current standard for flame retardants of polyester is a mixture of cyclic phosphate/phosphonates (Figure 5.5b) used in a dry-pad-heat set process of 190-210 °C for 0.5-2 minutes. Another approach to durable flame retardant finishes is the use of highly brominated chemicals as topical finishes. One example of these compounds that is particularly useful is hexabromocyclodecane (Figure 5.5c).

Chemical Dyes

Dyes used in coloring textiles absorb discrete packages or quanta of light raising the dye molecule to an excited, higher energy state.⁸⁵ Normally, the absorbed energy is dissipated through increased vibration within the dye molecule as heat. If the energy

absorbed is transferred to the fibers of the textile, chemical damage will occur. Organic molecules that contain a series of unsaturated double bonds can absorb light within a given wavelength range, usually in the ultraviolet (UV). With sufficient conjugation, the molecule will absorb light in a lower energy visible wavelength range and be considered a dye or a pigment. The Munsell system uses standard hues and numerical values for a textile's chroma and lightness to define color.

The dyes used to color textiles must have an affinity for the fibers in the textile to have a uniform application. Physical forces such as hydrogen bonding, van der Waals interactions, and in certain cases chemical covalent bonding are responsible for the bonding of dyes to textile fibers. The chemical composition and basis of application are the most widely used methods of dye classification. The basic classes of dyes are anionic, cationic, and chemically reactive.

1. Dyes containing anionic functional groups. Acid dyes are large dyes containing one or more sulfonic or carboxylic acid salt functional groups. The driving force behind dye diffusion and migration into the textile fibers when using acid dyes is

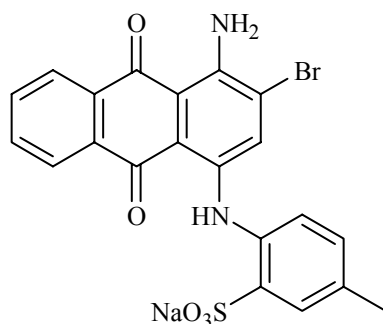


Figure 5.6. Chemical structure for anionic acid dye Acid Blue 78.

the development of a positive charge within a fiber in acidic solution. Fibers suitable for acid dyes include wool, silk, other protein fibers, and nylon. An example of an acid dye, acid blue 78, can be seen in Figure 5.6.

Direct dyes are another subclass of anionic dyes. Typically, direct dyes have a long, narrow, or flat molecular structure. This makes direct dyes highly suitable for coloring cellulosic fibers. This group of dyes often contains azo groups connecting aromatic chromophores. The driving force of diffusion is not based on the development of charge in the fibers, and as a result, direct dyes are applied in basic media, where cellulose is more stable. An example of a direct dye is Direct Red 185 (Figure 5.7).

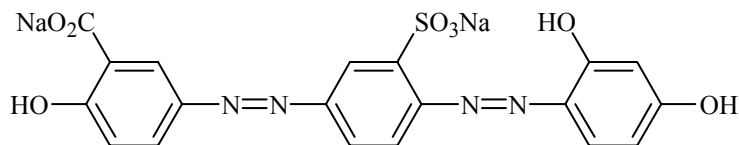


Figure 5.7. Chemical structure for anionic dye Direct Red 185

Mordant dyes are a type of acid dyes that have special sites other than acid salt anionic groups that can react with a metal salt mordant. Common mordants for these dyes include salts of chromium, aluminum, copper, iron, tin, and cobalt. The mordant dye Brown 35 can be seen in Figure 5.8.

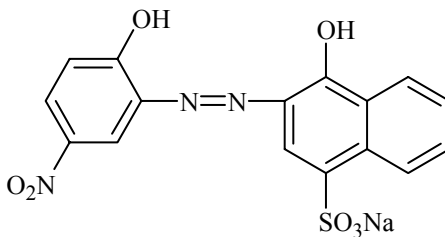


Figure 5.8. Chemical structure of anionic dye Mordant Brown 35

The final group of anionic dyes are called reactive dyes. These molecules have the same basic chemical structures of other acid, direct, and mordant dyes; however, they contain functional groups that are capable of chemically bonding the dye to the fiber of the textile. The dye, Reactive Red 1, seen in Figure 5.9 is a common reactive dye. Cellulosic, protein, and nylon fibers are routinely dyed with this class.

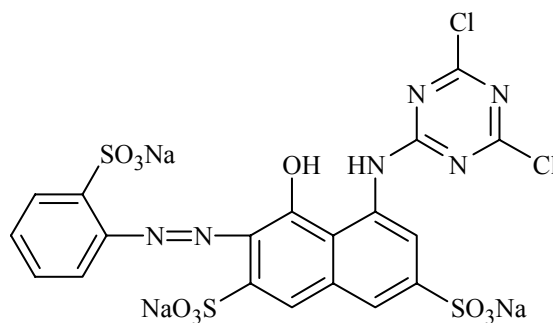


Figure 5.9. Chemical structure of anionic dye Reactive Red 1

2. *Dyes containing cationic functional groups.* Cationic or basic dyes are colored cationic salts of amine derivatives.⁸⁵ The cations will migrate toward negative charges inside the fibers. Cellulose, protein, nylon, acrylic, and specially modified synthetic

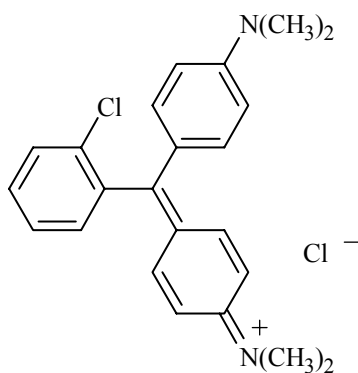


Figure 5.10. Chemical structure of cationic dye Basic Blue 1

fibers perform well when dyed with cationic agents. One drawback to basic dyes is the colorfastness of the finished textile is usually poor. The dye Basic Blue 1, a cationic dye, can be seen in Figure 5.10.

3. *Dyes requiring chemical reaction before application.* The final type of dye is those that require a chemical reaction to proceed application of the dye to the textile fibers. Vat dyes are usually water-insoluble compounds that can be chemically reduced in the presence of base to form a water-soluble and colorless leuco-form of the dye. This species is then applied to the desired fiber. Vat dyes work well with cellulosic and most synthetic fibers. However, care must be taken when using vat dyes with wool and other protein based materials due to the high basicity of the leuco dyer solution, which can damage protein fibers. One common example of a vat dye is Indigo (Figure 5.11).

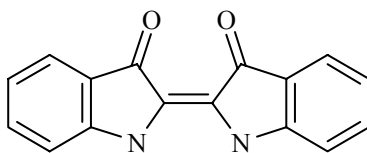


Figure 5.11. Chemical structure of vat dye Blue I (Indigo)

Sulfur dyes are inexpensive complex reaction mixtures of selected aromatic compounds with sodium polysulfide. These dyes are chemically reduced in the presence of base before the application to the textile fibers, and they are then oxidized while on the fiber in the presence of oxygen or by the application of hydrogen peroxide or other mild oxidizing agent. An example of a sulfur dye, Sulfur Green 6, can be seen in Figure 5.12.

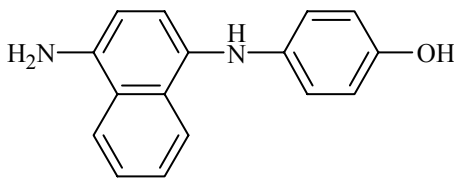


Figure 5.12. Chemical structure of sulfur dye Green-6

Azoic dyes or naphthol dyes formed in situ on the textile fabric through a coupling reaction of an aromatic alcohol or amine, such as naphthol (Figure 5.13a), with a diazonium salt (Figure 5.13b).

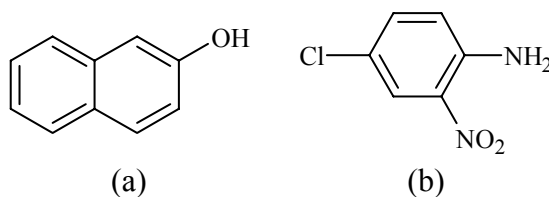


Figure 5.13. Chemical structures of coupling component naphthol (a) and diazonium salt (b) of an azoic dye

Methods and Materials

Experimental

Textile flame retardants ammonium bromide and ammonium sulfamate were obtained from Aldrich Chemical Co., while ammonium phosphate was obtained from Fluka Chemical Co. The textile softeners dimethylpolysiloxane and sodium dodecyl sulfate were also obtained from Aldrich Chemical Co. Chemical dyes hexamethylpararosaniline chloride (Gentian Violet) and Indigo-5,5'-disulfonic acid (Indigo Carmine) were obtained from Fisher Chemical, and 5-(4-Nitrophenylazo) salicylic acid (Mordant Orange) was acquired from Aldrich Chemical Co. Commercially available fabric dyes, Rit scarlet and dark brown, distributed by Phoenix Brands, were purchased at local area stores.

The ammonium sulfamate, ammonium bromide, sodium dodecyl sulfate, Mordant Orange, Indigo Carmine, and Gentian Violet finishing solutions were prepared by dissolving an appropriate amount of solid into 100 ml of deionized water to produce an approximately 1.0 M solution. A 40-ml aliquot of the 2.5 M stock solution of ammonium phosphate was diluted to 100 ml, to afford a concentration of approximately 1.0 M. The Rit scarlet and dark brown dyes were prepared by diluting a 5-ml aliquot to a total volume of 30 ml. A 20% (v/v) solution of dimethylpolysiloxane in chloroform was prepared by mixing 20 ml of dimethylpolysiloxane with 80 ml of chloroform.

Untreated fabric samples of natural linen, 65%/35% polyester (Dacron)/cotton blend, cotton twill, and textured polyester (Dacron) were acquired from Dr. Judith Lusk of the Baylor University Department of Family and Consumer Science. A group five, untreated cotton textiles, desized Texcellona, Texas, pima desized, organic desized, and natural color organic desized, were obtained from the textiles lab at Texas Tech University. The remaining untreated cotton fabrics, organdy, #4 Greige duck, fine filtercloth, bleached combed 80x84, percale sheeting, and bleached mercerized poplin were acquired from Testfabrics, Inc. The fabrics were cut into strips, approximately 1.5 inches wide and 3 inches long.

All the fabric samples were submerged into the different prepared solutions and allowed to soak for 5 to 10 minutes. The samples were then removed and allowed to dry on paper towels in a fume hood for a period of no less than 48 hours. Each sample was then subjected to analysis using NIR reflectance spectroscopy.

NIR Spectroscopy

NIR analysis of textile samples was carried out on a dual-channel spectrometer assembled in our laboratory.¹⁸ The textile samples were placed in the sample holder at the sample window of the integrating sphere. NIR reflectance scans were run from 1100 to 2200 nm, with a sampling interval of 2.0 nm. Measurements were reported as $\log 1/R$, where R is the reflectance. Data were written into ASCII tables by the LabView software and converted to Microsoft® Excel spreadsheets used in chemometric analysis.

Chemometric Analysis

The spectral data, contained in the Excel spreadsheet files, were imported into the Unscrambler© software package (versions 9.1 and 9.6 – CAMO, Inc., Woodbridge, NJ). The data were then transposed to list the wavelengths as the x -variables and the samples as the y -variables. The multivariate principal components analysis was performed on the data. Raw spectral data as well as data subjected to a Savitzky-Golay smoothing routine, using a zero degree polynomial with 5 averaging points on both ends, were used in the principal components analysis.

Results and Discussion

Patterning in the data sets from the principal components analyses had been seen to varying degrees throughout the expanded study discussed in Chapter 4. For instance, in Figure 5.14, the three-dimensional scores plot output for the complete set of wool samples, totaling 192 samples, shows some degree of scatter within the data. A small collection samples (a) can be seen segregating from the main cluster designated (b). Figure 5.15 shows the scores plot analysis for the 287 cotton samples in the data set. The

samples take on the shape of the letter “V”, with the vertex about the cluster labeled (a). The remaining samples in the data swarm fan out towards the points designated (b) and (c).

Both cotton and wool are natural fibers, with cotton being cellulose based and wool is a keratin protein. Similar clustering behaviors can also be seen in manufactured fibers. Figure 5.16 demonstrates a separation between two subclasses of rayon samples. The scores plot clearly shows two groupings at points (a) and (b) with a plane of separation between them.

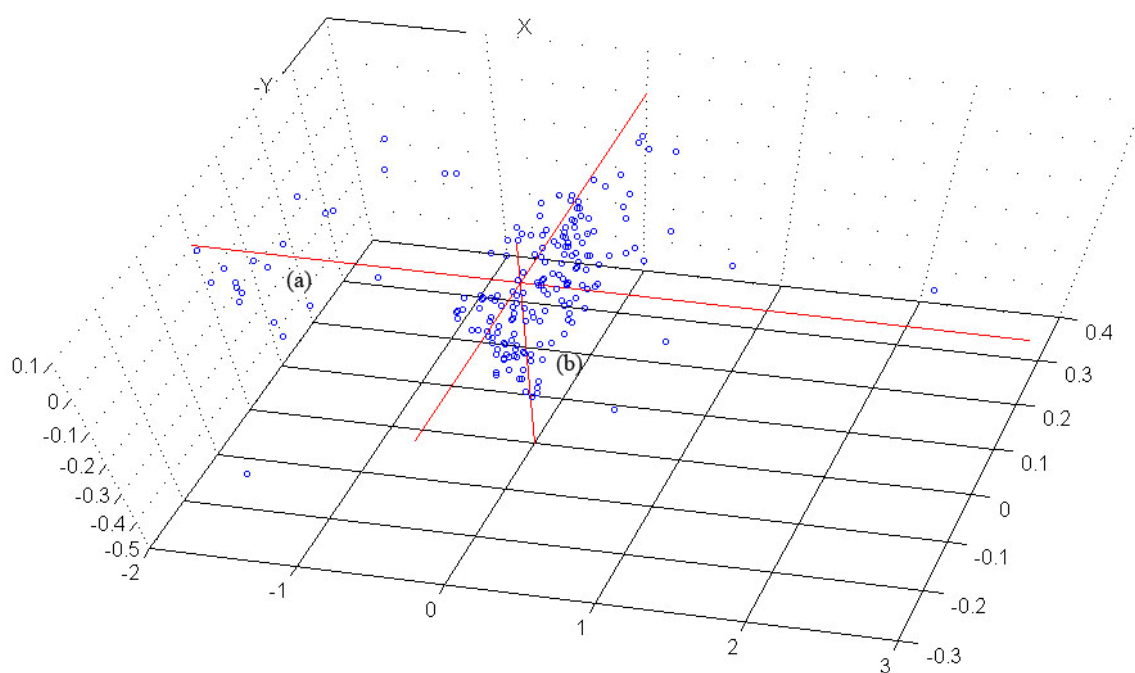


Figure 5.14. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for wool samples with a wavelength window of 1334 to 1906 nm

The behavior seen in rayon is also mirrored in the performance of polyester samples in the principal components analysis (Figure 5.17). Rayon is a regenerated

cellulose fiber, while polyester is completely synthetic. In principal components analysis, the samples in a data set are plotted on a new coordinate system, based on the principal components (PCs) calculated.

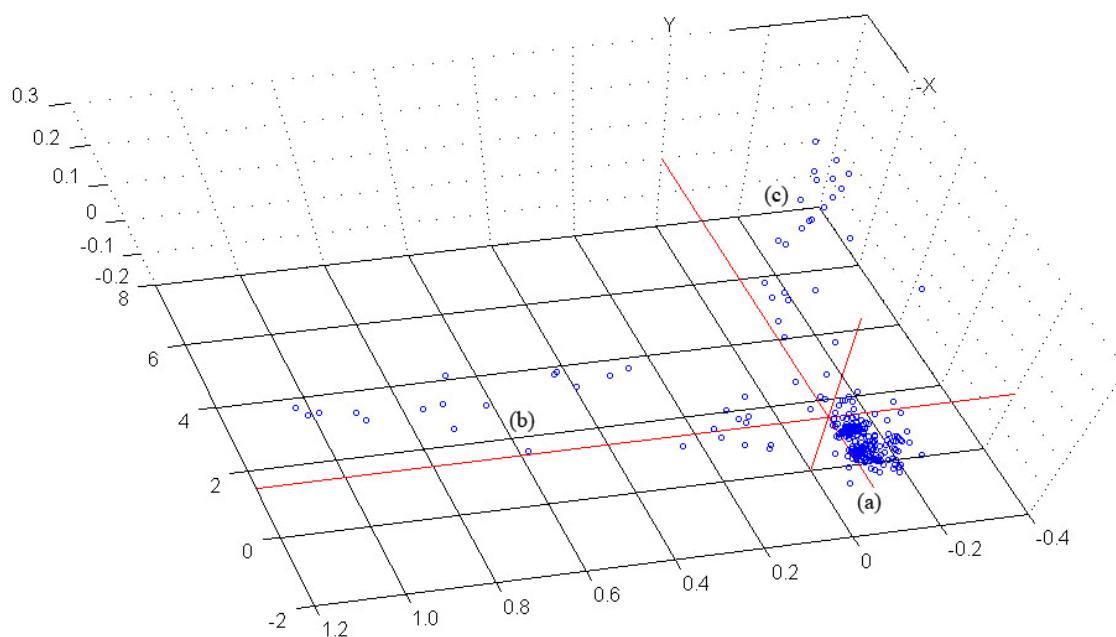


Figure 5.15. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples with a wavelength window of 1334 to 1906 nm

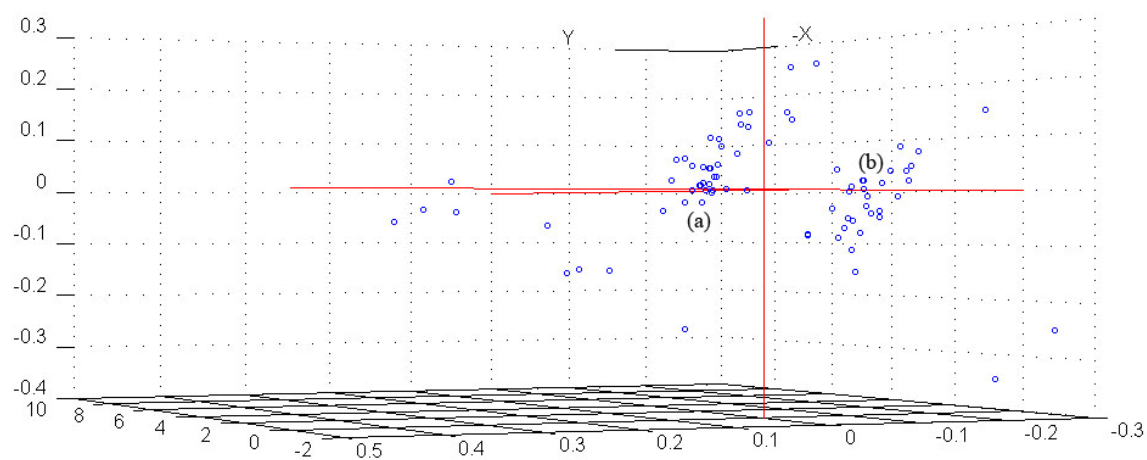


Figure 5.16. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for rayon samples with a wavelength window of 1334 to 1906 nm

The axes for the coordinate system are chosen so that the largest amount of variation in the samples is explained by the first PC. The next largest amount of variation is explained by the second PC, and so on. The differentiation of samples within a single fiber class could be the result of variability in the samples expressed by their relative positions within the scores plot. The variability of the subclasses could be caused by the fiber content, the fiber's origin, manufacturing process residues, topical finishes, dyes, weave pattern, or other factors.

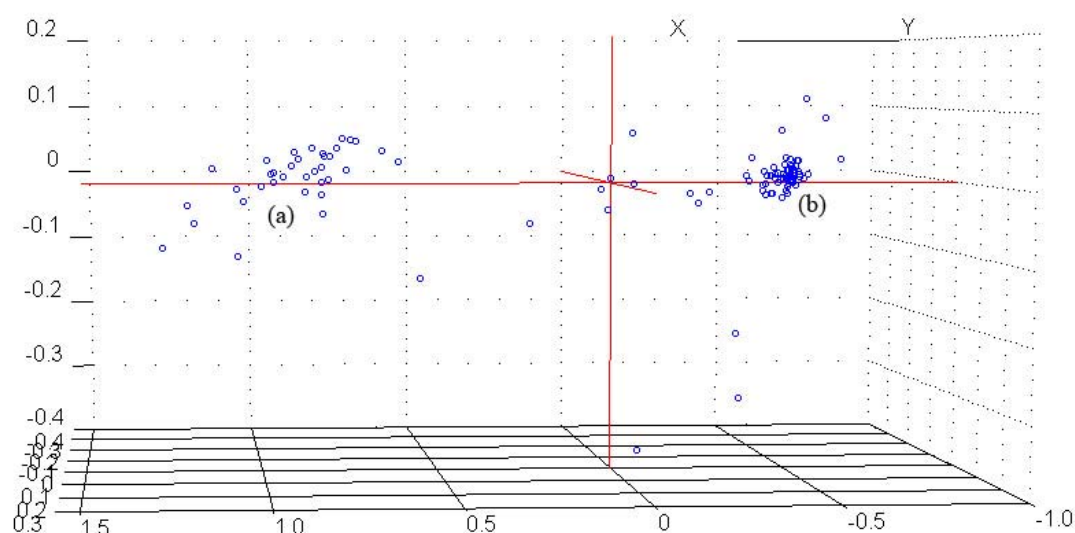


Figure 5.17. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for polyester samples with a wavelength window of 1334 to 1906 nm

Finishes and Groupings

Untreated samples of raw polyester, cotton, linen, and a 65:35 polyester/cotton blend were treated with ammonium bromide and ammonium phosphate. One sample of each textile was left untreated to be used as a control. The diffuse near-infrared

reflectance spectra for these samples can be seen in Figure 5.18. The samples of the different fiber types stack on top of one another from polyester up to linen. The polyester/cotton blend sample falls between the samples of 100% cotton and polyester. Between approximately 1400 to 1650 nm, the spectra for the blend samples are analogous to that of the pure cotton and linen samples, which are both cellulosic fibers. The blend fibers then diverge from the pure samples, and the characteristic peak of polyester at approximately 1672 nm is seen.

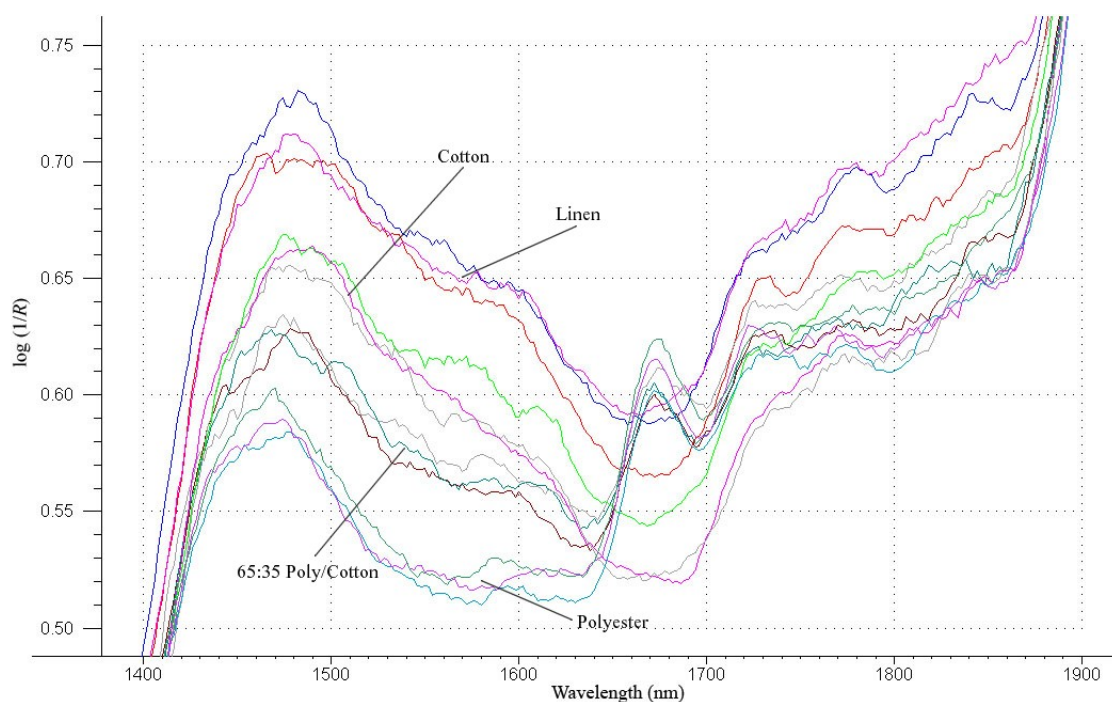


Figure 5.18. Diffuse near-infrared reflectance spectra for cotton, linen, 65:30 polyester/cotton blend, and polyester (untreated, ammonium bromide, ammonium phosphate) from 1400 to 1900 nm

The three-dimensional scores plot for the four treated sample sets can be seen in Figure 5.19. The samples cluster together in a linear fashion by fiber type, as was seen in the preliminary and expanded studies in the previous Chapter. The first two principal

components in the plot are along the X- and Y-axes. The bulk of the information contained in both PCs seems to deal with the fiber type of the given sample, which gives rise to the clusters. The third principal component, the Z-axis, differentiates the samples based on the topical finish applied.

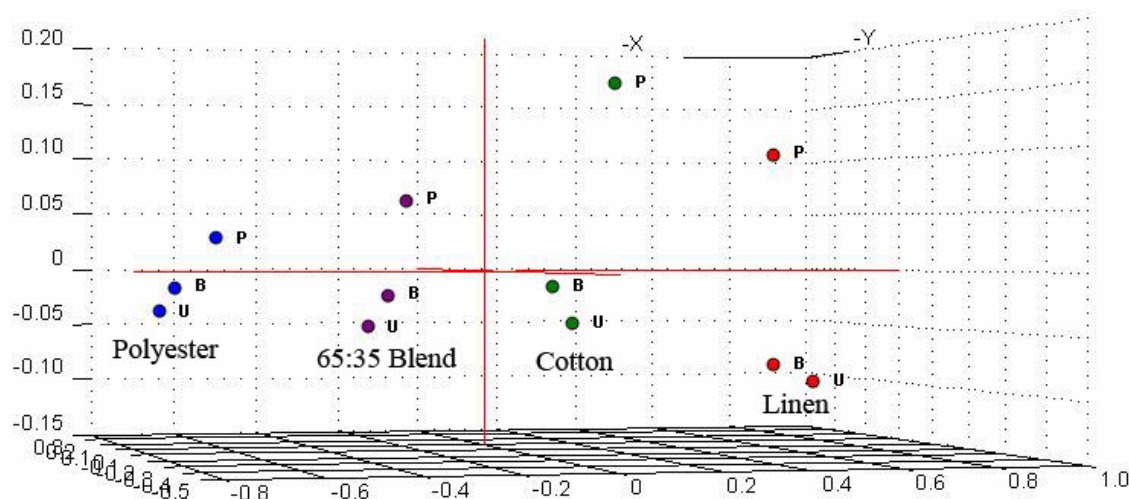


Figure 5.19. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton, linen, 65:30 polyester/cotton blend, and polyester samples that were untreated (U), and coated with ammonium bromide (B) and ammonium phosphate (P) with a wavelength window of 1334 to 1906 nm

The next step was to add an additional finishing treatment to see if the trend observed would continue in the four fiber types with the introduction more variability. Ammonium sulfamate, also a flame retardant, was coated onto samples of the cotton, linen, polyester, and blend fabrics. The near-infrared reflectance spectra were collected and compared to those of the previous analysis (Figure 5.18). The spectra of the samples are grouping together based on their fiber type, as was seen in Figure 5.20. A PCA was performed on the spectral data, and it can be seen in Figure 5.21. The results, as would be expected, are similar to the PCA done for bromide, phosphate, and untreated

samples. The introduction of the ammonium sulfamate samples did not disrupt the clustering of the samples based on fiber type. The relationship of the samples to one another in the Z-direction (third PC) is still roughly linear.

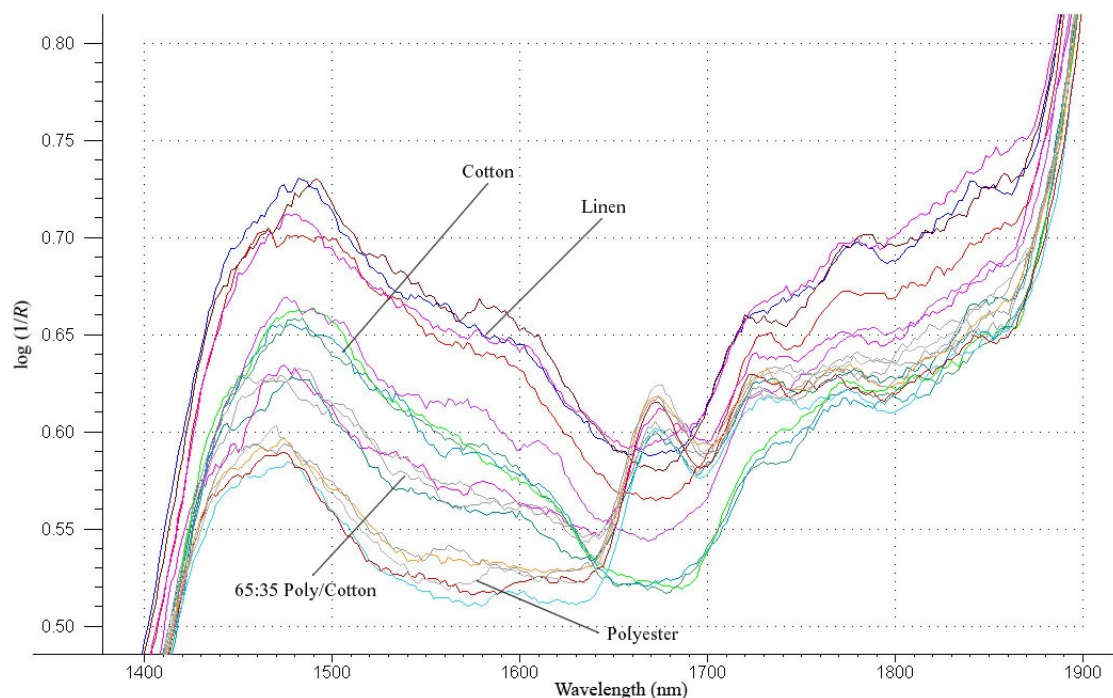


Figure 5.20. Diffuse near-infrared reflectance spectra for cotton, linen, 65:30 polyester/cotton blend, and polyester (untreated, ammonium bromide, ammonium phosphate, ammonium sulfamate) from 1400 to 1900 nm

The samples in the previous two examples were from four different fiber classes. The underlying goals of this study also include the explanation of fragmentation for samples in a data swarm of a single fiber type. The fiber chosen for this analysis was cotton, due to ease of obtaining samples as well as the importance of cotton in the global market. A group of untreated cotton samples, 11 in all, were gathered from the Baylor University Department of Family and Consumer Science, Texas Tech University, and

Testfabrics, Inc. The samples were each treated with a variety of chemical finishes, and their NIR reflectance spectra were taken.

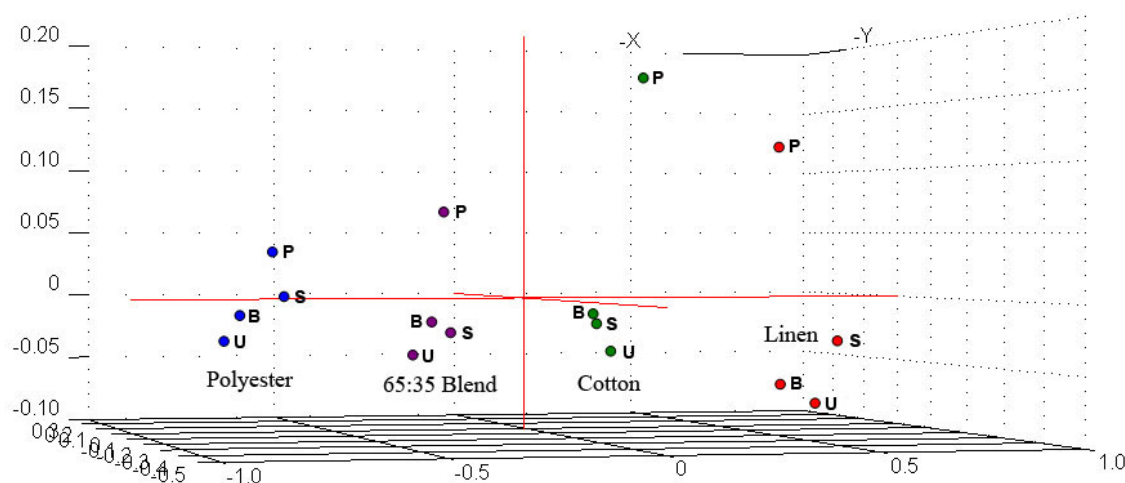


Figure 5.21. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton, linen, 65:30 polyester/cotton blend, and polyester samples that were untreated (U) and coated with ammonium bromide (B), ammonium phosphate (P), and ammonium sulfamate (S) with a wavelength window of 1334 to 1906 nm

In Figure 5.22, a three-dimensional scores plot for cotton samples of the 11 types, each coated with ammonium phosphate and an uncoated sample as a control. There is inherent structure to the resulting data swarm. The control samples, designated (a), are partitioned from the phosphate treated samples centered around point (b). This is the first instance where it can be clearly seen that samples of a given fiber type will segregate based upon the finishing the textile receives.

The situation is next complicated by the introduction of a new fiber coating, dimethylpolysiloxane (DMPS). DMPS is used as a textile softener, often used to impart a particular feel and weight to a fabric. The resulting PCA scores plot from the spectra of the samples can be seen in Figure 5.23. As before, the untreated samples (a) partition

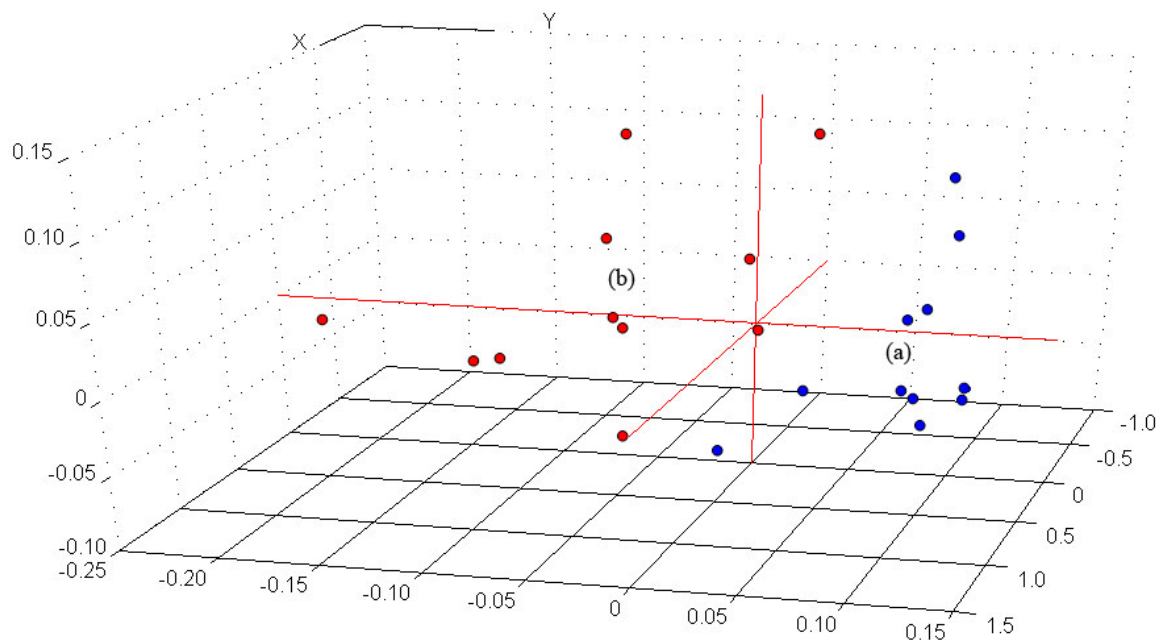


Figure 5.22. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples that were (a) untreated and (b) treated with ammonium phosphate for a wavelength window of 1334 to 1906 nm

themselves away from the ammonium phosphate coated samples (b). The DMPS coated samples (c) have formed a disparate cluster, removed from both the other treated or control samples.

Figure 5.24. shows the three-dimensional scores plot for the PCA of cotton samples coated with Gentian violet dye, DMPS, and a control. The data set fragments into groupings consistent with each of the fabric treatments. The structure of the clusters formed is linear and not a spherical shape as was seen in the previous example. The clusters stack against one another from the untreated (a), to the Gentian violet dyed, and finally to the DMPS coated. This analysis demonstrates that a chemical textile softener and dye can be distinguished from one another, as well as from an untreated fiber of the sample class.

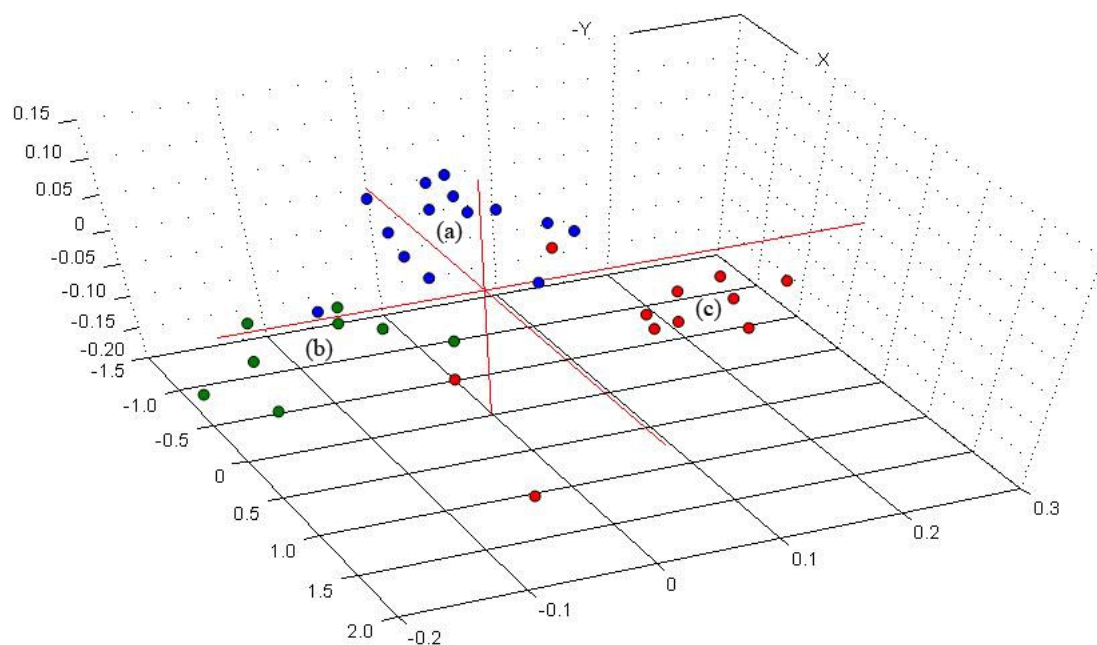


Figure 5.23. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples that were (a) untreated, (b) treated with ammonium phosphate, and (c) dimethylpolysiloxane for a wavelength window of 1334 to 1906 nm

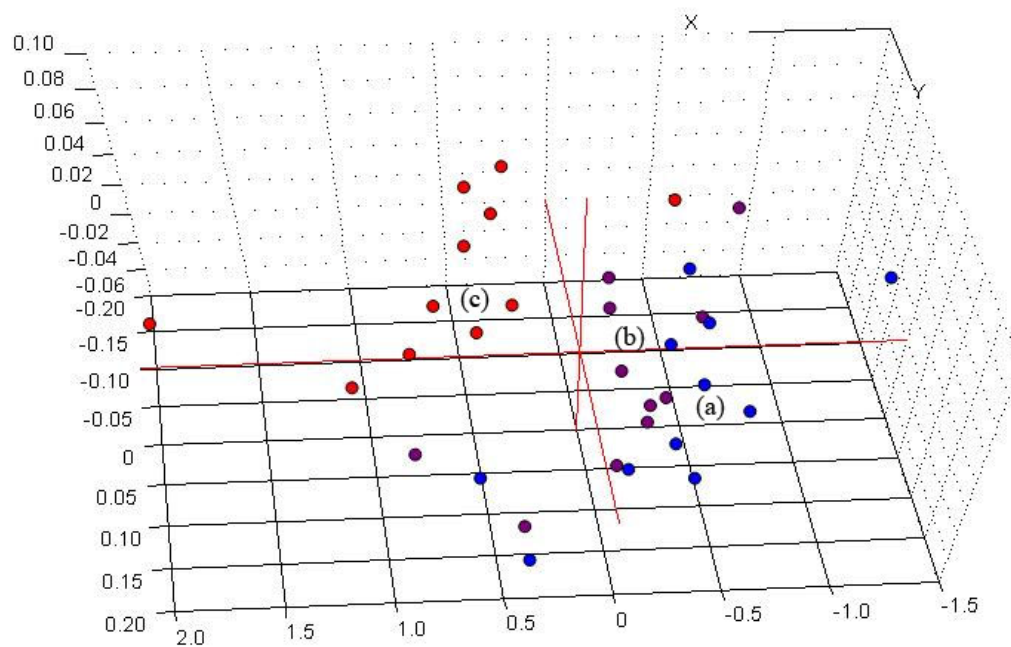


Figure 5.24. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples that were (a) untreated, (b) treated with Gentian Violet, and (c) dimethylpolysiloxane for a wavelength window of 1334 to 1906 nm

The final PCA was conducted on cotton samples that were untreated, as well as coated with DMPS and sodium dodecyl sulfate (DSS). Both of the coatings are softening agents used in the textile industry. The scores data can be seen in Figure 5.25. The samples segregated themselves based on the finishing treatment applied, as was seen in previous examples. The delineation of the sample group is reminiscent of planes, stacked side by side.

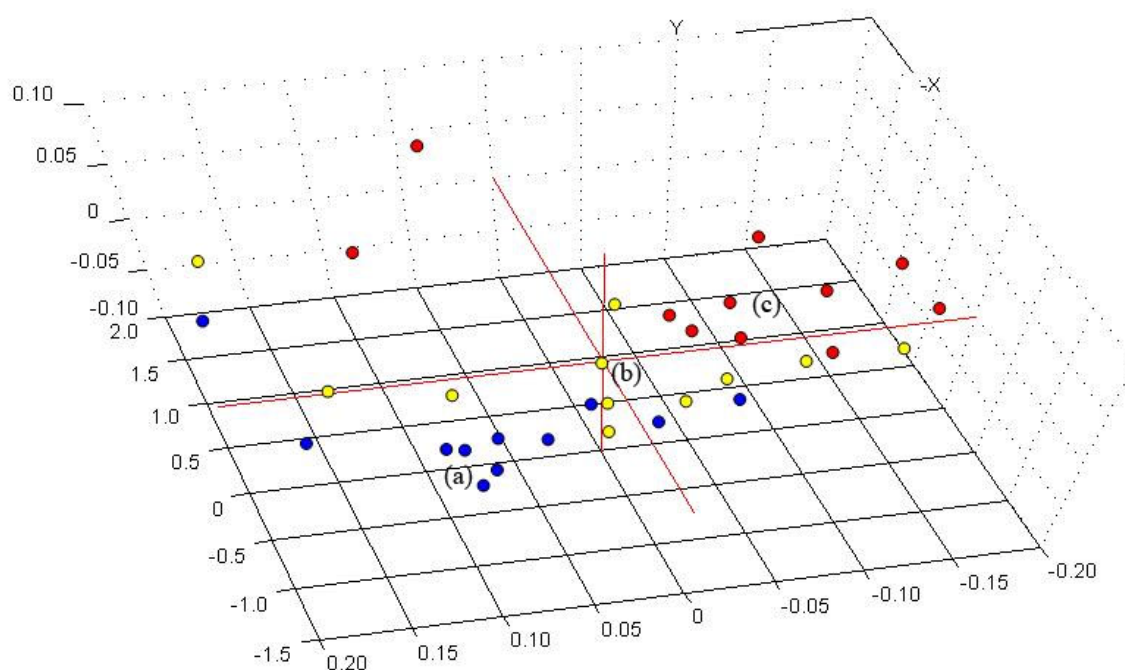


Figure 5.25. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples that were (a) untreated, (b) treated with sodium dodecyl sulfate, and (c) dimethylpolysiloxane for a wavelength window of 1334 to 1906 nm

The patterning of the spectral data in the scores plot of a principal components analysis is not always dictated by the topical treatment applied. The modeling of the data onto a new coordinate system is different for each group of samples introduced. The principal component axes for the scores plot could have been influenced, by varying

degrees, by variables other than the chemical finish. When examining the scores plot for cotton samples treated with Gentian violet and Indigo Carmine, the samples were not clearly being decomposed into clusters based on the dye that was selected. The scores plot is depicted in Figure 5.26. At first glance, there was no clear pattern to the data points in the swarm. However, upon a closer assessment, there was logic to the data distribution. The samples were correlating to one another, based upon which type of cotton the textile sample was cut from. For example, the samples of cotton organdy, treated with the two different dyes, were plotted very close to one another. The oval shapes in Figure 5.26 group the samples of the same cotton type together.

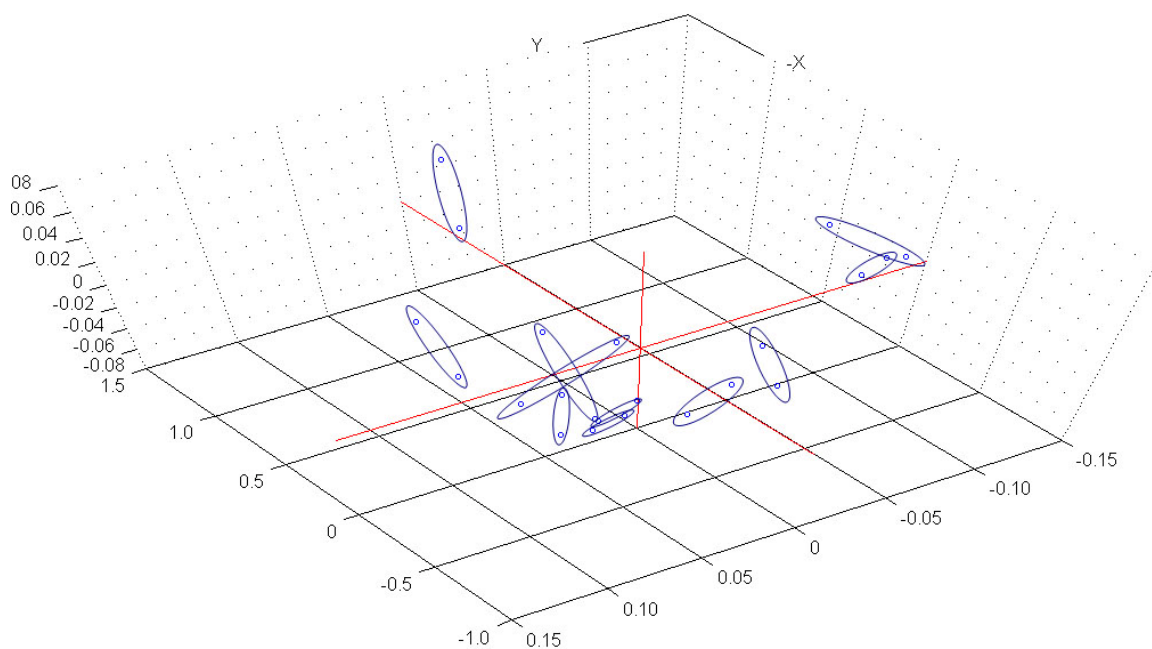


Figure 5.26. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for cotton samples that were treated with Gentian Violet and Indigo Carmine for a wavelength window of 1334 to 1906 nm

Water Content Modeling

There was no attempt to control or monitor the relative humidity or temperature of the laboratory when the NIR measurements for these studies were carried out. Official textile analytical laboratories have their temperature and relative humidity levels moderated to afford relatively constant conditions. The Baylor Sciences Building, where the Center for Analytical Spectroscopy is housed and these spectral measurements were recorded, does not internally regulate humidity. The overtones and combination bands of water play a major role in the NIR spectrum. Therefore, it could be a concern that the spectral variations seen and modeled in these studies could be merely a result of modeling water content variations in the textile fibers. Samples of the same class had their NIR reflectance spectra collected at roughly the same period on the same day. Some sample sets were analyzed over several days due to the large number of samples in the set. Barring any sudden change in climatological conditions, the water content of the samples would have remained nearly constant. The following data analysis was conducted to determine if water was the overriding factor involved in the SIMCA classifications.

Three cotton types, cotton organdy, Greige cotton duck, and cotton fine filtercloth were selected as the basis for the water-content modeling analysis. A total of ten cotton samples from each type were treated with ammonium bromide, ammonium phosphate, ammonium sulfamate, dimethylpolysiloxane, sodium dodecyl sulfate, Gentian violet, Mordant orange, Indigo Carmine, Rit® scarlet, and Rit® dark brown. The spectra for each treated group were collected on the same day. The complete set of spectra can be seen in Figure 5.27. If the clustering pattern of the samples, and in turn, the resulting

classification based on that patterning, was based on the moisture content of the fibers themselves, the samples would group themselves according to the day their spectra were collected.

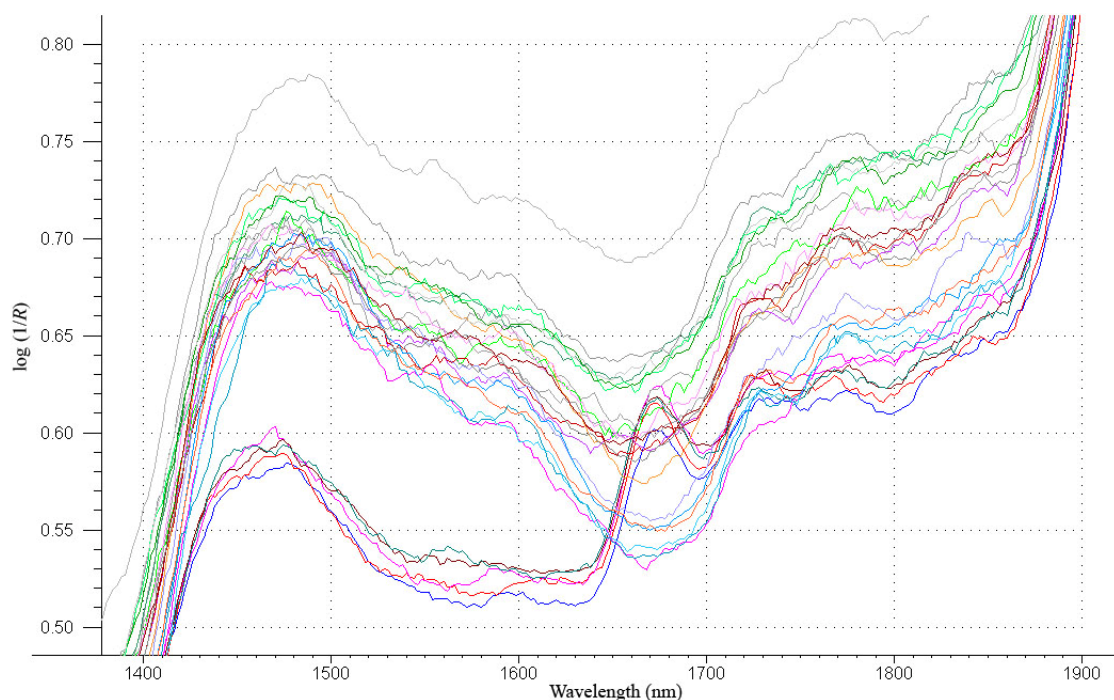


Figure 5.27. Diffuse near-infrared reflectance spectra for cotton organdy, Greige cotton duck, cotton fine filtercloth, and polyester coated with various fabric treatments from 1400 to 1900 nm

The resulting PCA scores plot from the spectral data (Figure 5.28) shows that the samples clustered based upon the cotton type and not the type of finish applied. This is consistent with the findings when Gentian violet and Indigo Carmine were applied to other cotton samples. The samples of the same treatment were analyzed on the same day. The moisture content for those samples would be similar, and as such, would be expected to cluster together. The reality is, however, that the samples clustered together based solely on the cotton type, no matter what the treatment was, or when the sample's spectrum was collected.

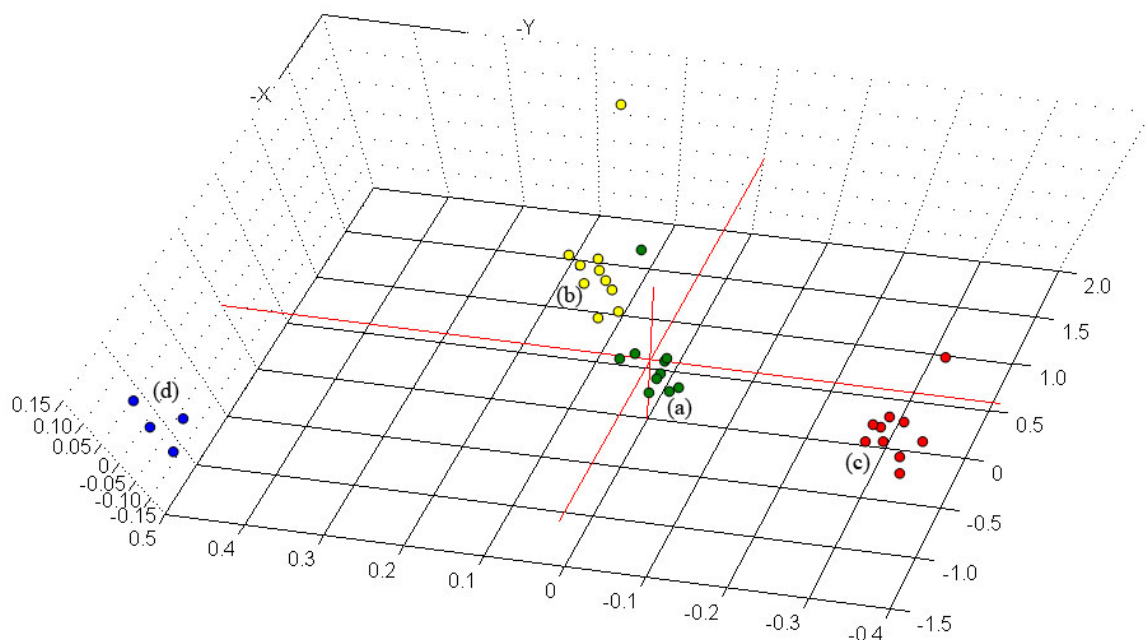


Figure 5.28. Groupings in three-dimensional scores plot for principal components analysis of diffuse near-infrared reflectance spectra for (a) cotton fine filtercloth, (b) cotton organdy, (c) Greige cotton duck, and (d) polyester samples treated with various finishes for a wavelength region of 1334 to 1906 nm

Conclusion

The clustering patterns in principal components analysis scores plots derived from diffuse near-infrared reflectance spectra of textiles can be affected by the different finishes and dyes applied to the textile's surface. Laboratory-prepared samples from different fiber categories were treated with a variety of finishes including flame retardants and had their NIR spectra collected. The data swarm of the samples clustered into groups based upon the fiber type in the sample. Secondary structure in the clusters was attributed to increased influence on the third principal component calculated by the factors relating to the flame retardant treatment.

Further work demonstrated that subclasses within a data cluster for a particular fiber type can be related to the finish treatment applied to the textile. The behavior is seen when cotton textiles of different types are coated with a select group of flame

retardants, softeners, and dyes. Sometimes, the clustering of the sample data is more heavily influenced by the type of cotton fiber used. The subclass structure of the data cluster is often more heavily influenced by a variable other than chemical finish. This is shown clearly when Gentian violet and Indigo Carmine were applied to two types of cotton samples.

The SIMCA classifications completed throughout this work are based upon the fiber type of the textile being analyzed. Moisture levels in the sample set could vary depending on the ambient temperature and relative humidity of the laboratory. But, when samples that had their NIR spectra collected on a given day were compared to samples taken on different days, the clustering of the samples was in accordance with the type of fabric that was present and not moisture content.

CHAPTER SIX

Conclusions

The modeling of spectral data with chemometric methods can aid in the analysis of complex matrices of information to elucidate important information that would otherwise be buried. Determining the hotness level of chili peppers is of importance in agricultural, as well as in the food industry as a means of quality control for spicy condiments. The variations in pungency levels of raw materials become important in food processing operations designed to produce final food products rated as mild, medium, and hot.¹⁸ Previous work done in our laboratory showed a need for larger sample populations, smaller sample size, and ease of preparation.

A collection of partial-least-squares (PLS) regression models were prepared from the ultraviolet spectra of alcoholic habanero pepper extracts. Total capsaicinoid content (capsaicin and dihydrocapsaicin) was determined independently by high-performance liquid chromatography to be included in the multivariate analyses. Other studies investigating pepper hotness have used a much larger sample size, which is detrimental to the analysis if the end goal is to determine the pungency of a single chili. The variations between peppers of the same type are lost due to the homogeneous nature of the samples after they have been blended together. The sample size for this study was reduced to approximately 4 g of a single habanero pepper. The validation samples of 9 test sets were shown to have, on average, a root mean square error of prediction (RMSEP) of ± 5

ppm. The models were derived from 86 wavelengths from the UV spectra of 21 calibration samples and used 4 PLS-components,

A full independent validation set of pepper extracts was prepared to assess the integrity of the PLS regression models. The samples had an average RMSEP of ± 4 ppm. The samples were then stored for a period of 25 weeks to evaluate the stability of the model over time. The UV lamp of the spectrophotometer failed and was replaced during the 25 week period. The average RMSEP for the samples was again only ± 5 ppm, which demonstrates that the model developed to predict total capsaicinoid concentrations was stable enough, not only over time, but also due to instrumental adjustments.

Chemometric analysis was also instrumental in the development of a diffuse near-infrared reflectance database used in the soft independent modeling of class analogy (SIMCA) to classify commercially available textiles. The total number of samples in the database was 826. There were 13 fiber categories, which included acetate, acrylic, cotton, linen, mohair, nylon, olefin, polyester, PVC, rayon, silk, wool, and a collection of fiber blends. No sample pre-treatments were done for the samples in the preliminary and expanded studies reported on in Chapter 4.

The preliminary study was completed to gauge whether or not it was feasible to use the SIMCA classification methodology with NIR reflectance spectra to classify textiles. SIMCA is a disjoint principal components analysis (PCA) technique, where samples of the same type are analyzed by PCA. The resulting models are then combined with others derived from spectra of different fibers to make classification decisions. The limited number of samples in the preliminary study was an obstacle; however, it was still possible to show that samples of different fiber types segregated into different regions of

space in the scores plot data of the PCA. It was possible to construct PCA models, though limited, to correctly classify cotton, silk, and polyester textiles.

The expanded study in the latter portion of Chapter 4 shows that the expanded database, with more of each individual kind of fiber type, improved the predictive abilities of the PCA models in the SIMCA classification. The number of principal components (PC) to be used for each fiber type model was of particular importance. Increasing the number of PCs in a model will often limit or eliminate any misclassifications in the validation set. Models were developed that correctly classified acetate, cotton, polyester, rayon, silk, and wool fibers. The classifications for these six fiber types were 94% accurate, on average.

Clustering of samples by fiber type was the anticipated and desired result for the PCA of the diffuse near-infrared spectra of the textile samples in both the preliminary and expanded studies in Chapter 4. The samples in the data sets for a single category would segregate themselves into different clusters. This behavior could be the result of a variety of factors, including: fiber type and origin, manufacturing process residues, topical finishes, weave pattern, and dye content. The clustering patterns were investigated by first analyzing laboratory-prepared samples of different fiber types coated with various finishes, and then by examining various finishes on only one fiber type. The results in Chapter 5 show that subclasses within a given data cluster can be the result of a particular finishing treatment applied to a textile.

APPENDICES

APPENDIX A

Selection of Principal Components for SIMCA Classification

The principal components analyses used for the soft independent modeling of class analogy classifications of textiles had a total of six principal components (PC) calculated. The selection process for the optimum number of PCs used in the final predictions is discussed in Chapter 4. Depicted in this appendix are the x-loadings plots for the six principal components calculated for the acetate, cotton, polyester, and rayon samples, Figures A.1 to A.5, respectively. Each variable has a loading along each model principal component. The loadings show how well a variable is taken into account by the model components. The larger the magnitude of the loading, the more it contributes to the meaningful variation in the data. Loadings can also be useful in interpreting the meaning of each model component.

The information contained within each variable, and subsequent PC, can improve the predictive abilities of the PCA models used in SIMCA. The residual validation variance of a model plotted as a function of principal components can help to determine how many PCs are necessary to explain the spectral variation in the sample set. Often times, the residual validation plot will indicate the use of fewer PCs than were calculated to explain the data set. However, as was seen in Chapter 4, classifications generally performed better when more PCs were used than were recommended.

When Figure A.1 is examined, the loadings plots for PCs two through five are fairly smooth. It is not until the sixth principal component is reached that the structure of

the plot begins to deteriorate slightly. This demonstrates that the sixth principal component still has information that could be useful in classifying samples of similar origin (similar spectra) such as cotton and rayon, which are both cellulose. Conversely, an examiner must still be careful that all the PCs selected to be used in a classification are explaining spectral variations and not simply instrument noise.

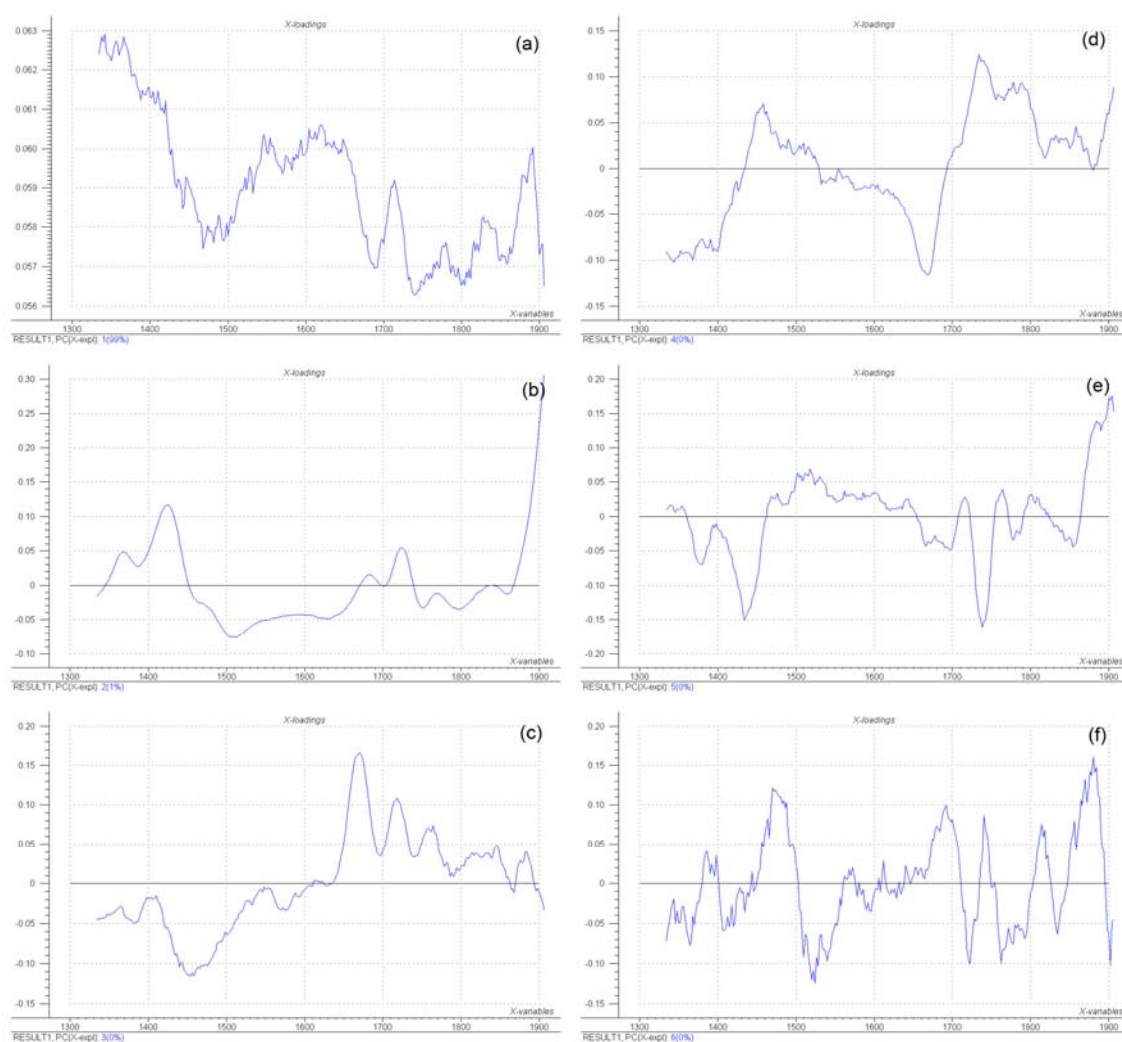


Figure A.1. Principal components analysis X-loadings plots for acetate at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components

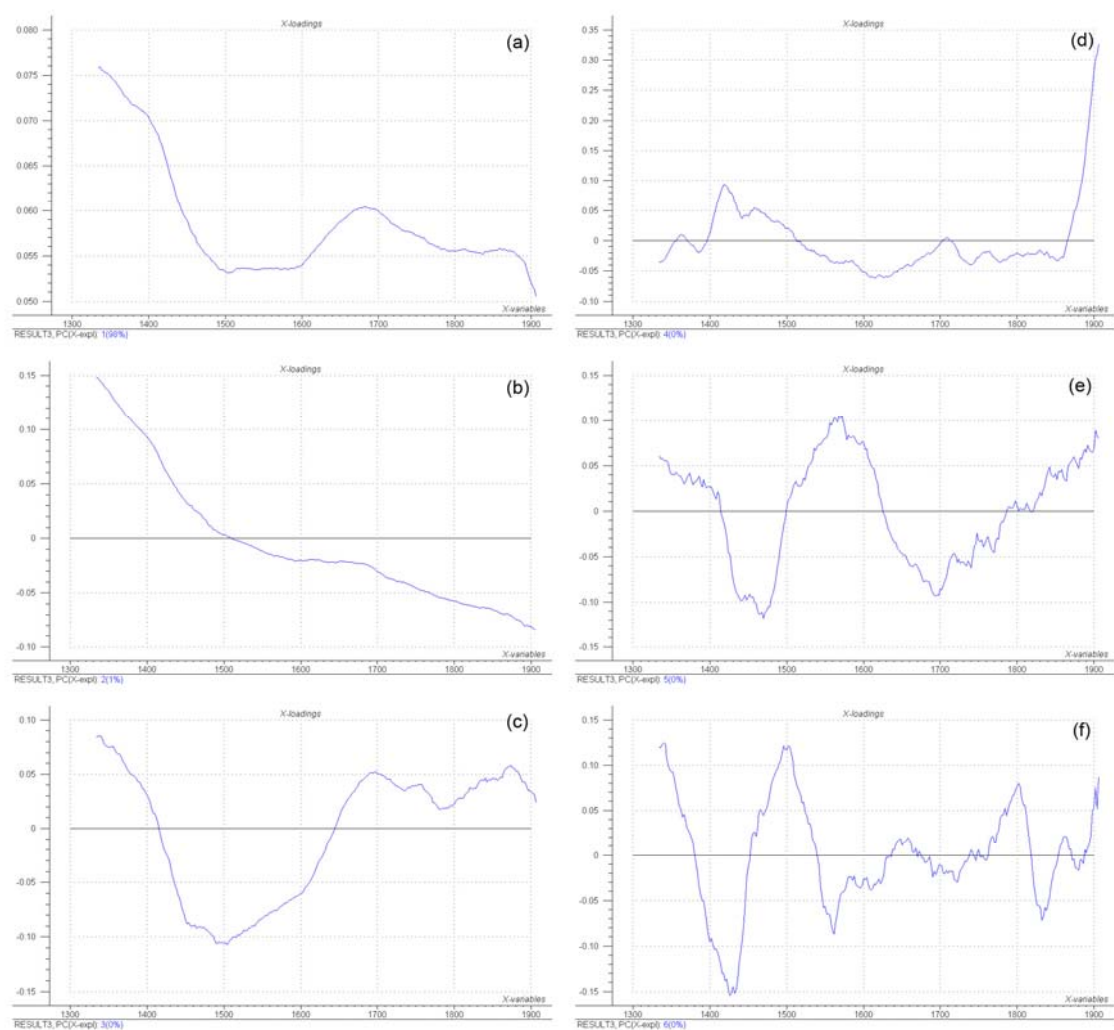


Figure A.2. Principal components analysis X-loadings plots for cotton at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components

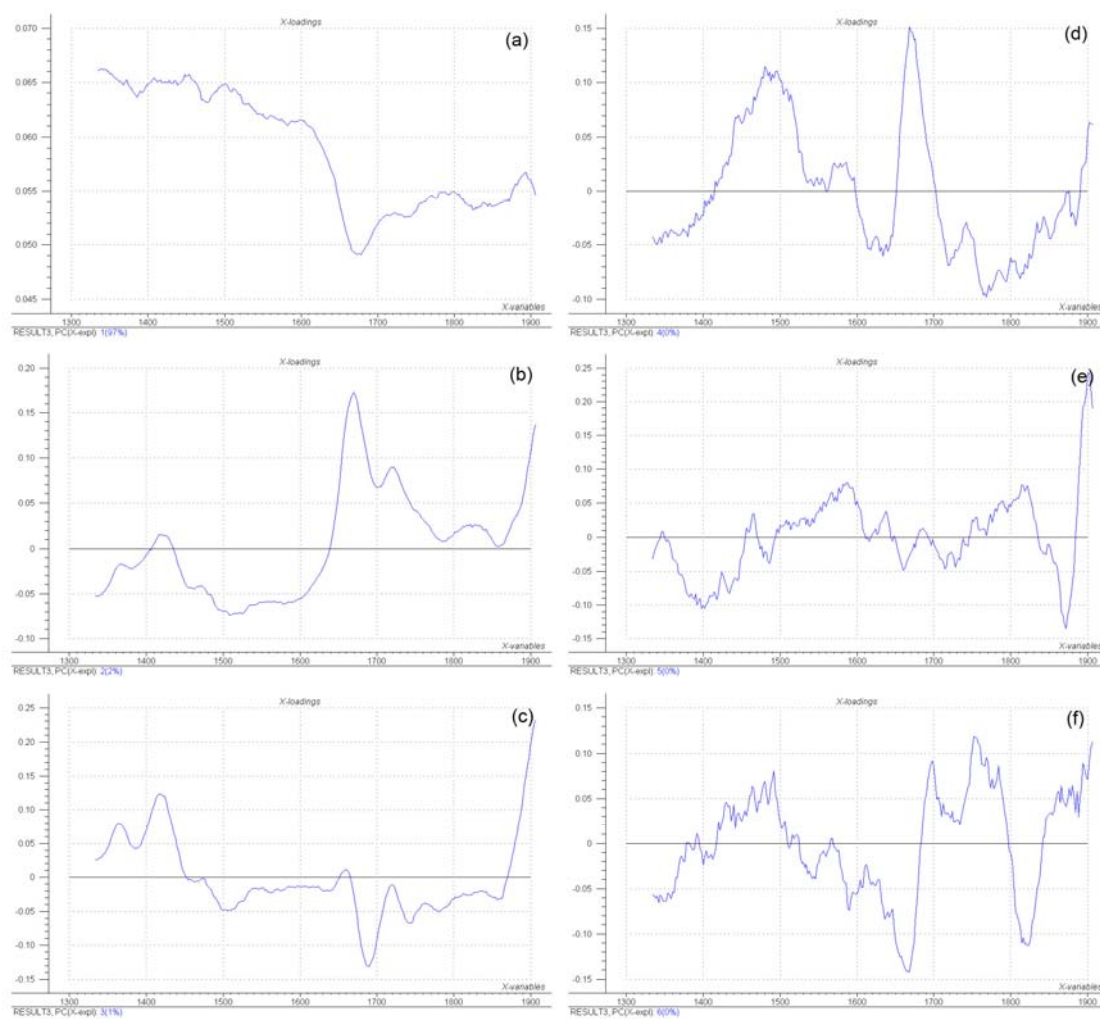


Figure A.3. Principal components analysis X-loadings plots for polyester at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components

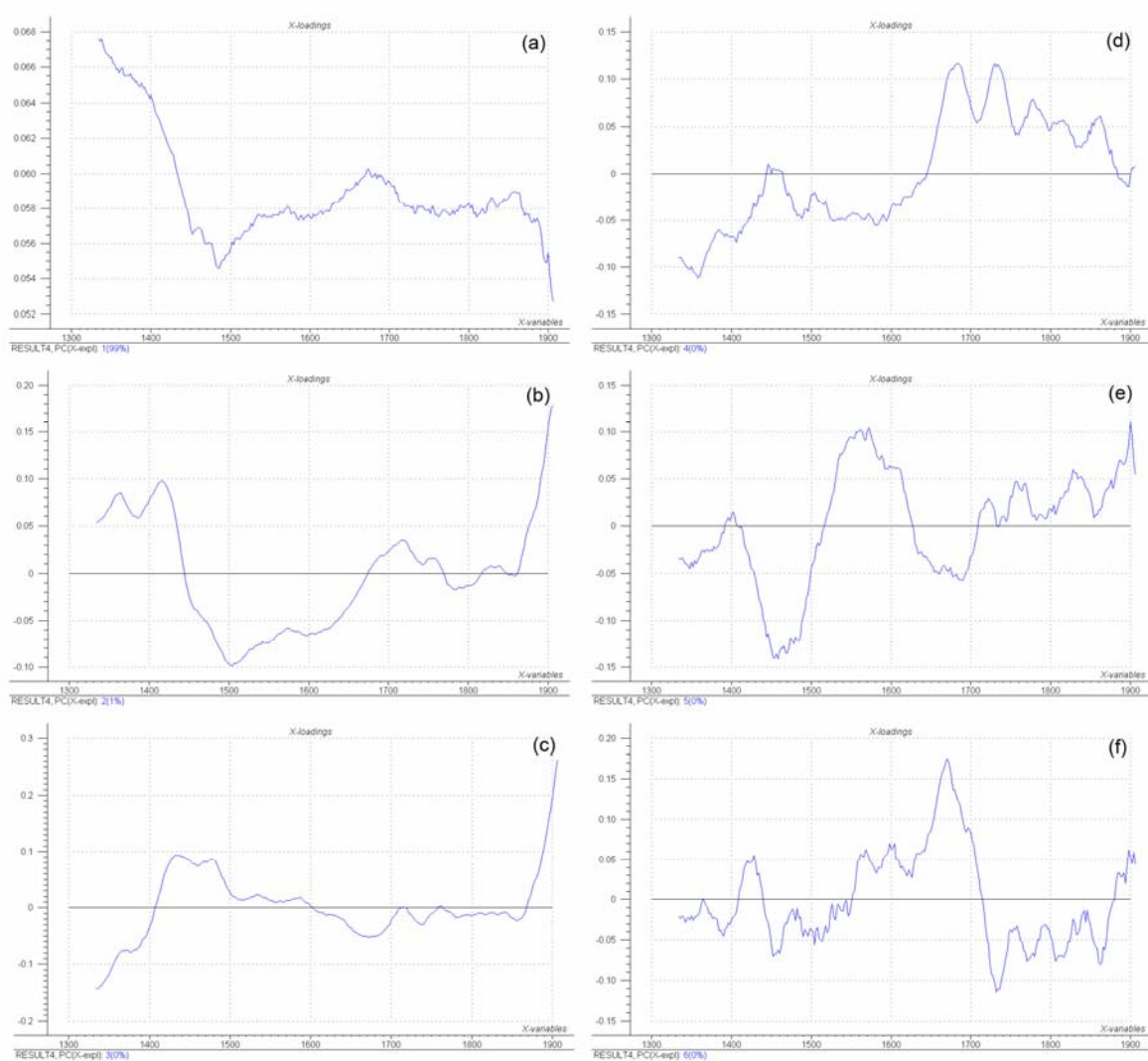


Figure A.4. Principal components analysis X-loadings plots for rayon at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components

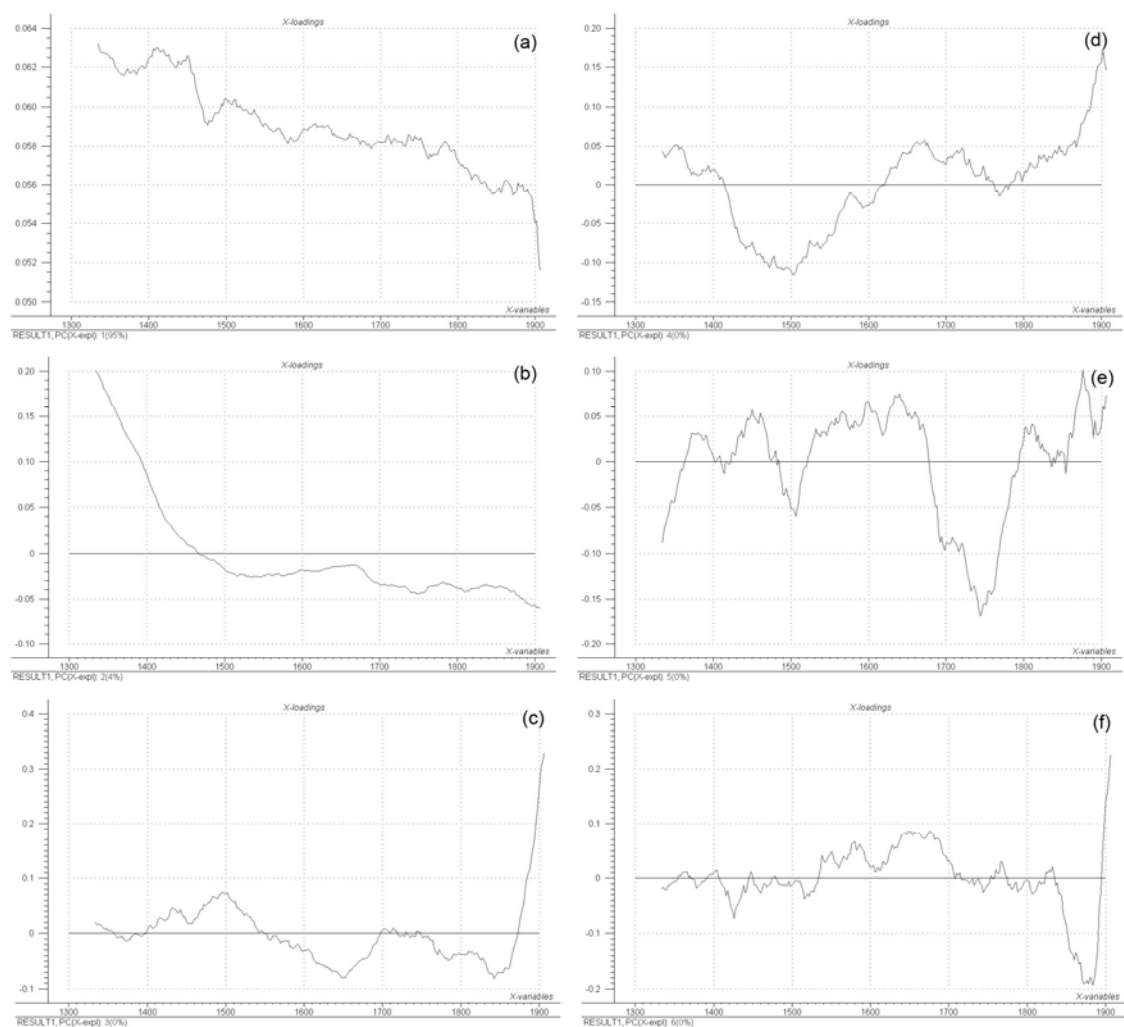


Figure A.5. Principal components analysis X-loadings plots for wool at the (a) first, (b) second, (c) third, (d) fourth, (e) fifth, and (f) sixth principal components

APPENDIX B

Publisher Permissions

CAMO Software Inc, 1480 Route 9 North, Suite 209, Woodbridge, NJ 07095



March 27, 2007

Christopher B. Davis, PhD
 Busch Research Group
 Center for Analytical Spectroscopy
 Baylor Sciences Building Rm E.155R.1
 Office: (254) 710-2009 Mobile: (281) 684-0479
 Baylor University - Department of Chemistry & Biochemistry

Dear Dr. Davis,

Per your request, the purpose of this letter is to give permission for you to use the following information:

Title of Book/Periodical/Journal: Multivariate Data Analysis - in practice:
 Full Name of Author(s): Kim H. Esbensen
 Volume/Issue/Edition #: 5th Edition
 Date of Publication (copyright date or issue date): 2002 CAMO AS
 ISBN (books only): 82-993330-3-2
 ISSN (journals/periodicals only): NA
 Beginning on page NA, line NA, with the words NA
 Ending on page NA, line NA, with the words NA
 Illustration/Figure No. 3.29 on page 66
 Chart No. NA on page NA

Thank you for referencing the Introducer book.

Best regards,

Dongsheng Bu, PhD
 Principal Scientist

CAMO Software Inc.
 Aspen Corporate Park One, Suite 209
 1480 Route 9 North
 Woodbridge, NJ 07095

(Office) 732 602 8886 ext. 32
 (Fax) 732 602 8887

Email: dbu@camo.com

Confidential

Davis, Christopher B

From: Sayo Fakayode [sayo@lsu.edu] **Sent:** Sat 3/31/2007 7:45 PM
To: Davis, Christopher B
Cc:
Subject: Re: Congratulation
Attachments:

Dear Dr. Christopher B. Davis,

Congratulation on the defense of your dissertation. I wish you the very best on your future academic and professional endeavor.

Concerning the permission to use the figures in my dissertation, you have my unreserved permission to reproduced, copy or use the requested Figures in your dissertation.

Sayo Fakayode, PhD

From: "Davis, Christopher B" <Christopher_B_Davis@baylor.edu> on 03/30/2007 17:15 EST

Sent by: "Davis, Christopher B" <Christopher_B_Davis@baylor.edu>

To: <sayo@lsu.edu>
cc:

Subject: Hey Sayo!

Hey Sayo,

I don't know if you heard, but I defended my dissertation. The entire private defense took less than an hour. Dr. Gipson only asked me one question, and didn't give me any corrections! I think he was either really busy and didn't have time to read my dissertation, or the fact he was heading out of town the next day might have had something to do with it.

Anyway, I wanted to ask you, can you send me an email back stating that I can reproduce some figures that are in your dissertation. One of the new guidelines for writing our dissertations is that we have to include written approval for the use of any other sources figures. I want to use two from your dissertation.

Davis, Christopher B

From: Moss, Marion (ELS-OXF) [m.moss@elsevier.co.uk] **Sent:** Tue 3/27/2007 6:20 AM
To: Davis, Christopher B
Cc:
Subject: RE: Obtain Permission
Attachments:



Dear Mr Davis

We hereby grant you permission to reproduce the material detailed below at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.
2. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier".
3. Your thesis may be submitted to your institution in either print or electronic form.
4. Reproduction of this material is confined to the purpose for which permission is hereby given.
5. This permission is granted for non-exclusive English rights only. For other languages please reapply separately for each one required.
6. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely

Marion Moss
Senior Rights Assistant
Elsevier Ltd
The Boulevard
Langford Lane
Kidlington
Oxford
OX5 1GB

Tel : +44 1865 843280
Fax : +44 1865 853333
E-mail : m.moss@elsevier.com

-----Original Message-----

From: Christopher_B_Davis@baylor.edu [mailto:Christopher_B_Davis@baylor.edu]
Sent: 26 March 2007 19:56
To: Rights and Permissions (ELS)
Subject: Obtain Permission

This Email was sent from the Elsevier Corporate Web Site and is related to Obtain Permission form:

Product: Customer Support
Component: Obtain Permission
Web server: <http://www.elsevier.com>
IP address: 10.10.24.148
Client: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; .NET CLR 1.0.3705; .NET CLR 1.1.4322; InfoPath.1)
Invoked from: http://www.elsevier.com/wps/find/obtainpermissionform.cws_home?isSubmitted=yes&navigateXmlFileName=/store/prod_webcache_act/framework_support/obtainpermission.xml

Request From:
Graduate Student Christopher B Davis
Baylor University
P.O. Box 97348
Waco
TX
United States

Contact Details:
Telephone: 254-710-2009
Fax: 254-710-4272
Email Address: Christopher_B_Davis@baylor.edu

To use the following material:
ISSN/ISBN: 0-444-82853-2
Title: Handbook of Chemometrics and Qualimetrics: Part B
Author(s): B.G.M. Vandeginste, D.L. Massart
Volume: 20B
Issue: N/A
Year: 1998
Pages: 229 - 229
Article title: N/A

How much of the requested material is to be used:
Figure 33.17 ONLY

Are you the author: No
Author at institute: No

How/where will the requested material be used: [how_used]

Details:

Dissertation Title: Application of Chemometric Analysis to UV-Visible and Diffuse Near-Infrared Reflectance

https://fs-exchange.baylor.edu/exchange/Christopher_B_Davis/Inbox/RE:%20Obtain%20P... 3/27/2007



Legal/Permissions
One Lake Street
Upper Saddle River, NJ 07458
Fax: 201-236-3290
Phone: 201-236-3564
Cheryl.Freeman@pearsoned.com

Apr 11, 2007

PE Ref # 128303

CHRISTOPHER DAVIS
Center for Analytical Spectroscopy
Baylor University
PO Box 97348
Waco, TX 76798

Fax #: 254-710-4272

Dear Christopher:

You have our permission to include content from our text, *FABRIC REFERENCE, 1st Ed. by HUMPHRIES, MARY*, in your doctoral dissertation for your course at Baylor University.

Content to be included is:

p. 5 Figure 1.7
p. 14 Figure 2.3
p. 15 Figure 2.4
p. 22 Figure 2.8

Please credit our material as follows:

HUMPHRIES, MARY, FABRIC REFERENCE, 1st Edition, © 1996. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ

Sincerely,

Cheryl Freeman, Permissions Administrator



Legal/Permissions
One Lake Street
Upper Saddle River, NJ 07458
Fax: 201-236-3290
Phone: 201-236-3564
Cheryl.Freeman@pearsoned.com

Apr 11, 2007

PE Ref # 128305

CHRISTOPHER DAVIS
Center for Analytical Spectroscopy
Baylor University
PO Box 97348
Waco, TX 76798

Fax #: 254-710-4272

Dear Christopher:

You have our permission to include content from our text, *TEXTILES, 9th Ed.* by **KADOLPH, SARA J.; LANGFORD, ANNA L.**, in your dissertation for your course at Baylor University.

Content to be included is:

pps. 36,37,42,43,52,55,56,275 (Figs. 4.4,4.7,4.12,4.13,5.3,5.7,5.8,16.4)

Please credit our material as follows:

KADOLPH, SARA J.; LANGFORD, ANNA L., TEXTILES, 9th Edition, © 2002. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ

Sincerely,

Cheryl Freeman, Permissions Administrator

04/13/2007 09:31 FAX

JOHN WILEY & SONS INC

001/002



republication
<republication@wiley.com>

03/26/2007 02:42 PM

Please respond to
republication
<republication@wiley.com>

To <republication@wiley.com>

cc

bcc

Subject Republication/Electronic Request Form

A01_First Name: Christopher
A02_Last Name: Davis
A03_Company Name: Baylor University
A04_Address: P.O. Box 97348
A05_City: Waco
A06_State: TX
A07_Zip: 76798
A08_Country: USA
A09_Contact_Phone_Number: 254-710-2009
A10_Fax: 254-710-4272
A11_Emails: Christopher_B_Davis@baylor.edu
A12_Reference:
A13_Book Title: Chemometrics: A Practical Guide
A14_Book_or_Journal: Book
A14_Book_Author: Kenneth R. Beebe
A15_Book_ISBN: 0-471-12451-6
A16_Journal_Month:
A17_Journal_Year:
A18_Journal_Volume:
A19_Journal_Issue_Number:
A20_Copy_Pages: p. 63-67 (Figures 4.5-4.9) and p. 187 (Figure S.3)
A21_Maximum_Copies: 5
A22_Your_Publisher: Baylor University
A23_Your_Title: Applications of Chemometric Analysis to UV-Visible and Diffuse
Near-Infrared Reflectance Spectra
A24_Publication_Date: 4/15/2007
A25_Format: print
A31_Print_Run_Size:
A41_Ebook_Reader_Type:
A26_If_WWW_URL:
A27_If_WWW_From_Adopted_Book:
A28_If_WWW_Password_Access: No
A45_WWW_Users:
A29_If_WWW_Material_Posted_From:
A30_If_WWW_Material_Posted_To:
A42_If_Intranet_URL:
A32_If_Intranet_From_Adopted_Book:
A33_If_Intranet_Password_Access: No
A48_Intranet_Users:
A34_If_Intranet_Material_Posted_From:
A35_If_Intranet_Material_Posted_To:
A36_If_Software_Print_Run:
A37_Comments_For_Request:

PERMISSION GRANTED

BY: *[Signature]* 4/2/07
Global Rights Dept., John Wiley & Sons, Inc.

NOTE: No rights are granted to use content that
appears in the work with credit to another source

04/13/2007 09:32 FAX

JOHN WILEY & SONS INC

002/002



republication
<republication@wiley.com>

03/26/2007 02:49 PM

Please respond to
republication
<republication@wiley.com>

To <republication@wiley.com>

cc

bcc

Subject Reproduction/Electronic Request Form

A01_First_Name: Christopher
A02_Last_Name: Davis
A03_Company_Name: Baylor University
A04_Address: P.O. Box 97348
A05_City: Waco
A06_State: TX
A07_Zip: 76798
A08_Country: USA
A09_Contact_Phone_Number: 254-710-2009
A10_Fax: 254-710-7272
A11_Emails: Christopher_B_Davis@baylor.edu
A12_Reference:
A13_Book_Title: Encyclopedia of Textiles, Fibers, and Nonwoven Fabrics
A40_Book_or_Journal: Book
A14_Book_Author: Martin Grayson
A15_Book_ISBN: 0-471-81461-X
A16_Journal_Month:
A17_Journal_Year:
A18_Journal_Volume:
A19_Journal_Issue_Number:
A20_Copy_Pages: p. 401 (Figure 1), p. 413 (Figure 6)
A21_Maximum_Copies: 5
A22_Your_Publisher: Baylor University
A23_Your_Title: Application of Chemometric Analysis to UV-Visible and Diffuse
Near-Infrared Reflectance Spectra
A24_Publication_Date: 4/15/2007
A25_Format: print
A31_Print_Run_Size:
A41_Ebook_Reader_Type:
A26_If_WWW_URL:
A27_If_WWW_From_Adopted_Book:
A28_If_WWW_Password_Access: No
A45_WWW_Users:
A29_If_WWW_Material_Posted_From:
A30_If_WWW_Material_Posted_To:
A42_If_Intranet_URL:
A32_If_Intranet_From_Adopted_Book:
A33_If_Intranet_Password_Access: No
A48_Intranet_Users:
A34_If_Intranet_Material_Posted_From:
A35_If_Intranet_Material_Posted_To:
A36_If_Software_Print_Run:
A37_Comments_For_Request:

PERMISSION GRANTED
BY: *[Signature]* 4/2/07
Global Rights Dept., John Wiley & Sons, Inc.

NOTE: No rights are granted to use content that
appears in the work with credit to another source

REFERENCES

- (1) ^aMassart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., Eds.; In *Handbook of Chemometrics and Qualimetrics, Part A*. Vandeginste, B. G. M., Rutan, S. C., Eds.; Data Handling in Science and Technology; Elsevier Science B.V.: Amsterdam, The Netherlands, 1997; Vol. 20A.
- ^bMassart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., Eds.; In *Handbook of Chemometrics and Qualimetrics, Part B*. Vandeginste, B. G. M., Rutan, S. C., Eds.; Data Handling in Science and Technology; Elsevier Science B.V.: Amsterdam, The Netherlands, 1998; Vol. 20B.
- (2) Wold, S. *Kemisk Tidskrift* **1972**, *84*, 34-37.
- (3) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. *Anal. Chem.* **1969**, *41*, 690-695.
- (4) Jurs, P. C.; Kowalski, B. R.; Isenhour, T. L.; Reilley, C. N. *Anal. Chem.* **1969**, *41*, 1949-1953.
- (5) Kowalski, B. R.; Jurs, P. C.; Isenhour, T. L.; Reilley, C. N. *Anal. Chem.* **1969**, *41*, 695-700.
- (6) Wold, S. *Chemometrics Intellig. Lab. Syst.* **1995**, *30*, 109-115.
- (7) Esbensen, K. H. *Multivariate Data Analysis: An Introduction to Multivariate Data Analysis and Experimental Design, 5th Edition*; Camo Process AS: Woodbridge, NJ, 2004.
- (8) Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley & Sons: New York, 1989.
- (9) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Brereton, R. G., Ed.; John Wiley & Sons Ltd.: UK, 2003.
- (10) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*, John Wiley & Sons: New York, 1998.
- (11) Fakayode, S. O., Doctoral Dissertation, Baylor University, Waco, TX, 2004.
- (12) Wold, S.; Sjostrom, M. *ACS Symp. Ser.* **1977**, *52*, 243-282.

- (13) Haswell, S. J., Ed.; In *Practical Guide to Chemometrics*; Haswell, S. J., Ed.; Marcel Dekker, Inc.: New York, NY, 1992.
- (14) Krajewska, A. M., Doctoral Dissertation, University of Georgia, Athens, GA, 1984.
- (15) Busch, K. W.; Swamidoss, I. M.; Fakayode, S. O.; Busch, M. A. *J. Am. Chem. Soc.* **2003**, *125*, 1690-1691.
- (16) Fakayode, S. O.; Busch, M. A.; Bellert, D. J.; Busch, K. W. *Analyst* **2005**, *130*, 233-241.
- (17) Fakayode, S. O.; Swamidoss, I. M.; Busch, M. A.; Busch, K. W. *Talanta* **2005**, *65*, 838-845.
- (18) Soyemi, O. O., Doctoral Dissertation, Baylor University, Waco, TX, 2000.
- (19) Humphrey, K. J., Doctoral Dissertation, Baylor University, Waco, TX, 2002.
- (20) Markey, C.; Busch, M.; Busch, K. In *In Determination of Hotness Levels in Capsicum Fruits by Chemometric Modeling of UV Spectral Data*; 30th Federation of Analytical Chemistry and Spectroscopy Societies; Ft. Lauderdale, FL, 2003, Paper No. 570.
- (21) Govindarajan, V. S.; Sathyanarayana, M. N. *Crit. Rev. Food Sci. Nutr.* **1991**, *29*, 435-475.
- (22) Govindarajan, V. S. *Crit. Rev. Food Sci. Nutr.* **1986**, *24*, 245-355.
- (23) Reilly, C. A.; Crouch, D. J.; Yost, G. S. *J. Forensic Sci.* **2001**, *46*, 502-509.
- (24) Kirschbaum-Titze, P.; Mueller-Seitz, E.; Petz, M. *J. Agric. Food Chem.* **2002**, *50*, 1264-1266.
- (25) Govindarajan, V. S.; Rajalakshmi, D.; Chand, N. *Crit. Rev. Food Sci. Nutr.* **1987**, *25*, 185-282.
- (26) Perry, L.; Dickau, R.; Zarrillo, S.; Holst, I.; Pearsall, D. M.; Piperno, D. R.; Berman, M. J.; Cooke, R. G.; Rademaker, K.; Ranere, A. J.; Raymond, J. S.; Sandweiss, D. H.; Scaramelli, F.; Tarble, K.; Zeidler, J. A. *Science* **2007**, *315*, 986-988.
- (27) Knapp, S. *Science* **2007**, *315*, 946-947.
- (28) Cordell, G. A.; Araujo, O. E. *Ann. Pharmacother.* **1993**, *27*, 330-336.
- (29) Kopec, S. E.; DeBellis, R. J.; Irwin, R. S. *Pulm. Pharmacol. Ther.* **2002**, *15*, 529-534.

- (30) Reilly, C. A.; Ehlhardt, W. J.; Jackson, D. A.; Kulanthaivel, P.; Mutlib, A. E.; Espina, R. J.; Moody, D. E.; Crouch, D. J.; Yost, G. S. *Chem. Res. Toxicol.* **2003**, *16*, 336-349.
- (31) Govindarajan, V. S. *Crit. Rev. Food Sci. Nutr.* **1986**, *23*, 207-288.
- (32) Perucka, I.; Oleszek, W. *Food Chem.* **2000**, *71*, 287-291.
- (33) Henry, C. J.; Emery, B. *Human nutrition. Clinical nutrition* **1986**, *40*, 165-168.
- (34) Yoshioka, M.; Lim, K.; Kikuzato, S.; Kiyonaga, A.; Tanaka, H.; Shindo, M.; Suzuki, M. *J. Nutr. Sci. Vitaminol.* **1995**, *41*, 647-656.
- (35) Tanaka, N. Patent Application Country: Application: JP; Patent Country: JP Patent 2001213772, 2001.
- (36) Molina-Torres, J.; Garcia-Chavez, A.; Ramirez-Chavez, E. *J. Ethnopharmacol.* **1999**, *64*, 241-248.
- (37) Chowdhury, B.; Mukhopadhyay, S.; Bhattacharayay, D.; De, A. K. *Med. Sci. Res.* **1996**, *24*, 669-670.
- (38) Mori, A.; Lehmann, S.; O'Kelly, J.; Kumagai, T.; Desmond, J. C.; Pervan, M.; McBride, W. H.; Kizaki, M.; Koeffler, H. P. *Cancer Res.* **2006**, *66*, 3222-3229.
- (39) Gibbs, H. A. A.; O'Garro, L. W. *HortScience* **2004**, *39*, 132-135.
- (40) Athanasiou, A.; Smith, P. A.; Vakilpour, S.; Kumaran, N. M.; Turner, A. E.; Bagiokou, D.; Layfield, R.; Ray, D. E.; Westwell, A. D.; Alexander, S. P. H.; Kendall, D. A.; Lobo, D. N.; Watson, S. A.; Lophatanon, A.; Muir, K. A.; Guo, D.; Bates, T. E. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 50-55.
- (41) Jang, J. J.; Kim, S. H.; Yun, T. K. *In Vivo* **1989**, *3*, 49-53.
- (42) Law, M. W. *J. Assoc. Off. Anal. Chem.* **1983**, *66*, 1304-1306.
- (43) Scoville, W. L. *Journal of the American Pharmaceutical Association (1912-1977)* **1912**, *1*, 453-454.
- (44) Wall, M. M.; Bosland, P. W. *Developments in Food Science* **1998**, *39*, 347-373.
- (45) North, H. *Anal. Chem.* **1949**, *21*, 934-936.
- (46) DiCecco, J. J. *J. Assoc. Off. Anal. Chem.* **1976**, *59*, 1-4.
- (47) Hollo, J.; Kurucz, E.; Bodor, J. *Lebensmittel-Wissenschaft und -Technologie* **1969**, *2*, 19-21.
- (48) Hartman, K. T. *J. Food Sci.* **1970**, *35*, 543-547.

- (49) Todd, P. H., Jr.; Bensinger, M. G.; Biftu, T. *J. Food Sci.* **1977**, *42*, 660-5, 680.
- (50) Betts, T. A. *J. Chem. Educ.* **1999**, *76*, 240-244.
- (51) Chalchat, J. C.; Garry, R. P.; Michet, A. *Parfums, Cosmetiques, Aromes* **1988**, *84*, 69-70, 75.
- (52) Chiang, G. H. *J. Food Sci.* **1986**, *51*, 499-503.
- (53) Cooper, T. H.; Guzinski, J. A.; Fisher, C. *J. Agric. Food Chem.* **1991**, *39*, 2253-2256.
- (54) Kaale, E.; Van Schepdael, A.; Roets, E.; Hoogmartens, J. *J. Pharm. Biomed. Anal.* **2002**, *30*, 1331-1337.
- (55) Maillard, M.; Giampaoli, P.; Richard, H. M. *J. Flavour Fragrance J.* **1997**, *12*, 409-413.
- (56) Weaver, K. M.; Luker, R. G.; Neale, M. E. *J. Chromatogr.* **1984**, *301*, 288-291.
- (57) Weaver, K. M.; Awde, D. B. *J. Chromatogr.* **1986**, *367*, 438-442.
- (58) Wood, A. B. *Flavour Fragrance J.* **1987**, *2*, 1-12.
- (59) Woodbury, J. E. *J. Assoc. Off. Anal. Chem.* **1980**, *63*, 556-558.
- (60) Yao, J.; Nair, M. G.; Chandra, A. *J. Agric. Food Chem.* **1994**, *42*, 1303-1305.
- (61) Iwai, K.; Suzuki, T.; Fujiwake, H.; Oka, S. *J. Chromatogr.* **1979**, *172*, 303-311.
- (62) Nyberg, N. T.; Baumann, H.; Kenne, L. *Magn. Reson. Chem.* **2001**, *39*, 236-240.
- (63) Tice, L. F. *Am. J. Pharm.* **1933**, *105*, 320-325.
- (64) DiCecco, J. J. *J. Assoc. Off. Anal. Chem.* **1979**, *62*, 998-1000.
- (65) Luise, M. *Industrie Alimentari (Pinerolo, Italy)* **1968**, *7*, 80-84.
- (66) Nikolaeva, D. A. *Biokhim. Metody Analiza Plodov, Kishinev* **1984**, 99-102.
- (67) Polesello, A.; Pizzocaro, F. *Rivista di Scienza e Tecnologia degli Alimenti e di Nutrizione Umana* **1976**, *6*, 305-306.
- (68) Ramos Palacio, J. J. *J. Assoc. Off. Anal. Chem.* **1977**, *60*, 970-972.
- (69) Ramos Palacio, J. J. *J. Assoc. Off. Anal. Chem.* **1979**, *62*, 1168-1170.
- (70) Suzuki, J. I.; Tausig, F.; Morse, R. E. *Food Technol.* **1957**, *11*, 100-104.

- (71) Pryakhin, O. R.; Tkach, V. I.; Golovkin, V. A.; Gladyshev, V. V.; Kuleshova, N. D. Patent Application Country: Application: SU; Patent Country: SU Patent 1783404, 1992.
- (72) Laskaridou-Monnerville, A. *J. Chromatogr. A* **1999**, 838, 293-302.
- (73) Korel, F.; Bagdatlioglu, N.; Balaban, M. O.; Hisil, Y. *J. Agric. Food Chem.* **2002**, 50, 3257-3261.
- (74) Way, R. M. In *Official Analytical Methods of the American Spice Trade Association*; Way, R. M., Ed.; Official Analytical Methods of the American Spice Trade Association; American Spice Trade Association: Washington, D.C., 1985.
- (75) Skoog, D. A.; Leary, J. J. In *Principles of Instrumental Analysis. 4th Ed.*; Harcourt Brace Jovanovich: New York, 1992.
- (76) Kortum, G.; Seiler, M. *Angew. Chem.* **1939**, 52, 687-693.
- (77) Sharpe, M. R. *Anal. Chem.* **1984**, 56, 339A-340A, 342A, 344A, 348A, 350A, 356A.
- (78) Elving, P. J.; Meehan, E. J.; Kolthoff, I. M.; Editors **1981**, 816.
- (79) Pavia, D.; Lampman, G.; Kriz, G.; *Introduction to Spectroscopy*; Harcourt Brace Jovanovich: New Cork, 1996.
- (80) Lorber, A.; Wangen, L. E.; Kowalski, B. R. *J. Chemometrics* **1987**, 1, 19-31.
- (81) Reis, M. S.; Saraiva, P. M. *J. Chemometrics* **2004**, 18, 526-536.
- (82) Sales, F.; Callao, M. P.; Rius, F. X. *Chemometrics Intellig. Lab. Syst.* **1997**, 38, 63-73.
- (83) Grayson, M., Ed.; In *Encyclopedia of Textiles, Fibers, and Nonwoven Fabrics*; Grayson, M., Ed.; Encyclopdeia Reprint Series; John Wiley & Sons, Inc.: New York, NY, 1984; Vol. 1.
- (84) Humphries, M. In *Fabric Reference*; Humphries, M., Ed.; Prentice Hall: Upper Saddle River, NJ, 1996; Vol. 1.
- (85) Needles, H. L. In *Textile Fibers, Dyes, Finishes, and Processes: A Concise Guide*; Needles, H. L., Ed.; Noyes Publications: New Jersey, 1986.
- (86) Kadolph, S. J.; Langford, A. L. In *Textile Fibers and Their Properties*; Helba, S., Ed.; Textiles; Prentice Hall: New Jersey, 2002.
- (87) Rendle, D. F. *Chem. Soc. Rev.* **2005**, 34, 1021-1030.

- (88) Mark, H.; Whitby, G. In *Collected Papers of W. H. Carothers*; Interscience Publishers, Inc.: New York, 1940.
- (89) AnonymousHerodotus. <http://en.wikipedia.org/wiki/Herodotus> (accessed 2/07, 2007).
- (90) Wikipedia Herodotus. <http://en.wikipedia.org/wiki/Herodotus> (accessed 2/13, 2007).
- (91) Ghosh, S.; Rodgers, J. In *NIR analysis of textiles*. Burns, D. A., Ciurczak, E. W., Eds.; Handbook of Near-Infrared Analysis (2nd Edition); Marcel Dekker, Inc.: New York, NY, 2001.
- (92) King, R. R.; *Textile Identification, Conservation, and Preservation*; Noyes Publications: Parkridge, NJ, 1985.
- (93) AnonymousAsbestos. <http://en.wikipedia.org/wiki/Asbestos> (accessed 2/07, 2007).
- (94) Bohn, C. R.; Schaefgen, J. R.; Statton, W. O. *J. Polym. Sci.* **1961**, 55, 531-549.
- (95) Krigbaum, W. R.; Tokita, N. *J. Polym. Sci.* **1960**, 43, 467-488.
- (96) Sohn, M.; Himmelsbach, D. S.; Akin, D. E.; Barton, F. E., II *Text. Res. J.* **2005**, 75, 583-590.
- (97) Del Rio, J. C.; Gutierrez, A.; Martinez, A. T. *Rapid Commun. Mass Spectrom.* **2004**, 18, 1181-1185.
- (98) Morrison, W. H. I.; Archibald, D. D. *J. Agric. Food Chem.* **1998**, 46, 1870-1876.
- (99) Perkins, H. H., Jr. *Textile Bulletin* **1971**, 97, 21, 25, 34.
- (100) Olson, E. S.; *Textile Wet Processes*, Vol. 1: Preparation of Fibers and Fabrics; Noyes Publications: Parkridge, NJ, 1983.
- (101) Edelstein, S. M. *Am.Dyestuff Repr.* **1936**, 25, P458-66.
- (102) Burns, D. A.; Ciurczak, E. W., Eds.; In *Handbook of Near-Infrared Analysis*; Burns, D. A., Ciurczak, E. W., Eds.; Practical Spectroscopy Series; Marcel Dekker, Inc.: New York, 2001.
- (103) Busch, K. W.; Soyemi, O.; Rabbe, D.; Humphrey, K.; Dundee, B.; Busch, M. A. *Appl. Spectrosc.* **2000**, 54, 1321-1326.
- (104) Jasper, W. J.; Kovacs, E. T. *Text. Res. J.* **1994**, 64, 444-448.
- (105) Cleve, E.; Bach, E.; Schollmeyer, E. *Anal. Chim. Acta* **2000**, 420, 163-167.

- (106) Church, J. S.; O'Neill, J. A.; Woodhead, A. L. *Text. Res. J.* **1999**, *69*, 676-684.
- (107) Schindler, W. D. In *Chemical Finishing of Textiles*. Section Title: Textiles and Fibers; Woodhead Publishing Limited: Cambridge, England, 2004.