12-14-2015

Optical Spectroscopy and Chemometrics for Discrimination of Dyed Textile Fibers and Magnetic Audio Tapes

Nathan C. Fuenffinger
University of South Carolina - Columbia

Optical Spectroscopy and Chemometrics for Discriminations of Dyed Textile Fibers and Magnetic Audio Tapes

by

Nathan C. Fuenffinger

Bachelor of Science
Gannon University, 2009

Master of Science
University of Kentucky, 2012

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry

College of Arts and Sciences

University of South Carolina

2015

Accepted by:

Stephen L. Morgan, Major Professor

S. Michael Angel, Committee Chair

Michael L. Myrick, Committee Member

Brian T. Habing, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

DEDICATION

In loving memory of my late grandparents, Patricia and Robert Fuenffinger.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mother, Terri, my father, Curtis Sr., and my brother, Curtis Jr., for their endless support. Next, I would like to offer my gratitude to my advisor, Dr. Morgan, for giving me the resources and the freedom to work on projects that I enjoy doing and allow me to develop my skillset. Additionally, I would like to thank my dissertation committee, Dr. Angel, Dr. Myrick, and Dr. Habing for their oversight and advice throughout this process. Finally, I am grateful for my high school chemistry teacher, Miss Carla Andrews, for laying the groundwork in the field which would ultimately become my career.

ABSTRACT

This dissertation focuses on the application of both novel and standard chemometric approaches toward societal problems of interest in the areas of forensic science and cultural heritage preservation.

Microspectrophotometry (MSP), a technique enabling measurements of absorption of electromagnetic radiation by microscopic materials in the ultraviolet-visible (UV-Vis) region, is widely used by forensic examiners for comparisons of metameric textile fibers. These comparisons are often hindered, however, by the raw or normalized spectra showing little detail or having few points of comparison. Derivative preprocessing can enhance structure in some instances. We have demonstrated through the use of multivariate statistics that derivatives are an effective tool for discriminating dyed textile fibers. The Fiber Spectra Comparison Tool developed in this work is an easy-to-use program designed for comparing multiple fibers simultaneously.

Microspectrofluorimetry (MSF) is another useful technique, often used as a follow-up method to MSP, for studying fibers that absorb and emit in the UV-Vis region. Results found after applying MSP and MSF to the same set of fibers suggest that the discrimination power of MSP measurements are slightly higher than those obtained from MSF for most colors and fiber-types. In some instances, MSP and MSF provide complimentary information which can be taken advantage of by fusing the measurements. A low-level (i.e., data level) fusion strategy has been developed which provides increased discrimination over the individual techniques.

The ability to transfer multivariate classification models between laboratories having differing instruments has also been investigated in this work. Such efforts could save time and resources in forensic analyses and help identify variability between examiners. A set of 12 blue acrylic fibers was analyzed by MSP at three academic institutions and two certified forensic laboratories. Using a six-step preprocessing procedure combined with quadratic discriminant analysis, a transferrable classification model was developed which, when tested, produced a classification accuracy of 93.2%. This percentage was only slightly lower than the 96.3% accuracy resulting from intra-laboratory models. This outcome speaks to the consistency of results obtained on the same samples in different laboratories.

Multivariate classification strategies similar to those applied to colored fibers have also proven useful for determining the playability of magnetic audio tapes, a popular recording medium from the 1950s to the 1990s. Attempting to play degraded tapes during the digitization process can cause damage to the playback instrument, and to the tapes themselves, often leading to significant downtime for museums and archives. A reliable and non-destructive technique for determining the playability of a tape without ever actually playing the tape would be beneficial. This work has shown that with attenuated total reflectance Fourier transform infrared spectroscopy and machine learning algorithms, playability of quarter-inch magnetic audio tapes can be determined with greater than 90% accuracy. This finding led to the creation of the Magnetic Audio Tape Spectra Analysis program, a user-friendly software program allowing tape custodians to visualize data and determine which tapes need to be subjected to restoration processes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

"The three great essentials to achieve anything worthwhile are, first, hard work; second, stick-to-itiveness; third, common sense."

— Thomas Edison

CHAPTER 1

COMPARISON OF MULTIVARIATE PREPROCESSING TECHNIQUES FOR THE FORENSIC DISCRIMINATION OF COTTON FIBERS BY UV-VISIBLE MICROSPECTROPHOTOMETRY

ABSTRACT

Color plays a critical role when analyzing natural fibers such as cotton in forensic investigations. Ultraviolet-visible microspectrophotometry is a non-destructive technique capable of providing objective color measurements on fibers in the form of absorption or transmission spectra. Forensic fiber examinations are often hindered, however, by spectra with little detail or points of comparison. In this work, samples of reactive, direct, and vat dyed cotton fibers were analyzed and spectra were preprocessed using multiple methods including baseline correction, normalization, and derivatives. Principal component analysis followed by linear discriminant analysis was employed to discriminate the cotton samples.

Direct dyed fibers exhibited almost featureless and low absorbing spectra compared to those of reactive and vat dyed fibers. As a result, classification accuracies for direct dyed fibers were lower than those calculated for reactive and vat dyed fibers. The results of this study show that derivative spectra can significantly enhance structure in spectra and are especially useful where spectra only have subtle features such as those

exhibited by direct dyed cotton fibers. No single method was best for all classes of fibers in the study, and the shapes and intensities of the curves are important when determining if derivative calculations are auspicious.

1. INTRODUCTION

Cotton is the most abundant fiber in the world with an estimated 25 million tons produced annually.[1] Much of that amount is used in clothing manufacturing.[2] The likelihood of recovering cotton fibers from a crime scene is the highest of any fiber type, as population studies have shown cotton to be the most common textile fiber found on indoor[3,4] and outdoor surfaces[5], as well as human head hair.[6] Certain types of cotton fibers, such as indigo dyed and colorless cotton, are found in such abundance that they are often considered of little significance for use by forensic analysts.

Cotton fibers can be categorized based on the method in which they were dyed. Fibers dyed using reactive dyes make up the majority of all cotton fibers, and their dominance over direct and vat dyed fibers is expected to continue due to the excellent wetfastness properties of reactive dyed fibers and the range of brilliant colors which can be made using these dyes.[7] Although the strength of the covalent bonding between reactive dyes and the fiber allows for some superior properties, these forces also make removing the dye from the fiber very challenging in cases where one wishes to use chromatographic methods of analysis. Direct dyed fibers seemingly account for only about 10% of all colored cotton fibers[8], and vat dyed fibers share a similar percentage. Due to the decreasing popularity of direct and vat dyes, the evidentiary value of fibers colored with reactive dyes should significantly increase.

Color-based techniques such as thin layer chromatography, Raman spectroscopy, and ultraviolet-visible (UV-Vis) microspectrophotometry (MSP) are popular methods used to analyze cotton fibers due to the lack of other defining characteristics in most natural fibers.[9-11] UV-Vis MSP provides a simple, non-destructive method for analyzing fibers *in situ*. The technique is often beneficial for excluding fibers which are indistinguishable by other approaches such as comparison light microscopy and fluorescent light microscopy.[12]

The traditional method used to compare two fibers by UV-Vis MSP might have a forensic examiner overlaying representative absorption or transmission spectra and comparing them based on the locations and shapes of the peaks. Spectral preprocessing techniques can be performed on UV-Vis data to potentially present the data in a more useful way. For example, differentiation has been used for many years in analytical chemistry for various applications described elsewhere.[13,14] The use of derivatives has only been used sparingly in forensics, however, for analyzing textile fibers[15-18] and fiber dyes.[19,20]

The aim of this work is to investigate the extent to which the ability to discriminate cotton fibers is influenced by various multivariate preprocessing techniques. Feature extraction by principal component analysis (PCA) will be used for the purpose of reducing complex datasets down to the most significant variables which may not be readily visible by examining the spectra. Classification accuracies will be obtained using linear discriminant analysis (LDA), a technique which has been used in conjunction with UV-Vis MSP in previous studies to discriminate colored textile fibers of cotton, acrylic, nylon, and polyester.[21,22]

2. EXPERIMENTAL

2.1 Materials

Cotton fibers were collected from fabric obtained from a textile-related manufacturer in the southeast United States. Dyeing of the fabric was performed at the North Carolina State School of Textiles pilot facility (Raleigh, NC). A total of 121 cotton fiber samples known to have been dyed using direct, reactive, or vat dyes were analyzed. The fibers were then placed into the groups in Table 1.1 based on their observed color for subsequent analysis by UV-Vis MSP.

Procedures from the Scientific Working Group on Materials Analysis (SWGMAT) were followed for fiber analysis by UV-Vis MSP.[23] Single fibers were positioned on quartz slides (CRAIC Technologies, Altadena, CA, and Esco Products Inc., Oak Ridge, NJ) using micro tweezers. Each fiber was mounted using spectral grade glycerin (Spectrum Chemical Mfg. Corp., Gardena, CA) and quartz cover slips.

2.2 Instrumentation

UV-Vis spectra were obtained using a Quantum Detection Instrument (QDI) 1000 microspectrophotometer (CRAIC Technologies, San Dimas, CA). Data was processed using GRAMS/AI version 700 software (Thermo Galactic, Salem, NH). The microspectrophotometer was operated in transmission mode using a xenon light source. A 15× collecting objective was used to focus the source light onto an area within the diameter of the fiber samples, and replicate spectra were taken along the length of the same fiber. Spectra were obtained by taking an average of 100 scans across a spectral region of 200-850 nm with a bandwidth of 10 nm. Integration time for the charge coupled device (CCD) was approximately 4 ms.

2.3 Data Analysis

Data was saved as comma separated variable (CSV) files and analyzed using Fiber Spectrum Explorer, an in-house program written in MATLAB version 8.1 (The Mathworks, Inc., Natick, MA). By convention, each dataset explored consisted of a matrix with *n* (number of samples) rows and *p* (number of variables) columns. For discrimination by multivariate analysis, wavelength ranges for all spectra were truncated to a wavelength range of 380 to 700 nm. The spectra were then preprocessed using the methods described below.

2.3.1 Baseline Correction

It is common to have offsetting baselines from spectrum to spectrum in UV-Vis measurements. Numerous methods are available for correcting offsetting baselines in spectroscopy. Although there are more elaborate techniques for estimating the baseline of a spectrum, the method used here involves a simple rescaling of each spectrum by assuming the lowest non-zero intensity across a spectrum is in a region where there is zero signal. That intensity is then subtracted from all other points in that spectrum. For the purpose of discrimination, this method seemingly works at least as well as some polynomial fitting algorithms.

2.3.2 Normalization

Normalization to unit area is achieved by dividing each observation, X, in the ith row and jth column, by the sum of the absolute value of all elements in that row, also called 1-norm, as in Equation 1.[24]

$$X_{ij,norm} = \frac{X_{ij}}{\sum_{j=1}^{n} |X_{ij}|} \quad (1)$$

The result is the total area under the curve of each resulting vector being set equal to one. Normalization to unit area is used to account for scaling differences arising from variations in concentration, amount, and sample size as well as instrumental intensity variations caused by changes in fiber thickness.

2.3.3 Standard Normal Variate

The standard normal variate (SNV) transformation is a method of preprocessing similar to that of the normalization technique described previously and is calculated using Equation 2, where $X_{ij,SNV}$ is the SNV transformed point.[25]

$$X_{ij,SNV} = \frac{X_{ij} - \overline{X_i}}{s_i} \quad (2)$$

The sample mean spectrum, $\overline{X_i}$, used in SNV calculations is not used in normalization to unit area, however, and can instead be thought of as being set to zero. In addition, normalization uses a scaling factor (1-norm for the calculation used in this study) in place of the standard deviation, s_i, of the sample-spectrum.

2.3.4 Autoscale

Autoscaling is a method of preprocessing which involves subtracting the column mean from each element of each column and dividing that result by the standard deviation of the column, s_j.

$$X_{ij,auto} = \frac{X_{ij} - \overline{X_j}}{s_j} \quad (3)$$

This results in the variance of all columns being equal to one. Because all values are given equal weighting, small variations in the data are emphasized. When using data with a low signal to noise ratio, however, noise and signal are treated equally, and this approach becomes less useful.

2.3.5 Derivatives

First derivatives (FDs) are an effective tool for correcting baseline offsets. By using a normalization technique such as a normalization to unit area or SNV following a FD calculation, a slope correction can also be gained. FD spectra of data recorded using evenly spaced intervals along the x-axis can be obtained by calculating the difference between two features, *n* and *n* + 1, where *y* is signal intensity and λ is wavelength.

$$\frac{\partial y}{\partial \lambda} = \frac{y_{n+1} - y_n}{\lambda_{n+1} - \lambda_n} \tag{4}$$

The sharpest features in UV-Vis absorbance spectra are caused by noise in the measured signal. This results in a decrease in the signal-to-noise ratio when the FD is calculated. Noise enhancement by derivative spectroscopy is often dealt with by spectral smoothing before differentiation. For all FD spectra in this study, a line was fitted to a 23 point moving window using a least-squares approximation.

A second approach to dealing with the increasing noise amplification that is associated with calculating derivatives is to use the gap-segment method. Unlike the FD calculations collected by taking the difference of values over two adjacent points, the second derivative (SD) calculations were made by calculating the derivative over a number of variables (i.e., segments). The user can then define the number of variables between those segments (i.e., gaps). A practical method for determining a “good” gap-

segment combination is to try multiple combinations on one or more datasets, and select the one which gives the best results. From this process, a gap size of 31 points and a segment size of 35 points were used for all SD calculations in this study.

2.3.6 Feature extraction and classification

After preprocessing, all sets of spectra were subjected to PCA. PCA is a technique used to reduce the dimensionality of large data sets.[26] In PCA, the original correlated variables (wavelengths) are reduced to a new set of uncorrelated variables, or principal components (PCs). PCs are linear combinations of the original variables and are arranged in such a way that the first PC accounts for the largest variation in the data set, and the amount of variance captured decreases with each successive PC. Selection of the number of relevant PCs to be used in the models was chosen via scree plots which display the percent variance captured by each PC. The greatest number of PCs before the captured variance begins to level off was selected for use in LDA.

After the appropriate number of PCs are selected, LDA, a supervised classification technique, was used to maximize the separation between groups in the reduced PC space. This is carried out by projecting the data into the space of the canonical variates. These axes differ from those in PCA in that they account for the within-group and between-group variances after the groups are specified by the user. Classification accuracies of each PCA-LDA model were determined internally using leave-one-out cross-validation. In this cross-validation technique, LDA is performed on the data set with one sample omitted thus becoming the training set. An attempt is then made to allocate the omitted sample back into the training set. This process is repeated for each sample in the data set, and classification accuracies are obtained by assigning

each ‘unknown’ spectrum to the group of which the Mahalanobis distance[27] is the shortest.

3. RESULTS AND DISCUSSION

3.1 Comparing preprocessing methods for fiber analysis

The impact of using multiple forms of preprocessed UV-Vis spectra for fiber discrimination with PCA-LDA was investigated. Though there is no hard-and-fast rule for classifying cotton fibers based on the class of the dye, the classes themselves may have characteristic changes in absorbance (ΔA) across each spectrum as calculated using Equation 5. A_{max} and A_{min} are the respective maximum and minimum absorbance values in a spectrum as seen in Figure 1.1. For Gaussian-shaped curves, such as the ones encountered in UV-Vis MSP, this calculation can serve as an indicator of the evenness of the features across the spectra.

$$\Delta A = A_{max} - A_{min} \tag{5}$$

Most direct dyed fibers in this study (80.8%) had values for ΔA between 0.005 and 0.020. The ΔA values for vat dyed fibers were mainly (85.7%) within a range of 0.020 and 0.130. Cotton fibers dyed using reactive dyes showed a broad range of ΔA values. However, 89.2% of these fibers had a ΔA between 0.020 and 0.456. This suggests that the majority of reactive dyed fibers show equal or greater absorption than vat dyed fibers. UV-Vis spectra of fibers consistent with this trend are shown in Figure 1.2.

Multivariate statistical methods were used to compare each fiber sample to other fibers of both the same color and dye class. A group of six yellow fibers dyed using reactive dyes was selected to demonstrate the methodology used in this study. The UV-

Vis absorbance spectra obtained for this group of fibers are shown in Figure 1.3. The hues for these fibers were obtained using one to three dyes, the typical range for all fibers in this study. From a visual examination of the absorbance spectra of this set of six fibers, it was concluded that three of the fibers could easily be discriminated. As seen in Figure 1.4, the broad peaks in the averaged raw absorbance spectra of the three remaining fibers demonstrate the challenge associated with fiber comparisons based on unprocessed spectra. It should be noted that none of the samples in this group which appeared to have very similar absorbance spectra shared any of the same dyes.

After all UV-Vis absorbance spectra had been preprocessed using the techniques described in the previous section, PCA was performed to reduce the dimensionality of the data. A requirement for LDA calculations is to have more samples than variables. If this is not the case, inversion of the within-groups sum of squares and cross-products matrix cannot occur. Selection of an appropriate number of PCs to include in the LDA model is important for achieving the best separation between classes. When too many PCs are included, the resulting classification model may generalize poorly. Overfitting is avoided or reduced by using a Scree plot such as the one in Figure 1.5. This plot shows the percent variance captured by each PC after performing PCA on FD spectra of the six yellow reactive dyed fibers. Four PCs, containing 91.7% of the total variance in the data, were chosen as the number used as input for LDA, since the variance appears to be relatively flat for all PCs greater than four.

Table 1.2 shows the confusion matrix resulting from leave-one-out cross-validation on the PC-LDA model of the six yellow reactive dyed fibers after FD preprocessing. A classification accuracy of 96.7% (58 correctly classified spectra out of

60) was obtained for the group of six yellow reactive dyed fibers after FD preprocessing. As was already shown, samples RY03 and RY05 have very similar absorbance spectra. The PCA scores plot in Figure 1.6 shows that even after FD preprocessing, there is still significant overlap of the 95% confidence ellipses calculated for the two groups of fibers. No preprocessing method used in this study allowed for complete discrimination of samples RY03 and RY05.

The positive effect that each preprocessing method has on UV-Vis absorbance spectra is perhaps best shown using samples RY03 and RY04 (Figure 1.7). These fibers gave very similar raw spectra, but were easily distinguishable after several types of preprocessing. Feature by feature t-tests were used to measure the difference between the group means at each observation point. The largest t-statistic value and the wavelength at which it occurs are indicated following each preprocessing technique. As seen by the t-statistic values from the raw and autoscaled spectra, autoscaling provided no improvement in separating the two groups. In general, the changes in classification accuracies for all groups before and after autoscaling were insignificant. Separation of the two groups of fibers was achieved to the greatest extent by calculating SDs. This is not surprising, as SDs correct for both baseline offsets and changes in slope. The PCA scores plots in Figure 1.8 are consistent with the t-statistic values for each preprocessing technique. The greatest separation between the two groups of fibers was gained using FD (Figures 1.8g, 1.8h, and 1.8i), SD (Figure 1.8j), and SNV preprocessing techniques (Figure 1.8d).

Classification accuracies of PCA-LDA, obtained after numerous methods of preprocessing, for the three dye classes of cotton fibers studied are shown in Table 1.3.

Methods involving FD and SD calculations show increased classification accuracies of direct dyed cotton fibers by as much as 11%. This gap in discrimination ability is lessened for vat dyed fibers and is non-existent in the analysis of reactive dyed fibers. Normalized spectra are slightly more discriminating, in the case of reactive dyed fibers, than FD spectra. Because, in general, these fibers had larger changes in absorbance relative to those of direct dyed fibers, this suggests there may be some cutoff value of ΔA at which calculating derivatives provides no further benefit over using other methods of preprocessing. The baseline correction method used in this study was found to have no real advantage over the other preprocessing techniques used, and therefore is not recommended. Classification accuracies for the entire dataset are shown in Table 1.4. All FD and SD preprocessing methods used in this study, in addition to normalization, can be considered effective methods for discriminating cotton fibers.

3.2 Fiber Spectra Comparison Tool

Using MATLAB, a Fiber Spectra Comparison Tool (FISCOTO) was developed which would allow an examiner to view absorbance spectra of textile fibers and calculate the previously described two sample single feature t-statistics. The design of the interface makes the program convenient for rapidly skimming through spectra of numerous questioned and known fibers and selecting out those fibers which require a more thorough comparison. The user-friendly interface for the FISCOTO application is displayed in Figure 1.9.

Spectra from multiple fibers may be loaded into FISCOTO as .CSV files. FISCOTO uses the .CSV filenames and number of classes specified by the user to associate replicate spectra with a particular fiber sample. These samples are listed in the

Select fibers to compare list box on the left-hand side of the main user interface. Although the user can opt to select multiple fibers to compare, a t-test for the equality of the means is only performed when two samples are selected. The maximum t-statistic for all features is made available for the user on the main interface. The t-statistics for all features can be found by accessing the *View t-test* option located on the top right portion of the main window.

To show an example of the single feature t-tests in FISCOTO, the same two vat-dyed blue cotton fibers selected in Figure 1.9 will be used. After clicking the *View t-test* button, the two plots in Figure 1.10 are shown. Here, the calculated t-statistics for each feature of the raw data and data processed using Savitzky-Golay smoothing, the FD, and standard normal variate transformation are shown in black. The horizontal green threshold line denotes the critical value (t_{crit}) of Student's *t* at the 0.05 level of significance, and 9 degrees of freedom (df) (due to 10 samples in each group). When the black line is above the green line for any comparison of means at a single feature value, the null hypothesis of equal means can be rejected at the 95% level of confidence. The plots of calculated *t*-values across the feature domain also have red lines at t_{calc} = ~4.3, representing a conservative choice for a threshold value. The most significant difference between groups in the processed data is seen near 416 nm. This corresponds to the large separation shown at that wavelength shown on the *After processing* panel in Figure 1.9.

In addition to hypothesis testing, this software allows users to remove undesired features and visualize raw or processed data. FISCOTO currently makes available 12 separate methods of processing including those used in this work prior to multivariate classification. A summary of all processing performed on the data and the order in which

it was performed is listed in the box in the bottom right corner of the interface. This serves as a reminder to the user as to how the data has been manipulated. The box will fill in automatically when previously saved data is loaded back into the program. The *Reset* option located by the Summary box undoes all performed preprocessing steps and restores the original data.

4. CONCLUSION

Performing PCA-LDA on derivative spectra can improve discrimination of cotton fibers over other methods of spectral preprocessing. Significant increases in discrimination of fibers with mostly flat spectra with small changes in absorbance are possible using derivative spectra. Direct dyed cotton fibers are one class of fibers that would seemingly benefit significantly from utilizing derivative spectra, since these fibers had distinctively low ΔA values. It should be noted that the effect of smoothing the spectra using a Savitzky-Golay polynomial[28] prior to calculating the FD was examined. Although Savitzky-Golay polynomial smoothing may be advantageous for visual examinations, increases in classification accuracies were not gained by using a higher-order polynomial smooth rather than a linear smooth.

As was stated by Wiggins *et al.*[19], there is a risk of FD spectra misclassifying matching fibers with large variations in absorbance. This resulted in classification accuracies of FD spectra being slightly lower in the analysis of reactive dyed cotton fibers when compared to the normalized spectra. Still, the high classification accuracies (greater than 90%) achieved using all methods of preprocessing are significant due to the difficulty of extracting these dyes for analysis by other techniques such as thin-layer chromatography or liquid chromatography. Because no single method of preprocessing is

best for all types of spectra, a convenient software application with many preprocessing options available, called FISCOTO, was developed for rapid comparisons of fiber spectra. The program is freely available and can be attained by sending a request to the author.

ACKNOWLEDGMENTS

This project was supported by Award No. 2010-DN-BX-K245 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Portions of this work were presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Pittcon) on March 6, 2014 in Chicago, IL. The co-authorship of Jessica M. McCutcheon, John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan is also acknowledged. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect those of the Department of Justice.

REFERENCES

1. van Dam, J. In *Environmental benefits of natural fibre production and use,* Proceedings of the Symposium on Natural Fibres, Rome, Italy, 2008; pp 3-17.

2. Biermann, T.; Grieve, M. A Computerized Data Base of Mail Order Garments: a Contribution Toward Estimating the Frequency of Fibre Types Found in Clothing. Part 2: the Content of the Data Bank and its Statistical Evaluation, *Forensic Sci. Int.* **1996**, 77, 75-71.

3. Cantrell, S.; Roux, C.; Maynard, P.; Robertson, J. A Textile Fibre Survey as an Aid to the Interpretation of Fibre Evidence in the Sydney Region, *For. Sci. Int.* **2001**, 123, 48-53.

4. Watt, R.; Roux, C.; Robertson, J. The Population of Coloured Textile Fibres in Domestic Washing Machines, *Sci. Justice* **2005**, 45, 75-83.

5. Grieve, M.; Biermann, T. The Population of Coloured Textile Fibres on Outdoor Surfaces, *Sci. Justice* **1997**, 37, 231-239.

6. Palmer, R.; Oliver, S. The Population of Coloured Fibres in Human Head Hair, *Sci. Justice* **2004**, 44, 83-88.

7. Christie, R. *Colour Chemistry*; Royal Society of Chemistry: Cambridge, U.K., 2001.

8. Wiggins, K. Thin Layer Chromatographic Analysis for Fibre Dyes. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K., 1999, pp 291-310.

9. Adolf, F.; Dunlop, J. Microspectroctrophotometry/Colour Measurement. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K., 1999, pp 251-289.

10. Goodpaster, J.; Liszewski, E. Forensic Analysis of Dyed Textile Fibers, *Anal. Bioanal. Chem.* **2009**, 394, 2009-2018.

11. Massonnet, G.; Buzzini, P.; Monard, F.; Jochem, G.; Fido, L.; Bell, S.; Stauber, M.; Coyle, T.; Roux, C.; Hemmings, J.; Leijenhorst, H.; Van Zanten, Z.; Wiggins, K.; Smith, C.; Chabli, S.; Sauneuf, T.; Rosengarten, A.; Meile, C.; Ketterer, S.; Blumer, A. Raman Spectroscopy and Microspectrophotometry of Reactive Dyes on Cotton Fibres: Analysis and Detection Limits, *For. Sci. Int.* **2012**, 222, 200-207.

12. Houck, M. Inter-Comparison of Unrelated Fiber Evidence, *For. Sci. Int.* **2003**, 135, 146-149.

13. Bosch Ojeda, C.; Sanchez Rojas, F. Recent Applications of Derivative Ultraviolet/Visible Absorption Spectrophotometry: 2009-2011 a Review, *Microchem. J.* **2013**, 106, 1-16.

14. Karpinska, J. Derivative Spectrophotometry – Recent Applications and Directions of Developments, *Talanta* **2004**, 64, 801-822.

15. Coyle, T.; Larkin, A.; Smith, K.; Mayo, S.; Chan, A.; Hunt, N. Fibre Mapping – a Case Study, *Sci. Justice* **2004**, 44, 179-186.

16. Grieve, M.; Biermann, T.; Schaub, K. The Individuality of Fibres Used to Provide Forensic Evidence – Not all Blue Polyesters are the Same, *Sci. Justice* **2005**, 45, 13-28.

17. Grieve, M.; Biermann, T.; Schaub, K. The Use of Indigo Derivatives to Dye Denim Material, *Sci. Justice* **2006**, 46, 15-24.

18. Wiggins, K.; Palmer, R.; Hutchinson, W.; Drummond, P. An Investigation into the Use of Calculating the First Derivative of Absorbance Spectra as a Tool for Forensic Fibre Analysis, *Sci. Justice* **2007**, 47, 9-18.

19. Almeida, V.; Vargas, A.; Garcia, J.; Lenzi, E.; Oliveira, C.; Nozaki, J. Simultaneous Determination of the Textile Dyes in Industrial Effluents by the First-Order Derivative Spectrophotometry, *Anal. Sci.* **2009**, 25, 487-492.

20. Bridge, T.; Wardman, R.; Fell, A. Novel Digital Methods for the Qualitative Characterization of Some Acid Dyes Applied to Wool and Nylon, *Analyst* **1985**, 110, 1307-1312.

21. Morgan, S.; Niewland, A.; Mubarak, C.; Hendrix, J.; Enlow, E.; Vasser, B. In *Forensic Discrimination of Dyed Textile Fibres Using UV-VIS and Fluorescence Microspectrophotometry*, Proceedings of the 12th Meeting of the European Fibres Group, Prague, Czech Republic, 2004.

22. Morgan, S.; Hall, S.; Hendrix, J.; Bartick, E. In *Pattern Recognition Methods for Classification of Trace Evidence Textile Fibers from UV/Visible and Fluorescence Spectra*, Proceedings of the National Institute of Justice Trace Evidence Symposium, Kansas City, Missouri, 2011.

23. SWGMAT Forensic Fiber Examination Guidelines. http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf (accessed May 5, 2014).

24. Beebe, K.; Pell, R.; Seasholtz, M. *Chemometrics: a Practical Guide*; John Wiley & Sons: New York, 1998.

25. Barnes, R.; Dhanoa, M.; Lister, S. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, *Appl. Spectrosc.* **1989**, 43, 772-777.

26. Gemperline, P. *Principal Component Analysis*, 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006, pp 69-104.

27. Mahalanobis, P. On the Generalized Distance in Statistics, *Proc. Nat. Instit. Sci. India* **1936**, 2, 49-55.

28. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Square Procedures, *Anal. Chem.* **1964**, 36, 1627-1639.

Table 1.1 Groups of studied fibers categorized by dye and color.

Subclass	Color	Spectra examined	Subclass	Color	Spectra examined
Direct	Blue	60	Reactive	Black	80
	Green	60		Blue	110
	Grey	50		Brown	150
	Pink	30		Green	100
	White	20		Grey	20
	Yellow	40		Orange	40
Vat	Blue	40		Pink	20
	Brown	60		Purple	80
	Green	70		Red	80
	Pink	20		Yellow	60
	Yellow	20			

Table 1.2 Confusion matrix for discrimination of yellow reactive dyed fibers.

	Actual					
Predicted	RY01	RY02	RY03	RY04	RY05	RY06
RY01	10	0	0	0	0	0
RY02	0	10	0	0	0	0
RY03	0	0	8	0	0	0
RY04	0	0	0	10	0	0
RY05	0	0	2	0	10	0
RY06	0	0	0	0	0	10

Table 1.3 Performance of PCA-LDA models following different preprocessing techniques.

Dye Class	Preprocessing technique	Total spectra	Correctly classified	Classification accuracy (%)
Direct	None	260	212	82
	Autoscale	260	220	85
	Normalization	260	218	84
	SNV	260	218	85
	Baseline correction	260	211	81
	Baseline correction + normalization	260	208	80
	First derivative	260	236	91
	First derivative + normalization	260	237	91
	First derivative + SNV	260	231	89
	Second derivative	260	232	89
Vat	None	210	185	88
	Autoscale	210	184	88
	Normalization	210	188	90
	SNV	210	187	89
	Baseline correction	210	181	86
	Baseline correction + normalization	210	187	89
	First derivative	210	188	90
	First derivative + normalization	210	186	89
	First derivative + SNV	210	185	88
	Second derivative	210	193	92
Reactive	None	740	671	91
	Autoscale	740	677	91
	Normalization	740	712	96
	SNV	740	668	90
	Baseline correction	740	687	93
	Baseline correction + normalization	740	673	91
	First derivative	740	709	96
	First derivative + normalization	740	705	95
	First derivative + SNV	740	697	94
	Second derivative	740	694	94

Table 1.4 Correctly classified spectra in each fiber dye class and color based category, with the number of principal components used for each model in parentheses, following different preprocessing techniques.

Dye Class	Color	Preprocessing									
		None	Autos.	Norm.	SNV	Baseline correction	BC + norm.	FD	FD + norm.	FD + SNV	SD
Direct	Blue	50 (6)	51 (6)	46 (6)	41 (7)	51 (6)	46 (7)	55 (4)	53 (6)	51 (4)	50 (4)
	Green	43 (5)	43 (5)	50 (7)	50 (7)	46 (7)	44 (7)	49 (6)	51 (7)	48 (7)	50 (8)
	Grey	44 (5)	44 (5)	43 (6)	36 (7)	42 (6)	39 (8)	47 (5)	49 (4)	46 (5)	50 (8)
	Pink	29 (7)	29 (7)	28 (6)	26 (6)	25 (6)	27 (6)	30 (5)	30 (5)	29 (5)	29 (5)
	White	17 (4)	17 (4)	19 (6)	19 (6)	18 (4)	18 (5)	20 (5)	20 (4)	20 (5)	19 (2)
	Yellow	34 (7)	34 (7)	32 (7)	32 (8)	29 (6)	34 (8)	35 (7)	37 (7)	37 (7)	34 (5)
Vat	Blue	40 (4)	40 (4)	40 (4)	40 (6)	40 (4)	39 (4)	40 (5)	40 (5)	40 (5)	40 (3)
	Brown	42 (5)	41 (5)	40 (5)	41 (7)	38 (5)	43 (8)	44 (8)	43 (6)	43 (8)	50 (6)
	Green	63 (5)	63 (5)	69 (5)	66 (5)	63 (5)	65 (5)	63 (5)	62 (5)	62 (5)	64 (8)
	Pink	20 (4)	20 (4)	19 (3)	20 (4)	20 (5)	20 (6)	20 (5)	20 (5)	19(6)	20 (2)
	Yellow	20 (3)	20 (3)	20 (2)	19 (2)	19 (3)	20 (4)	20 (3)	20 (3)	20 (3)	19 (3)
Reactive	Black	74 (6)	75 (5)	76 (6)	77 (7)	77 (6)	78 (5)	77 (5)	74 (5)	76 (5)	75 (5)
	Blue	95 (6)	98 (5)	108 (5)	102 (4)	98 (5)	103 (4)	107 (5)	101 (5)	104 (5)	103 (5)
	Brown	129 (5)	127 (5)	138 (4)	108 (4)	131 (4)	105 (5)	137 (3)	137 (4)	126 (3)	137 (6)
	Green	89 (4)	90 (4)	96 (4)	91 (6)	91 (4)	94 (6)	93 (4)	97 (4)	96 (3)	92 (5)
	Grey	20 (3)	20 (3)	19 (4)	19 (6)	20 (3)	19 (6)	20 (2)	20 (2)	20 (2)	20 (2)
	Orange	40 (4)	40 (4)	40 (4)	40 (4)	40 (3)	40 (4)	40 (2)	40 (2)	39 (2)	40 (2)
	Pink	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)	20 (2)
	Purple	72 (5)	74 (5)	78 (6)	79 (3)	77 (5)	78 (5)	77 (3)	78 (3)	78 (3)	77 (3)
	Red	75 (5)	75 (5)	79 (5)	79 (4)	77 (4)	78 (6)	79 (5)	78 (5)	79 (5)	76 (6)
	Yellow	57 (7)	58 (7)	58 (4)	53 (7)	57 (7)	58 (7)	58 (4)	58 (7)	58 (6)	58 (6)
Total Correct		**1073**	**1079**	**1118**	**1060**	**1079**	**1068**	**1131**	**1128**	**1111**	**1123**
Classification (%)		**89**	**89**	**92**	**88**	**89**	**88**	**93**	**93**	**92**	**93**

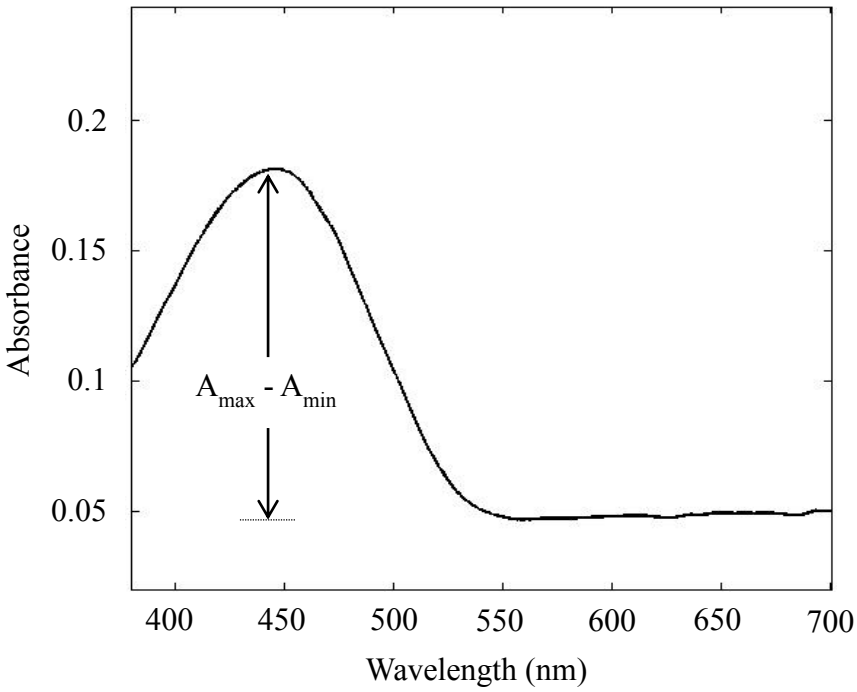

Figure 1.1 Gaussian shaped absorbance spectra and associated A_{max} and A_{min} locations.

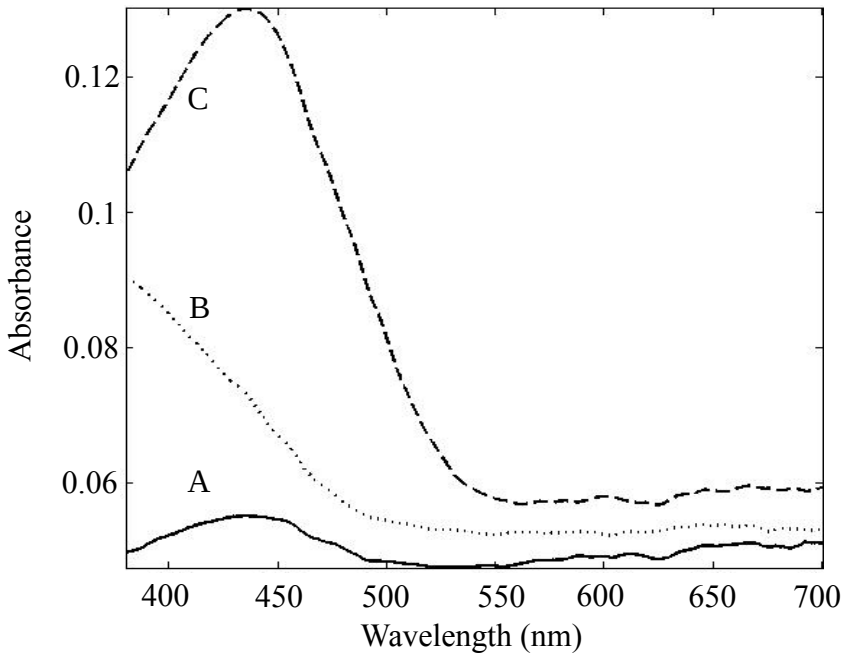

Figure 1.2 Absorbance spectra of three cotton fibers containing A) direct blue 86 and direct direct yellow 106, B) vat black 25, vat brown 81 and vat yellow 33, and C) reactive yellow 206.

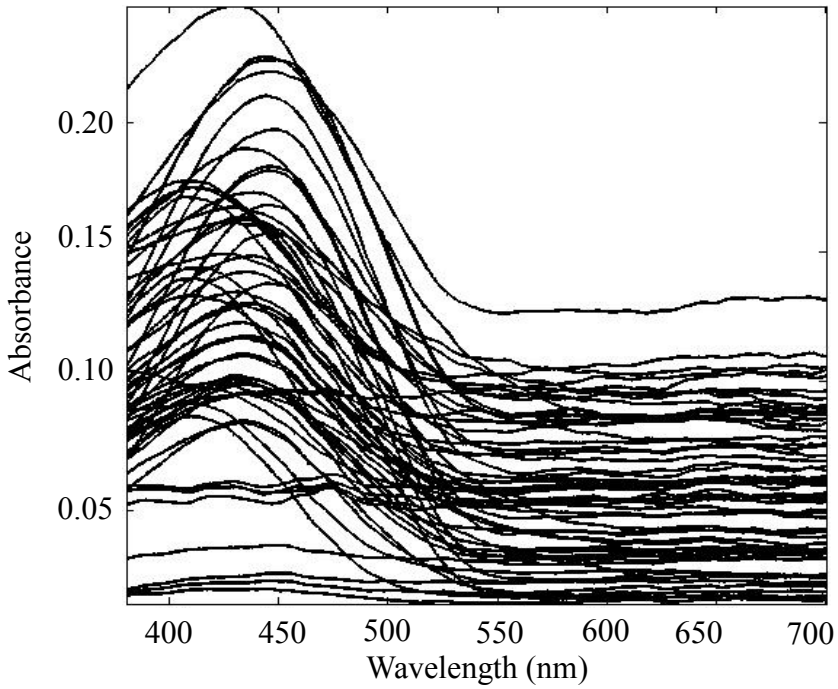

Figure 1.3 Absorbance spectra of six reactive dyed yellow cotton fibers (10 replicates each).

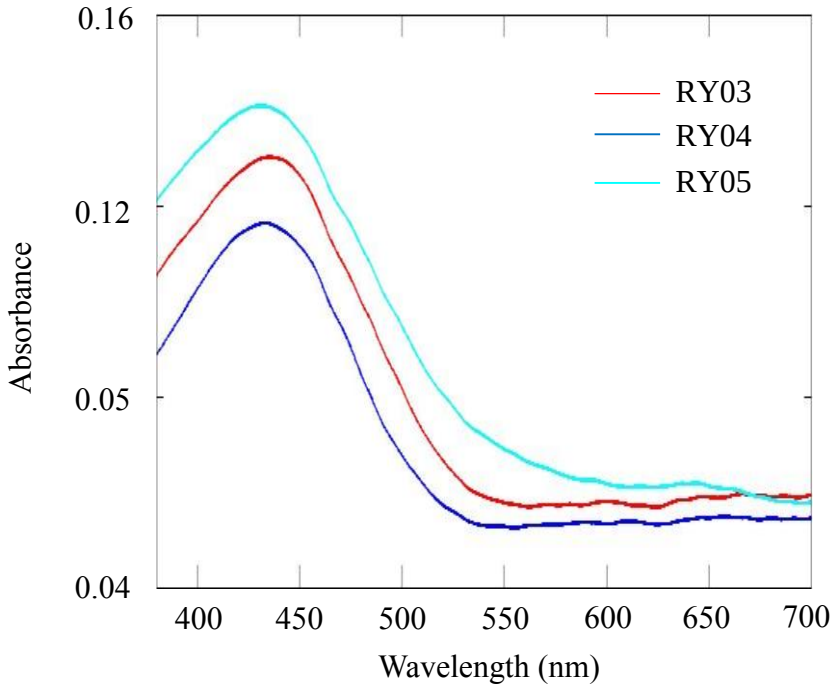

Figure 1.4 Averaged absorbance spectra for three of six yellow reactive dyed cotton fiber samples.

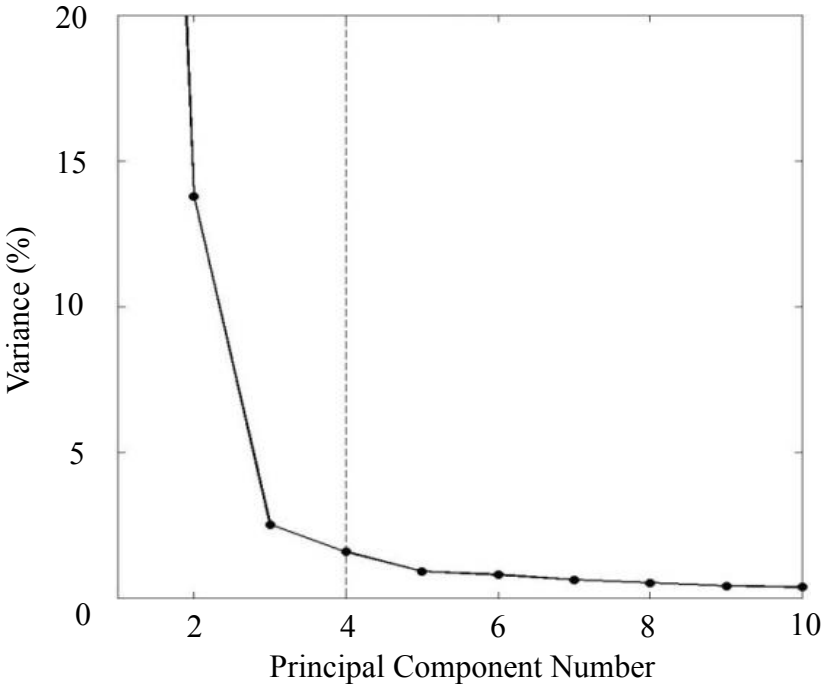

Figure 1.5 Scree plot obtained following PCA on first derivative spectra of six yellow reactive dyed cotton fibers.

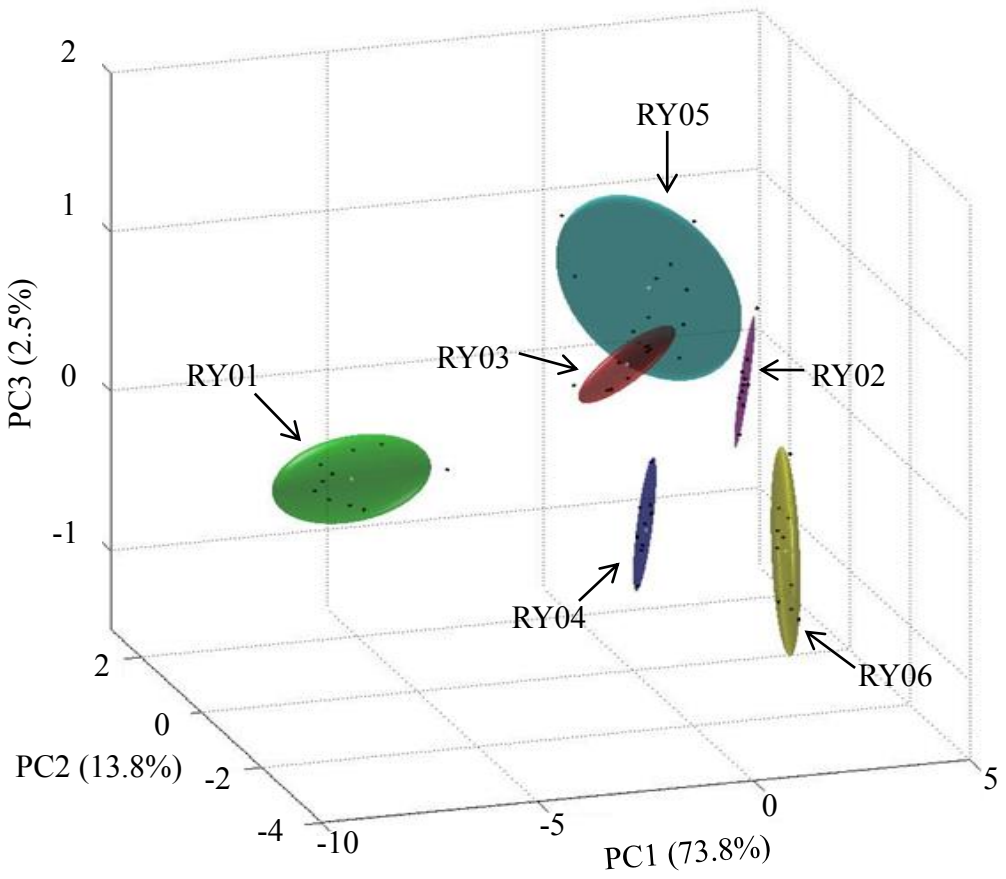

Figure 1.6 PCA scores plot for six reactive dyed yellow cotton fibers after first derivative preprocessing.

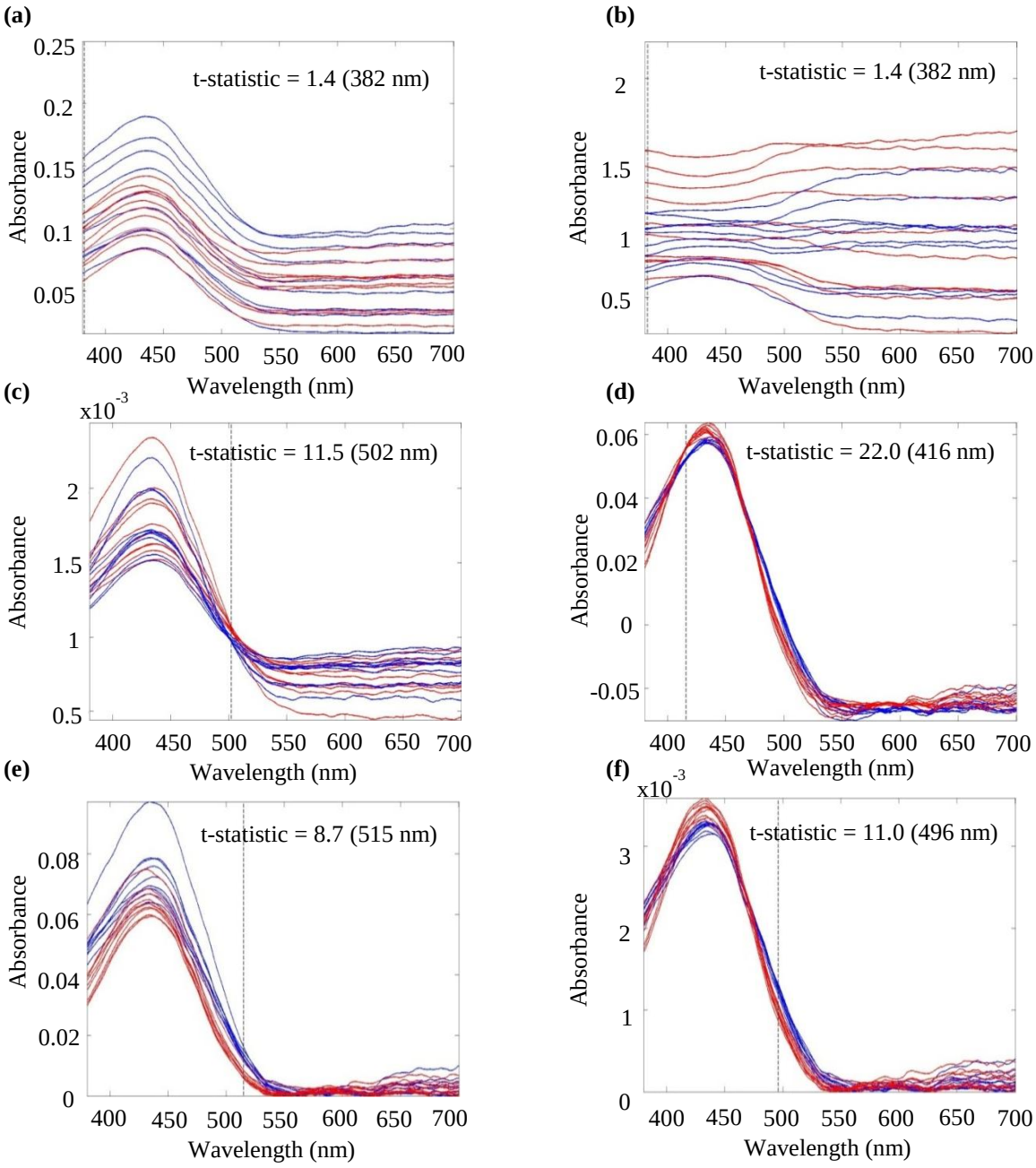

Figure 1.7 Absorbance spectra (10 replicates each) of samples RY03 and RY04 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.

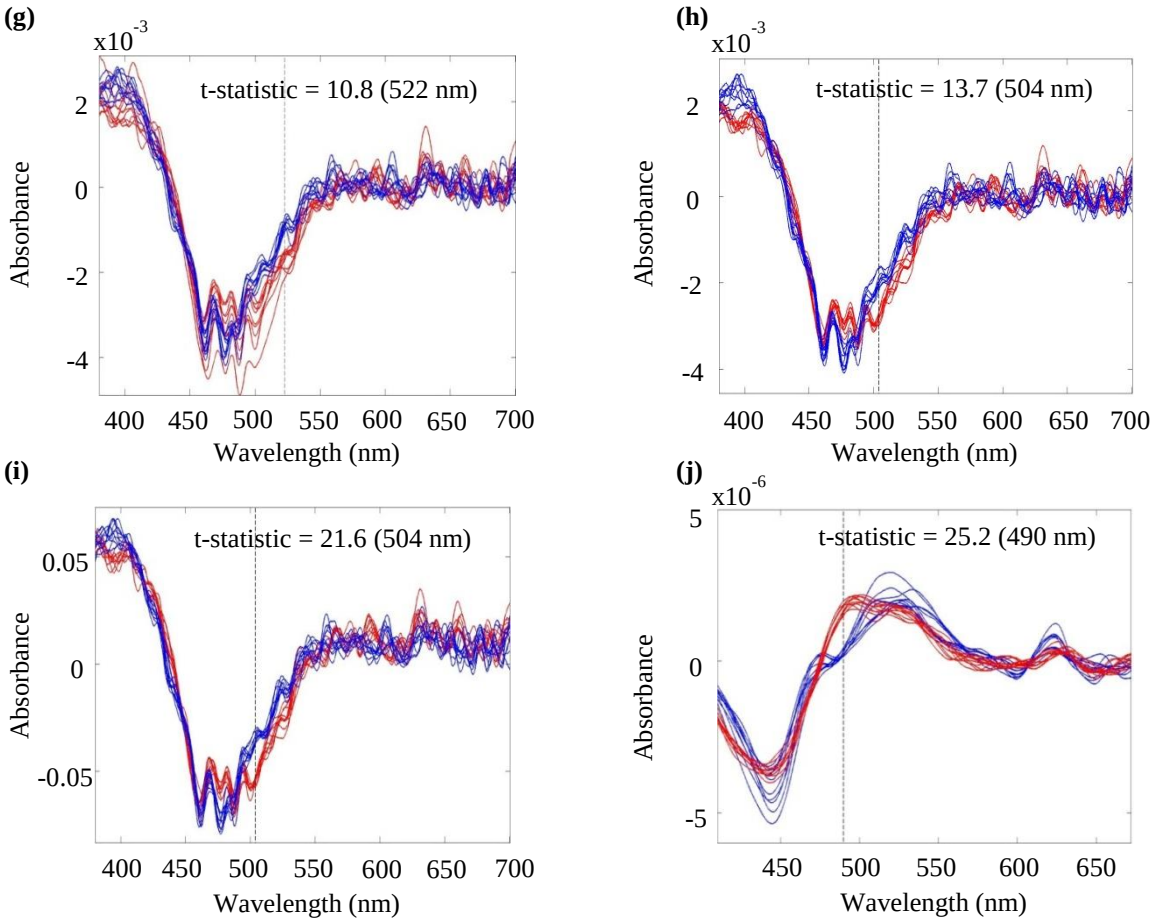

Figure 1.7 Absorbance spectra (10 replicates each) of samples RY03 and RY04 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.

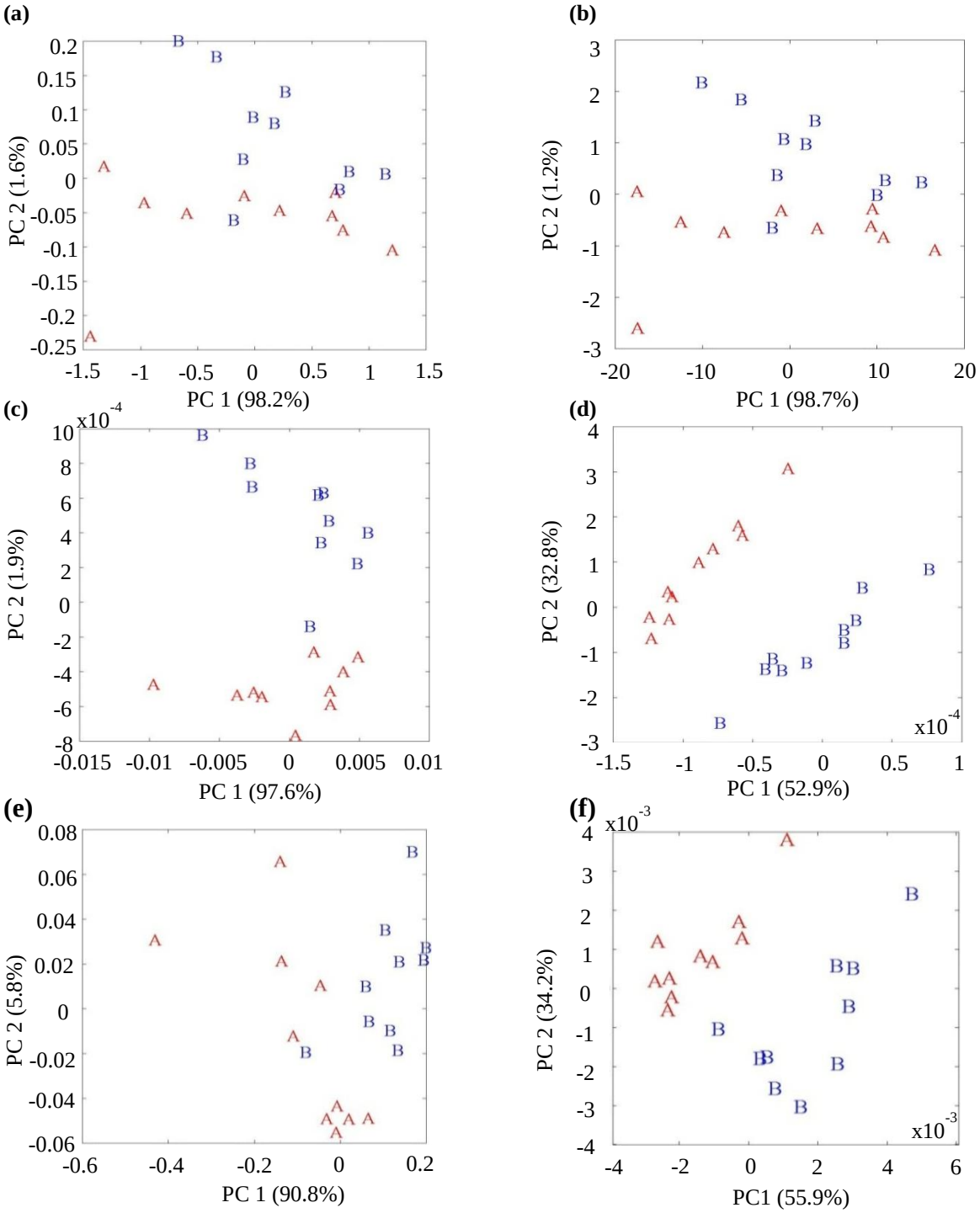

Figure 1.8 PCA scores plot resulting from absorbance spectra (10 replicates each) of samples RY03 and RY04 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.

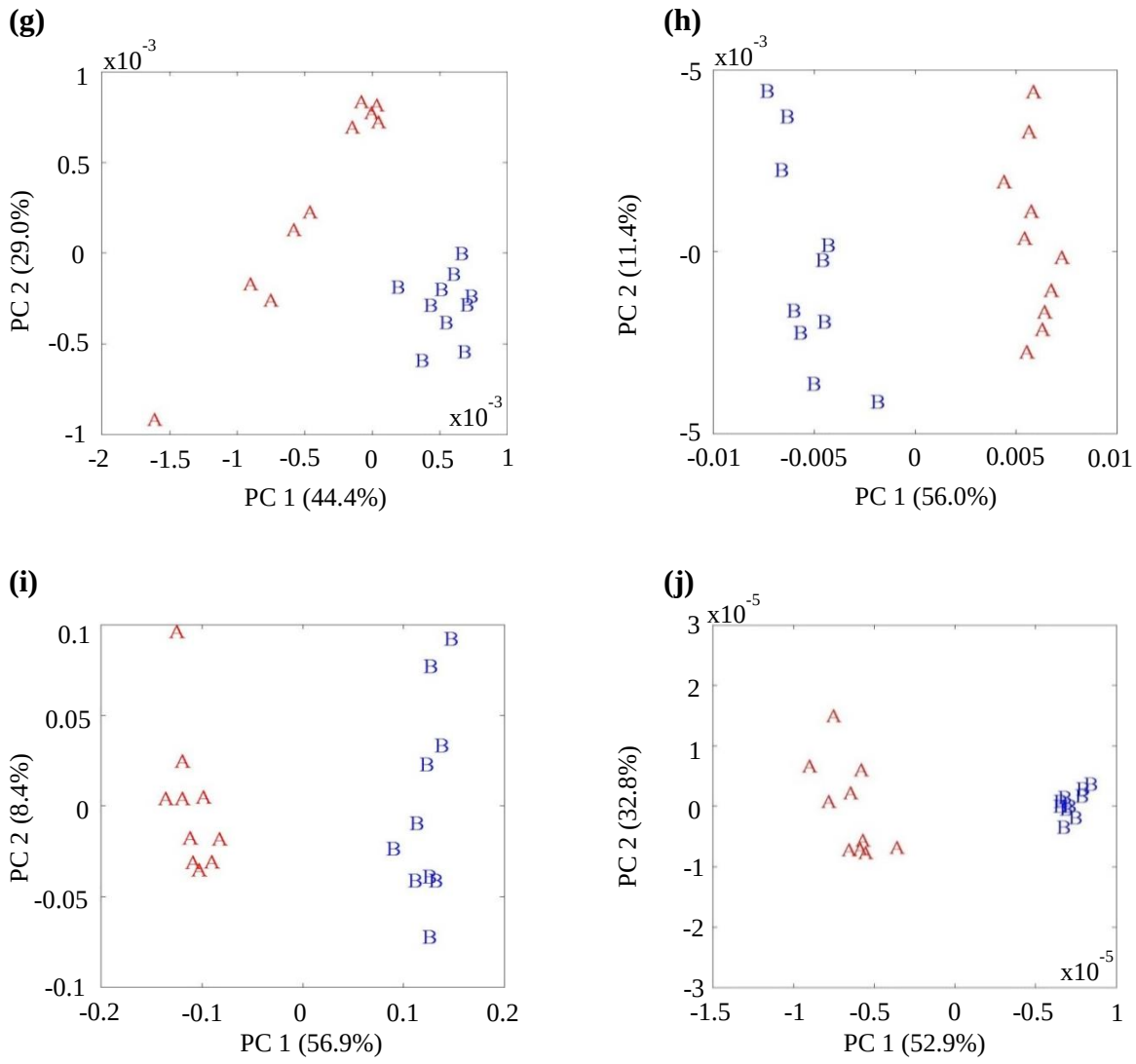

Figure 1.8 PCA scores plot resulting from absorbance spectra (10 replicates each) of samples RY04 and RY05 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.

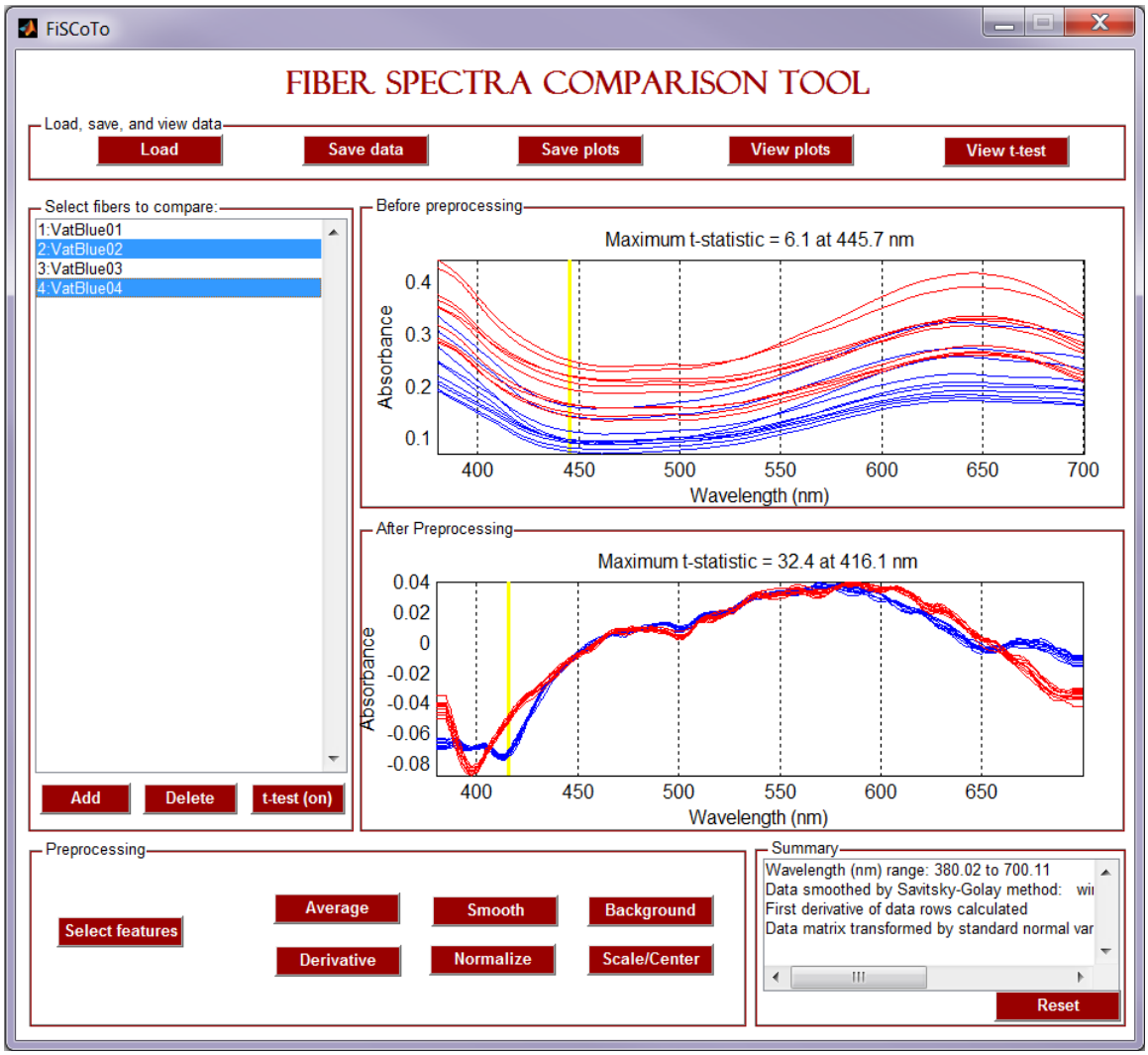

Figure 1.9 Fiber spectra comparison tool graphical user interface.

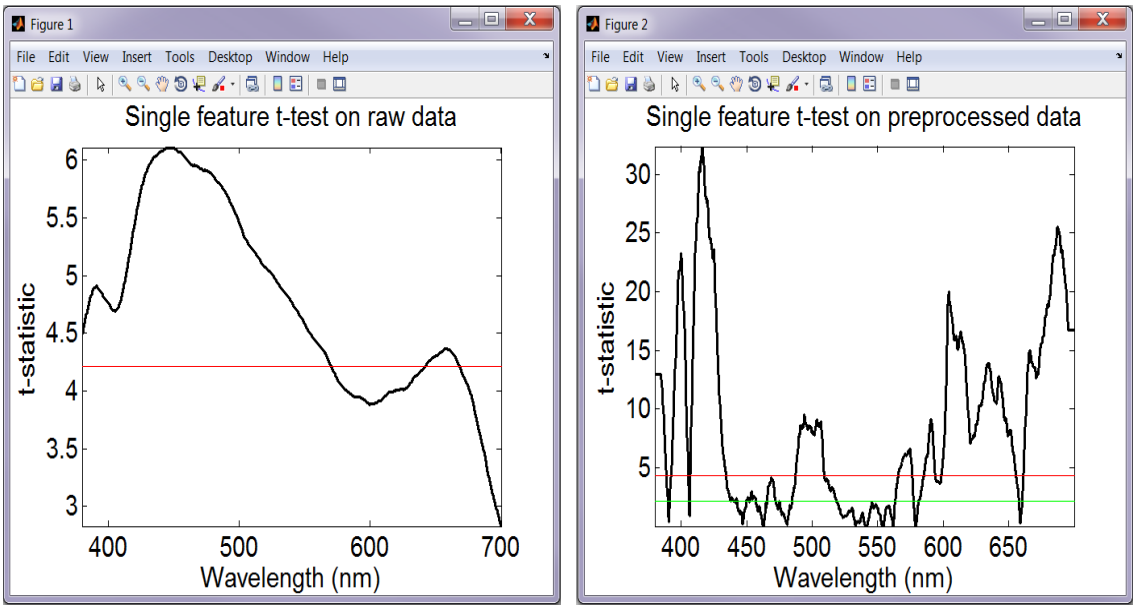

Figure 1.10 Calculated feature-by-feature t-statistics for raw (left) and processed (right) data.

CHAPTER 2

CLASSIFICATION STRATEGIES FOR FUSING UV-VISIBLE ABSORBANCE AND FLUORESCENCE MEASUREMENTS FROM TEXTILE FIBERS

ABSTRACT

A recent emphasis in forensic science has been placed on the development of statistical methods for improving the interpretability of trace evidence analyses. Determining which non-destructive analytical methods will have the highest discrimination power for trace evidence examinations is significant to forensic laboratories to save time and assets. Knowing which analytical techniques provide complimentary information on the evidence is also useful for ensuring that the data collected is utilized in the optimum manner.

This study compares the discrimination ability of ultraviolet-visible (UV-Vis) microspectrophotometry (MSP) and microspectrofluorimetry (MSF), two common techniques used by forensic analysts to study textile fibers. Fusion of MSP and MSF data was also evaluated. Low-, intermediate-, and high-level data fusion strategies were employed in discriminations of over 400 dyed textile fiber samples of cotton, acrylic, nylon 6,6, and polyester, resulting in correct classification rates of 97.8%, 94.6%, and 93.8%, respectively. Comparatively, classification rates of 89.5%, 87.7% and 87.6% resulted from quadratic discriminant analysis models built from isolated absorbance measurements, fluorescence measurements with 405 nm excitation, and fluorescence

measurements with 546 nm excitation, respectively. The results suggest that data fusion is useful for providing additional discriminatory information on textile fibers when compared to single technique data evaluations.

1. INTRODUCTION

Textile fibers are a frequently encountered form of class evidence in forensic investigations of incidents involving personal contact.[1,2] Cotton is the most abundant natural source of fibers in the world[3], and nylon, polyester, and acrylic fibers are three of the most common classes of synthetic fibers likely to be encountered in forensic investigations.[4] As a whole, these fibers are found as trace evidence in more than 80% of all criminal cases pertaining to textile fibers.[5]

Initial fiber analysis is often carried out using forms of microscopy. Polarized light microscopy is beneficial towards determining the polymer class (especially for synthetic fibers), and stereomicroscopy enables an examiner to document the physical characteristics of a fiber such as diameter, color, and luster.[6,7] If the studied fibers, however, are a metameric match, they may not be excluded as originating from the same source. In cases such as these, further analyses can be carried out using optical spectroscopy techniques.

Ultraviolet-visible (UV-Vis) microspectrophotometry (MSP) is a widely accepted technique for discriminating fibers based on color.[8] Color is often the most discriminating characteristic of dyed fibers and is the only distinctive feature of many natural fibers such as cottons due to a lack of variation in morphology.[8,9] To visualize and interpret the vast amount of data that can be collected using one or more spectroscopic techniques, chemometric tools are required.

Data fusion is a chemometric approach which merges data from multiple sources with the expectation that a better interpretation can be gained from the combined data in comparison to any single sensor. Data fusion has been examined for multiple applications in analytical chemistry and has been heavily used in the food and drink realm.[10-16] A fused data process falls into one of three categories: low-level fusion (LLF), intermediate-level fusion (ILF), and high-level fusion (HLF). The fusion level is determined by the stage of data processing at which fusion is carried out. LLF involves merging the raw or preprocessed data signals of each set of input data.[10,11,17] Similarly, ILF also involves fusion at the data-level but occurs after feature selection or feature extraction is employed.[18] For this reason, ILF is also known as feature-level fusion. In HLF, a separate model is built for each set of input data, and the responses of each model are then "fused" together to create one final response.[19]

With relative ease, fusible data can be provided by modern MSP instruments capable of collecting transmittance, absorbance, reflectance, and fluorescence information from fibers. Though most fiber comparisons are carried out by a simple examination of UV-Vis transmittance or absorbance spectra, microspectrofluorimetry (MSF) is a tool often used following MSP to provide additional discriminatory information. This is especially true in cases where the absorbance and transmittance spectra resulting from multiple fibers appear to match. In addition to the dye components, dye-bath additives, and the garment material may contribute to a fiber's fluorescence spectrum.[8]

This study investigates various strategies for the fusion of UV-Vis absorbance and fluorescence data collected from acrylic, cotton, nylon 6,6, and polyester fibers. The

ability to discriminate these fibers was examined using multivariate classification techniques in the form of quadratic discriminant analysis (QDA) or naïve Bayes classification (NBC). Multivariate approaches make it possible to extract as much information as possible from the data, and have been applied previously to color-based discriminations of textile fibers.[20,21] Lastly, the classification accuracies from fused UV-Vis MSP and MSF data were compared to the results found using single modality.

2. EXPERIMENTAL

2.1 Materials

A total of 482 dyed textile fibers collected from various manufacturers were examined in the study. All fibers were classified as belonging to one of four classes, acrylic (at least 85% acrylonitrile), cotton, nylon 6,6, and polyester, based on their polymer component. The fibers were further placed in subgroups based on perceived color of source material to reduce the number of overall comparisons. Single fibers were then removed from the source using razor blades and centered on quartz microscope slides (CRAIC Technologies, Altadena, CA, and Esco Products Inc., Oak Ridge, NJ) using micro tweezers. Spectral grade glycerin (Spectrum Chemical Mfg. Corp., Gardena, CA) and quartz coverslips were used to mount the fibers on the slides for analysis by MSP and MSF.

2.2 Instrumentation

UV-Vis absorbance and fluorescence spectra were acquired using a CRAIC Technologies Quantum Detection Instrument (QDI) 1000 MSP. Data acquisition was carried out using GRAMS/AI 7.0 (Thermo Galactic, Salem, NH). For absorbance measurements, the MSP was operated in transmission mode using a xenon lamp and an

integration time for the charge coupled device (CCD) detector of ~4 ms. Fluorescence was carried out using a mercury light source with a CCD integration time of ~200 ms. Fluorescence spectra were acquired after 365, 405, 436, and 546 nm excitations. All absorbance and fluorescence spectra were calculated by performing an average of 100 scans over a range of 200-850 nm with a 10 nm bandwidth.

2.3 Data Analysis

All data was analyzed using MATLAB version 8.3 (The Mathworks, Inc., Natick, MA) and the statistics toolbox. The 'classify' function with discriminant types 'quadratic' and 'diagquadratic' was used to perform QDA and NBC, respectively.

2.3.1 Preprocessing

All absorbance spectra were truncated to the region of 380 to 700 nm. Truncation of fluorescence spectra was based on the excitation cube used. The lower wavelength cutoff was 390, 444, 470, 581 nm for 365, 405, 436, 546 nm excitation, respectively, with an upper wavelength cutoff of 850 nm.

Further preprocessing to be performed on the datasets was dependent on whether the preprocessed data would be analyzed individually or as part of a fused data classification model. Preprocessing of the absorbance data to be analyzed individually was carried out by first calculating the first derivative of each spectrum. Noise reduction was accomplished using a Savitzky-Golay[22] numerical algorithm with a second order polynomial and nine point moving window. The first derivative is followed by a standard normal variate transformation. Fluorescence data to be analyzed individually was preprocessed using a linear smoothing algorithm with a window width of 31 points. The

final preprocessing step for the individual absorbance and fluorescence datasets is mean centering, a technique which scales each feature to a mean of zero.

Fused absorbance and fluorescence datasets were preprocessed using the second derivative. The second derivative spectra were calculated using the gap-segment method. In the gap-segment technique, derivatives are calculated over a number of variables (i.e., segments) as opposed to adjacent points. The gap, also defined by the user, is the number of variables between the segments. Different types of spectra from multiple datasets were examined using various gap-segment combinations. A gap size of 31 points and segment size of 35 points were chosen as the parameters to be used based on the resulting absorbance and fluorescence spectra which appeared to strike a balance between feature retention and spectral smoothing. As with the single sensor data, the datasets which would ultimately be fused are also mean centered.

2.3.2 Classification

2.3.2.1 Single Modality

Preprocessed absorbance and fluorescence spectra were subjected to principal component analysis (PCA). Using PCA, new sets of uncorrelated variables (PCs) are generated for each dataset.[23] The number of PCs from each set of data to be used for QDA was selected by plotting the variance captured by each PC (i.e., scree plots). QDA is similar to the popularly used discriminant method, linear discriminant analysis, developed by R. A. Fisher in 1936.[24] In QDA, however, a separate variance/covariance matrix is calculated for each group as opposed to the pooled variance/covariance matrix used in LDA. The result of using stratified covariance matrices is a quadratic decision boundary. Classification accuracies for single modality models (the percentage of the

total number of classification attempts which were performed correctly) were determined using leave-one-out cross-validation.

2.3.2.2 Low-level Fusion

A schematic of the LLF process is shown in Figure 2.1. To avoid increased complexity, only fluorescence measurements collected at two excitation wavelengths (405 and 546 nm) were used in the fused data models. The original data collected utilizing MSP and MSF was viewed as three separate $n \times p$ matrices, where n is the number of samples and p is the number of measured wavelengths. Before fusion, p is equal to 940, 1193, and 790 for absorbance, fluorescence-405, and fluorescence-546, respectively. After all data has been preprocessed as described previously, the three matrices are horizontally concatenated, generating a single matrix with n rows and 2923 (940 + 1193 + 790) columns.

In LLF, no feature extraction or selection techniques were carried out prior to classification. Discrimination of the fiber samples in LLF was carried out utilizing the naïve Bayes method of classification which uses probabilistic responses. The naïve Bayes approach assumes that the features for a given class are independent (i.e., the covariance matrices are diagonal ones). Though this assumption is generally not valid, the technique remains a popular one due the surprisingly good results that are often obtained on real-world datasets.[25] The substantial bias generated by assuming the features of a class are independent is usually countered by savings in variance. Classification accuracies for fused data models were determined using 100 iterations of stratified 10-fold cross-validation.

2.3.2.3 Intermediate-level Fusion

Figure 2.2 shows the process by which ILF was carried out. Feature extraction using PCA was performed on the individual preprocessed data matrices to reduce the total number of variables. The significant PCs from each dataset, again determined using scree plots, were then horizontally concatenated to form a final matrix containing *n* rows with the number of columns equaling the total number of PCs selected from all three techniques. The final step in the ILF procedure is to classify each sample using the same naïve Bayes approach and stratified iterated10-fold cross-validation presented for LLF.

2.3.2.4 High-level Fusion

The HLF approach used in this study is described in Figure 2.3. PCA was performed on the preprocessed data collected from each instrumental technique as in ILF. In HLF, however, three individual multivariate classification models based on QDA are built using the resulting PCs from each set of data. Because QDA lacks the naïve assumption that the covariance is zero (i.e., the variables are independent), the method would seemingly be a more prudent approach for all fusion experiments in this study. However, including covariance estimates for each class reduces the total number of PCs that can be used for classification. Because HLF involves dimensionality reduction and the creation of separate classification models for each set of measurements, the number of variables in HLF is expected to be much lower than the number of variables used in LLF and ILF.

Each fiber was given a final classification based on a majority vote[26] of the three classification models. In instances where no majority was obtained, the fiber was designated a group based on the classification resulting from the visible absorbance

technique. Absorbance measurements were given more weight towards the final classifications, because MSP is the preferred technique for discriminating dyed textile fibers in forensic analyses.

3. RESULTS AND DISCUSSION

3.1 Absorbance versus fluorescence

The benefit of using first derivative preprocessing on absorbance spectra was observed during the analysis of a set of five yellow acrylic fibers. Leave-one-out cross-validation was performed on the PCA-QDA models of both the raw spectra and first derivative spectra. The resulting confusion matrices displayed in Table 2.1 show the difficulty of separating yellow acrylic fiber samples numbered two and four based on their raw spectra. Calculation of first derivative, however, resulted in a classification accuracy of 100% for the same set of fibers. The inability of the PCA-QDA model to discriminate yellow acrylic samples two and four is shown visually using the canonical variate scores plot in Figure 2.4. Nearly 100% of the variation in the raw and first derivative datasets can be explained using the axes of the first three canonical variates. Although these axes can be rotated in all directions, no such rotation is able to avoid overlap of the 95% confidence ellipses resulting from the raw spectra of samples two and four. However, the QDA scores plot produced by first derivative data shows a clear separation between these two samples.

Figure 2.5 shows the UV-Vis absorbance spectra for samples two and four of the yellow acrylic dataset before and after calculating the first derivative. The strong similarity between the raw absorption spectra results from the samples having the same three dyes: Yellow 29, Red 29, and Blue 147. Variance between these two groups of

spectra is more clearly displayed using first derivative preprocessing. The 460 to 490 nm region of the first derivative spectra, which corresponds to points of inflection in the original spectra, shows two distinct patterns for each sample.

Classification accuracies of UV-Vis MSP and MSF for all fibers in this study are listed in Table 2.2. Entries marked with "x" indicate groups of fibers in which QDA was not performed due to the limited number of samples that fell within a particular subclass. Numbers of correctly classified spectra based on leave-one-out cross validation are listed in columns for absorbance and fluorescence spectra at four excitation wavelengths.

As a whole, UV-Vis MSP had the highest discriminating power of any of the methods used in the study. Classification accuracies of UV-Vis MSP for acrylic, cotton, nylon 6,6, and polyester were 95.68%, 86.13%, 95.36%, and 81.38%, respectively. The lower discriminating power witnessed for cotton are not unexpected. Uneven dye uptakes by natural fibers such as cotton give higher degrees of variation between replicate spectra. The cotton fibers in this study also showed some of the lowest absorbance values of any class which further hindered classifications. Many such cotton fibers contained direct dyes, a selection of dyes which are popular due to their cost and ease of application but often display lower wetfastness and less brilliant colors when compared to other common dying classes for cotton such as reactive and vat dyes. Highest discrimination of absorbance spectra was seen in groups of orange (98.83%), red (97.40%), and purple (96.93%). Lowest discrimination of absorbance spectra was seen with brown and white fibers (especially acrylic and nylon 6,6) which mainly had low absorbing featureless spectra in the visible range studied. White fibers often only contain optical brighteners, compounds that do not absorb in the visible region.

The high number of misclassifications seen within brown polyester fibers contributed to the decreased discrimination ability of this method when applied to polyester. In addition to the difficulty of discriminating the UV-Vis spectra of brown polyester fibers, large sample sets tend to hinder the prediction accuracy of the QDA model. The PCA scores plot of the preprocessed absorbance spectra for all brown polyester fibers in this study is shown Figure 2.6. From this plot, it was determined that the large confidence ellipses associated with samples 3 and 18 are a result of significant variation between spectra within each group. The large within-group variation in these samples made misclassifications of their spectra likely to occur. The scores plot also indicates samples 8, 9, and 10 as being clearly distinguishable from all other brown polyester samples, and appears to show two separate clusters containing multiple samples.

There is a significant amount of overlap between samples 22, 24, 26, 28, and 30 (Cluster A in Figure 2.6). Replicate absorbance spectra of the samples in Cluster A were averaged and are shown in Figure 2.7. UV-Vis spectra of these samples were similar, but had multiple points of comparison. When analyzed separately, 46 of 50 (92%) of these spectra were correctly classified by PCA-QDA. All other brown polyester samples not previously mentioned (29 in all) are located in the second cluster (Cluster B in Figure 2.6). Many of the fibers in Cluster B share spectra which are either broad or low-absorbing, and have limited points of comparison. In general, it is difficult to discriminate these types of spectra regardless of the method used. Analysis of this group resulted in 190 of 290 (65.52%) spectra being classified correctly.

Fluorescence spectra collected with 405, 436, and 545 nm excitations all showed similar discrimination. The classification accuracies resulting from a lower excitation wavelength of 365 nm showed no distinct advantage over other methods for any group studied, and therefore by itself is not recommended for fiber analysis. As with UV-Vis MSP, discrimination was highest for MSF in groups of orange and purple fibers with each group having classification accuracies of over 90% at all wavelengths of excitation. Discrimination at all excitation wavelengths is lowest for nylon fibers. The issues with cotton fibers, mentioned previously with absorbance spectra, are apparent in the discrimination power of MSF as well.

Unlike absorption spectra, which are solely the result of the dyes contained on or within the fiber, fluorescence spectra can be influenced by a fiber's polymeric composition, its dye components, or any optical brighteners added to it. Two types of fibers which saw increases in discrimination after transitioning from absorption to fluorescence spectral evaluations were polyester (from 80.7% to 90.5%) and acrylic (from 95.8% to 97.5%). Application of fluorescent dyes is more commonly associated with synthetic fibers such as polyester and acrylic than natural fibers such as cotton. Both disperse and cationic (basic) dyes were used in the dying process of the polyester fibers, whereas the acrylic fibers were exclusively dyed using cationic dyes. Disperse dyes (especially greenish-yellow shades) include some of the most important industrial fluorescent dyes. Approximately 25% of the 123 polyester fibers studied in this work contained Disperse Yellow 108. All polyester fibers dyed with Disperse Yellow 108 showed high fluorescence intensities using 405 nm excitation. Similarly, all acrylic fibers

dyed with Basic Blue 3 showed considerable fluorescence using 546 nm excitation. Of the 125 acrylic fibers studied, only ~7% were noted to have been dyed with Basic Blue 3.

The ability to correctly classify many fluorescence spectra of brown polyester fibers which were misclassified based on their absorption spectra is one reason the classification rates between the two methods are so comparable. Figure 2.8 shows the fluorescence spectra of one such sample, which gave an almost featureless absorption spectrum, at three different excitation wavelengths. Although differing greatly in intensity, the fluorescence spectra of this sample have multiple points of comparison which aid in discriminant analysis. The double-peak shape resulting from emission in the blue region after excitation at 365 nm is consistent with the types of fluorescence emission spectra found for optical brighteners.[8] It is believed that the lower accuracies obtained from fluorescence at 365 nm are due to the limited variability of optical brightening agents and a reduction in fluorescence contribution of the dye components when shifting into the UV region.

3.2 Fusion of absorbance and fluorescence measurements

Absorbance and fluorescence measurements collected at two different excitation wavelengths (405 and 546 nm) generated a total of three matrices for data fusion analyses. In the previous section, it was determined using PCA-QDA to discriminate the same fibers that fluorescence data resulting from UV (365 nm) excitation resulted in significantly lower classification accuracies compared to visible wavelength fluorescence excitations. Based on that result, and the desire to limit the number of measurements which need to be collected, only two (violet and green) fluorescence excitations were selected for fusion.

A summary of classification results for all data fusion methods is presented in Table 2.3. When comparing the three levels of data fusion, LLF showed greater than three percent increases in classification accuracies over ILF and HLF. The likely cause of the better results obtained using LLF, is the lack of a feature extraction step in the process. By using PCA to reduce the dimensionality of a dataset, it is possible that valuable discriminatory information will be lost if the correct numbers of PCs are not retained. Because QDA was used in HLF for multivariate classification, there is a limit on the number of PCs which may be selected. The total number of PCs used towards classification must be at least two fewer than the number of replicate spectra collected per fiber.[27] Otherwise, singular covariance matrices which cannot be inverted can result for one or more classes. This restriction may become a problem when the size of the group is large and the number of replicate spectra collected per fiber is small. For all groups including 10 or more fiber samples, an average classification of 93.4% was determined for HLF. This result is significantly less than the 97.6% accuracy determined for LLF in which PCA was not performed.

One instance where HLF was largely outperformed by LLF was with a set of 39 brown polyester fibers, the largest group studied, of which 10 replicate spectra per fiber were collected. Previously, fibers in this group had been discriminated based on their isolated absorbance, fluorescence-405, and fluorescence-546 spectra using QDA. The classification accuracies obtained were 76.3%, 90.5%, and 88.4% for absorbance, fluorescence-405, and fluorescence-546 models, respectively. The problem of inadequate PCs which may have resulted in the lower discrimination ability of the individual absorbance model has less of an impact in fused models.

Figure 2.9 shows absorbance, fluorescence-405, and fluorescence-546 spectra before and after second derivative preprocessing followed by mean centering for five yellow acrylic fibers. Two of the fibers, AY02 and AY04, share similar raw absorbance and fluorescence-546 spectra. The preprocessed absorbance spectra for the same two fibers appear to show some slight differences between the groups. Conversely, the preprocessed fluorescence-546 spectra show no clear differences between AY02 and AY04. The raw fluorescence-405 spectral intensities observed by AY02 are much greater than those of AY04 and give very different second derivative/mean-centered spectra. Had one used only fluorescence-405 measurements for classification, however, it would have been difficult to discriminate AY01, AY02, and AY05. It is only by examining the fluorescence-546 spectra that the strong fluorescence of Basic Blue 3, one of the three cationic dyes bound to sample AY05, can be seen. By using data fusion to take all three techniques into account, the five yellow acrylic fibers were discriminated with 100% accuracy regardless of the level of fusion used.

With a difference in classification accuracy of less than one percent, it was concluded that the ILF and HLF processes gave very similar outcomes. In ILF and HLF, PCA is carried out on the preprocessed data. PCA scores plots resulting from a group of purple cotton fibers, one of the medium-sized groups in the study, are shown in Figure 2.10. In HLF fusion, classifications are based on the PCs calculated for each individual technique. The scores plot for the preprocessed absorbance data shows only three groups, CP03, CP04, and CP06, as being completely distinguishable based on the 95% confidence ellipses provided for each group. By examining the scores plot generated from the fluorescence-405 technique, it can be concluded that groups CP01 and CP05 are

distinguishable from the overlapping samples in the absorbance plot. This leaves four remaining fibers, two of which (CP02 and CP09) appear to be separable using fluorescence-546. Because CP07 and CP08 show overlap in the scores plots resulting from all techniques, it is not surprising that several CP08 samples were predicted as belonging to CP07, as indicated in Table 2.4, based on the simple majority vote used in HLF. It is also worth noting that fibers CP07 and CP08 were dyed using the same red, blue and yellow reactive dyes. The fused scores plot shows the high similarity between CP02 and CP08 which resulted in multiple CP02 samples being misclassified during the ILF process. Naïve Bayes classification on concatenated preprocessed spectra rather than PC scores resulted in zero misclassifications in the set of purple cotton fibers using LLF.

In instances where fibers were distinguishable using all three instrumental techniques, the HLF method can diminish the impact of spectral outliers. For example, if one of the replicate absorbance spectra obtained from a sample was an outlier, it is likely that the concatenated fluorescence measurements will not be outliers. As a result, the outlier in the absorbance spectra would not affect the overall classification accuracy. In instances where only one classifier is used, outliers can cause losses in accuracy if some form of multivariate outlier detection is not used beforehand.

4. CONCLUSION

Multivariate classification methods are an effective way to discriminate replicate spectra of multiple fibers. In forensics, it is important to identify which methods of analyses are most discriminating in order to ensure that as little time and resources are spent examining evidence as possible. While both MSP and MSF provided discriminating information, MSP collectively outperformed the fluorescence methods used. Two factors

which seemingly contributed to the lower classification accuracies obtained utilizing fluorescence spectra were the sizeable within-class variations between spectra as well as the increased noise in fluorescence spectra as compared to absorption spectra. Using PCA-QDA, a classification of nearly 90% was achieved for 482 fibers studied using UV-Vis MSP.

The effect of fusing absorbance and fluorescence measurements was also investigated for the purpose of discriminating dyed textile fibers. The classification accuracies resulting from LLF, ILF, and HLF were between four and eight percent higher than those found using only UV-Vis MSP measurements. This suggests that additional discriminatory information can be obtained about the fibers by combining UV-Vis absorbance and fluorescence data. Mainly, this is due to fluorescent dyes, but may also include dye bath additives and contributions from the garment. When comparing the three levels of data fusion, it was found that LLF outperformed ILF and HLF by greater than three percent. Clearly, the naïve Bayes classifier worked well for the full datasets in this study. Therefore, dimensionality reduction was unnecessary. By extracting features using PCA, there is a risk of throwing away valuable information.

It is clear from this study that the discrimination power of MSF was dependent on the number of fibers in our database containing fluorescent dyes. Future work in this area should focus on determining the likelihood of finding fibers with fluorescent dyes in forensic casework. This will allow forensic analysts to determine how often MSF is likely to provide useful information. In addition, some of the more complex methods for combining pattern classifiers in HLF, as opposed to the simple majority vote technique applied in this study, could be subjects of interest.

ACKNOWLEDGMENTS

This project was supported by Award No. 2010-DN-BX-K245 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Portions of this work were presented at the Great Scientific Exchange (SciX) conference at the Rhode Island Convention Center in Providence, Rhode Island on September 30, 2015. The co-authorship of Alexander A. Nieuwland, Jessica M. McCutcheon, John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan is also acknowledged. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect those of the Department of Justice.

REFERENCES

1. Rendle, D.; Wiggins, K. Forensic Analysis of Textile Fibre Dyes, *Rev. Progress. Coloration* **1995**, 25, 29-34.

2. Gaudette, B. The Forensic Aspects of Textile Fiber Examination, In *Forensic Science Handbook*, Vol. II.; Saferstein, R., Ed.; Prentice Hall: New Jersey, 1988.

3. van Dam, J. In *Environmental benefits of natural fibre production and use*, Proceedings of the Symposium on Natural Fibres, Rome, Italy, 2008; pp 3-17.

4. Tungol, M.; Bartick, E.; Montaser, A. Analysis of Single Polymer Fibers by Fourier Transform Infrared Microscopy: The Results of Case Studies," *J. Forensic Sci.* **1991**, 36, 1027-1043.

5. Stefan, A.; Dockery, C.; Nieuwland, A.; Roberson, S.; Baguley, B.; Hendrix, J.; Morgan, S. Forensic Analysis of Anthraquinone, Azo, and Metal Complex Acid Dyes from Nylon Fibers by Micro-extraction and Capillary Electrophoresis, *Anal. Bioanal. Chem.* **2009**, 394, 2077-2085.

6. Stoeffler, S. A Flowchart System for the Identification of Common Synthetic Fibers by Polarized Light Microscopy, *J. Forensic Sci.* **1996**, 41, 297-299.

7. Palenik, S. Microscopical Examinations of Fibers, In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M.; Taylor and Francis: London, U.K. 1999, pp. 153-177.

8. Adolf, P.; Dunlop, J. Microspectrophotometry/Colour Measurement. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K. 1999, pp. 251-289.

9. Palmer, R.; Hutchinson, W.; Fryer, V. The Discrimination of (Non-denim) Blue Cotton, *Sci. Justice* **2009**, 49, 12-18.

10. Di Natale, Paolesse, R.; Macagnano, A.; Mantini, A.; D'Amico, A.; Legin, A.; Lvova, L.; Rudnitskay, A.; Vlasov, Y. Electronic Nose and Electronic Tongue Integration for Improved Classification of Clinical and Food Samples, *Sens. Actuators B* **2000**, 64, 15-21.

11. Rudnitskaya, A.; Kirsanov, D.; Legin, A.; Beullens, K.; Lammertyn, J.; Nicolai, B.; Irudayaraj, J. Analysis of Apples Varieties – Comparison of Electronic Tongue with Different Analytical Techniques, *Sens. Actuators B* **2006**, 116, 23-28.

12. Vera, L.; Acena, L.; Guasch, J.; Boque, R.; Mestres, M.; Busto, O. Discrimination and Sensory Description of Beers through Data Fusion, *Talanta* **2011**, 87, 136-142.

13. Charve, J.; Chen, C.; Hegeman, A.; Reineccius, G. Evaluation of Instrumental Methods for the Untargeted Analysis of Chemical Stimuli of Orange Juice Flavor, *Flavour Fragr. J.* **2011**, 26, 429-440.

14. Subari, N.; Saleh, J.; Shakaff, A.; Zakaria, A. A Hybrid Sensing Approach for Pure and Adulterated Honey Classification. *Sensors* **2012**, 12, 14022-14040.

15. Pizzaro, C.; Rodriguez-Tecedor, S.; Perez-del-Notario, N.; Esteban-Diez, Gonzalez-Saiz, J. Classification of Spanish Extra Virgin Olive Oils by Data Fusion of Visible Spectroscopic Fingerprints and Chemical Descriptors, *Food Chem.* **2013**, 138, 915-922.

16. Biancolillo, A.; Bucci, R.; Magri, A. L.; Magri, A. D.; Marini, F. Data-Fusion for Multiplatform Characterization of an Italian Craft Beer at its Authentication, *Anal. Chim. Acta* **2014**, 820, 23-31.

17. Liu, Y.; Brown, S. Wavelet Multiscale Regression from the Perspective of Data Fusion: New Conceptual Approaches, *Anal. Bioanal. Chem.* **2004**, 380, 445-452.

18. Roussel, S.; Bellon-Maurel, V.; Roger, J.; Grenier, P. Fusion of Aroma, FT-IR and UV Sensor Data Based on the Bayesian Inference. Application to the Discrimination of White Grape Varieties." *Chemom. Intell. Lab. Syst.* **2003**, 65, 209-219.

19. Doeswijk, T.; Smilde, A.; Hageman, J.; Westerhuis, J.; van Eeuwijk, F. On the Increase of Predictive Performance with High-Level Data Fusion, *Anal. Chim. Acta* **2011**, 705, 41-47.

20. Morgan, S.; Niewland, A.; Mubarak, C.; Hendrix, J.; Enlow, E.; Vasser, B. In *Forensic Discrimination of Dyed Textile Fibres Using UV-VIS and Fluorescence Microspectrophotometry*, Proceedings of the 12th Meeting of the European Fibres Group, Prague, Czech Republic, 2004.

21. Morgan, S.; Hall, S.; Hendrix, J.; Bartick, E. In *Pattern Recognition Methods for Classification of Trace Evidence Textile Fibers from UV/Visible and Fluorescence Spectra*, Proceedings of the National Institute of Justice Trace Evidence Symposium, Kansas City, Missouri, 2011.

22. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.* **1964**, 36, 1627-1639.

23. Gemperline, P. Principal component analysis, In *Practical Guide to Chemometrics*, 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006, pp 69-104.

24. Fisher, R. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugenics.* **1936**, 179-188.

25. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 3rd ed.; Academic Press: California, 2006.

26. Kuncheva, L. Combining Pattern Classifiers; Wiley: New Jersey, 2004.

27. Brereton, R. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: West Sussex, U.K., 2003, p 238.

Table 2.1 Confusion matrices displaying results of leave-one-out cross-validation for five yellow acrylic samples (10 replicate spectra each).

Data	Group	Observed classifications					Classification accuracy (%)
		G1	G2	G3	G4	G5	
Raw	G1	10	0	0	0	0	100.00
	G2	0	7	0	3	0	70.00
	G3	0	0	10	0	0	100.00
	G4	0	0	0	10	0	100.00
	G5	0	0	0	0	10	100.00
	Total	10	7	10	13	10	94.00
First Deriv.	G1	10	0	0	0	0	100.00
	G2	0	10	0	0	0	100.00
	G3	0	0	10	0	0	100.00
	G4	0	0	0	10	0	100.00
	G5	0	0	0	0	10	100.00
	Total	10	10	10	10	10	100.00

Table 2.2 Numbers of correctly classified spectra for all fiber types and colors.

Color	*Fiber Type*	*Groups*	*Absorbance*	*FE 365*	*FE 405*	*FE 436*	*FE 546*
Black	Acrylic	6	60	58	57	55	60
	Cotton	8	77	64	66	72	73
	Nylon 6,6	13	129	91	96	92	105
	Polyester	14	120	105	122	117	104
Blue	Acrylic	30	294	251	282	281	296
	Cotton	21	182	147	164	165	149
	Nylon 6,6	26	245	208	217	215	242
	Polyester	19	171	171	173	168	183
Brown	Acrylic	17	163	160	159	161	168
	Cotton	22	179	172	185	195	186
	Nylon 6,6	16	144	117	115	111	101
	Polyester	39	261	310	335	331	345
Green	Acrylic	16	154	147	154	158	157
	Cotton	23	193	177	197	201	166
	Nylon 6,6	15	147	112	111	129	118
	Polyester	19	145	148	172	167	166
Grey	Acrylic	5	50	47	49	50	50
	Cotton	7	62	59	63	65	54
	Nylon 6,6	8	77	47	55	64	49
	Polyester	6	49	51	58	60	60
Orange	Acrylic	3	30	30	30	30	30
	Cotton	4	40	40	40	40	40
	Nylon 6,6	7	69	53	65	60	55
	Polyester	3	30	30	30	30	30
Pink	Acrylic	6	60	59	59	60	60
	Cotton	7	62	59	61	61	51
	Nylon 6,6	2	20	20	20	20	20
	Polyester	2	20	20	20	20	20
Purple	Acrylic	10	99	98	97	98	98
	Cotton	9	85	79	87	88	86
	Nylon 6,6	7	69	62	58	58	66
	Polyester	1	x	x	x	x	X
Red	Acrylic	16	159	154	156	153	154
	Cotton	8	78	72	78	69	78
	Nylon 6,6	12	116	92	100	95	105
	Polyester	10	96	78	85	93	100
White	Acrylic	11	80	89	101	99	96
	Cotton	2	20	18	16	17	15
	Nylon 6,6	4	33	36	32	39	39
	Polyester	7	69	61	62	67	67
Yellow	Acrylic	5	48	47	50	49	50
	Cotton	13	90	109	113	106	90
	Nylon 6,6	1	x	x	x	x	X
	Polyester	4	40	38	39	40	38
Total Spectra		4820	4314	3986	4229	4249	4220
%Classification Accuracy			89.50	82.70	87.74	88.15	87.55

Table 2.3 Correct classification percentages and number of principal components used from each instrumental technique for all fiber types and colors.

Color	*Fiber type*	*Groups*	*Classification (%)*			*Number of PCs*		
			LLF	*ILF*	*HLF*	*Abs.*	*F-405*	*F-546*
Black	Acrylic	6	100	100	100	3	4	4
	Cotton	8	97.34	97.31	97.04	5	6	5
	Nylon 6,6	13	99.23	94.45	94.70	4	4	5
	Polyester	14	92.37	90.61	90.94	4	5	4
Blue	Acrylic	30	100	96.53	98.94	4	6	5
	Cotton	21	97.74	92.21	89.43	4	6	5
	Nylon 6,6	26	99.32	98.29	96.92	5	5	5
	Polyester	19	99.47	97.96	97.21	4	6	3
Brown	Acrylic	17	100	97.91	99.04	5	6	5
	Cotton	22	96.69	90.37	89.09	5	6	4
	Nylon 6,6	16	96.23	91.88	88.87	6	6	4
	Polyester	39	94.37	85.18	88.55	5	5	6
Green	Acrylic	16	99.89	96.54	98.59	4	6	4
	Cotton	23	98.16	94.82	90.62	5	5	5
	Nylon 6,6	15	100	100	94.06	6	5	4
	Polyester	19	91.97	87.01	91.75	5	5	5
Grey	Acrylic	5	100	100	100	3	6	3
	Cotton	7	100	98.57	90.10	6	5	5
	Nylon 6,6	8	99.39	81.98	76.54	2	5	5
	Polyester	6	90.35	88.68	89.55	3	5	3
Orange	Acrylic	3	100	100	100	2	3	7
	Cotton	4	100	100	100	4	4	3
	Nylon 6,6	7	98.57	96.36	91.36	4	5	3
	Polyester	3	100	100	100	4	4	2
Pink	Acrylic	6	100	100	100	5	5	5
	Cotton	7	95.63	93.41	88.99	7	5	4
	Nylon 6,6	2	100	95.00	100	2	2	7
	Polyester	2	100	100	100	2	4	2
Purple	Acrylic	10	99.00	99.00	98.92	4	6	3
	Cotton	9	100	99.00	99.17	4	5	6
	Nylon 6,6	7	100	100	97.07	3	5	3
	Polyester	1	x	x	x	x	x	x
Red	Acrylic	16	99.38	99.38	99.70	5	6	5
	Cotton	8	98.75	96.84	98.81	5	6	5
	Nylon 6,6	12	97.68	94.94	95.47	6	6	5
	Polyester	10	98.94	95.29	98.17	4	5	4
White	Acrylic	11	97.38	92.76	80.85	3	5	5
	Cotton	2	88.45	96.70	89.60	7	2	2
	Nylon 6,6	4	97.18	98.33	91.08	2	6	3
	Polyester	7	100	100	97.99	3	3	5
Yellow	Acrylic	5	100	100	100	5	6	3
	Cotton	13	95.89	94.20	89.33	4	4	5
	Nylon 6,6	1	x	x	x	x	x	x
	Polyester	4	95.25	97.50	97.38	7	2	4
% Classification Accuracy			97.78	94.61	93.81			

Table 2.4 Confusion matrices of classification percentages based on 100 iterations of 10-fold cross-validation for low-level fusion, intermediate-level fusion, and high-level fusion on nine purple cotton fibers. Percentages equaling zero are omitted.

Actual class	Predicted class								
	CP01	CP02	CP03	CP04	CP05	CP06	CP07	CP08	CP09
Fusion type									
Low-level fusion									
CP01	100								
CP02		100							
CP03			100						
CP04				100					
CP05					100				
CP06						100			
CP07							100		
CP08								100	
CP09									100
Intermediate-level fusion									
CP01	100								
CP02		92.2						7.8	
CP03			100						
CP04				100					
CP05					100				
CP06						100			
CP07							100		
CP08								100	
CP09									100
High-level fusion									
CP01	100								
CP02		100							
CP03			100						
CP04				100					
CP05					100				
CP06						100			
CP07							100		
CP08							7.1	92.9	
CP09									100

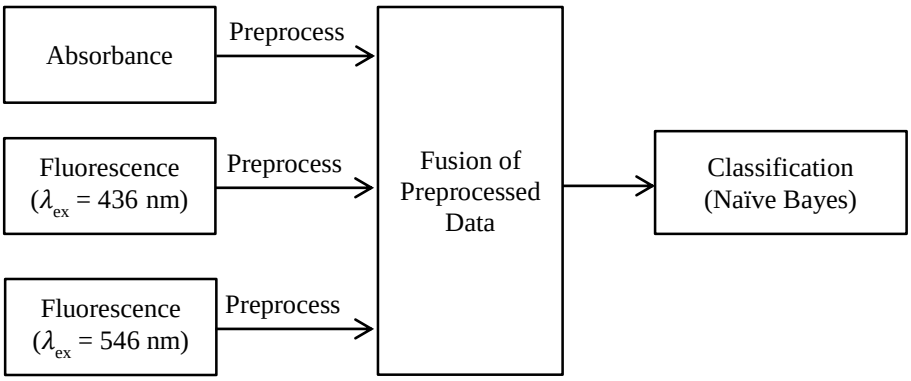

Figure 2.1 Schematic of low-level fusion process.

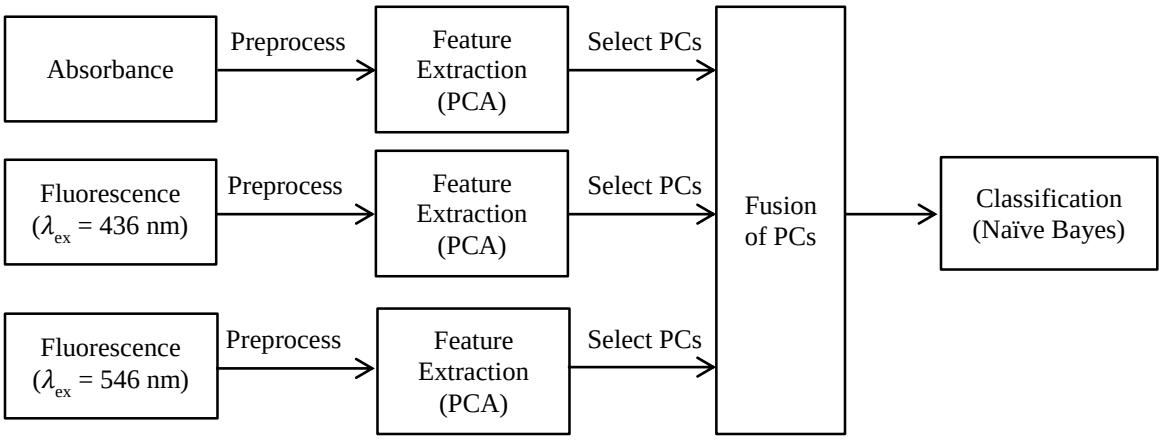

Figure 2.2 Schematic of intermediate-level fusion process.

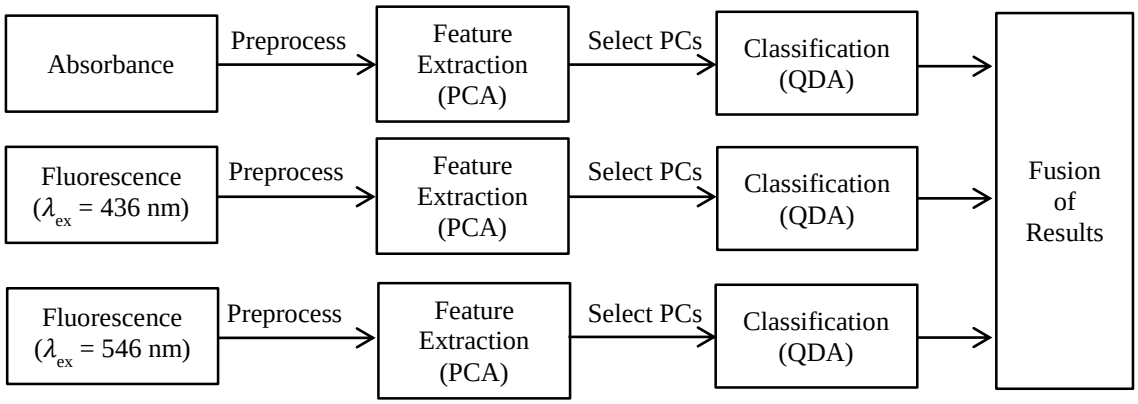

Figure 2.3 Schematic of high-level fusion process.

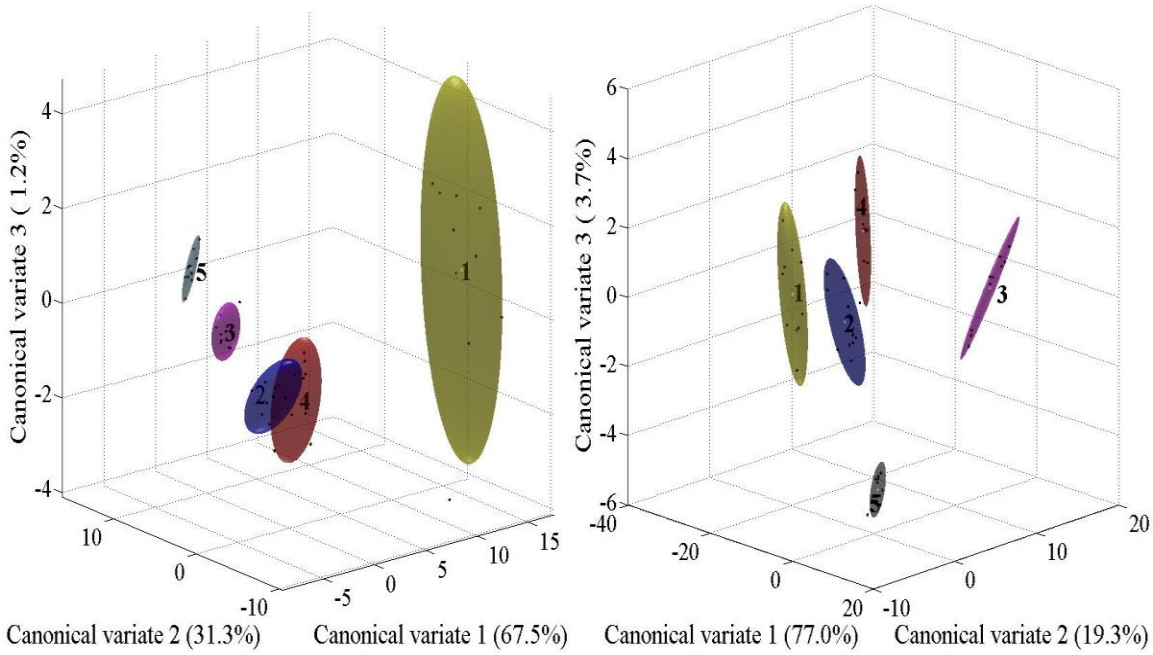

Figure 2.4 Canonical variate scores plot resulting from raw (left) and first derivative (right) spectra of five yellow acrylic fibers.

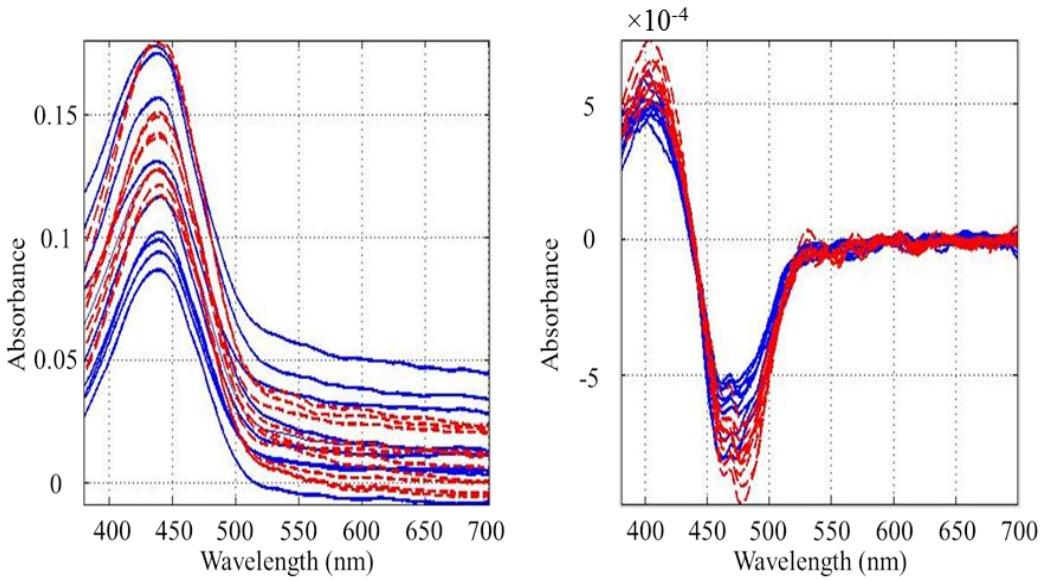

Figure 2.5 Raw (left) and first derivative (right) UV-Vis absorbance spectra for two yellow acrylic samples (10 replicate spectra for each sample).

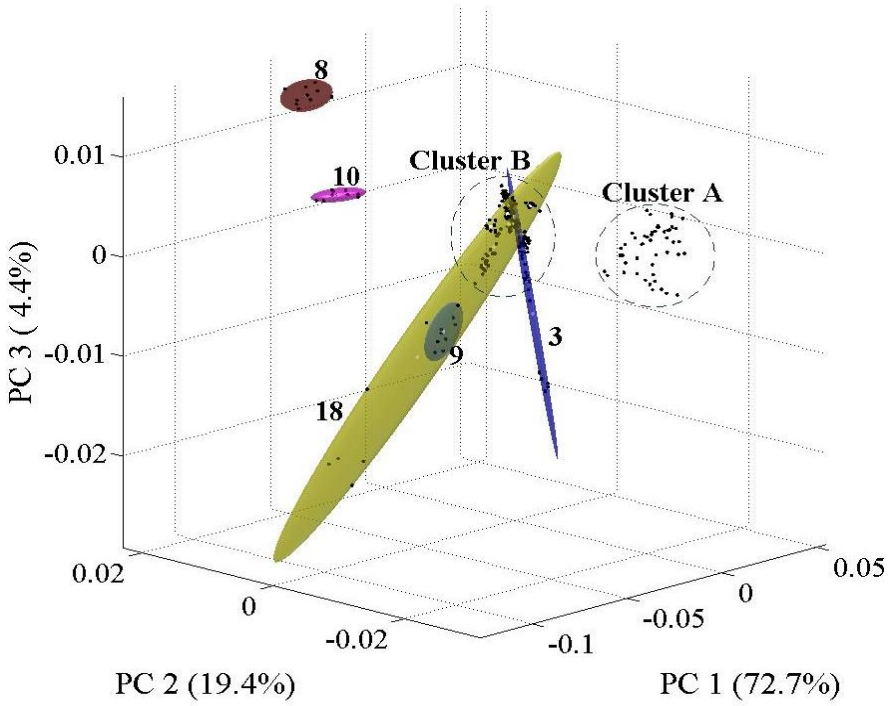

Figure 2.6 PCA scores plot of 39 brown polyester fibers with 95% elliptical confidence regions around clustered samples removed.

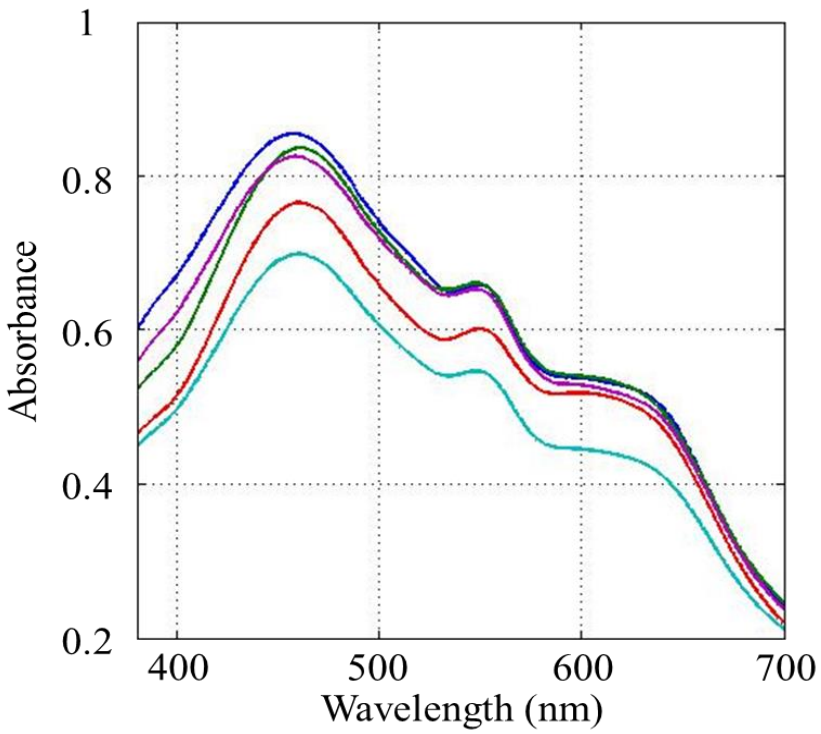

Figure 2.7 Averaged UV-Vis absorbance spectra for five brown polyester samples in Cluster B.

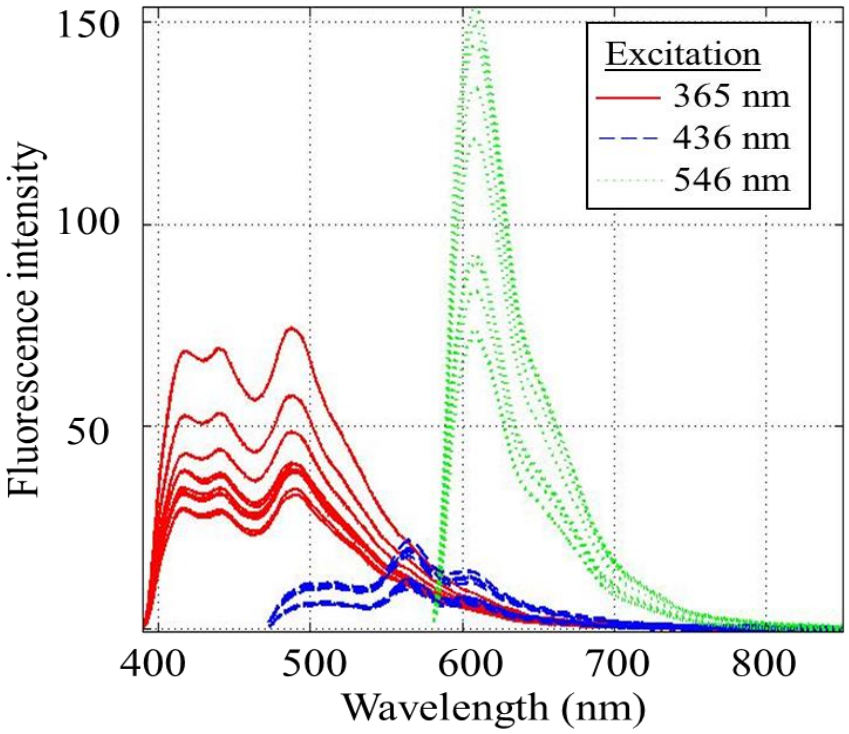

Figure 2.8 Replicate fluorescence spectra of brown polyester sample at three excitation wavelengths. Excitation at 405 nm not shown due to amount of overlap with 436 nm excitation.

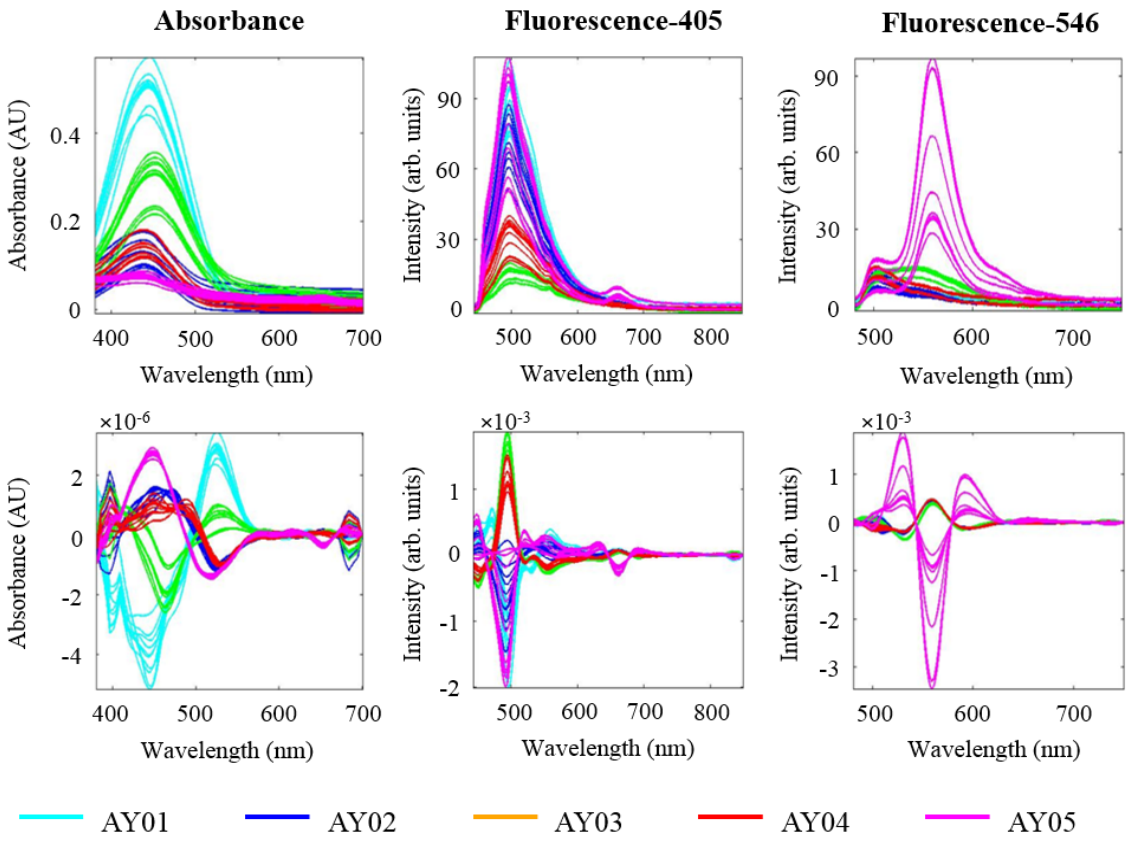

Figure 2.9 Visible absorbance and fluorescence (405 and 546 nm excitations) spectra of five yellow acrylic fibers before (top) and after (bottom) second derivative preprocessing followed by mean centering.

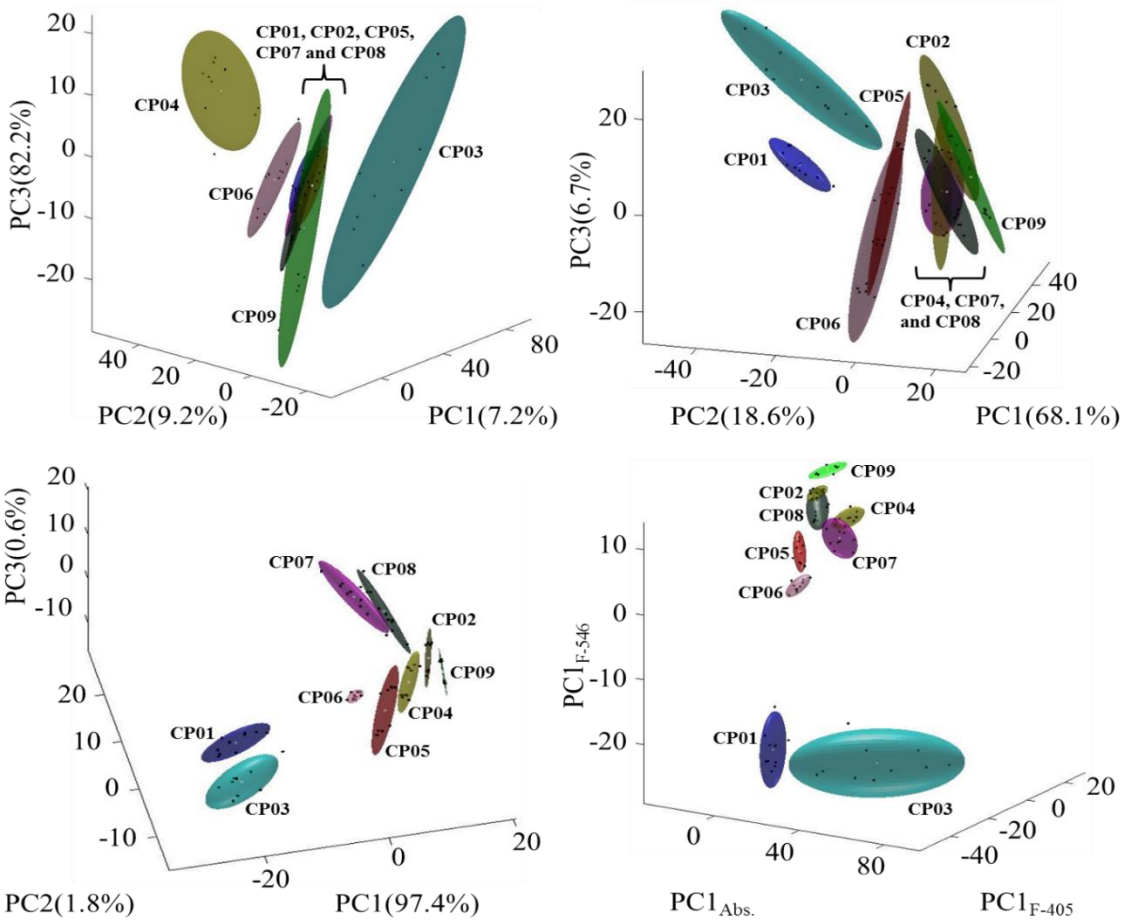

Figure 2.10 PCA scores plots resulting from absorbance (top left), fluorescence with 405 nm excitation (top right), fluorescence with 546 nm excitation (bottom left), and fusion of all three techniques (bottom right) for nine purple cotton fibers.

CHAPTER 3

MULTIVARIATE CLASSIFICATION MODEL TRANSFER OF UV-VISIBLE DATA FROM ACRYLIC FIBERS WITHOUT STANDARDS

ABSTRACT

Ultraviolet (UV)-visible microspectrophotometry has been used for decades as a means for discriminating metameric fibers in forensic casework. Recent studies have shown that multivariate classification techniques are an effective tool for characterizing such fibers. The ability to transfer multivariate classification models between laboratories could save time and resources in forensic analyses. However, issues transferring models of this type from one laboratory to another can arise as a result of differences in sample preparation, environmental conditions, and instrumental signal response.

In this study, UV-visible absorbance spectra of 12 blue acrylic fibers were examined at five separate locations including three academic research laboratories and two forensic laboratories. The data received from these facilities were analytically assessed in three manners. Multivariate classification models were initially constructed on each individual laboratory's dataset to evaluate intra-laboratory variability between samples. In a second subset of the study, discriminant analysis was performed after merging all data collected in the study. Lastly, the transferability of classification models was assessed by predicting class membership of samples analyzed at a single laboratory using models built from the spectra collected at the four remaining locations.

Principal component analysis (PCA) followed by linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), or support vector machine discriminant analysis (SMV-DA) was used to evaluate the agreement of results among the laboratories. An average classification accuracy of 93.2% was found after training the discriminant analysis models using data collected at four laboratories and using the information collected at the fifth laboratory as an external test set. For comparison, intra-laboratory studies carried out produced an average classification accuracy of 96.3%. The reduction in the discriminative abilities of the transferred models was likely due to the differences in spectral noise and peak intensities experienced between laboratories. On the whole, the errors generated by QDA were lower than those resulting from LDA and SMV-DA.

1. INTRODUCTION

Ultraviolet-visible (UV-Vis) microspectrophotometry (MSP) is an established technique for comparing metameric pairs of fibers in forensic casework.[1,2] MSP can be used following microscopy without having to remove small amounts of fibers from the microscope slides, providing a great convenience for examiners of trace evidence. A decision regarding the likelihood two fibers originated from the same source is often formed by a simple visual examination of the normalized or differentiated absorbance spectra. This process may be complicated by the fact that during the course of an investigation, numerous fibers of interest may be collected, and typical MSP protocols call for absorbance spectra to be collected at a minimum of five locations along each fiber to produce representative mean spectra and standard deviations.[3] Statistical software packages utilizing pattern-recognition techniques can be used for more robust analyses of

fibers, and are especially useful in instances where one wishes to examine a multitude of spectra simultaneously.[4]

A subdivision of the pattern recognition methodology includes multivariate classification techniques such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and support vector machine discriminant analysis (SMV-DA). These popular supervised methods of classification have been previously combined with optical spectroscopy techniques to study various forensic analytes of interest such as gunshot residue,[5] bodily fluids,[6,7] drugs[8,9] and ink.[10-12] The objective of LDA, introduced by Fisher[13] in 1936 (multi-class version developed by Rao[14] in 1948), is to find a linear projection which best separates the classes of objects studied. LDA calculations assume the data is normally distributed, and the covariance (a descriptive measure of the association between variables) matrices are equal. In instances of unequal covariance, the classes are best separated using quadratic discriminant functions generated with QDA. Originally developed for binary classification problems, the more robust SVM methods are able to handle both linear and nonlinear classification tasks through the use of kernels (e.g., polynomial or radial basis function). The kernel functions are used to reallocate the observations into a higher dimensional space. Once a hyperplane maximizing the distance between classes is constructed in high dimensional space, the function is returned to the original space appearing as a nonlinear boundary. The reader is referred to the numerous resources available covering supervised learning techniques for more in-depth discussions covering LDA, QDA, and SMV-DA.[15-18]

Ideally, the classification models built using LDA, QDA, or SMV-DA could be transferred from one instrument to another. This would be useful for routine

discriminations of textile fibers and save forensic laboratories the time and cost required to construct a new model each time new samples are introduced. Instrumental transfer of classification models, however, can be challenging due to differences in sample preparation, instrumental response, and environmental conditions experienced between laboratories. Although the issue of transferability of calibration models has been studied extensively,[19-23] much less attention has been focused toward transferring classification models.[24]

This research has three objectives: (a) to conduct interlaboratory experiments for the evaluation of decision making in forensic fiber examinations by MSP; (b) to evaluate intra-laboratory variability, inter-laboratory agreement, and error rate performance in designed experiments; and (c) to investigate the application of multivariate statistical measures for comparisons of UV-Vis spectra of fibers. To address these objectives, the agreement between classification accuracies among five laboratories and the transfer of multivariate classification models between the laboratories was evaluated using LDA, QDA, and SMV-DA for discrimination of spectra taken from a set of twelve blue acrylic fibers.

2. EXPERIMENTAL

2.1 Materials

Acrylic samples were donated from commercial sources in the southeastern United States. Procedures from the Scientific Working Group on Materials Analysis Fiber Subgroup (SWGMAT) were followed.[3] In the five laboratories, individual fibers were cut using a razor blade and positioned on glass or quartz microscope slides using micro-tweezers. The fibers from each exemplar were removed and mounted on the slides

with coverslips using either spectral grade Permount® mounting media (Fischer Scientific, Fairlawn, NJ) or glycerol. To assess apparent variations in dyeing depth, ten replicate spectra of each fiber were obtained at different locations along the length of the fiber.

2.2 Instrumentation

Spectral measurements in this study were made using slightly different instruments from the same company, and with slightly different instrumental settings.

Laboratory one obtained spectra with a CRAIC QDI 302 UV-Vis microspectrophotometer with a Carl Zeiss (Jena, Germany) Axioscope A1 and a 1.3 megapixel digital imaging system with CCD cooling. The spectral range was 400 to 800 nm with a resolution of 0.5 nm.

Analysis in laboratory two was conducted with a CRAIC 380 Perfect Vision™ UV/Vis microspectrophotometer with a Carl Zeiss (Thornwood, NY) Axioscope A1. Fibers were mounted on glass slides using Permount® (Fisher Scientific, Fair Lawn, NJ) and a coverslip. The spectral data acquisition range was 400 to 800 nm with a resolution of 0.5 nm.

Spectra in laboratory three were taken using a CRAIC Quantum Detection Instrument (QDI) 1000 microspectrophotometer operated in transmission mode with a xenon light source, a Carl Zeiss Axioscope A1 microscope, and a megapixel cooled charge coupled detector. A 15× collecting objective was used to focus an area within the diameter of the fibers. UV-Vis spectra of textile fibers were produced by collecting an

average of 100 scans over the spectral range of 200 to 850 nm at a 10 nm bandwidth and 4 ms integration time.

Spectra were collected in laboratory four using a CRAIC 20/20 Perfect Vision™ MSP version 1.1.0. The instrument was operated in transmission mode using a 40× collection objective with a detector integration time of ~8.6 ms. The spectrum of each fiber was calculated by averaging 10 scans over a 250 to 820 nm spectral region.

Laboratory five acquired spectra using a CRAIC 2000 microspectrophotometer (CRAIC Technologies, San Dimas, CA) in transmitted light mode at 150× magnification. Calibration of the spectrometer with NIST traceable standards was performed before each use, along with Köhler illumination for the microscope. Autoset optimization, a dark scan, and a reference scan were employed prior to each sample scan. MSP data was collected over the wavelength range of 350 to 800 nm.

2.3 Data Analysis

Analysis of the data obtained from the five laboratories was performed using MATLAB version 8.3 (The MathWorks, Natick, MA). The PLS_Toolbox version 7.8.2 (Eigenvector Research, Wenatchee, WA) was the source of the preprocessing routines used in this study. LDA and QDA classifiers were generated using the "fitcdiscr" routine located in the statistics toolbox in MATLAB. The SMV-DA algorithms used in this study can be found as part of the Library for Support Vector Machines (LIBSVM) developed by Chang and Lin.[25] For SMV-DA, multiple kernels and parameters were tried, and those resulting in the best average classification accuracies were utilized. The discussed results in the intra-laboratory study were collected using the C-SMV-DA algorithm with the cost parameter set to one. Combined laboratory studies were carried out using the nu-SVM

algorithm with a nu-parameter of 0.1. For intra-laboratory studies, a 2nd order polynomial kernel was used, whereas a linear kernel was used for fused data. Using LDA, QDA, and SMV-DA, three distinct data analysis procedures were performed and are described below.

2.3.1 Intra-laboratory evaluation

The datasets received from all laboratories, each containing 120 spectra (10 replicate spectra for each of 12 samples), were treated separately using the same preprocessing techniques. The spectra were first truncated to a wavelength range of 400 to 800 nm. Each sample spectrum was then subjected to Savitzky-Golay[26,27] smoothing by fitting a 2nd order polynomial to a moving window of 21 points. Following smoothing, a 2nd order weighted least squares baseline correction was performed. Finally, the data was mean-centered, a preprocessing technique typically recommended when performing PCA. In mean-centering, the average of each column is calculated and subtracted from all of the elements in that column.

After preprocessing, principal component analysis (PCA) was used to reduce the dimensionality of the data. PCA is used to calculate new uncorrelated variables called principal components (PCs), which are linear combinations of the original spectral variables.[28] Increasing amounts of variation in data are obtained with successive PCs. Contained within each PC are the scores (the projections of the spectra) and the loadings (the weights of the original variables). A scree plot, which plots the amount of variance captured by each PC, was used to select those PCs for removal which appear to capture noise rather than variations due to group differences. The retained PC scores were then subjected to LDA, QDA, and SMV-DA.

The predictive performances of the discriminant analysis models created in this study were determined by internal validation using the stratified 10-fold method. This method of cross-validation was chosen to assess our inter-laboratory models prediction ability, because it is often a good compromise between bias and variance.[15] In stratified 10-fold cross-validation, the data is partitioned into 10 nearly equal sized parts with approximately the same number of samples per class (i.e., per fiber). The discriminant functions are then calculated using the information from all but one of these subsets, and the left-out portion is used to test the classifier. This process is repeated until each subset of samples has been used for testing. The classification accuracy (the number of correctly classified samples divided by the total number of samples examined) for this part of the study was calculated after 100 iterations of stratified 10-fold cross-validation.

2.3.2 Discrimination of combined datasets

The methodology used to evaluate classification models built on merged data from all five laboratories is shown in Figure 3.1. The data received from each laboratory was fitted to the same wavelength axis by using a spline interpolation method that creates interpolated values by fitting a 3rd order polynomial to the neighboring points.[29] Feature values in all three data sets were interpolated in the region of 400 to 800 nm due to its commonality in all data received from each institution. The splined datasets were given a new wavelength spacing of 0.3407 nm, as this was the lowest interval used for data collection at the five laboratories. Following interpolation of the feature axis for each individual laboratory's data, the datasets were combined into one matrix for further analysis.

Prior to any further processing, approximately 20% of the data was partitioned out in a stratified manner to generate an internal test set containing samples from all classes and all laboratories. Thus, all samples (fiber spectra) in the internal test set will also have had replicate samples in the training set collected from that same laboratory. The training set contained 60% of the original amount of data. The remaining 20% was used as a validation set for the selection of the number of significant principal components to be used for testing. The selection of which samples would comprise the training set was done using the Kennard-Stone algorithm.[30] The goal of this method is to select the samples which constitute the boundaries in each class and are the most representative.

All spectra in the training and test sets were again subjected to the smoothing and baseline correction techniques described in the previous section. When dealing with combined datasets, the data was also treated using autoscaling and a standard normal variate (SNV) transformation. Autoscaling is performed by subtracting the mean spectral intensity $\bar{X}$ at wavelength *j* in spectrum *i* and dividing by the standard deviation *s* of all values at that wavelength.

$$X_{ij,auto} = \frac{X_{ij} - \overline{X_j}}{s_j} \quad (1)$$

The result of autoscaling is a matrix in which each column has a mean of zero and a standard deviation of one. The purpose of SNV is to remove the differences in the slope of the spectra which may result from scattering.[31] The transformation is applied to each individual spectrum using the following equation:

$$\chi_{ij,SNV} = \frac{X_{ij} - \overline{X_i}}{\sqrt{\frac{\sum_{j=1}^{f}(X_{ij} - \overline{X_i})^2}{n-1}}} \quad (2)$$

where $X_{ij,SNV}$ is the SNV corrected absorbance, X_{ij} is the original absorbance for the same element, $\overline{X_i}$ is the mean, and *n* is the number of wavelength variables. To preserve the integrity of the test set, autoscaling and SNV were performed on the training set, and the parameters used, such as the column means, spectral means, and standard deviations were then applied to the preprocessing of the test and validation sets.

Following preprocessing, PCA was used to reduce the dimensionality of the training set, and the samples in the validation set were projected into the resulting PC space. After repeating this process using anywhere from one to ten PCs, the number of significant PCs to use in the final training model before testing was selected based on the number of misclassifications of samples in the validation set. The final classification accuracy and standard deviation was calculated as the average of 100 iterations of this process.

2.3.3 Transfer of classification models between instruments

This part of the study only differs from the portion of the experiment mentioned previously in the manner in which the training and test sets were selected. The training set was built using splined spectra from only four of the five laboratories. The interpolated spectra from the remaining laboratory were used as an external test set. Following preprocessing, five separate classification models of each type were built, allowing datasets from each of the laboratories to be used as the external test set. Because the test sets were not selected at random, there was no need to iterate these processes.

3. RESULTS AND DISCUSSION

3.1 UV-visible spectra

Twelve blue acrylic fiber samples were analyzed at five separate laboratories using UV-Vis MSP and are shown in Figure 3.2. In general, each sample was very similar in color to at least one of the other remaining samples in the dataset. One exception is fiber 114 which is a visually distinguishable shade of blue. As expected, sample 114 was also one of the least misclassified samples throughout the study. Averaged UV-Vis spectra for each of the 12 fibers obtained at all five laboratories are shown in Figure 3.3. The majority of the fibers in this study can be discriminated upon a simple visual examination of their absorbance spectra based on the locations and sizes of the peaks, troughs, and shoulders. Using this methodology, an initial survey of those spectra which could be difficult to discriminate was carried out. For example, absorbance spectra of samples 086, 098, 112, and 145 have peak locations at approximately 650 nm. A shoulder located between 600 and 620 nm is also characteristic of these samples, though it may be possible in some instances to discriminate the samples based on the intensity of the shoulder. The similarities in the absorbance spectra of all four samples are seemingly due to these fibers containing the same blue dye, CI Blue 6. In addition, samples 086 and 112 contain all three of the same cationic dyes, as indicated in Table 3.1. Blue acrylic sample fibers 087 and 088 also appear to share similar spectra, as do samples 113 and 114. Finally, it should be noted that the lowest absorbing and most noisy spectra are associated with fiber sample 092. Because sample 092 was the only fiber dyed using the cationic dye, CI Blue 60, it is believed that this dye is responsible for the lack in color strength displayed by the fiber.

3.2 Intra-laboratory evaluation

Dimensionality reduction of the spectral data is often a necessary step before performing discriminant analysis. To examine the discriminative ability of each laboratory separately, PCA was used prior to LDA, QDA, and SMV-DA. By examining the percentage of variation in the spectra captured by each PC, it was determined that either four or five PCs, depending on the dataset, would be sufficient to build the discriminant analysis models. The PCs used represent 99.83%, 99.94%, 99.84%, 98.35%, and 99.49% of the data in laboratories one through five, respectively. Most of the information used to discriminate the 12 blue acrylic fibers is contained within the first two PCs as shown in Figure 3.4. The ellipses around groups of spectra represent, with 95% confidence, distances that are statistically equidistant from the group mean.

The proximity of the groups of similar spectra mentioned in the previous section is also evident in Figure 3.4. Of particular significance is the amount of overlapping that occurs between samples 086 and 112. As seen in Table 3.2, these two samples could not be distinguished (accuracies of 80% and 58.2% for samples 086 and 112, respectively) through the use of LDA on spectral data from laboratory 3. Classification accuracies of 100% were obtained on these samples using QDA, however. The decision boundaries generated by LDA and QDA for 086 and 112 are shown in Figure 4 in the coordinates of PCs three and four (the PCs which show the most discrimination between these two classes). The case of equal covariance between classes, which is assumed by LDA, does not hold true in the laboratory three dataset. The result of this misassumption is a poor decision boundary in a case where the classes are clearly separable. Because unequal covariance is assumed in QDA, an optimal boundary separating 086 and 112 is drawn. A

reexamination of Figure 3.4 shows the samples responsible for the faulting of LDA in this situation. The variance in sample 099 is much greater than those of the remaining eleven samples. By not including the covariance for 099 when calculating the pooled-in covariance for LDA, a better separation is achieved as shown in Figure 3.5. An even better fit could be made by removing 098, the group with the second largest variance.

The results of the intra-laboratory study suggest that the classes of fibers are best separated using classification techniques useful for solving non-linear classification problems. The best results were seen with QDA which gave a correct classification rate of 96.3%. With a classification accuracy of 95.9%, SMV-DA using the quadratic polynomial kernel was only slightly lower. Although, the lowest accuracies were obtained using LDA at 94.8%, the technique gave the most consistent results across all five laboratories. The measured standard deviations between laboratories were ±2.0%, ±2.9%, and ±4.1% for LDA, QDA, and SMV-DA, respectively. It was clear after using SMV-DA on all sets of data in this study that when used correctly SMV-DA can be a very powerful discriminatory tool. However, the results varied to a great extent depending on the parameters used to train the support vectors. It is recommended that these parameters be optimized using a "grid search" for every set of data that is to be examined.[25]

3.3 Discrimination of combined datasets

The purpose of the second portion of the study was to determine how well the fibers could be discriminated by including data obtained at all five laboratories in the training and test sets. Because the data received from each institution was collected at differing wavelengths and wavelength intervals, a spline interpolation was used to

generate three new sets of data with the same wavelength region (400 to 800 nm) and intervals (0.3407 nm). It should be noted that the splined datasets were compared to the original datasets by once again performing intra-laboratory discriminant analyses on the splined datasets. No significant differences were found between the intra-laboratory results of the splined data and those obtained from the original data. After all sets of data had been placed on the same wavelength axis, preprocessing was carried out in the form of smoothing and baseline correction. Column (variable) scaling was performed to eliminate inter-laboratory variations in signal intensities. Finally, row scaling by SNV transformation is used to eliminate the changes in sample-to-sample peak locations.

After splining and fusing five datasets from each of the laboratories, the resulting matrix included 600 total spectra. Because a much greater amount of data is involved in the interlaboratory studies (50 replicate spectra for each fiber as opposed to 10 in the intra-laboratory studies), the decision was made to use training, validation, and test sets. By initially removing 120 spectra by stratified partitioning for later use as a test set, it was ensured that the same number of samples would be tested on in the analysis of fused data as well as when attempting to transfer the model to one instrument. To make sure the samples which are most representative of the data as a whole are included in the training set, the Kennard-Stone algorithm was used. Conversely, this means that the validation set is likely to be overoptimistic since the samples selected for it are known to fall within those of the training set. Nevertheless, the validation set in this study is only used for the purpose of model selection, and the inherent bias is not likely to influence the number of components used to train the final model.

Classification models were built using the LDA, QDA, and SMV-DA algorithms on the training set containing 360 (72 from each of the five laboratories with 30 total samples per class) pretreated spectra. The optimal classification models chosen from the validation results were then applied to the test sets where each spectrum was assigned to one of the 12 classes. Average classification accuracies after 100 iterations of this process in its entirety were 94.6±1.9%, 96.1±1.3%, and 91.9±2.1% for LDA, QDA, and SMV-DA, respectively. Table 3.3 shows the extent to which the 12 fibers in this study were misclassified using the "best" method for this set of data, QDA. There was once again a significant inability of the model to discriminate between samples 086 and 112, two fibers containing the same three cationic dyes. This scheme, however, was very useful in achieving its objective of obtaining similar classification accuracies to those of the intra-laboratory studies.

3.4 Transfer of classification models between instruments

The goal of the final portion of the study was to determine how well blue acrylic fibers could be discriminated using models built on data from instruments separate from the one used as the test set. To do this, discriminant models were built using data from four laboratories, and the data from one remaining laboratory was used as an external test set. This process was repeated four more times allowing each laboratory's data a chance to be the test set, while the other four laboratories comprised the training set. The results of each trial are displayed in Table 3.4. As expected, the accuracy of predicting samples from a laboratory in which no samples had been included in the training set was lower across the board compared to the other processes in this study. The average classification accuracies obtained after attempting to transfer LDA, QDA, and SMV-DA models to an

external laboratory were 92.0±3.2%, 93.2±5.3%, and 88.7±6.0%, respectively. The poorest performances of classification models occurred when trying to predict samples from laboratory four based on the data from laboratories one, two, three, and five. The poor signal-to-noise ratio of the spectra from sample 092 led to multiple misclassifications with that group and was a significant contributor to the lower classification accuracy obtained when using laboratory four as a test set. The same sample was also responsible for multiple misclassifications in both the intra-laboratory and combined dataset studies.

The choice of preprocessing techniques and the order in which they are performed is critical to set all laboratories on as close to the same scale as possible. Figure 3.6 shows how the ability to transfer classification models is influenced by each preprocessing step. The five step preprocessing technique of spline interpolation, smoothing, baseline correction, autoscaling, and standard normal variate transform worked the best on four out of the five laboratories tested with the one exception being laboratory five. QDA model testing on laboratories four and five saw the most improvement with respective gains in classification ability of 17.3% and 39.2%. Such a sizeable increase in discrimination was seemingly due to the noticeable scaling differences in Figure 3.3. The spectra obtained from laboratory four gave, in many instances, significantly lower absorbance values as compared to those in laboratories one, two, and three, and contained significantly more noise. Laboratory five gave noticeably low absorbing spectra as well. As a consequence, some peak shapes (e.g., samples 086 and 112) are slightly different when compared to those same samples collected at the other four laboratories. The preprocessing used in this study does not correct for differences in peak shape, and

therefore would explain why no gain in discrimination power was gained using these preprocessing steps. Fiber spectra from laboratories one, two, and three appear to show a high degree of correlation between datasets. It is therefore not surprising that scaling had less of an impact on these sets of data, since they were not only closest to the mean of the data as a whole, but there were also existing similar samples in the training set.

4. CONCLUSION

UV-Vis microspectrophotometry is widely accepted as a valid analytical approach for characterization of trace evidence fibers. The reliability of a spectrophotometer in a specific laboratory is dependent on whether it has functional capability and performance that meet the requirements of the task. Qualification of a spectrophotometer for operation typically involves testing for wavelength accuracy and reproducibility, photometric accuracy, presence of stray light, baseline flatness, stray light levels, stability, and linearity. These issues from the viewpoint of quality control all involve defining performance characteristics that are targeted to insuring reliability, and comparability of measurements among laboratories. The laboratories involved in this study were all facilities with a well-documented history of using MSP for fiber characterization.

Transfer of calibration models, particularly in near infrared spectroscopy, has been a topic of continuing discussion in the literature, with the focus on methods involving robust calibration and spectral preprocessing to correct for inter-laboratory variability (e.g., reference 23). In the present study, classification models based on QDA in separate laboratories produced an average classification accuracy of 96.3%, an outcome that is indicative of good operational control of methodology and practice within each laboratory. Little change was observed in discriminating ability when laboratory

data sets were fused together and used to predict classification over all samples. Permutations were also performed to combine data from four of the laboratories for classification of the same samples from the fifth laboratory used as an external test set. The average classification accuracy of the best method used, QDA, was slightly lower at 93.1%. Upon detailed visual inspection of the spectra, this observed loss in classification accuracy was attributed to noisy spectra, and differences in intensity or peak shape. Alternative scaling techniques should continue to be examined for the purpose of eliminating these factors which cause accuracy reduction in transferred classification models.

ACKNOWLEDGMENTS

This project was supported by Award No. 2010-DN-BX-K245 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Portions of this work were presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy (Pittcon) on March 11, 2015 in New Orleans, LA. The co-authorship of John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan is also acknowledged. The authors are grateful for the participation in this study from the South Carolina Law Enforcement Division and the Boston Crime Lab. The opinions, findings, and conclusions or recommendations expressed in this work are those of the author(s) and do not necessarily reflect those of the Department of Justice.

REFERENCES

1. Adolf, P.; Dunlop, J. Microspectrophotometry/Colour Measurement. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K. 1999, pp. 251-289.

2. Gaudette, B. The Forensic Aspects of Textile Fiber Examination, In *Forensic Science Handbook*, Vol. II.; Saferstein, R., Ed.; Prentice Hall: New Jersey, 1988.

3. SWGMAT Forensic Fiber Examination Guidelines. http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf (accessed May 5, 2014).

4. Morgan, S.; Bartick, E. Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In *Forensic Analysis on the Cutting Edge*; Blackledge, R., Ed.; John Wiley & Sons: New Jersey, 2007, pp. 333-374.

5. Bueno, J.; Sikirzhytski, V.; Lednev, I. Raman Spectroscopic Analysis of Gunshot Residue Offering Great Potential for Caliber Differentiation. *Anal. Chem.* **2012**, 84, 4334-4339.

6. Li, B.; Beveridge, P.; O'Hare, W.; Islam, M. The Estimation of the Age of a Blood Stain Using Reflectance Spectroscopy with a Microspectrophotometer, Spectral Pre-processing and Linear Discriminant Analysis. *Forensic Sci. Int.* **2011**, 212, 198-204.

7. Sikirzhytski, V.; Sikirzhytski, A.; Lednev, I. Advanced Statistical Analysis of Raman Spectroscopic Data for the Identification of Body Fluid Traces: Semen and Blood Mixtures. *Forensic Sci. Int.* **2012**, 222, 259-265.

8. Broseus, J.; Vallat, M.; Esseiva, P. Multi-class Differentiation of Cannabis Seedlings in a Forensic Context. *Chemom. Intell. Lab. Syst.* **2011**, 107, 343-350.

9. Dingari, N.; Barman, I.; Myakalwar, A.; Tewari, S.; Gundawar, M. Incorporation of Support Vector Machines in the LIBS Toolbox for Sensitive and Robust Classification Amidst Unexpected Sample and System Variability. *Anal. Chem.* **2012**, 84, 2686-2694.

10. Kher, A.; Mulholland, M.; Green, E.; Reedy, B. Forensic Classification of Ballpoint Pen Inks Using High Performance Liquid Chromatography and Infrared Spectroscopy with Principal Components Analysis and Linear Discriminant Analysis. *Vib. Spectrosc.* **2006**, 270-277.

11. Hoehse, M.; Paul, A.; Gornushkin, I.; Panne, U. Multivariate Classification of Pigments and Inks Using Combined Raman Spectroscopy and LIBS. *Anal. Bioanal. Chem.* **2012**, 402, 1443-1450.

12. Silva, C.; Borba, F.; Pimentel, M.; Pontes, M.; Honorato, R.; Pasquini, C. Classification of Blue Pen Ink Using Infrared Spectroscopy and Linear Discriminant Analysis. *Microchem. J.* **2013**, 122-127.

13. Fisher, R. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugenics.* **1936**, 179-188.

14. Rao, R. The Utilization of Multiple Measurements in Problems of Biological Classification. *J. R. Stat. Soc. Series B.* **1948**, 10, 159-203.

15. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Series in Statistics: Springer: New York, 2001.

16. Rencher, A. *Methods of Multivariate Analysis*, 2nd ed.; Wiley: New York, 2002.

17. Krzanowski, W. *Principles of Multivariate Analysis*; Oxford Statistical Science Series 3; Clarendon Press: Oxford, U.K., 1988.

18. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, U.K., 2000.

19. Blank, T.; Sum, S.; Brown, S. Transfer of Near-infrared Multivariate Calibrations without Standards. *Anal. Chem.* **1996**, 68, 2987-2995.

20. Dreassi, E.; Caramelli, G.; Perruccio, P.; Corti, P. Transfer of Calibration in Near-Infrared Reflectance Spectrometry. *Analyst.* **1998**, 123, 1259-1264.

21. Fearn, T.; Standardisation and Calibration Transfer for Near-infrared Instruments: a Review. *J Near Infrared Spectrosc.* **2001**, 9, 229-244.

22. Feudale, R.; Woody, N.; Tan, H.; Myles, A.; Brown, S.; Ferre, J. Transfer of Multivariate Calibration Models: a Review. *Chemom. Intel. Lab. Syst.* **2002**, 64, 181-192.

23. Sahni, N.; Isaksson, T.; Naes, T. Comparison of Methods for Transfer of Calibration Models in Near-infrared Spectroscopy: a Case Study Based on Correcting Path Length Differences Using Fiber-Optic Transmittance Probes in In-line Near-infrared Spectroscopy. *Appl. Spectrosc.* **2005**, 59, 487-495.

24. Myles, A.; Zimmerman, T.; Brown, S. Transfer of Multivariate Classification Models between Laboratory and Process Near-infrared Spectrometers for the Discrimination of Green Arabica and Robusta Coffee Beans. *Appl. Spectrosc.* **2006**, 60, 1198-1203.

25. Chang, C.; Lin, C. LIBSVM: a Library for Support Vector Machines. *ACM TIST*, **2011**, 2, 1-27.

26. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.* **1964**, 36, 1627-1639.

27. Steiner, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Squares Procedure. *Anal. Chem.* **1972**, 44, 1906-1909.

28. Gemperline, P. *Principal Component Analysis*, 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006, pp 69-104.

29. Boor, C. *A Practical Guide to Splines*; Springer-Verlag: Germany, 1978.

30. Kennard, R.; Stone, L. Computer Aided Design of Experiments, *Technometrics* **1969**, 11, 137-148.

31. Barnes, R.; Dhanoa, M.; and Lister, S. Standard Normal Variate Transformation and De-trending of Near-infrared Diffuse Reflectance Spectra. *Appl. Spectrsoc.* **1989**, 43, 772-777.

Table 3.1 Cationic dye composition 12 blue acrylic fibers examined. 'Y' indicates presence of dye.

	Cationic dye									
Fiber	Blue	Blue	Blue	Blue	Red	Red	Red	Yellow	Yellow	Yellow
086	Y				Y				Y	
087		Y					Y		Y	Y
088		Y					Y		Y	
091		Y				Y		Y		
092			Y			Y			Y	
095				Y		Y			Y	
098	Y			Y						
099				Y			Y		Y	
112	Y				Y				Y	
113		Y				Y			Y	
114		Y			Y				Y	
145	Y						Y		Y	

Table 3.2 Comparison of classification accuracies between laboratories by method.

				Percent of correctly classified spectra by sample											
Method	Lab.	Acc. (%)	SD	086	087	088	091	092	095	098	099	112	113	114	145
LDA	1	95.9	0.40	89.2	100	100	91.9	100	100	100	100	70	100	100	100
	2	97.3	0.53	80	100	100	100	100	100	100	100	91.2	100	100	100
	3	94.0	0.31	80	100	100	100	100	100	100	90	58.2	100	100	100
	4	92.0	0.83	88.8	100	100	61.6	75.6	91.4	100	100	100	100	100	100
	5	94.9	0.52	70	98.9	100	100	100	98.3	100	100	81.5	90.1	100	100
QDA	1	99.2	0.43	100	100	100	90	100	100	100	100	100	100	100	100
	2	95.3	0.51	87.6	100	100	100	100	100	100	100	55.8	100	100	100
	3	98.2	0.30	100	100	100	100	88.3	100	100	100	100	90	100	100
	4	91.8	0.89	100	95.4	100	72.8	81	79.8	100	100	81.9	90.1	100	100
	5	97.2	0.56	92.5	100	100	100	100	100	100	100	73.8	100	100	100
SMV-DA	1	99.2	0.41	100	100	100	90.9	100	100	100	100	100	100	98.9	100
	2	98.3	0.26	89.9	100	100	100	100	100	100	99.9	90.3	100	100	100
	3	98.3	0.46	99.3	100	100	100	100	100	90.9	90	99.3	100	100	100
	4	88.8	1.02	90.4	99.5	90	33.7	91.3	79.6	100	100	100	100	90.2	98.8
	5	94.9	1.00	70	98.9	100	100	100	98.3	100	100	81.5	90.1	100	100

Table 3.3 Confusion matrix resulting from QDA on test set composed of combined laboratory data. Percentages of correctly classified spectra are in bold. Percentages equaling zero are omitted.

	Actual class											
Predicted class	086	087	088	091	092	095	098	099	112	113	114	145
086	**85.4**								14.6			
087		**100**										
088		0.1	**99.9**									
091				**96**	0.3	3.4				0.3		
092				2.3	**92.3**	5.4						
095				2.2	2.8	**95**						
098							**100**					
099								**100**				
112	15.8								**84.2**			
113						0.2				**99.8**		
114											**100**	
145												**100**

Table 3.4 Classification accuracies resulting from LDA, QDA, and SMV-DA models trained using datasets from four laboratories and tested on a fifth laboratory's dataset.

Training lab.	Testing lab.	LDA (%)	QDA (%)	SMV-DA (%)
2,3,4,5	1	94.2	97.5	87.5
1,3,4,5	2	93.3	95.8	86.7
1,2,4,5	3	95.0	97.5	99.2
1,2,3,5	4	87.5	86.7	85.0
1,2,3,4	5	90.0	88.3	85.0

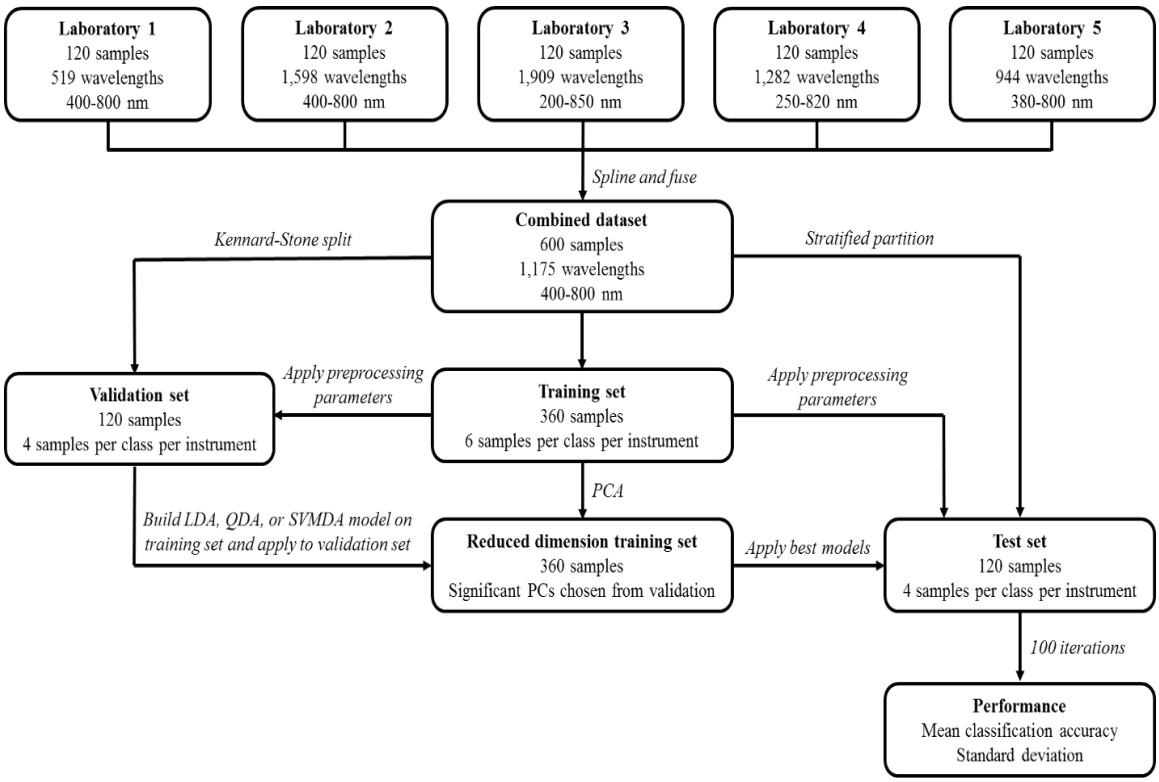

Figure 3.1 Schematic of the methodology used for the classification of combined laboratory datasets.

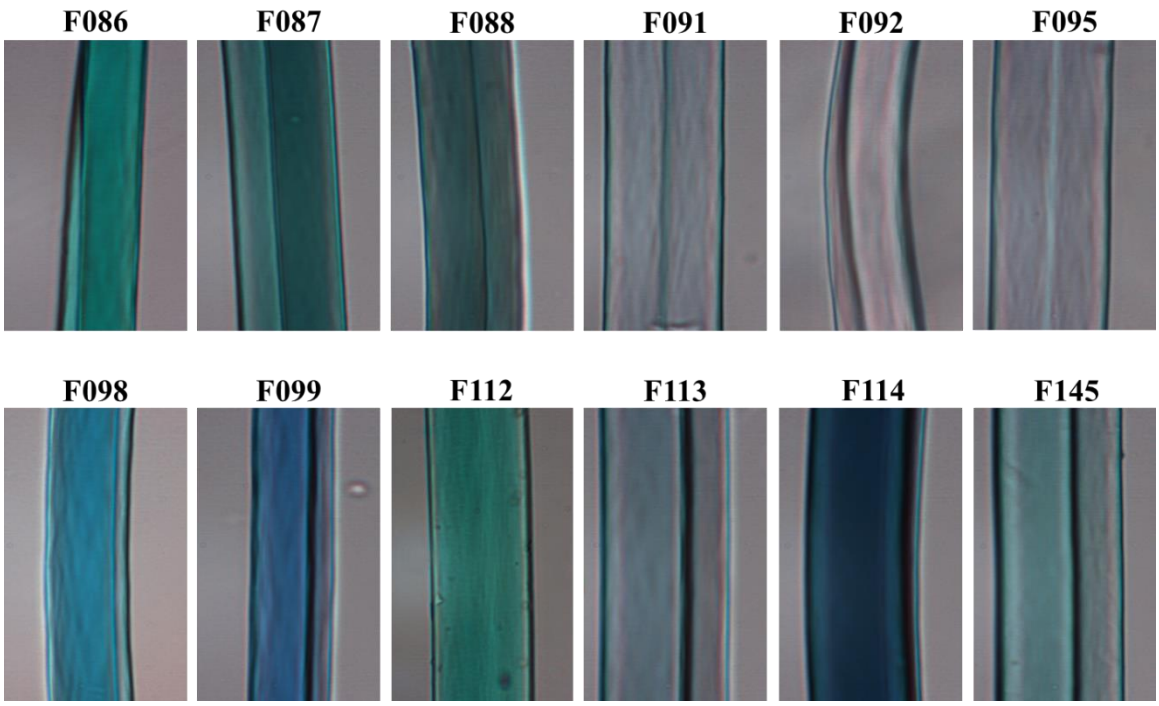

Figure 3.2 Microscopic images of 12 blue acrylic fibers under 40× magnification.

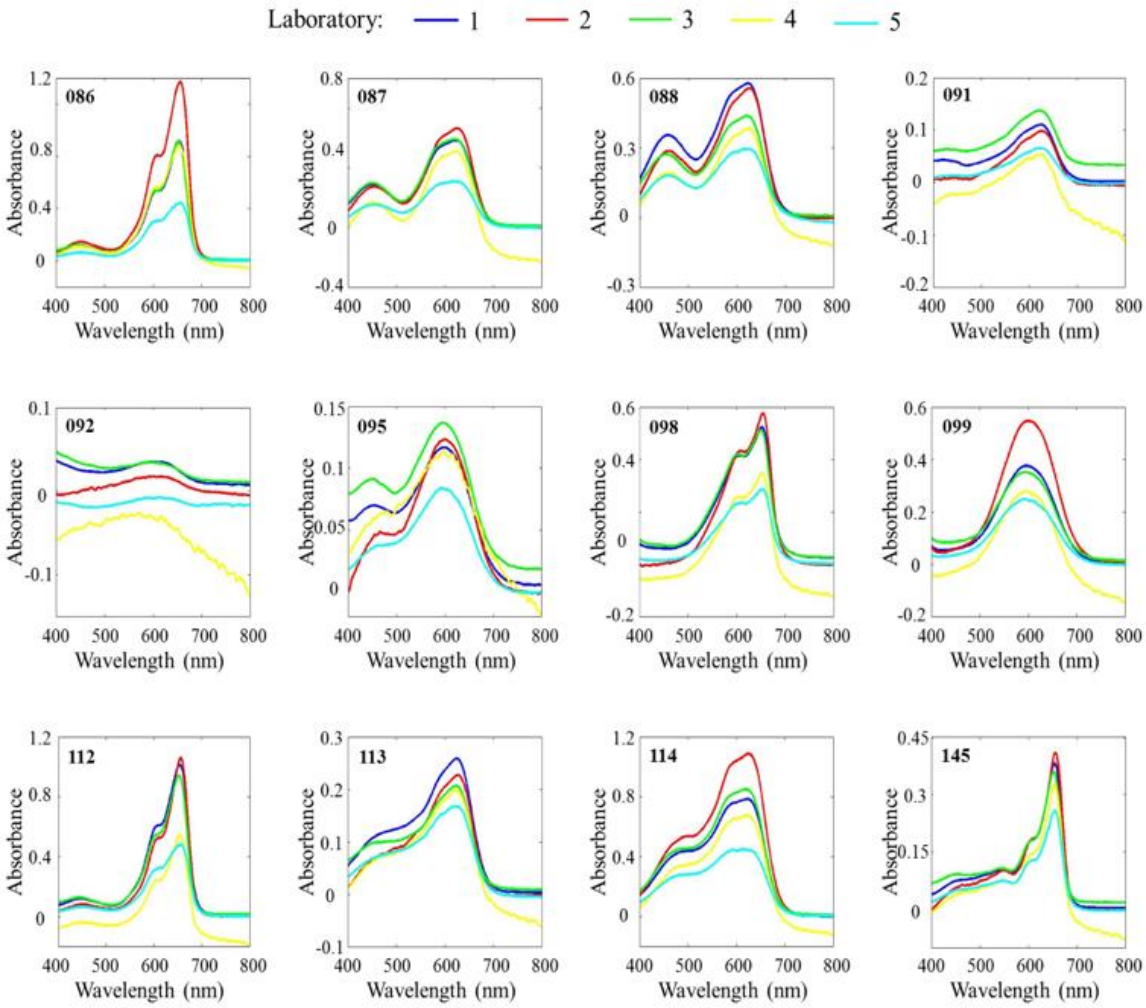

Figure 3.3 Mean absorbance spectra of 12 blue acrylic fibers collected at five laboratories.

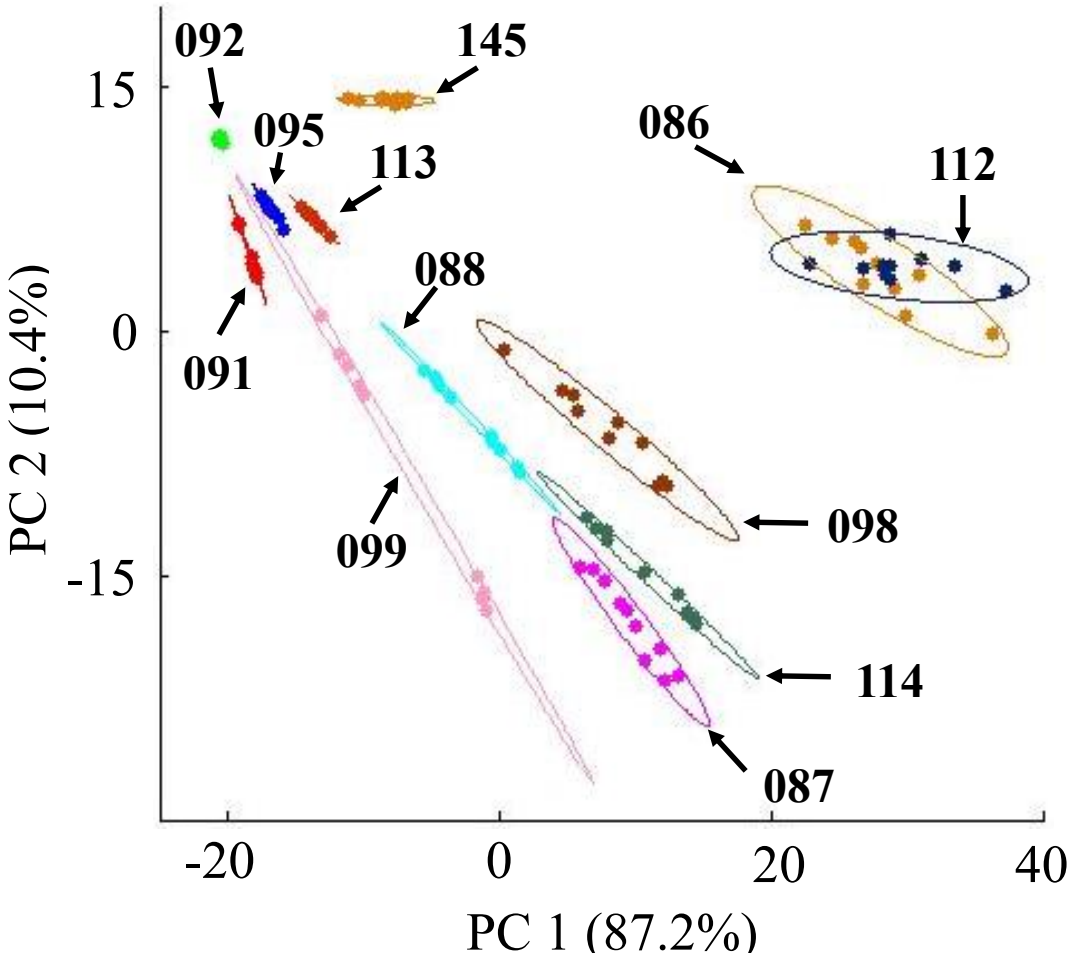

Figure 3.4 PCA scores plot of 12 blue acrylic samples (10 replicates each) collected at laboratory three.

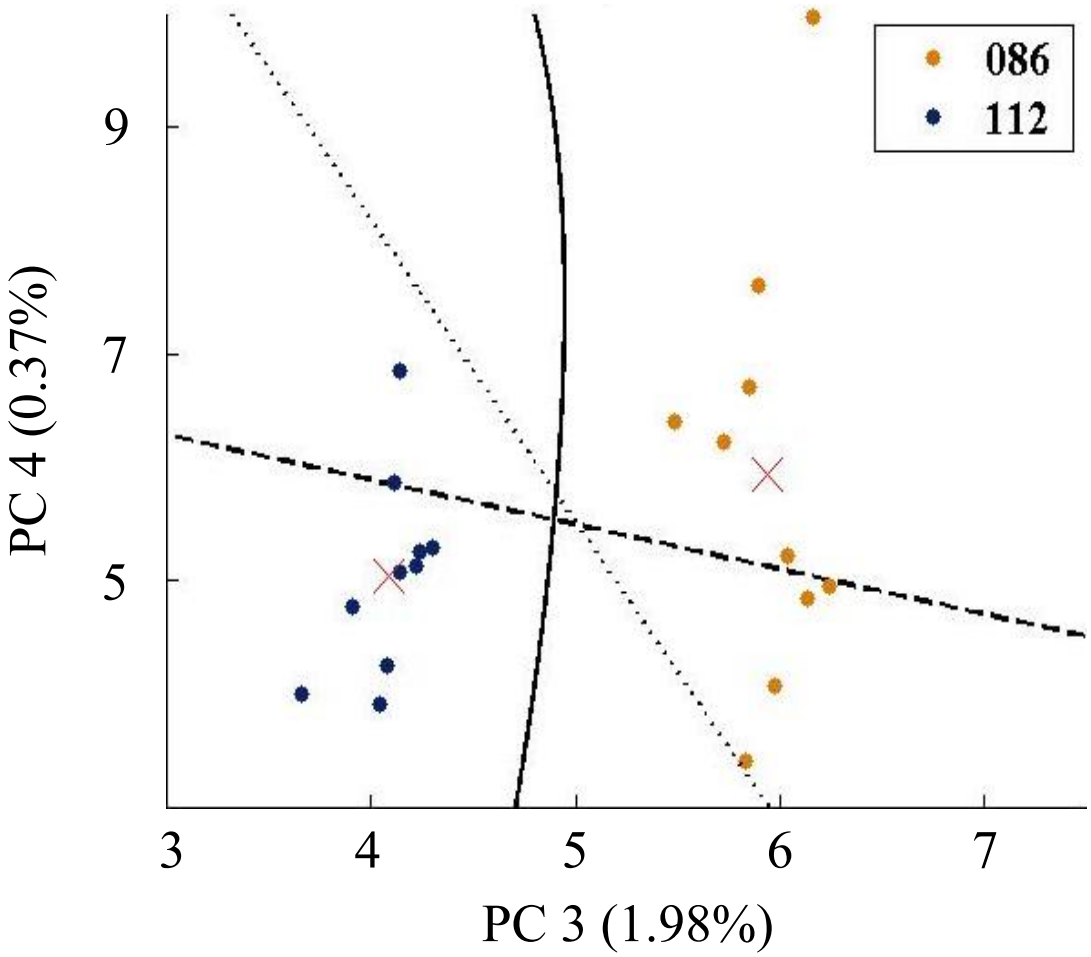

Figure 3.5 Decision boundaries used to separate samples 086 and 112 generated by QDA (solid line), LDA using pooled-covariance from all 12 samples (dashed line), and LDA with sample 099 excluded from pooled-covariance (dotted line). Class means are indicated by 'X'.

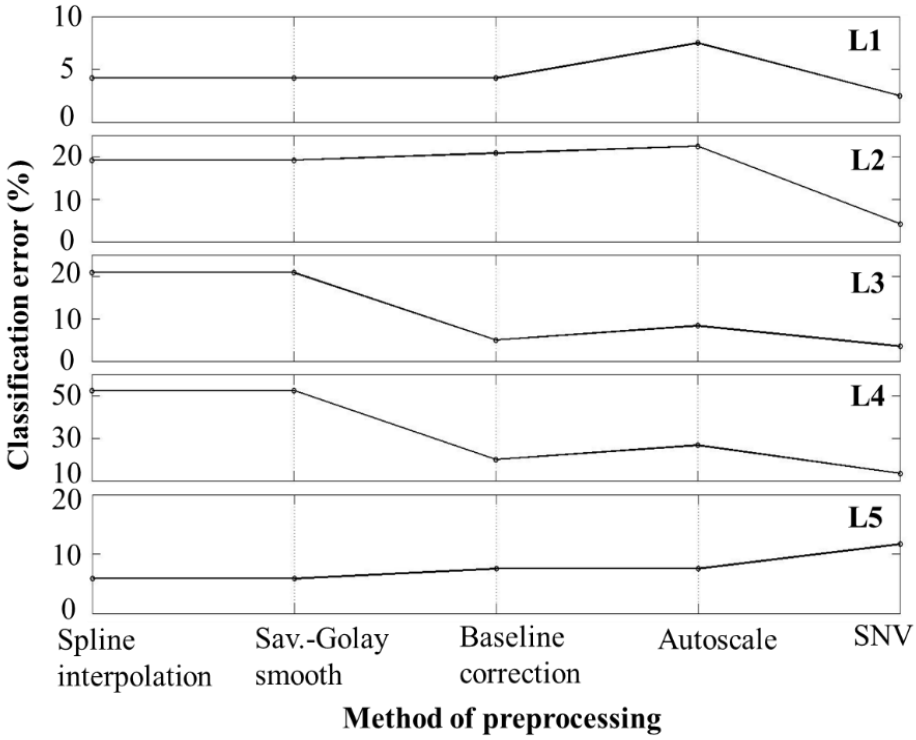

Figure 3.6 Misclassification percentages of LDA resulting from transfer of classification models to five laboratories. Each 'o' label involves that method of preprocessing combined with all those, if any, to the left of it.

CHAPTER 4

SUPERVISED MACHINE LEARNING IN PLAYABILITY DETERMINATIONS OF POLYESTER-URETHANE MAGNETIC AUDIO TAPES

ABSTRACT

As magnetic audio tapes age, the playability of the tapes can be affected by various chemical and physical degradation processes. The extent that a tape is degraded, however, is often difficult to observe without actually playing the tape. Attempting to play an extensively degraded tape can result in damage to the instrument used to play the tape as well as the tape backing. Recent work has focused on less destructive methods for determining the degree to which a tape is degraded before playing. The development of such methods would allow conservationists to predetermine if a tape should be subjected to a restoration process such as baking and eliminate any further damage from occurring.

In this study, attenuated total reflectance (ATR) Fourier transform infrared (FT-IR) spectroscopy was used as a rapid and objective method for determining the playability of polyester-urethane magnetic tapes. A collection of 95 tapes was used to generate chemometric models based on partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), artificial neural networks (ANN), naïve Bayes classification (NBC), and decision trees (DT). We were able to correctly classify a separate collection of 50 tapes with greater than 90% accuracy using PLS-DA, SVM-DA, and ANN.

1. INTRODUCTION

Magnetic tape was patented by Fritz Pfleumer in Germany in 1928. In the 1930s, the German company, AEG, utilized Pfleumer's invention to develop the Magnetophone, one of the world's first practical tape recorders. For the next several decades, magnetic audio tape was a favored recording medium for the entertainment industry and for personal use. Though magnetic tape recording has largely been replaced by digital recording in the 21st century, cultural heritage institutions in the United States have amassed collections totaling greater than 46 million tapes in various forms (reel to reel, VHS, DAT, cassette, etc.).[1] As tape aging progresses, they are susceptible to multiple chemical and physical degradation processes that could potentially render them unplayable. Because much of the information on magnetic tapes is now being digitized, a decision must be made on the tapes playability beforehand, either by visual inspection or by attempting to play the tape. The latter method is not ideal due to the potential harm that can result. If a tape is sufficiently degraded, it can fall-apart or stick to the playback device. When this occurs, valuable data may be lost, and the playback device may need extensive maintenance before being used again. With approximately 40% of all tapes being held in U.S. institutions in unknown condition,[1] an easy-to-use, reliable, and non-destructive method for determining the playability of these tapes would be beneficial and would allow archivists to identify tapes which should be restored using a process such as the baking method.[2,3]

Infrared (IR) spectroscopy meets the previously mentioned criteria and has been used previously to study magnetic audio tapes with "sticky shed syndrome."[3,4] Magnetic tape components which can be studied in the IR region include the finely dispersed

coating of magnetic oxide particles and polymeric binder materials. The magnetic particles, cobalt-doped iron oxide (Fe_2O_3) or chromium (IV) oxide (CrO_2), are responsible for sound and image reproduction, whereas the polymeric component determines the longevity of the tape along with other performance characteristics.[5] Some tapes have an additional third layer containing carbon particles which are dispersed on a polymeric material. This extra layer of support improves winding characteristics and reduces static build-up.[6] Since the 1970s, the most commonly used substrates in the construction of magnetic tapes belong to a class of polymers called polyester-urethanes.[7] The tapes may consist of one or multiple polyester-urethanes to achieve the desired properties. Some common degradation pathways for magnetic tape include hydrolysis of the polyester-urethane elastomers, binder oxidation, and peroxide decomposition.[5,6,8-11]

Examining the differences between IR spectra of degraded and non-degraded magnetic tape samples can be achieved using multivariate statistics. Algorithms which are capable of handling large amounts of data and have the ability to adjust their parameters (i.e., learning algorithms) as reference tapes are made available are logical choices for determining if a tape is degraded to the extent that it becomes non-playable using common tape playing machines. Because IR spectroscopy is already widely used in cultural heritage institutions and machine learning software is readily available, collection custodians potentially will be able to carry out non-invasive tape evaluations in a fast manner by combining the two approaches.

The objectives of this study were as follows: (1) determine if attenuated total reflectance (ATR) Fourier transform (FT) IR spectroscopy is a suitable technique for distinguishing playable from non-playable magnetic tapes, and (2) compare performance

of different models based on partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), artificial neural networks (ANN), decision trees (DT), and naïve Bayes classification (NBC).

2. EXPERIMENTAL

2.1 Samples

The magnetic audio tapes obtained for this study were used either to train classification models or as part of an external dataset to test the models after construction. The training set included 95 polyester-urethane-based quarter-inch reel-to-reel tapes. These samples were contributed by the Motion Picture Broadcasting and Recorded Sound Division (MBRS) at the Library of Congress (Washington, DC). For model testing, 38 quarter-inch tapes from the same MBRS collection were utilized along with an additional 12 tapes supplied by the University of Maryland (College Park, MD). Though brand and model information was sometimes found on the box in which the tapes were contained, or on the tape hub, this information cannot be considered reliable due to common rehousing practices. Only tapes in the original unopened packaging could be reliably associated with the brand and model identifiers. Tapes in this state would not have any recorded information on them.

2.2 Instrumentation

ATR FT-IR spectra of the magnetic side of all tape samples were collected using a Nexus 670 FT-IR (Thermo-Nicolet, Madison, WI) running Omnic version 8.2. The spectrometer was equipped with a Thunderdome ATR accessory (Thermo Spectra Tech, Inc., Shelton, CT) with a germanium crystal and an incident angle of 45°. IR detection was carried out using a deuterated triglycine sulfate (DTGS) detector.

Sections of tape were placed without preparation onto the germanium crystal. The tape was covered with a ~0.1 mm thick Mylar film to protect the tape surface from the metal anvil of the ATR pressure tower. In addition, a ~1 mm thick steel plate was placed on top of the Mylar film to disperse the pressure. A total of 20 replicate spectra of each tape were collected. Ten spectra were collected ~50 cm from the beginning of the tape with ~ 3 mm distances between the scanning areas on the tape. This process was repeated at ~100 cm from the beginning of the tape to obtain the remaining ten replicates. The spectral range investigated was 4000-500 cm^{-1} with 32 scans at 4 cm^{-1} resolution.

FT-IR spectra for the 95 tape training set were obtained by a single analyst in 2011, while spectra for the 50 tape test collection were obtained by two separate analysts in 2013.

2.3 Playability of Magnetic Tape Samples

Following ATR FT-IR analysis, attempts were made to play the tapes using a Scully 280 tape player (Scully Recording Instruments, Bridgeport, CT) by an audio engineer at MBRS. This subjective measurement made using a tape player is currently the standard approach for determining the playability of magnetic tapes.[6] The playabilities of the tapes were determined after attempting to pass the tape from one wheel to the next. Magnetic tapes were determined to be unplayable if there was noticeable friction between the tape player and any one of the six stationary guides, audible squealing noises were produced, material from the tape shed to the player guides, or fast-forward/rewind transitions were slowed due to increasing friction.

2.4 Data pretreatment

All preprocessing steps were performed using MATLAB version 8.1 (The MathWorks, Natick, MA). Each FT-IR spectrum was first truncated to a spectral region of 1750 to 950 cm^{-1} (416 variables). Following truncation, a least-squares polynomial smoothing of the signal was performed using the Savitzky-Golay (SG) algorithm with a 4^{th} order polynomial and 5-point moving window.[12,13] A standard normal variate (SNV) transformation was used to remove slope differences in the FT-IR spectra resulting from light scattering.[14] Following SNV transformation, the data was mean-centered. Mean-centering involves subtracting the average observation of a single feature from all observations made at that feature and is typically recommended when performing principal component analysis (PCA) or PLS.

2.5 Machine Learning

PLS-DA was performed using the PLS_Toolbox version 7.0.3 (Eigenvector Research, Wenatchee, WA). ANN, DT, NBC, and SVM-DA models were developed using the neural network and statistics toolboxes in MATLAB.

2.5.1 Partial least squares discriminant analysis

PLS-DA is a supervised classification technique based on algorithms for partial least squares (PLS) regression.[15] In PLS-DA, the dimensionality of the data is reduced according to:

$$X = TP^T + E \quad (1)$$

$$Y = UQ^T + F \quad (2)$$

where X is a matrix for model training containing the instrumental responses of 1900 samples (20 replicate spectra of 95 tapes) at 416 separate wavenumbers, Y is a categorical

vector containing the group memberships of the tapes determined by the playability testing method previously described, T and U are the resulting scores matrices, and P^T and Q^T are the orthogonal loadings matrices. Finally, contained in matrices E and F is the information not explained by the scores and loadings matrices of X and Y, respectively.

The factors, also called latent variables (LVs), used to decompose X and Y are created in such a way as to maximize the variance between spectra that is relevant for predicting classes. To avoid the inclusion of instrumental noise into the model, not all LVs are included in PLS-DA model building. The number of LVs to be retained was based on the error resulting from stratified 10-fold cross validation of all samples included in the calibration dataset. Because PLS is a useful compression tool for the purpose of making multivariate classification techniques more stable and less prone to overfitting, the LVs selected in this process were used as the input for all machine learning algorithms mentioned in this work.

For the binary classification problem in this study, Y was a dummy vector generated using zeroes and ones representing non-playable and playable tape samples, respectively. Therefore, the values calculated for each spectrum by PLS-DA read between zero and one. The Bayesian theorem, which assumes the predicted values will follow a distribution similar to those of future samples, was used to determine the threshold value between zero and one for classifying the tape samples as non-playable or playable.

2.5.2 Support vector machines

SVMs were built using a library for support vector machines (LIBSVM)[16] and a C-SVM classification algorithm with a linear kernel.[17,18] The objective of the SVM is to

find a hyperplane which correctly separated the non-playable tape LVs from the playable tape LVs. The hyperplane is created in such a way as to maximize the distance between the two classes of samples in the training set. The optimal hyperplane for separation is found using the minimization expression:

$$min\left(\frac{1}{2}\|\alpha_i y_i\|^2 + C\sum_{i=1}^{n}\xi_i\right)$$

where α_i are the Lagrange multipliers (parameters learned from the data), y_i are reference values (±1) indicating class membership, C is the cost parameter, n is the 1900 data points in the training set with index labels, i, and ξ_i are slack variables which can be thought of as the noise in the data causing individual training points to have smaller or even negative margins.[19] The minimization expression is subject to the following constraints:

$$y_i[\alpha_i y_i \phi(x_i) + b] \geq 1 - \xi_i$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, n$$

where $\phi(x_i)$ denotes the mapping to feature space, and b is the bias. The *C* parameter was optimized by maximizing the training performance of values between 1 and 1000.

2.5.3 Artificial neural networks

A three-layer feedforward ANN consisting of one input layer, one hidden layer, and one output layer was employed. The input layer is comprised of weighted values for the significant LVs determined during PLS-DA. The hidden layer is responsible for taking the weighted data in the input layer and passing it through a sigmoid function.[20] The transformed results are reweighted and passed through another sigmoid function in the final output layer which contains two neurons. The error in the corrected weight

measurements calculated at the output side of the neural network is propagated backward from the output layer to the hidden layer and finally to the input layer. Backpropagation was carried out using gradient descent with momentum and adaptive learning rate.[21]

To develop the ANNs in this study, spectra from 20 magnetic tapes were used for internal validation. The spectra used for this purpose were selected based on the algorithm developed by Kennard and Stone.[22] Training performance of the ANN was monitored using the cross-entropy method.[23] Training was allowed to progress until the cross-entropy of the validation set did not reach a minimum for six consecutive epochs. Finally, the epoch which resulted in minimum cross-entropy of the validation set was used to determine the weights for ANN testing.

2.5.4 Naïve Bayes

Naïve Bayesian classifiers are based on Bayes theorem with independence assumptions between predictors.[24] Although this assumption is often false, naïve Bayes is a popular classification method which has shown to outperform other more sophisticated techniques on a number of real-world datasets.[25,26] Bayes theorem is stated mathematically

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)} \tag{3}$$

where $P(c|x)$ is the posterior probability of class, c, given predictor, x, and $P(x|c)$ is the likelihood of observing x with the corresponding c. $P(c)$ and $P(x)$ are the prior probabilities of belonging to c and observing x, respectively.

One of the advantages of using naïve Bayes classification is the simplified training of the Bayesian parameters. The independent predictor assumption allows $P(x|c)$ to be calculated using the simplified product rule.[24]

$$P(x|c) = \prod_{i=1}^{D} P(x_i|c) \tag{4}$$

Here, x_i is the i^{th} of D predictors, and is either zero or one. A result of using Equation 4 is many of the parameters that would need to be estimated if the independence assumption was not made have now been eliminated. Assuming a normal distribution enables $P(x_i|c)$ to be calculated from the number of samples, n, in each class along with the group mean, μ, and standard deviation, σ.

$$\mu = \frac{1}{n}\sum_{i=l}^{n} x_i \tag{5}$$

$$\sigma = \left[\frac{1}{n-1}\sum_{i=l}^{n}(x_i - \mu)^2\right]^{0.5} \tag{6}$$

$$P(x_i|c) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} \tag{7}$$

In this study, the class prior probabilities for non-playable and playable magnetic tape were uniform ($P(c)$ = 0.5).

2.5.5 Decision trees

Classification trees in this study were built using the top-down (i.e., tree begins with best predictor) approach with each of the input LVs representing the nodes in the tree. The 'node error' method was used as the criterion to split the items in the tree. Node error is calculated simply as one minus the fraction of misclassified samples at a given node. In general, the criterion used for splitting has a much smaller impact on the

predictive nature of the tree when compared to the criterion used for determining the number of branches which should be retained.[27] Here, new nodes were created until all samples in the training set have been placed into one of the two classes. Stratified 10-fold cross-validation was used to determine which branches were insignificant for classification and should be subjected to 'pruning'.

2.5.6 Misclassification Cost

Predicted classifications, $\hat{y}$, are made so as to minimize the expected classification cost:

$$\hat{y} = \underset{y=1,\ldots,K}{\arg\min} \sum_{k=1}^{K} \hat{P}(k|x)\, C(y|k) \qquad (8)$$

where *K* is the number of classes, $\hat{P}(k|x)$ is the posterior probability of observation x belonging to class *k*, and $C(y|k)$ is the cost of misclassifying an observation as *y* when its true class is *k*. *C* is a zero-diagonal square matrix with the cost of falsely classifying a playable tape as non-playable in the first row/second column, and the cost of falsely classifying a non-playable tape as playable in the second row/first column. For example, setting *C* equal to [0,1;2,0] would indicate a 2× higher cost for classifying a tape as playable when it is actually not playable. For ease of discussion, in all discussions of *C*, only the non-zero digits will be referred to.

3. RESULTS AND DISCUSSION

3.1 Infrared analysis

A total of 95 tapes were used for the purpose of classification model training. Of the 95 tapes in the training set, 55 tapes were considered playable, and the remaining 44 were determined to be non-playable based on playability testing. Twenty replicate ATR

FT-IR spectra were collected for each tape in the training set, and the averaged spectra calculated from the raw ATR FT-IR data are shown on the top panel of Figure 4.1. As can be seen on the bottom panel, the differences in the spectra of playable and non-playable tapes are significantly enhanced by mean centering after SG smoothing and SNV. It is also worth noting that peak shoulders such as those at 1710 and 1515 cm^{-1} become more pronounced after preprocessing.

The chemical differences which lead to distinctive ATR FT-IR spectra between classes is revealed in Table 4.1. The vibrational assignments made were based on previous studies on the degradation of polyester-urethane using FT-IR spectroscopy.[28-30] Intensity increases at 1730, 1710, and 1140 cm^{-1} for non-playable tapes appear to indicate hydrolysis product formation. Hydrolysis of polyester-urethane yields carboxylic acid, consistent with the rise in intensity of the C=O vibrational peak, and alcohol. Other differences between classes include peaks at 1363 and 1254 cm^{-1} which only show up in the spectra of non-playable tapes. In a study by Zelst,[31] these peaks were shown to be consistent with the IR spectrum of poly(1,4-butylene adipate) (PBA) using spectral subtraction. PBA is the base of polyester-urethane binder material and is believed to be increasingly exposed on the surface of the tape as polyester-urethane degrades.

3.2 Partial least squares discriminant analysis

Models based on PLS-DA, SVM-DA, ANN, DT, and NBC were developed using the 95 magnetic tape samples selected for use as training data to determine if playable tapes could be distinguished from non-playable tapes using ATR FT-IR. Training data was preprocessed using a SG smooth followed by SNV and mean centering.

Approximately 76.1% of the variance in the data is explained using the first three LVs obtained from PLS-DA. The scores for these LVs are plotted against each other in Figure 4.2. The scores plot shows separation into two distinct groups based on the chemical composition of the tapes. What cannot be seen from this figure are the subgroupings of the tapes. In most instances, all 20 replicate spectra of a tape are located in the same area of LV space. However, there are several instances where the 10 spectra taken at 50 cm are clustered in a separate location from the 10 spectra of the same tape collected at 100 cm. This indicates that some tapes are not chemically homogenous over the entire length of the tape. Therefore, different predictions about the playability of a single tape could be made by simply collecting spectra at two different locations.

PLS-DA loadings plots for the first three LVs are shown in Figure 4.3. As expected, the loading for the first LV shows the most contribution from the free carbonyl groups likely resulting from carboxylic acid formation due to hydrolysis. Hydrolysis of the polyester-urethane binder has been considered by many researchers to be the main degradation mechanism leading to sticky shed syndrome.[5,8,9] Contributions from the hydrogen bonded carbonyl groups appear in the second LV. The ester components (e.g., PBA and poly(ethylene terephthalate)) which aid in distinguishing playable from non-playable tapes in the region between 1244 and 1095 cm^{-1} are visible in the second LV but are most prominent in the third LV.

Poor generalization of discriminant analysis models can occur as a result of fitting noisy data that are not relevant for predicting the classes. To avoid overfitting, the number of LVs to be used as input for all models built in this study was specified by plotting the average classification error of the calibration and stratified 10-fold cross-

validation results of PLS-DA versus the LV number as shown in Figure 4.4. Due to the increase in the average error of both calibration and cross-validation when transitioning from 11 to 12 LVs, it was determined that 11 LVs would be used to build the models. Also noteworthy in Figure 4.4 is the role that the first three LVs, which have already been described through the use of scores and loadings plots, have for the prediction of playable and non-playable tapes. The greatest reductions in classification error result from the use of those particular LVs.

3.3 Support vector machine discriminant analysis

SVM-DA algorithms can apply linear or radial basis function (RBF) kernels for linear or nonlinear separations, respectively. Because the kernel function defines the transformed feature space in which the separation will occur, selection of the appropriate kernel is critical. In this study, the linear approach was selected for use toward playability classification of polyester magnetic tapes. The reasons for selecting the linear kernel over the RBF kernel were twofold. First, previous work in this lab using linear discriminant analysis has shown that it is possible to discriminate non-playable tapes from playable tapes using a linear function. Second, when compared to linear SVM-DA, RBF SVM-DA is much more susceptible to overfitting. This problem often causes poor generalization even though little error may be seen during calibration and cross-validation.

An additional advantage to using linear SVM-DA is the significantly less computation required in comparison to RBF SVM-DA. The linear SVM-DA model used here was trained in less than 12 seconds. For the purpose of comparison, an attempt was made to generate an SVM-DA model using the RBF kernel. Training of RBF SVM-DA was completed in ~4 minutes. The additional time needed for RBF SVM-DA results from

the amount of parameter optimization involved. While both methods are performed by optimizing the SVM-DA cost parameter, *C*, a linear profile $K(x_i, x_j) = x_i^T x_j + c$ requires no additional optimization. Conversely, the σ value used to define the Gaussian profile $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/(2\sigma^2)$ in RBF SVM-DA must be optimized. The optimal *C* parameter for the linear SVM-DA model was ~316 and included 126 support vectors. A relatively small value for *C* (i.e., closer to 1 than 1000), a small number of support vectors (using only 126 of the 1900 training samples), and performance on an external test set all suggest the risk of overfitting has been minimized.

3.4 Artificial neural networks

The three-layer ANN configuration used in this study is shown in Figure 4.5. The first layer consists of 11 nodes corresponding to the LVs selected in PLS-DA. The number of hidden layers was set to nine based on the trial and error method. The two output possibilities represent non-playable or playable group membership. Internal validation was used to determine the weights of the hidden layers used to classify the LV scores. In this process, 20 (10 from both non-playable and playable classes) of the 95 tapes were removed from the training data and used as a validation set. The 75 tapes to be used for building the ANN models were determined using the Kennard-Stone algorithm to ensure that samples defining the outer boundaries of the two classes were maintained for training. Figure 4.6 indicates, based on the minimum cross-entropy of the validation set, the weights used at epoch 123 should be used in the final model. The similar validation and test curves suggest the Kennard-Stone method is appropriate for selecting validation samples which will act similarly to newly introduced samples. In addition, the two curves being similar indicate that it is doubtful that overfitting has occurred.

3.5 Naïve Bayes classification

One of the advantages to using NBC is the transparency of the technique. Kononecko[32] related the probabilistic nature of NBC to those used by physicians as a way of diagnosing their patients in a replicative fashion. Because the prior probability of magnetic tape being classified as either playable or non-playable is uniform, playability was entirely based on the posterior probabilities of belonging to the classes as determined by Equation 6. Posterior probabilities were calculated using the class measures listed in Table 4.2 to estimate a separate normal distribution for each class. In this dataset, negative LV scores are indicative of non-playable tapes, whereas positive LV scores suggest the tape is playable. Of the five classification techniques mentioned in this study, NBC resulted in a calibration model with the highest misclassification rate at ~8.6%. The large standard deviations between group scores resulted from this relatively small portion of data.

Figure 4.7 shows plots of the logarithm of the posterior probability of each spectrum belonging to playable and non-playable groups. Points lying along the top edge of these plots colored red are spectra corresponding to playable tapes, determined by playability testing, and were correctly assigned using NBC. The same can be said about the non-playable tape data shown along the right side in blue. The area in the top right corner of the plots (magenta in color) is where significant overlap of the class memberships occurs. This area can be thought of as the 'borderline' region where the tapes transition from playable to non-playable. The described region corresponds to samples of which the posterior probability of belonging to either class does not exceed 0.95. Approximately 24% of the training data and 17% of the testing data fell in the

borderline region. Based on the calibration data, it appears that one non-playable tape (20 spectra) shared similar chemical characteristics with many of the non-playable tapes and as a result was misclassified. An examination of the test data shows 20 replicate spectra from a playable tape showing up in the group of non-playable tape spectra. Because these misclassifications occur mostly near the borderline region, it is likely these errors are due to the subjective nature of playability testing as previously discussed.

3.6 Decision trees

Decision trees are a vast class of nonlinear classifiers.[25,27] A full DT using all 11 LV scores is given in Figure 4.8. The root node, the node which begins the tree, is the predictor which best separates the data into classes of non-playable and playable tapes. In this instance, the root node consists of the scores from LV 1, which takes into account much of the intensity differences seen between classes near 1730 cm^{-1}. The data is then split into two nodes based on a threshold value which minimizes the classification error of the training data. LV 3, which is heavily influenced by class differences between 1244 and 1095 cm^{-1} is the second most useful predictor in the tree. DT predictions about the playability of magnetic tape may be made using only scores from LVs 1 and 3. The results of cross-validation suggest that the full DT with scores from all 11 LVs should be used for prediction. Cross-validation misclassification rates of 0.01 and 0.09 were obtained using the full and pruned trees, respectively. The misclassification rate of the full tree when applied to the test data, however, was higher (0.19) than that of the pruned tree (0.16). This result suggests that the full-tree carries some level of overfitting, or some samples in the test set are significantly different than those used to train the DT. This issue should become clearer with more test samples. It should be noted that all values

communicated during the performance comparison of all model-types were obtained using the full DT as opposed to the shortened version.

3.7 Performance comparison

Table 4.3 gives the sensitivity (true positive rate) and fall-out (false positive rate) results of calibration, stratified 10-fold cross-validation, and external testing for each model-type used in the study. For binary classifications, the specificity (true negative rate) of one class is equal to the sensitivity of the opposing class. For this reason, specificity values are not directly mentioned. Here, a 'good' model is judged as being one that gives high sensitivity and low-fallout in both classes throughout calibration, cross-validation, and testing. All models showed good performance during calibration and cross-validation giving sensitivity values greater than 0.9 and fall-out values less than 0.1 for both non-playable and playable tapes. The generalization results show more separation between the methods, however. Even though the same 11 LVs were used as input to train all models, when applied to external data, only SVM-DA and ANN met our criteria to be considered good working models. Having the lowest across-the-board sensitivity and fall-out values, classification using DT was considered unfavorable. Though it depends heavily on the method, the sensitivity and fall-out results, especially in the cases of calibration and cross-validation, show that the likelihood of classifying a playable tape as non-playable is very similar to that of classifying a non-playable tape as playable.

To lower the likelihood of classifying a non-playable tape as playable, a scenario most tape custodians would like to avoid if possible, a cost parameter can be added to the statistical methods described above. In this work, the cost ratio, the cost of misclassifying

a non-playable tape as playable divided by the cost of misclassifying a playable tape as non-playable, was increased until a false positive rate for playable tapes in the calibration set reached a value of ~0.01, or ~1%. Because two of five methods, ANN and DT, already had false positive rates of playable tapes below 0.01, the cost parameter was only adjusted for three methods, PLS-DA, SVM-DA, and NBC. The results of this analysis are shown in Table 4.4. The instances of falsely classifying a non-playable tape as playable decreased by 4.9%, 2.9%, and 3.3% for PLS-DA, SVM-DA, and NBC, respectively. The false positive rates for playable tapes in the test sets are all greater than 0.1. With all tapes in the calibration set originating from only one institution, this appears to be due to tape formulation differences (discussed more in Chapter 5) or storage condition differences between institutions. The effect that changing the cost parameter had on the overall classification accuracy of magnetic tape samples was highly dependent on the method. SVM-DA and NBC models displayed an increase in classification of 1.2% and 1.0%, respectively, whereas, the classification of the PLS-DA model decreased by 7.7%.

It is often prudent to determine which classification techniques are best by examining how well they perform on an external test set, provided that test set is representative of the population, and the amount of computation required for model training. These metrics are displayed in Table 4.5. Although multiple techniques (PLS-DA, SVM-DA, ANN, and NBC) resulted in a correct prediction rate greater than 90%, ANNs were clearly the best discriminator of this set of data based on the higher prediction accuracy and Matthew's correlation coefficient.[33] Epochs required for ANN training, and parameter optimization required for the training of SVM-DA (especially in the case of nonlinear SVM-DA) led to increases in computation time. Conversely, logic

based algorithms (e.g., DTs) and statistical learning algorithms (e.g., PLS-DA or NBC), though the prediction accuracies were significantly lower in most instances, required less time for training.

4. CONCLUSION

In this paper, various supervised machine learning techniques have been successfully applied to ATR FT-IR spectra for the purpose of determining playability of polyester-urethane magnetic tapes. No learning algorithm can perform best on all types of datasets, and each technique has inherent strengths and weaknesses. Some of the highest prediction accuracies were gained using ANN and SVM-DA models. Some pitfalls of these methods are poor interpretability and sometimes hefty computation times. The opposite holds true for DT and NBC which are easily interpretable and train very quickly, but tend to give less accurate predictions. NBC typically has high bias which occurs as a result of assuming the data follows a single (e.g., normal) probability distribution. Conversely, the hard cut-off values used to separate classes using DTs make the method susceptible to high variance. Finally, the other statistical method used in this research, PLS-DA, seemingly strikes a balance between all of the methods we have studied. The method gave slightly lower prediction accuracies compared to those of ANNs, but the scores and loadings determined in PLS-DA are useful for studying the chemical nature of which class separation is based upon. As shown in this study, PLS-DA has also proven to be a useful tool for reducing the dimensionality of ATR FT-IR magnetic tape data. In future work, algorithms similar to the ones described here could be utilized as part of a software package that enables conservationists to screen for tapes which require restoration, without risking further damage to the tape or playing instrument.

ACKNOWLEDGMENTS

This project was supported by Grant LG-06-12-056912 from the Institute of Museum and Library Services to the University of South Carolina. Portions of this work were presented at the Great Scientific Exchange (SciX) conference at the Rhode Island Convention Center in Providence, Rhode Island on September 29, 2015. In addition, the work presented here was made to support a published article in Analytical Chemistry (volume 87) entitled, "Minimally Invasive Identification of Degraded Polyester-Urethane Magnetic Tape Using Attenuated Total Reflection Fourier Transform Infrared Spectroscopy and Multivariate Statistics." The co-authorship of Brianna M. Cassidy, Zhenyu Lu, Michael L. Myrick, Eric M. Breitung, and Stephen L. Morgan is acknowledged. The authors are grateful for support from Heather M. Heckman, Lydia C. Pappas, and Gregory J. Wilsbacher at Moving Image Research Collections at the University of South Carolina, and Gene DeAnna, Larry Miller, Fenella France, and Linchi Nguyen (Summer 2013, ACS Project SEED research participant) at the Library of Congress. The contributions made by University of South Carolina undergraduate researchers, Samantha E. Skelton and Eric J. Bringley are also recognized.

REFERENCES

1. *A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections*, Heritage Preservation, Inc., Washington, D.C., 2005.

2. Hobaica, S. Analysis of Audio Magnetic Tapes with Sticky Shed Syndrome by ATR FT-IR, *J. Appl. Polym. Sci.* **2013**, *128*, 1962-1973.

3. Richardson C. Process for Restoring Magnetic Recording Tape Damaged by "Sticky Shed" Syndrome. United States Patent No. 6797072 B1 (28 Sept. 2004).

4. Breitung, E.; Skelton, S.; Morgan, S. *Rapid Identification of Sticky Shed Syndrome in Magnetic Tape Using ATR-FTIR and Multivariate Statistics*, 40th Annual American Institute of Conservation Meeting, Albuquerque, New Mexico, 2012.

5. Edge, M.; Allen, S.; Hayes, M.; Jewitt, T.; Horie, C.; and Brems, K. Degradation of Magnetic Tape: Support and Binder Stability, *Polym. Degrad. Stabil.* **1993**, 39, 207-214.

6. Thiébaut, B.; Vilmont, L-B.; Lavédrine, B. Characterization of U-matic Videotape Deterioration by Size Exclusion Chromatography and Pyrolysis Gas Chromatography/Mass Spectrometry and the Role of Adipic Acid, *J. Cultural Heritage* **2009**, 10, 183-197.

7. Hess R. Tape Degradation Factors and Challenges Predicting Tape Life, *J. Assoc. Rec. Sound Collect.* **2008**, 39, 240-274.

8. Bertram, H.; Cuddihy, E. Kinetics of the Humid Aging of Magnetic Recording Tape. *IEEE Trans. Magn.* **1982,** 18, 993-999.

9. Cuddihy, E. Aging of Magnetic Recording Tape, *IEEE Trans. Magn.* **1980**, 16, 558-568.

10. Volkmann, H. Film Preservation: a Report of the Preservation Committee of the International Federation of Film Archives, *FIAF* **1965**, 7.

11. Eilers, D. Polyester and Acetate as Magnetic Tape Backings, *J. Audio Engng. Soc*. **1969**, 17, 303-308.

12. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.* **1964**, 36, 1627–39.

13. Steiner, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Squares Procedure. *Anal. Chem.* **1972**, 44, 1906-1909.

14. Barnes, R.; Dhanoa, M.; Lister, S. Standard Normal Variate Transformation and De-trending of Near-infrared Diffuse Reflectance Spectra, *Appl. Spectrosc.* **1989**, 43, 772-777.

15. Barker, M.; Rayens, M. Partial Least Squares for Discrimination, *J. Chemom*. **2003**, 17, 166-173.

16. Chang, C.; Lin, J. LIBSVM: a Library for Support Vector Machines, *ACM TIST*, **2011**, 2, 1-27.

17. Boser, B.; Guyon, I.; Vapnik, V. A Training Algorithm for Optimal Margin Classifiers, *COLT* **1992**, 5, 144-152.

18. Cortes, C.; Vapnik, V. Support-Vector Network, *Mach. Learn.* **1995**, 20, 273-297.

19. Cristianini, N.; Shawe-Taylor, J. In *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*; Cambridge University Press: Cambridge, U.K., 2000.

20. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Series in Statistics: Springer: New York, 2001.

21. Neural Network ToolboxTM User's Guide. http://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf (accessed Jan. 16, 2015).

22. Kennard, R.; Stone, L. Computer Aided Design of Experiments, *Technometrics* **1969**, 11, 137-148.

23. Reubinstein, R.; Kroese, D. *The Cross-Entropy Method: A Unified Apporach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*; Springer-Verlag: New York, 2001.

24. Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, U.K., 2012.

25. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 3rd ed.; Academic Press, California, 2006.

26. Domingos, P.; Pazzani, M. Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier, *Mach. Learn.* **1997**, 29, 103-130.

27. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Wadsworth International Group: California, 1984.

28. Thompson, D.; Osborn, J.; Kober, E.; Schoonover, J. Effects of Hydrolysis-Induced Molecular Weight Changes on the Phase Separation of a Polyester Urethane, *Polym. Degrad. Stab.* **2006**, 91, 3360-3370.

29. Schoonover, J.; Dattelbaum, D.; Osborn, J.; Bridgewater, J.; Kenney, J. Pressure-Dependent Fourier Transform Infrared Spectroscopy of a Polyester-Urethane, *Spectrochim. Acta A* **2003**, 59, 309-319.

30. Wilhelm, C.; Gardette, J. Infrared Analysis of the Photochemical Behavior of Segmented Polyurethanes: 1. Aliphatic Polyester-Urethane, *Polymer* **1997**, 38, 4019-4031.

31. Zelst, L. *Sticky Shed Sydrome: Characterization, Diagnosis, and Treatment,* Internal Library of Congress Report, March, 2008.

32. Kononenko, I. Inductive and Bayesian learning in Medical Diagnosis, *Appl. Artif. Intell.* **1993**, 7, 317-337.

33. Matthews, B. Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme, *Biochim. Biophys. Acta-Protein Struct. M*. **1975**, 405, 442-451.

Table 4.1 Peak assignments for infrared absorbance spectrum of magnetic tape. Wavenumbers are approximates based on average spectrum of tapes.

Wavenumber (cm^{-1})	Tape type	Assignment[a]	Tape component	References
1730	Both	ν(C=O) free	Poly(ester-urethane)	28-30
1710	Both	ν(C=O) H-bonded	Polyurethane	29
1600	Both	ν(C=C) Phenyl ring	Polyurethane	29,30
1533, 1311, 1220	Both	ν(C=N) + δ(N–H)	Polyurethane	28-30
1475	Non-playable	$\delta(CH_2)$	Poly(ester-urethane)	28-30
1460	Both	$\delta(CH_2)$	Poly(ester-urethane)	28-30
1414	Both	ν(C–C) Phenyl ring	Polyurethane	29,30
1394	Non-playable	$\omega(CH_2)$	Poly(ester-urethane)	29,30
1363	Non-playable	ω (CH_2)	Poly(ester-urethane)	29,30
1254	Non-playable	ν(C–O–C); $\omega(CH_2)$	Polyester	29,30
1174	Both	ν(C–O–C)	Poly(ester-urethane)	28,30
1142	Non-playable	ν(C=O) + $\nu(O–CH_2)$	Poly(ester-urethane)	30
1076	Non-playable	ν(C–O–C)	Poly(ester-urethane)	29,30
1065	Playable	ν(C–O–C)	Poly(ester-urethane)	29,30
1018	Both	ρ(C–H)	Polyurethane	29

[a]Vibrational modes are: ν = stretching, δ = in-plane bending (scissoring), ρ = in-plane bending (rocking), ω = out-of-plane bending (wagging).

Table 4.2 Naïve Bayes classification parameters.

Latent variable	Mean		Standard deviation	
	Non-playable	Playable	Non-playable	Playable
1	-2.5500	2.7163	2.0930	2.3732
2	-0.5085	0.5416	1.5761	1.4034
3	-0.3334	0.3552	1.3509	1.7732
4	-0.2424	0.2582	2.0735	0.9426
5	-0.1106	0.1178	0.7496	0.7325
6	-0.0869	0.0926	1.0153	0.6894
7	-0.0696	0.0741	0.6776	0.6868
8	-0.0306	0.0326	0.2943	0.2817
9	-0.0243	0.0259	0.2431	0.3565
10	-0.0264	0.0281	0.3204	0.4095
11	-0.0215	0.0229	0.2178	0.2663

Table 4.3 Sensitivity and fall-out of calibration, stratified 10-fold cross-validation, and external testing with equal cost of misclassifications between classes.

Model	Method[a]	True positive rate		False positive rate	
		Non-playable	Playable	Non-playable	Playable
Calibration	PLS-DA	0.957	0.959	0.041	0.043
	SVM-DA	0.977	0.967	0.033	0.023
	ANN	1.000	0.999	0.001	0.000
	NBC	0.898	0.929	0.070	0.102
	DT	0.986	0.999	0.001	0.004
Cross-validation	PLS-DA	0.955	0.959	0.041	0.045
	SVM-DA	0.977	0.966	0.034	0.023
	ANN	0.992	0.995	0.005	0.008
	NBC	0.893	0.929	0.089	0.107
	DT	0.986	0.987	0.013	0.014
Test	PLS-DA	0.898	0.921	0.079	0.102
	SVM-DA	0.900	0.911	0.089	0.100
	ANN	0.950	0.905	0.095	0.050
	NBC	0.895	0.911	0.089	0.104
	DT	0.829	0.771	0.229	0.171

[a]Method abbreviations: partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), artificial neural networks (ANN), naïve Bayes classification (NBC), and decision trees (DT).

Table 4.4 Sensitivity and fall-out of calibration, stratified 10-fold cross-validation, and external testing with higher cost of falsely classifying a non-playable tape as playable.

Model	Method[a]	Cost ratio	True positive rate		False positive rate	
			Non-playable	Playable	Non-playable	Playable
Calibration	PLS-DA	1199	0.865	0.992	0.135	0.008
	SVM-DA	6.201	0.989	0.919	0.099	0.011
	NBC	10.07	0.992	0.820	0.180	0.008
Cross-validation	PLS-DA	1199	0.862	0.985	0.138	0.015
	SVM-DA	6.201	0.989	0.914	0.086	0.011
	NBC	10.07	0.982	0.817	0.183	0.017
Test	PLS-DA	1199	0.758	0.947	0.242	0.053
	SVM-DA	6.201	0.929	0.887	0.113	0.071
	NBC	10.07	0.968	0.818	0.182	0.032

[a]Method abbreviations: partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), and naïve Bayes classification (NBC).

Table 4.5 Model testing classification accuracies and Matthew's correlation coefficients with equal cost of misclassifying playable and non-playable tapes.

Method	Correct classification	MCC	Time (s)
PLS-DA	0.907	0.820	1.357
SVM-DA	0.906	0.814	11.94
ANN	0.923	0.842	4.522
NBC	0.901	0.795	0.170
DT	0.807	0.595	0.890

[a]Method abbreviations: partial least squares discriminant analysis (PLS-DA), support vector machine discriminant analysis (SVM-DA), artificial neural networks (ANN), naïve Bayes classification (NBC), and decision trees (DT).

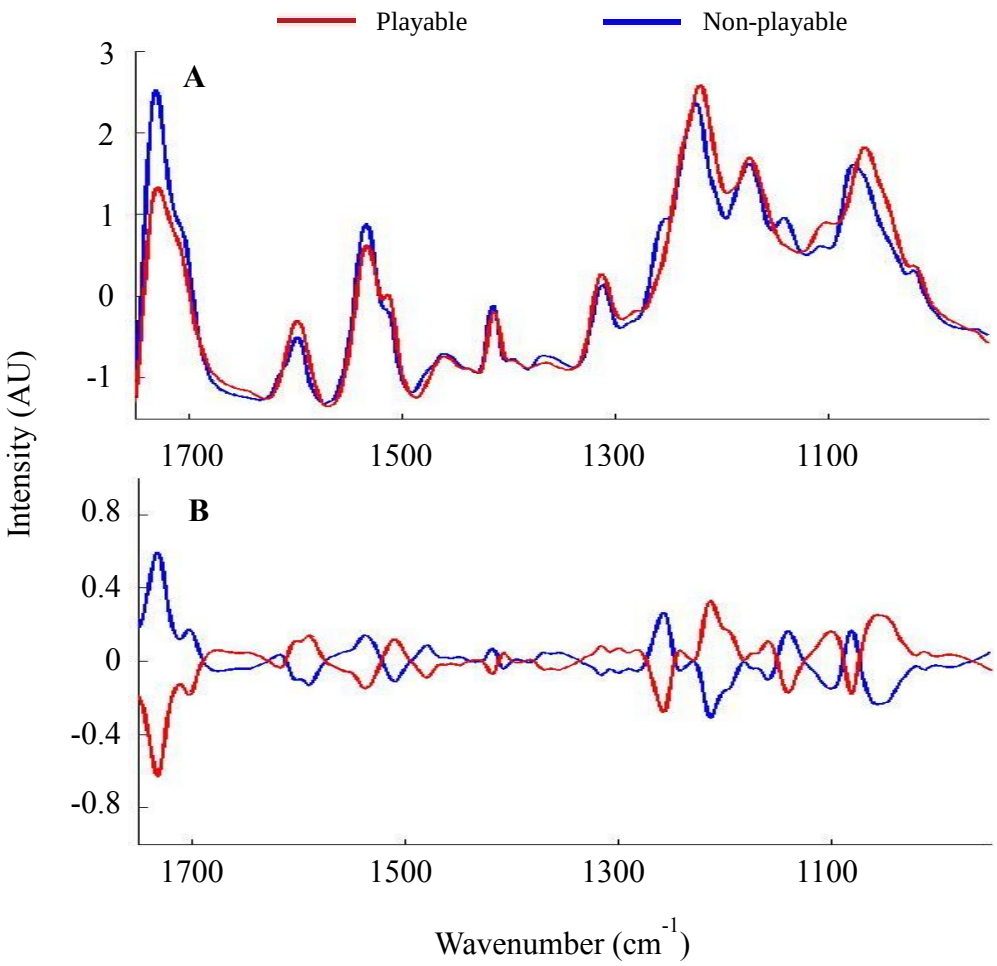

Figure 4.1 Absorbance spectra from magnetic tape samples (a) before preprocessing and (b) after preprocessing with Savitzky-Golay smoothing followed by standard normal variate transform and mean centering.

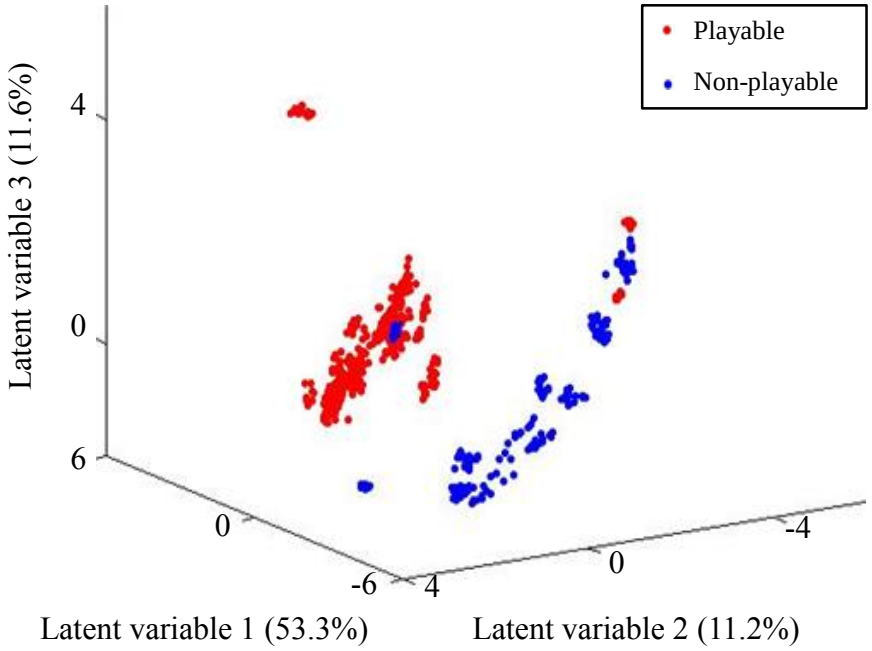

Figure 4.2 Partial least squares discriminant analysis scores plot for playable and non-playable magnetic tape samples.

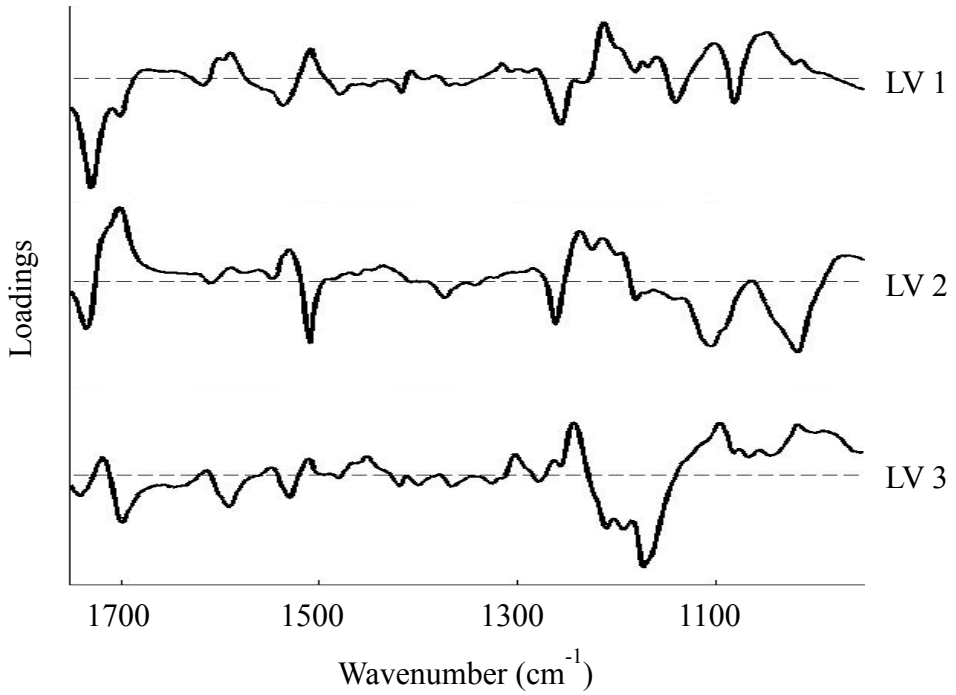

Figure 4.3 Partial least squares discriminant analysis loadings plot for the first three latent variables.

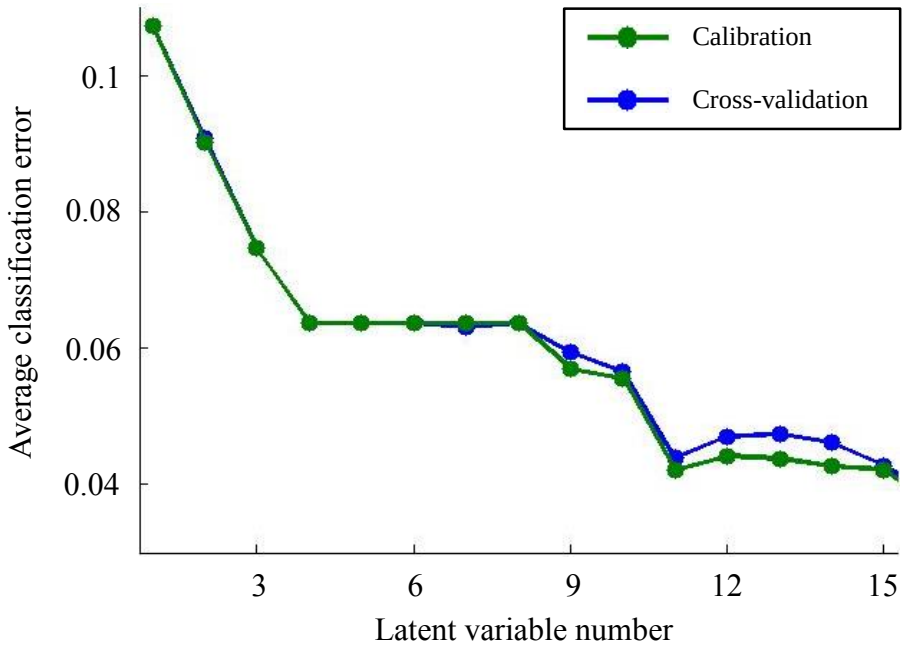

Figure 4.4 Average classification error of calibration and stratified 10-fold cross-validation models by number of latent variables used in partial least squares discriminant analysis.

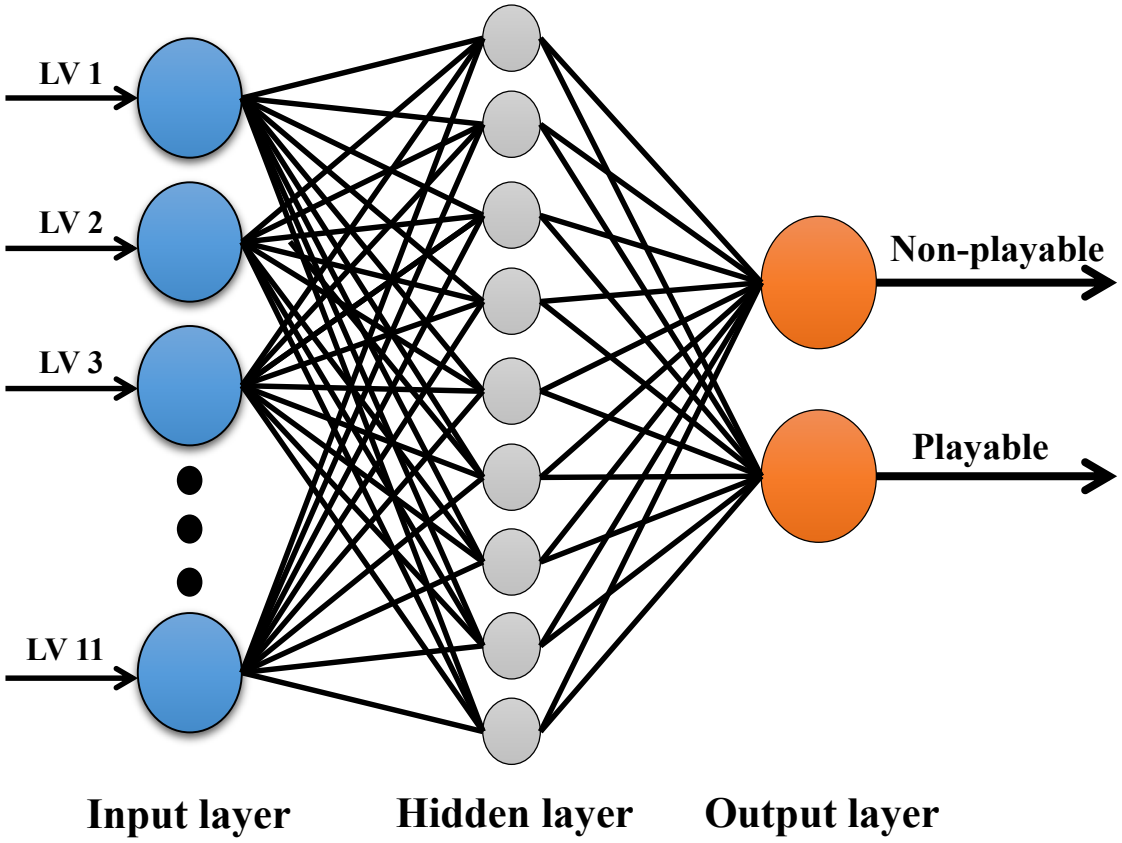

Figure 4.5 Three layer feed-forward artificial neural network.

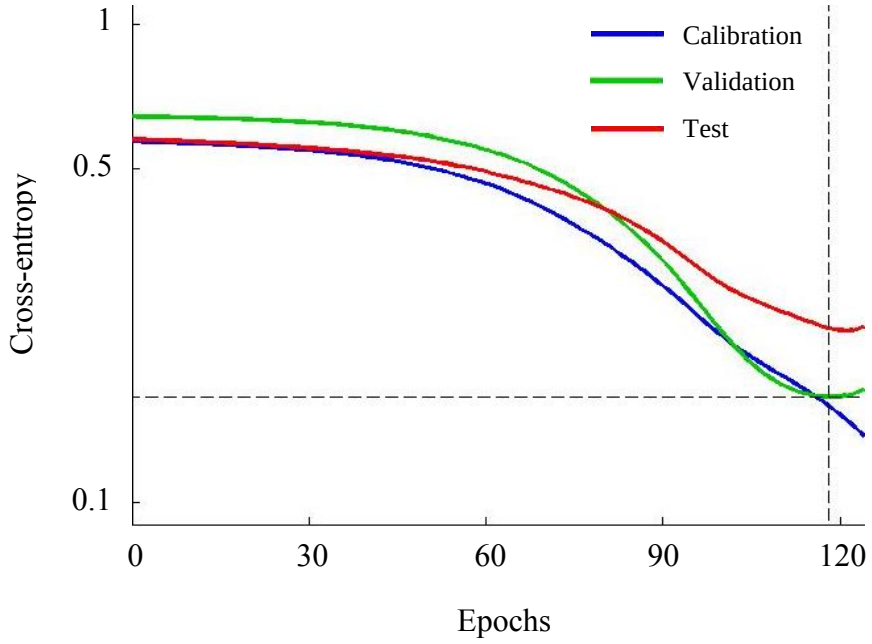

Figure 4.6 Artificial neural network performance progress of training, validation, and test data. Best validation performance is indicated by intersecting dashed lines.

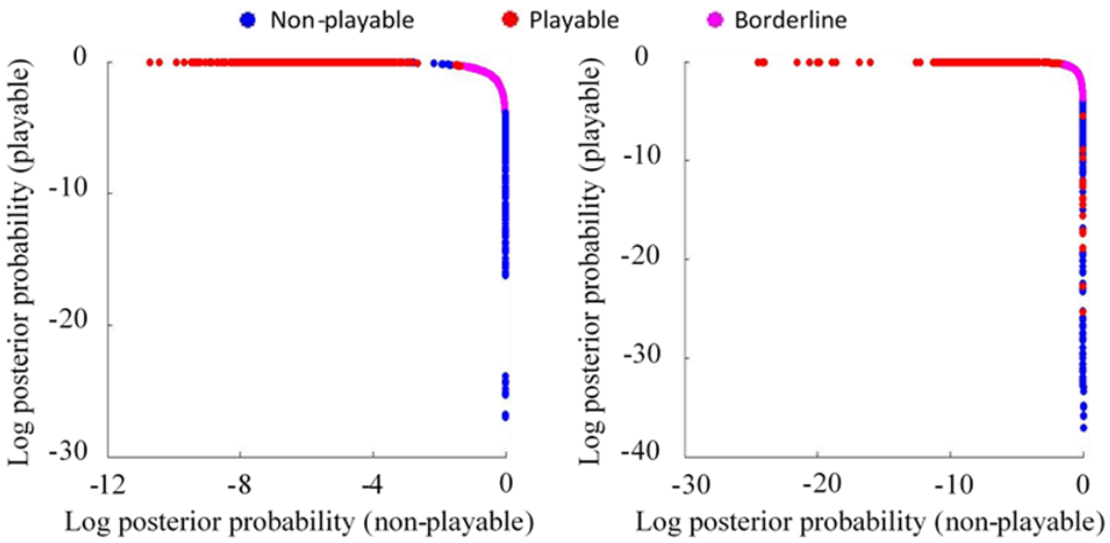

Figure 4.7 Log posterior probabilities of magnetic tapes included in the calibration (left) and test (right) sets belonging to playable and non-playable classes as determined by naïve Bayes classification. Blue and red markers indicate class determined by playability testing. Magenta markers indicate region of significant overlap of classes.

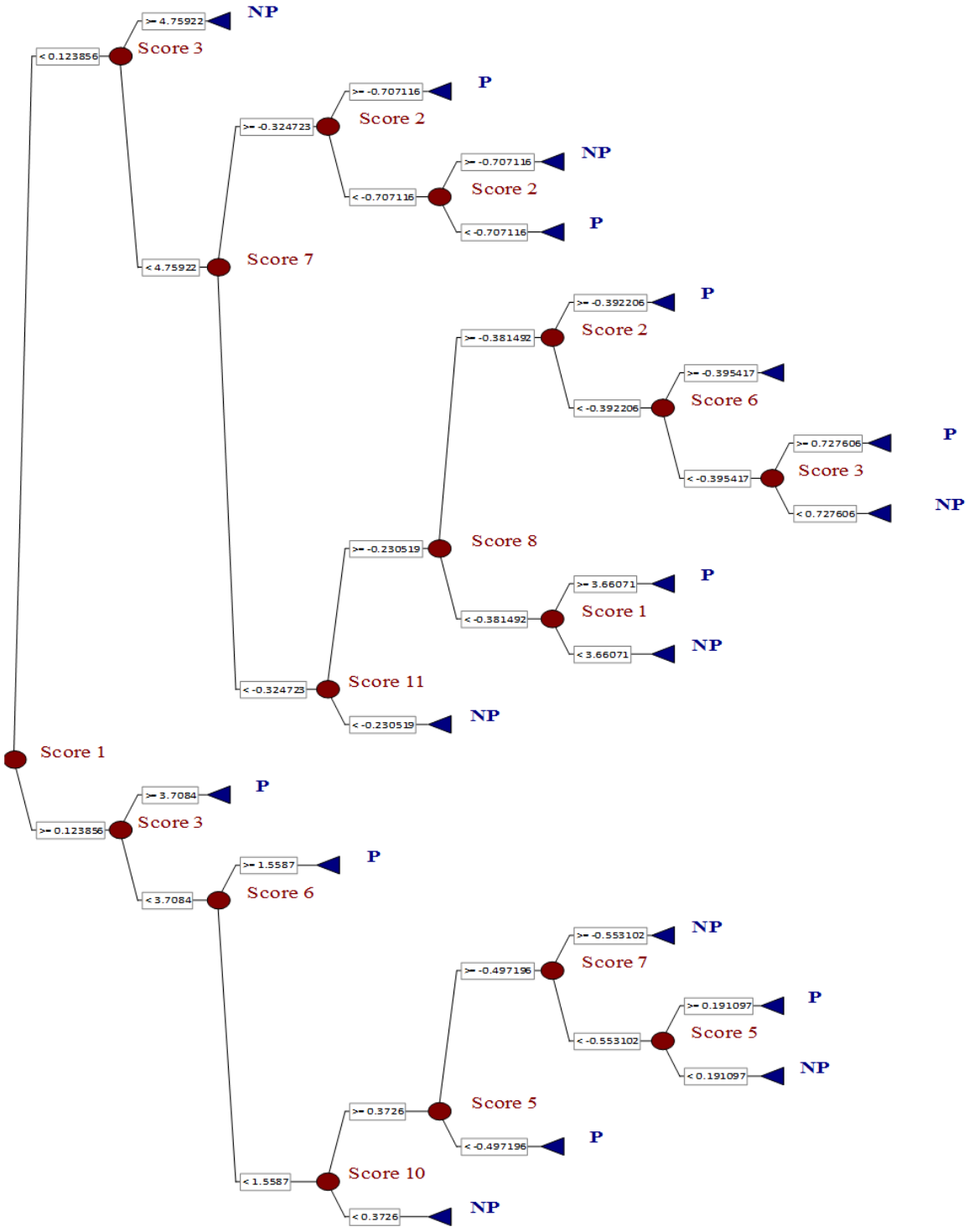

Figure 4.8 Decision tree used to classify playable (P) and non-playable (NP) magnetic tapes using scores from 11 latent variables selected by partial least squares discriminant analysis.

CHAPTER 5

MATSA: A USER FRIENDLY SOFTWARE PROGRAM FOR MAGNETIC AUDIO TAPE SPECTRA ANALYSIS

ABSTRACT

For both personal and professional purposes, magnetic audio tapes experienced widespread use for decades as a recording medium and are an invaluable source of cultural and historical information. Unfortunately, the data stored on magnetic audio tapes is at risk of being destroyed due to chemical deterioration. In recent years, a massive effort has been put in place by many cultural heritage institutions to transfer the media from the magnetic tapes to a digital format, thereby preserving the endangered information for future generations. The first step in the digitization process involves examining the collection of tapes and removing any degraded tapes which, during playback by an audio engineer, may be further impaired or cause damage to the tape recorder. A fast, objective, and non-destructive method of assessing a tape's playability is ideal, and previous work in our laboratory has shown that combining attenuated total reflectance Fourier transform infrared spectroscopy (ATR FT-IR) and multivariate statistics can be used for this purpose.

Here, we have developed a user-friendly software package which would enable tape custodians to visualize spectral data and determine the playability of magnetic audio tapes. Among the methods currently implemented in the Magnetic Audio Tape Spectral

Analysis (MATSA) program are two feature extraction techniques, principal component analysis (PCA) and partial least squares (PLS), in addition to six classification algorithms including artificial neural networks, decision trees, naïve Bayes classification, support vector machines (SVM), and linear and quadratic discriminant analysis. The program was built using a reference collection of ATR FT-IR spectra from 95 quarter-inch audio tapes. An internal validation accuracy of 93% was achieved after applying PCA-SVM classification to the reference data. Two separate test collections, one containing 50 tapes and another containing 31 tapes, were classified with 88% and 90% accuracies, respectively.

1. INTRODUCTION

Magnetic audio tapes earned widespread use as a recording medium beginning in the 1950s. Their reputation as being a durable and high-quality storage medium enabled magnetic tapes to remain popular for nearly a half century. The usage of magnetic tapes has been steadily declining recently, though many recording artists and engineers still prefer the sound of analog over digital for recording. At this time, manufactures have virtually halted the production of all high-end analog audio tape recorders, and collection managers are left with several decades' worth of cultural and historical information stored on magnetic audio tapes which are threatened by numerous mechanisms of deterioration.[1-6] As a result, many institutions are now converting analog recordings to the more stable digital medium.

The digitization process, which requires the tapes to be played, can be challenging for tapes produced from the 1970s to the 1990s. During this time, select manufacturers adjusted the formulation of magnetic recording tapes in order to provide a greater

dynamic range. These changes produced a new side-effect called sticky-shed syndrome (also referred to as sticky-tape or archival-shed syndrome). Sticky-shed syndrome is a condition in which the polyester-urethane binder, used to hold the magnetic oxide particles onto the polyester substrate, is subjected to hydrolysis due to humidity in the storage location.[7,8] Tapes displaying sticky-shed syndrome often make high-pitched noises when played, move slower in fast-forward mode compared to non-deteriorated tapes, and hinder movement of the reels when in rewind mode. More importantly, shedding of the magnetic layer onto the playback device can result in permanent data loss and downtime while the instrument is cleaned.

Fortunately for conservationists, the hydrolysis reaction responsible for the tape's deterioration is reversible. Therefore, if tapes are determined to have sticky-shed syndrome, a temporary restoration may be made prior to playback by 'baking' the tapes using a dehydrator or convection oven.[9] With collections of more than 46 million recording tapes in the U.S. alone,[10] the development of fast and nondestructive tools to assess magnetic tape collections is a priority in order to establish preservation strategies. To date, no such tools are available, and determinations on whether recording tapes should undergo restoration are often based on the ability of a tape machine to play the tapes.

Recent studies by Zelst[11] and Hobaica[12] suggest that sticky-shed syndrome can be detected without causing harm to the tape or recording instrument using attenuated total reflectance Fourier transform infrared spectroscopy (ATR FT-IR). In addition, by applying a decision-making algorithm based on multivariate statistical analysis such as quadratic discriminant analysis (QDA), tapes that are not likely to be playable by

common recording instruments can be separated from those tapes that are unlikely to cause significant complications when attempting to play.[13] Improvements in the accuracy of playability determinations can be gained using more sophisticated machine learning algorithms such as artificial neural networks (ANN) and support vector machine discriminant analysis (SMV-DA).

The objective of this work is to develop a user-friendly software application that tape custodians can apply to IR measurements to rapidly determine playability of tapes in a collection. The Magnetic Audio Tape Spectra Analysis (MATSA) program made for this purpose was developed from the graphical user interface development environment (GUIDE) in MATLAB, a high-level language and interactive environment used in the scientific and engineering communities. The MATSA program is available for free upon request by contacting the corresponding author. A general description of the application and the variety of options contained within MATSA are described in this paper. A more in-depth explanation of the software and algorithms contained therein can be found in the user's manual provided with the software.

2. PROGRAM DESCRIPTION

The MATSA program was created using GUIDE in MATLAB version 8.3. MATSA is comprised of subroutines within the statistics toolbox in MATLAB, in-house written functions, and other functions freely available on the Web. The current version of MATSA was programmed and tested on a system with a Windows operating system. Future releases of the program can be made available for other operating systems such as Macintosh and Linux after slight modifications.

2.1 Data importation

The MATSA title menu and main user interface in Figure 5.1 are accessed by running the 'MATSA_v1_1' m-file and importing data. Data is imported by following a series of commands issued after clicking anywhere on the title menu or through the use of the 'File' menu. To make MATSA fully functional, two types of datasets are imported. A reference dataset contains spectra from tapes with known playabilities determined through playability testing. Reference data is provided with the software but may be added to or replaced completely by the user. All samples of unknown playability are imported into MATSA as prediction data.

MATSA accepts data in Microsoft Excel format and requires three inputs. The first input requirement is a data array in which each row corresponds to a magnetic tape sample, and each column is a transmittance, T, or absorbance, A, measurement. To ensure that all reference and prediction data is comparable, any transmittance measurements that are input into MATSA are converted to absorbance either by $A = -\log(T)$ for fractional transmittance data or $A = -\log(T/100)$ for percent transmittance values. The second input must be a vector comprised of wavenumber (cm^{-1}) frequencies or nanometer (nm) wavelengths corresponding to the measurements in the input data array. By default, MATSA converts nm values to cm^{-1}. The final piece of required information is a two column list identifying names and numbers of samples measured. The list of identifiers allows the program to discern which samples in the data array correspond to different tapes and which correspond to replicate measurements of the same tape. These identifiers are also listed alongside the playability results in the final table that is made available by

MATSA, thus making it easy for the user to organize the results for their current collection of magnetic audio tapes.

2.2 Working with the graphical user interface

After one or more datasets have been loaded, the MATSA GUI main menu shown in Figure 5.1 will become available. From the main menu, the following functions can be performed:

1. View absorbance spectra of reference collection and samples to be predicted.
2. Edit examined frequency region.
3. Set classification parameters (dimensionality reduction method, number of factors to be retained, and prediction algorithm).
4. Determine quality of selected model through validation.
5. Obtain graphical and tabular representations of playability prediction results.

Each of these functions will be explained in greater detail in the following sections utilizing an example collection of data that was used to test the program during the development process. See reference 13 for a description of the datasets, including the methodologies for ATR FT-IR spectra collection and tape recorder playability testing. The reference data contained 1900 spectra from 95 (20 replicate spectra collected per tape) magnetic audio tapes. Only one spectrum per tape (50 total spectra) was used for a 50 tape collection of tapes with unknown playabilities.

2.2.1 Spectral visualization and preprocessing

Raw and preprocessed absorbance spectra can be obtained using the 'Visualization' pop-up menu. For simplicity, data preprocessing in MATSA is currently a

four-step static process. First, a cubic spline interpolation of the data is carried out using the 'interp1' function in MATLAB.[14] Using this function, new absorbance measurements are predicted at given wavenumber values (equally spaced increments of 1 cm^{-1}). The result is all reference and prediction data sharing a common frequency axis regardless of the original intervals used during data collection. The interpolated signal is smoothed further using the Savitzky-Golay method (4^{th} order polynomial, 5 point moving window).[15,16] Next, multiplicative interference from scatter and particle size is removed using a standard normal variate (SNV) transformation.[17] Finally, each column in the data array is mean centered before performing one of the dimensionality reduction methods in MATSA. It should be noted that the user also has the option to not perform any processing beyond interpolation. This makes it possible to perform data analysis on spectra preprocessed using alternative software.

Figure 5.2 shows ATR FT-IR spectra of magnetic audio tapes before and after preprocessing. Changes occurring in the ATR FT-IR spectrum of aging polyester-urethanes have been described on multiple occasions. In general, non-playable tapes will show very strong absorbance around 1730 cm^{-1} due to the C=O stretching. The strength of this peak is lower in tapes that have not undergone significant degradation. Characteristic bands in spectra of non-playable tapes are located near 1255 cm^{-1} and 1140 cm^{-1}, corresponding to the respective C-O and O-H stretching vibrations likely due to carboxylic acid formation as a result of hydrolysis. Non-degraded tapes typically display peaks due to saturated ester stretching from polyester near 1210 cm^{-1} (C-C-O) and 1050 cm^{-1} (O-C-C) which are not present in degraded samples. With degradation being a continuous process, it can be difficult to interpret numerous mean centered spectra

simultaneously on a single panel in MATSA. It is, perhaps, easier to select spectra which represent the extremes of the reference dataset, and compare a single spectrum of interest to those extremes to determine the extent of degradation of a particular tape.

2.2.2 Input parameters

When new frequencies are specified by the user in the 'input parameters' section on the GUI, MATSA searches the features array of all available datasets, finds the values nearest to the user-specified values, and truncates the data at those points. Any change in the frequency region is reflected in all subsequent MATSA plots and calculations. In addition to the editable frequency boxes, the respective numbers of samples in the reference and prediction datasets are displayed. These values are determined using the list of identifiers imported by the user.

2.2.3 Prediction model algorithms

Before playability predictions can be made in MATSA, the user must specify the algorithms used by the program to reduce data dimensionality and perform classifications. Several benefits are gained by reducing the dimensionality of ATR FT-IR data prior to analysis. Feature selection or extraction methods are often effective ways of dealing with the "curse of dimensionality," which states that the number of samples required for statistical significance increases exponentially with the number of dimensions.[18] In addition, dimensionality reduction can be used to reduce noise, reduce computational cost, and remove correlated variables, all of which can negatively influence the outcome of classification models.

The GUI allows for feature extraction by principal component analysis (PCA) or partial least squares (PLS). In PCA, the original wavenumber variables are reduced to a

new set of factors, called principal components (PCs).[19,20] Each PC consists of scores, transformed variable values of each data point, and loadings, weights used to multiply each variable to get the component score. The scores are orthogonal and therefore, by definition, are well-conditioned. Moreover, the first PC describes the direction of maximum variance, provided the input data is mean centered.

Plots of the scores and loadings from the first two PCs representing 77.7% of the total variation in the reference dataset are depicted in Figure 5.3. The scores plot shows the clustering of spectra from non-degraded audio tapes and their separation from spectra associated with degraded samples of mostly non-playable tapes. Spectra from playable tapes located on the left side of the scores plot are an indication that tapes can be degraded without showing sticky-shed syndrome to the extent that would cause difficulty when playing. Linear discriminant analysis (LDA) was used to draw a linear decision boundary which best separated the two classes in two-dimensional space. The boundary drawn in Figure 5.3 separates the reference samples into the two playability classes with 88.3% (1,678 of 1,900 spectra) accuracy. The chemical information captured by the first two PCs is shown in the loadings plots. The first PC captures the C=O stretching vibration near 1730 cm^{-1} which, as mentioned previously, increases as the tapes degrade. The largest contributor to the second PC is the ester C-O-C stretching vibration band near 1170 cm^{-1} which shifts to higher frequencies as the polyester-urethane binder deteriorates.

PC scores and loadings are useful for describing the variance in the absorbance spectra but do not necessarily capture all relevant information for predicting playability. A solution to this problem can be found using PLS. PLS is similar to PCA, but the

factors, called latent variables (LVs), are created in such a way as to maximize the amount of variation in the spectra that is relevant for predicting the tape's playability.[21] This typically means that fewer factors are needed for PLS models compared to PCA models.[22]

Because no single technique works best for all possible tasks and forms of data, there are currently six supervised classification algorithms available in MATSA to predict playability of magnetic tapes after performing PCA or PLS on ATR FT-IR data. The algorithms included are naïve Bayes classification (NBC), decision trees (DT), LDA, QDA, ANN, and SMV-DA. Three of the six algorithms, LDA, QDA, and NBC have an explicit underlying statistical model which provides the probability that each tape is either playable or non-playable. Along with decision trees, a logic-based algorithm, these methods have the advantages of being easily interpretable, less prone to overfitting, and requiring less computation when compared to more sophisticated techniques such as ANN and SMV-DA. However, ANN and SMV-DA remain valuable tools, and are included in this program because of their tendency to perform better on large datasets with continuous features. Due to the popularity of the six supervised methods just mentioned, many detailed descriptions of how these techniques carry out classification tasks are available.[23-25] Also, a comprehensive review of the strengths and weaknesses of each technique has been provided by Kotsiantis.[26]

MATSA is capable of running within the MATLAB environment in which it was created or as a stand-alone application. Both versions include all six of the previously described prediction algorithms. However, due to the current capabilities of the MATLAB Compiler used to create the stand-alone form of MATSA, the ability to train

ANN and DT models is not included in the edition that runs outside of MATLAB. Instead, the stand-alone version uses pre-trained ANN and DT models based on the default settings (five PCs) in addition to the other four fully functional prediction algorithms of which models can be built and run in MATSA.

2.2.4 Prediction model validation

To examine the potential performance of the user selected prediction method on new magnetic audio tapes, internal validation of the reference dataset is applied by clicking on the 'Calculate' button on the main user interface. The hold-out cross-validation method used in MATSA is performed by temporarily removing ~40% of the reference data. An attempt is then made to classify the validation samples, which are selected in a stratified manner according to the tape identifiers. The user can use the hold-out method to determine a 'good' combination of parameters to use for prediction, or as a measure of the default model's ability to predict new samples. The default parameters were selected as a result of the work carried out in reference 13. In that work, only the first five PCs were needed to build a classification model using QDA.

Alternatively, an appropriate model can be selected using the 'Autoselect' button on the main window. The critical first step of the automated model selection process is to determine the number of PCs or LVs that should be used to build the classification models. The goal of this step is to include all factors which are useful for classifying the tapes, but leave out those which capture spectral noise. The inclusion of noise during model training can lead to incorrect predictions of unknown samples. For PCs, this is determined using Joliffe's modification of the Guttman-Kaiser criterion, which retains only those PCs with an eigenvalue greater than the product of 0.7 and the average of all

eigenvalues generated.[20] For PLS, the number of optimum number of LVs was selected as the number which correctly classified the most samples in the reference set using a minimum Mahalanobis distance classifier.[27,28] The 'Autoselect' algorithm then cycles through 20 iterations of hold-out cross-validation with each prediction method using only the selected factors. The technique with the highest average classification accuracy is displayed by MATSA and is suggested for predicting playability of new data.

SVM classification using 16 PCs was selected by MATSA after using the 'Autoselect' function on the 95 tape reference collection. The average hold-out cross-validation accuracy for the model is displayed in the GUI shown in Figure 5.4. Also displayed in the figure is a box-and-whisker plot showing the distribution of results obtained by each prediction method during the iterative cross-validation process. Here, the box edges represent the 25th and 75th percentile, the median of each group is shown using a horizontal green line, and black dashed lines (whiskers) extend to the most extreme points not considered outliers. Those points considered outliers such as seen in the ANN and DT results, have red markings. In most instances where the 'Autoselect' feature was used on the example dataset to select an appropriate model for prediction, less variance in the validation results was seen with LDA and SVM classification. ANN, DT, and NBC typically showed higher variance in the results and tended to be more prone to outliers.

2.2.5 Playability predictions

There are multiple ways to view playability of magnetic audio tapes in MATSA. This can be done graphically, by viewing PCA or PLS scores of samples with unknown playabilities. Figure 5.5 shows the tapes to be predicted in the space of the first two LVs

calculated from the reference data. Projected with the prediction samples, is the linear decision boundary used by LDA to classify the reference samples using only the first two PCs. The user has the option of plotting a quadratic decision boundary, as opposed to the linear boundary displayed, using QDA. Using a built-in pop-up menu, a specific tape can be selected and highlighted to view its location in PC or LV space. Here, the first recording tape is highlighted, and by recalling the scores plot of the reference samples in Figure 5.3, it can be seen that the ATR FT-IR absorbance spectrum of tape number one more closely resembles the spectra of the non-playable reference tapes.

Because prediction will often be accomplished using an alternative prediction algorithm or occur in a higher dimension, there are no guarantees that the graphical representations of playability will match the final prediction results given by MATSA. The final predictions made using the specified dimensionality reduction and prediction techniques are given in tabular form (example shown in Figure 5.5). Here, each sample name and number is associated with a color-labeled playability prediction. A 'green' prediction is indicative of a recording tape unlikely to cause significant problems when attempting to play using a tape machine. To ensure the tape may be safely copied to a different format without causing harm to the recording instrument or losing valuable information, temporary restoration should be considered for any sample given a 'red' prediction.

The predictions of the 50 tape test collection generated by PCA-SMV-DA in MATSA were compared to the predetermined designations resulting from physically attempting to play the tapes. Of the 50 tapes, 44 (88%) were classified correctly. In only one instance, a playable tape was falsely classified as non-playable. The remaining five

tapes were classified as non-playable after being previously designated as playable. It should be noted that all six misclassified tapes came from the same institution from which the reference data originated. No misclassifications were found in the collection of tapes donated from the second institution.

2.2.6 Dealing with the cost of misclassifications

Because of the potential value of the information contained on magnetic audio tapes, it will often be wise to error on the side of caution and try to avoid falsely classifying non-playable tapes as being playable. This comes with the obvious drawback that more playable tapes will be incorrectly classified as being non-playable, resulting in more tapes which need to be restored before digitization. However, this is typically not as big of a concern as the data loss or instrument maintenance issues that could result from attempting to play a sticky tape. By specifying 'cost' parameters when using statistical learning algorithms in MATSA, the probability of getting a particular result can be increased or decreased.

Shown in Figure 5.6 is a comparison of the LDA decision boundaries used to classify samples of the reference dataset in the space of the first three PCs with a non-playable to playable cost of 1:1 versus a cost of 1:0.2 (i.e., the cost of classifying a non-playable tape as playable is five times greater than the cost of classifying a playable tape as non-playable). Three-dimensional plots are fully rotatable in MATSA using the 'Make Interactive' button on the user interface. Here, clustered samples representing replicate spectra from two tapes, which were nearly equidistant from the centroids of the two groups, switched playability designations when the cost of misclassifying non-playable samples was increased. Because the two tapes were assigned to different classes initially

using playability testing, the overall change in classification accuracy was minimal. As the cost parameter was altered from 1:1 to 1:0.2, the classification accuracy of the reference samples only increased from 87.8% to 88.7%. However, a 35.8% reduction in the number of reference spectra that were falsely classified as being playable was achieved. The 50 tape test data was largely unaffected after performing the same analysis using LDA, as only one of the 50 tapes was given a new playability designation after changing the cost from 1:1 to 1:0.2. Overall, when compared to the SMV-DA results mentioned in the last section, LDA with a cost of 1:02 had a 6% (3 tapes) lower accuracy with each of those tapes being falsely classified as non-playable.

3. ADDITIONAL PERFORMANCE TESTING

A collection of 41 tapes from various locations was obtained and analyzed using ATR FT-IR to further test the ability of MATSA to predict tape playability. Unlike the 145 tape collection used to build the MATSA program, this test collection was comprised of tapes of which the brand and model were known and available to use as part of the data exploration process. Exploratory analysis in the form of agglomerative hierarchical cluster analysis (AHCA) was performed using PLS_Toolbox (Eigenvector Research Inc., Manson, WA) software to determine if the spectra from the 41 tape assorted collection showed similar spectral characteristics with the 95 polyester-urethane tapes used as the default training model in MATSA. The AHCA method starts with all spectra apart (i.e., the number of clusters equals the total number of spectra). The clusters progressively merge based on a user-selected criterion until all samples are in a single cluster.[29,30] In this work, the AHCA linkage criterion is Ward's method which minimizes the total variance within each cluster.[31]

For simplification, the preprocessed replicate spectra of each tape were averaged before performing AHCA, and the tapes were identified according to the manufacturer and model number. The dendrogram resulting from AHCA on the averaged spectra is shown in Figure 5.8. Immediately noticeable is the similarity between tapes with the same manufacturer and model number. This suggests the slight variations in the formulations of the binder used between the manufacturers can be detected in the IR region being studied. The dendrogram is also indicative of two very distinct groups (labeled one and two on the left side of the figure) being present in this collection.

To study the differences between the clusters, spectra from the ten tapes in cluster one were averaged and compared to the average spectrum calculated from the 31 tapes in the second cluster. This comparison between spectra can be seen in Figure 5.9. It is clear that the spectra in the second cluster more closely resemble the spectra from the 95 polyester-urethane tapes which were used to build the classification models in MATLAB. Therefore, useful predictions could only be made on the tapes in the second cluster. Many of the tape brands located in cluster one were known to be introduced in the early 1960s, before the switch made by many manufactures to use polyester-urethane as the binder, in an era where acetate- and polyvinyl chloride-based binders were common. To predict the playability of tapes in cluster one, it would be necessary to create a new model using tapes with similar formulations as the current classifier is likely to place all of these tapes into one class, and any accuracy obtained would be based completely on chance.

To assess the prediction ability of the MATSA program on the tapes in cluster two, the playability of each tape was predetermined by attempting to play the tape with a tape machine. The same 95 tape model and automatically selected method of prediction

(SMV-DA with 16 PCs) discussed in previous sections were used to classify each replicate spectrum (20 replicates per tape) of the 31 tapes. Of 620 total spectra, 559 spectra were correctly associated with the predetermined playabilities of the tapes resulting in a classification accuracy for this dataset of 90.16%. With no cost parameter adjustment, MATSA falsely classified non-playable tape spectra as belonging to playable tapes 3.39% of the time. Playable tape spectra were falsely classified as belonging to non-playable tapes 6.45% of the time.

4. CONCLUSION

A stand-alone software application has been presented for determining the playability of magnetic audio tapes following ATR FT-IR spectroscopy, an instrumental analysis technique already used by many cultural heritage institutions to identify organic compounds and polymeric materials. The easy-to-use interface gives tape custodians the ability to determine, within minutes after spectra collection, if tapes in a lot are immediately ready to be digitized or should be subjected to restoration. MATSA is a flexible program allowing data with differing numbers of features and observations to be used for reference and prediction. Therefore, six different prediction algorithms are included in the program to help achieve a quality separation between data from playable and non-playable tapes. Additional flexibility is available using the cost parameter which enables one to adjust the decision boundaries between classes, thus making it less likely for an undesired outcome (i.e., predicting a tape is playable when it is not) to occur. Visualization tools such as decision boundaries in the space of PC or LV scores, loadings plots, and spectral plots, all make it possible to conveniently explore the chemical differences between samples which resulted in the prediction given by the program.

The MATSA program was built and tested using ATR FT-IR data collected from over 250 quarter-inch reel-to-reel audio tapes stored at four separate institutions. Quarter-inch tapes were selected as the desired format for analysis in MATSA after a survey disseminated to over 50 archives, museums, and libraries in the U.S. identified quarter-inch tapes as the most common format in need of restoration. Exploratory analysis on ATR FT-IR data from multiple collections of quarter-inch tapes suggests that predictions made in MATSA can only be reliable when the brand of the prediction tapes are known and placed into models comprising only tapes with similar formulations. However, because manufacturers sometimes change the model number without making significant changes to the formulation of the tape, one could suspect that in many instances a single classification model could be valuable for multiple tape model numbers.

ACKNOWLEDGMENTS

This project was supported by Grant LG-06-12-056912 from the Institute of Museum and Library Services to the University of South Carolina. Portions of this work were presented at the Great Scientific Exchange (SciX) conference at the Rhode Island Convention Center in Providence, Rhode Island on September 29, 2015. In addition, the work presented here contributed to a published article in Against the Grain (volume 27, number 4) entitled, “Magnetic Tapes, Playable or Not?” The co-authorship of Brianna M. Cassidy, Zhenyu Lu, Michael L. Myrick, Eric M. Breitung, and Stephen L. Morgan is acknowledged. The authors are grateful for support from Heather M. Heckman, Lydia C. Pappas, and Gregory J. Wilsbacher at Moving Image Research Collections at the University of South Carolina, and Gene DeAnna, Larry Miller, Fenella France, and Linchi Nguyen (Summer 2013, ACS Project SEED research participant) at the Library of

Congress. The contributions made by University of South Carolina undergraduate researchers, Samantha E. Skelton and Eric J. Bringley are also recognized.

REFERENCES

1. Eilers, D. Polyester and Acetate as Magnetic Tape Backings*, J. Audio Engng. Soc*. **1969**, 17, 303-308.

2. Volkmann, H. Film Preservation: a Report of the Preservation Committee of the International Federation of Film Archives, *FIAF* **1965**, 7.

3. Cuddihy, E. Aging of Magnetic Recording Tape, *IEEE Trans. Magn.* **1980**, 16, 558-568.

4. Bertram, H.; Cuddihy, E. Kinetics of the Humid Aging of Magnetic Recording Tape. *IEEE Trans. Magn.* **1982**, *18*, 993-999.

5. Edge, M.; Allen, S.; Hayes, M.; Jewitt, T.; Horie, C.; and Brems, K. Degradation of Magnetic Tape: Support and Binder Stability, *Polym. Degrad. Stabil.* **1993**, 39, 207-214.

6. Thiébaut, B.; Vilmont, L-B.; Lavédrine, B. Characterization of U-matic Videotape Deterioration by Size Exclusion Chromatography and Pyrolysis Gas Chromatography/Mass Spectrometry and the Role of Adipic Acid, *J. Cultural Heritage* **2009**, 10, 183-197.

7. Smith, L. Factors Governing the Long-Term Stability of Polyester-Based Recording Media, *Restaurator* **1991**, 12, 201-218.

8. Hess R. Tape Degradation Factors and Challenges Predicting Tape Life, *J. Assoc. Rec. Sound Collect.* **2008**, 39, 240-274.

9. Richardson C. Process for Restoring Magnetic Recording Tape Damaged by "Sticky Shed" Syndrome. United States Patent No. 6797072 B1 (28 Sept. 2004).

10. *A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections*, Heritage Preservation, Inc., Washington, D.C., 2005.

11. Zelst, L. *Sticky Shed Sydrome: Characterization, Diagnosis, and Treatment,* Internal Library of Congress Report, March, 2008.

12. Hobaica, S. Analysis of Audio Magnetic Tapes with Sticky Shed Syndrome by ATR FT-IR, *J. Appl. Polym. Sci.* **2013**, *128*, 1962-1973.

13. Cassidy, B.; Lu, Z.; Fuenffinger, N.; Skelton, S.; Bringley, E.; Nguyen, L.; Myrick, M.; Breitung, E.; Morgan, S. Rapid and Minimally Invasive Identification of Degraded Polyester-Urethane Magnetic Tape Using Attenuated Total Reflection Fourier Transform Infrared (ATR FT-IR) Spectroscopy and Multivariate Statistics, *Anal. Chem.* **2015**, 87, 9265-9272.

14. Boor, C. *A Practical Guide to Splines*; Springer-Verlag: Germany, 1978.

15. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedure, *Anal. Chem.* **1964**, 36, 1627–39.

16. Steiner, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Squares Procedure. *Anal. Chem.* **1972**, 44, 1906-1909.

17. Barnes, R.; Dhanoa, M.; Lister, S. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, *Appl. Spectrosc.* **1989**, 43, 772-777.

18. Bellman, R. *Adaptive Control Processes: A Guided Tour*; Princeton University Press: New Jersey, 1961.

19. Gemperline, P. Principal Component Analysis. In *Practical Guide to Chemometrics,* 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006. pp 69-104.

20. Joliffe, I. *Principal Component Analysis,* 2nd ed.; Springer Series in Statistics; Springer: New York, 2002.

21. Barker, M.; Rayens, M. Partial Least Squares for Discrimination, *J. Chemom*. **2003**, 17, 166-173.

22. Kalivas, J.; Gemperline, J. Calibration. In *Practical Guide to Chemometrics,* 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006, pp 105-162.

23. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction;* Springer Series in Statistics; Springer-Verlag: New York, 2001.

24. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 3rd ed.; Academic Press: California, 2006.

25. Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, U.K., 2012.

26. Kotsiantis, S. Supervised Machine Learning: a Review of Classification Techniques, *Informatica* **2007**, 31, 249-268.

27. Mahalanobis, P. On the Generalized Distance in Statistics, *Proc. Nat. Instit. Sci. India* **1936**, 2, 49-55.

28. De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. The Mahalanobis Distance, *Chemom. Intell. Lab. Syst.* **2000**, 50, 1-18.

29. Kaufman, L.; Rousseeuw, P. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley Series in Probability and Mathematical Statistics; Wiley: New York, 1990, pp 199-252.

30. Jain, A. K.; Dubes, R. C. *Algorithms for Clustering Data*; Prentice Hall Advanced Reference Series; Prentice Hall: New Jersey, 1988, pp 58-89.

31. Ward, J. Hierarchical Grouping to Optimize an Objective Function, *J. Amer. Statist. Assoc.* **1963**, 58, 236-244.

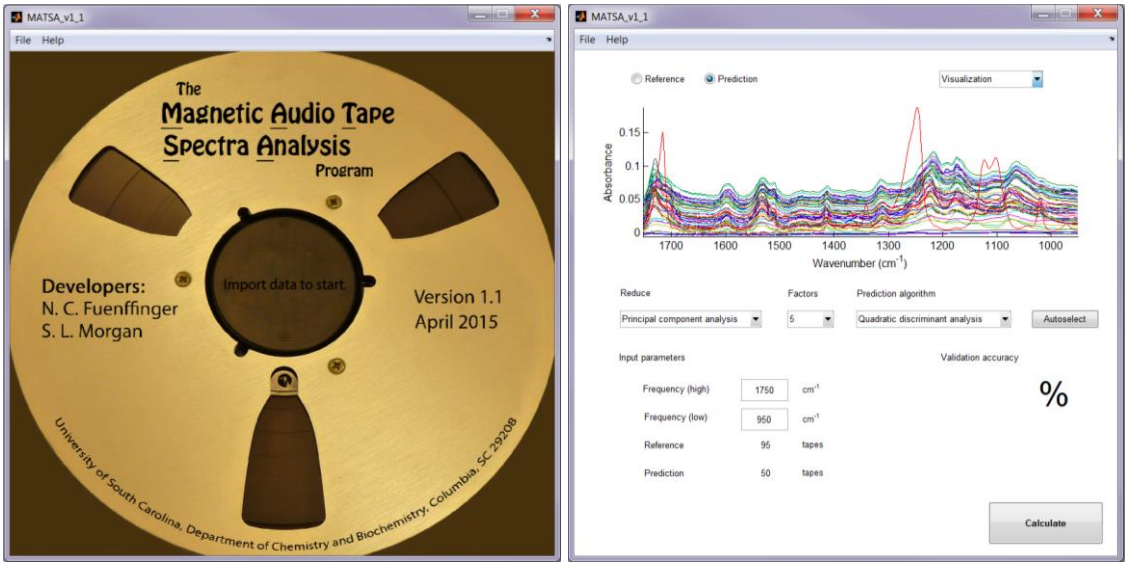

Figure 5.1 MATSA program title window (left) and main user interface (right).

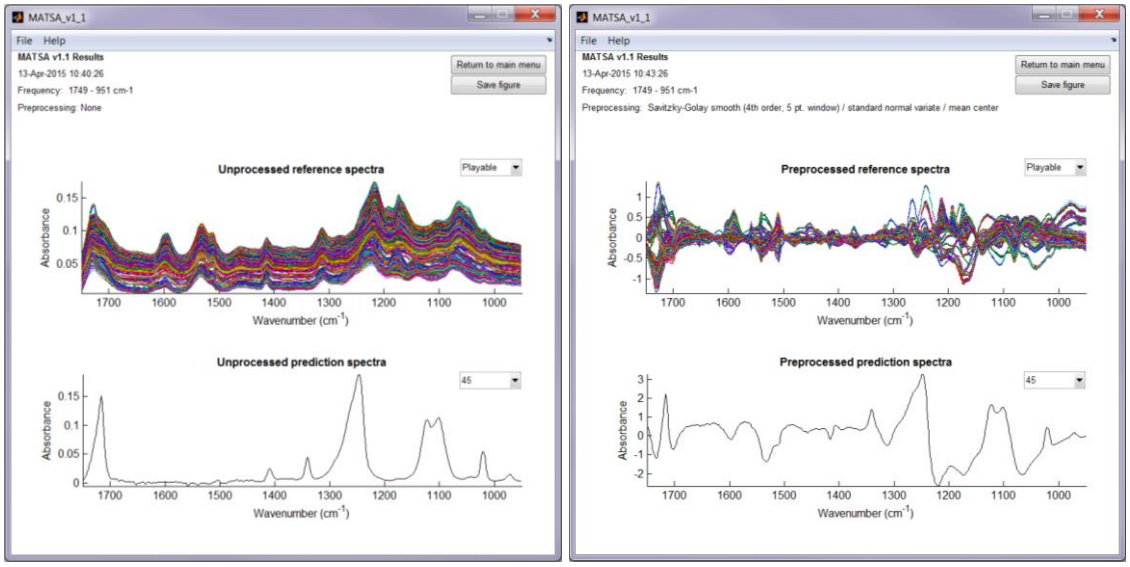

Figure 5.2 MATSA spectral plot window showing examples of magnetic audio tape raw absorbance spectra (left) and spectra preprocessed using interpolation, Savitzky-Golay smoothing, standard normal variate transform, and mean centering (right).

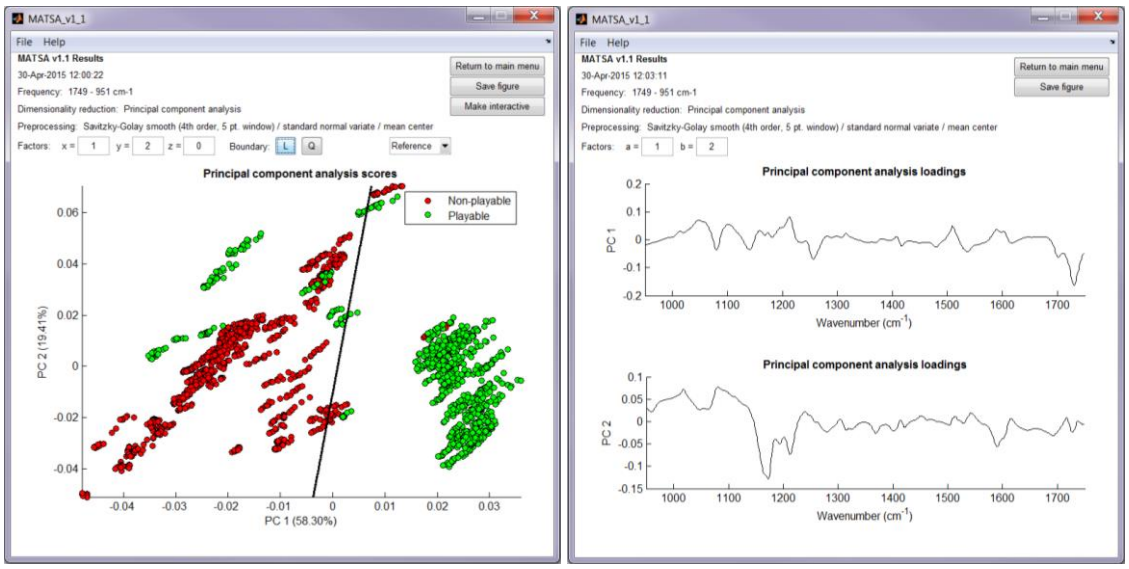

Figure 5.3 MATSA scores (left) and loadings (right) plots resulting from feature extraction by principal component analysis.

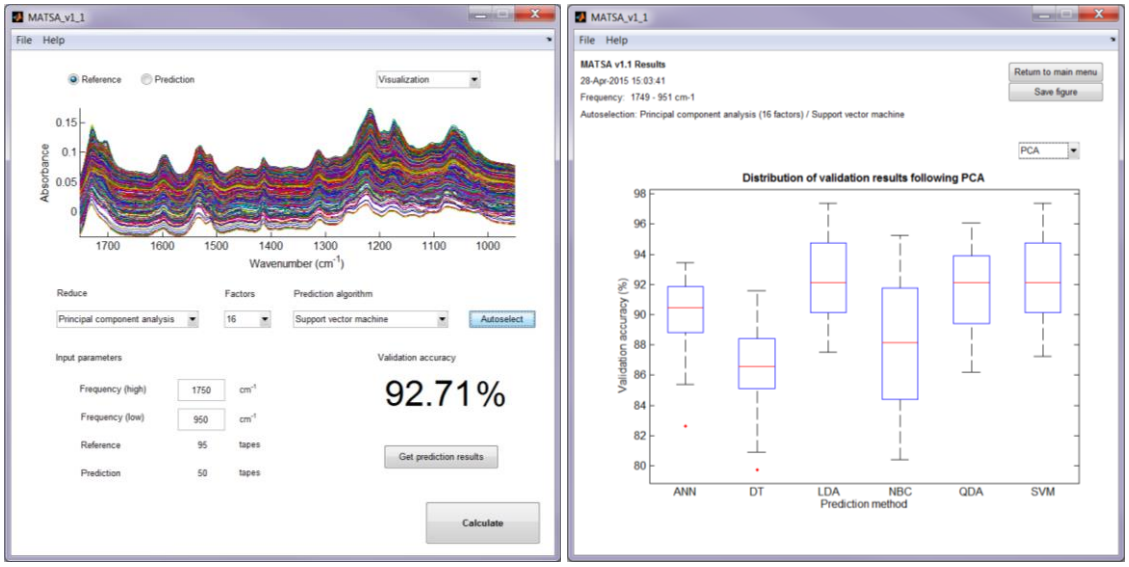

Figure 5.4 MATSA displays of validation accuracy with newly selected algorithms (left) and validation results for all algorithms based on the number of input latent variables from partial least squares (right).

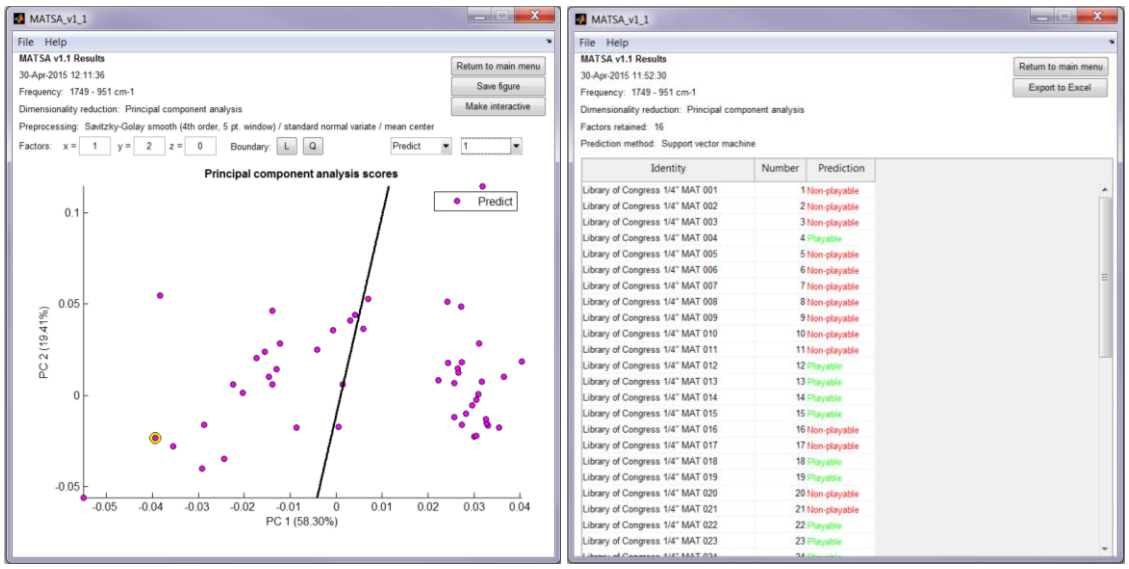

Figure 5.5 MATSA displays showing samples to be predicted in the space of first two latent variables with linear boundary separating playable tapes from non-playable tapes (left) and playability predictions for each sample in tabular form (right).

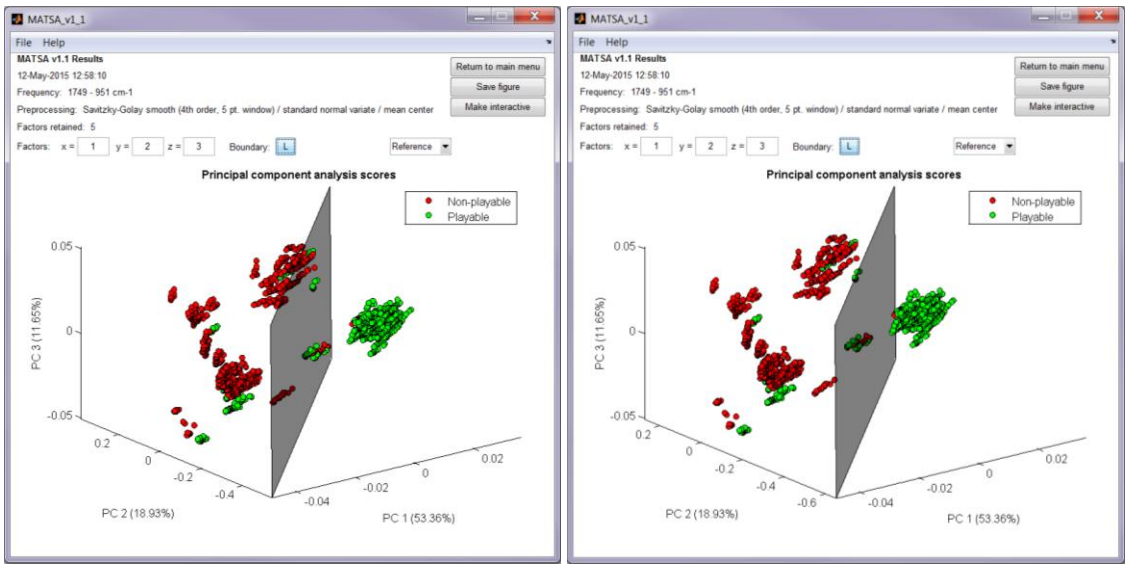

Figure 5.6 MATSA three-dimensional plots showing decision planes resulting from linear discriminant analysis on the principal component scores with an equal cost of misclassifying tapes (left) and with a higher cost of falsely classifying non-playable tapes as playable (right).

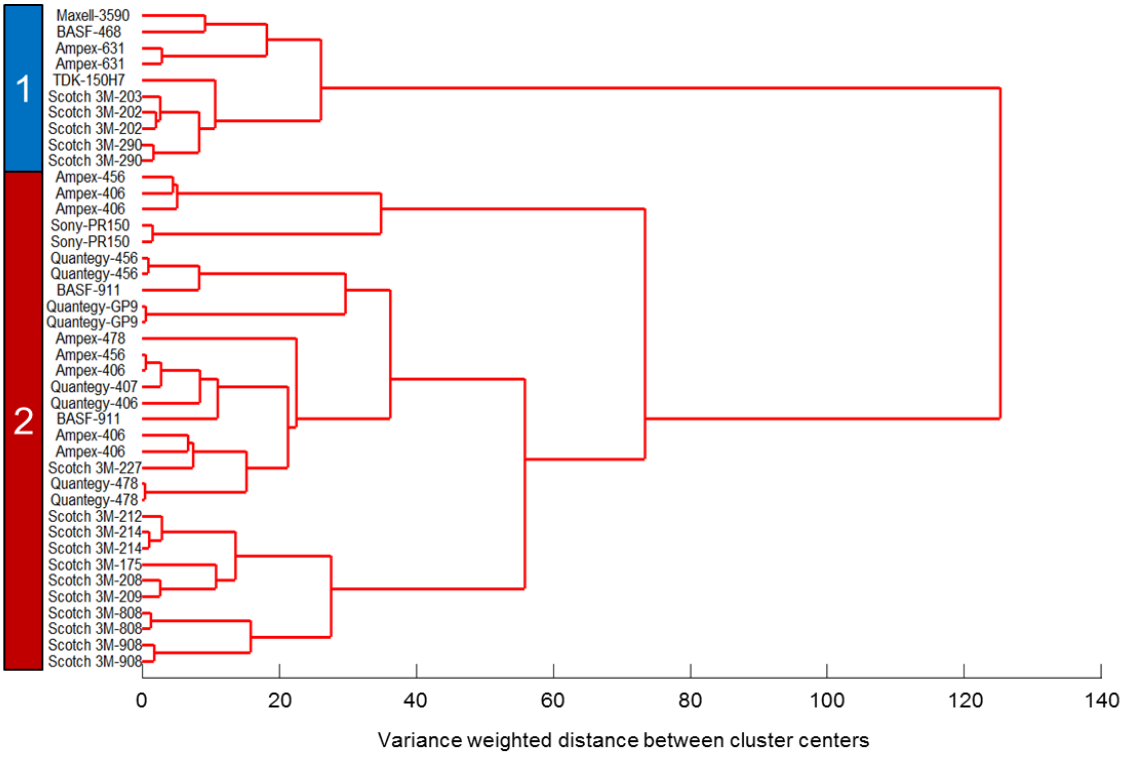

Figure 5.7 Dendrogram showing two clusters from hierarchical cluster analysis using Ward's linkage on the average ATR FT-IR absorbance spectra from 41 tapes in an assorted collection listed by brand and model number.

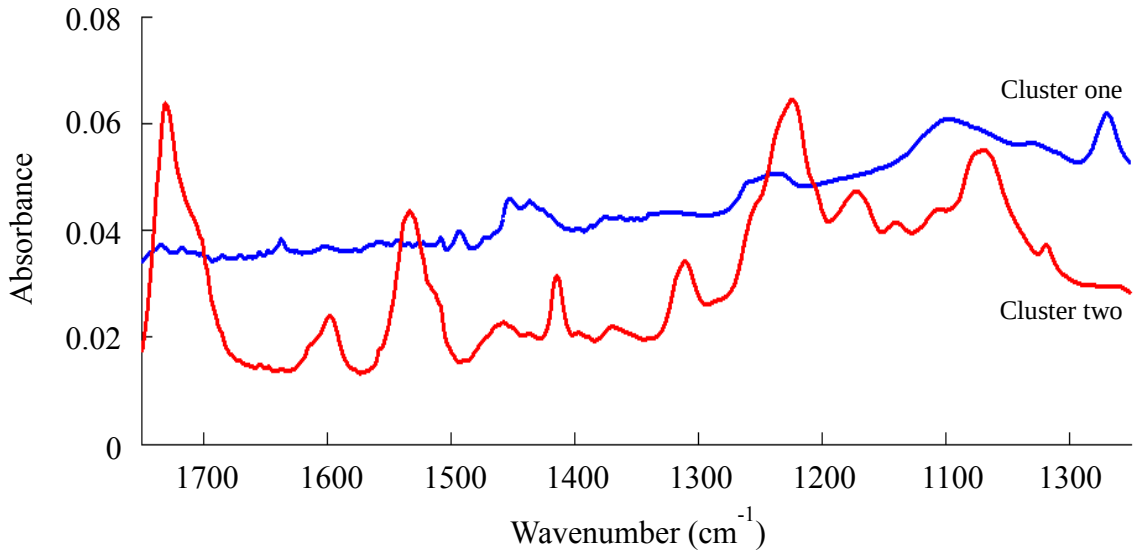

Figure 5.8 Average spectra of two clusters resulting from hierarchical cluster analysis.

BIBLIOGRAPHY

A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections, Heritage Preservation, Inc., Washington, D.C., 2005.

Adolf, P.; Dunlop, J. Microspectrophotometry/Colour Measurement. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K. 1999, pp. 251-289.

Almeida, V.; Vargas, A.; Garcia, J.; Lenzi, E.; Oliveira, C.; Nozaki, J. Simultaneous Determination of the Textile Dyes in Industrial Effluents by the First-Order Derivative Spectrophotometry, *Anal. Sci.* **2009**, 25, 487-492.

Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, U.K., 2012.

Barker, M.; Rayens, M. Partial Least Squares for Discrimination, *J. Chemom*. **2003**, 17, 166-173.

Barnes, R.; Dhanoa, M.; and Lister, S. Standard Normal Variate Transformation and De-trending of Near-infrared Diffuse Reflectance Spectra. *Appl. Spectrsoc.* **1989**, 43, 772-777.

Beebe, K.; Pell, R.; Seasholtz, M. *Chemometrics: a Practical Guide*; John Wiley & Sons: New York, 1998.

Bellman, R. *Adaptive Control Processes: A Guided Tour*; Princeton University Press: New Jersey, 1961.

Bertram, H.; Cuddihy, E. Kinetics of the Humid Aging of Magnetic Recording Tape. *IEEE Trans. Magn.* **1982**, *18*, 993-999.

Biancolillo, A.; Bucci, R.; Magri, A. L.; Magri, A. D.; Marini, F. Data-Fusion for Multiplatform Characterization of an Italian Craft Beer at its Authentication, *Anal. Chim. Acta* **2014**, 820, 23-31.

Biermann, T.; Grieve, M. A Computerized Data Base of Mail Order Garments: a Contribution Toward Estimating the Frequency of Fibre Types Found in Clothing. Part 2: the Content of the Data Bank and its Statistical Evaluation, *Forensic Sci. Int.* **1996**, 77, 75-71.

Blank, T.; Sum, S.; Brown, S. Transfer of Near-infrared Multivariate Calibrations without Standards. *Anal. Chem.* **1996**, 68, 2987-2995.

Boor, C. *A Practical Guide to Splines*; Springer-Verlag: Germany, 1978.

Bosch Ojeda, C.; Sanchez Rojas, F. Recent Applications of Derivative Ultraviolet/Visible Absorption Spectrophotometry: 2009-2011 a Review, *Microchem. J.* **2013**, 106, 1-16.

Boser, B.; Guyon, I.; Vapnik, V. A Training Algorithm for Optimal Margin Classifiers, *COLT* **1992**, 5, 144-152.

Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Wadsworth International Group: California, 1984.

Breitung, E.; Skelton, S.; Morgan, S. *Rapid Identification of Sticky Shed Syndrome in Magnetic Tape Using ATR-FTIR and Multivariate Statistics*, 40th Annual American Institute of Conservation Meeting, Albuquerque, New Mexico, 2012.

Brereton, R. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: West Sussex, U.K., 2003, p 238.

Bridge, T.; Wardman, R.; Fell, A. Novel Digital Methods for the Qualitative Characterization of Some Acid Dyes Applied to Wool and Nylon, *Analyst* **1985**, 110, 1307-1312.

Broseus, J.; Vallat, M.; Esseiva, P. Multi-class Differentiation of Cannabis Seedlings in a Forensic Context. *Chemom. Intell. Lab. Syst.* **2011**, 107, 343-350.

Bueno, J.; Sikirzhytski, V.; Lednev, I. Raman Spectroscopic Analysis of Gunshot Residue Offering Great Potential for Caliber Differentiation. *Anal. Chem.* **2012**, 84, 4334-4339.

Cantrell, S.; Roux, C.; Maynard, P.; Robertson, J. A Textile Fibre Survey as an Aid to the Interpretation of Fibre Evidence in the Sydney Region, *For. Sci. Int.* **2001**, 123, 48-53.

Cassidy, B.; Lu, Z.; Fuenffinger, N.; Skelton, S.; Bringley, E.; Nguyen, L.; Myrick, M.; Breitung, E.; Morgan, S. Rapid and Minimally Invasive Identification of Degraded Polyester-Urethane Magnetic Tape Using Attenuated Total Reflection Fourier Transform Infrared (ATR FT-IR) Spectroscopy and Multivariate Statistics, *Anal. Chem.* **2015**, 87, 9265-9272.

Chang, C.; Lin, J. LIBSVM: a Library for Support Vector Machines, *ACM TIST*, **2011**, 2, 1-27.

Charve, J.; Chen, C.; Hegeman, A.; Reineccius, G. Evaluation of Instrumental Methods for the Untargeted Analysis of Chemical Stimuli of Orange Juice Flavor, *Flavour Fragr. J.* **2011**, 26, 429-440.

Christie, R. *Colour Chemistry*; Royal Society of Chemistry: Cambridge, U.K., 2001.
Cortes, C.; Vapnik, V. Support-Vector Network, *Mach. Learn.* **1995**, 20, 273-297.
Coyle, T.; Larkin, A.; Smith, K.; Mayo, S.; Chan, A.; Hunt, N. Fibre Mapping – a Case Study, *Sci. Justice* **2004**, 44, 179-186.

Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*; Cambridge University Press: Cambridge, U.K., 2000.

Cuddihy, E. Aging of Magnetic Recording Tape, *IEEE Trans. Magn.* **1980**, 16, 558-568.

De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. The Mahalanobis Distance, *Chemom. Intell. Lab. Syst.* **2000**, 50, 1-18.

Di Natale, Paolesse, R.; Macagnano, A.; Mantini, A.; D'Amico, A.; Legin, A.; Lvova, L.; Rudnitskay, A.; Vlasov, Y. Electronic Nose and Electronic Tongue Integration for Improved Classification of Clinical and Food Samples, *Sens. Actuators B* **2000**, 64, 15-21.

Dingari, N.; Barman, I.; Myakalwar, A.; Tewari, S.; Gundawar, M. Incorporation of Support Vector Machines in the LIBS Toolbox for Sensitive and Robust Classification Amidst Unexpected Sample and System Variability. *Anal. Chem.* **2012**, 84, 2686-2694.

Doeswijk, T.; Smilde, A.; Hageman, J.; Westerhuis, J.; van Eeuwijk, F. On the Increase of Predictive Performance with High-Level Data Fusion, *Anal. Chim. Acta* **2011**, 705, 41-47.

Domingos, P.; Pazzani, M. Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier, *Mach. Learn.* **1997**, 29, 103-130.

Dreassi, E.; Caramelli, G.; Perruccio, P.; Corti, P. Transfer of Calibration in Near-Infrared Reflectance Spectrometry. *Analyst.* **1998**, 123, 1259-1264.

Edge, M.; Allen, S.; Hayes, M.; Jewitt, T.; Horie, C.; and Brems, K. Degradation of Magnetic Tape: Support and Binder Stability, *Polym. Degrad. Stabil.* **1993**, 39, 207-214.

Eilers, D. Polyester and Acetate as Magnetic Tape Backings, *J. Audio Engng. Soc*. **1969**, 17, 303-308.

Fearn, T.; Standardisation and Calibration Transfer for Near-infrared Instruments: a Review. *J Near Infrared Spectrosc.* **2001**, 9, 229-244.

Feudale, R.; Woody, N.; Tan, H.; Myles, A.; Brown, S.; Ferre, J. Transfer of Multivariate Calibration Models: a Review. *Chemom. Intel. Lab. Syst.* **2002**, 64, 181-192.

Fisher, R. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugenics.* **1936**, 179-188.

Gaudette, B. The Forensic Aspects of Textile Fiber Examination, In *Forensic Science Handbook*, Vol. II.; Saferstein, R., Ed.; Prentice Hall: New Jersey, 1988.

Gemperline, P. Principal Component Analysis, In *Practical Guide to Chemometrics*, 2nd ed., Gemperline, P., Ed.; Taylor & Francis: Florida, 2006.

Goodpaster, J.; Liszewski, E. Forensic Analysis of Dyed Textile Fibers, *Anal. Bioanal. Chem.* **2009**, 394, 2009-2018.

Grieve, M.; Biermann, T. The Population of Coloured Textile Fibres on Outdoor Surfaces, *Sci. Justice* **1997**, 37, 231-239.

Grieve, M.; Biermann, T.; Schaub, K. The Individuality of Fibres Used to Provide Forensic Evidence – Not all Blue Polyesters are the Same, *Sci. Justice* **2005**, 45, 13-28.

Grieve, M.; Biermann, T.; Schaub, K. The Use of Indigo Derivatives to Dye Denim Material, *Sci. Justice* **2006**, 46, 15-24.

Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Series in Statistics: Springer: New York, 2001.

Hess R. Tape Degradation Factors and Challenges Predicting Tape Life, *J. Assoc. Rec. Sound Collect.* **2008**, 39, 240-274.

Hobaica, S. Analysis of Audio Magnetic Tapes with Sticky Shed Syndrome by ATR FT-IR, *J. Appl. Polym. Sci.* **2013**, *128*, 1962-1973.

Hoehse, M.; Paul, A.; Gornushkin, I.; Panne, U. Multivariate Classification of Pigments and Inks Using Combined Raman Spectroscopy and LIBS. *Anal. Bioanal. Chem.* **2012**, 402, 1443-1450.

Houck, M. Inter-Comparison of Unrelated Fiber Evidence, *For. Sci. Int.* **2003**, 135, 146-149.

Jain, A. K.; Dubes, R. C. *Algorithms for Clustering Data*; Prentice Hall Advanced Reference Series; Prentice Hall: New Jersey, 1988, pp 58-89.

Joliffe, I. *Principal Component Analysis*, 2nd ed.; Springer Series in Statistics; Springer: New York, 2002.

Kalivas, J.; Gemperline, J. Calibration. In *Practical Guide to Chemometrics*, 2nd ed.; Gemperline, P., Ed.; Taylor & Francis: Florida, 2006, pp 105-162.

Karpinska, J. Derivative Spectrophotometry – Recent Applications and Directions of Developments, *Talanta* **2004**, 64, 801-822.

Kaufman, L.; Rousseeuw, P. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley Series in Probability and Mathematical Statistics; Wiley: New York, 1990, pp 199-252.

Kennard, R.; Stone, L. Computer Aided Design of Experiments, *Technometrics* **1969**, 11, 137-148.

Kher, A.; Mulholland, M.; Green, E.; Reedy, B. Forensic Classification of Ballpoint Pen Inks Using High Performance Liquid Chromatography and Infrared Spectroscopy with Principal Components Analysis and Linear Discriminant Analysis. *Vib. Spectrosc.* **2006**, 270-277.

Kononenko, I. Inductive and Bayesian learning in Medical Diagnosis, *Appl. Artif. Intell.* **1993**, 7, 317-337.

Kotsiantis, S. Supervised Machine Learning: a Review of Classification Techniques, *Informatica* **2007**, 31, 249-268.

Krzanowski, W. *Principles of Multivariate Analysis*; Oxford Statistical Science Series 3; Clarendon Press: Oxford, U.K., 1988.

Kuncheva, L. *Combining Pattern Classifiers*; Wiley: New Jersey, 2004.

Li, B.; Beveridge, P.; O'Hare, W.; Islam, M. The Estimation of the Age of a Blood Stain Using Reflectance Spectroscopy with a Microspectrophotometer, Spectral Pre-processing and Linear Discriminant Analysis. *Forensic Sci. Int.* **2011**, 212, 198-204.

Liu, Y.; Brown, S. Wavelet Multiscale Regression from the Perspective of Data Fusion: New Conceptual Approaches, *Anal. Bioanal. Chem.* 2004, 380, 445-452.

Mahalanobis, P. On the Generalized Distance in Statistics, *Proc. Nat. Instit. Sci. India* **1936**, 2, 49-55.

Massonnet, G.; Buzzini, P.; Monard, F.; Jochem, G.; Fido, L.; Bell, S.; Stauber, M.; Coyle, T.; Roux, C.; Hemmings, J.; Leijenhorst, H.; Van Zanten, Z.; Wiggins, K.; Smith, C.; Chabli, S.; Sauneuf, T.; Rosengarten, A.; Meile, C.; Ketterer, S.; Blumer, A. Raman Spectroscopy and Microspectrophotometry of Reactive Dyes on Cotton Fibres: Analysis and Detection Limits, *For. Sci. Int.* **2012**, 222, 200-207.

Matthews, B. Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme, *Biochim. Biophys. Acta-Protein Struct. M.* **1975**, 405, 442-451.

Morgan, S.; Bartick, E. Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In *Forensic Analysis on the Cutting Edge*; Blackledge, R., Ed.; John Wiley & Sons: New Jersey, 2007, pp. 333-374.

Morgan, S.; Hall, S.; Hendrix, J.; Bartick, E. In *Pattern Recognition Methods for Classification of Trace Evidence Textile Fibers from UV/Visible and Fluorescence Spectra*, Proceedings of the National Institute of Justice Trace Evidence Symposium, Kansas City, Missouri, 2011.

Morgan, S.; Niewland, A.; Mubarak, C.; Hendrix, J.; Enlow, E.; Vasser, B. In *Forensic Discrimination of Dyed Textile Fibres Using UV-VIS and Fluorescence Microspectrophotometry*, Proceedings of the 12th Meeting of the European Fibres Group, Prague, Czech Republic, 2004.

Myles, A.; Zimmerman, T.; Brown, S. Transfer of Multivariate Classification Models between Laboratory and Process Near-infrared Spectrometers for the Discrimination of Green Arabica and Robusta Coffee Beans. *Appl. Spectrosc.* **2006**, 60, 1198-1203.

Neural Network ToolboxTM User's Guide. http://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf (accessed Jan. 16, 2015).

Palenik, S. Microscopical Examinations of Fibers, In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M.; Taylor and Francis: London, U.K. 1999, pp. 153-177.

Palmer, R.; Hutchinson, W.; Fryer, V. The Discrimination of (Non-denim) Blue Cotton, *Sci. Justice* **2009**, 49, 12-18.

Palmer, R.; Oliver, S. The Population of Coloured Fibres in Human Head Hair, *Sci. Justice* **2004**, 44, 83-88.

Pizzaro, C.; Rodriguez-Tecedor, S.; Perez-del-Notario, N.; Esteban-Diez, Gonzalez-Saiz, J. Classification of Spanish Extra Virgin Olive Oils by Data Fusion of Visible Spectroscopic Fingerprints and Chemical Descriptors, *Food Chem.* **2013**, 138, 915-922.

Rao, R. The Utilization of Multiple Measurements in Problems of Biological Classification. *J. R. Stat. Soc. Series B.* **1948**, 10, 159-203.

Rencher, A. *Methods of Multivariate Analysis*, 2nd ed.; Wiley: New York, 2002.

Rendle, D.; Wiggins, K. Forensic Analysis of Textile Fibre Dyes, *Rev. Progress. Coloration* **1995**, 25, 29-34.

Reubinstein, R.; Kroese, D. *The Cross-Entropy Method: A Unified Apporach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*; Springer-Verlag: New York, 2001.

Richardson C. Process for Restoring Magnetic Recording Tape Damaged by "Sticky Shed" Syndrome. United States Patent No. 6797072 B1 (28 Sept. 2004).

Roussel, S.; Bellon-Maurel, V.; Roger, J.; Grenier, P. Fusion of Aroma, FT-IR and UV Sensor Data Based on the Bayesian Inference. Application to the Discrimination of White Grape Varieties." *Chemom. Intell. Lab. Syst.* **2003**, 65, 209-219.

Rudnitskaya, A.; Kirsanov, D.; Legin, A.; Beullens, K.; Lammertyn, J.; Nicolai, B.; Irudayaraj, J. Analysis of Apples Varieties – Comparison of Electronic Tongue with Different Analytical Techniques, *Sens. Actuators B* **2006**, 116, 23-28.

Sahni, N.; Isaksson, T.; Naes, T. Comparison of Methods for Transfer of Calibration Models in Near-infrared Spectroscopy: a Case Study Based on Correcting Path Length Differences Using Fiber-Optic Transmittance Probes in In-line Near-infrared Spectroscopy. *Appl. Spectrosc.* **2005**, 59, 487-495.

Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Square Procedures, *Anal. Chem.* **1964**, 36, 1627-1639.

Schoonover, J.; Dattelbaum, D.; Osborn, J.; Bridgewater, J.; Kenney, J. Pressure-Dependent Fourier Transform Infrared Spectroscopy of a Polyester-Urethane, *Spectrochim. Acta A* **2003**, 59, 309-319.

Sikirzhytski, V.; Sikirzhytski, A.; Lednev, I. Advanced Statistical Analysis of Raman Spectroscopic Data for the Identification of Body Fluid Traces: Semen and Blood Mixtures. *Forensic Sci. Int.* **2012**, 222, 259-265.

Silva, C.; Borba, F.; Pimentel, M.; Pontes, M.; Honorato, R.; Pasquini, C. Classification of Blue Pen Ink Using Infrared Spectroscopy and Linear Discriminant Analysis. *Microchem. J.* **2013**, 122-127.

Smith, L. Factors Governing the Long-Term Stability of Polyester-Based Recording Media, *Restaurator* **1991**, 12, 201-218.

Stefan, A.; Dockery, C.; Nieuwland, A.; Roberson, S.; Baguley, B.; Hendrix, J.; Morgan, S. Forensic Analysis of Anthraquinone, Azo, and Metal Complex Acid Dyes from Nylon Fibers by Micro-extraction and Capillary Electrophoresis, *Anal. Bioanal. Chem.* **2009**, 394, 2077-2085.

Steiner, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Squares Procedure. *Anal. Chem.* **1972**, 44, 1906-1909.

Stoeffler, S. A Flowchart System for the Identification of Common Synthetic Fibers by Polarized Light Microscopy, *J. Forensic Sci.* **1996**, 41, 297-299.

Subari, N.; Saleh, J.; Shakaff, A.; Zakaria, A. A Hybrid Sensing Approach for Pure and Adulterated Honey Classification. *Sensors* **2012**, 12, 14022-14040.

SWGMAT Forensic Fiber Examination Guidelines. http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf (accessed May 5, 2014).

Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 3rd ed.; Academic Press: California, 2006.

Thiébaut, B.; Vilmont, L-B.; Lavédrine, B. Characterization of U-matic Videotape Deterioration by Size Exclusion Chromatography and Pyrolysis Gas Chromatography/Mass Spectrometry and the Role of Adipic Acid, *J. Cultural Heritage* **2009**, 10, 183-197.

Thompson, D.; Osborn, J.; Kober, E.; Schoonover, J. Effects of Hydrolysis-Induced Molecular Weight Changes on the Phase Separation of a Polyester Urethane, *Polym. Degrad. Stab.* **2006**, 91, 3360-3370.

Tungol, M.; Bartick, E.; Montaser, A. Analysis of Single Polymer Fibers by Fourier Transform Infrared Microscopy: The Results of Case Studies," *J. Forensic Sci.* **1991**, 36, 1027-1043.

van Dam, J. In *Environmental benefits of natural fibre production and use,* Proceedings of the Symposium on Natural Fibres, Rome, Italy, 2008; pp 3-17.

Vera, L.; Acena, L.; Guasch, J.; Boque, R.; Mestres, M.; Busto, O. Discrimination and Sensory Description of Beers through Data Fusion, *Talanta* **2011**, 87, 136-142.

Volkmann, H. Film Preservation: a Report of the Preservation Committee of the International Federation of Film Archives, *FIAF* **1965**, 7.

Ward, J. Hierarchical Grouping to Optimize an Objective Function, *J. Amer. Statist. Assoc.* **1963**, 58, 236-244.

Watt, R.; Roux, C.; Robertson, J. The Population of Coloured Textile Fibres in Domestic Washing Machines, *Sci. Justice* **2005**, 45, 75-83.

Wiggins, K. Thin Layer Chromatographic Analysis for Fibre Dyes. In *Forensic Examination of Fibres*, 2nd ed.; Robertson, J.; Grieve, M., Eds.; Taylor and Francis: London, U.K., 1999, pp 291-310.

Wiggins, K.; Palmer, R.; Hutchinson, W.; Drummond, P. An Investigation into the Use of Calculating the First Derivative of Absorbance Spectra as a Tool for Forensic Fibre Analysis, *Sci. Justice* **2007**, 47, 9-18.

Wilhelm, C.; Gardette, J. Infrared Analysis of the Photochemical Behavior of Segmented Polyurethanes: 1. Aliphatic Polyester-Urethane, *Polymer* **1997**, 38, 4019-4031.
Zelst, L. *Sticky Shed Sydrome: Characterization, Diagnosis, and Treatment,* Internal Library of Congress Report, March, 2008.