

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: Evaluation of Statistical Measures for Fiber Comparisons: Interlaboratory Studies and Forensic Databases

Author(s): Stephen L. Morgan, Ph.D.

Document No.: 248386

Date Received: October 2014

Award Number: 2010-DN-BX-K220

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

FINAL REPORT

Federal Agency and Organization Element to Which Report is Submitted:

U. S. Department of Justice, National Institute of Justice

Federal Grant or Other Identifying Number Assigned by Agency:

NIJ award number 2010-DN-BX-K220

Project Title:

Evaluation of Statistical Measures for Fiber Comparisons: Interlaboratory Studies and Forensic Databases

PI Name, Title and Contact Information (e-mail address and phone number):

Dr. Stephen L. Morgan, Professor
Department of Chemistry and Biochemistry, 631 Sumter Street,
University of South Carolina, Columbia, SC 29208
Voice: 803.777.2461, FAX: 803.777.9521
Email: morgansl@mailbox.sc.edu

Administrative Point Of Contact:

Lumi Bakos, Administrator
Sponsored Awards Management
901 Sumter Street-5th Floor, Byrnes Building
Columbia, SC 29208 University of South Carolina, Columbia, SC 29208
Voice: 803.777.2274, FAX: 803.777.4136
Email: bakos@mailbox.sc.edu

Submission Date:

21 May 2014

DUNS and EIN Numbers:

DUNS number 11-131-0249
EIN number 57-0967350

Recipient Organization (Name and Address):

South Carolina Research Foundation, 5th Floor Byrnes Building,
901 Sumter Street, Columbia, SC 29208

Project/Grant Period (Start Date, End Date):

1 October 2010-30 April 2014

Abstract

This project is a collaborative effort between three groups of forensic researchers in academic laboratories (Stephen Morgan at the University of South Carolina (USC), Columbia, SC; Edward Bartick at Suffolk University in Boston, MA; John Goodpaster at Indiana University-Purdue University Indianapolis (IUPUI), IN). The major objectives of this project : (a) conduct interlaboratory experiments to evaluate decision making in forensic fiber examinations by polarized light microscopy measurements, UV/visible microspectrophotometry, and IR spectroscopy; (b) investigate the application of multivariate statistical measures for evaluation of comparisons of questioned (Q) vs. known (K) fibers; (c) to evaluate intra-laboratory variability, inter-laboratory agreement, and error rate performance in designed experiments; (d) document good laboratory practices relevant to achieving acceptable levels of intra- and inter-laboratory consistency in fiber data; (e) development and use of a prototype forensic data management system for fiber examinations that will integrate electronic signatures for documentation on data stored, data validity checking, and relational database searching.

The general focus of the project has been to collect microspectrophotometry (MSP) and supporting infrared microspectrophotometry (IR) data (for polymer identification) from a collection of well characterized textile fibers. Visible MSP analysis has been used to determine how well we can discriminate fibers based on multivariate statistical methods applied to the spectral characteristics that result from dye components present on the fibers. Fibers have been identified as to their generic and sub-generic classes and microscopy examinations have been conducted to determine fiber diameters and the cross-sectional shape. Statistical graphics and, multivariate analysis using principal component analysis (PCA) and linear discriminant analysis (LDA) have been used to determine the discriminating ability between fibers of different colors and chemical compositions using the analysis methods of MSP and IR.

Table of contents

Title page, technical contact and administrative contact information	page 1
Abstract	2
Table of contents	3
Executive summary	4
Technical report	8
A. Microspectrophotometric analysis of yellow polyester fiber dye loadings with utilization of chemometric techniques	8
B. A statistical basis for significance evaluation of fibers as class evidence	22
C. Comparison of multivariate preprocessing techniques for the forensic discrimination of cotton fibers by UV/visible microspectrophotometry	36
D. Multivariate discrimination of dyed textile fibers from UV/visible and fluorescence spectra	54
E. Model transfer for multivariate discrimination of textile fibers by UV/visible microspectrophotometry	65
F. Design of an extensible forensic database for textile fibers	80
Overall Results	92
Conclusions	92
1. Discussion of findings	92
2. Implications for policy and practice	97
3. Dissemination of research findings	102

EXECUTIVE SUMMARY

This project started with reorganization of the USC fiber collection. We started this project with a collection of over 600 fibers resulting from previous research funded by the FBI Laboratory. During the second six-month period (1 January 2011-30 June 2011), we expanded our fiber collection to more than 1,400 fiber samples in two steps. The collection was initially expanded to 914 samples with the addition of 300 fiber samples from textile companies. This collection consists mostly of the four most common fibers encountered in forensic trace evidence: acrylic, cotton, nylon, and polyester. Fairly large quantities of these stock samples (in some cases, large swatches of fabric) are stored in filing cabinet folder in a darkroom, along with chemical samples of most of the dyes used. Secondly, The Trace Evidence Section of the State Law Enforcement Division Forensic Services Laboratory in Columbia, SC, contributed about 500 fibers samples from their copy of the Consolidated Testing Services collection that was distributed by the National Bureau of Standards (which is no longer available). By August 2011, this collection increased with additional donations from SLED to almost 3,000 fibers. The samples include polymer staple materials and undyed and dyed fibers, but dyes samples are not available for all fabrics. A 'working collection' of about 1,300 fibers containing smaller representative samples of the combined fiber collection are stored in acid-free protective pockets in three-ring binders, with relevant information on polymer and dye composition, referenced by assigned fiber identification and dye identification numbers, and organized in a Microsoft Access database. Our laboratory kept one of these working collections, and a second copy was given to SLED for their use as casework comparison samples. Copies of these physical samples were also sent to Drs. Bartick and Goodpaster for their use during the project and afterwards.

The premise of our fiber collection and data base efforts is that if enough is known about the distribution of a population from which questioned fibers from a suspect and known fibers from a crime scene are class members, multiple associated characteristics (physical, optical, or spectroscopic) decrease the random probability of a match occurring solely by chance. To consider a database usable and realistic when dealing with fibers as evidence, it is necessary to have a large number of representative samples that are typical within the geographic region where the crime occurred. With a good database, the use of instrumental measurements of multiple characteristics can be applied to a range of class materials to achieve some measures of statistical significance for trace evidence matches.

Our research accomplishments are documented in the five draft papers in the following Technical Report section.

The first study involves the analysis of lightly dyed fibers and the discrimination of fibers with varying dye loadings; discrimination in this situation is a challenging task for the fiber examiner. Microspectrophotometry is a quick, accurate, and reproducible method to compare colored fibers for forensic purposes. The use of chemometric techniques applied to spectroscopic data can provide valuable information, especially when looking at a complex dataset. In our first study, background subtracted and normalized visible spectra from ten yellow polyester exemplars dyed with different concentrations of the same dye ranging from 0.1-3.5% (w/w), and analyzed by agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). Three classes of fibers with a classification accuracy of approximately 96% were found, representing low, medium, and high dye loadings. Exemplars with similar dye loadings showed discrimination with classification accuracy of 90% or higher

and an area under the receiver operating characteristic curve of 0.9 or greater. Calibration curves based upon a proximity matrix of dye loadings between 0.1-0.75% (w/w) were developed that challenged the accuracy and precision of a traditional approach. Multivariate statistical analysis of visible spectra was found to provide an objective means of discriminating similar fibers with different dye loadings. Chemometric treatment of visible spectra from fibers with different dye loadings has shown to be a reliable and effective way of discriminating between yellow dye loadings when fibers were placed into low, medium, and high classes. Calibration curves, based on proximity matrices, can be produced and accurately predict the dye loading of an unknown fiber. Comparisons of two groups of fibers can provide discriminating information. Overall, forensic fiber examiners can have a more objective way of comparing a known and questioned fiber using chemometric techniques on visible spectra.

The objective of the second study reported here was to demonstrate the statistical significance of measurement variance of like and unlike fibers and to demonstrate how the estimations of match significance can be made from fiber evidence. In addition, the process can be applied to class evidence in general. The work was conducted through the use of multivariate statistical analysis and the product rule of independent variables. We have shown how to increase the ability to classify and discriminate synthetic and cotton fibers. A library of 923 well characterized fibers was developed with known dye components. With the use of visual light microscopy, UV/visible microspectrophotometry (MSP), and Fourier transform infrared (FTIR) microspectroscopy, data was collected on fibers. Studies were done on twenty-one red cotton and twenty-one red acrylic fibers using multivariate statistical analysis. The absorption spectra of fibers from 10 replicate UV/visible microspectrophotometry scans on each fiber were compared by using principal component (PCA) and linear discriminant analysis (LDA). With the aid of multivariate statistics, fibers that are difficult to distinguish by visual comparison were distinguished. In addition, for the acrylic fibers, FTIR spectra were used to identify generic and sub-generic class. The characteristics of cross sectional shape and diameter were determined. Finally, using the product rule of probability, knowing the number of fibers in the database with specific color, diameter, cross-sectional shape, and chemical composition, the percentage occurrence of each fiber was determined. The product of the percentages was then calculated to determine the probability of two fibers matching randomly with those characteristics. Probabilities on the order of 1 in 0.5 million are obtainable with such comparisons between fibers, provided a sufficiently large and representative database of fiber characteristics is accessible. The improved understanding of sources of variability and decision-making processes gained from this research will serve to advance the forensic significance of involving fiber and other class evidence material examinations.

The third study originated with a 2007 paper by Ken Wiggins in *Science & Justice* in which taking the first derivative of spectra was recommended to enhance discrimination. However, quantitative data to back up this claim was not provided. This work is important because fiber examiners do not routinely use derivatives for visual examination of differences in fiber spectra. Color plays a critical role when analyzing natural fibers such as cotton in forensic investigations. Ultraviolet-visible (UV/vis) 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. We have found that derivative preprocessing can enhance structure in spectra. Samples of reactive, direct, and vat dyed cotton fibers were analyzed and spectra were preprocessed using multiple methods including baseline correction, normalization, and first and

second derivatives. Principal component analysis followed by linear discriminant analysis was employed to discriminate 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 classification accuracy when analyzing spectra with only subtle features such as those seen with 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.

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 λ values. It should be noted that the effect of smoothing the spectra using a Savitzky-Golay polynomial [27] prior to calculating the first derivative 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 first derivative spectra misclassifying matching fibers with large variations in absorbance. This resulted in classification accuracies of first derivative 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 percent) 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, the analyst must show caution when selecting cases in which derivatives should be used.

The fourth research paper reported here presents a comprehensive comparison of discrimination of 482 fibers by UV/visible and fluorescence microspectrophotometry. Determining which analytical method will have the highest discrimination power for trace evidence examinations is significant to forensic laboratories to save time and resources. This study compares the discrimination ability of ultraviolet-visible (UV-VIS) microspectrophotometry (MSP) and fluorescence MSP, two common techniques used by forensic analysts to study fibers and fiber dyes. Dyed textile samples of cotton, acrylic, nylon 6,6, and polyester were analyzed using UV-VIS MSP and fluorescence MSP at four wavelengths (365, 405, 436, and 546 nm). All spectra were preprocessed and classified using principal component analysis followed by linear discriminant analysis. Leave-one-out cross validation was used to test the ability of principal component-linear discriminant analysis (PC-LDA) to discriminate fibers in each color and polymer based group. The highest discrimination power was obtained by UV-VIS MSP. PC-LDA correctly classified 89.50% of the UV-VIS MSP spectra examined.

Our fifth research paper discusses the ability to transfer multivariate classification models between laboratories. Such efforts might save time and resources in forensic analyses. 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, ultraviolet (UV)-visible absorbance spectra of 12 blue acrylic fibers were examined. The agreement of results among three separate laboratories was evaluated by testing the

transferability of multivariate classification models was based on principal component analysis-linear discriminant analysis (PCA-LDA) and partial least squares-discriminant analysis (PLS-DA). An average classification accuracy of 88.06% was found after training the PLS-DA models using data collected at two laboratories and using the information collected at the third laboratory as an external test set. For comparison, intra-laboratory studies carried out using PCA-LDA produced an average classification accuracy of 98.33%. These results suggest some limitations to the transferability of classification models between laboratories. Researchers are advised to use caution and to follow the traditional adage of "Trust, but verify."

The last section of the Technical Report discusses the design and implementation of web-based forensic fiber database that facilitates archiving fiber data such as polarized microscopy measurements (birefringence, sign of elongation), physical characteristics (diameter, shape), and spectral data. One immediate advantage is the ability to store data in a documented manner and access this information on demand from a relational database. Aspects of the implementation are provided to document our efforts to create a product to demonstrate the potential of such a system. Although there is little likelihood of establishing a truly comprehensive fiber database because of fast moving trends in manufacturing and globalization of production, a combined data archiving and statistical graphics and analysis system offers both data management and decision-making support to the forensic fiber examiner.

TECHNICAL REPORT

A. Microspectrophotometric Analysis of Yellow Polyester Fiber Dye Loadings with Utilization of Chemometric Techniques.

Eric J. Reichard¹, John V. Goodpaster^{1*}, Edward G. Bartick², Stephen L. Morgan³

¹Indiana University Purdue University Indianapolis (IUPUI), Department of Chemistry and Chemical Biology, Forensic & Investigative Sciences Program, 402 N. Blackford Street, LD 326, Indianapolis, IN 46202, ereichar@iupui.edu, jvgoodpa@iupui.edu

²Suffolk University, Department of Chemistry and Biochemistry, Boston, MA, ebartick@suffolk.edu

³University of South Carolina, Department of Chemistry and Biochemistry, Columbia, SC 29208, morgansl@mailbox.sc.edu

*author to whom correspondence should be addressed.

Abstract. Microspectrophotometry is a quick, accurate, and reproducible method to compare colored fibers for forensic purposes. The use of chemometric techniques applied to spectroscopic data can provide valuable information, especially when looking at a complex dataset. In this study, background subtracted and normalized visible spectra from ten yellow polyester exemplars dyed with different concentrations of the same dye ranging from 0.1-3.5% (w/w), were analyzed by agglomerative hierarchical clustering (AHC), principal component analysis (PCA), and discriminant analysis (DA). Three classes of fibers with a classification accuracy of approximately 96% were found, representing low, medium, and high dye loadings. Exemplars with similar dye loadings showed discrimination with classification accuracy of 90% or higher and an area under the receiver operating characteristic curve of 0.9 or greater. Calibration curves based upon a proximity matrix of dye loadings between 0.1-0.75% (w/w) were developed that challenged the accuracy and precision to that of a traditional approach. Multivariate statistical analysis of visible spectra provides an objective means of discriminating similar fibers with different dye loadings.

Keywords: multivariate statistics; dye loading; polyester; fibers; microspectrophotometry; forensic science.

Introduction. The Locard Exchange Principle states that when two objects come into contact, there is always a transfer of material¹. This principle is especially relevant to trace evidence involving textile fibers. Textile fibers can be identified and compared based on their macroscopic and microscopic characteristics, optical characteristics, chemical composition, and color^{1,2}. In the comparison of transferred fibers during violent crimes, the fiber color is a very important point of comparison.

Color comparisons by eye are subjective and not always discriminating for closely dyed fiber colors. Spectral comparisons by microspectrophotometry (MSP) have become important in transferred fiber cases because the technique provides quick, non-destructive, and objective color comparisons for dyed fibers. Microspectrophotometry can discriminate between two colored fibers that are visually similar based upon the different chromophores in the dye's molecular

structure. Research in color comparisons with MSP have been conducted and show the viability of this technique^{1,3-6}. While the comparison of spectra has been very helpful, even this can become difficult in closely dyed fibers that have the same dye composition, but different concentrations of individual dyes. Fiber comparisons of this nature could be of importance in cases where a residence had carpets with the same colors, but shaded differently in different rooms. UV/visible spectra of textile fibers are dominated by the dyes present and tend to have broad peaks with few distinctive features to aid visual comparisons. Some authors have suggested calculating first derivatives of spectra to enhance the ability to pick out discriminating features⁷. First derivatives, however, tend to magnify noise in spectra, which could impede interpretation. Other common analytical techniques employed for dye analysis include thin-layer chromatography (TLC)^{9,10}, liquid chromatography-mass spectrometry (LC-MS)^{11,12}, and capillary electrophoresis (CE)¹³. All of these techniques are destructive to the sample, however. A detailed discussion of forensic textile fiber examination techniques can be found in Houck² and Robertson and Grieve¹⁴.

Multivariate statistics utilizes multiple variables to describe complex datasets. Chemometrics identifies patterns and groupings from large complex datasets more accurately than visual examination alone. Three techniques can be used for the analysis of colored fibers: Agglomerative Hierarchical Clustering (AHC), Principal Component Analysis (PCA), and Discriminant Analysis (DA). AHC clusters samples based upon their relative similarity/dissimilarity, which is often expressed as a multi-variate Euclidean distance¹⁵. An initial understanding of the classes that may be present in a dataset can be obtained from AHC. PCA reduces the dimensionality of the dataset by concentrating the total amount of variance into a smaller number of latent variables, which are linear combinations of the original variables^{15,16}. These new variables, which represent the directions of maximum variation, are called principal components, or PCs. If most of the variation can be explained in the first few PCs, a plot of the data projected on the first several PC's may reveal clustering behavior of the spectral data from several fibers and allow visualization of the grouping relationships. DA, just like PCA, produces linear combinations of variables called canonical variates (CVs). However, CVs are designed to maximize between-group variability and minimize within-group variability¹⁶. DA generates a model that best classifies the original data and which can predict the classification of new samples into the established model. Multivariate statistics have been applied to a number of relevant evidence types, including dyes¹⁷, inks¹⁸, automotive paint¹⁹, electrical tape²⁰, and fire debris²¹. Overall, chemometrics has established itself as a viable method of analyzing complex chemical data.

The purpose of this study was to discriminate between different yellow polyester fiber visible spectra based solely upon their dye loadings using chemometric analysis. A dye loading is the concentration of a dye, usually in weight percent, applied to a fiber. Research has shown that visually similar yellow polyester fibers with different dye compositions can be discriminated based on their UV-visible spectra⁸. However, no one has determined if fibers dyed with the same dye, but different dye loadings, can be reliably discriminated by their visible spectra alone using a chemometric approach. Being able to discriminate between dye loadings would provide a higher level of discriminating power to forensic fiber examiners in their everyday casework. This study will address issues brought about by the National Academy of Sciences (NAS) report for strengthening forensic science in the United States²². For example, the use of chemometric analysis would address issues with the accuracy and reliability of observer interpretations of spectroscopic data.

Materials and Methods. Ten yellow polyester exemplars dyed with various amounts of Dianix Yellow 5-6G (Disperse Yellow 114) dye, were supplied from a collection of textile fibers housed at the University of South Carolina (Columbia, SC). Preliminary measurements using a Leica DM EP PLM (Leica Microsystems, Buffalo Grove, IL) and a Perkin Elmer Spectrum One FT-IR spectrometer with a universal ATR sampling accessory (Perkin Elmer, Waltham, MA) were performed to confirm that the set of polyester fibers of round cross-sections and diameters varying within a range of 20-27.5 μm . Cargille oils (R.P. Cargille Laboratories, Cedar Grove, NJ) were used to determine refractive indices by PLM; Michel-Levy charts were used to determine birefringence values. FT-IR spectra were acquired, based on averaging sixteen scans at a resolution of 4.00 cm^{-1} , and recorded in percent transmittance over the wavenumber range of 4000-650 cm^{-1} . Table I shows the amount of dye applied, in weight percent, to each exemplar and the naming system employed for the study. All fibers in the study were of the same fiber type, cross-sectional shape, and dye.

Procedures from the Scientific Working Group on Materials Analysis: Fiber Subgroup (SWGMAF) were followed²³. A number of fibers and locations along those fibers must be analyzed due to assess real and apparent variations in dyeing depth at different locations along the fiber. Ten fibers from each exemplar were removed and mounted on glass microscope slides using Permunt (Fischer Scientific, Fairlawn, NJ) mounting media. A CRAIC QDI 2000 microspectrophotometer (Craic Technologies, San Dimas, CA) was used in transmitted light mode at a total magnification of 150 \times . 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. Fifty scans were taken at a resolution factor of five for each sample spectrum. Five spectra were taken at different locations along each fiber to account for intra-fiber variation. A total of 50 spectra were collected for each exemplar to provide sufficient information on dye loading variations within each exemplar. Visible MSP data was collected over the wavelength range of 350-800 nm.

Preprocessing techniques were employed before subjecting the data to statistical treatment. Background subtraction was performed on each spectrum by subtracting the minimum absorbance value for each sample from all absorbance values to eliminate the effects of scattered light and bring the baseline for each spectrum down to zero. Next, each background-subtracted spectrum was normalized to unit vector length by dividing each absorbance value by the square root of the sum of squares of all absorbance values. This normalization step accounts for differences in path lengths due to varying fiber diameters. In theory, normalizing would remove any concentration effects, but this may not always be the case as our results show otherwise (see below). Absorbance values at all wavelengths were used for statistical analysis because the normalized spectra exhibited noticeable differences between fibers at wavelengths other than the maximum wavelength. All subsequent data analyses were performed on the background subtracted and normalized data utilizing all wavelengths.

Calibration plots based on the absorbance at the maximum wavelength (424.98 nm) and the Euclidean distance between an exemplar and a blank were generated on the fibers with dye loading percentages from 0.1-0.75%. Three calibrators (0.2%, 0.4%, and 0.5%) were left out of the curve one at a time and considered an unknown to determine the efficiency of the calibration curve.

All chemometric techniques were performed using XLSTAT Pro (AddinSoft, New York, NY), an add-in software for Microsoft Excel (Redmond, WA). For Agglomerative Hierarchical Clustering analysis, the five spectra for each fiber were averaged. This was done to produce a readable dendrogram. The proximity between two samples was measured by Euclidean distance¹⁵ and the aggregation method used for grouping samples was Ward's Method²⁴. The truncation line was set just higher than the most dissimilar exemplar's replicates. Values were determined by locating the node where all the replicates for each exemplar met. From that truncation line, exemplars were placed into classes.

For principal component analysis all scans were utilized instead of the averaged scans. The PCA algorithm used was singular value decomposition (SVD)²⁴, which utilizes a correlation matrix of the original variables to produce principal components. A factor loadings plot and observations plot were generated from the first two principal components. Three PCs were retained for subsequent discriminant analysis based upon a scree plot. A scree plot provides a visual representation of the decreasing variation in each principal component by plotting the eigenvalues against the principal component number. A sudden break in the plot indicates the number of significant PCs to retain. Any PC after that break is considered noise.

For discriminant analysis all scans were utilized instead of the averaged scans. A Box M test was performed to determine if the covariance matrices were unequal. This test was derived from the likelihood-ratio test and uses an F approximation to compute its significance. From that test it was determined if subsequent analysis of data should utilize quadratic discriminant analysis (QDA) or linear discriminant analysis (LDA). The first three PCs were used as variables instead of the original variables because the number of variables must be less than the number of samples when using DA.

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

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

Results and Discussion. Ten yellow polyester fibers with round cross sections and the same dye composition were analyzed using multivariate statistical techniques to evaluate whether dye loadings could be used to discriminate them. Visual examination of the fibers showed a slight difference in saturation between exemplars. Discrimination of similar dye loadings was not apparent, however. A preliminary examination of the exemplars by use of PLM and ATR-FTIR was conducted to confirm the identity of the fibers and to determine if the dye loading had any effect on the analysis. PLM was used to determine diameters, refractive indices, birefringence values, and signs of elongation. All exemplars had similar optical characteristics and dye loadings did not affect the results other than visual differences based on saturation. IR spectra were subjected to chemometric analysis and resulted in no significant differences between

spectra of different fiber dye loadings. Library searches of the spectra were conducted, and they were highly correlated to exemplars of polyester.

Normalization of the spectra was appropriate due to the different diameters of the fibers. Normalizing spectral data can eliminate concentration differences, but in this case it does not affect the results due to the underlying background spectrum. Spectra superimposed on a sloping background contain varying signal to background ratios. Fibers with a strong absorbance have had a larger signal to background ratio than fibers with a weak absorbance. The background spectrum consists of the fiber polymer and possibly the delustering agent and decreases in absorbance from short to long wavelengths. This is due to light scatter at short wavelengths leading to an absorbance artifact at short wavelengths in the background spectrum. Thus, normalizing changes the shape of spectra due to the sloping background (see Figure 1).

Calibration plots. The absorbance at the maximum wavelength of the spectra was plotted against the exemplars' dye loading percentage. Proximity matrices were also created and the generated distances between dyed exemplars and an undyed exemplar were plotted against the exemplars' dye loadings. Both plots exhibited a curve with a linear portion at lower dye loadings that produced a negative deviation at higher dye loadings. This negative deviation can be attributed to several reasons, including the use of polychromatic light from the MSP source. Beer's law is considered invalid if polychromatic light is used as a light source at higher concentrations²⁵. The linear portion of the curve ends at a dye loading of 0.75%. Therefore, calibration plots were created from 0.1% - 0.75% dye loading. This assertion complements previous work that characterized different pigments loaded into polypropylene fibers²⁶. The results from Table II indicate that statistical treatment of the data provides better results than simply using the maximum wavelength. Based on our results, using proximity matrices, which utilize every variable, are more accurate and precise than using one absorbance value for generating a calibration curve. This is understandable because the maximum wavelength shifts to a higher wavelength as dye concentration increases for this dataset, thus underestimating the absorbance at that wavelength.

Training set. The ten exemplars from Table I were subjected to AHC, PCA, and DA after being background subtracted and normalized. Exemplars A and B exhibit a slight difference in their spectral shape because of the low dye loading on those fibers. The remaining exemplars exhibited the same general spectral curve, but at different absorbance values at the maximum wavelength. The maximum wavelength slightly shifted to higher wavelengths as dye loading increased (see Figure 2).

Agglomerative hierarchical clustering was performed in order to detect any outliers and classes of fibers. The intra-fiber spectra were averaged in order to visualize the classes in the dendrogram. This brought the replicates down from fifty to ten. The dendrogram in Figure 3 shows how the data were grouped. The truncation line was set just higher than the highest level of dissimilarity between replicates. The replicates of exemplars C and D were determined to be the most dissimilar by locating the node where all the replicates for each exemplar met. Everything then to the right of the dendrogram was grouped into classes. From the truncation line, three distinct classes of data were produced, representing low, medium, and high dye loadings. Class 1 are low dye loadings (0.1%, 0.2% w/w) and include exemplars A and B. Class 2 are medium dye loadings (0.4%, 0.5%, 0.75% w/w) and includes exemplars C, D, and E. Class 3 are high dye loadings (1.5%, 2.0%, 2.5%, 3.0%, 3.5% w/w) and includes exemplars F, G, H, I, and J. Class 1 exemplars are further separated into their own respective groups because of their

high reproducibility. Class 2 and class 3 exemplars show low reproducibility and are intermixed in their respective classes.

Principal Component Analysis was performed on every spectrum for each exemplar. Observation plots of the individual exemplars and the three classes captured 92.59% of the total variance in two PCs. Figure 4 shows the ten exemplars spread out from low to high dye loadings. PC1 separates out exemplars A-E while PC2 slightly separates exemplars F-J. Figure 5 shows the observation plot of the three classes produced from AHC. PC1 separates out classes 1, 2, and 3, while PC2 separates samples within each class. From this plot and the results from AHC, it is apparent that these three classes produce three distinct levels of dye loading that correspond to low, medium, and high dye concentrations.

A factor loadings plot, seen in Figure 6, was also generated to understand the relationship between the original wavelengths and the new principal components. Factor loadings of 1 and -1 correspond to high positive and negative correlations between the original wavelengths and the PCs. PC1 showed a strong negative correlation around the maximum wavelength (400-475 nm) of the spectra and strong positive correlations between 350-375 nm and 500-700 nm. PC2 showed positive correlations around the leading and trailing edges of the fiber's spectral curve. The importance of these regions explains how the exemplars were separated. The lower dye loadings were separated based on the maximum absorbance, which corresponded to their maximum wavelength. The higher dye loading exemplars were separated based on the sides of their respective spectral curves, not the maximum wavelength as expected because their spectral curves plateaued and became broader. This was especially seen with an absorbance above 1.

Discriminant analysis was performed several times on all the spectra. First, the exemplars were assigned to their own classes. The ten exemplars showed the same general trend from low to high dye loadings, as seen in Figure 7. This observation plot had a total captured variance of 99.82% using two canonical variates. A "leave-one-out" cross validation was performed on the training set and a subsequent confusion matrix was produced. Here, DA is performed on all but one of the samples and a model is created. The left out sample is then added into the model and is classified to a class. This step is repeated until every sample in the dataset is classified. Cross validating the training set gave a total classification accuracy of 50.80%, which can be seen in Table III. The classification accuracy generally decreased as dye loading increased. This decrease can be attributed to the limitations of the instrument. Higher dye loadings produced absorbance values above 1, which pushed the limits of the detector. The detector reaches its limits as the sample scans intensity approaches the intensity of the dark scan.

The exemplars were also assigned to their AHC classes and DA was performed. Three classes were formed representing low, medium, and high dye loadings. The observation plot in Figure 8 had a total captured variance of 100% using two canonical variates. Although the 95% ellipses did overlap, the centroids for each class were not captured by the other ellipses. A leave-one-out cross validation was performed and gave a total classification accuracy of 95.60%, which can be seen in Table IV. The medium dye loadings, representing class 2, were confused the most because of the confusion with both the low and high dye loading exemplars.

External Validation. Three additional fibers from each exemplar were analyzed by PCA and DA to evaluate the accuracy of the model predicted in the training set. These additional samples were then placed into the class its spectra most resembled. This was based on the probabilities calculated during DA analysis. The samples were placed into one of the ten exemplar classes for

the first DA analysis, and then placed into one of the three AHC classes for the second DA analysis for comparison purposes. Table V shows the classification accuracies of the external validation. A total classification accuracy of 56.67% was seen when the fibers were placed into the ten classes and 98.67% when the fibers were placed into the three classes. There was confusion between exemplars when the fibers were placed into the individual exemplar classes. This was expected because there was significant confusion in the training set especially as dye loading increased. However, low confusion was seen when the additional fibers were predicted into the AHC class model. Exemplars E and F were the only two confused when exemplars were placed into the three classes. Both misclassified one sample into a wrong class.

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

Conclusions. Chemometric treatment of visible spectra from fibers with different dye loadings has shown to be a reliable and effective way of discriminating between yellow dye loadings when fibers were placed into low, medium, and high classes. Calibration curves, based on proximity matrices, can be produced and accurately predict the dye loading of an unknown fiber. Comparisons of two groups of fibers can provide discriminating information. Overall, forensic fiber examiners can have a more objective way of comparing a known and questioned fiber using chemometric techniques on visible spectra.

Acknowledgements. The authors acknowledge Dana Bors and Maria Diez for their contributions in acquiring spectra for the external validation of the training dataset. This project was supported by Award No. 2010-DN-BX-K220 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.

References

1. M. B. Eyring, B. D. Gaugette. "An Introduction to the Forensic Aspects of Textile Fiber Examination". In: R. Saferstein, editor. *Forensic Science Handbook*. Englewood Cliffs, NJ: Prentice Hall, 2005. Vol. II, Pp. 231-296.
2. S. Walkbridge-Jones. "Microspectrophotometry for textile fiber color measurement". In: M. M. Houck, editor. *Identification of textile fibers*. Boca Raton, FL: CRC Press, 2009, Pp. 165-180.
3. P. Adolf and J. Dunlop. "Microspectrophotometry/Colour Measurement". In: J. Robertson, M. C. Grieve, editors. *Forensic Examination of Fibres*. London, England: Taylor and Francis, 1999. 2nd ed., Pp. 251-287.
4. M. Eng, P. Martin, C. Bhagwandin. "The Analysis of Metameric Blue Fibers and Their Forensic Significance". *J. Forensic Sci.* 2009, 54(4): 841-845.
5. M. C. Grieve, T. Biermann, M. Davignon. "The occurrence and individuality of orange and green cotton fibres". *Sci. Justice.* 2003, 43(1): 5-22.
6. R. Marcrae, R. J. Dudley, K. W. Smalldon. "The Characterization of Dyestuffs on Wool Fibers with Special Reference to Microspectrophotometry". *J. Forensic Sci.* 1979, 24(1): 117-129.

7. K. Wiggins, R. Palmer, W. Hutchinson, P. Drummond. "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.
8. S. L. Morgan, A. A. Nieuwland, C. R. Mubarak, J. E. Hendrix, E. M. Enlow, B. J. Vasser. E. G. Bartick, "Forensic Discrimination of Dyed Textile Fibers using UV-VIS and Fluorescence Microspectrophotometry". Proceedings of the European Fibres Group Annual Meeting, May 25 2004, Prague, Czechoslovakia.
9. K. G. Wiggins, S. R. Crabtree, B. M. March. "The Importance of Thin Layer Chromatography in the Analysis of Reactive Dyes Released from Wool Fibers". *J. Forensic Sci.* 1996, 41(6): 1042-1045.
10. R. Resua, P.R. DeForest, H. Harris. "The Evaluation and Selection of Uncorrelated Paired Solvent Systems for Use in the Comparison of Textile Dyes by Thin-Layer Chromatography". *J. Forensic Sci.* 1981, 26(3): 515-534.
11. M. Huang, J. Yinon, M. Sigman. "Forensic Identification of Dyes Extracted from Textile Fibers by Liquid Chromatography Mass Spectrometry (LC-MS)". *J. Forensic Sci.* 2004, 49(2): 238-249.
12. L. M. Petrick, T. A. Wilson, W. R. Fawcett. "High-Performance Liquid Chromatography-Ultraviolet-Visible Spectroscopy-Electrospray Ionization Mass Spectrometry Method for Acrylic and Polyester Forensic Fiber Dye Analysis". *J. Forensic Sci.* 2006, 51(4): 771-779.
13. C. R. Dockery, A. R. Stefan, A. A. Nieuwland, S. N. Roberson, B. M. Baguley, J. E. Hendrix, S. L. Morgan. "Automated extraction of direct, reactive, and vat dyes from cellulosic fibers for forensic analysis by capillary electrophoresis". *Anal. Bioanal. Chem.* 2009, 394: 2095-2103.
14. J. Robertson, M. Grieve, editors. *Forensic Examination of Textile Fibers*. London: England. Taylor and Francis, 1999, 2nd Ed.
15. K. R. Beebe, R. J. Pell, M. B. Seasholtz. *Chemometrics: A Practical Guide*. New York, NY: John Wiley and Sons, 1998.
16. S. L. Morgan, E. G. Bartick. Discrimination of Forensic Analytical Chemical Data Using Multivariate Statistics. In: R. D. Blackledge, Ed.. *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*. New York: NY: John Wiley & Sons. 2007. Pp. 333-374.
17. J. A. Barrett, J. A. Siegel, J. V. Goodpaster. "Forensic Discrimination of Dyed Hair Color: II. Multivariate Statistical Analysis". *J. Forensic Sci.* 2011, 56(1): 95-101.
18. A. Kher, M. Mulholland, E. Green, B. Reedy. "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, 40: 270-277.
19. B. K. Kochanowshi and S. L. Morgan. "Forensic Discrimination of Automotive Paint Samples Using Pyrolysis-Gas Chromatography-Mass Spectrometry with Multivariate Statistics". *J. Chromatogr. Sci.* 2000. 38: 100-108.
20. J.V. Goodpaster, A. B. Sturdevant, K. L. Andrews, E. M. Briley, L. Brun-Conti. "Identification and Comparison of Electrical Tapes Using Instrumental and Statistical Techniques: II. Organic Composition of the Tape Backing and Adhesive". *J. Forensic Sci.* 2009, 54(2): 328-338.
21. L. J. Marshall, J. W. McIlroy, V. L. McGuffin, R. W. Smith. "Association and discrimination of diesel fuels using chemometric procedures". *Anal. Bioanal. Chem.* 2009, 394: 2049-2059.
22. N. R. Council, *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press, 2009.
23. SWGMAT, *Forensic Fiber Examination Guidelines*. *Forensic Science Communications* 1999, 1(1).
24. K. Varmuza and P. Filzmoser. *Introduction to Multivariate Statistical Analysis in Chemometrics*. Boca Raton, FL: CRC Press, 2009.
25. J.D. Ingle Jr. and S.R. Crouch. *Ultraviolet and Visible Molecular Absorption Spectrophotometry*. In: *Spectrochemical Analysis*, Englewood Cliffs, NJ: Prentice Hall, 1988. 372-380.
26. S.P. Bouffard, A.J. Sommer, J.E. Katon, S. Godber. "Use of Molecular Microspectroscopy to Characterize Pigment-Loaded Polypropylene Single Fibers". *Appl. Spectrosc.* 1994, 48(11): 1387-1393.

Table I. Training dataset with their respective naming designations and dye loadings in weight percent.

Fiber ID	Naming System	% Dye Applied (w/w)
674	A	0.10
675	B	0.20
676	C	0.40
677	D	0.50
678	E	0.75
679	F	1.50
680	G	2.00
681	H	2.50
682	I	3.00
683	J	3.50

Table II. Calibration curve results for three unknown dye loadings using a statistical approach and non-statistical approach.

Statistical Treatment			
Expected Dye Loading	Calculated Dye Loading	Std. Dev.	R ²
0.20%	0.20%	0.058	0.981
0.40%	0.45%	0.038	0.992
0.50%	0.53%	0.050	0.986
Non-statistical Treatment			
Expected Dye Loading	Calculated Dye Loading	Std. Dev.	R ²
0.20%	0.23%	0.122	0.913
0.40%	0.50%	0.094	0.951
0.50%	0.57%	0.119	0.924

Table III. “Leave-one-out” cross-validation confusion matrix of training set exemplars.

From\to	A	B	C	D	E	F	G	H	I	J	Total	%correct
A	46	4	0	0	0	0	0	0	0	0	50	92.00%
B	5	40	5	0	0	0	0	0	0	0	50	80.00%
C	0	4	23	18	5	0	0	0	0	0	50	46.00%
D	0	0	18	18	14	0	0	0	0	0	50	36.00%
E	0	0	1	10	33	6	0	0	0	0	50	66.00%
F	0	0	0	0	7	30	8	5	0	0	50	60.00%
G	0	0	0	0	0	9	27	9	2	3	50	54.00%
H	0	0	0	0	0	7	10	11	5	17	50	22.00%
I	0	0	0	0	0	1	10	17	8	14	50	16.00%
J	0	0	0	0	0	2	7	15	8	18	50	36.00%
Total	51	48	47	46	59	55	62	57	23	52	500	50.80%

Table IV. Leave-one-out cross-validation confusion matrix of the classes generated from AHC dendrogram.

From\ to	1	2	3	Total	%correct
1	94	6	0	100	94.00%
2	3	145	2	150	96.67%
3	0	11	239	250	95.60%
Total	97	162	241	500	95.60%

Table V. External validation results: percentage of positive classifications for fibers grouped into respective classes, and fibers grouped into respective classes generated from AHC dendrogram.

Individual Fiber Exemplars			Class Fiber Exemplars		
Additional Fiber	Class	Accuracy	Additional Fiber	Class	Accuracy
A	A	100 %	A	1	100 %
B	B	93.33 %	B	1	100 %
C	C	46.67 %	C	2	100 %
D	D	46.67 %	D	2	100 %
E	E	60 %	E	2	93.33 %
F	F	86.67 %	F	3	93.33 %
G	G	40 %	G	3	100 %
H	H	40 %	H	3	100 %
I	I	26.67 %	I	3	100 %
J	J	26.67 %	J	3	100 %

Table VI. Classification accuracies and ROC AUC scores for the eleven pair-wise comparisons

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

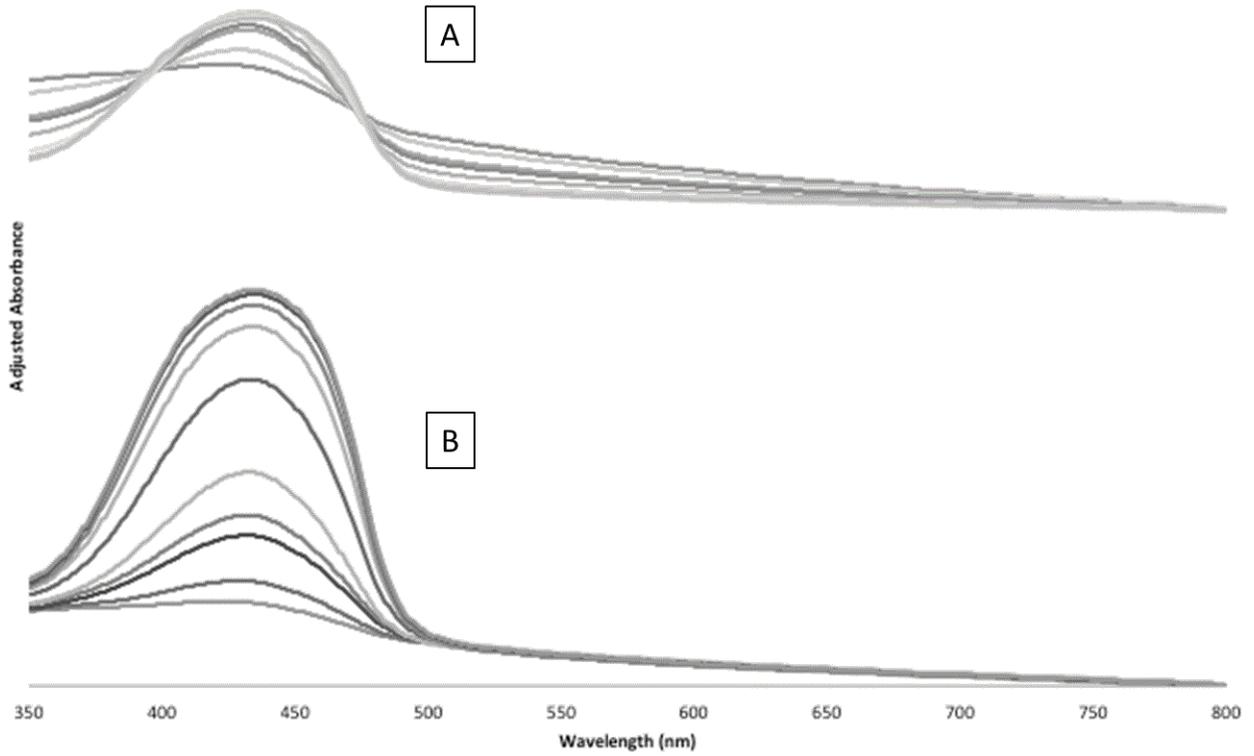


Figure 1. Fiber spectra with adjusted absorbance values for A) background subtracted/normalized spectra and B) background subtracted only spectra

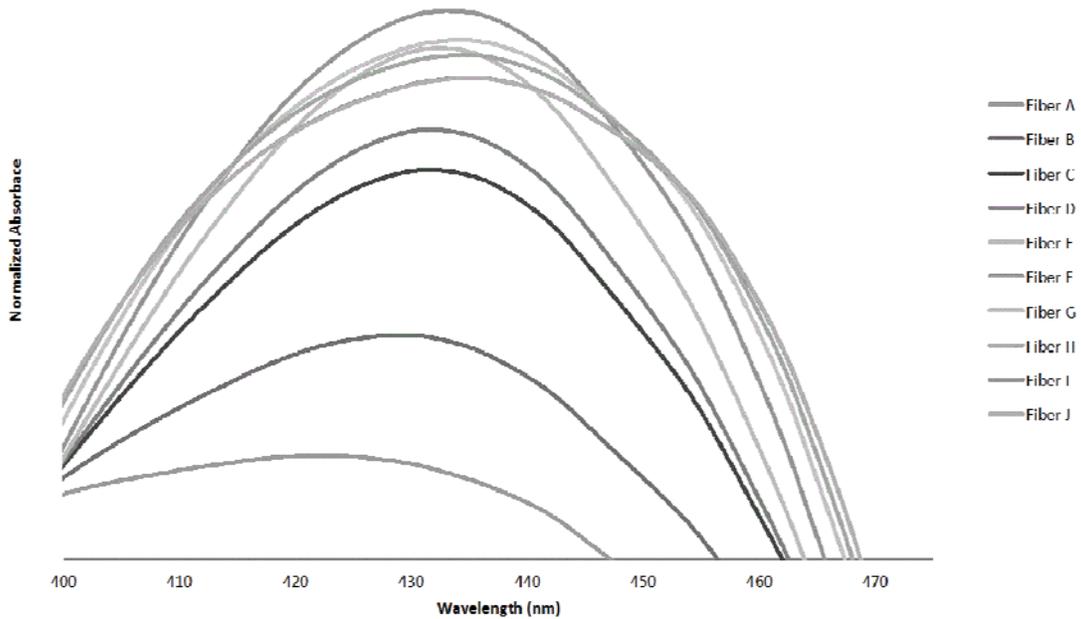


Figure 2. Close-up view of shift in maximum wavelength as dye loading increases

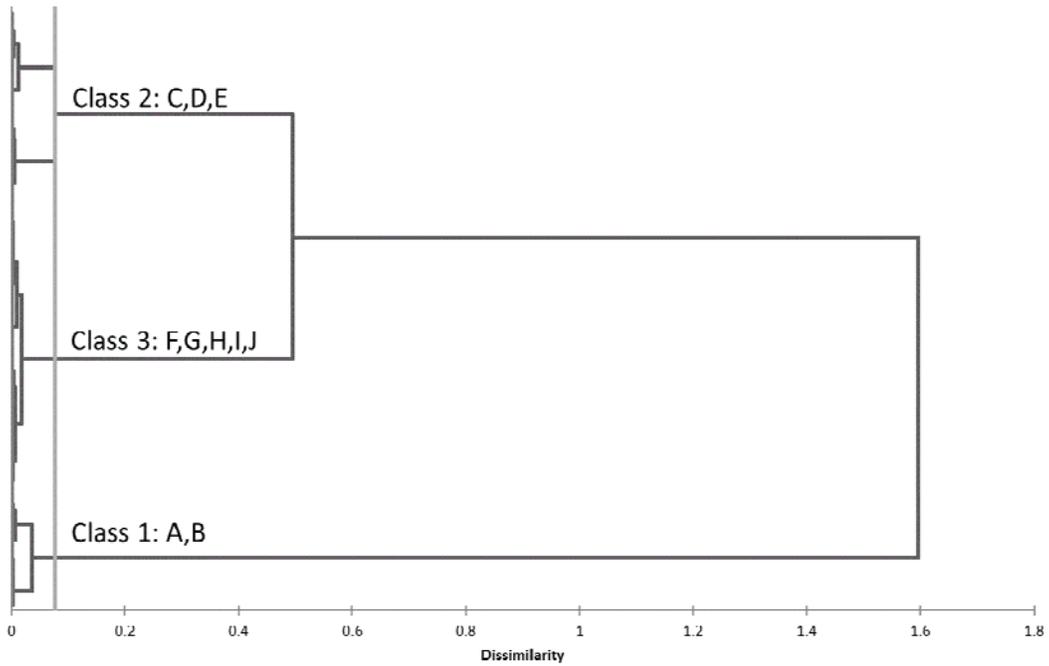


Figure 3. AHC Dendrogram of the 10 exemplars from the training set.

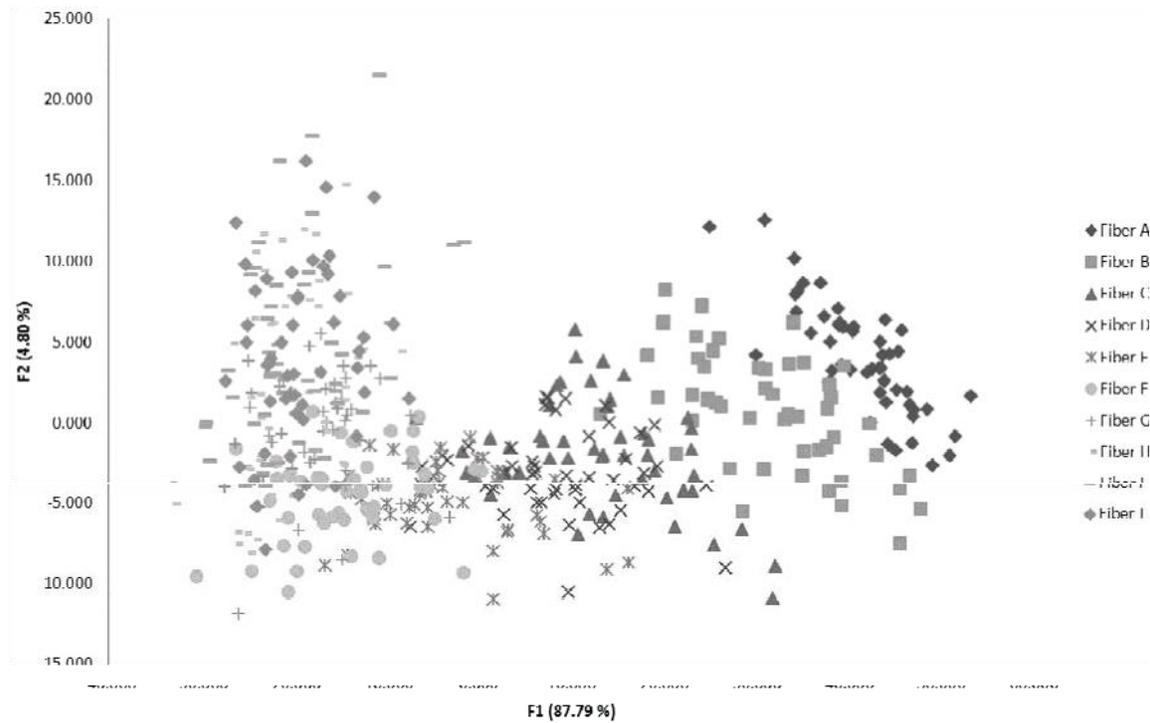


Figure 4. Principal component projections of the ten exemplars from training set, with 92.59% of the variation about the mean accounted for by two principal components.

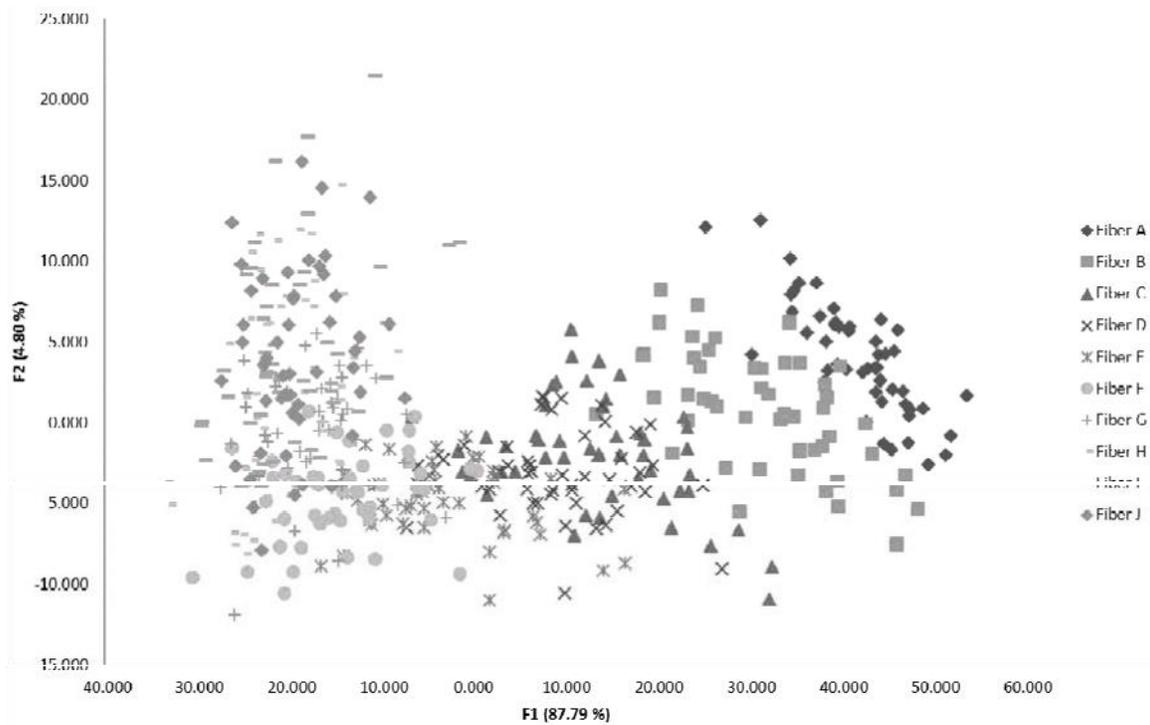


Figure 5. Projections of the three classes generated from the dendrogram (Class 1: A,B; Class 2: C,D,E; Class 3: F,G,H,I,J) onto the principal components shown in Figure 4.

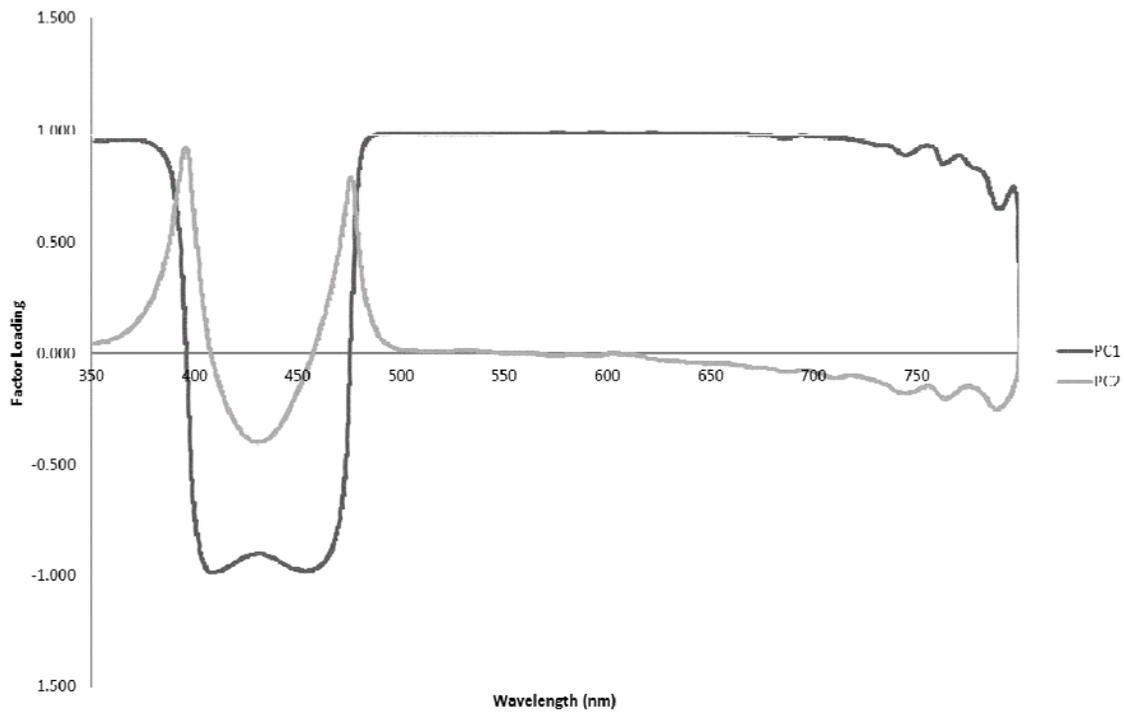


Figure 6. Factor loadings plot for first two principal components (PC1, high correlation around maximum wavelength; also high correlation from 350-375 nm and 500-800 nm; PC2, high correlation along leading and trailing edge of maximum wavelength peak).

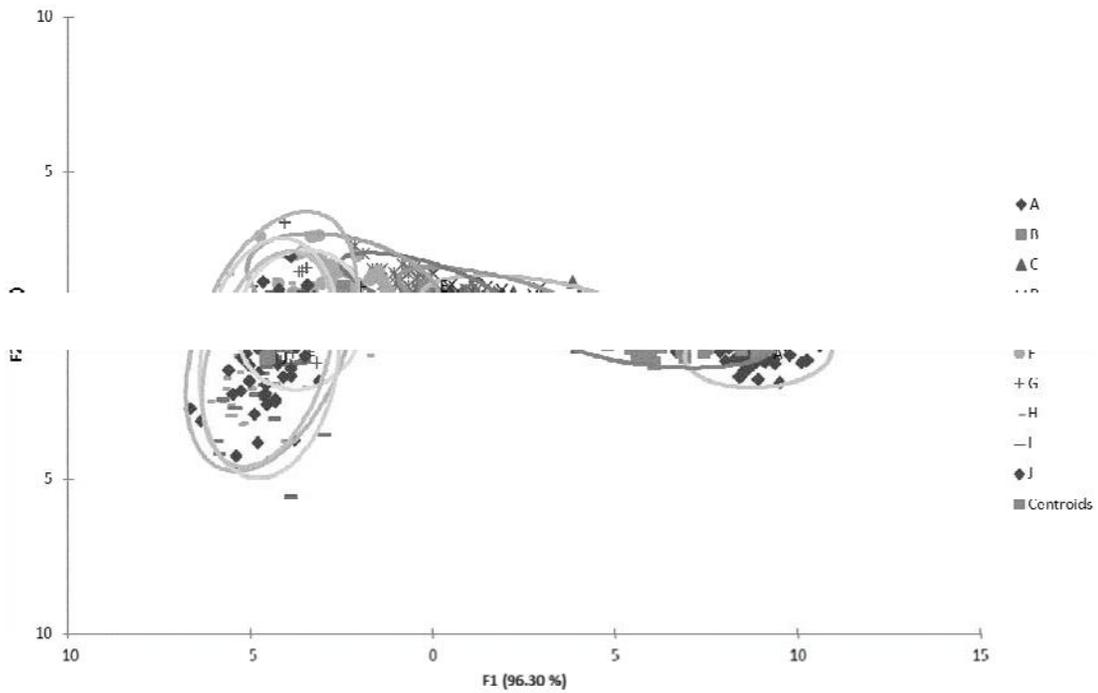


Figure 7. Discriminant projections on the first two canonical variates of training set exemplars.

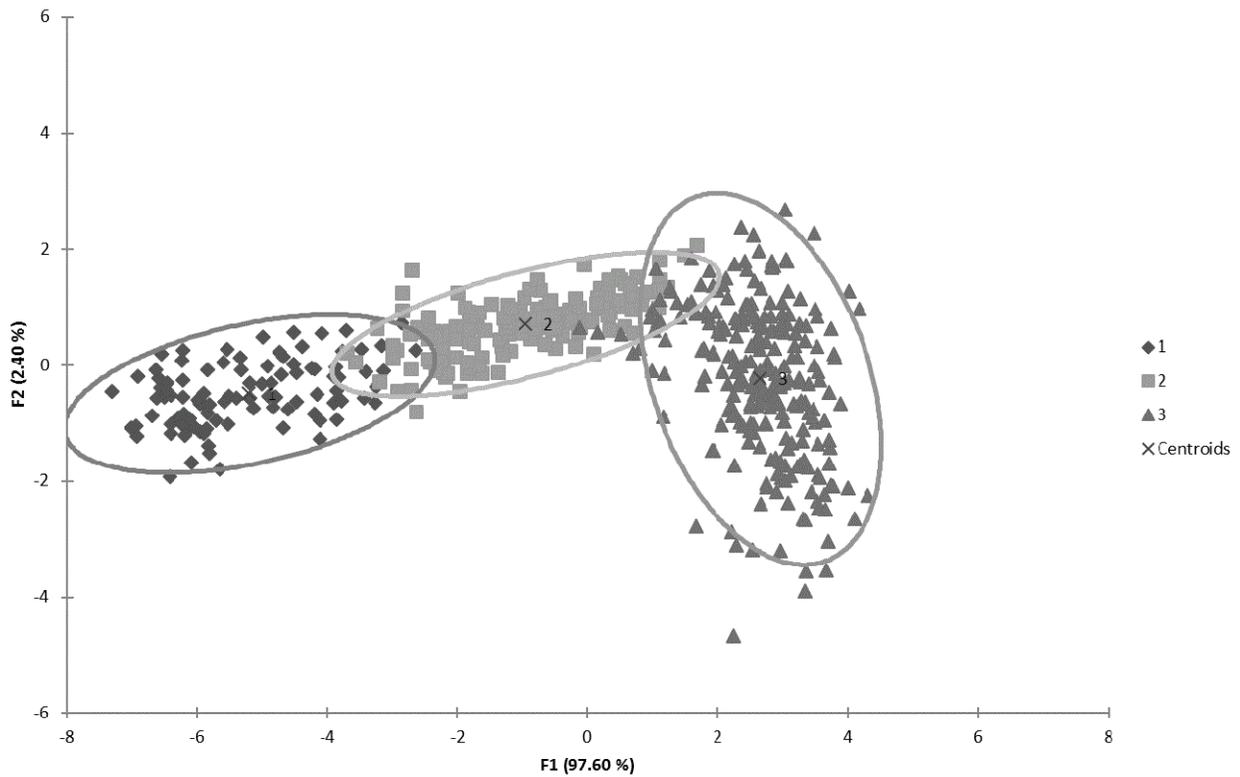


Figure 8. Discriminant projections on the first two canonical variates of classes generated from dendrogram (Class 1: A,B; Class 2: C,D,E; Class 3: F,G,H,I,J).

B. A Statistical Basis for Significance Evaluation of Fibers as Class Evidence

Edward G. Bartick^{1,4*}, Ph.D.; Kevin Roberts¹, B.S.; Laura Schneider¹, B.S.; Stephen L. Morgan², Ph.D.; and John V. Goodpaster³, Ph.D.

¹Forensic Science Program, Department of Chemistry and Biochemistry, Suffolk University, 41 Temple Street, Boston, MA 02114.

²Department of Chemistry and Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC 29208.

³Department of Chemistry and Chemical Biology, Forensic and Investigative Sciences Program, Indiana University Purdue University Indianapolis (IUPUI), 402 North Blackford Street, LD 326, Indianapolis, IN 46202.

⁴Current address: Department of Forensic Science, The George Washington University, 2100 Foxhall Rd. NW, Washington, District of Columbia 20007.

*author to whom correspondence should be addressed.

Key words: fibers, trace evidence, class evidence, statistical analysis, multivariate analysis,

Abstract. The objective of this research was to demonstrate the statistical significance of measurement variance of like and unlike fibers and to demonstrate how the estimations of match significance can be made from fiber evidence. In addition, the process can be applied to class evidence in general. The work was conducted through the use of multivariate statistical analysis and the product rule of independent variables. We have shown how to increase the ability to classify and discriminate synthetic and cotton fibers. A library of 923 well characterized fibers was developed with known dye components. With the use of visual light microscopy, UV/visible microspectrophotometry (MSP), and Fourier transform infrared (FTIR) microspectroscopy, data was collected on fibers. Studies were done on twenty-one red cotton and twenty-one red acrylic fibers using multivariate statistical analysis. The absorption spectra of fibers from 10 replicate UV/visible microspectrophotometry scans on each fiber were compared by using principal component (PCA) and linear discriminant analysis (LDA). With the aid of multivariate statistics, fibers that are difficult to distinguish by visual comparison were distinguished. In addition, for the acrylic fibers, FTIR spectra were used to identify generic and sub-generic class. The characteristics of cross sectional shape and diameter were determined. Finally, using the product rule of probability, knowing the number of fibers in the database with specific color, diameter, cross-sectional shape, and chemical composition, the percentage occurrence of each fiber was determined. The product of the percentages was then calculated to determine the probability of two fibers matching randomly with those characteristics. Probabilities on the order of 1 in 0.5 million are obtainable with such comparisons between fibers, provided a sufficiently large and representative database of fiber characteristics is accessible. The improved understanding of sources of variability and decision-making processes gained from this research will serve to advance the forensic significance of involving fiber and other class evidence material examinations.

Introduction. Fiber evidence can be vital in forensic case work. The transfer of fibers during violent crimes often occurs when victims have a physical altercation, their bodies are dragged over residential carpets and/or subsequently placed in automobiles.¹⁻⁴ Perhaps the highest profile case involving fibers was the murder of 30 African American young men and children in Atlanta,

during 1979 to 1980. Multiple transferred carpet fibers, residential and automotive, found on the victims played a pivotal role in the conviction of Wayne Williams for the murder of two of the victims.^{5,6}

Fibers found at crime scenes or potentially related to the crime often share properties or characteristics due to common or similar sources, resulting from manufacturing methods, or treatments. The standard positive conclusion typically derived from a forensic fiber examination is the equivocal statement that “The questioned fiber exhibits the same physical, optical and chemical properties as the known sample. Therefore, these fibers could have originated from the same source as the known sample or another fiber source composed of fibers with the same properties”.⁷⁻⁹ While this conclusion is accurate, and is the current accepted statement for dealing with class evidence, no statistical basis of evidential significance is currently accepted.

With development of forensic databases such as CODIS for DNA, PDQ for automobile paint, and AFIS for fingerprints, the creation and maintenance of databases for trace evidence is recognized to be of paramount importance in addressing issues of class evidence. The Research, Testing, Development and Evaluation Interagency Working Group (RTD & E IWG) of the National Science and Technology Council Subcommittee on Forensic Science asked, “What databases are most needed in the field of fiber analysis?”¹⁰ The Scientific Working Group for Materials Analysis (SWGMA) responded: “An up-to-date, comprehensive automotive carpet fiber database along the lines of the PAINT Data Query (PDQ)—searchable for investigative leads”. SWGMA also stated that “a fiber population database is not recommended for statistical use at this time due to the ever changing colors of fibers and textiles due to style and season as well as post-manufacture changes from exposure to sun, laundering, *etc.* The wide range of countries that manufacture fibers would also make it almost impossible to get a truly representative sample—it would not be advisable to reference a fiber found in a case against a database that may be limited by fiber manufacturer”. Databases are usually specific to manufacturers, and other publications paint only broad strokes in regard to manufacturing output; *e.g.*, *Fiber Organon* reports on world fiber production.¹¹ Houck was among the first to voice dissatisfaction with lack of appropriate statistics in trace evidence evaluation.¹² The 2009 National Academy of Sciences highlighted this issue when recommending, “A statistical framework that allows quantification of these claims is greatly needed” (p. 189).¹³ In regard to fiber analyses, the National Academy of Sciences Report states that “it would be possible in principle to develop an understanding of the uncertainties associated with those analyses. However, to date, this has not been done” (p. 163).¹³

This work aims to: (1) develop a comprehensive, well characterized fiber database; and, (2) to investigate statistical approaches that can provide improved decision-making support for forensic fiber comparisons. We have sought to achieve this goal by establishing a large collection of physical samples of fibers of known dye components. Measurements have included physical characteristics such as cross sectional shape and diameter determined by visual light microscopy, PLM measurements such as birefringence and sign of elongation, and color characteristics and chemical composition as determined by UV/visible and infrared (IR) microspectrophotometry. Cross-validation comparisons of these measurements, interpretation of matches and exclusions via statistical methods have been conducted in this work.

For a given type of class evidence to be truly probative, it must have a reasonable frequency of occurrence in matters of legal relevance, significant and well-documented diversity must be present in the population under study, and laboratory methods must exist that can reliably discern

this diversity. For the evidence to be of forensic value, the probability of a coincidental association between an unknown and known exhibit should be low, and the burden is on the forensic scientist to evaluate this risk. This can be done by demonstrating that the sample type in question is both diverse and differentiable using the following key steps: (1) Understand the product population, including manufacturing and distribution; (2) Obtain a large, representative collection; (3) Analyze samples using multiple, orthogonal techniques to maximize information content; (4) avoid microheterogeneity with appropriate sample sizes; (5) Assess diversity of the sample collection with rigorous quantitative methods and employ appropriate statistical hypothesis tests for evaluation of match probabilities. The diversity and depth of a database is critical to its usefulness as a foundation for establishment of the statistical significance of decision-making. We use our well characterized fiber database to test the reliability of our measurements. Future development of a sizable representative database of specific material types will be necessary to conduct statistics on case evidence.

The “product rule for independent events” was used in the Williams trial to show improved evidential strength based on occurrence of multiple independent events.^{5,6} This rule states that for two or more independent events, the probability of all events occurring simultaneously is the product of the individual probabilities.¹⁴ Independent events are specified, meaning that the outcome of the one event cannot influence another in any way. This is true for 'orthogonal' measurements that characterize different aspects of an evidence item, but may not be true of all measurements. While such probabilities could be very small, fibers are class evidence. It cannot be stated unequivocally that fibers originated from the same source. However, this approach can provide support for small random match probabilities of finding like fibers on both a suspect and at a crime scene.

For decision-making in the comparison of individual fibers, both parametric (assuming specific probability distributions) and nonparametric statistical methods, along with multivariate methods such as principal component and linear discriminant analysis and associated hypothesis testing have been employed with practical tools for determining the significance of a match of questioned and known fibers. We define a match of two materials as having measured physical and spectral properties within a statistical tolerance of variability, informed by the outcome of statistical hypotheses tests. The practical significance of such a result will be enhanced by information provided from the database on the discriminating power of various measurements. Ultimately, by combining a statistically based approach for judging the match quality of individual fibers with database information, this report will provide a scientific basis for significance of fiber associations in forensic practice.

Statistical evaluation of fiber comparisons. Instrumental methods produce data of high dimensionality. Microspectrophotometry produces spectral intensities (absorbance) at hundreds to thousands of wavelengths. Therefore, multivariate methods are required rather than the commonly used univariate methods of statistical analysis. Means and standard deviations (or variances) are also employed in multivariate statistics, but must be calculated for each variable; additionally, covariances (or correlations), measuring the strength of the linear relationships between variables, are calculated. These statistics are not single numbers, but are arrays (matrices) of numbers describing the correlations between all the measured variables. Increases in computing power have made computationally intensive data analysis feasible, and increased availability of software for multivariate statistics has made these techniques accessible.

Principal component analysis (PCA)^{14,15}, first discussed by Hotelling¹⁶, projects multivariate data into a reduced dimensional display that retains as much of the original variability as possible in as few dimensions as possible. Linear discriminant analysis constructs a lower dimensional data display which best separates predefined groups (e.g., groups of replicate spectra of the same fiber) by defining axes that maximize the ratio of their between- to within-group variances.⁵⁸⁻⁶² If a sufficiently large proportion of the variability associated with the first few discriminant axes, a projection of the data points (the spectra) in the two- or three-dimensional space of the discriminant vectors permits the researcher to visualize clustering and similarity of the data. Clustering of similar samples can be assessed by comparison to the distances between spectra judged different from one another. The multivariate generalization of the univariate Student's *t* test is Hotelling's T^2 test for the equivalence of means.¹⁷

Materials and Methods

Materials. The fiber collection at the University of South Carolina (USC) began 2003 and currently contains over 1200 fibers, including samples of dyed and undyed acrylic (polyacrylonitrile, co-acrylate and co-methacrylate, sulfonates), cotton, nylon, (-6 and -6,6), polyester, and polyester/cotton blend fabrics, along with smaller of other less common textiles. The sample fabrics contain one to three fiber variants and are dyed with up to six different dyestuffs, often representing several dye classes. Some of the samples are dyed with up to four dyes of a single dye class. The collection presently contains fibers from the textile industry dyed with 273 different dyes with chemical samples of the dyes employed.

Chemical information and physical characteristics are available for most of the fibers (polymer, color, diameter, cross sectional shape, denier, dye class, dyes used, PLM measurements, etc.). Currently, 10 replicate UV/visible absorbance spectra (200-800 nm). IR spectra are present for some samples. Large stock samples of fabrics, from which many of the fibers in the original USC collection were sampled, are maintained in acid-free plastic storage bags located in filing cabinets. Small representative samples are stored in acid-free protective pockets in three-ring binders. All these fiber samples are kept in a dark air-conditioned room.

A second, reference collection of fibers from Microtrace, LLC (Elgin, IL) was also used in this work. This collection contains 201 fibers, many undyed, having a wide range of chemical composition. The FBI FTIR fiber library, version 4.0(), consisting of 83 different polymeric compositions of fibers used to aid in sub-generic composition classification of the fibers.

Methods. Microspectrophotometry analysis was conducted with a CRAIC (San Dimas, CA) 380 PV UV/visible spectrophotometer with a Carl Zeiss (Thornwood, NY) Axioscope A1. Fibers were mounted on glass slides using Permout® (Fisher Scientific, Fair Lawn, NJ) and a coverslip. The spectral data acquisition range was 400-800 nm with a resolution of 0.5 nm.

FTIR analysis was conducted with a Bruker (Billerica, MA) Tensor 27 Bench and a Hyperion Microscope. All fibers were flattened for FTIR analysis using a Port-A-Press (International Crystal Laboratories, Garfield, NJ) with a 4 mm die while the fibers were suspended across a loose leaf binder hole reinforcement ring attached by the adhesive on the ring. A resolution of 4 cm^{-1} was used and 10 spectra were collected for each sample. Sixty-four scans were single averaged to produce each spectrum. Prior to acquiring sample spectra, a background scan was obtained. Spectra were obtained from 4000–650 cm^{-1} . The fiber diameters were measured 10 times along the length and averaged using a microscope stage micrometer. The tolerance when

comparing fibers was ± 2 μm . Cross-sectional shapes were determined when samples were entered into the USC database.

Multivariate statistical analysis was conducted using a software package, Spectrum eXplorer (SPX), developed at in-house in MatLab ver. 7, 2013 (MathWorks, Natick, MA). PCA and LDA were implemented as described above. Leave-one out cross validation was used to estimate the classification accuracy of the LDA model by removing each sample from the data set in turn and recomputing discriminant functions based on the remaining samples. Thus, estimates of the classification error for each sample are obtained without using that sample to calculate the discriminant model. The similarity expressed as Mahalanobis distance, between the jackknifed sample and the mean vectors for each group, calculated to assign group membership. Elliptical confidence ellipses representing distances statistically equidistant from the group mean for a predetermined level of probability (95% in this case) were based on Hotelling's T^2 statistic (the multivariate generalization of Student's t statistic).¹⁶⁻¹⁸

Results and Discussion. Eighty UV/visible spectra of red cotton fibers are shown in Figure 1, along with the projection of the spectra following PCA/LDA.¹⁸ Each of ten replicate spectra from eight different fibers are shown in Figure 8 (top); clearly, replicate spectra of the first five fibers are clearly distinguished from one another, whereas spectra of fibers 6-8 are not distinguishable. When the spectra are normalized, subjected to PCA for dimensionality reduction, then LDA for discrimination, projections of the 80 spectra into the space of the first three linear discriminants analysis are shown in Figure 8 (bottom). This modeling process is straightforward and performed in just a few minutes using our menu-driven statistics software. The resulting map depicts the similarity of the 8 fibers, as points plotted in a 3-dimensional space with 95% confidence ellipses for the 10 spectra corresponding to each fiber. For the present data, the classification accuracy is 88.75%, with 71/80 spectra correctly classified, and with misclassifications occurring between spectra of groups 6, 7, and 8 (at the right side of the LDA plot). This outcome is exactly what was expected from visual inspection of the spectra. The multivariate analysis program is a tool that provides statistical rationale for the excellent discrimination of fibers 1-5, and for a decision that fibers 6-8 are more difficult to distinguish from one another.

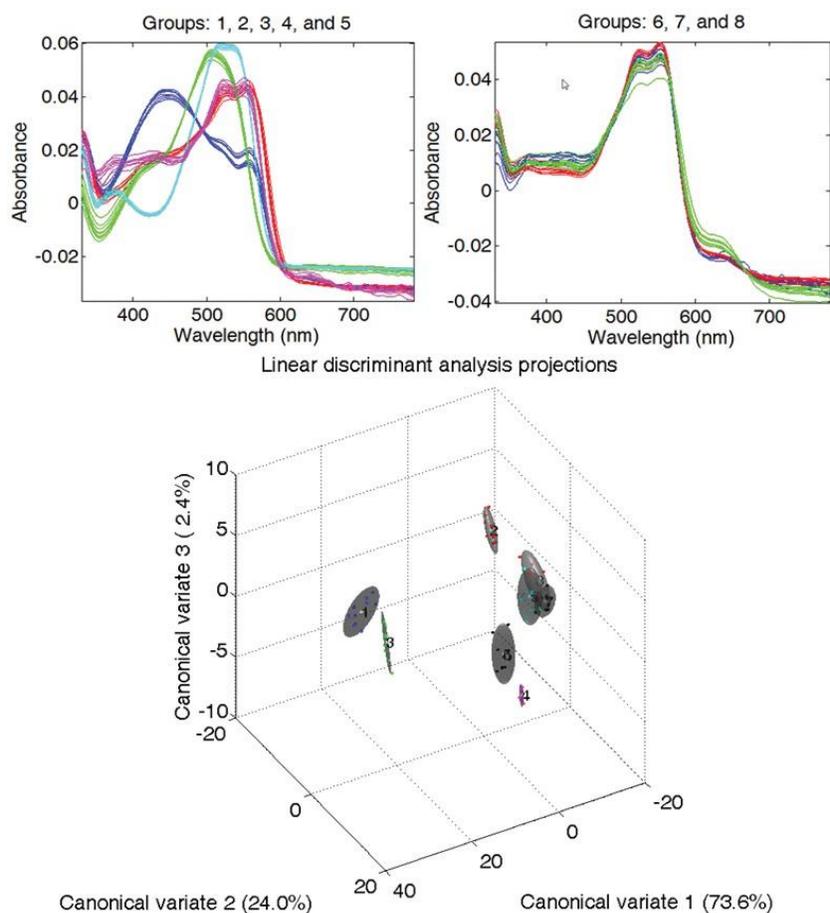


Figure 1. (top) plots of UV/visible spectra for red cotton fibers 1-5, and fibers 6-8. (bottom) Linear discriminant projections after PCA (for 10 replicate spectra of 8 red cotton fibers).

The second example, presented at the SWGMAT, NEAFS 2012 and AAFS 2013 Meetings,¹⁹ is based on red acrylic fibers selected from a database of 849 total fibers. The fibers in the database have been characterized in terms of dyes used, cross-sectional shape, and diameter. The fibers were grouped by those parameters with the diameters at a tolerance of $\pm 2 \mu\text{m}$ (Table 1). There was one group of round fibers and one group that had a bean shaped cross-section. Fibers were then categorized by their dye composition from LC/MS. With the different combinations of dyes, various shades of red are produced as shown in Table 2, thus producing 9 groups. Of the 21 red acrylic fibers, two pairs of red fibers with near identical physical characteristics and visual appearance were examined (see photos in Figure 2). These two fibers (fibers 74 and 76) are bilobal, $17.7 \mu\text{m}$ in diameter and are dyed with CI Yellow 28, CI Violet 16, and CI Blue 41 dyes. Fibers 73 and 75 differ, having a slightly different diameter of $20 \mu\text{m}$ and dyed with a combination of CI Yellow 28, CI Red 48, and CI Blue 147 dyes. In Figure 3, the visible spectra fibers of 74 and 76 are similar, with differences in the 600-700 nm region which appear to be a variation in the red component. The same red component variation appears in the spectra for fibers 73 and 75.

Table 1. Cross sectional shape and diameter of acrylic fibers.

Bilobal Cross Section			Round Cross Section		
Group	FID	Diameter (± 2)	Group	FID	Diameter (± 2)
1	60	16.25	9	78	23.75
	72	15		132	23.75
2	60	16.25	Bean Cross Section		
	74	17.5	Group #	FID	Diameter (± 2)
	76	17.5	10	123	22.5
	79	17.5		129	21.25
3	74	17.5			
	76	17.5			
	77	18.75			
	79	17.5			
4	73	20			
	75	20			
	77	18.75			
5	73	20			
	75	20			
	80	21.25			
6	80	21.25			
	81	22.5			
7	106	25			
	125	25			
8	107	27.5			
	126	28.75			
	131	26.25			

Table 2. Dyes present on acrylic fibers, tabulated by fiber number.

Dye 1	Dye 2	Dye 3	Dye 4	Dye 5
Yellow 28, Red 18, Blue 41	Yellow 28, Violet 16, Blue 41	Yellow 28, Red 46, Blue 147	Yellow 28, Violet 16	Yellow 28, Violet 16, Blue 60
60	72	73	77	78
106	74	75	132	80
107	76	123	133	130
	129	162		
Dye 6	Dye 7	Dye 8	Dye 9	
Yellow 28, Violet 16, Blue 147	Yellow 28, Red 46, Blue 41	Yellow 28, Yellow 21, Red 18, Blue 147	Doracryl Red GL	
79	159	164	887	
125	161	165	888	
126			889	
			890	

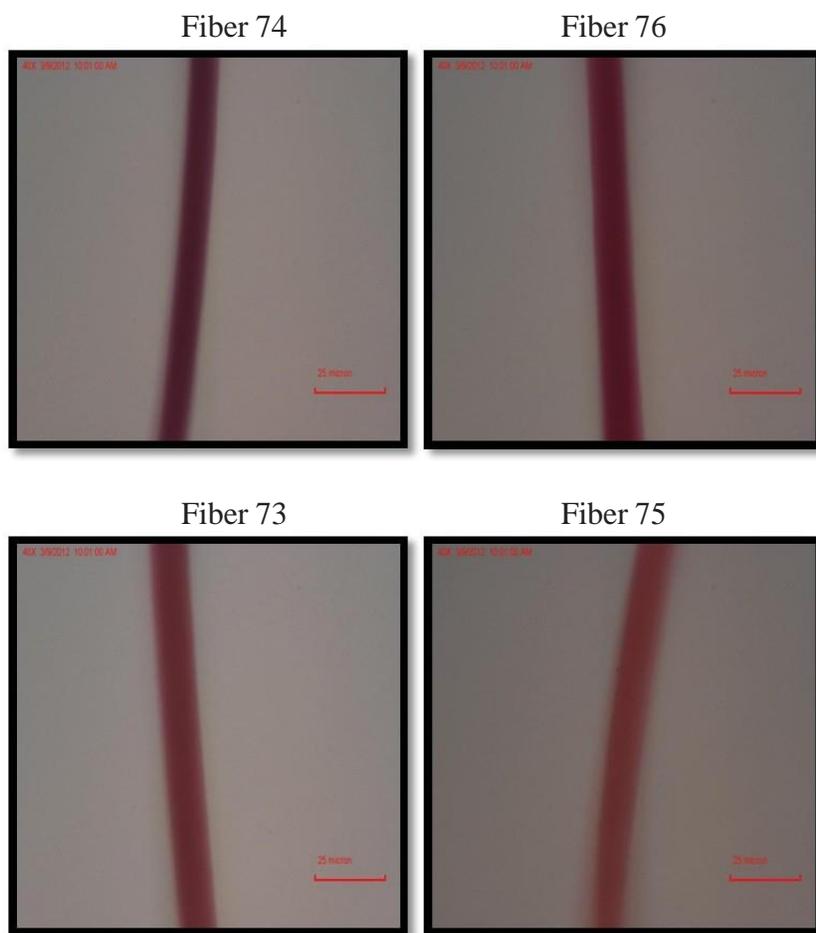


Figure 2. Photographs (40u) of two pairs of fibers (74 / 76, 73 / 75) with near identical physical characteristics and visual appearance.

The PCA analysis shown in Figure 4 confirms that each pair of spectra are significantly different from one another. However, to determine the ability to discriminate these fibers from the additional red acrylic fibers within the database, we must conduct PCA and LDA on all 21 of the red acrylics.

Figure 5 shows the LDA plot of all the red acrylics in the database. With no preprocessing only a 70% discrimination classification was achieved. By subtracting non-zero values, mean centering and normalized to the unit area, the overall discrimination classifications was 97%, which is excellent for such a large number of comparisons. The circled fibers are the pair in question from Figure 2. Three of four were clearly separated from the rest. However, when looking at fiber 76 (Group 6) it is close to fiber 125 (Group 15). Upon examining the cross validation results, two samples between Groups 6 and 12 are misclassified. Thus, a 90% correct classification was achieved between the 20 scans for the two groups.

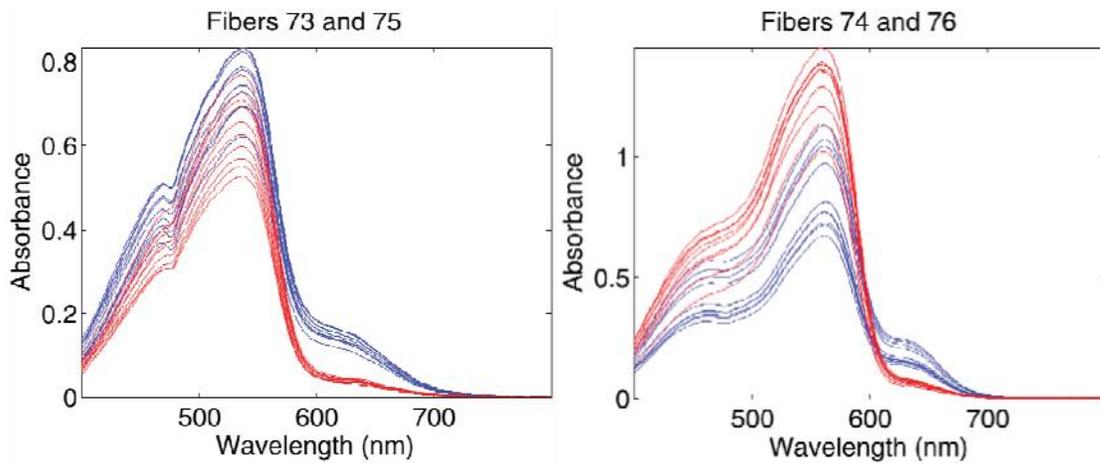


Figure 3. UV/visible spectra of the two pairs of fibers shown in Figure 2. (Right) fibers 73 (blue trace) and 75 (red trace); (Left): fibers 74 (blue trace) and 76 (red trace).

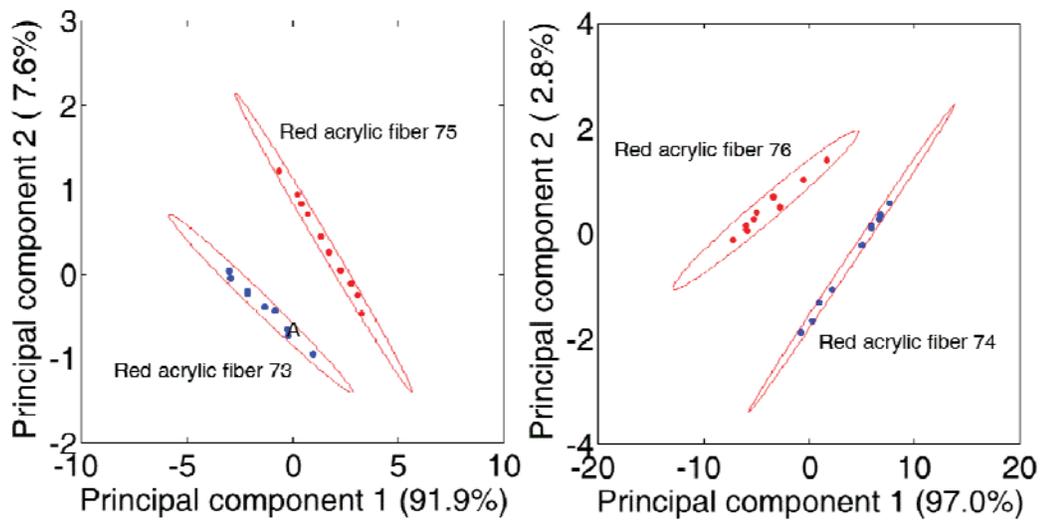


Figure 4. PCA discrimination of UV/visible spectra of each pair of physically indistinguishable fibers.

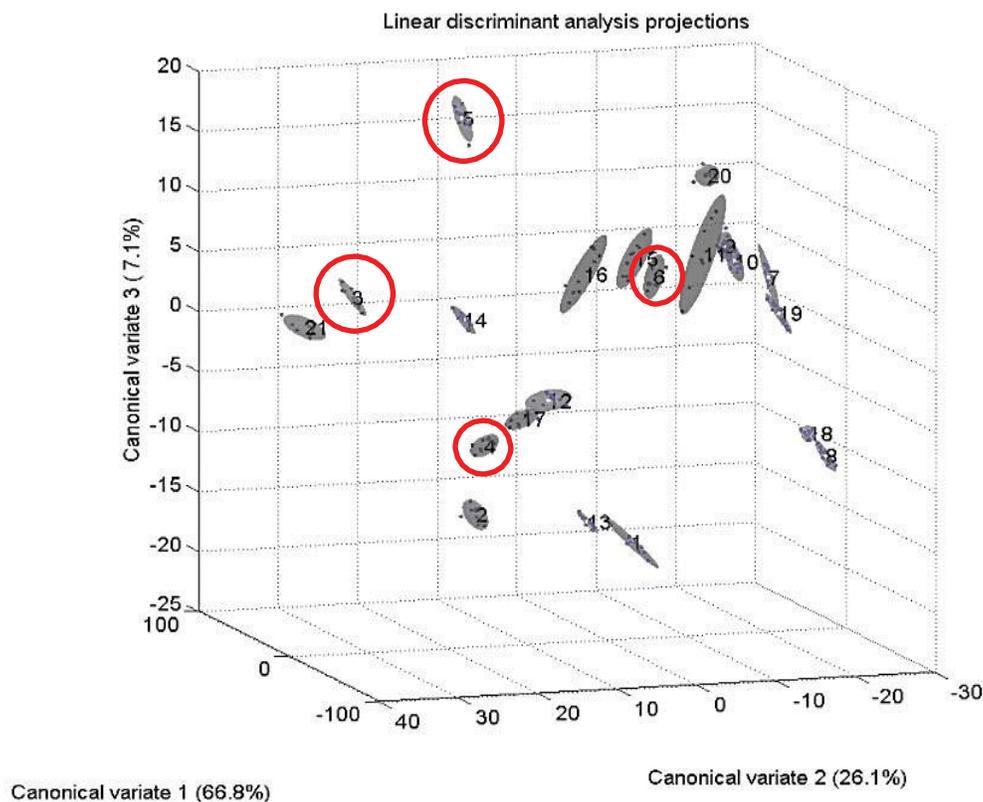


Figure 5. Linear discriminant analysis projections of 21 acrylic fibers. Circled Groups: upper group 5 is fiber 75; middle left group 3 is fiber 73; middle right group 6 is fiber 76; lower group 4 is fiber 74.

For DNA analysis in the CODIS database, 13 alleles of short tandem repeats (STRs) are used to determine the random probability of two individuals having the same STRs. The probability is determined by the product rule of independent events, calculated by multiplying the decimal ratio of each STR determined by analyzing a specific number (about 100 ea.) of representative individuals. For example for Caucasian Americans, that number is 1.718×10^{-15} . That number is then divided into one to produce the probability of one in 582 trillion. Such low random match probabilities calculated for two individuals is the statistical basis for the discriminating power of DNA analysis.²⁰

In the present example of fiber discrimination by UV/visible MSP, two red fibers with different diameters have been distinguished from all others in the database of 849 fibers. If a sufficiently diverse and representative data base of fibers were available, the probability of this outcome might be determined within the given population of that database by similar means to that done with DNA. Fibers are manufactured with many different characteristics such as generic class, sub-generic class, dyes used, diameter, cross-sectional shape, and delusterants. Each potentially discriminating feature can contribute, often independently to the probability of a random match. The following discussion is not meant to be realistic, only illustrative of the potential for statistical significance in fiber matches.

Table 3 presents the fractional representations of several fiber characteristics in our current database. With 202 acrylic fibers in the database of 849 fibers total, the fraction of acrylic fibers is 0.2379. The 116 red fibers were narrowed down to one using MSP and PCA analysis, producing a fraction of 0.00118. A number 57 of the 849 fibers are bilobal, giving a fraction of 0.0671. Fibers having 20 μm diameters produced a fractional amount 0.0766. The product of these fractional probabilities is 1.441×10^{-6} . Thus, the probability of occurrence of a red acrylic, bilobal, 20 μm diameter, fiber at random in the current database with the specified characteristics is one in 694,000 as shown in Table 3.

Table 3. Probability of occurrence of fibers 73 and 75 in databae.

Property	No. of Fibers in Database	Match Probability
Acrylic	202/849	0.2379
Red	1/849	0.00118
Bilobal	57/849	0.06714
Diameter 20 \pm 2 μm	65/849	0.07660
Product Probability		0.000001441 One in 693,854

In the present study, only the generic class of acrylic fibers was considered. Ten separate FTIR fiber spectra of each sample were collected each scanned 64 times, base-line flattened and then averaged to one spectrum. These spectra enable sub-classification of the fiber composition as being poly(acrylonitrile: co-vinyl acetate, sulfonate). Acrylic fibers constitute one of approximately 22 sub-classes.²¹ Based on a study by Tungol, *et al.*, acrylic fibers produced were found to occur in approximately one-third of case work.²¹ In Figure 6, two circles highlight additional minor differences in the spectra suggesting that a study of the sub-generic frequencies and variations might produce even lower probabilities of randomly finding specific fibers.

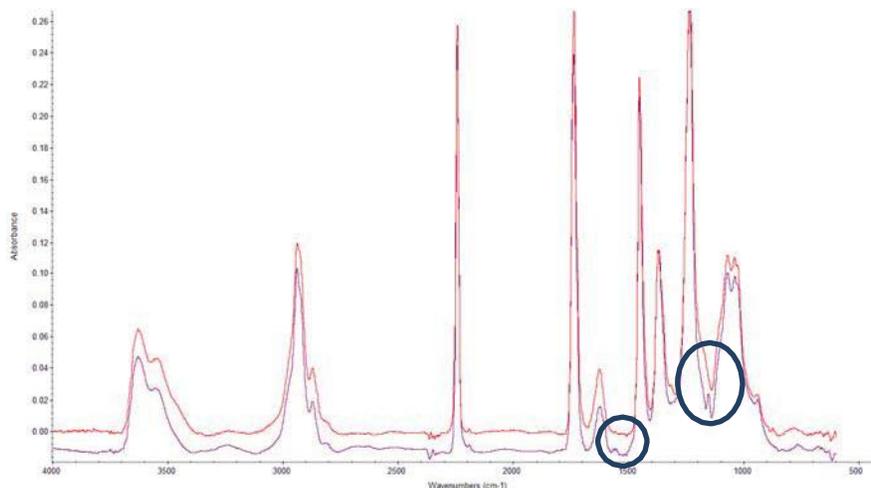


Figure 6. FTIR spectra fibers identified as poly(acrylonitrile: co-vinyl acetate, sulfonate).
Upper trace: fiber 73; Lower trace: fiber 75.

Conclusion. The premise of this work is that if enough is known about the distribution of a population from which questioned fibers from a suspect and known fibers from a crime scene are class members, multiple associated characteristics (physical, optical, or spectroscopic) decrease the random probability of a match occurring solely by chance. However, it is important to understand that this work is an illustration using a limited database of controlled characteristics. To consider a database usable and realistic when dealing with fibers as evidence, it is necessary to have a large number of representative samples that are typical within the geographic region where the crime occurred. With a good database, the use of instrumental measurements of multiple characteristics can be applied to a range of class materials to achieve some measures of statistical significance for trace evidence matches.

Acknowledgement. This research was supported provided by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice Award No. 2010-DN-BX-K220. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

1. Ballou, S. "Wigs and the significance of one fiber." In: *Mute Witnesses: Trace Evidence Analysis*, Houck, M. M. Ed., Academic Press, San Diego, CA, 2001, pp. 21-48.
2. Deedrick, D. W. "Searching for the source: Car carpet fibres in the O.J. Simpson case," *Contact* 1998, 26, 14-16.
3. Houck, M. M. "A case of cross-transfer." In: *Mute Witnesses: Trace Evidence Analysis*, Houck, M. M. Ed.; Academic Press: San Diego, CA, 2001, pp. 175-186.
4. Houck, M. M. "My roommate is using the refrigerator." In: *Trace Evidence Analysis: More Cases from Mute Witnesses*, Houck, M. M., Ed., Academic Press, San Diego, CA, 2003, pp. 233-250.
5. Post, H.; Hilder, D. B. "Fibers found on victims form links in Williams case," *The Atlanta Journal*, 2 February 1982, p. 1A.
6. Deadman, H. "Fiber evidence and the Wayne Williams trial." *Law Enforcement Bulletin*; U.S. Government Document J1.14/8a:F44, Federal Bureau of Investigation, U.S. Department of Justice, FBI, March and May, 1984.
7. Scientific Working Group for Materials Analysis (SWGMAT), Forensic Fiber Examination Guidelines, 1999; URL: <http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf>
8. Harmon, R. P.; Clarke, G.; Michaud, A. L.; Plourd, C. J. "Daubert Presentation," *Trace Evidence Symposium*; Clearwater Beach, FL, 6 August 2009; URL: <http://projects.nfstc.org/trace/2009/day4.htm>.
9. SWGMAT, Fiber Evidence, "Courtroom Education and Admissibility Response" , URL: <http://SWGMA.org/>.
10. SWGMAT, "Response to Fiber Analysis Question List by the Research, Testing, Development and Evaluation Interagency Working Group (RTD&E IWG)," National Science and Technology Council Subcommittee on Forensic Science; URL: <http://www.swgmat.org/fiber.htm>.
11. *Fiber Organon*, Fiber Economics Bureau, Washington, DC, URL: <http://www.fibereconomics.com/>.

12. Houck, M. M. "Statistics and trace evidence: the tyranny of numbers," *Forensic Sci. Commun.* 1999, 1(3); URL: <http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct1999/houck.htm>.
13. National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press, Washington, DC, 2009.
14. Jackson, J. E. (1991) *A User's Guide to Principal Components*, John Wiley & Sons, New York, NY.
15. Jolliffe, I. T. (2002) *Principal Component Analysis*, 2nd ed., Springer-Verlag, New York, NY.
16. Hotelling, H. "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, 1933, 10, 69-79.
17. Hotelling, H. "The generalization of Student's t-ratio," *Annals of Math. Statist.* 1931, 2 (3), 360-378.
18. Morgan, S. L.; Bartick, E.G.; Goodpaster, J. V.; Birt, D. L.; Burnip, M. R.; Reichard, E. J.; Roberts, K., "Chemometrics and databases for comparisons of spectral data from trace evidence," paper at SciX 2012 (sponsored by the Federation of Analytical Chemistry and Spectroscopy Societies), Kansas City, MO, Kansas City, MO, 2 October 2012.
19. Bartick, E. G.; Roberts, K.; Morgan, S. L.; and Goodpaster, J. V., "A Statistical Approach to Discrimination and Match Capability to Provide Scientific Basis for Estimating Significance of Fiber Associations in Forensic Practice", Annual Meeting of the American Academy of Forensic Sciences, Washington, DC, February 2013.
20. Aitken, C. G. G. *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester, 1995.
21. Grieve, M. C. "Another look at the classification of acrylic fibres, using FTIR microscopy", *Sci. Justice.* 1995, 35, 179-190.
22. Tungol, M. W.; Monteser, A.; Bartick, E. G. "The analysis of single polymer fibers by Fourier transform infrared microscopy: The results of case studies", *J. Forensic Sci.* 1991, 36, 1027-1043.

C. Comparison of Multivariate Preprocessing Techniques for the Forensic Discrimination of Cotton Fibers by UV/visible Microspectrophotometry

Nathan C. Fuenffinger¹, John V. Goodpaster², Edward G. Bartick³, and Stephen L. Morgan^{1*}

¹Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208

²Department of Chemistry and Chemical Biology, Forensic And Investigative Sciences Program, Indiana University Purdue University Indianapolis (IUPUI), 402 North Blackford Street, LD 326, Indianapolis, IN 46202

³Department of Forensic Science, The George Washington University, 2100 Foxhall Rd. NW, Washington, DC 20007.

*author to whom correspondence should be addressed.

Keywords. cotton, fibres, discrimination, derivative, microspectrophotometry, multivariate statistics

Abstract. Color plays a critical role when analyzing natural fibers such as cotton in forensic investigations. Ultraviolet-visible (uv/vis) 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. We have found that derivative preprocessing can enhance structure in spectra. samples of reactive, direct, and vat dyed cotton fibers were analyzed and spectra were preprocessed using multiple methods including baseline correction, normalization, and first and second derivatives. Principal component analysis followed by linear discriminant analysis was employed to discriminate 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 classification accuracy when analyzing spectra with only subtle features such as those seen with 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.

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 [5] surfaces, as well as human head hair [6]. The abundance of cotton is so great that these fibers (especially black/grey and blue cotton) 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 with these types of 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 [9-12]. 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 [13].

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. By applying chemometric methods to UV/Vis spectra, complex data can be reduced down to the most significant variables which may not be readily visible by examining the spectra. 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 [14,15]. The use of differentiation has only been used sparingly in forensics, however, for analyzing textile fibers [16-19] and fiber dyes [20,21]. 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. Classification accuracies will be obtained by using principal component analysis (PCA) followed by 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 [22,23].

Materials and methods

2.1. Samples. Cotton fibers were collected from fabric obtained from a textile-related manufacturer in the southeastern United States. 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 based on their observed color for subsequent analysis by UV/visible MSP.

For analysis by UV/Vis MSP, single fibers were positioned on quartz slides (CRAIC Technologies, Altadena, CA, and Esco Products Inc., Oak Ridge, New Jersey) using micro tweezers. Each fiber was mounted using spectral grade glycerin (Spectrum Chemical Mfg. Corp., Gardena, CA) and quartz cover slips.

2.2. UV/Vis Microspectrophotometry. 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, New Hampshire). The microspectrophotometer was operated in transmission mode using a xenon light source. A 15x 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, a program written in MATLAB (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 380-700 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 simply 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 *i*th row and *j*th column by the sum of the absolute value of all elements in that row, also called 1-norm, as in Equation 1 [24].

$$X_{ij} = \frac{X_{ij}}{\sum_{k=1}^p |X_{ik}|} \quad (1)$$

The total area under the curve of each resulting vector is therefore equal to one. Normalization by 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} is the SNV transformed point [25].

$$X_{ij} = \frac{X_{ij} - \bar{X}_i}{\sigma_i} \quad (2)$$

The sample mean spectrum, \bar{X}_i used in SNV calculations is not used in 1-norm normalization, 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, σ_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, σ_j .

$$X_{ij} = \frac{X_{ij} - \bar{X}_j}{\sigma_j} \quad (3)$$

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. *First Derivative.* First derivative (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.

$$\frac{dA}{d\lambda} \approx \frac{A_{n+1} - A_n}{\lambda_{n+1} - \lambda_n} \quad (4)$$

The sharpest features in 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. FDs are an effective tool for correcting baseline offsets. By using normalization to unit area or SNV following a FD calculation, a slope correction can also be gained.

2.3.6. *Second Derivative.* One approach to dealing with the increasing noise amplification that is associated with calculating second and higher order 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.

2.3.7. *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 highest variation in the data set, and the variance decreases with each successive PC. Selection of the number of relevant PCs to be used in the models was chosen via Scree plots which indicate 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 number of PCs is selected, LDA, a supervised 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 linear discriminant axes (also called 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 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 to which the Mahalanobis distance is the shortest.

Results and discussion. 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 (\bar{A}) 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. For a curve which is Gaussian in shape, such as the ones encountered in UV/Vis MSP, this calculation can be used as an indicator of the evenness of the features across the spectra.

$$\bar{A} = \frac{A_{\max} + A_{\min}}{2} \quad (5)$$

Most direct dyed fibers in this study (80.8%) had values for \bar{A} between 0.005 and 0.020. The \bar{A} 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 \bar{A} values. However, 89.2% of these fibers had a \bar{A} 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 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 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 4, the broad peaks in the averaged raw absorbance spectra of the three remaining fibers demonstrate the challenge associated with fiber comparisons based on non-preprocessed spectra. It is worth noting 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 previously, 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 accurate classification percentages. When too many PCs are included, noise hinders the results gained. Overfitting is avoided or reduced by using a scree plot such as the one in Figure 5. This plot shows the percent variance captured by each PC after PCA was performed on FD spectra of the six yellow reactive dyed fibers. Four PCs, containing 91.7 percent 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 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 6 shows that even after FD preprocessing, there is still significant overlap of the 95 percent 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 effect that each preprocessing method has on UV/Vis absorbance spectra is perhaps best shown using samples RY03 and RY04 (Figure 7). These fibers gave very similar raw spectra, but were easily distinguishable after several types of preprocessing. A t-test can be used to measure the difference between the group means at a single 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 SD. This is not surprising as SD accounts for both baseline offsets and changes in slope. The PCA scores plots in Figure 4 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 8g, 8h, and 8i), SD (Figure 8j), and SNV preprocessing techniques (Figure 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 3. Methods involving FD and SD calculations show increased classification accuracies of direct dyed cotton fibers by as much as 11 percent. This gap in discrimination ability is lessened for vat dyed fibers, and is non-existent in the analysis of reactive dyed fibers. In the case of reactive dyed fibers, normalized spectra are slightly more discriminating 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 4. All FD and SD preprocessing methods used in this study, in addition to normalization, can be considered effective methods for discriminating cotton fibers.

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 [27] prior to calculating the first derivative 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 first derivative spectra misclassifying matching fibers with large variations in absorbance. This resulted in classification accuracies of first derivative 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 percent) 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, the analyst must show caution when selecting cases in which derivatives should be used.

Acknowledgement. This research was supported provided by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice Award No. 2010-DN-BX-K220. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

- [1] J.E.G. van Dam, Environmental benefits of natural fibre production and use, in: Proceedings of the Symposium on Natural Fibres, Rome, Italy, 2008, 3-17.
- [2] T.W. Biermann, M.C. Grieve, 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.* 77 (1996) 75-91.
- [3] S. Cantrell, C. Roux, P. Maynard, J. Robertson, A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region, *Forensic Sci. Int.* 123 (2001) 48-53.
- [4] R. Watt, C. Roux, J. Robertson, The population of coloured textile fibres in domestic washing machines, *Sci. Justice* 45 (2005) 75-83.
- [5] M.C. Grieve, T.W. Biermann, The population of coloured textile fibres on outdoor surfaces, *Sci. Justice* 37 (1997) 231-239.
- [6] R. Palmer, S. Oliver, The population of coloured fibres in human head hair, *Sci. Justice* 44 (2004) 83-88.
- [7] R.M. Christie, *Colour Chemistry*, Royal Society of Chemistry, Cambridge, 2001.
- [8] M.C. Grieve, T.W. Biermann, M. Davignon, The evidential value of black cotton fibres, *Sci. Justice* 41 (2001) 245-260.
- [9] K.J. Wiggins, Thin layer chromatographic analysis for fibre dyes, in: J. Robertson, M. Grieve (Eds.), *Forensic Examination of Fibres*, second ed., Taylor and Francis, Forensic Science Series, London, 1999, pp. 291-310.
- [10] F.P. Adolf, J. Dunlop, Microspectrophotometry/colour measurement, in: J. Robertson, M. Grieve (Eds.), *Forensic Examination of Fibres*, second ed., Taylor and Francis, Forensic Science Series, London, 1999, pp. 251-287.
- [11] J.V. Goodpaster, E.A. Liszewski, Forensic analysis of dyed textile fibers, *Anal. Bioanal. Chem.* 394 (2009) 2009-2018.
- [12] G. Massonnet, P. Buzzini, F. Monard, G. Jochem, L. Fido, S. Bell, M. Stauber, T. Coyle, C. Roux, J. Hemmings, H. Leijenhors, Z. Van Zanten, K. Wiggins, C. Smith, S. Chabli, T. Sauneuf, A. Rosengarten, C. Meile, S. Ketter, A. Blumer, Raman spectroscopy and microspectrophotometry of reactive dyes on cotton fibres: analysis and detection limits, *Forensic Sci. Int.* 222 (2012) 200-207.
- [13] M.M. Houck, Inter-comparison of unrelated fiber evidence, *Forensic Sci. Int.* 135 (2003) 146-149.
- [14] C. Bosch Ojeda, F. Sanchez Rojas, Recent applications in derivative ultraviolet/visible absorption spectrophotometry: 2009-2011 a review, *Microchem. J.* 106 (2013) 1-16.
- [15] J. Karpinska, Derivative spectrophotometry – recent applications and directions of developments, *Talanta* 64 (2004) 801-822.
- [16] T. Coyle, A. Larkin, K. Smith, S. Mayo, A. Chan, N. Hunt, Fibre mapping – a case study, *Sci. Justice* 44 (2004) 179-186.
- [17] M.C. Grieve, T.W. Biermann, K. Schaub, The individuality of fibres used to provide forensic evidence – not all blue polyesters are the same, *Sci. Justice* 45 (2005) 13-28.
- [18] M.C. Grieve, T.W. Biermann, K. Schaub, The use of indigo derivatives to dye denim material, *Sci. Justice* 46 (2006) 15-24.

- [19] K. Wiggins, R. Palmer, W. Hutchinson, P. Drummond, An investigation into the use of calculating the first derivative of absorbance spectra as a tool for forensic fibre analysis, *Sci. Justice* 47 (2007) 9-18.
- [20] V.C. Almeida, A.M.M. Vargas, J.C. Garcia, E. Lenzi, C.C. Oliveira, J. Nozaki, Simultaneous determination of the textile dyes in industrial effluents by first-order derivative spectrophotometry, *Anal. Sci.* 25 (2009) 487-492.
- [21] T.P. Bridge, R.H. Wardman, A.F. Fell, Novel digital methods for the qualitative characterization of some acid dyes applied to wool and nylon, *Analyst* 110 (1985) 1307-1312.
- [22] S.L. Morgan, A.A. Nieuwland, C. R. Mubarak, J. E. Hendrix, E. M. Enlow, B. J. Vasser, Forensic discrimination of dyed textile fibres using UV-VIS and fluorescence microspectrophotometry, in: *Proceedings of the 12th meeting of the European Fibres Group, Prague, Czech Republic, 2004.*
- [23] S.L. Morgan, S.H. Hall, J.E. Hendrix, E. Bartick, Pattern recognition methods for the classification of trace evidence textile fibers from UV/visible and fluorescence spectra, in: *National Institute of Justice Trace Evidence Symposium; Kansas City, Missouri, 2011.*
- [24] K. Beebe, R. Pell, M. Seasholtz, *Chemometrics: a practical guide*, John Wiley & Sons, New York, 1998.
- [25] R. Barnes, M. Dhanoa, S. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.* 43 (1989) 772-777.
- [26] P.J. Gemperline, Principal component analysis, in: P.J. Gemperline (Ed.), *Practical Guide to Chemometrics*, second ed., Taylor & Francis, Florida, 2006, pp. 69-104.
- [27] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least square procedures, *Anal. Chem.* 36 (1964) 1627-1639.

Table 1. Groups of studied fibers.

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 2. Confusion matrix for discrimination of yellow reactive dyed fibers.

Predicted	Actual					
	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 3. Performance of PC-LDA 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 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	Auto. ^a	Norm. ^b	SNV	BC ^c	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

^a Autoscale

^b Normalization to unit area

^c Baseline correction

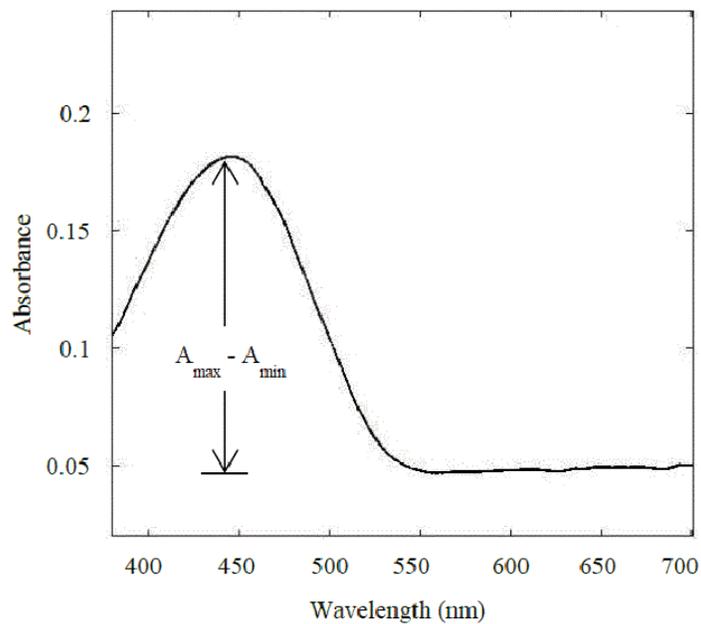


Figure 1. Gaussian shaped absorbance spectra and associated A_{\max} and A_{\min} locations.

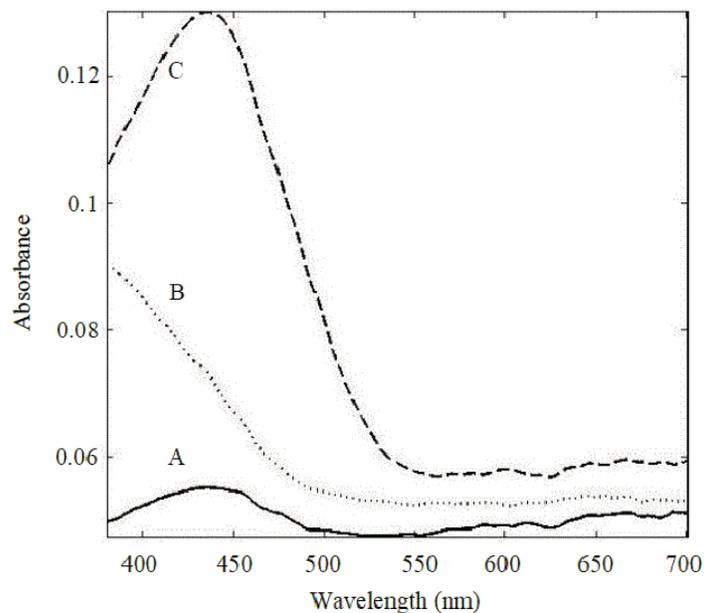


Figure 2. Absorbance spectra of three cotton fibers containing A) direct blue 86 and direct yellow 106, B) vat black 25, vat brown 81 and vat yellow 33, and C) reactive yellow 206.

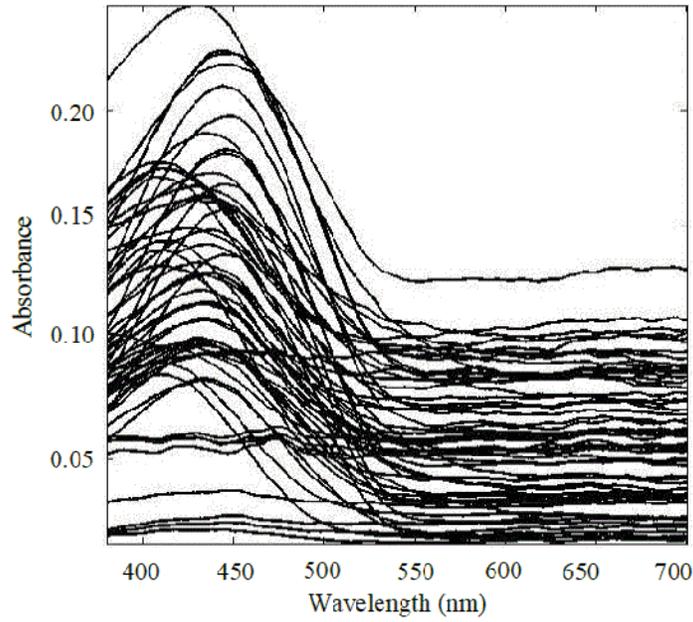


Figure 3. Absorbance spectra of six reactive dyed yellow cotton fibers (10 replicates each).

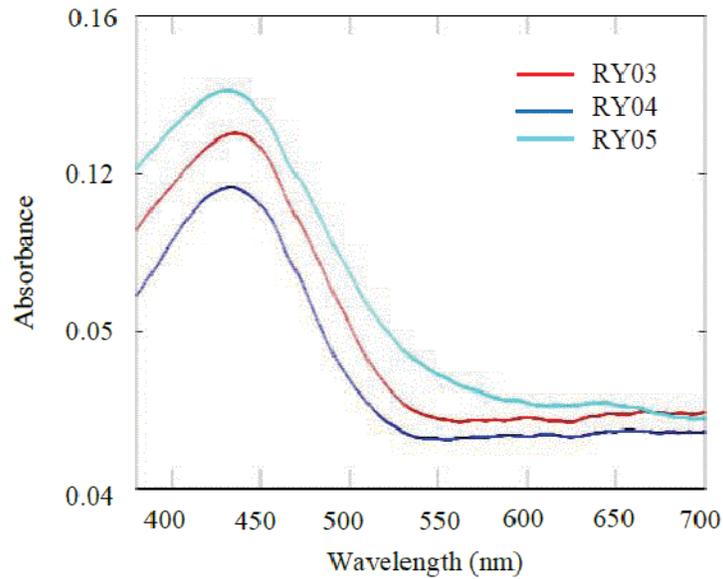


Figure 4. Averaged absorbance spectra for three of six yellow reactive dyed cotton fiber samples.

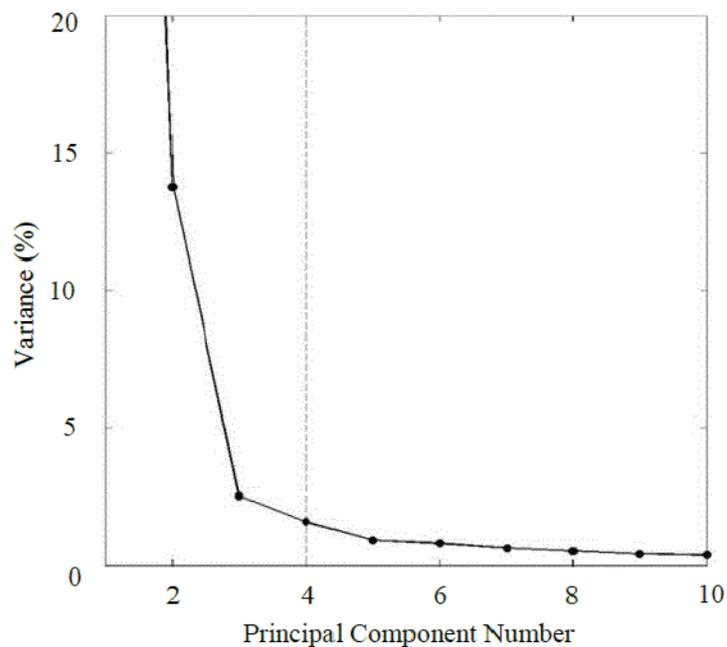


Figure 5. Scree plot obtained following PCA on first derivative spectra of six yellow reactive dyed cotton fibers.

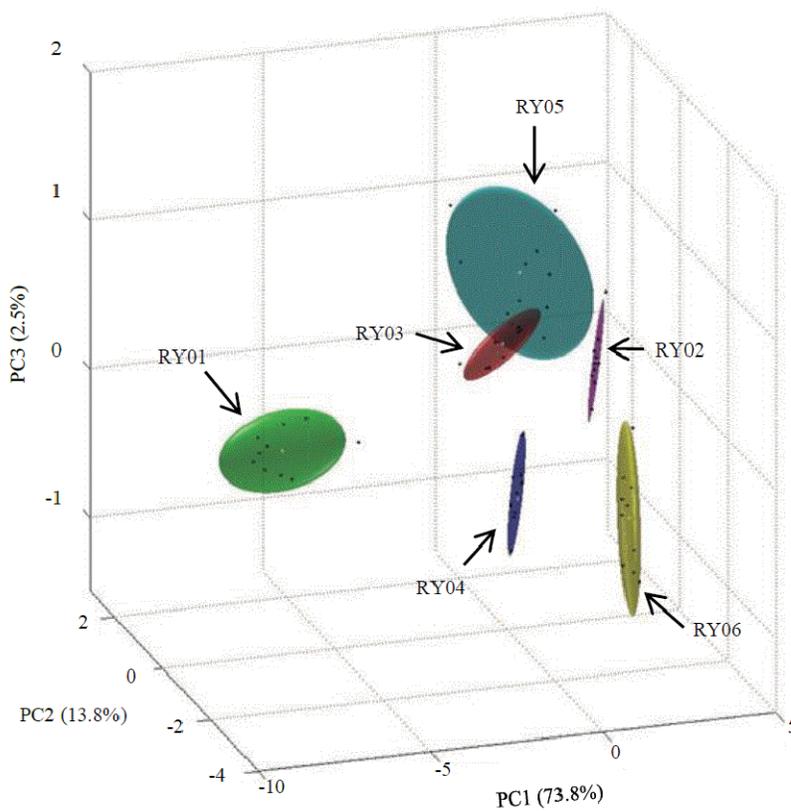
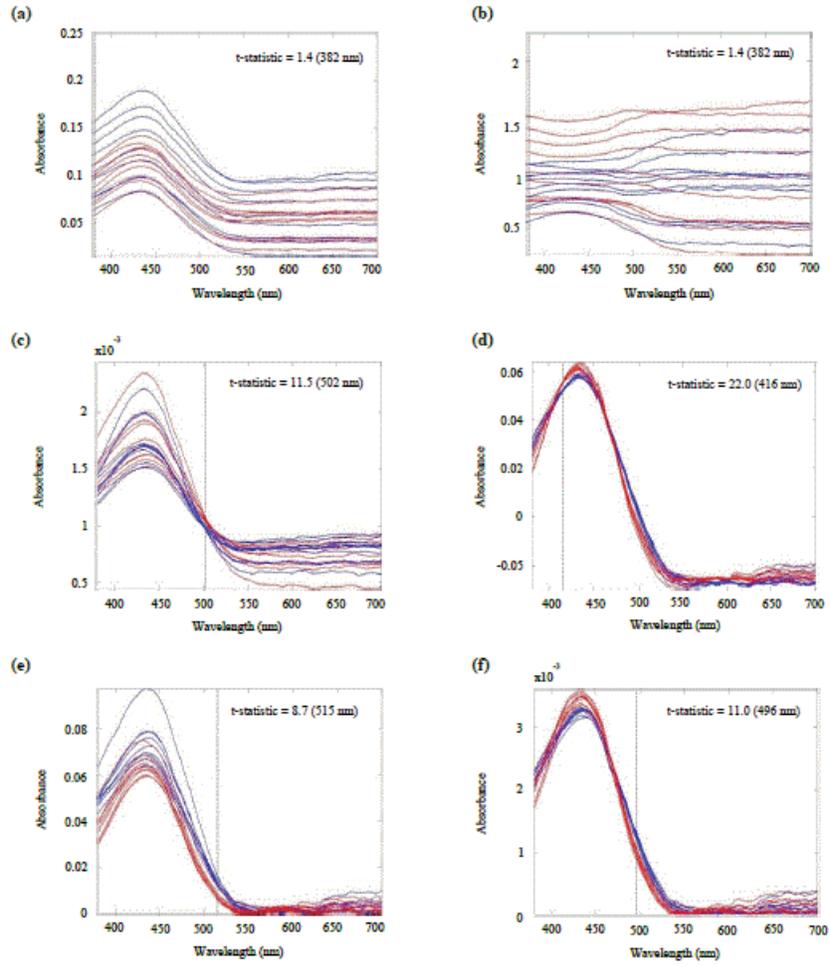


Figure 6. PCA scores plot for six reactive dyed yellow cotton fibers after first derivative preprocessing.



[Figure 7 continued on next page]

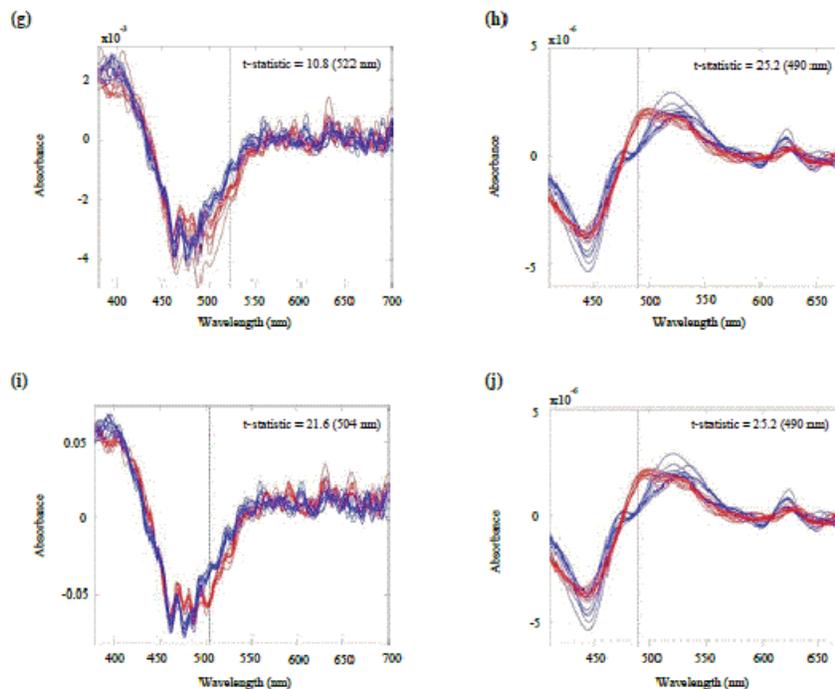
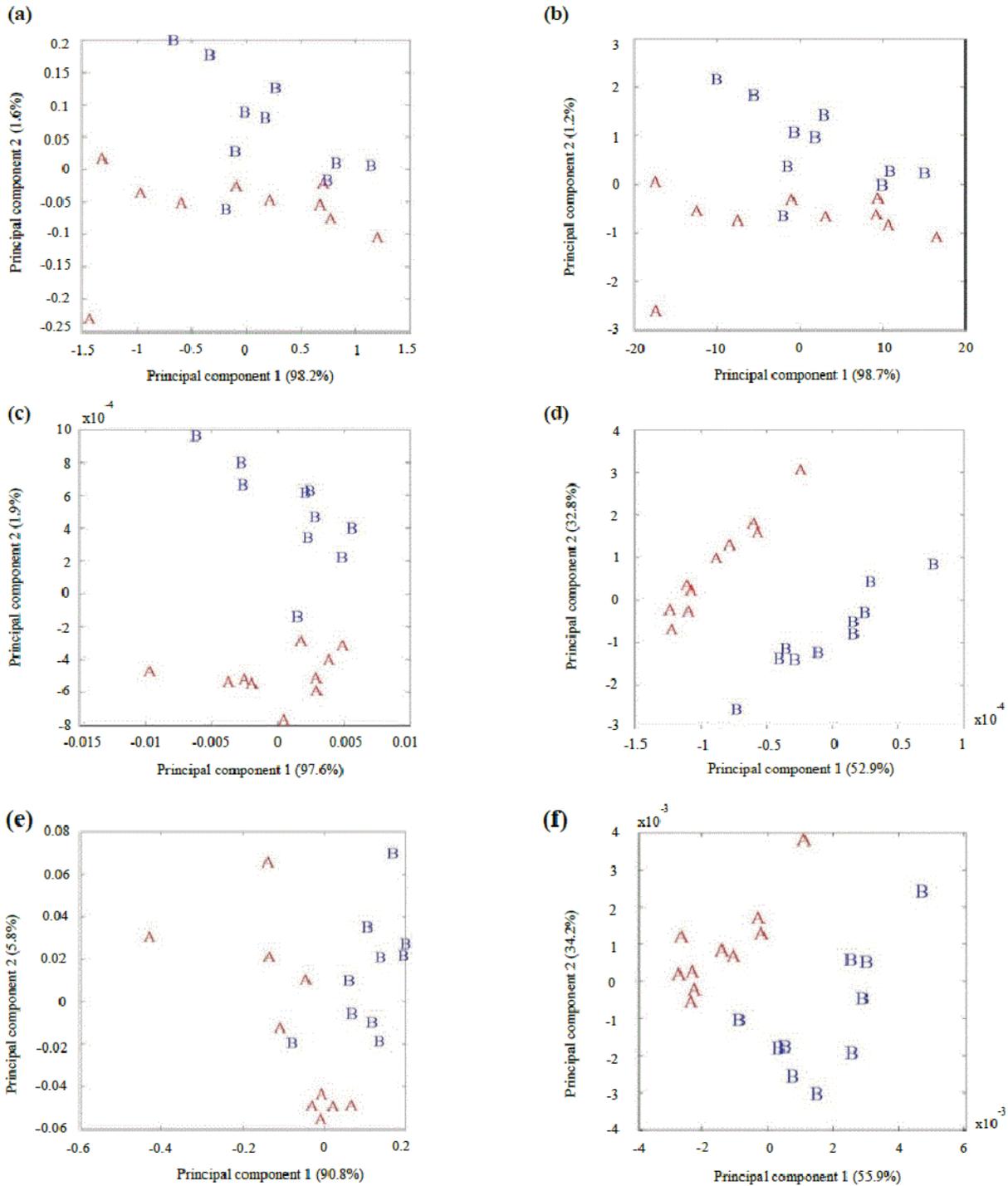


Figure 7. Absorbance spectra (10 replicates each) of samples RY03 and RY04 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.



[Figure 8 contined on next page]

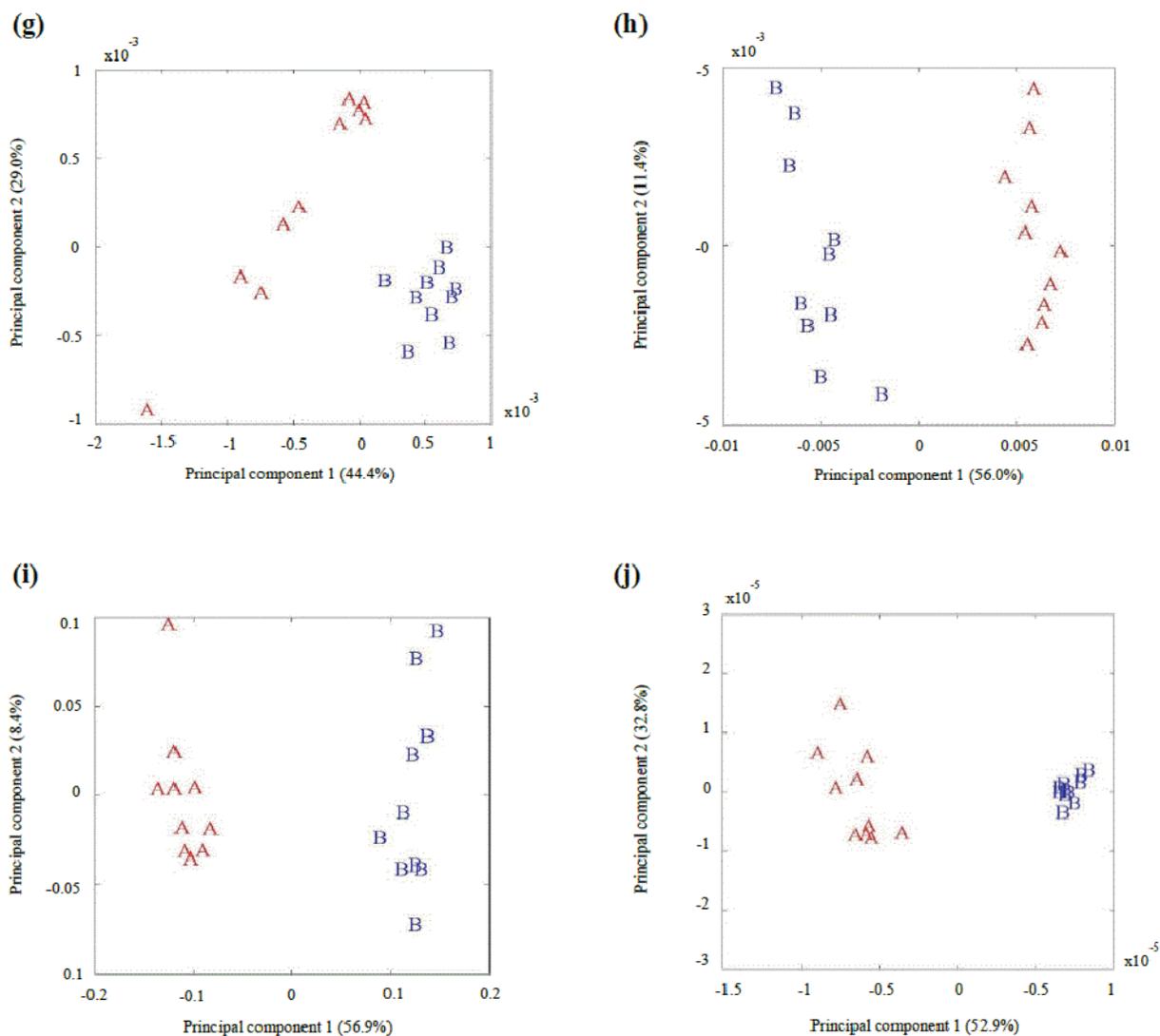


Figure 8. PCA scores plot resulting from absorbance spectra (10 replicates each) of samples RY03 and RY04 after a) no preprocessing, b) autoscaling, c) normalization, d) SNV, e) baseline correction, f) baseline correction plus normalization, g) first derivative, h) first derivative plus normalization, i) first derivative plus SNV, and j) second derivative.

D. Multivariate Discrimination of Dyed Textile Fibers from UV/Visible and Fluorescence Spectra

Nathan C. Fuenffinger¹, Jessica N. McCutcheon¹, John V. Goodpaster², Edward G. Bartick³, and Stephen L. Morgan¹

¹Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208

²Department of Chemistry and Chemical Biology, Forensic And Investigative Sciences Program, Indiana University Purdue University Indianapolis (IUPUI), 402 North Blackford Street, LD 326, Indianapolis, IN 46202

³Department of Forensic Science, The George Washington University, 2100 Foxhall Rd. NW, Washington, DC 20007.

Abstract. Importance has been placed on developing statistical methods for evaluation of trace evidence analysis in forensic science. Determining which analytical method will have the highest discrimination power for trace evidence examinations is significant to forensic laboratories to save time and resources. This study compares the discrimination ability of ultraviolet-visible (UV-VIS) microspectrophotometry (MSP) and fluorescence MSP, two common techniques used by forensic analysts to study fibers and fiber dyes.

Over 400 dyed textile samples of cotton, acrylic, nylon 6,6, and polyester were analyzed using UV-VIS MSP and fluorescence MSP at four wavelengths (365, 405, 436, and 546 nm). All spectra were preprocessed and classified using principal component analysis followed by linear discriminant analysis. Leave-one-out cross validation was used to test the ability of principal component-linear discriminant analysis (PC-LDA) to discriminate fibers in each color and polymer based group. The highest discrimination power was obtained by UV-VIS MSP. PC-LDA correctly classified 89.50% of the UV-VIS MSP spectra examined.

Introduction. Textile fibers are a significant form of trace evidence in forensics. Fibers are often transferred as a result of contact between people or people and objects. Cotton is the most abundant source of fibers in the world (1), and nylon, polyester, and acrylic fibers are a few of the most common classes of synthetic fibers likely to be encountered in forensic investigations (2). Optimized methods used to discriminate fibers such as these are necessary in order to say with confidence whether or not a particular fiber may be linked to person or crime scene.

Microscopic techniques such as light microscopy (3-6), polarized light microscopy (PLM) (4,7-9), and fluorescence microscopy (4,5,8) continue to play an important role in forensic fiber examinations. Alternatively, spectroscopic techniques can provide a more objective method of evaluating the chemical composition of textile fibers. Fourier transform infrared (FTIR) spectroscopy is used in fiber analysis to determine the class of polymer (acrylic, nylon, polyester, etc.) (2,5,8,10-19). Visible microspectrophotometry (MSP) (4,5,9,20-25), ultraviolet-visible (UV-VIS) MSP (3,6,26-29), fluorescence MSP (30), Raman spectroscopy (12,31-34), and thin

*author to whom correspondence should be addressed.

Abbreviations:

FD – First derivative

SD – Second derivative

layer chromatography (3,5,6,23,26) are all common techniques used to discriminate fibers based on color or dye composition. MSP techniques are particularly advantageous as they are non-destructive, require little sample preparation, and repeatable. UV-VIS MSP, in particular, has shown the ability to exclude fibers which are indistinguishable by comparison light microscopy and fluorescent light microscopy (4).

Evaluation of spectroscopic data can be carried out by the use of multivariate statistical techniques. These methods are designed to recognize patterns in complex sets of data and assign objects to classes (35). Multivariate classification methods have been used in forensics for spectroscopic analyses of explosives (36-37), pharmaceuticals (38), coins (39), inks (40), toners (41), paints (42), paper (43) and fibers (19,44,45). One such classification method involves principal component analysis (PCA) followed by linear discriminant analysis (LDA). In PCA-LDA, the original number of variables is reduced by PCA, and the new smaller set of variables is used in LDA to define and predict classes (46). In this paper, methods of evaluating common textile fibers by UV-VIS and fluorescence MSP using PCA-LDA are presented.

Materials and methods

Samples. Textile samples of commercially dyed cotton, nylon 6,6, polyester, and acrylic were obtained from commercial sources. The fibers were placed into 11 different groups based on their observed color with blue, brown, and green being the largest groups. Single fibers were removed from the fabric by using micro tweezers and razor blades. The individual fibers were then centered on glass microscope slides. Spectral grade Permout (Fisher Scientific, Fair Lawn, NJ) and glass cover slips were used to mount the fiber on the microscope slides.

UV/VIS Microspectrophotometry. UV-VIS spectra were obtained using a Quantum Detection Instrument (QDI) 1000 microspectrophotometer (CRAIC Technologies, San Dimas, CA). Data was acquired using GRAMS/AI version 700 software (Thermo Galactic, Salem, New Hampshire). The MSP was operated in transmission mode using a xenon source. A 15x 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.

Data Analysis. Data was analyzed using a program, Spectral Explorer (SPX), written in MATLAB (The Mathworks, Inc., Natick, MA). After spectral collection, ranges of all absorbance spectra were truncated to the region of 380-700 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. Continued preprocessing of absorbance spectra was carried out by calculating the first derivative. Noise reduction was accomplished for absorbance using a Savitzky-Golay numerical algorithm (47) with a second order polynomial and nine point moving window. Fluorescence spectra were preprocessed using a linear smoothing algorithm with a window width of 31 points.

Groups of spectra were then subjected to PCA. PCA is a technique used to reduce the dimensionality of large data sets. In PCA, the original correlated variables (wavelengths) are reduced into a new set of uncorrelated variables, or principal components (PCs). These PCs are linear combinations of the original variables and are arranged in such a way that the first PC accounts for the highest variation in the data set, and the variance decreases with each successive

PC (48). Selection of the number of relevant PCs to be used in the models was chosen visually by examining the percent variance captured by each PC. The highest number of PCs before the captured variance begins to level off was selected for use in LDA.

After the PCs are selected, LDA, a supervised 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 linear discriminant axes (also called 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 LDA model were determined 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 to which the Mahalanobis distance is the shortest.

Results and Discussion. 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-LDA models of both the raw spectra and first derivative spectra. The resulting confusion matrices displayed in Table 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-LDA model to discriminate yellow acrylic samples two and four is shown visually using the scores plot in Figure 1. 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 LDA scores plot produced by first derivative data shows a clear separation between these two samples.

Figure 2 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 and fluorescence MSP for all fibers in this study are listed in Table 2. Entries marked with "x" indicate groups of fibers in which LDA 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.

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.

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 LDA model. The PCA scores plot of the preprocessed absorbance spectra for all brown polyester fibers in this study is shown Figure 3. 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 extent to which within-group variation took place 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 3). Replicate absorbance spectra of the samples in Cluster A were averaged and are shown in Figure 4. 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-LDA. All other brown polyester samples not previously mentioned (29 in all) are located in the second cluster (Cluster B in Figure 3). 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 MSP spectra collected at 405, 436, and 545 nm 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 is not recommended for fiber analysis. As with UV-VIS MSP, discrimination was highest for fluorescence MSP in groups of orange and purple fibers with each group having classification accuracies of over 90% at all wavelengths of excitation. Discrimination at all fluorescence wavelengths is lowest for nylon fibers. The issues with cotton fibers mentioned previously with absorbance spectra are apparent in the discrimination power of fluorescence MSP as well.

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 UV-VIS and fluorescence MSP provided discriminating information, UV-VIS MSP collectively outperformed the fluorescence methods used. With the aid of first derivative preprocessing and PC-LDA, a classification of nearly 90% was achieved for 482 fibers studied using UV-VIS MSP.

Acknowledgement. This research was supported provided by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice Award No. 2010-DN-BX-K220. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

- [28] van Dam, J. E. G. "Environmental benefits of natural fibre production and use," Proceedings of the Symposium on Natural Fibres, Rome, Italy, 2008; pp 3-17.
- [29] Tungol, M. W.; Bartick, E. G.; Montaser, A. "Analysis of single polymer fibers by Fourier transform infrared microscopy: The results of case studies," *J. Forensic Sci.* **1991**, 36, 1027-1043.
- [30] Rendle, D. F.; Wiggins, K. G. "Forensic analysis of textile fibre dyes," *Rev. Prog. Coloration.* **1995**, 25, 29-34.
- [31] Houck, M. "Inter-comparison of unrelated fiber evidence," *For. Sci. Int.* **2003**, 135, 146-149.
- [32] Palmer, R.; Chinherende, V. "A target fiber study using cinema and car seats as recipient items," *J. Forensic Sci.* **1996**, 41, 802-803.
- [33] Wiggins, K. G.; Holness, J. A.; March, B. M. "The importance of thin layer chromatography and UV microspectrophotometry in the analysis of reactive dyes released from wool and cotton fibers," *J. Forensic Sci.* **2005**, 50, 364-368.
- [34] Stoeffler, S. F. "A flowchart system for the identification of common synthetic fibers by polarized light microscopy," *J. Forensic Sci.* **1996**, 41, 297-299.
- [35] 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.
- [36] Coyle, T.; Larkin, A.; Smith, K.; Mayo, S.; Chan, A.; Hunt, N. "Fibre mapping – a case study," *Sci. Justice.* **2004**, 44, 179-186.
- [37] Pandey, G. C. "Fourier-transform infrared microscopy for the determination of the composition of copolymer fibers – acrylic fibers," *Analyst.* **1989**, 114, 231-232.
- [38] White, G. W. "A simple high-pressure anvil and template device for the production of infrared spectra from microfiber samples," *J. Forensic Sci.* **1992**, 37, 620-631.
- [39] Lang, P. L.; Katon, J. E.; O'Keefe, J. F.; Schiering, D. W. "The identification of fibers by infrared and Raman microspectroscopy," *Microchem. J.* **1986**, 34, 319-331.
- [40] Grieve, M. C. "Another look at the classification of acrylic fibres using FTIR microscopy," *Sci. Justice.* **1995**, 35, 179-190.
- [41] Kirkbride, K. P.; Tungol, M. W. "Infrared microspectroscopy of fibres." In: Robertson, J.; Grieve, M. (Ed). *Forensic Examination of Fibres*, 2nd ed.; Taylor and Francis, London, 1999.
- [42] Tungol, M. W.; Bartick, E. G.; Montaser, A. "The development of a spectral data base for the identification of fibers by infrared microscopy," *Appl. Spectrosc.* **1990**, 44, 543-549.
- [43] Tungol, M. W.; Bartick, E. G.; Montaser, A. "Forensic analysis of acrylic copolymer fibers by infrared microscopy," *Appl. Spectrosc.* **1993**, 47, 1655-1658.
- [44] Flynn, K.; O'Leary, R.; Roux, C.; Reedy, B. "Forensic analysis of bicomponent fibers using infrared chemical imaging," *J. Forensic Sci.* **2006**, 51, 586-596.
- [45] Causin, V.; Marega, C.; Guzzini, G.; Marigo, A. "The effect of exposure to the elements on the forensic characterization by infrared spectroscopy of poly(ethylene terephthalate) fibers," *J. Forensic Sci.* **2005**, 50, 887-893.
- [46] Enlow, E. M.; Kennedy, J. L.; Nieuwland, A. A.; Hendrix, J. E.; Morgan, S. L. "Discrimination of nylon polymers using attenuated total reflection mid-infrared spectra and multivariate statistical techniques," *Appl. Spectrosc.* **2005**, 59, 986-992.

- [47] Wiggins, K.; Palmer, R.; Hutchinson, W.; Drummond, W. "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.
- [48] Grieve, M. C.; Biermann, T. W.; Schaub, K. "The use of indigo derivatives to dye denim material," *Sci. Justice*. **2006**, 46, 15-24.
- [49] Suzuki, S.; Suzuki, Y.; Ohta, H.; Sugita, R.; Marumo, Y. "Microspectrophotometric discrimination of single fibres dyed by indigo and its derivatives using ultraviolet-visible transmittance spectra," *Sci. Justice*. **2001**, 41, 107-111.
- [50] Kelly, E.; Griffin, R. M. E. "A target fibre study on seats in public houses," *Sci. Justice*. **1998**, 38, 39-44.
- [51] Palmer, R.; Hutchinson, W.; Fryer, V. "The discrimination of (non-denim) blue cotton," *Sci. Justice*. **2009**, 49, 12-18.
- [52] Palmer, R.; Turnbull, L. D. "A survey of dye batch variation," *Sci. Justice*. **1995**, 35, 59-64.
- [53] Biermann, T. W. "Blocks of colour IV: The evidential value of blue and red cotton fibres," *Sci. Justice*. **2007**, 47, 68-87.
- [54] Grieve, M. C.; Biermann, T.; Davignon, M. "The occurrence and individuality of orange and green cotton fibres," *Sci. Justice*. **2003**, 43, 5-22.
- [55] Grieve, M. C., Biermann, T. W. "The evidential value of black cotton fibres," *Sci. Justice*. **2001**, 41, 245-260.
- [56] Grieve, MC, Biermann, TW, Schaub, K. The individuality of fibres used to provide forensic evidence – not all blue polyesters are the same. *Sci Justice* **2005**; 45:13-28.
- [57] Adolf P, Dunlop J. Microspectrophotometry/colour measurement. In: Robertson, J.; Grieve, M. (Ed). *Forensic Examination of Fibres*, 2nd ed.; Taylor and Francis, London, 1999.
- [58] 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.
- [59] Miller, J. V.; Bartick, E. G. "Forensic analysis of single fibers by Raman spectroscopy," *Appl. Spectrosc.* **2001**, 55, 1729-1732.
- [60] Keen, I. P.; White, G. W.; Fredericks, P. M. "Characterization of fibers by Raman microprobe spectroscopy," *J. Forensic Sci.* **1998**, 43, 82-89.
- [61] Edwards, H. G. M.; Farwell, D. W.; Webster, D. "FT Raman microscopy of untreated natural plant fibres," *Spectrochim. Acta A.* **1999**, 53, 2383-2392.
- [62] Marini, F. "Classification methods in chemometrics," *Curr. Anal. Chem.* **2010**, 6, 72-79.
- [63] Hwang, J.; Choi, N.; Park, A.; Park, J. Q.; Chung, J. H.; Baek, S.; Cho, S. G.; Baek S. J.; Choo, J. "Fast and sensitive recognition of various explosive compounds using Raman spectroscopy and principal component analysis," *J. Mol. Struct.* **2013**, 1039, 130-136.
- [64] Banas, K.; Banas, A.; Moser, H. O.; Bahou, M.; Li, W.; Yang, P.; Cholewa, M.; Lim, S. K. "Multivariate analysis techniques in the forensics investigation of the postblast residues by means of Fourier transform-infrared spectroscopy," *Anal. Chem.* **2010**, 82, 3038-3044.

- [65] Anzanello, M. J.; Ortiz, R. S.; Limbergerb, R. P.; Mayorga, P. "A multivariate-based wavenumber selection method for classifying medicines into authentic or counterfeit classes," *J. Pharm. Biomed. Anal.* **2013**, 83, 209-214.
- [66] Hida, M.; Sato, H.; Sugawara, H.; Mitsui, T. "Classification of counterfeit coins using multivariate analysis with X-ray diffraction and X-ray fluorescence methods," *For. Sci. Int.* **2001**, 115, 129-134.
- [67] 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**, 40, 270-277.
- [68] Egan, W. J.; Morgan, S. L.; Bartick, E. G.; Merrill, R. A., Taylor, H. J. "Forensic discrimination of photocopy and printer toners. II. Discriminant analysis applied to infrared reflection-absorption spectroscopy," *Anal. Bioanal. Chem.* **2003**, 376, 1279-1285.
- [69] Muehlethaler, C.; Massonnet, G.; Esseiva, P. "The application of chemometrics on infrared and Raman spectra as a tool for the forensic analysis of paints," *For. Sci. Int.* **2011**, 209, 173-182.
- [70] Kher, A.; Mulholland, M.; Reedy, B.; Maynard, P. "Classification of document papers by infrared spectroscopy and multivariate statistical techniques," *Appl. Spectrosc.* **2001**, 55, 1192-1198.
- [71] Yu, M. M. L.; Sanderock, P. M. L. "Principal component analysis and analysis of variance on the effects of Entellan New on the Raman spectra of fibers," *J. Forensic. Sci.* **2012**, 57, 70-74.
- [72] Morgan, S. L.; Bartick, E. G. "Discrimination of forensic analytical chemical data using multivariate statistics," In: Blackledge, R. D. (Ed). *Forensic Analysis on the Cutting Edge*; John Wiley & Sons, New Jersey, 2007.
- [73] Mendlein, A.; Szkudlarek, C.; Goodpaster, J. V. "Chemometrics," In: Siegel JA, Saukko PJ, editors. *Encyclopedia of Forensic Sciences*, 2nd ed.; Elsevier, 2013.
- [74] Savitzky, A.; Golay, M. "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.* **1964**, 36, 1627-1639.
- [75] Gemperline, P. J. "Principal Component Analysis," In: Gemperline, P. J. (Ed). *Practical Guide to Chemometrics*, 2nd ed.; Taylor & Francis: Florida, 2006.

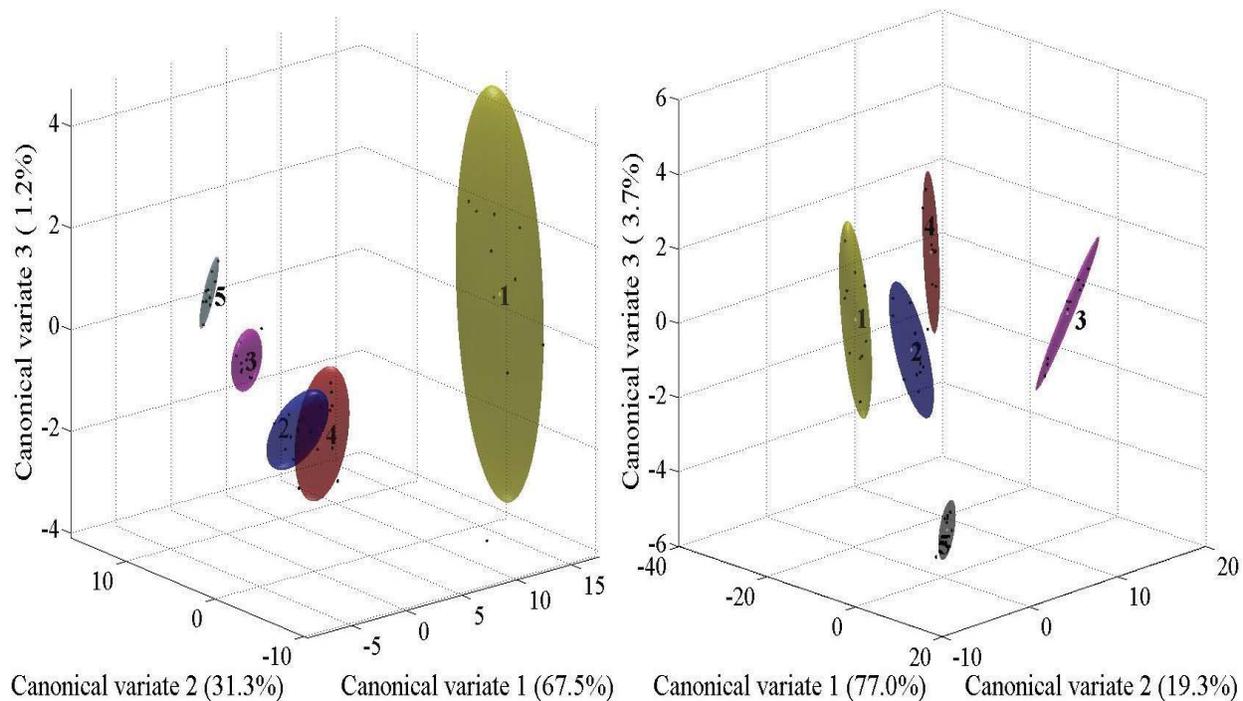


Figure 1. LDA scores plot resulting from raw (left) and first derivative (right) spectra of five yellow acrylic fibers.

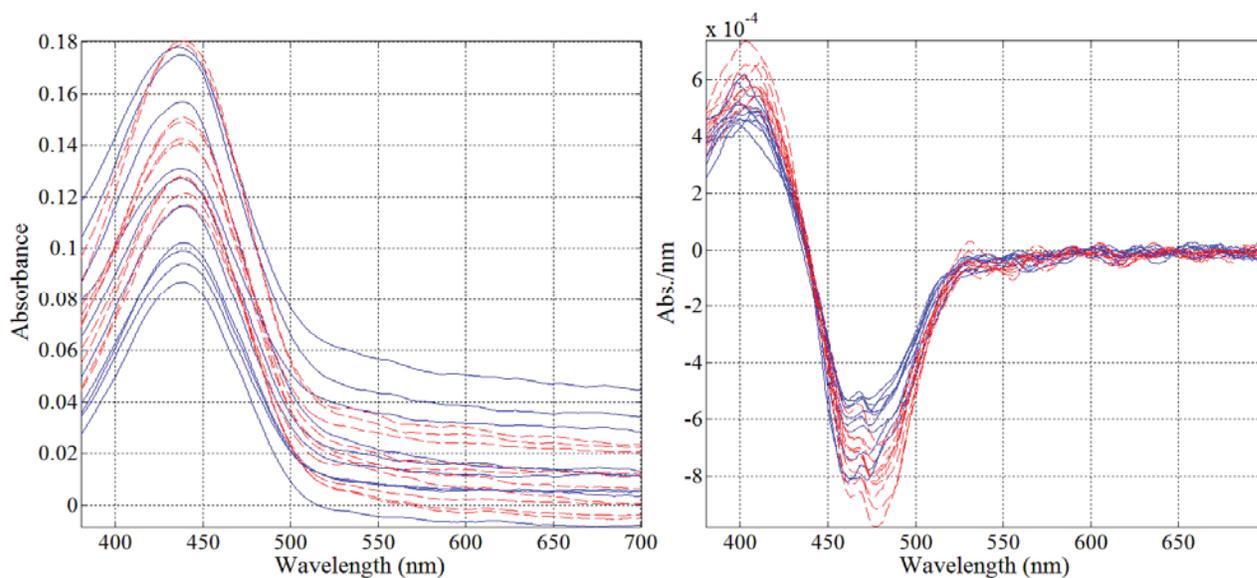


Figure 2. Raw (left) and first derivative (right) UV-VIS absorbance spectra for two yellow acrylic samples (10 replicate spectra for each sample).

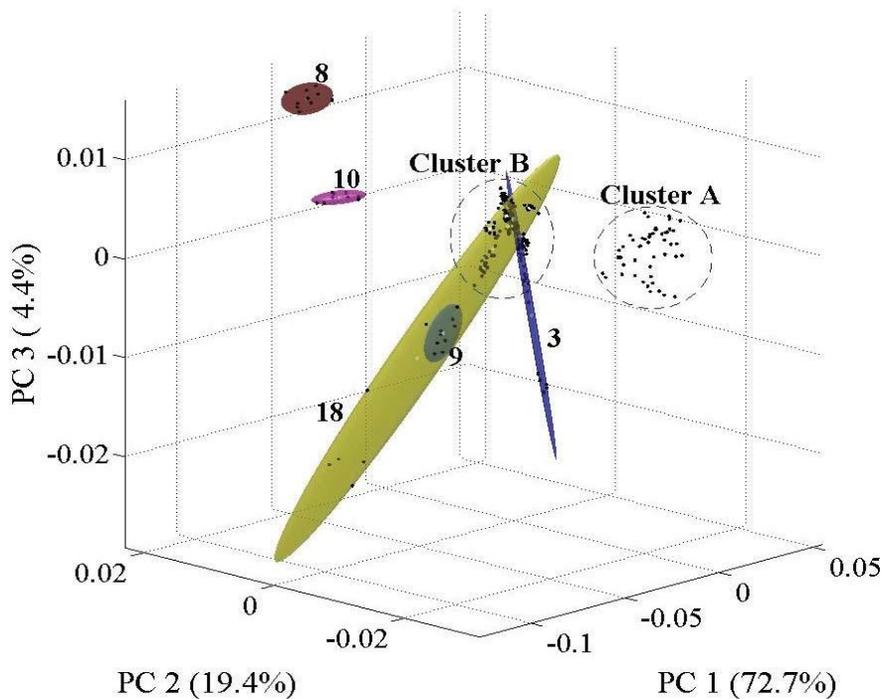


Figure 3. PCA scores plot of 39 brown polyester fibers with elliptical confidence regions around clustered samples removed.

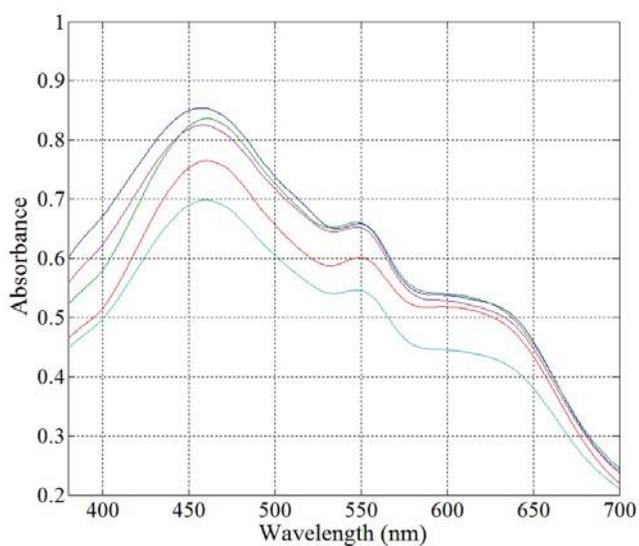


Figure 4. Averaged UV-VIS absorbance spectra for five brown polyester samples in Cluster B.

Table 1. Confusion matrices displaying results of cross-validation for five yellow acrylic samples (10 replicate spectra each).

Data	Group	<u>Observed classifications</u>					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. Classification accuracy for all fiber types and colors.

<i>Color</i>	<i>Fiber Type</i>	<i>Groups</i>	<i>Absorbance</i>	<i>FE 365</i>	<i>FE 405</i>	<i>FE 436</i>	<i>FE 546</i>
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
<i>Total Spectra</i>		4820	4314	3986	4229	4249	4220
<i>%Classification Accuracy</i>			89.50	82.70	87.74	88.15	87.55

OVE

E. Model Transfer for Multivariate Discrimination of Textile Fibers by UV/Visible Microspectrophotometry

Nathan C. Fuenffinger,¹ Eric J. Reichardt,² Edward G. Bartick,³ John G. Goodpaster,² and Stephen L. Morgan¹

¹Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208

²Department of Forensic Sciences, The George Washington University, Washington, DC 20007

³Indiana University Purdue University Indianapolis (IUPUI), Department of Chemistry and Chemical Biology, Forensic and Investigative Sciences Program, Indianapolis, IN 46202

Abstract. The ability to transfer multivariate classification models between laboratories can save time and resources in forensic analyses. 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, ultraviolet (UV)-visible absorbance spectra of 12 blue acrylic fibers were examined. The agreement of results among three separate laboratories was evaluated by testing the transferability of multivariate classification models based on principal component analysis-linear discriminant analysis (PCA-LDA) and partial least squares-discriminant analysis (PLS-DA). An average classification accuracy of 88.06% was found after training the PLS-DA models using data collected at two laboratories and using the information collected at the third laboratory as an external test set. For comparison, intra-laboratory studies carried out using PCA-LDA produced an average classification accuracy of 98.33%. Researchers are advised to use caution and to follow the traditional adage of "Trust, but verify."

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 principal component analysis-linear discriminant analysis (PCA-LDA) and partial least squares-discriminant analysis (PLS-DA). While PCA captures the variation between each sample, the technique does not attempt to determine the directions in the data that allow the classes to be discriminated. As a result, discrimination using PCA can be difficult when the within-class variance meets or exceeds the between-class variance. The aim of PLS-DA is to alleviate this problem by selecting out the variation in the spectra which is most useful for the

Corresponding author.

E-mail address: morgansl@mailbox.edu

Telephone: +1 803 777 2461

discrimination of classes. Though essentially the same results can be achieved using PCA-LDA and PLS-DA, the latter includes the advantages of noise reduction and variable selection.⁵ Previously, PCA-LDA and PLS-DA methodologies have been combined with optical spectroscopy techniques to study various forensic analytes of interest such as gunshot residue,⁶ explosives,^{7,8} oil,⁹⁻¹¹ soil,¹²⁻¹⁴ and fibers.^{4,15,16}

Ideally, classification models built using PCA-LDA and PLS-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 a new sample is introduced. Instrumental transfer of classification models, however, can be challenging due to differences in sample preparation, instrumental response, and environmental conditions experienced between laboratories.

This research has three objectives: (a) to conduct interlaboratory experiments for the evaluation of decision making in forensic fiber examinations by microspectrophotometry; (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/visible spectra of fibers. To address these objectives, the agreement between classification accuracies among three laboratories and the transfer of multivariate classification models between the laboratories was evaluated using PCA-LDA and PLS-DA for discrimination of spectra taken from a set of twelve blue acrylic fibers.

Materials and Methods

Fiber samples. Acrylic samples were donated from commercial sources in the southeastern United States. Procedures from the Scientific Working Group on Materials Analysis Fiber Subgroup (SWGMAF) were followed.²³ A number of fibers and locations along the fibers were analyzed to assess real and apparent variations in dyeing depth at different locations along the fiber. In the three collaborating laboratories, individual fibers were cut using a razor blade and positioned on glass microscope slides using micro-tweezers. Ten fibers from each exemplar were removed and mounted on glass microscope slides with a coverslip using spectral grade Permount[™] mounting media (Fischer Scientific, Fairlawn, NJ).

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

Laboratory 1 acquired spectra using a Craic QDI 2000 microspectrophotometer (Craic Technologies, San Dimas, CA) in transmitted light mode at a magnification of 150 \times . 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 350-800 nm.

Laboratory 2 obtained spectra with a Craic Quantum Detection Instrument (QDI) 302 UV/visible microspectrophotometer with a Carl Zeiss (Germany) Axioscope A1 and a 1.3 megapixel digital imaging system with CCD cooling. The spectral range was 400-800 nm with a resolution of 0.5 nm.

Spectra in Laboratory 3 were taken using a Craic Quantum Detection Instrument (QDI) 1000 microspectrophotometer operated in transmission mode with a xenon light source, a Carl Zeiss (Thornwood, NY) 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/visible spectra of textile fibers were produced by collecting an average of 100 scans over the spectral range of 240-850 nm at a 10 nm bandwidth and 4 ms integration time.

Data Analysis

Examining the discriminative ability of each individual laboratory using PCA-LDA.

Preprocessing of the data obtained from the three laboratories was performed using MATLAB version 8.1 (The MathWorks, Natick, MA, USA). The first derivative of each sample spectrum was calculated using a Savitzky-Golay algorithm^{17,18} to fit a 3rd order polynomial to a 15 point moving window. All spectra were then subjected to a standard normal variate (SNV) transformation.¹⁹ The purpose of SNV is to remove the differences in the slope of the spectra which may result from scattering. The transformation is applied to each individual spectrum using the following equation:

$$\hat{a}_{ij} = \frac{a_{ij} - \bar{a}_i}{\sigma_i} \quad (1)$$

where \hat{a}_{ij} is the SNV corrected absorbance at wavenumber j in spectrum i , a_{ij} is the original absorbance for the same element, \bar{a}_i is the mean of spectrum i , and n is the number of variables (wavenumbers). Finally, the data was mean-centered, a preprocessing technique typically recommended when performing PCA or PLS. 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.²⁰ 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 scores from the retained PCs were then projected into linear discriminant axes which seek to maximize the between- to within-group variance based on groups specified by the user.^{21,22}

The predictive performance of the PCA-LDA models created in this study was determined by internal validation using the leave-one-out method. For leave-one-out cross-validation, discriminant functions were calculated using information from all but one of the samples. The left-out sample was then assigned to the group for which its Mahalanobis distance was the shortest. This process was then repeated until each of the samples has been used for validation.

Discrimination using PLS-DA on combined data sets. 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.²³ Feature values in all three data sets were interpolated in the region of 400 to 800 nm. This region was selected because it was common to the data from each laboratory. A wavelength interval of 0.3407 nm was the lowest interval used for data collection at the three laboratories and was chosen as the new wavelength spacing. The MATLAB code shown in Table 1 was used to generate the interpolated data.

Table 1. MATLAB code for interpolation of wavelength features from data sets with incompatible wavelengths to a common wavelength axis.

```
x = features;           % original wavelength axis
y = transpose(X);      % transpose original n×p data matrix 'X'
xi = (400:0.3407:700); % define new axis and wavelength interval
yi = interp1(x,y,xi,'spline'); % interpolation using spline method
```

Following interpolation of the feature axis for each of the datasets from the three laboratories, training (containing 60% of the samples) and test sets (containing 40% of the samples) were generated. The training and test sets were developed assuring that each laboratory contributed the same number of samples in each class to the training and test sets. Thus, all samples (fiber spectra) in the test set also had replicate samples (spectra) in the training set collected from that same laboratory.

Data pretreatment and PLS-DA analyses of combined datasets were performed using PLS_Toolbox version 7.0.3 (Eigenvector Research, Wenatchee, WA, USA). All spectra in the training and test sets were again subjected to first derivative and SNV transformations as described previously. When dealing with combined datasets, a class centroid centering and scaling (CCCS) algorithm was used following first derivative and SNV. In CCCS, the class centroid (the mean of the class means) is removed from each variable within the class and scaled by the pooled standard deviation of the class.

After the data was pretreated, PLS-DA models were built using the samples in the training set. PLS-DA is a supervised classification technique based on algorithms for partial least squares (PLS) regression.⁵ In PLS-DA, the dimensionality of the data is reduced according to:

$$X = TP^T + E \quad (2)$$

$$Y = UQ^T + F \quad (3)$$

where $X(n \times p)$ is a matrix containing the instrumental responses of n samples at p different variables, Y is a matrix or vector containing the group memberships of all n samples, T and U are the scores matrices, and P^T and Q^T are the orthogonal loadings matrices. Finally, contained in matrices E and F is the information in X and Y , respectively, which is not explained by the scores and loadings matrices.

The factors, or 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. Selection of LVs used was based on the error resulting from leave-one-out cross validation of all samples included in the calibration dataset.

Using data from two laboratories to discriminate samples at a third using PLS-DA. 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 two of the three laboratories. The interpolated spectra from the remaining laboratory

were used as an external test set. Following preprocessing, three separate PLS-DA models were built, allowing datasets from each of the laboratories to be used as the external test set.

Results and Discussion

UV/visible spectra. Twelve blue acrylic fiber samples were analyzed at three separate laboratories using UV-Vis MSP. The resulting spectra obtained at each of these laboratories are shown in Appendix A. 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, 145 each 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 sample 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 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 sample fiber 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.

Examining the discriminative ability of each individual laboratory using PCA-LDA.

Dimensionality reduction of the spectral data is a necessary step for performing discriminant analysis. To examine the discriminative ability of each laboratory separately, PCA was used prior to LDA. Through the use of scree plots (not shown), it was determined that in each instance four PCs would be sufficient to build the LDA models. The four PCs represent 93.89%, 80.51%, and 89.78% of the total variation in the spectra collected by laboratories one, two, and three, respectively. The PC scores were then projected into the space of the linear discriminant axes (also called canonical variates) which account for the within-group and between-group variances. Because most of the discriminating information is often contained within the first few canonical variates, a three dimensional plot such as the one shown in Figure 1, resulting from the PCA-LDA analysis of laboratory one, can be used to view the separation between the 12 blue acrylic fibers. The ellipses around groups of spectra represent, with 95% confidence, distances that are statistically equidistant from the group mean. The size of the ellipse is related to the degree of within-class spectral variations. For example, as was mentioned previously, sample fiber 092 gave noisy, low-absorbing spectra. The amount of intra-sample variation between spectra resulted in a characteristically large confidence ellipse.

The proximity of the groups of similar spectra mentioned in the previous section is also evident in Figure 1. Of particular significance are the relatively short Mahalanobis distances between samples 086 and 112. As seen in Table 2, these two samples could be distinguished completely through the use of PCA-LDA on spectral data from laboratory 1. The significant spectral similarities between samples 086 and 112 did, however, contribute to the lower classification accuracy, resulting from leave-one-out cross-validation, experienced during the analysis of data from laboratory 2. Even bigger factors causing decreased discrimination ability in laboratory 2 were the relatively large amounts of noise and intra-sample spectral variations witnessed. These influences often have a negative impact on the classification ability of multivariate models.

Discrimination using PLS-DA on 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 three laboratories in the training set. 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). After all sets of data had been placed on the same wavelength axis, first derivative and SNV preprocessing was performed. Additionally, CCCS was used in an attempt to keep the differing scales of the instruments from confusing interpretation.

Classification models were built using the PLS-DA algorithm on a training set containing 216 (72 from each of the three laboratories with 18 total samples per class) pretreated spectra. The optimum number of LVs used to build the model was based on the average classification error of leave-one-out cross-validation. As displayed in Figure 2, the error of cross validation levels off after 6 LVs. As a result, six LVs were chosen as the optimum number of LVs to base the PLS-DA model on. The classification errors corresponding to calibration and cross-validation of the training set were 5.09% and 13.89%, respectively.

The optimal model was then validated on a test set containing the remaining 144 (48 from each laboratory with 12 total samples per class) pretreated spectra. Each spectrum in the test set was assigned to one of the 12 classes. The classification results for the test set are reported in Table 3. A classification accuracy of 93.75% (135 out 144 spectra correctly classified) was obtained. There was a significant inability of the model to discriminate between samples 086 and 112, two fibers containing the same three cationic dyes. The nearly identical spectra of these samples led to a false negative rate of 75% and false positive rate of 6.81% samples 086 and 112, respectively.

Using data from two laboratories to discriminate samples at a third using PLS-DA. 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, PLS-DA models were built using the data from laboratories one and two, and the data from laboratory 3 was used as an external test set. This process was repeated two more times allowing each laboratory's data a chance to be the test set while the other two laboratories comprised the training set. The parameters used to build the PLS-DA models along with the results of each trial are stated in Table 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 accuracy for the three models was 88.06%. The poorest performance of a PLS-DA model occurred when trying to predict samples from laboratory 2 based on the data from laboratories 1 and 3. 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 2 as a test set.

As seen in Figure 3, there were numerous misclassifications involving groups 086 and 112. As mentioned before, because these samples have the same dye compositions and share extremely similar spectra, it is challenging to discriminate the spectra of these samples visually or by PLS-DA. The other groups of fibers which showed the most misclassifications were samples 091, 092, 113, and 145. Unlike the misclassified fibers mentioned previously, the samples in this group appear to be visibly distinguishable in most instances. The number of misclassifications with these samples can be attributed to noisy spectra, significant intensity differences between

samples of the same class, and, in some instances, the spectra collected at Laboratories 1 and 3 seemed much more comparable in curve shape when compared to the spectra collected at Laboratory 2.

Conclusions. UV/visible 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 whether it has functional capability and performance that meet task 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 microspectrophotometry for fiber characterization.

Transfer of calibration models, particularly in Near-IR spectroscopy, has been a topic of continuing discussion in the literature, with the focus on methods for methods for robust calibration and spectral preprocessing to correct for inter-laboratory variability (*e.g.*, ref. 24). In the present study, classification models based on PCA-LDA in separate laboratories produced an average classification accuracy of 98.33%, an outcome that is indicative of good operational control of methodology and practice within each laboratory. Nevertheless, differences in discriminating ability became apparent when laboratory data sets were fused together and used to predict sample classification over all samples. Permutations were also performed to combine data from two of the laboratories for PLS-DA classification prediction of the same samples from the third laboratory (as an external test set). The average classification accuracy was significantly lower, at 88.06%. Upon detailed visual inspection of the spectra, this observed loss in classification accuracy was attributed to noisy spectra, and/or slight differences in intensity or peak shape.

Acknowledgement. This research was supported provided by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice Award No. 2010-DN-BX-K220. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

References

1. Adolf, P.; Dunlop, J. Microspectrophotometry/colour measurement. In: Robertson, J. Grieve, M. (Ed). *Forensic Examination of Fibres*, 2nd ed.; Taylor and Francis: London, 1999, pp. 251-289.
2. Gaudette, B. The forensic aspects of textile fiber examination. In: Saferstein, R. (Ed). *Forensic Science Handbook*, vol. 2; Prentice Hall: New Jersey, 1988, pp. 209-272.
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: Blackledge, R. (Ed). *Forensic Analysis on the Cutting Edge*; John Wiley & Sons, New Jersey, 2007, pp. 333-374.
5. Barker, M.; Rayens W. Partial least squares for discrimination. *J. Chemometr.* **2003**, *17*, 166-173.

6. Bueno, J.; Sikirzhyski, V.; Lednev, I. Raman spectroscopic analysis of gunshot residue offering great potential for caliber differentiation. *Anal. Chem.* **2012**, 84, 4334-4339.
7. Gottfried, J.; De Lucia, F.; Miziolek, A. Discrimination of explosive residues on organic and inorganic substrates using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2009**, 24, 288-296.
8. Gottfried, J.; De Lucia, F.; Munson, C.; Miziolek, A. Strategies for residue explosives detection using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2008**, 23, 205-216.
9. Orzel, J.; Daszykowski, M.; Grabowski, I.; Zaleszczyk, G.; Sznajder, M. Identifying the illegal removal from diesel oil of certain chemical markers that designate excise duty. *Fuel* **2014**, 117, 224-229.
10. Pontes, M.; Pereira, C.; Pimental, M.; Vasconcelos, F.; Silva, A. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta* **2011**, 85, 2159-2165.
11. Corgozinho, C.; Pasa, V.; Barbeira, P. Determination of residual oil in diesel oil by spectrofluorimetric and chemometric analysis. *Talanta* **2008**, 76, 479-484.
12. Jantzi, S.; Almirall, J. Characterization and forensic analysis of soil samples using laser-induced breakdown spectroscopy (LIBS). *Anal. Bioanal. Chem.* **2011**, 400, 3341-3351.
13. Baron, M.; Gonzalez-Rodriguez, J.; Croxton, R.; Gonzalez, R.; Jimenez-Perez, R. Chemometric study on the forensic discrimination of soil types using their infrared spectral characteristics. *Appl. Spectrosc.* **2011**, 65, 1151-1161.
14. Thanasoulis, N.; Pilouris, E.; Kotti, M.; Evmiridis, N. Application of multivariate chemometrics in forensic soil discrimination based on the UV-Vis spectrum of the acid fraction of humus. *For. Sci. Int.* **2002**, 130, 73-82.
15. Schenk, E.; Almirall, J. Elemental analysis of cotton by laser-induced breakdown spectroscopy. *Appl. Opt.* **2010**, 49, C153-C160.
16. Enlow, E.; Kennedy, J.; Nieuwland, A.; Hendrix, J.; Morgan, S. Discrimination of nylon polymers using attenuated total reflection mid-infrared spectra and multivariate statistical techniques. *Appl. Spectrosc.* **2005**, 59, 986-992.
17. Savitzky, A.; Golay, M. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, 36, 1627-1639.
18. Steiner, J.; Termonia, Y.; Deltour, J. Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.* **1972**, 44, 1906-1909.
19. Barnes, R.; Dhanoa, M.; and Lister, S. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, 43, 772-777.
20. Gemperline, P. Principal component analysis. In: Gemperline, P. (Ed). *Practical Guide to Chemometrics*, 2nd ed., Taylor & Francis: Florida, 2006, pp. 69-104.
21. Fisher, R. The statistical utilization of multiple measurements. *Ann. Eug.* **1938**, 8, 376-386.
22. Egan, W. J.; Morgan, S. L.; Bartick, E. G.; Merrill, R. A.; Taylor, H. J. Forensic discrimination of photocopy and printer toners. II. Discriminant analysis applied to infrared reflection-absorption spectroscopy. *Anal. Bioanal. Chem.* **2003**, 376, 1279-1285.
23. Boor, C. *A Practical Guide to Splines*, Springer-Verlag: Germany, 1978.
24. Sahni, N. S.; Isaksson, T.; Naes, T. Comparison of methods for transfer of calibration models in near-IR 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.

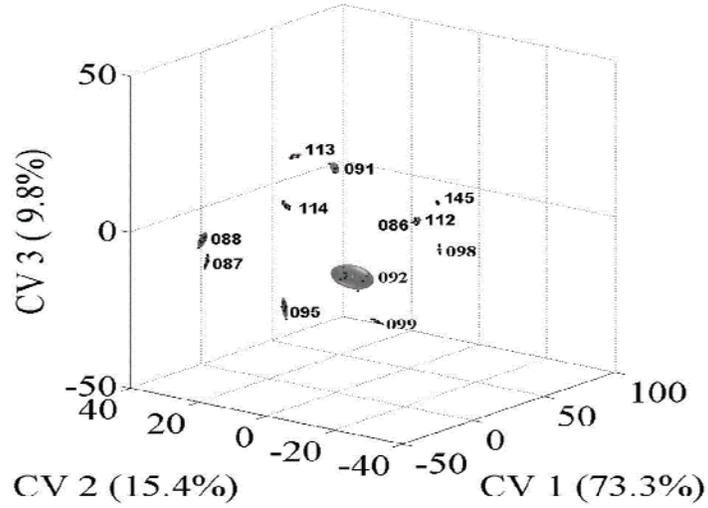


Figure 1. Twelve blue acrylic fiber samples from laboratory one projected into the first three canonical variates representing 98.5 percent of the data.

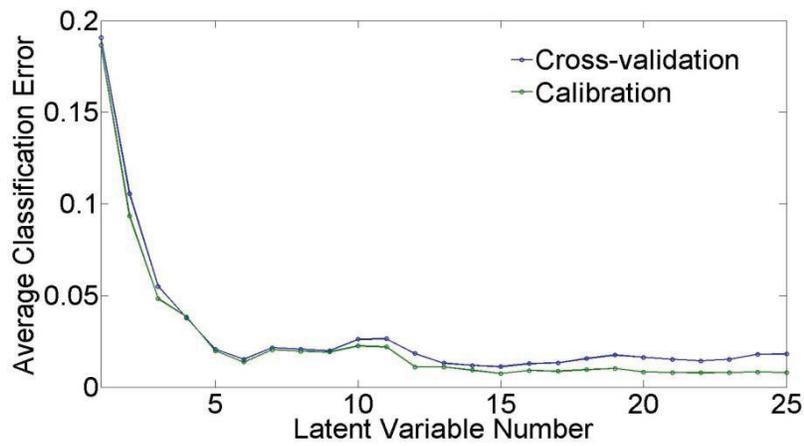


Figure 2. Average classification error of leave-one-out cross validation and calibration for training set containing data from three laboratories as a function of number of latent variables.

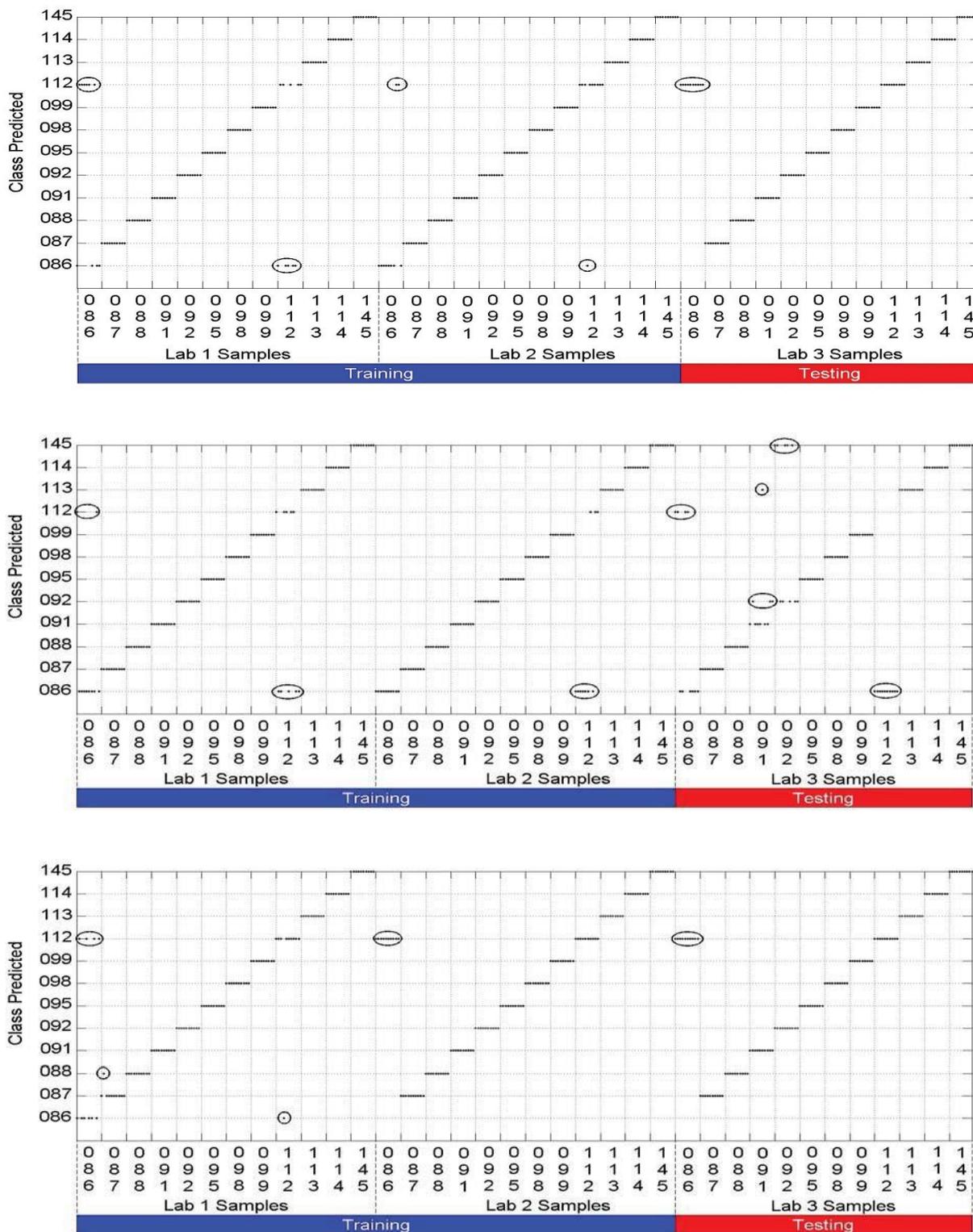
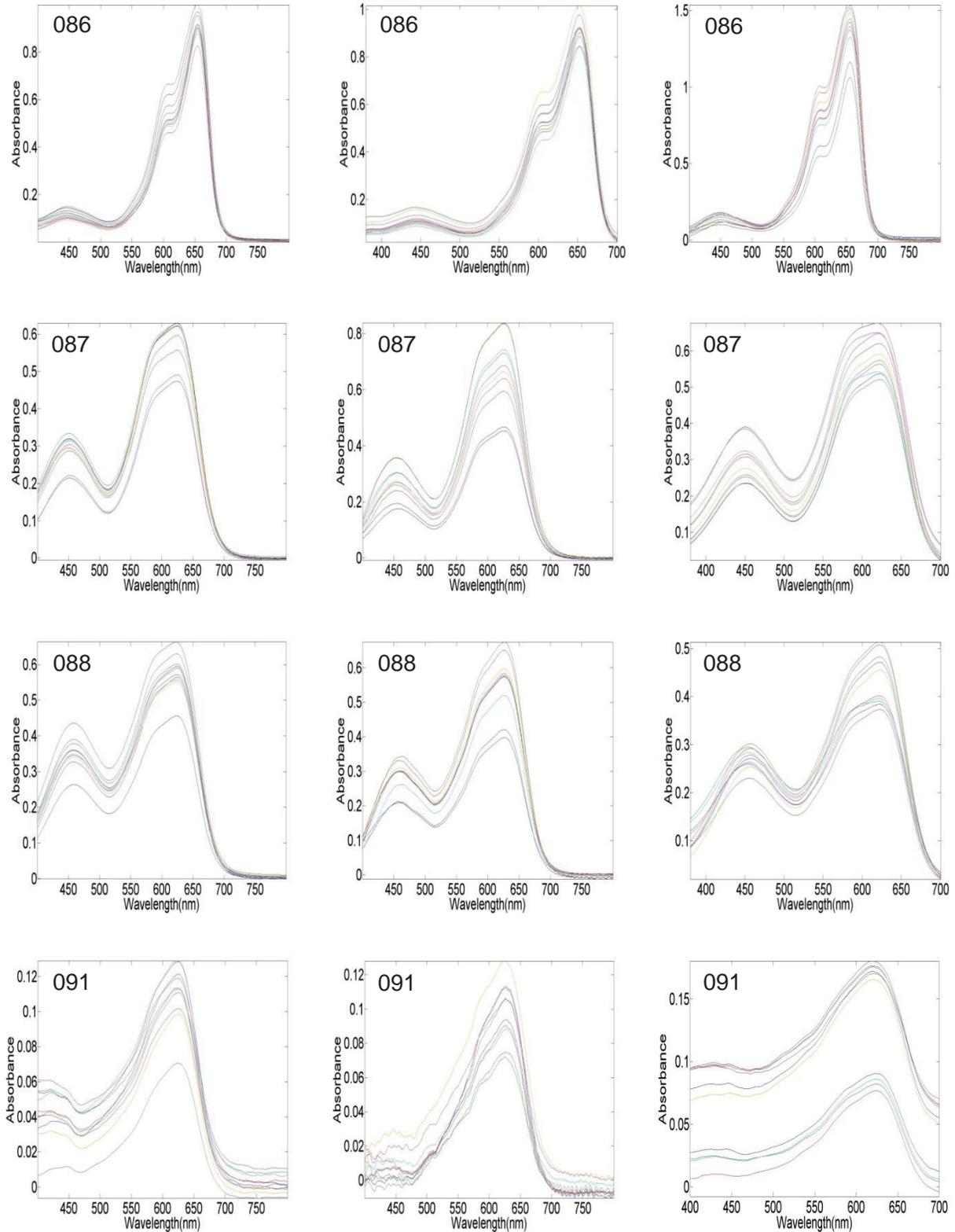
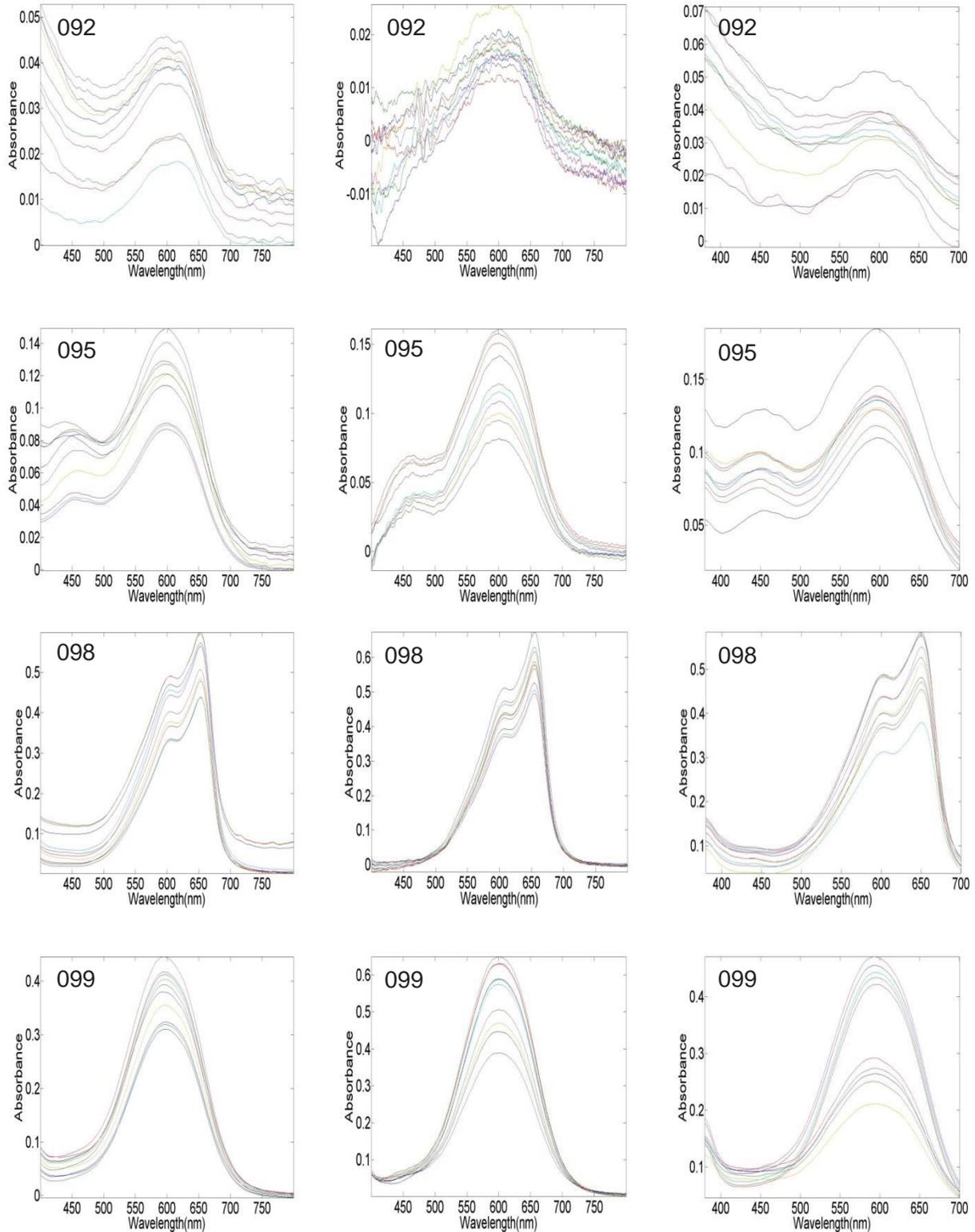


Figure 3. PLS-DA results using data collected from laboratory three (top), two (middle), and one (bottom) as an external test set and using data from the remaining laboratories as the training set.

APPENDIX A. UV/Vis spectra of 12 blue acrylic fibers (10 replicates each) collected at laboratories one (left), two (middle), and three (right) sorted by fiber identification number.



APPENDIX A (Contd.). UV/Vis spectra of 12 blue acrylic fibers (10 replicates each) collected at laboratories one (left), two (middle), and three (right) sorted by fiber identification number.



APPENDIX A (Cont.). UV/Vis spectra of 12 blue acrylic fibers (10 replicates each) collected at laboratories one (left), two (middle), and three (right) sorted by fiber identification number.

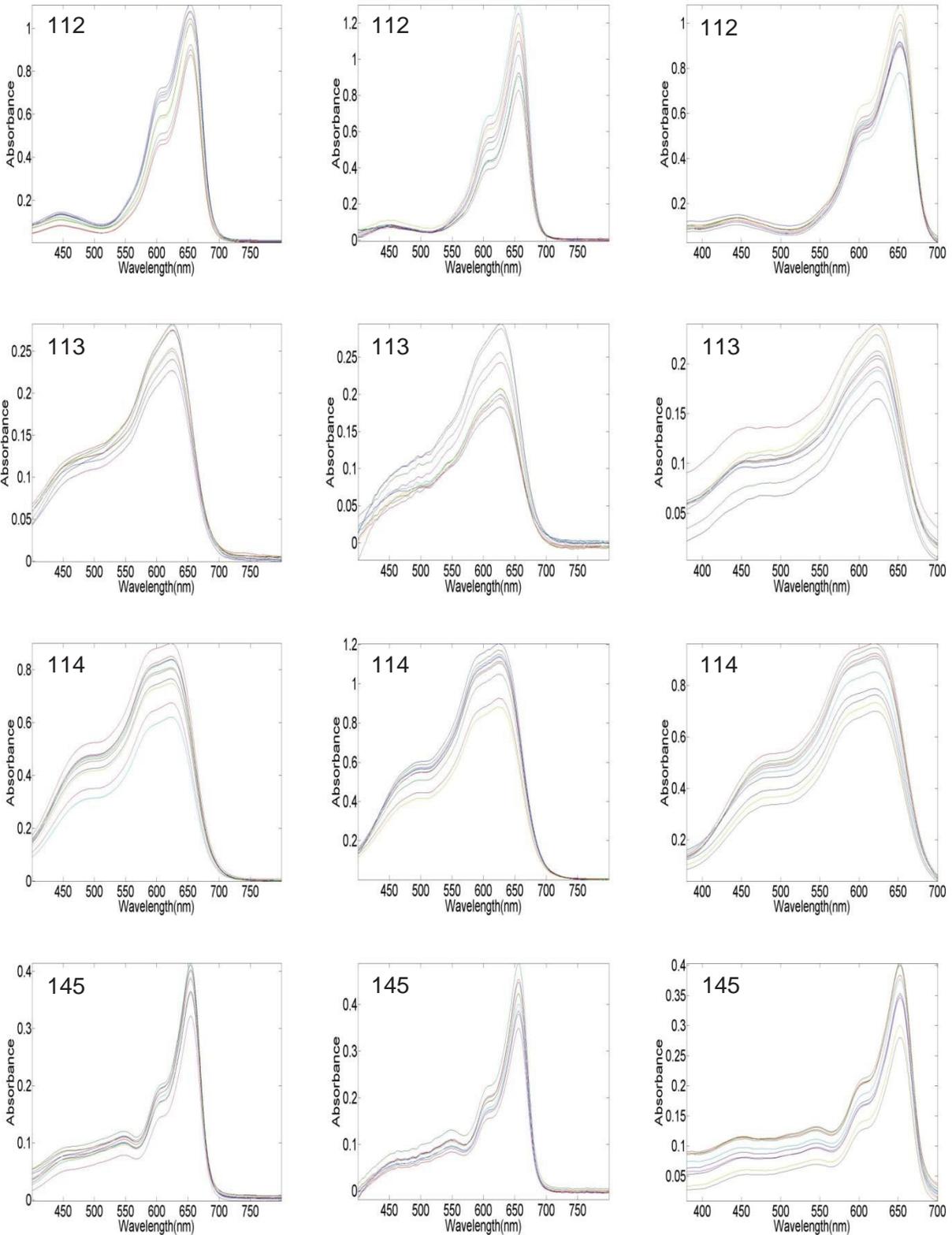


Table 1. Cationic dye composition of the fibers examined in this study. ‘Y’ indicates presence of dye.

Fiber	Cationic dye									
	Blue 3	Blue 41	Blue 60	Blue 147	Red 18	Red 29	Red 46	Yellow 21	Yellow 28	Yellow 29
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 2. Comparison of correct classification rates by PCA-LDA between laboratories.

Laboratory	Classification (%)	Number of correctly classified spectra by sample											
		086	087	088	091	092	095	098	099	112	113	114	145
1	99.17	10	10	10	10	10	9 ^a	10	10	10	10	10	10
2	95.83	10	10	9 ^b	10	10	9 ^a	10	10	9 ^c	9 ^d	10	10
3	100	10	10	10	10	10	10	10	10	10	10	10	10

^aMisclassifications between fibers 095 and 092.

^bMisclassification between fibers 088 and 113.

^cMisclassification between fibers 112 and 086.

^dMisclassification between fibers 113 and 091

Table 3. Confusion matrix resulting from PLS-DA on test set composed of combined laboratory data. Correctly classified spectra are in bold.

Predicted class	<u>Actual class</u>											
	086	087	088	091	092	095	098	099	112	113	114	145
086	3											
087		12										
088			12									
091				12								
092					12							
095						12						
098							12					
099								12				
112	9								12			
113										12		
114											12	
145												12

Table 4. Classification accuracies resulting from calibration, cross-validation, and external testing with the number latent variables used to train each PLS-DA model.

Training lab.	Testing lab.	LVs	Calibration (%)	CV (%)	Test (%)
1 and 2	3	9	94.17	93.75	91.67
1 and 3	2	6	94.17	86.25	80.83
2 and 3	1	6	93.33	82.08	91.67

F. Design of an extensible forensic database for textile fibers.

Stephen L. Morgan^{1*}, David L. Birt¹, Edward G. Bartick², and John V. Goodpaster³.

¹Department of Chemistry and Biochemistry, University of South Carolina, Columbia, SC 29208

²Department of Forensic Sciences, The George Washington University, Washington, DC 20007

³Indiana University Purdue University Indianapolis (IUPUI), Department of Chemistry and Chemical Biology, Forensic and Investigative Sciences Program, Indianapolis, IN 46202

Abstract. The design and implementation of web-forensic fiber database is discussed that facilitates archiving fiber data such as polarized microscopy measurements (birefringence, sign of elongation), physical characteristics (diameter, shape), and spectral data. One immediate advantage is the ability to store data in a documented manner and access this information on demand from a relational database. Although there is little likelihood of establishing a truly comprehensive fiber database because of fast moving trends in manufacturing and globalization of production, a combined data archiving and statistical graphics and analysis system offers both data management and decision-making support to the forensic fiber examiner.

Introduction. As stated by Moore [5], “Long-term data archiving has much value..., not only to retain access to research and product development records, but also to enable new developments and new discoveries. There are some recent regulatory requirements (e.g., US FDA 21 CFR Part 11), but good science and good business both benefit regardless. A particular example of the benefits of and need for long-term data archiving is the management of data from spectroscopic laboratory instruments. The sheer amount of spectroscopic data is increasing at a scary rate, and the pressures to archive come from the expense to create the data (or re-create it if it is lost) as well as its high information content. The goal of long-term data archiving is to save and organize instrument data files as well as any needed meta data (such as sample ID, LIMS information, operator, date, time, instrument settings, sample type, and relevant environmental parameters).”

USC Fiber Collection. The fiber collection at USC began with a collection of 500 fibers from research funded by the FBI Laboratory during 2003-2007. Samples of dyed and undyed acrylic, cotton, nylon, (-6 and -6,6), polyester, and polyester/cotton blend fabrics were obtained, along with smaller of other less common textiles. The sample fabrics contain one to three fiber variants and are dyed with up to six different dyestuffs, often representing several dye classes. Some of the samples are dyed with up to four dyes of a single dye class. The collection presently contains 923 fibers from the textile industry dyed with 273 different dyes (of which we also have chemical samples). Figure 1 shows a photo of a subset of dyed nylon-6,6 samples (along with a roll of undyed fabric) as received.



Figure 1. Nylon-6,6 samples.

Chemical information and physical characteristics are available for most of the fibers (polymer, color, diameter, cross sectional shape, denier, dye class, dyes used, PLM measurements, etc.). Currently, 10 replicate UV/visible absorbance spectra (200-800 nm), and 10 replicate fluorescence spectra (200-800 nm) at four different excitation wavelengths (365, 405, 436, 546 nm) also available for each fiber. IR spectra are present for some spectra. Large stock samples of fabrics, from which many of the fibers in the original USC collection were sampled, are maintained in acid-free plastic storage bags located in filing cabinets. Small representative samples are stored in acid-free protective pockets in three-ring binders. All these fiber samples are kept in a dark air-conditioned room.

The number of fiber samples at USC has expanded by contributions from four sources, although information on new samples is not in the data base files at this time:

- (a) Jennifer Stoner of the Trace Evidence Laboratory at the SC State Law Enforcement Division (SLED) donated about 500 fibers, many undyed, from their copy of the Consolidated Testing Services collection that was distributed by the National Bureau of Standards in the 1980's;
- (b) SLED also donated more than 1,400 residential carpet samples obtained from Lowe's Home Improvement (Columbia, SC) consisting of multiple shades of different colors of different fiber polymers.
- (c) Because of its coverage of additional polymer types, we purchased copies of the Microtrace fiber collection (collected by Skip Palenik, Microtrace, and Michael Grieve, previously at the Federal Police Laboratory, Germany). This collection contains 201 fibers, many undyed, having a wide range of chemical composition. These samples are not in our data base.
- (d) Dr. Hal Deadman (The George Washington University) donated a collection of 200 auto carpet fibers collected from junk yards to Northern Virginia. Automobile models were identified and VIN numbers recorded. We have not analyzed these samples yet.

Database design. Our current database is a demonstration prototype. XML, a standardized data format, is used throughout the implementation to achieve three goals: (1) forensic analytical fiber data characterizing fibers is documented in a human-readable and portable format for text data files (to insure data integrity, clarity of content, and portability between different labs); (2) fiber data is organized in a hierarchical relational database structure, for efficient data storage and data retrieval, using a variety of stored procedures (database queries); (3) automated selection and export of fiber data from the database to facilitate further statistical analysis. Data input on the

front end insures that proprietary data formats from different instruments are translated to a standardized format that can be read by simple programs. Following a database query, for example, to list all red cotton fibers with a specified combination of characteristics), data can be exported for statistical analysis in formats compatible with commercial or in-house written software for statistical comparisons. XML is an extensible markup language endorsed by the World Wide Web Consortium (W3C), and has grown through adoption by the Microsoft Corporation as a file format for Microsoft Office documents. NIST has long recognized the significance of XML. Interagency efforts involving data format standardization using XML include: (a) NIST ITL American National Standards for Biometrics (fingerprints, etc.); (b) ANSI/NIST-ITL 1-2000, Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT) Information; (c) NASA Common Data Format (CDF), an XML data format for the storage and manipulation of scalar and multidimensional data in a platform- and discipline-independent fashion; and (d) the NIST standard format for machine tool performance test data.

The web-based SQL/ASP database currently holds information on 600+ fibers at this time, and facilitates queries for rapid retrieval of data and interactive visualization, export of existing raw data, and import of new fiber data sets. XML is used as a data documentation standard in conjunction with Microsoft SQL. The database is accessed through a Microsoft ASP-driven web site using a standard web browser (e.g., Internet Explorer, Firefox). Data management will encompass electronic signatures on all data entered, validation of data integrity, storage of diverse types of analytical data, including user-written notes providing descriptive information on measurement parameters, spectra, etc. A relational database with a hierarchical structure facilitates queries for rapid retrieval of records, interactive visualization and summary of data using graphical and statistical techniques, and searching/reporting for samples that match user-specified queries. Data export options allow external statistical analysis to be conducted for assessing the degree of matching of patterns of data. We initially envisioned two types of database operation: multiple users in a single facility; multiple users over the Internet, for collaborative sharing of database information across the Web. A secure web-based approach satisfies both possibilities and offers simple upgrade options.

Microsoft SQL server was selected for the database because of its continuing history of development and support, availability of documentation, ability to handle a variety of data storage formats (including XML import and export), extensibility to manage database growth as new data is submitted, and ability to retrieve data on request from other software or from web applications. We are currently using Microsoft SQL Server 2008; upgrading to future versions of Microsoft SQL Server should be easy if desired.

The fiber database contains UV/visible, fluorescence, and IR spectra, as well as dye information, and other physical and chemical data. Figure 1 shows a structure diagram and the object browser summary for the database. The data includes physical characteristics for many of the fibers (color, diameter, cross sectional shape, denier, dye class, dyes used, PLM measurements, and other data if available), absorbance spectra, and fluorescence spectra. Currently, 10 replicate absorbance spectra (200-800 nm), and 10 replicate fluorescence spectra (200-800 nm) at four different excitation wavelengths (365, 405, 436, 546 nm) are in the database for each fiber. IR spectra of the fibers can also be added, but are not presently available for all fibers. This information is stored in tables, accessible by SQL database queries on keys that point to various aspects of the information. For example, each fiber is referenced by a fiber identification number

(Fiber ID), the dye class table contains all dye classes present in the database, the fiber type table contains the list of all textile polymers present in the database, and so on. Relationships among these elements establish the data structure. Thus, each physical fiber sample in our collection has a fiber ID, with which is associated the data available for that fiber sample such as fiber polymer type, fiber dye class, specific dyes with which the fiber is dyed, and other characteristics that are known.

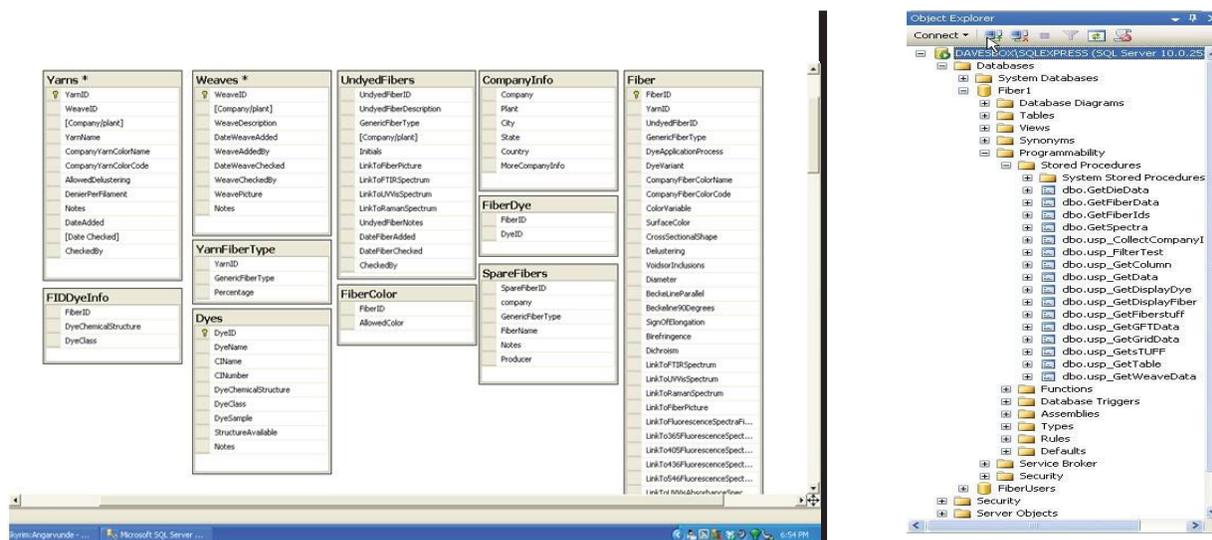


Figure 1. Fiber database diagram (left) and object browser summary (right).

Constraints on proprietary information from manufacturing companies may be limit releasing dye information in a few cases; we have agreed not to release dye information for fibers that are dyed with colorants that are not in the *Colour Index* (from the Society of Dyers and Colourists and the American Association of Textile Chemists and Colorists, Raleigh, NC). However, this constraint only applies to a small number of the dyes currently in the database.

Microsoft SQL enables functionality for handling the data, passing stored data into and from the database with other applications. Stored procedures (programs) can be executed to display data in a variety of ways, or perform, and to export the data in several formats. Not all of the possibilities are currently programmed, but will be implemented as needed. The fiber database is hosted by a Microsoft ASP.NET web application which provides users with the ability to interact with the data in various ways. Specifically, ASP.NET contains common web forms, built-in controls, and other elements that together define the client-side user interface elements. Because ASP.NET applications run in a web browser that communicates transparently with the server-based database, the user does not have to download a ‘fat’ client program that occupies extensive storage space or that requires frequent updating. Users simply access the web application URL to use the latest version of the software. ASP.NET also provides the developer with tools to make changes to the user interface for updating forms and data.

Database functionality. A Login account (Figure 2) is required to access the database. The login uses Microsoft Windows authentication to verify prerequisites concerning the password length or strength, and then checks a user database for credential verification. The SQL database with the user’s personal information is stored with the fiber database on the server behind a firewall. The

stored procedure that writes and reads this data does so under 128-bit encryption to secure access to this data. For new users, a sign-up link redirects the user to a sign-up page; the login indicator at the bottom indicates whether the user is logged in.

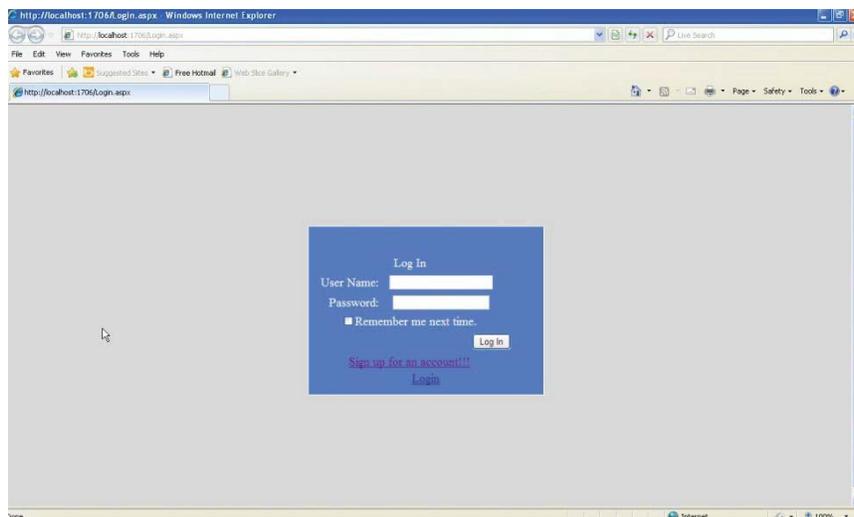


Figure 2. User login web page.

Web pages displayed are development versions and final pages will have identical color palette, style, and font for a consistent 'look and feel.' The Sign-up page (Figure 3) displays a Windows control that takes user account information and writes to the user account SQL database (FiberUser) after hashing and salting sensitive information. An information page (Figure 4) then solicits information from the user to validate account approval. Policies will provide access control for legitimate forensic laboratory users only and will be firmed up at a later date; a privacy policy will also be formulated. The information page is a validation page that requires that all fields be filled in before the user can proceed. Upon successful completion, user information is displayed in a summary, and an email is sent to the database administrator for approval to create a new user account.

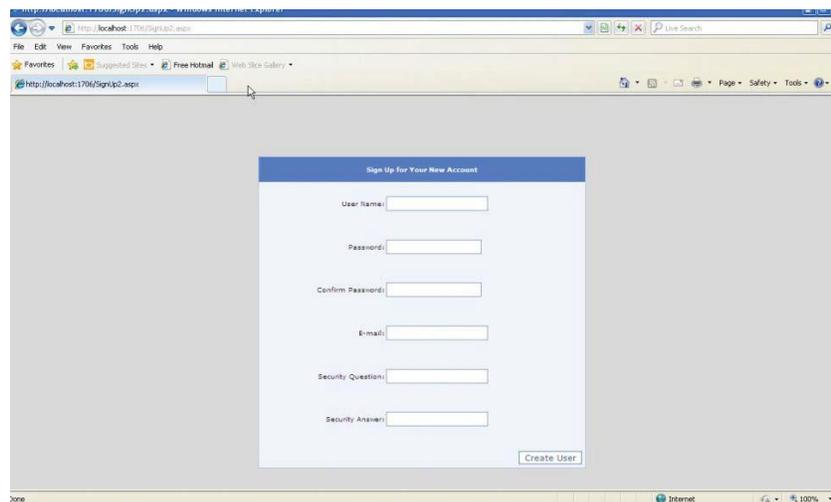


Figure 3. Account information web page.

The screenshot shows a web browser window with the address bar displaying 'http://localhost:1706/SignUp.aspx'. The browser's menu bar includes 'File', 'Edit', 'View', 'Favorites', 'Tools', and 'Help'. The address bar also shows 'http://localhost:1706/SignUp.aspx'. The main content area features a blue rectangular form titled 'Tell Us About Yourself'. The form contains the following fields: 'First Name:' with a single-line text input; 'Last Name:' with a single-line text input; 'Email:' with a single-line text input; 'Company:' with a single-line text input; and 'Purpose:' with a larger multi-line text area. A 'Submit' button is located at the bottom of the form. The browser's status bar at the bottom indicates 'Internet' and '100%' zoom.

Figure 4. User information web page.

Once the user account is verified, the Main fiber selection page is displayed upon login (Figure 5). This page is basically a search page that solicits, from the user, from 1 to 5 different values. All the fibers in the database with the specified characteristics are returned in a Selected Fiber window grid (Figure 6). This page displays the result set returned from the database—a list of fibers that match the selected characteristics previously selected. The check list at the bottom left of the screen enables further filtering of the fields that are returned in the results grid.

Figure 6 shows the first page of a listing of all fibers in the database. Figure 7 shows a search report for a request to show all fibers in the database having a diameter of 15 mm. To view database information on any fiber that is listed in the right hand display pane of the Selected Fiber grid of Figure 6, the user clicks on the “Select” link to the left of the fiber ID, and the Fiber Details web page (Figure 7) shows all the characteristics in the database for the fiber ID selected. The stored procedure that plots spectra has recently been replaced by a new routine written in Microsoft Visual Basic (see Figure 8 for new plot output) and connected, as are other functions, to the an on-demand ASP.NET web link. At present, a fiber ID window allows plotting of spectra for any selected fibers. A future revision will implement plots enabling comparison of spectra from different fibers.



Figure 5. Main fiber selection web page.

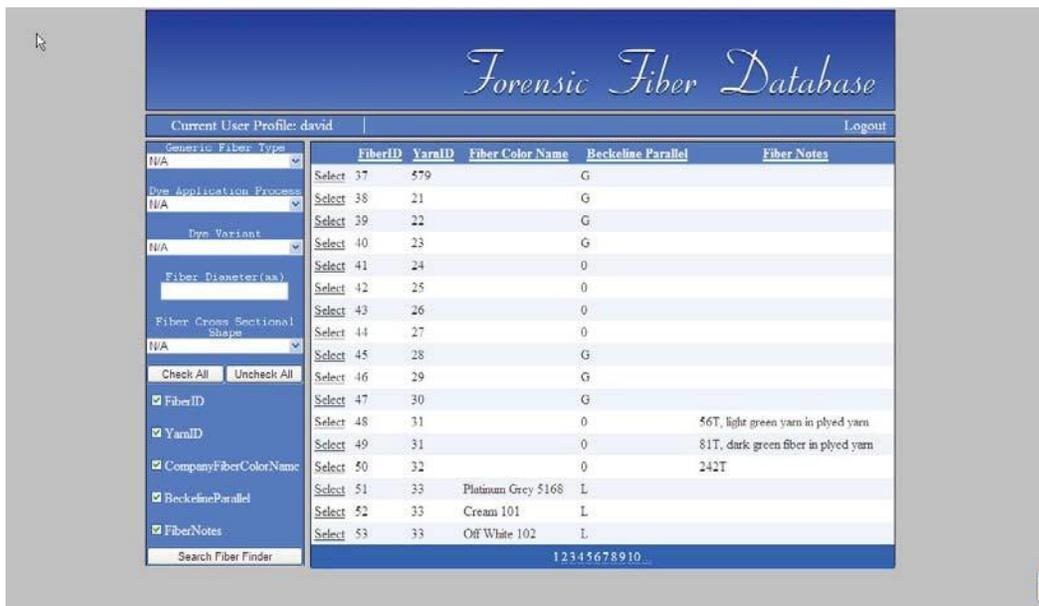


Figure 6. Selected fiber web page, listing all fibers.



Figure 7. Search results for fiber diameter 15 mm.

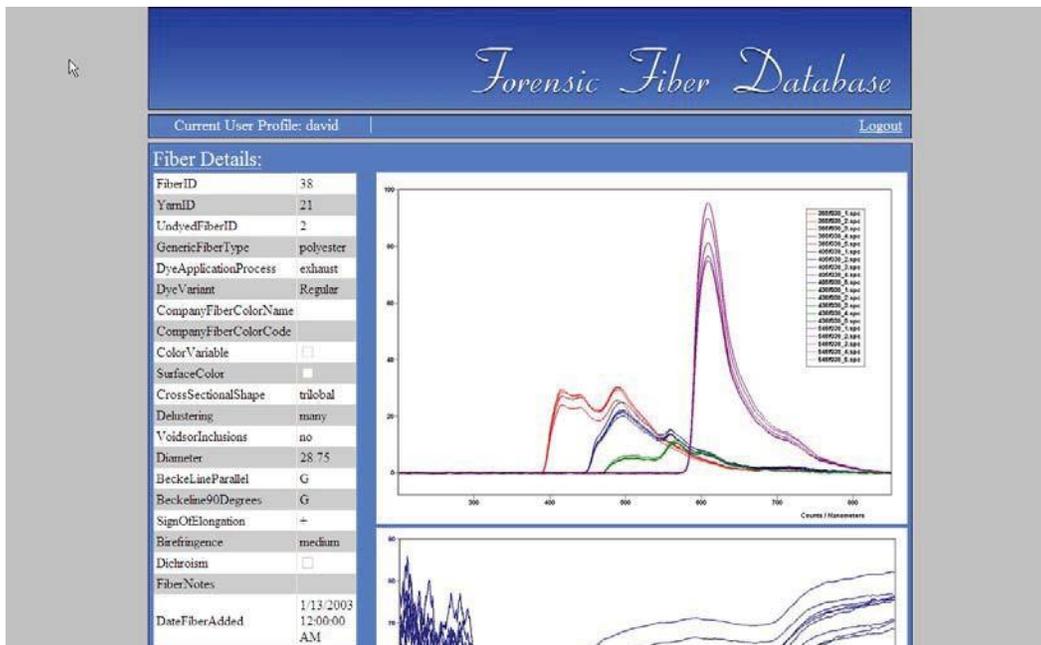


Figure 8. Fiber Details web page for fiber ID 38.

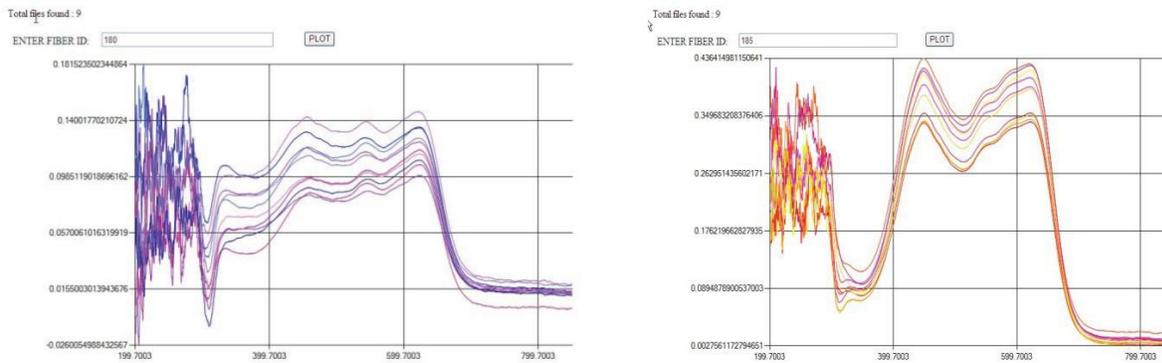


Figure 8. Spectral plots on-demand from database.

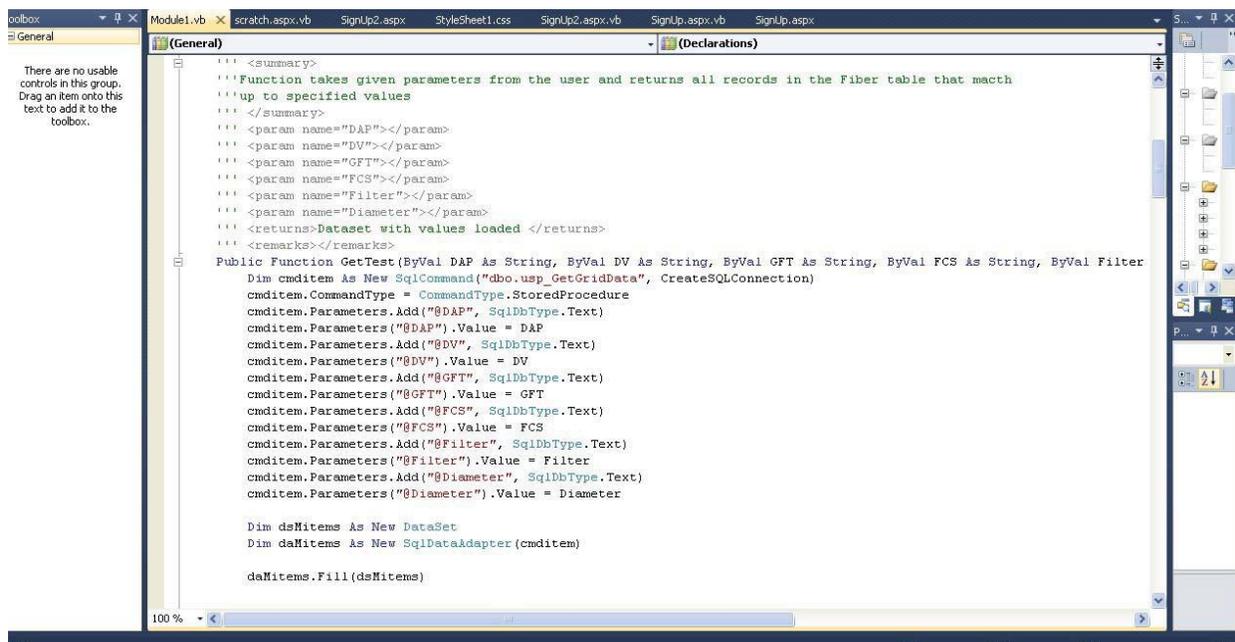
The Dye Details web page (Figure 9) is accessed in a similar fashion. The dyes present on the selected fiber (if known) are displayed in a Windows gridview control. At the right of the display, the “Select” link for a given dye uses the unique Dye ID to load a Dye Information web page containing all the details for a specific dye. A future link in the DyeChemicalStructure column will load a web page with a ChemDraw® image of the molecular structure for the selected dye and its molecular weight (if both available). Image files for all dyes in the database to populate these displays have been created. If the fiber was collected as part of a cloth or yarn, the weave and yarn detail are populated with pertinent information. A photograph of a single fiber, the fabric sample from which it originated, and the sample submission checklist can also be added to the database (Figure 10).

Dye Details:							
	DyeID	DyeName	CIName	CINumber	DyeChemicalStructure	DyeClass	Notes
Select	20	Dianix Pink AM-REL	not in C.I.			disperse	old name: Palanil Brill Pink E-REL
Select	22	Intrasil (Brill) Blue BNS	not in C.I.			disperse	
Select	109	Terasil Yellow W-6GS	C.I. Disperse Yellow 114			disperse	powder and paste available in our sample library

Figure 9. Dye Details web page.

Figure 10. Weave Details web page.

Programming in Visual Basic with SQL. Figures 11-14 display several of the functions written to run the “business rules”, or executable database procedures for code that: returns records in the Fiber table that match specified values selected by the user; writes an XML schema for a specified table and loads the data for output; returns a row of data from a specified Table Name and Fiber ID number; and, returns all dyes associated with a given Fiber ID number and retrieves the associated dye data. These functions serve to automate tasks such as showing all cotton fibers in the database, show all cotton fibers dyed with vat dyes, show all vat dyes, and show all cotton fibers dyed with vat dyes that are colored red. The requested information can be viewed as output on the appropriate web page (see above), and spectra can be examined. All reports, data, and plots will be able to be exported to external text files.



```

<summary>
    '''Function takes given parameters from the user and returns all records in the Fiber table that match
    up to specified values
    '''
</summary>
    '''
    <param name="DAP"></param>
    <param name="DV"></param>
    <param name="GFT"></param>
    <param name="FCS"></param>
    <param name="Filter"></param>
    <param name="Diameter"></param>
    <returns>DataSet with values loaded </returns>
    <remarks></remarks>
Public Function GetTest(ByVal DAP As String, ByVal DV As String, ByVal GFT As String, ByVal FCS As String, ByVal Filter
Dim cmditem As New SqlCommand("dbo.usp_GetGridData", CreateSqlConnection)
cmditem.CommandType = CommandType.StoredProcedure
cmditem.Parameters.Add("@DAP", SqlDbType.Text)
cmditem.Parameters("@DAP").Value = DAP
cmditem.Parameters.Add("@DV", SqlDbType.Text)
cmditem.Parameters("@DV").Value = DV
cmditem.Parameters.Add("@GFT", SqlDbType.Text)
cmditem.Parameters("@GFT").Value = GFT
cmditem.Parameters.Add("@FCS", SqlDbType.Text)
cmditem.Parameters("@FCS").Value = FCS
cmditem.Parameters.Add("@Filter", SqlDbType.Text)
cmditem.Parameters("@Filter").Value = Filter
cmditem.Parameters.Add("@Diameter", SqlDbType.Text)
cmditem.Parameters("@Diameter").Value = Diameter

Dim dsMItems As New DataSet
Dim daMItems As New SqlDataAdapter(cmditem)

daMItems.Fill(dsMItems)

```

Figure 11. Code for Function GetTest returns records in the Fiber table that match specified values selected by the user.

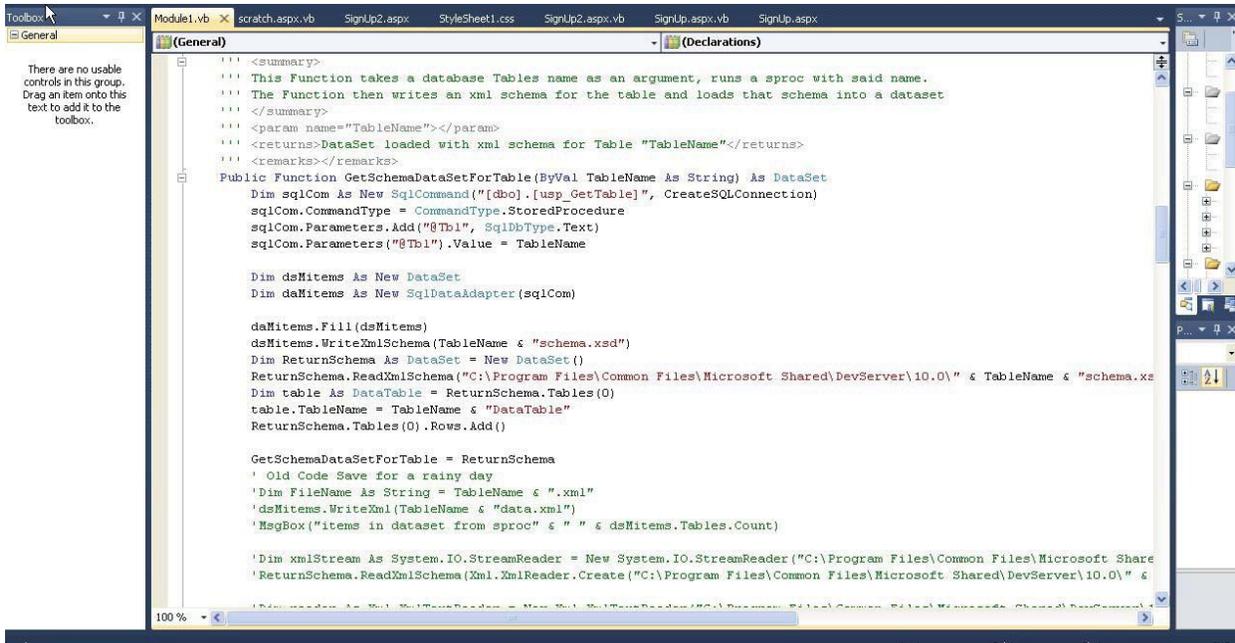


Figure 12. Code for Function GetSchemaDataSetForTable writes an XML schema for a specified table and loads the data into DataSet for output.

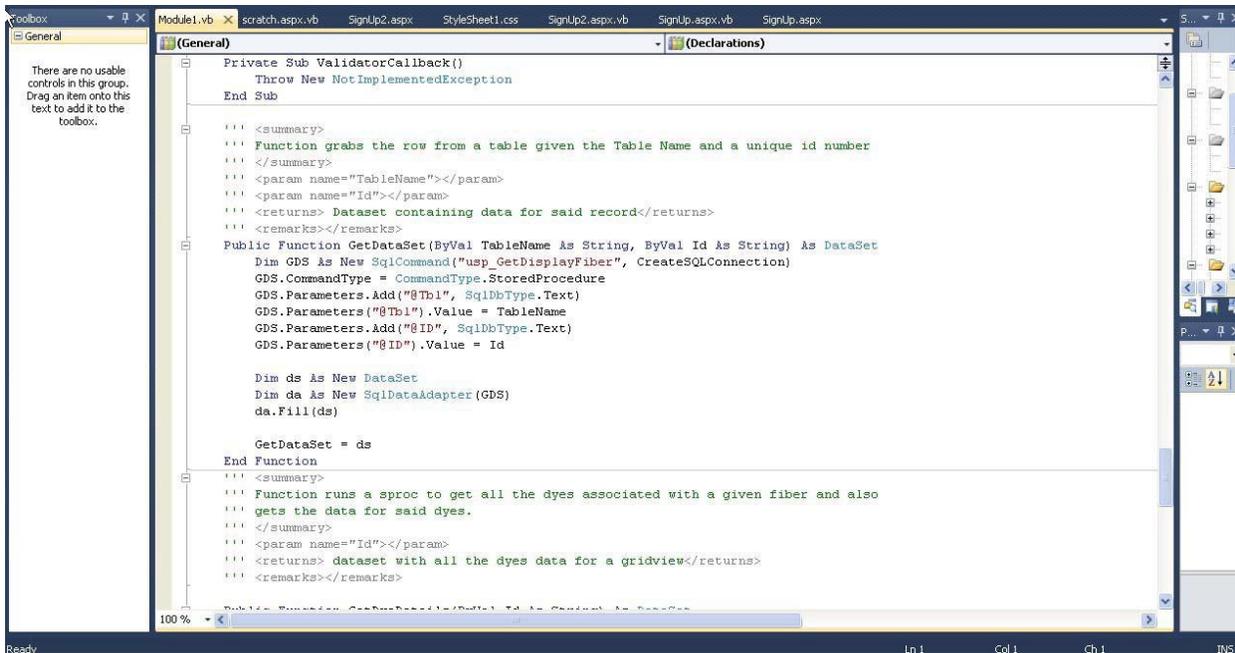


Figure 13. Code for Function GetDataSet returns a row of data from a specified Table Name and Fiber ID number.

```

Module1.vb
scratch.aspx.vb
SignUp2.aspx
StyleSheet1.css
SignUp2.aspx.vb
SignUp.aspx.vb
SignUp.aspx

General
(Declarations)

    GetDataSet = ds
End Function
    <summary>
    <<Function runs a sproc to get all the dyes associated with a given fiber and also
    <<gets the data for said dyes.
    </summary>
    </summary>
    <param name="Id"></param>
    <returns> dataset with all the dyes data for a gridview</returns>
    <remarks></remarks>

Public Function GetDyeDetails(ByVal Id As String) As DataSet
    Dim GDS As New SqlCommand("usp_GetDisplayDye", CreateSqlConnection)
    GDS.CommandType = CommandType.StoredProcedure

    GDS.Parameters.Add("@ID", SqlDbType.Text)
    GDS.Parameters("@ID").Value = Id

    Dim ds As New DataSet
    Dim da As New SqlDataAdapter(GDS)
    da.Fill(ds)

    GetDyeDetails = ds
End Function

End Module
  
```

Figure 14. Code for Function GetDyeDetails returns all dyes associated with a given Fiber ID number and retrieves the associated dye data.

Approval of a request for permission from the USC Information Technology department is required to host a web server on the USC web backbone (<http://www.sc.edu/>). At present, the database is only available to local users within the Morgan laboratory at USC.

Conclusion. We have made an excellent start on a usable forensic fiber database and a demonstration of what is possible with a web-based system. We have the resources to complete the database and to host a site on the Internet, but have at least six more months of work ahead of us to complete database programming, testing, and publishing to the web. We expect to continue database refinement and produce a description of the database capabilities as a research publication for a forensic journal.

Acknowledgement. This research was supported provided by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice Award No. 2010-DN-BX-K220. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

OVERALL RESULTS

The objectives of the proposed research include: (a) to conduct interlaboratory experiments to evaluate decision making in forensic fiber examinations by polarized light microscopy measurements, UV/visible microspectrophotometry, and IR spectroscopy; (b) to investigate the application of multivariate statistical measures for evaluation of comparisons of questioned (Q) vs. known (K) fibers; (c) to evaluate intra-laboratory variability, inter-laboratory agreement, and error rate performance in designed experiments; (d) to document good laboratory practices relevant to achieving acceptable levels of intra- and inter-laboratory consistency in fiber data; (e) development and use of a prototype forensic data management system for fiber examinations that will integrate electronic signatures for documentation on data stored, data validity checking, and relational database searching. We expect the outcome of this project will establish a performance baseline that will be relevant to discussions of fiber discrimination and will improve the ability to express the significance of this important type of class evidence.

CONCLUSIONS

1. Discussion of findings. Fiber evidence in general, can be vital in forensic case work. The transfer of fibers during violent crimes often occurs when violent contact causing tearing of clothing on a suspect or a victim, or when victims are dragged over residential carpets and/or subsequently placed in automobiles.¹⁻⁴ Perhaps the highest profile case involving fibers was the murder of 30 African American young men and children in Atlanta, during 1979 to 1980. Multiple transferred carpet fibers, residential and automotive, found on the victims played a pivotal role in the conviction of Wayne Williams for the murder of two of the victims.^{5,6}

Fibers often share properties or characteristics due to common or similar sources, manufacturing methods, or treatments. The standard positive conclusion typically derived from a forensic fiber examination is the equivocal statement that “The questioned fiber exhibits the same physical, optical and chemical properties as the known sample. Therefore, these fibers could have originated from the same source as the known sample or another fiber source composed of fibers with the same properties”.⁷⁻⁹ While this conclusion is accurate, and is the current accepted statement for dealing with class evidence, no statistical basis of evidential significance is currently accepted.

With successful development of forensic databases such as CODIS for DNA, PDQ for automobile paint, and AFIS for fingerprints, the creation and maintenance of databases for trace evidence is recognized to be of paramount importance in addressing issues of class evidence. The Research, Testing, Development and Evaluation Interagency Working Group (RTD&E IWG) of the National Science and Technology Council Subcommittee on Forensic Science asked, “What databases are most needed in the field of fiber analysis?”¹⁰ The Scientific Working Group for Materials Analysis (SWGMA) responded: “An up-to-date, comprehensive automotive carpet fiber database along the lines of the PAINT Data Query (PDQ)—searchable for investigative leads”. SWGMA also stated that “a fiber population database is not recommended for statistical use at this time due to the ever changing colors of fibers and textiles due to style and season as well as post-manufacture changes from exposure to sun, laundering, *etc.* The wide range of countries that manufacture fibers would also make it almost impossible to get a truly representative sample—it would not be advisable to reference a fiber found in a case against a database that may be limited by fiber manufacturer. □ Databases are usually specific to

manufacturers, and other publications paint only broad strokes in regard to manufacturing output; *e.g.*, *Fiber Organon* reports on world fiber production.¹¹

Houck was among the first to voice dissatisfaction with lack of appropriate statistics in trace evidence evaluation.¹² The 2009 National Academy of Sciences highlighted this issue when recommending, “A statistical framework that allows quantification of these claims is greatly needed” (p. 189).¹³ In regard to fiber analyses, the National Academy of Sciences Report states that “it would be possible in principle to develop an understanding of the uncertainties associated with those analyses. However, to date, this has not been done” (p. 163).¹³ It is our belief that research in the appropriate application of statistical methods can also lead to improved decision-making for class materials in general. Cross-validation comparisons of these measurements in our laboratories, interpretation of matches and exclusions via statistical methods, and dissemination of database information will benefit forensic fiber examiners by supporting scientific rigor in trace fiber examinations.

For a given type of class evidence to be truly probative, it must have a reasonable frequency of occurrence in matters of legal relevance, significant and well-documented diversity must be present in the population under study, and laboratory methods must exist that can reliably discern this diversity. For the evidence to be of forensic value, the probability of a coincidental association between an unknown and known exhibit should be low, and the burden is on the forensic scientist to evaluate this risk. This can be done by demonstrating that the sample type in question is both diverse and differentiable using the following key steps: (1) Understand the product population, including manufacturing and distribution; (2) Obtain a large, representative collection; (3) Analyze samples using multiple, orthogonal techniques to maximize information content; (4) avoid microheterogeneity with appropriate sample sizes, (5) Assess diversity of the sample collection with rigorous quantitative methods; (6) Monitor changes in the sample population over time. (7) Employ appropriate statistical hypothesis tests for evaluation of probabilities.

The diversity and depth of the proposed database is critical to its usefulness as a foundation for establishment of the statistical significance of decision-making. Our proposed database will reveal the natural variation in fiber characteristics of representative carpet fibers as well as the discriminating power of those characteristics relevant to fiber comparisons. The “product rule for independent events” was used in the Williams trial to show improved evidential strength based on occurrence of multiple independent events,^{5,6} This rule states that for two or more independent events, the probability of all events occurring simultaneously is the product of the individual probabilities.¹⁴ Independent events are specified, meaning that the outcome of the one event cannot influence another in any way. This is true for 'orthogonal' measurements that characterize different aspects of an evidence item, but may not be true of all measurements. While such probabilities could be very small, fibers are class evidence. It cannot be stated unequivocally that fibers originated from the same source. However, this approach can provide support for small random match probabilities of finding like fibers on both a suspect and at a crime scene.

Understanding the significance of fiber evidence must be based on a thorough background of textile manufacturing practices and of the prevalence of fiber types in various regions of the world.^{11,47} Mass production has resulted in the presence of textile fibers in numerous different and abundant commercial products. Further, when combinations of polymer types, colors, morphology, etc., are all taken into account, enormous numbers of different fibers exist.

Establishing a collection of fibers that is representative of all possibilities is complicated by rapid changes in manufacturing practices and globalization of textile production: the population is a moving target of indeterminate size and evolving diversity.

Most scientists are familiar with statistics associated with a single type of measurement variable. For example, birefringence measurements from a fiber might be compared with that of fibers found at a crime scene. The resulting data values have only one variable (birefringence) measured for a number of different objects (different fibers). Statistics used for summarizing univariate measurements include sample means and standard deviations. Univariate procedures (calculations of means, standard deviations, confidence intervals, two-sample *t*-test, *etc.*) for small data sets can be conducted by hand, on small calculators, or with a spreadsheet. However, instrumental methods produce data of high dimensionality. Microspectrophotometry produces spectral intensities (absorbance) at several thousand wavelengths. Means and standard deviations (or variances) are also employed in multivariate statistics, but must be calculated for each variable; additionally, covariances (or correlations), measuring the strength of the linear relationships between variables, are calculated. These statistics are not single numbers, but are arrays (matrices) of numbers describing the correlations between all the measured variables. Increases in computing power have made computationally intensive data analysis feasible, and increased availability of software for multivariate statistics has made these techniques accessible.

Statistical techniques of applicability to fiber examinations include statistical hypothesis testing methods, such as analysis of variance (ANOVA), and ROC graphs for comparison of univariate data and derived statistics. Multivariate methods include cluster analysis, principal component analysis, discriminant analysis, and multivariate analysis of variance (MANOVA). To use statistics effectively, multiple replicate spectra must be obtained from each individual fiber. Replicate spectra assess experimental variability (which is required for each of the statistical procedures), and facilitate detection of unrepresentative spectra. Whereas a two-sample *t*-test is a parametric test comparing two means for evidence of a statistically significant difference, ANOVA is a method for testing the equivalence of a group of means.⁵⁰ Whether the null hypothesis of equality can be rejected at a stated level of confidence is attracting growing interest in tests of equivalence (e.g., bioequivalence for drug formulations).^{51,52} Spiegelman demonstrated a test of equivalence on bullet lead data in another NRC report.⁵³ Cluster analysis is useful as an exploratory technique to identify natural groupings (i.e., fibers that might be discriminated from one another).^{54,55} Principal component analysis (PCA)^{56,57}, first discussed by Hotelling⁵⁸, projects multivariate data into a reduced dimensional display that retains as much of the original variability as possible in as few dimensions as possible. Linear discriminant analysis constructs a lower dimensional data display which best separates predefined groups (e.g., groups of replicate spectra of the same fiber) by defining axes that maximize the ratio of their between- to within-group variances.⁵⁸⁻⁶² If a sufficiently large proportion of the variability associated with the first few discriminant axes, a projection of the data points (the spectra) in the two- or three-dimensional space of the discriminant vectors permits the researcher to visualize clustering and similarity of the data. Clustering of similar samples can be assessed by comparison to the distances between spectra judged different from one another. The multivariate generalization of the univariate Student's *t* test is Hotelling's T^2 test for the equivalence of means.⁶³

Our draft manuscripts in the Technical Report section document our applications of multivariate statistics to fiber discrimination that have resulted from NIJ funding. As is often said about the problem of educating scientists to use statistics, the issues most discussed are often about which

statistical approaches are 'best'. In fact, the majority of the benefit of statistics, when applied to understanding complex data, arises from the use of simple systematic comparisons with supporting descriptive statistics. It is our belief that if simple graphics do not show discrimination, no amount of statistical machinery will be convincing.

To emphasize this point further, our first example is simple, but illustrates the power of effective statistical graphics. Figure 1 (left) shows sets of ten replicate UV/visible spectra from two similar fibers (groups 4 and 5 on the left), and from two dissimilar fibers (groups 1 and 7 on the right). Next to each figure showing the spectra is a plot of all the absorbance intensities from the two fibers being compared against one another in a correlation plot. The interpretation is straightforward: Groups 4 and 5 are spectra from two indistinguishable fibers—the correlation plot is nearly a straight line with an overall correlation coefficient of 0.9981; Groups 1 and 7 are clearly distinguishable (the spectra are visually different) and the correlation plot (at the far right) indicates that the spectral intensities do not 'track one another well.' Profound knowledge (in this case, knowing the dye composition on these fibers) tells us that group 1 and group 2 spectra that from a fibers dyed with the same dye, but having slightly different dye loading. In another simple graphic (Figure 2), the distribution of correlation coefficients for pairwise comparisons of spectra belonging to the same group (with group) and belonging to different groups (among groups) are shown. The clear gap between these two distributions for groups 1 and 7 signals dissimilarity; the overlap of these two distributions for groups 4 and 5 signals similarity. The percent area overlap of the two distributions measures the fractional false positive and false negative errors rates; clearly these errors rates are high for the comparisons shown on the right for group 4 and 5.

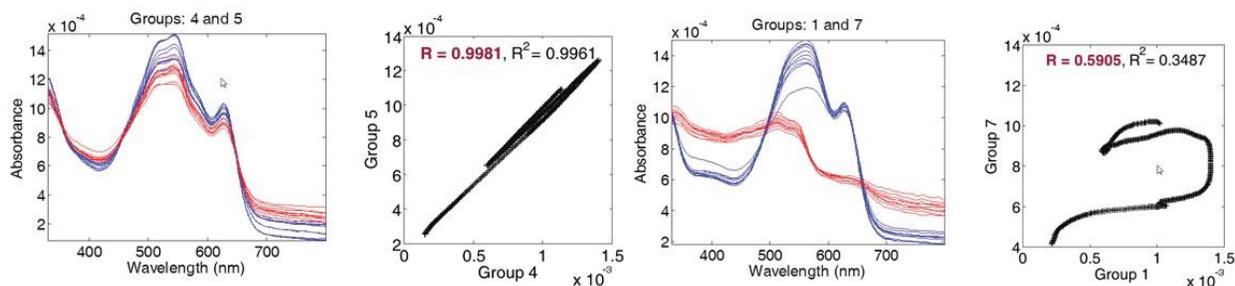


Figure 1. Correlation plots of similar fibers (left), and dissimilar fibers (right).

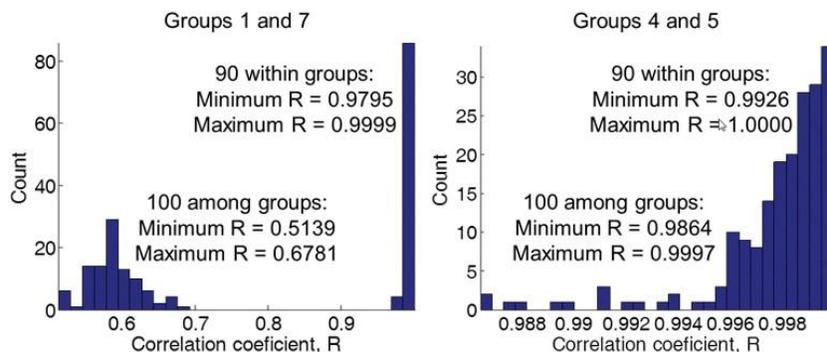


Figure 2. Plots comparing distributions of correlation coefficients from fiber groups 1 and 7 (left side) and groups 4 and 5 (right side).

A further analysis of this data, shown in Figure 3, is a plot of single feature-at-a-time tests for the difference in means for the two groups of spectra across the 2000+ wavelength features. The green horizontal lines represent the critical value of the t -statistic for rejecting the null hypothesis the means are equal. The red horizontal line is the Bonferroni-corrected critical value of Student's t (after accounting for the loss of confidence in conducting multiple hypothesis tests). Note that comparing the spectra between groups 1 and 7 the calculated t -statistic (blue line) jumps over 20 in the 350-500 nm range and over 60 in the 600 nm range. For the comparison of spectra from groups 4 and 5, the calculated t -statistic barely rises above the critical values over the whole wavelength ranges; while the means of the two groups of replicate spectra are slightly different, there is considerable overlap and reliable discrimination is not possible. The plots shown in these three figures are examples of easily interpreted statistical graphics for support of decision-making in such spectral comparisons.⁶⁴

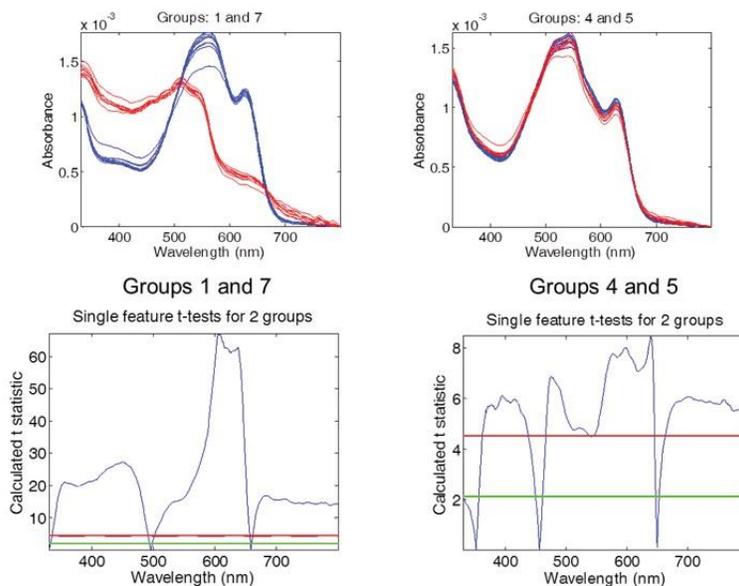


Figure 3. Single feature t -tests from groups 1 and 7 (left side) and groups 4 and 5 (right side).

The next example is based on the 80 UV/visible spectra of red acrylic fibers discussed in part B of the technical Report.⁶⁵ Each of ten replicate spectra from eight different fibers are shown in Figure 4 (top); clearly, replicate spectra of the first five fibers are clearly distinguished from one another, whereas spectra of fibers 6-8 are not distinguishable. When the spectra are normalized, subjected to PCA for dimensionality reduction, then LDA for discrimination, projections of the 80 spectra into the space of the first three linear discriminants analysis are shown in Figure 4 (bottom). This modeling process is straightforward and performed in just a few minutes using our menu-driven statistics software. The resulting map depicts the similarity of the 8 fibers, as points plotted in a 3-dimensional space with 95% confidence ellipses for the 10 spectra corresponding to each fiber. The ability to classify each of the 80 spectra is evaluated by leaving each spectrum out of the data matrix one-at-a-time (treating it as an unknown), repeating the LDA model, then assigning the 'unknown' spectrum into the group closest to it (by Mahalanobis distance). For the present data, the classification accuracy is 88.75%, with 71/80 spectra correctly classified, and with misclassifications occurring between spectra of groups 6, 7, and 8 (at the right side of the LDA plot). This outcome exactly what was expected from visual inspection of the spectra. Multivariate analysis offers easy-to-use tools that provide statistical

rationale for the discrimination of fibers 1-5, and for the decision that fibers 6-8 are more difficult to reliably distinguish from one another.

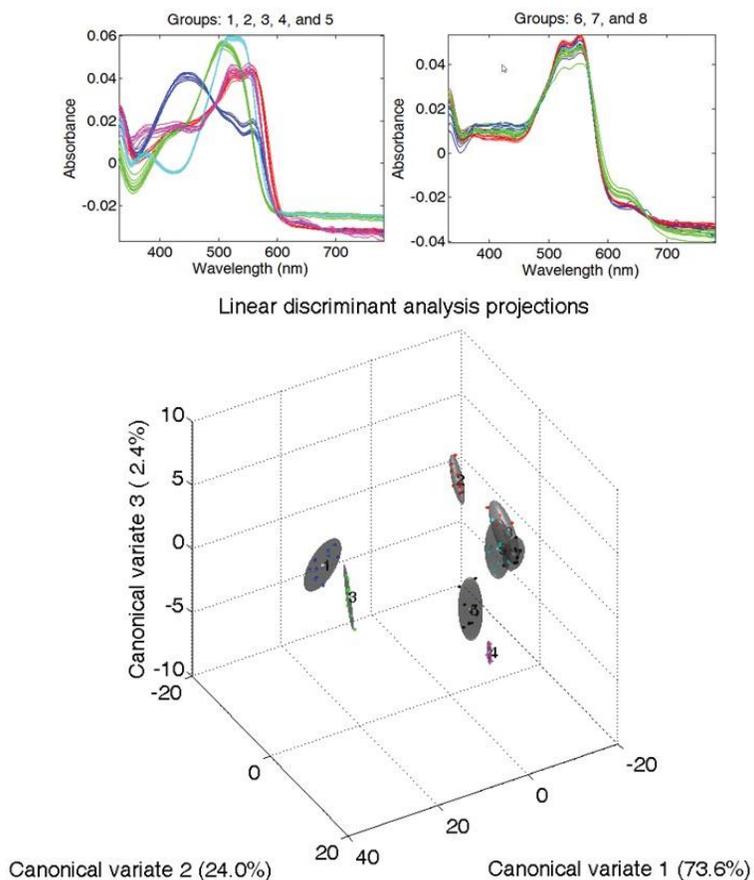


Figure 4. (top) Eighty spectra of red cotton fibers. (bottom) Linear discriminant analysis results.

2. Implications for policy and practice. The recent publication of the National Academy of Sciences report on the status of forensic science was critical of the lack of statistical reasoning and quantitative decision-making in forensic comparisons.⁴ Specifically criticized in the report was the prevalence of testimony involving trace evidence, such as hair and fibers, to be based on qualitative comparisons that were often judgment calls on the part of a forensic investigator. When a forensic examiner is called to testify in court, questions may arise in a Daubert hearing regarding the validity of comparisons. *Daubert vs. Merrill Dow Pharmaceuticals Inc.*^{66,67} established a checklist for assessing the reliability of scientific testimony, an important aspect of which is the assessment of error rates in practice. Other significant questions raised in the 2009 National Academy of Sciences report¹³ include: (a) [SWGMAT] “has produced guidelines, but no set standards, for the number and quality of characteristics that must correspond to conclude that two fibers came from the same manufacturing batch. (b) [T]here have been no studies that characterize either reliability or error rates in the procedures.” Stoney stated the case well: “Failure to use state of the art techniques for fibre identification and comparison can lead to a reduction in evidential value, as the number of potential alternative sources will rise considerably if all comparative possibilities are not exhausted.”⁶⁸ Likewise, the failure to take advantage of statistical techniques can also lead to losses in evidential value.

References

1. Ballou, S. "Wigs and the significance of one fiber." In: *Mute Witnesses: Trace Evidence Analysis*, Houck, M. M. Ed., Academic Press, San Diego, CA, 2001, pp. 21-48.
2. Deedrick, D. W. "Searching for the source: Car carpet fibres in the O.J. Simpson case," *Contact* **1998**, 26, 14-16.
3. Houck, M. M. "A case of cross-transfer." In: *Mute Witnesses: Trace Evidence Analysis*, Houck, M. M. Ed.; Academic Press: San Diego, CA, 2001, pp. 175-186.
4. Houck, M. M. "My roommate is using the refrigerator." In: *Trace Evidence Analysis: More Cases from Mute Witnesses*, Houck, M. M., Ed., Academic Press, San Diego, CA, 2003, pp. 233-250.
5. Post, H.; Hilder, D. B. "Fibers found on victims form links in Williams case," *The Atlanta Journal*, 2 February 1982, p. 1A.
6. Deadman, H. "Fiber evidence and the Wayne Williams trial." *Law Enforcement Bulletin*; U.S. Government Document J1.14/8a:F44, Federal Bureau of Investigation, U.S. Department of Justice, FBI, March and May, 1984.
7. Scientific Working Group for Materials Analysis (SWGMA), Forensic Fiber Examination Guidelines, 1999; URL: <http://www.swgmat.org/Forensic%20Fiber%20Examination%20Guidelines.pdf>
8. Harmon, R.P.; Clarke, G.; Michaud, A. L.; Plourd, C. J. "Daubert Presentation," *Trace Evidence Symposium*; Clearwater Beach, FL, 6 August 2009; URL: <http://projects.nfstc.org/trace/2009/day4.htm>.
9. SWGMA, Fiber Evidence, "Courtroom Education and Admissibility Response" , URL: <http://SWGMA.org/>.
10. SWGMA, "Response to Fiber Analysis Question List by the Research, Testing, Development and Evaluation Interagency Working Group (RTD&E IWG)," National Science and Technology Council Subcommittee on Forensic Science; URL: <http://www.swgmat.org/fiber.htm>.
11. *Fiber Organon*, Fiber Economics Bureau, Washington, DC, URL: <http://www.fibereconomics.com/>.
12. Houck, M. M. "Statistics and trace evidence: the tyranny of numbers," *Forensic Sci. Commun.* 1999, 1(3); URL: <http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/oct1999/houck.htm>.
13. National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press, Washington, DC, 2009.
14. Moore, D. S.; McCabe, G. P. *Introduction to the Practice of Statistics*, 2nd ed., W. H. Freeman, New York, 1993, 294-295.
15. Eyring, M. B.; Gaudette, B. D. "An introduction to the forensic aspects of textile fiber examination," in *Forensic Science Handbook*, Volume II, 2nd ed., Saferstein, R. D. (Ed.), Pearson Education, Inc., Upper Saddle River, NJ, 2005, p. 233.
16. Grieve, M.C. "A survey on the evidential value of fibres and on the interpretation of the findings in fibre transfer cases. Part 1-fibre frequencies," *Science & Justice* **2000**, 40, 189-200.
17. Grieve, M.C. "A survey on the evidential value of fibres and on the interpretation of the findings in fibre transfer cases. Part 2 - interpretation and reporting," *Science & Justice* **2000**, 40, 201-209.

18. Eyring, M. B.; Gaudette, B. D. "An introduction to the forensic aspects of textile fiber examination," in *Forensic Science Handbook, Volume II*, 2nd ed., Saferstein, R. D. (Ed.), Pearson Education, Inc., Upper Saddle River, NJ, 2005, pp 248-254.
19. Palenik, S. "Microscopical examination of fibres," in: *Forensic Examination of Fibres*, 2nd edition, Robertson J.; Grieve, M., Eds.; CRC Press: Boca Raton, 1999; p. 153-177.
20. Houck, M. M. Forensic Fiber Examination and Analysis. *Forens. Sci. Rev.* 2005, 17(1), 29-49.
21. Houck, M. M. "The forensic identification of textile fibers." In: *Identification of textile fibers*; Houck, M. M., Ed.; Woodhead Publishing: Cambridge, U.K., 2009.
22. Rendle, D.F.; Wiggins, K.G. "Forensic analysis of textile fibre dyes," *Review of Progress in Coloration and Related Topics*; **1995**; 25, 29-34.
23. Eyring, M. B. "Visible Microscopical Spectrophotometry in the Forensic Sciences." In: *Handbook of Forensic Science*, Vol. I, 2nd ed.; R. Saferstein, Ed.; Prentice Hall: Upper Saddle River, NJ, 2002; pp. 321-387.
24. Tungol, M. W.; Bartick, E. G.; Montaser, A. "Forensic Examination of Synthetic Fibers by Microscopic Infrared Spectroscopy," In: *Practical Guide to Infrared Spectroscopy*; H. J. Humecki, Ed., Marcel Dekker: New York, 1995; pp. 245-285.
25. Tippett, C. F.; Emerson, V. J.; Fereday, M. J.; Lawton, F.; Jones, L. T.; Lampert, S.M. "The Evidential Value of the Comparison of Paint Flakes from Sources other than Vehicles," *J. Forensic Sci. Soc.* **1993**, 8, 61-65 (1968).
26. Jones, D. A., Blood Samples: Probability of Discrimination. *J. Forensic Science Soc.* **1972**, 12, 355-359.
27. Smalldon, K. W.; Moffat A. C. *J. Forensic Science Society*, 1973, 13, 291-295 (1973).
28. Aitken, C. G. G.; Stoney, D. A., *The Use of Statistics in Forensic Science*, Ellis Horwood: New York, 1991.
29. Roux, C.; Novotny, M.; Evans. I.; Lennard, C. "A study to investigate the evidential value of blue and black ballpoint pen inks in Australia," *Forensic Sci. Int.* **1999**, 101, 167-176.
30. Grieve, M. C.; Biermann, T.W.; Schaub, K. "The individuality of fibres used to provide forensic evidence – not all blue polyesters are the same," *Sci. Justice*, 2005, 45(1), 13-28.
31. Almer, J.; McAnsh, E.; Doupe, B. "Forensic Fibre Analysis by UV-Visible Microspectrophotometry," *Can. Soc. of Forensic Sci. J.* **2010**, 43(1), 16-30.
32. Karlsson, T. Multivariate analysis ('Forensiometrics')-a new tool in forensic medicine: Differentiation between sharp force homicide and suicide, *Forensic Sci. Int.* 1998, 94, 183-200.
33. Karlsson, T. Multivariate analysis ('forensiometrics')-a new tool in forensic medicine. Differentiation between firearm-related homicides and suicides. *Forensic Sci. Int.* **1999**, 101, 33-41.
34. Karlsson, T. Multivariate analysis ('forensiometrics')-a new tool in forensic medicine. Findings on the victim of sharp-force homicide can predict the inter-relationship with the perpetrator. *Forensic Sci. Int.* **1999**, 101, 33-41.
35. Kochanowski, B. K.; Morgan, S. L. "Forensic discrimination of automotive paint samples using pyrolysis-gas chromatograph/mass spectrometry with multivariate statistics," *J. Chromatogr. Sci.* **2000**, 38(3), 100-108.
36. Egan, W. J.; Morgan, S. L.; Bartick, E. G.; Merrill, R, A.; Taylor, H. J. "Forensic Discrimination of Photocopy and Printer Toners. II. Discriminant Analysis Applied to

- Infrared Reflection-Absorption Spectroscopy", *J. Anal. Bioanal. Chem.* **2003**, 376, 1279-1285.
37. Egan, W. J.; Galipo, R. C.; Kochanowski, B. K.; Morgan, S. L.; Bartick, E. G.; Miller, M. L. Ward, D. C.; Mothershead II, R. F. "Forensic Discrimination of Photocopy and Printer Toners. III. Multivariate Statistics Applied to Scanning Electron Microscopy and Pyrolysis Gas Chromatography/Mass Spectrometry," *J. Anal. Bioanal. Chem.* **2003**, 376, 1286-1297.
 38. Hida M.; Sato H.; Sugawara H.; Mitsui T. Classification of counterfeit coins using multivariate analysis with X-ray diffraction and X-ray fluorescence methods. *Forensic. Sci. Int.* **2001**, 115, 129-134.
 39. Thanasoulas, N. C.; Piliouris, E. T.; Kotti, M. S.; Evmiridas, N. P. Application of multivariate chemometrics in forensic soil discrimination based on the UV-Vis spectrum of the acid fraction of humus. *Forensic. Sci. Int.* **2002**, 130, 73-82.
 40. Hida, M.; Mitsui, T. Classification of prepaid cards based on multivariate treatment of data obtained by X-ray fluorescence analysis. *Forensic. Sci. Int.* **2001**, 119, 305-309.
 41. Rajer-Kanduk, K.; Zupan, J.; Majcen, N. *Chemom. Intell. Lab. Sys.* **2003**, 65, 221-229.
 42. Introna, F.; Jr.; Vella, G. D.; Campobasso, C.P. *Forensic Sci. Int.* **1998**, 95, 39-45.
 43. Thanasoulas, N.C.; Parisi, N.A.; Evmiridas, N.P. *Forensic Sci. Int.* **2003**, 138, 75-84.
 44. Enlow, E. M.; Kennedy, J. L.; Nieuwland, A. A.; Hendrix, J. E.; Morgan, S. L. Discrimination of Nylon Polymers Using Attenuated Total Reflection Mid-Infrared Spectra and Multivariate Statistical Techniques, *Applied Spectroscopy* **2005**, 59(8), 986-992.
 45. Morgan, S. L.; Hall, S. H.; Hendrix, J. E.; Bartick, E. G. "Pattern recognition methods for the classification of trace evidence textile fibers from UV/visible and fluorescence spectra," Proceedings of the FBI Trace Evidence Symposium, Clearwater, FL, 15 August 2007. [See: <http://projects.nfstc.org/trace/>].
 46. Morgan, S.L.; Bartick, E.G., "Discrimination of forensic analytical chemical data using multivariate statistics," in: *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*, Blackledge RD, Ed., John Wiley & Sons, New York, 2007; pp. 331-372.
 47. Collier, B. J.; Bide, M. J.; Tortora, P. G. *Understanding Textiles*, 7th Ed.; Pearson Prentice Hall: Upper Saddle River, NJ, 2009.
 48. Moore, D. S. "Long-term data archiving," *Anal. Bioanal. Chem.* **2010**, 396, 189-192.
 49. Abendshien, L. C.; Brown, C. J.; Williams, D. K.; Shaw, S. Forensic Automotive Carpet Fiber Identification Database (FACID): Preliminary Validation and Evaluation. *Trace Evidence Symposium*, (Poster); Clearwater, FL; Aug. 13-17, 2007.
 50. Harris, R. J. *ANOVA: An Analysis of Variance Primer*, Peacock Publishers, Itasca, IL, 1994.
 51. Riffenburgh, R. H. *Statistics in Medicine*, Academic Press, San Diego, 1993.
 52. Wellek, S. *Testing Statistical Hypotheses of Equivalence*, Chapman and Hall/CRC Press LLC, Boca Raton, FL, 2003.
 53. National Research Council of the National Academies. *Forensic Analysis: Weighing Bullet Lead Evidence*, The National Academies Press: Washington, DC, 2004.
 54. Sokal, R. R.; Sneath, P. H. A. "Principles of Numerical Taxonomy," Freeman: San Francisco, 1963.
 55. Kaufmann, L.; Rousseeuw, P. J. *Finding Groups in Data*, John Wiley & Sons, Inc.: New York, 1990.
 56. Jackson, J. E. (1991) *A User's Guide to Principal Components*, John Wiley & Sons, New York, NY.
 57. Jolliffe, I. T. (2002) *Principal Component Analysis*, 2nd ed., Springer-Verlag, New York, NY.

58. Hotelling, H. "Analysis of a complex of statistical variables into principal *components*," *J. Educ. Psychol.*, **1933**, *10*, 69-79.
59. Fisher, R.A. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* **1936**, *7*, 179-188.
60. Krzanowski, W. J. *Principles of Multivariate Analysis: A User's Perspective*, Revised ed., Oxford University Press, Inc., New York, NY, 2000.
61. Rencher, A. C. *Methods of Multivariate Analysis*, 2nd ed., John Wiley & Sons, New York, NY, 2002.
62. Huberty, C. J., Olejnik, S. *Applied MANOVA and Discriminant Analysis*, 2nd ed., John Wiley & Sons, New York, NY, 2006.
63. Hotelling, H. "The generalization of Student's t-ratio," *Annals of Math. Statist.* **1931**, *2* (3), 360-378.
64. Morgan, S. L.; Bartick, E.G.; Goodpaster, J. V.; Birt, D. L.; Burnip, M. R.; Reichard, E. J.; Roberts, K., "Chemometrics and databases for comparisons of spectral data from trace evidence," paper at SciX 2012 (sponsored by the Federation of Analytical Chemistry and Spectroscopy Societies), Kansas City, MO, Kansas City, MO, 2 October 2012.
65. Roberts, K.; Bartick, E.G.; Morgan, S. L.; Goodpaster, J. V. "A Statistical Approach Using Multivariate Analysis on Visible Spectra to Determine the Matching and Discriminating Capabilities for the Forensic Examination of Question and Known Fibers," poster presented at the Annual Meeting of the American Academy of Forensic Sciences, Atlanta, GA, 23 February 2012.
66. *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993).
67. *Reference Manual on Scientific Evidence*, Federal Judicial Center, West Publishing Company: St. Paul, MN, 1994; p 71.
68. Stoney, D. A. "Relaxation of the assumption of relevance and an application to one-trace and two-trace problem," *J Forensic Sci. Soc.* **1994**, *34*, 17-21.

Dissemination of research findings

Progress reports

Six-monthly progress reports to the National Institute of Justice are up-to-date and approved. A final six-month project report from 1 July 2013 to 31 December 2013 completes these progress reports.

Scientific meeting presentations

The following oral presentations have been made at scientific meetings and workshops during this current project, for which NIJ was acknowledged for full support. Abstracts in meeting proceedings were published for all of these presentations.

1. Stephen L. Morgan, Jessica N. McCutcheon, Megan R. Baranowski, Heather Brooke, Michael L. Myrick, "Multivariate analysis of variance for forensic trace evidence decision-making," invited paper, "Chemometrics in Forensics" symposium, Federation of Analytical Chemistry & Spectroscopy Societies, Raleigh, NC, 19 October 2010.
2. Stephen L. Morgan, "Every contact leaves a trace: Forensic analytical chemistry and CSI," invited talk for the 2011 South Carolina Chemist of the Year Award, South Carolina Section of the American Chemical Society, Claflin College, Orangeburg, SC, 20 April 2011.
3. Patrishia Shelley and Stephen L. Morgan, "Receiver operator characteristics graphs for validation of forensic decision-making," poster at the University of South Carolina Discovery Day, Undergraduate Research Symposium, Columbia, SC, 22 April 2011.
4. Stephen L. Morgan, Oscar G. Cabrices, Scott J. Hoy, and James E. Hendrix, "Forensic discrimination of dyed textile fibers using UV/visible microspectrophotometry and micro-extraction/liquid chromatography/mass spectrometry," oral paper, California Association of Criminalists, spring meeting 2011, Long Beach, CA, 18 May 2011.
5. Stephen L. Morgan, Lecturer in full-day Workshop, "Introduction to Chemometrics for Forensic Scientists and Analytical Chemists," John Jay College of Criminal Justice, New York, NY, 7-8 June 2011.
6. Stephen L. Morgan, John V. Goodpaster, Edward G. Bartick, "Statistical Methods for Forensic Decision-Making in Trace Evidence Comparisons," invited one-day workshop at the 2011 NIJ/FBI Trace Evidence Symposium, St. Louis, MO, 8 August 2011. Professor Michael Risinger of the Seton Hall University School of Law (also lawyer for the Innocence Project), who attended the workshop, referred to this workshop in his talk at the meeting: [http://projects.nfstc.org/trace/2011/videos/Day1DebatingMerits_MichaelR_309.html]:
7. Stephen L. Morgan, John V. Goodpaster, Edward G. Bartick, "Evaluation of statistical measures for fiber comparisons by interlaboratory studies," oral paper, 2011 NIJ/FBI Trace Evidence Symposium, St. Louis, MO, 10 August 2011. URL: <http://projects.nfstc.org/trace/2011/agenda.htm>.
8. Stephen L. Morgan, John V. Goodpaster, and Edward G. Bartick, "Forensic comparisons of trace evidence fibers by infrared spectroscopy and UV/visible microspectrophotometry using statistical measures," oral paper, Federation of Analytical Chemistry and Spectroscopy Societies, Annual Meeting, 4 October 2011.
9. Laura Schneider, Edward G. Bartick, Stephen L. Morgan, John V. Goodpaster, Ph.D., "Still Another Look at the Classification of Acrylic Fibers by FTIR Microscopy," Northeast Association of Forensic Scientists (NEAFS), NEAFS Annual Meeting, Newport, RI, 2 November 2011.

10. Edward G. Bartick, Kevin Roberts and Laura Schneider, "An Approach to Discrimination and Match Quality of Questioned and Known Fiber Comparisons and using Databases to Provide a Statistical Significance Estimate of Fiber Associations," Northeast Association of Forensic Scientists (NEAFS), NEAFS Annual Meeting, Newport, RI, 2 November 2011.
11. Stephen L. Morgan, Michael L. Myrick, Anthony R. Trimboli, Jessica McCutcheon, and Megan Baranowski, "Validating Forensic Comparisons with Chemometrics," invited presentation in the Gerald S. Birth Award Session, International Diffuse Reflectance Conference, Chambersburg, PA, 2 August 2012.
12. Stephen L. Morgan, Edward G. Bartick, John V. Goodpaster, David L. Birt, Molly R. Burnip, Eric J. Reichard, and Kevin Roberts, "Chemometrics and databases for comparisons of spectral data from trace evidence," invited paper at SciX 2012 (sponsored by the Federation of Analytical Chemistry and Spectroscopy Societies), Kansas City, MO, Kansas City, MO, 2 October 2012.
13. Eric J. Reichard, John V. Goodpaster, Stephen L. Morgan, and Edward G. Bartick, "Differentiation of yellow polyester fibers with different dye uptakes using microspectrophotometry and chemometrics," paper at the 65th Annual Meeting of the American Academy of Sciences, Washington, DC, 18-23 February 2013.
14. Stephen L. Morgan, David L. Birt, Edward G. Bartick, and John V. Goodpaster, "Statistical measures for comparisons of fiber spectra: forensic database and statistical software," poster at the 65th Annual Meeting of the American Academy of Sciences, Washington, DC, 18-23 February 2013.
15. Edward G. Bartick, Kevin Roberts, Stephen L. Morgan, and John V. Goodpaster, "A Statistical approach to discrimination and match capability to provide scientific basis for estimating significance of fiber associations in forensic practice," oral paper at the 65th Annual Meeting of the American Academy of Sciences, Washington, DC, 18-23 February 2013.
16. Nathan Fuenffinger and Stephen L. Morgan, "Forensic Discrimination of Cotton Fibers by Derivative Preprocessing of UV/visible Spectra and Multivariate Statistics," paper 2270-8P. poster at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Philadelphia, PA, 6 March 2014.

Presentations at other venues (universities, high schools, organizations, etc.)

1. Stephen L. Morgan, "Avoiding univariate thinking in a multivariate world: forensic analytical chemistry perspectives" invited seminar at the Naval Research Laboratory, Washington, DC, 20 January 2011.
2. Stephen L. Morgan, "Forensic analytical chemistry: From research to CSI," invited keynote lecture, Faculty/Student Research Conference, Mercer University, Atlanta, GA, 16 April 2011.
3. Stephen L. Morgan, "Every contact leaves a trace: Forensic analytical chemistry and CSI" invited talk at the Gamma Sigma Chemistry Club, spring induction ceremonies, Catawba College, Salisbury, NC, 27 April 2011.
4. Stephen L. Morgan, "Every contact leaves a trace: Forensic analytical chemistry and CSI," invited talk at Strom Thurmond High School, Johnston, SC, 25 May 2011.
5. Stephen L. Morgan, "Forensic trace evidence research: fiber discrimination and blood detection," Invited seminar given to the Department of Chemistry & Biochemistry, Florida International University, Miami, FL, 14 October 2011.

6. Stephen L. Morgan, Michael L. Myrick, Eric M. Breitung, "Analytical chemistry research for forensic and cultural heritage decision-making," invited seminar, Department of Chemistry, University of South Carolina, Columbia, SC 18 January 2013.
7. Stephen L. Morgan, "Analytical chemistry for forensic trace analysis and crime scene blood imaging," invited seminar at the University of Mississippi, Oxford, MS, 28 March 2013.
8. Stephen L. Morgan, "Analytical chemistry for forensic trace analysis of fibers and imaging blood at crime scenes," invited seminar at the University of Albany (SUNY), Albany, NY, 30 April 2013.

Publications in print

None at this time.

Publications accepted or submitted for publication

None at this time, the manuscripts presented in the Technical Report section are being prepared for submission.

Manuscripts in preparation for publication

- (a) Eric J. Reichard, John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan, Microspectrophotometric Analysis of Yellow Polyester Fiber Dye Loadings with Utilization of Chemometric Techniques.
- (b) Edward G. Bartick, Kevin Roberts, Laura Schneider, Stephen L. Morgan, and John V. Goodpaster, A Statistical Basis for Significance Evaluation of Fibers as Class Evidence.
- (c) Nathan C. Fuenffinger, John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan, Comparison of Multivariate Preprocessing Techniques for the Forensic Discrimination of Cotton Fibers by UV/visible Microspectrophotometry.
- (d) Nathan C. Fuenffinger, Jessica N. McCutcheon, John V. Goodpaster, Edward G. Bartick, and Stephen L. Morgan, Multivariate Discrimination of Dyed Textile Fibers from UV-Visible and Fluorescence Spectra.
- (e) Nathan C. Fuenffinger, Eric J. Reichardt, Edward G. Bartick, John G. Goodpaster, and Stephen L. Morgan, Model Transfer for Multivariate Discrimination of Textile Fibers by UV/Visible Microspectrophotometry.
- (f) Stephen L. Morgan, David L. Birt, Edward G. Bartick, and John V. Goodpaster, Design of an extensible forensic database for textile fibers.

Graduate and undergraduate research students supported by this grant

Two graduate students have been supported by this project. One of these graduate students completed a Ph.D. degree in Chemistry at USC, and another completed an M.S. degree at IUPUI.

Scott J. Hoy Ph. D., Analytical Chemistry, University of South Carolina, August 2013, "Development and Figures of Merit of Microextraction and Ultra-Performance Liquid Chromatography for Forensic Characterization of Dye Profiles on Trace Acrylic, Nylon, Polyester, and Cotton Textile Fibers." 10 August 2013. Scott worked on the project in the summer of his first year in graduate school to organize the USC fiber collection and to build a computer to house data base software. Scott is presently a senior chemist at ExxonMobil, Baytown, TX.

Eric Reichard, M.S. Analytical and forensic chemistry, Indiana University-Purdue University Indianapolis, May 2013. Eric is currently at Eli Lilly, Indianapolis, IN.

Undergraduate research students who have worked on this project include:

Eric Reichard, (B.S., Chemistry, USC, 2011) worked as an undergraduate researcher for Dr. Morgan on the USC fiber database, then went to graduate school at IUPUI to work with Dr. Goodpaster for an M.S. degree (see above).

Molly R. Burnip (B. S., Chemistry, 2012). She worked as an undergraduate researcher for Dr. Morgan at SLED sampling fibers, organizing samples for our fiber database, and creating physical copies for SLED, IUPUI, and Suffolk labs, summer-fall 2011. She graduated with Phi Beta Kappa honors in Chemistry from USC. Molly is currently a second year graduate student in my research group working on UPLC analysis of dyes for her Ph. D. dissertation research.

David L. Birt (B. S., Chemistry, 2013). David has worked on data base programming in Microsoft SQL, ASP, and XML. He graduated in 2013 and now is working for a software company in Columbia, SC.

Nick Riley, B.S. Chemistry, USC, 2013. He worked as an undergraduate researcher for Dr. Morgan at SLED sampling fibers, organizing samples for our fiber database, and creating physical copies for SLED, IUPUI, and Suffolk labs, summer-fall 2011. Nick was a Fullbright Scholar and graduated with Phi Beta Kappa honors from USC.

Audrey Fennell, B.S. Chemistry, USC. She worked as an undergraduate researcher at SLED organizing samples for our fiber database and creating physical copies for SLED, IUPUI, and Suffolk labs, summer-fall 2011.

Kevin Roberts, B.S. forensic chemistry, Suffolk University. Kevin worked 6 months on UV/visible MSP of fibers with Dr. Bartick.

Matthew King, B.S., forensic chemistry, Suffolk University. Matthew worked 6 months on UV/visible MSP of fibers with Dr. Bartick.

Gianna Mancuso, forensic chemistry, Suffolk University. Gianna worked part time or 6 months on UV/visible MSP of fibers with Dr. Bartick.

Laura Schneider, B.S. forensic chemistry, Suffolk University. Laura worked 6 months on UV/visible MSP of fibers with Dr. Bartick.