Discrimination of Dyed Cotton Fibers Based on UV-visible Microspectrophotometry and Multivariate Statistical Analysis

Elisa Liszewski, Cheryl Szkudlarek and John Goodpaster
Department of Chemistry and Chemical Biology
Forensic and Investigative Sciences Program
Indiana University Purdue University Indianapolis (IUPUI)
Indianapolis, IN 46202

Forensic & Investigative Sciences
IUPUI

Objectives

1. Acquire a collection of dyed cotton fibers
2. Acquire spectra from all samples using UV-visible microspectrophotometers at IUPUI and the Indiana State Police
3. Carry out multivariate statistical analysis on the resultant spectra
4. Form a dye database

Questions We Can Answer

- How many classes of spectra can be reliably discerned in a population of dyed cotton fibers that have been analyzed by microspectrophotometry?
- What general features of the spectra represent the defined groups so that an unknown spectrum could be tentatively classified?
- What regions of the spectra are most important for discriminating these groups and therefore are the most reliable regions to inspect when comparing samples?
- To what extent can an unknown sample be correctly and quantitatively assigned to its member class?
- Can spectra obtained on one instrument serve as a database for a different instrument?
- If not, does the discrimination of spectra differ significantly depending on the instrument used?
- Does the calculation of first-derivative spectra or the use of chromaticity coordinates offer any advantages when comparing samples?
- Ultimately, to what extent do real fiber samples exhibit heterogeneity and what effect does this have on declaring a known and questioned sample to be indistinguishable with respect to their absorption characteristics?

Fibers as Trace Evidence

- Why are Fibers Important?
  - Widespread in the environment
  - Transferred easily
  - Many classifications and subtypes
  - Physically and chemically differentiable

- Key Characteristics
  - Morphology (cross-section)
  - Bulk Composition
  - Color *

*This is the main discriminating feature for cotton fibers

Fiber Dyes

- What is a Dye?
  - A colored substance that is able to absorb and reflect certain visible wavelengths of light
- Thousands of dyes produced worldwide
- Classification of Dyes:
  - By method of application
  - By chemical class
  - By type of fiber to which they are applied

Classification by Application Method

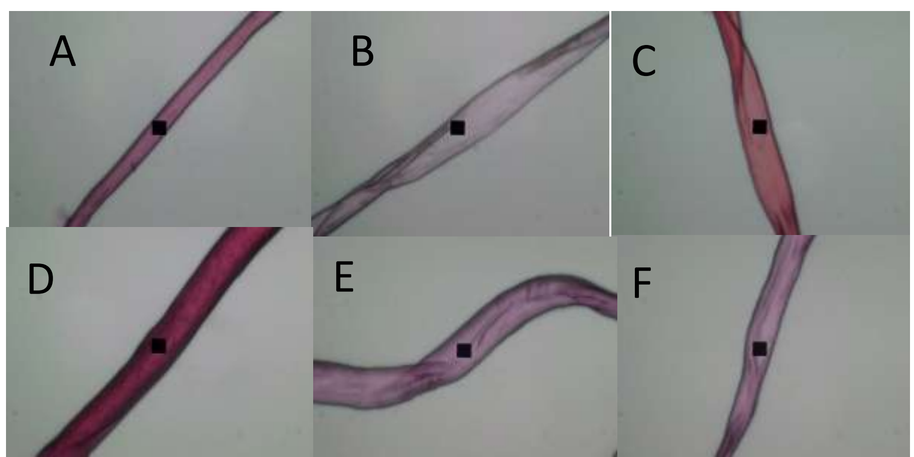
- Acid
- Azoic
- Basic
- Direct
- Dispersive
- Reactive
- Sulfur
- Vat

- Neutral or Acidic conditions forms ionic bonds
- Bonding between diazo and coupling component
- Acidic conditions forms ionic bonds
- Directly incorporated into cellulosic fiber with the presence of heat and an electrolyte
- Directly incorporated into synthetic fibers
- Form covalent bonds with the functional groups
- Require reducing agent to make dye soluble and then oxidize within the fiber to become insoluble
- Require reducing agent to make dye soluble and then oxidize within the fiber to become insoluble

The Dyes

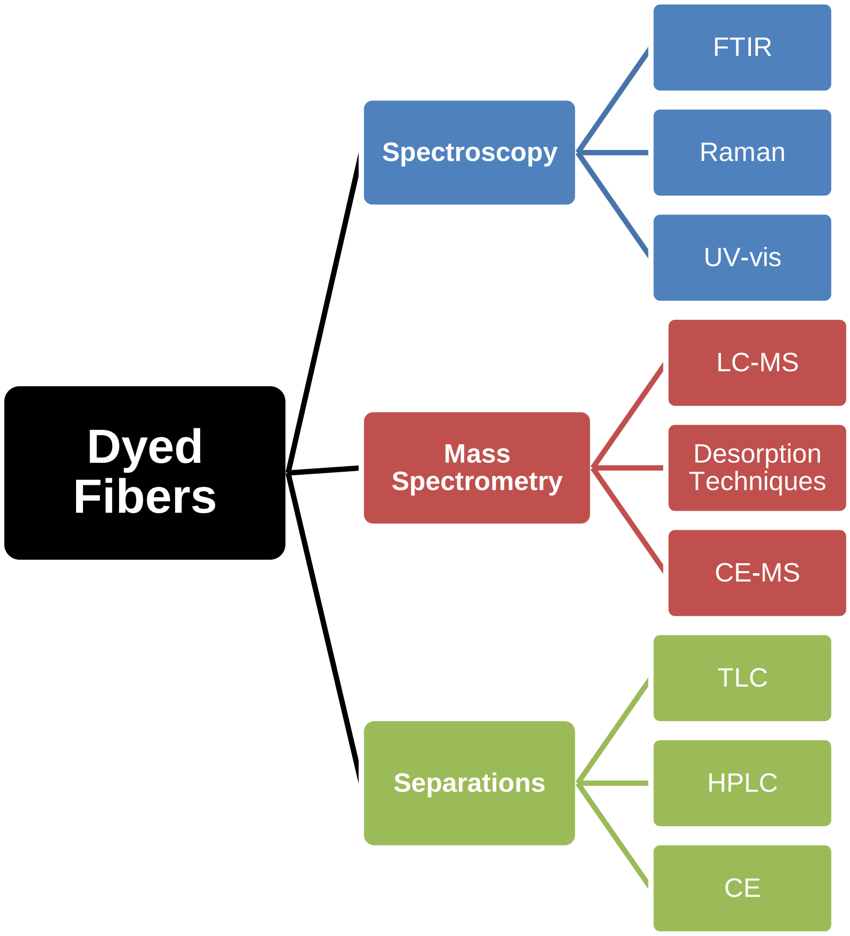
Label	Dye
A	Direct Red C-380
B	Reactive Red 120
C	Reactive Red 123
D	Reactive Red 195
E	Reactive Red 2
F	Reactive Red 228

*Provided by Testfabrics



Instrumental Analysis of Dyed Textile Fibers

- Relies upon principles of:
  - microscopy
  - spectroscopy
  - chromatography
  - mass spectrometry



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

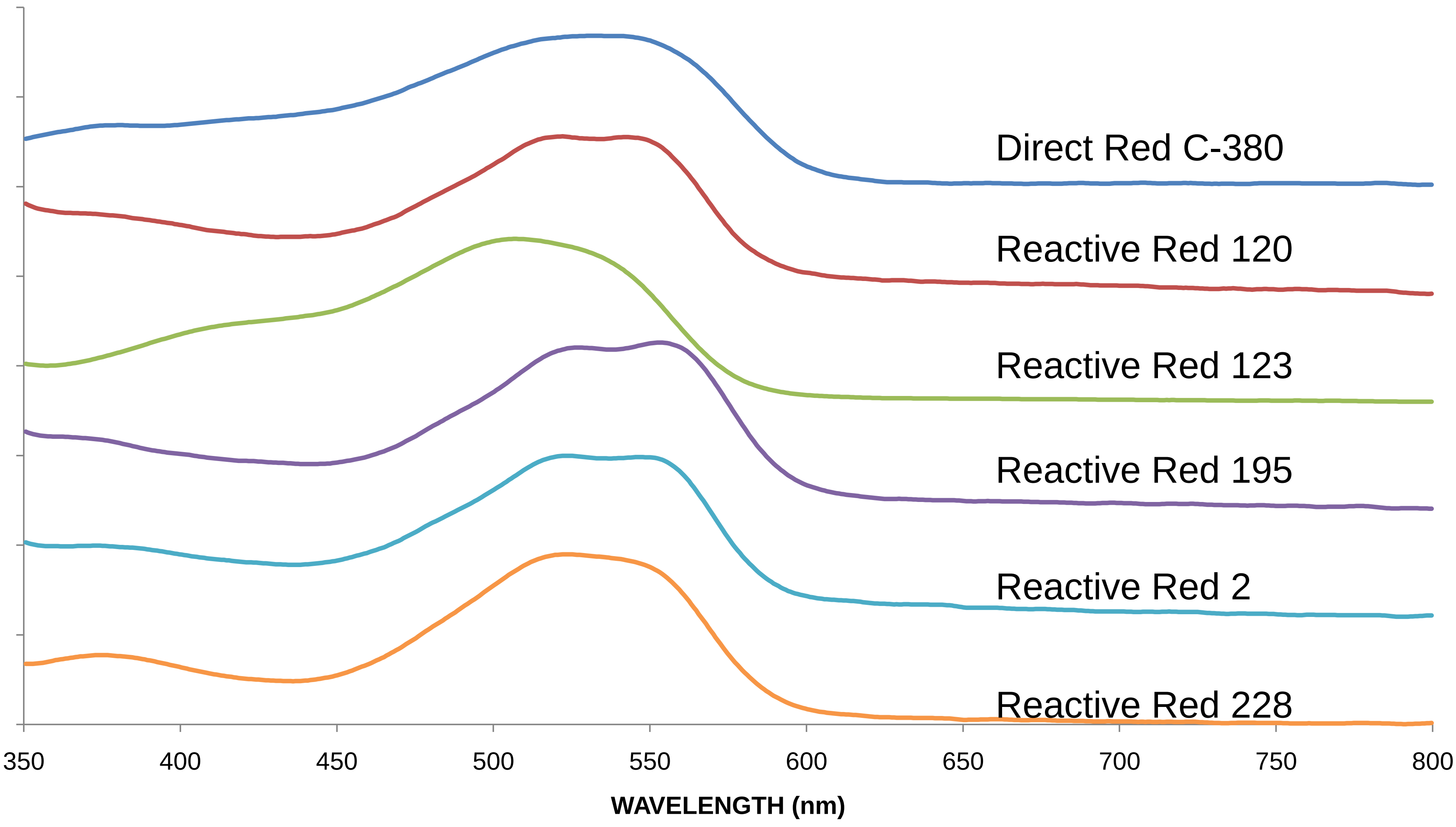
Sample Preparation

- 6 Dyed exemplars
- 10 fibers/dye
- Mounted in glycerin on glass slides
- CRAIC QDI 2000 MSP
- 35x magnification
- MSP calibrated prior to analysis
- 10 scans/fiber

Data Analysis

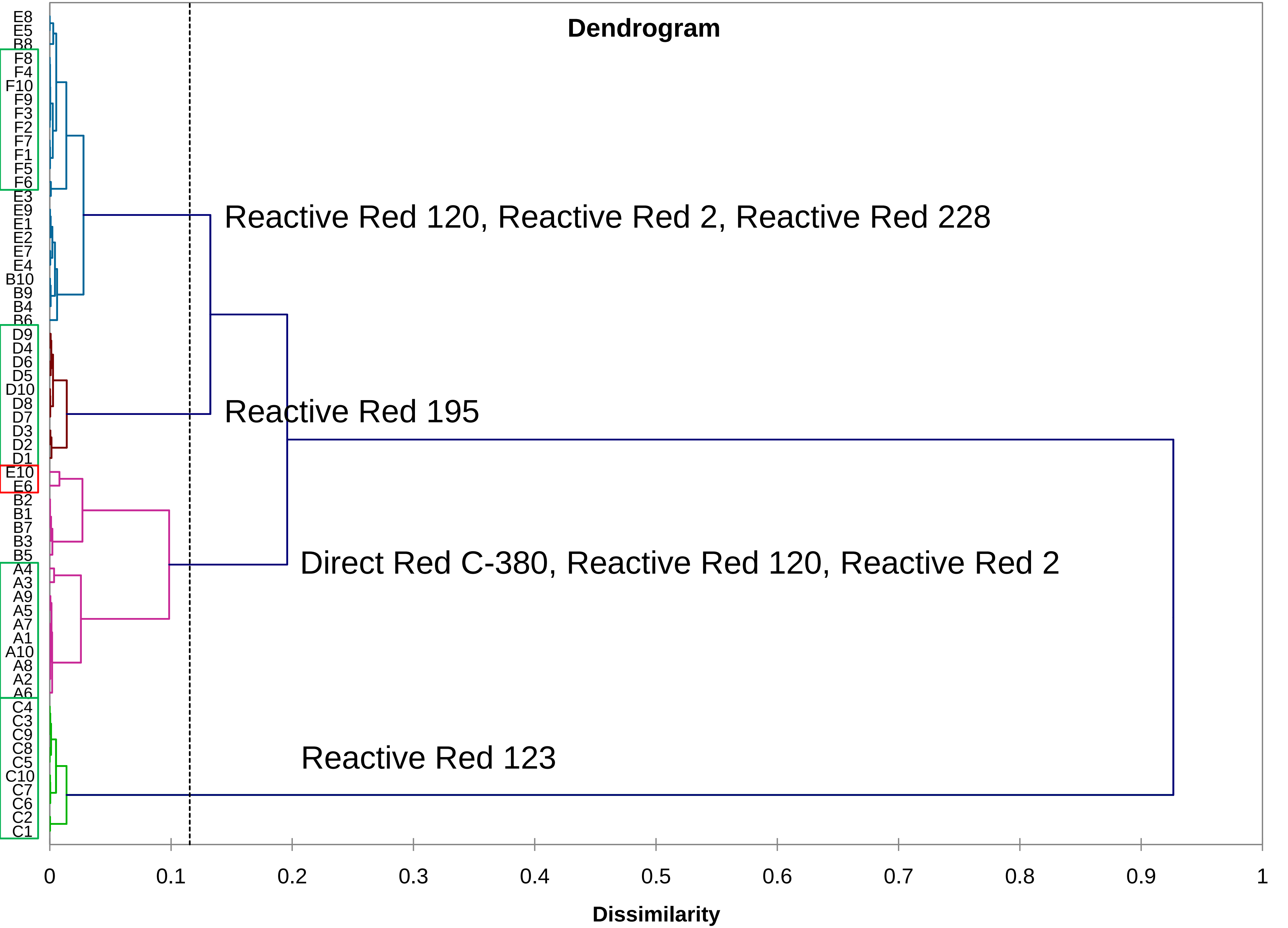
- Spectra truncated to 350-800 nm range
- Background subtracted
- Normalized
- Chemometric techniques run:
  - Agglomerative Hierarchical Clustering (AHC)
  - Principal Components Analysis (PCA)
  - Discriminant Analysis (DA)
  - Analysis of Variance (ANOVA)

Representative Spectra



Hierarchial Cluster Analysis (AHC)

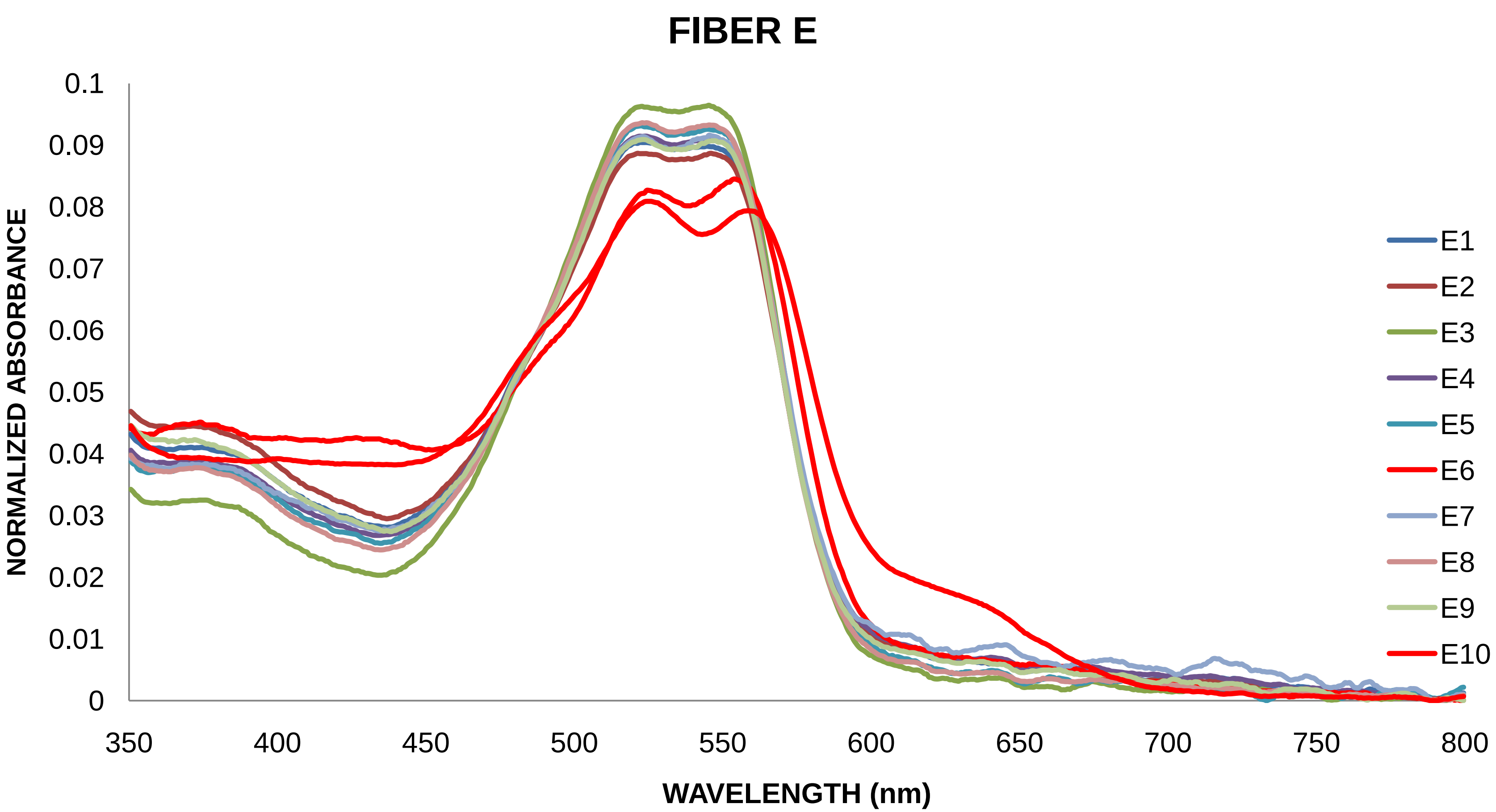
- Purpose:
  - Present data in a way that emphasize groupings
  - Detect outliers
- Procedure:
  - Calculate and Compare distances between samples
    - Small distances indicate similarity
- Result
  - Dendrogram
    - Display results according to dissimilarity
    - Truncation line
      - Groupings to the right significant


Dendrogram
Reactive Red 120, Reactive Red 2, Reactive Red 228
Reactive Red 195
Direct Red C-380, Reactive Red 120, Reactive Red 2
Reactive Red 123
E8
E5
B8
F8
F4
F10
F9
F3
F2
F7
F1
F5
F6
E3
E9
E1
E2
E7
E4
B10
B9
B4
B6
D9
D4
D6
D5
D10
D8
D7
D3
D2
D1
E10
E6
B2
B1
B7
B3
B5
A4
A3
A9
A5
A7
A1
A10
A8
A2
A6
C4
C3
C9
C8
C5
C10
C7
C6
C2
C1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Dissimilarity

Dendrogram Summary

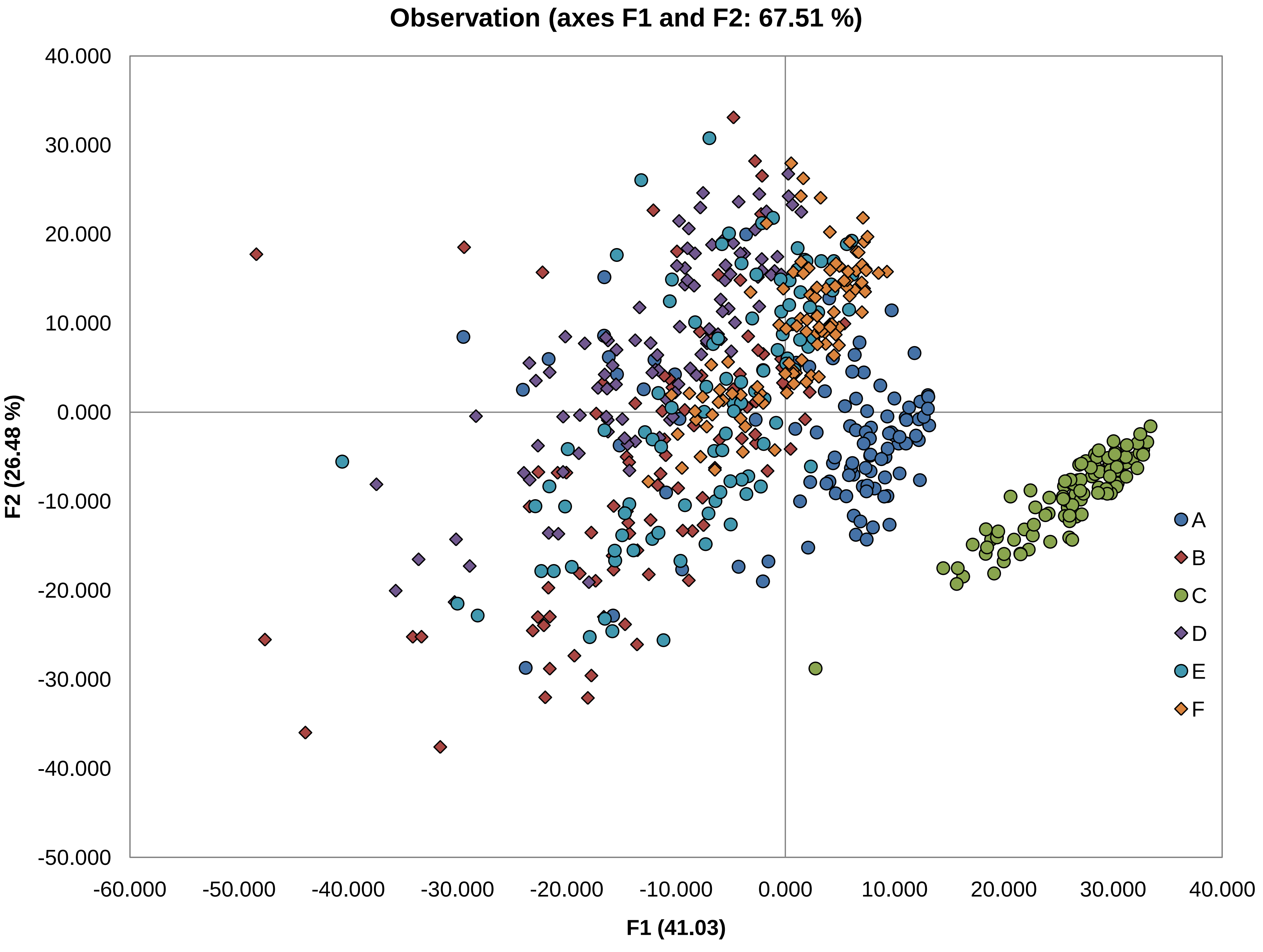
Class 1	Class 2	Class 3	Class 4
Direct Red C-380	Reactive Red 120	Reactive Red 123	Reactive Red 195
Reactive Red 120	Reactive Red 2		
Reactive Red 2	Reactive Red 228		

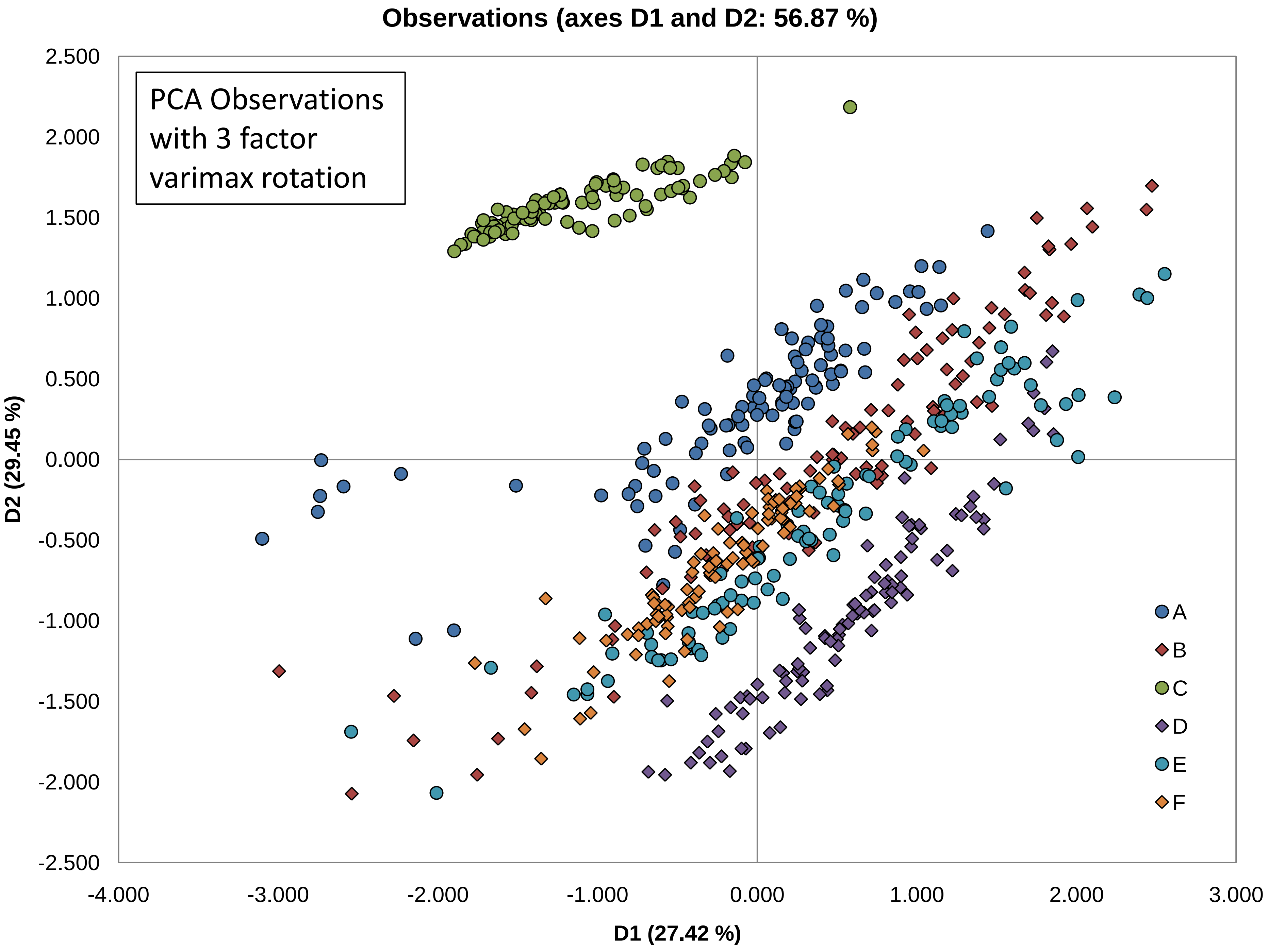
Unique Fiber Detection



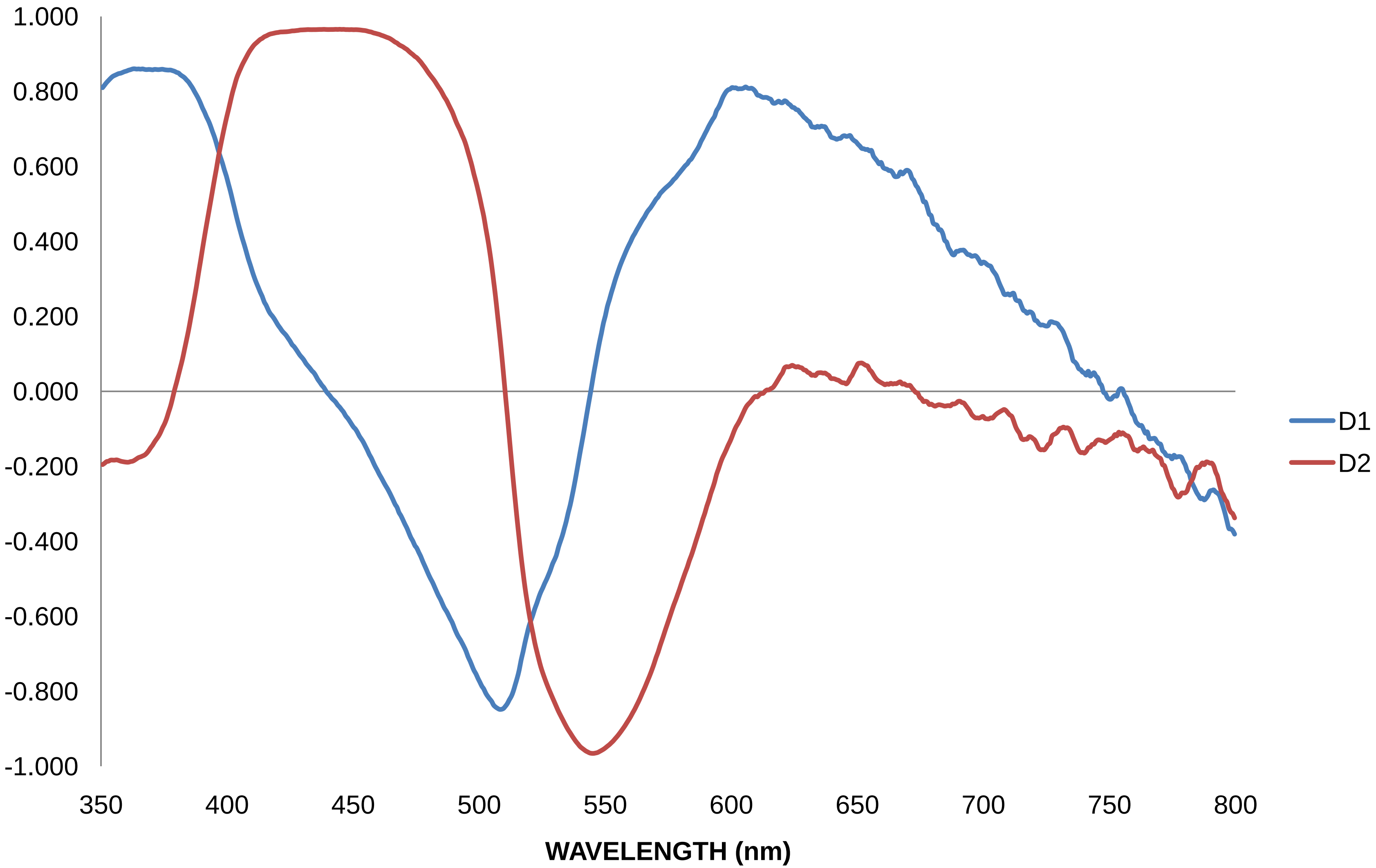
Principle Component Analysis (PCA)

- Purpose:
  - Reduce dimensionality of data
  - Detect patterns and outliers
- Procedure:
  - Form principle components
    - First pc captures the most variance
- Result:
  - Observation Plot
    - First two principle components
  - Factor Loading Plot
    - Contributions of the variables

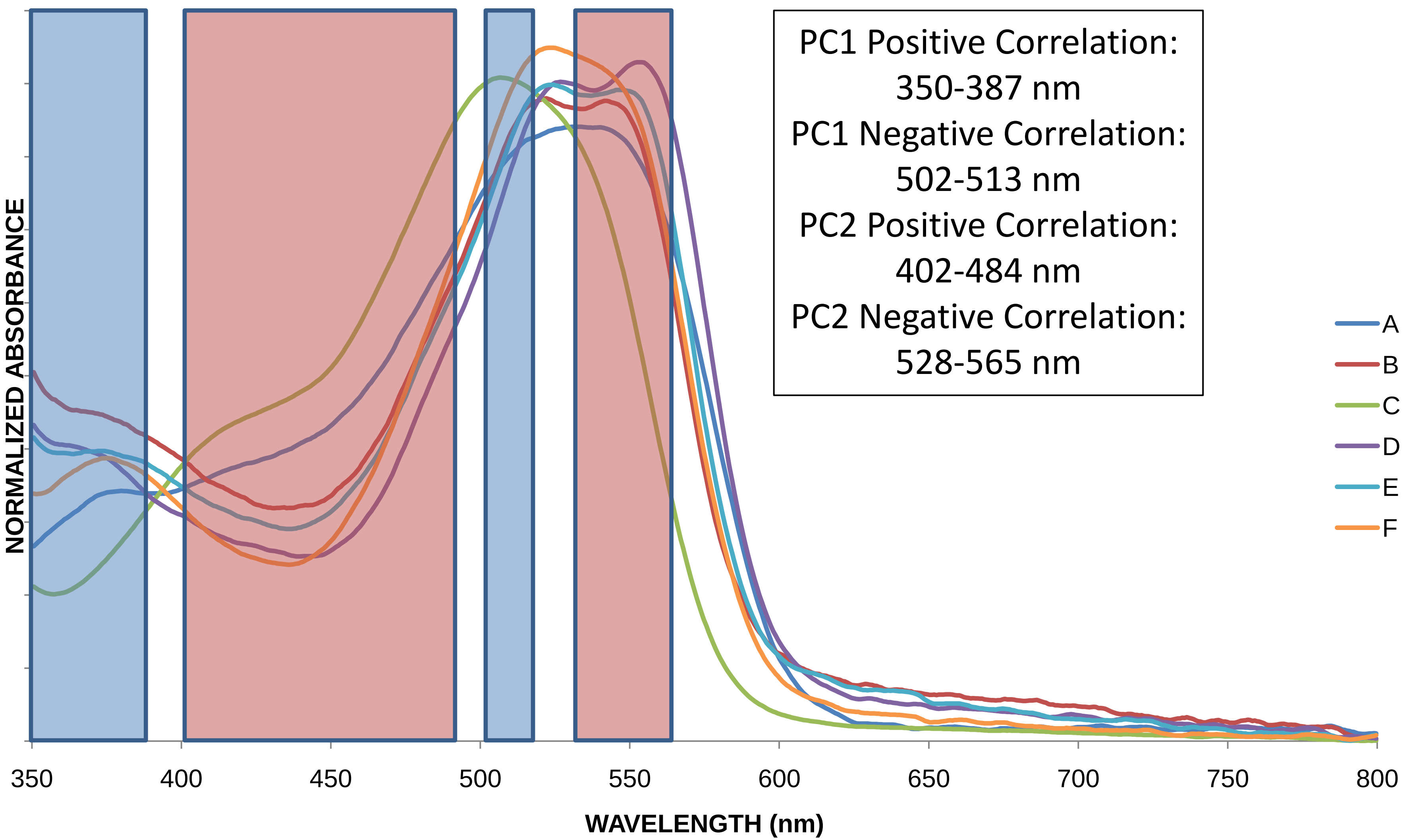

Observation (axes F1 and F2: 67.51 %)
40.000
30.000
20.000
10.000
0.000
-10.000
-20.000
-30.000
-40.000
-50.000
F2 (26.48 %)
-60.000
-50.000
-40.000
-30.000
-20.000
-10.000
0.000
10.000
20.000
30.000
40.000
F1 (41.03)
A
B
C
D
E
F


Observations (axes D1 and D2: 56.87 %)
PCA Observations
with 3 factor
varimax rotation
2.500
2.000
1.500
1.000
0.500
0.000
-0.500
-1.000
-1.500
-2.000
-2.500
D2 (29.45 %)
-4.000
-3.000
-2.000
-1.000
0.000
1.000
2.000
3.000
D1 (27.42 %)
A
B
C
D
E
F

Factor Loading

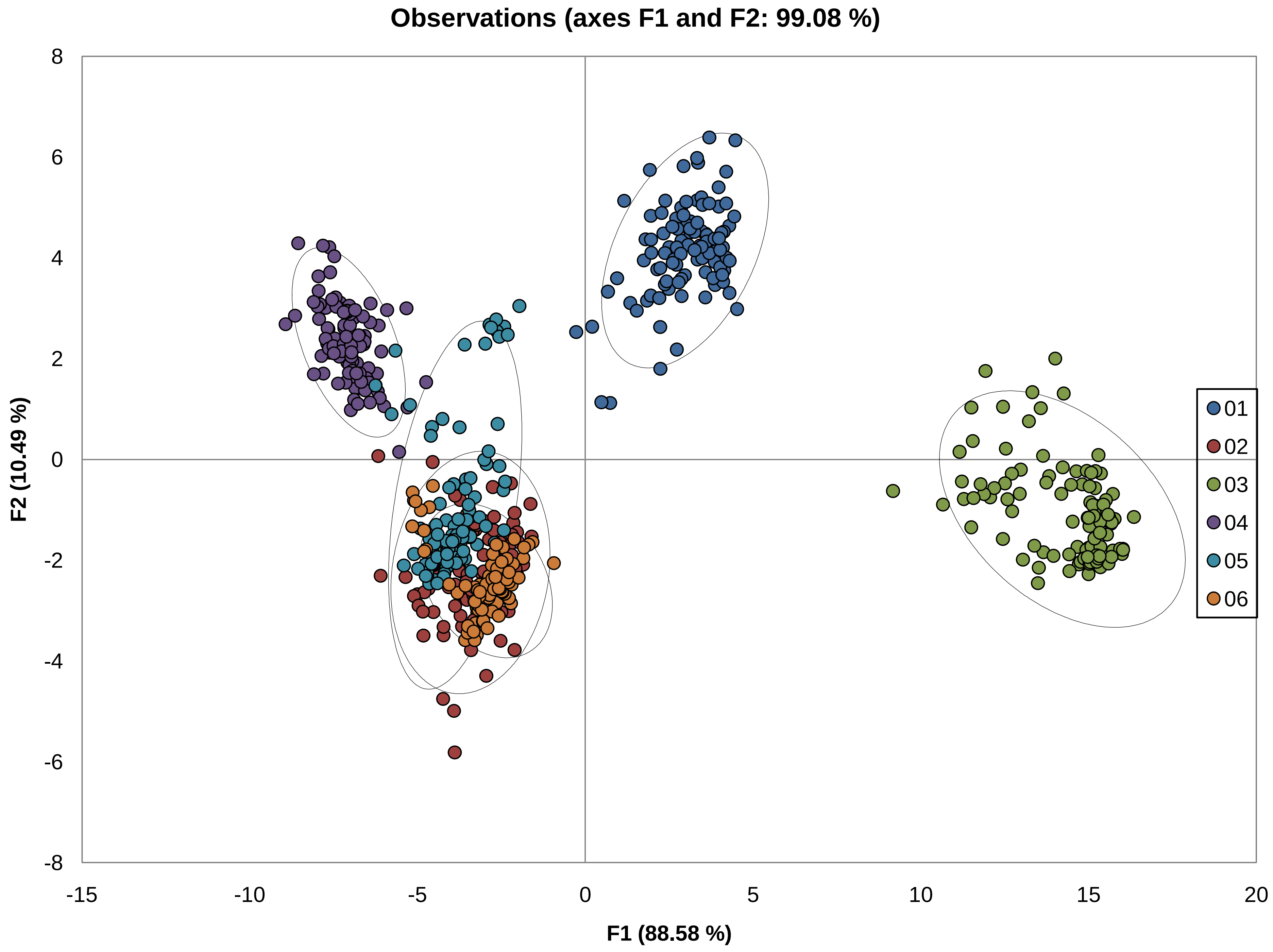


Most Informative Regions of Spectra



Linear Discriminant Analysis (DA)

- Purpose:
  - Predict group membership
  - Detect patterns
- Procedure:
  - Form canonical variates
  - Assign to class with highest probability
- Result:
  - Observations Plot
    - First two canonical variates
  - Confusion Matrix
    - Summary of the reclassification



Observations (axes F1 and F2: 99.08 %)
F2 (10.49 %)
F1 (88.58 %)
8
6
4
2
0
-2
-4
-6
-8
-15
-10
-5
0
5
10
15
20
01
02
03
04
05
06

DA Results

- The accuracy of DA can be measured by leave-one-out cross validation
- Groups listed with 100% accuracy had no errors in re-classification
- Groups with low accuracy are easily confused with other dyes

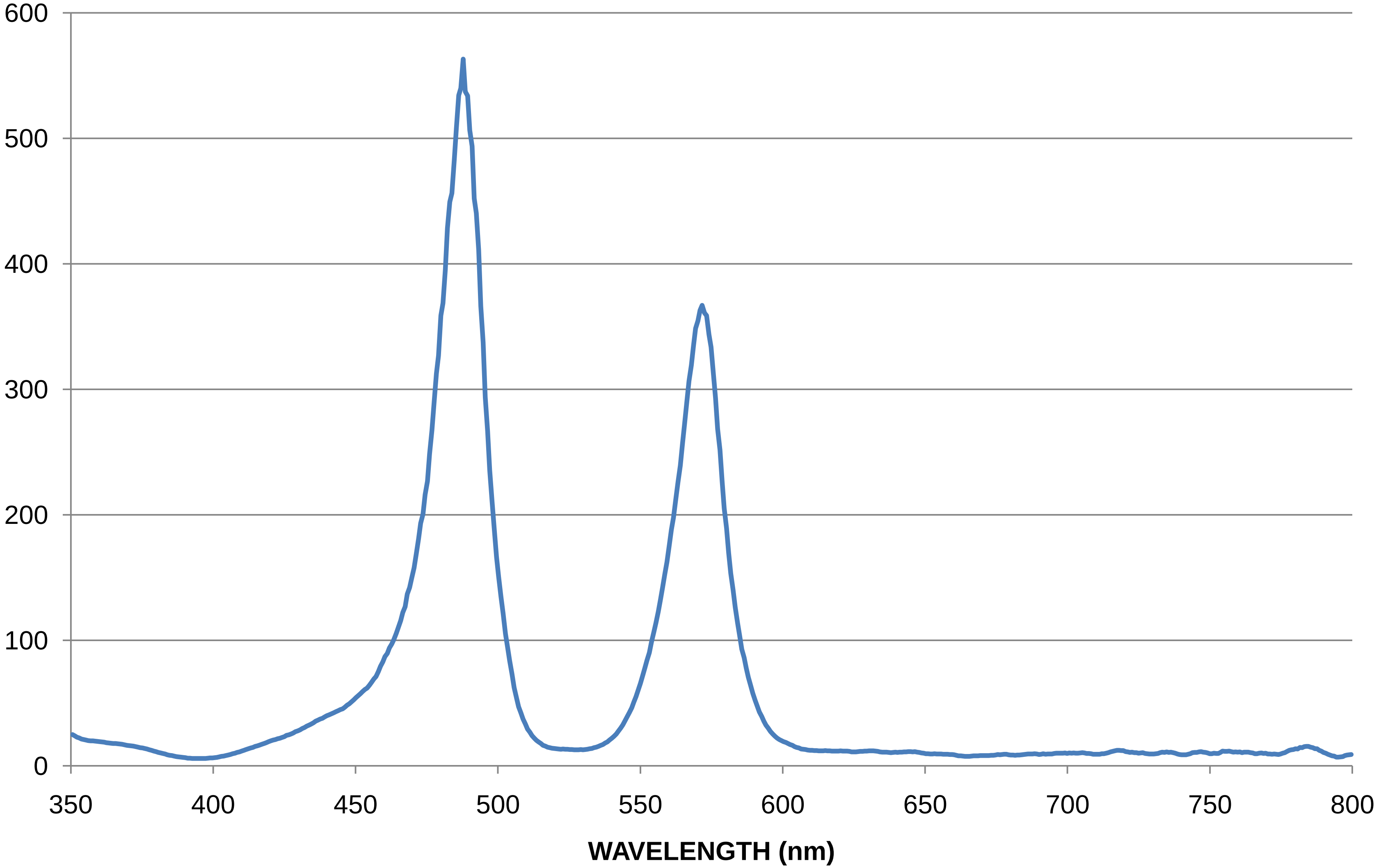
Fiber	% Correct
Direct Red C-380	100.00
Reactive Red 120	69.00
Reactive Red 123	100.00
Reactive Red 195	98.00
Reactive Red 2	59.00
Reactive Red 228	85.00
Total	**85.26**

Fiber	% Correct
Direct Red C-380	100.00
Reactive Red 120	74.23
Reactive Red 123	100.00
Reactive Red 195	100.00
Reactive Red 2	87.00
Reactive Red 228	91.00
Total	**92.13**

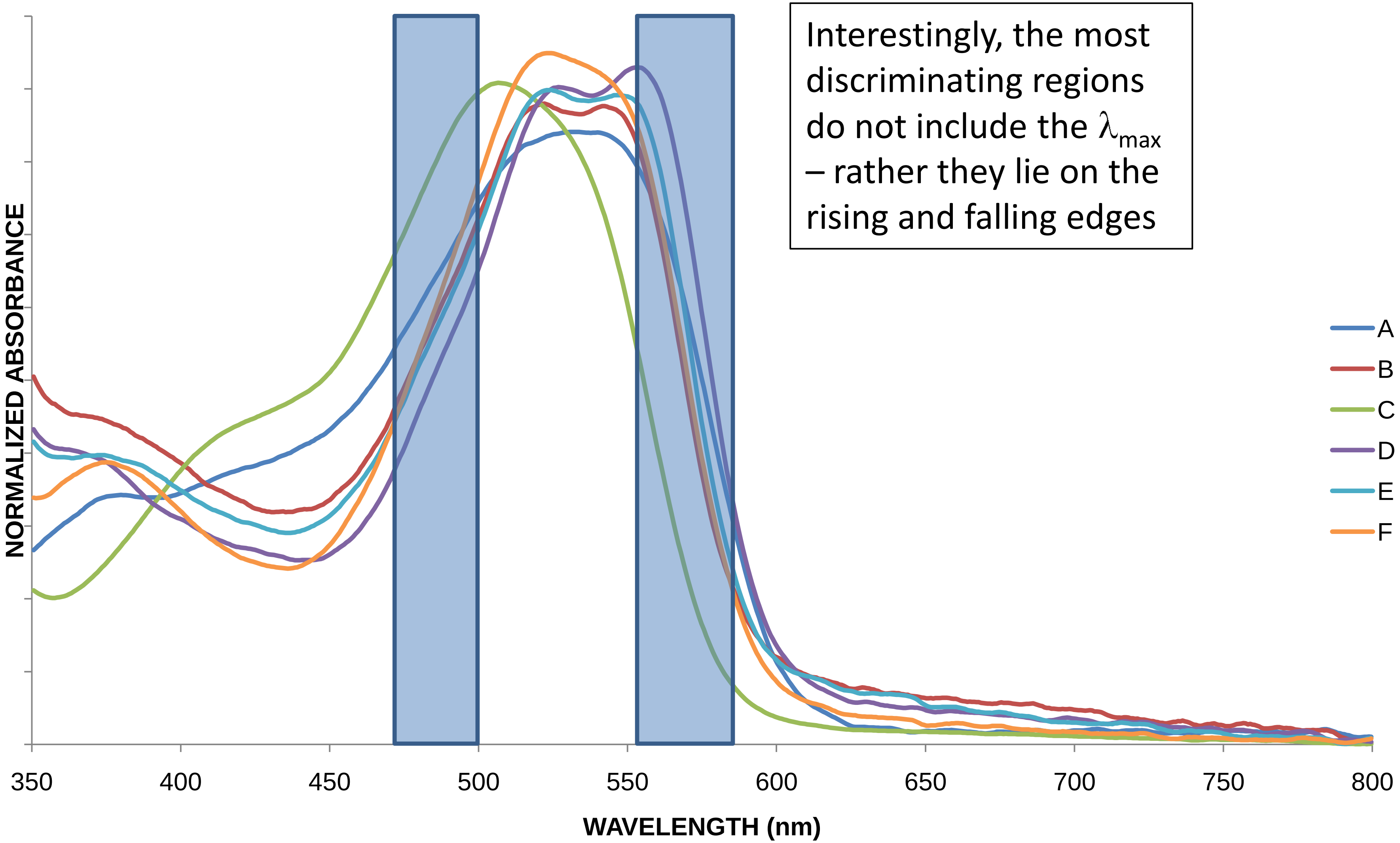
Analysis of Variance (ANOVA)

- Purpose:
  - Find parts of the spectra for maximum discrimination
- Procedure:
  - Assign F-values
    - $F = \dfrac{\text{Between Group Variance}}{\text{Within Group Variance}}$
- Result:
  - F-values
  - Fisher Ratio Plot
    - High value indicates difference

Univariate Fisher Ratios



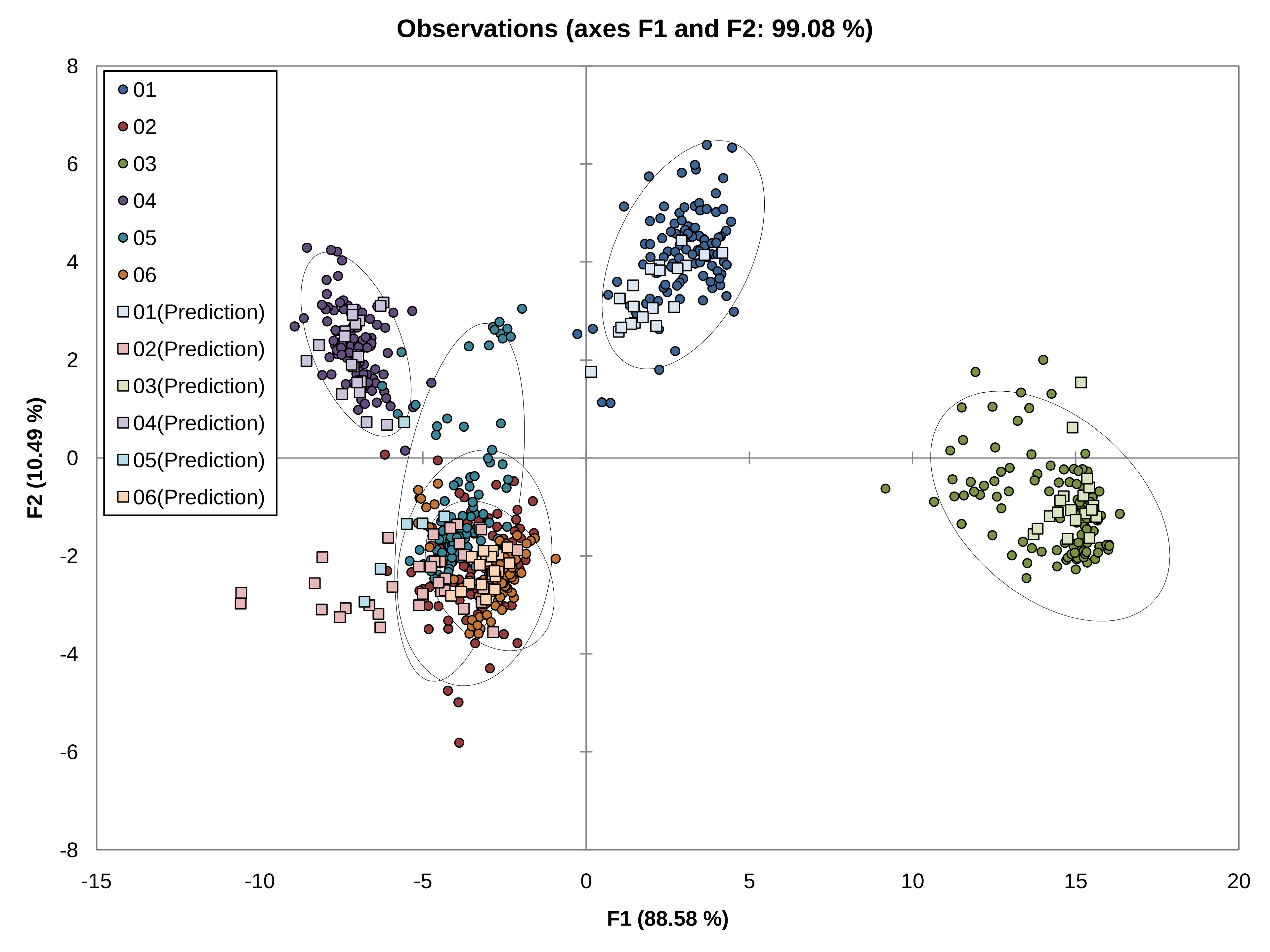
Most Discriminating Spectral Regions



Interestingly, the most discriminating regions do not include the λ_{max} – rather they lie on the rising and falling edges

The Results

EXTERNAL VALIDATION


Observations (axes F1 and F2: 99.08 %)
01
02
03
04
05
06
01(Prediction)
02(Prediction)
03(Prediction)
04(Prediction)
05(Prediction)
06(Prediction)
F1 (88.58 %)
F2 (10.49 %)
8
6
4
2
0
-2
-4
-6
-8
-15
-10
-5
0
5
10
15
20

DA Results

Fiber	% Correct
Direct Red C-380_1	100.00
Direct Red C-380_2	100.00
Reactive Red 120_1	90.00
Reactive Red 120_2	90.00
Reactive Red 123_1	100.00
Reactive Red 123_2	100.00
Reactive Red 195_1	100.00
Reactive Red 195_2	90.00
Reactive Red 2_1	10.00
Reactive Red 2_2	30.00
Reactive Red 228_1	90.00
Reactive Red 228_2	90.00
TOTAL	82.50

* 6 Class Assignment

Fiber	% Correct
Direct Red C-380_1	100.00
Direct Red C-380_2	100.00
Reactive Red 120_1	70.00
Reactive Red 120_2	90.00
Reactive Red 123_1	100.00
Reactive Red 123_2	100.00
Reactive Red 195_1	100.00
Reactive Red 195_2	100.00
Reactive Red 2_1	40.00
Reactive Red 2_2	10.00
Reactive Red 228_1	70.00
Reactive Red 228_2	90.00
TOTAL	80.83

* 60 Class Assignment

IUPUI versus ISP

INTER-LABORATORY STUDY

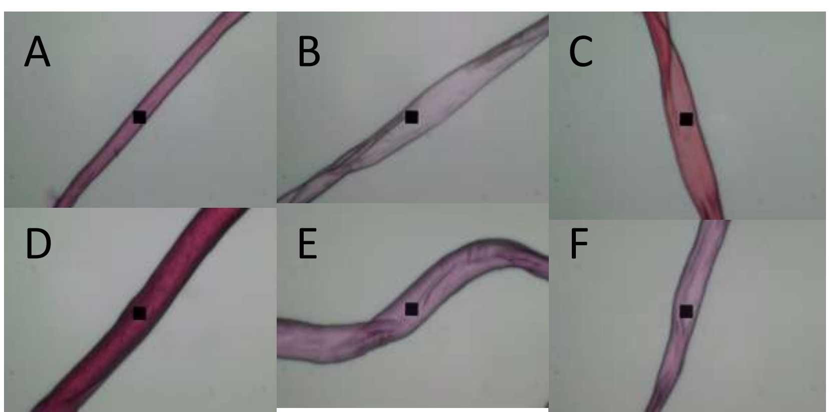
The Dyes

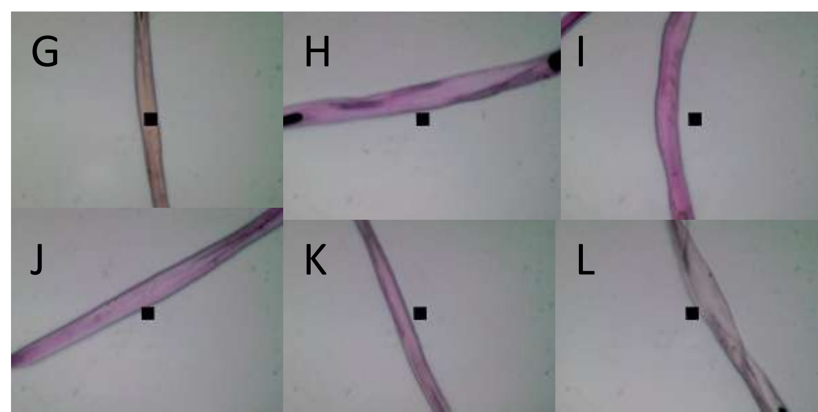
Label	Dye
A	Direct Red C-380
B	Reactive Red 120
C	Reactive Red 123
D	Reactive Red 195
E	Reactive Red 2
F	Reactive Red 228

*Provided by Testfabrics

Label	Dye
G	Direct Red 84
H	Reactive Red 180
I	Reactive Red 198
J	Reactive Red 239/241
K	Vat Red 10
L	Vat Red 15

*Provided by Dr. Stephen Morgan from University of South Carolina





Dye	% Correct
Direct Red C-380	100.00
Reactive Red 120	67.00
Reactive Red 123	100.00
Reactive Red 195	72.00
Reactive Red 2	50.00
Reactive Red 228	83.00
Direct Red 84	100.00
Reactive Red 180	89.00
Reactive Red 198	92.00
Reactive Red 239/241	67.00
Vat Red 10	100.00
Vat Red 15	100.00
Total	85.00

Dye	% Correct
Direct Red C-380	92.00
Reactive Red 120	61.00
Reactive Red 123	100.00
Reactive Red 195	93.00
Reactive Red 2	39.00
Reactive Red 228	77.00
Direct Red 84	99.00
Reactive Red 180	94.00
Reactive Red 198	70.00
Reactive Red 239/241	52.00
Vat Red 10	99.00
Vat Red 15	100.00
Total	81.25

Conclusions

- Training Set:
  - 4 main classes
  - 67.51% total variance captured by first 2 PCs
  - Direct Red C-380, Reactive Red 123, and Reactive Red 195 were correctly classified 100%
  - Overall accuracy of classification was above 90%
  - 463-502 nm and 554-585 nm were the most discriminating regions
  - Some “uniqueness was seen within the 10 fibers of Reactive Red 2

More Conclusions

- External Validation:
  - The overall classification accuracy was above 80%
- Inter-laboratory Study:
  - Five dyes were readily distinguished using instruments at IUPUI and ISP:
    - Direct Red C-380, Reactive Red 123, Direct Red 84, Vat Red 10, and Vat Red 15
  - The remaining dyes were not as readily distinguished and potentially confused with one another – the degree of confusion varied between laboratories
  - Overall, consistency was shown between the two instruments

Acknowledgements

- Microanalysis Unit of the Indiana State Police Laboratory
- Dr. Stephen Morgan (University of South Carolina)
- Tom Klass (Testfabrics, Inc.)
- Midwest Forensics Resource Center (MFRC)

INDIANA STATE POLICE

Testfabrics, Inc.

MFRC Midwest Forensics Resource Center