Assigning UNKNOWN FINGERNAIL POLISH to KNOWN MANUFACTURERS by Raman SPECTROSCOPY and MULTIVARIATE STATISTICS

KELSEY HARTT and GARY H. NAISBITT, Ph.D., Criminal Justice Department, Forensic Science Program, Utah Valley University, Orem, Utah

ABSTRACT

Using the R Project statistical platform[1], Raman spectra were compared by Principle Component Analysis and multivariate hierarchal clustering functions to associate unknown colored polishes to spectral clusters identified by manufacturer. A theoretical study predicted clear polishes were sufficiently different to cluster separately and that colored polishes by the same manufacturer would cluster with them. The theoretical study further predicted that at some point additives, such as dark pigments, resulted in greater spectral complexity that would dominate clustering and eventually exclude association to clear polishes. At that point assigning an unknown colored polish to a known manufacturer failed. These predictions were tested with clear and colored polishes from different manufactures that confirmed the predictions. Clears, creams, and light colors from the same manufacturer clustered together while reds and blues from any manufacturer tended to cluster together due to the spectral dominance of the dark pigments.

INTRODUCTION

Counterfeit products are part of industrial espionage in which proprietary research and development is stolen to produce a competing product, or by selling a non-genuine product. This report is a demonstration of associating an unknown fingernail polish to its manufactured brand by multivariate statistical analysis of Raman spectra.

MATERIALS AND METHODS

One centimeter squares of clear and colored nail polishes were obtained from local retail stores and painted on aluminum covered glass microscope slides.

Instrument setting were: 20x magnification, 100% laser power, 1 second exposure, 60 acquisitions and automatic cosmic ray elimination. While exciting the samples with 785nm light a Renishaw InVia Raman microscope[2] collected replicate spectra from 1724 to 614 wavenumbers in four wavenumber intervals.

A total of nine spectra were collected. Three spectra were collected from each location to assess instrument stability and three locations were sampled to assess sample homogeneity. To allow same-scale, sample-to-sample data comparison, raw data were normalized to a y-axis range of 0 to 1 and automatically baseline corrected with WiRE 3.4 instrument software[2]. The nine individual spectra were averaged and the average spectrum was used for statistical analysis.

THEORETICAL CLUSTER RESULTS

It was assumed that each manufacturer's clear polish was their fundamental formulation to which color dyes and pigments were added. Colored polishes were linked to their respective manufacturer by the similarity of their spectra to that manufacturer's clear polish. Results were visualized in dendrograms that indicated the degree of dis-similarity as the distance between clusters.

A theoretical truth table consisting of three different clear samples A, B, and C each represented by a single wavelength of the corresponding name Wave A for sample A and so forth for Wave B and Wave C was created by a 1 if the wave was present and a 0 if not present.

For each clear sample there were three colored samples named AC1, AC2, and AC3 and analogously for samples B and C. The expected results were that the clear samples would cluster separately and that colored samples would cluster with their corresponding clears polishes. To investigate how mixing wavelengths affected clustering outcomes, systematic changes were made to the truth tables and visualized in dendrograms.

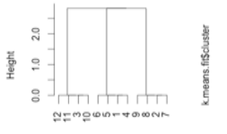
Because there were three known polishes, the results were viewed as the first three branches of the dendrogram with sub-clusters belonging to its respective branch. Selected experimental results are presented below.

EXPERIMENT 1

showed all three clears and their related colored samples grouped together as expected in their own clusters. Clear A is represented by unique wavelength Wave A and is associated with Unknown Wave A Colored Samples; AC_'s . The same pattern identifies Samples B and C.

Sample #		Wave 1	Wave 2	Wave 3
Clear Polish				
1	A	1	0	0
2	B	0	1	0
3	C	0	0	1
Unknown Colored				
4	AC1	1	0	0
5	AC2	1	0	0
6	AC3	1	0	0
7	BC1	0	1	0
8	BC2	0	1	0
9	BC3	0	1	0
10	CC1	0	0	1
11	CC2	0	0	1
12	CC3	0	0	1

CLUSTER DENDROGRAM

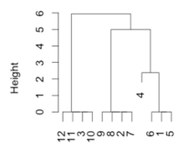


EXPERIMENT 2

showed that a colored AC sample, true for both Wave A and Wave B still clustered with clear A.

Sample #		Wave 1	Wave 2	Wave 3
Clear Polish				
1	A	1	0	0
2	B	0	1	0
3	C	0	0	1
Unknown Colored				
4	AC1	1	1	0
5	AC2	1	0	0
6	AC3	1	0	0
7	BC1	0	1	0
8	BC2	0	1	0
9	BC3	0	1	0
10	CC1	0	0	1
11	CC2	0	0	1
12	CC3	0	0	1

CLUSTER DENDROGRAM

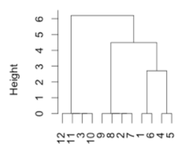


EXPERIMENT 3

showed that two colored samples true for Waves A and B clustered in together in a sub-cluster associated with cluster A.

Sample #		Wave 1	Wave 2	Wave 3
Clear Polish				
1	A	1	0	0
2	B	0	1	0
3	C	0	0	1
Unknown Colored				
4	AC1	1	1	0
5	AC2	1	1	0
6	AC3	1	0	0
7	BC1	0	1	0
8	BC2	0	1	0
9	BC3	0	1	0
10	CC1	0	0	1
11	CC2	0	0	1
12	CC3	0	0	1

CLUSTER DENDROGRAM

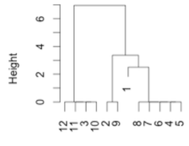


EXPERIMENT 4

showed when all colored AC's and two BC's were selected for both Wave A and Wave B, that all similar AC's and BC's clustered together, B and BC3 were the same and clustered together, while A clustered alone. This result suggests that spectral changes contributed by additives eventually dominate clustering.

Sample #		Wave 1	Wave 2	Wave 3
Clear Polish				
1	A	1	0	0
2	B	0	1	0
3	C	0	0	1
Unknown Colored				
4	AC1	1	1	0
5	AC2	1	1	0
6	AC3	1	1	0
7	BC1	1	1	0
8	BC2	1	1	0
9	BC3	0	1	0
10	CC1	0	0	1
11	CC2	0	0	1
12	CC3	0	0	1

CLUSTER DENDROGRAM

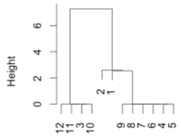


EXPERIMENT 5

supported the conclusions of Theory Experiment 4 that when BC3 was true for both Waves A and B it clustered with the AC-BC cluster. This suggest that at some point the similarity of the additives will dominate the less complex clear spectra. At that point the colored samples cluster together and exclude the clears. All C samples remained the same and clustered together as expected.

Sample #		Wave 1	Wave 2	Wave 3
Clear Polish				
1	A	1	0	0
2	B	0	1	0
3	C	0	0	1
Unknown Colored				
4	AC1	1	1	0
5	AC2	1	1	0
6	AC3	1	1	0
7	BC1	1	1	0
8	BC2	1	1	0
9	BC3	1	1	0
10	CC1	0	0	1
11	CC2	0	0	1
12	CC3	0	0	1

CLUSTER DENDROGRAM



ANALYTICAL CLUSTER RESULTS

The numerical values for the wavelengths of all except the most minor peaks in every clear spectrum were entered into a spreadsheet to create a matrix in which each clear polish was represented by unique wavelengths that were not present in other clear samples. The presence of these unique wavelengths in an unknown colored sample associated it with the manufacturer of the known clear. The numerical matrix was converted to a truth table in which a 1 indicated the present of that wavelength and a 0 when the wavelength is absent. The matrix was assembled by hand and was somewhat subjective because scans were taken in the fingerprint region of the infrared spectrum where peaks vary due to local neighborhood and molecular environment. Consequently, a peak was accounted as present in a range of +/- ten wavenumbers from the unique clear peak value.

EXPERIMENT 6

shows the elbow in the slope of the plots indicating that the number of clusters predicted by Principle Component Analysis is four. Although there are six reference clear polishes some were too similar to separate.

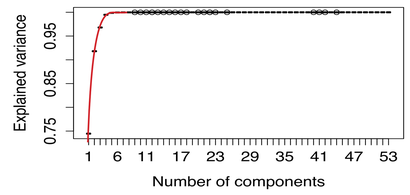


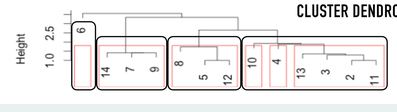
FIGURE 1

Spectra of a manufacturer's clear polishes, called reference clears, were used to associate unknown colored samples to their respective manufacturers. Note the similarity of the L'Oreal, Wet n Wild, and Sally Hansen spectra.

EXPERIMENT 7

shows a dendrogram with six predicted boxes with a table of sample names in each cluster. Cluster 2 successfully grouped all Sinful Colors samples together, while the other clusters were mixed. Blue colored boxes show the same data grouped into four clusters. The list identifies the samples.

CLUSTER DENDROGRAM



CLUSTER 1
6 OPI Clear

CLUSTER 2
14 Sinful Blue
7 Sinful Clear
9 Sinful Red

CLUSTER 3
8 Wet n Wild Red
5 L'Oreal Clear
12 OPI Red

CLUSTER 4
10 Sally Hansen Purple

CLUSTER 5
4 Wet n Wild Clear

CLUSTER 6
13 Wet n Wild Peach
3 INM clear
2 Sally Hansen Clear
11 Sally Hansen Cora

EXPERIMENT 8

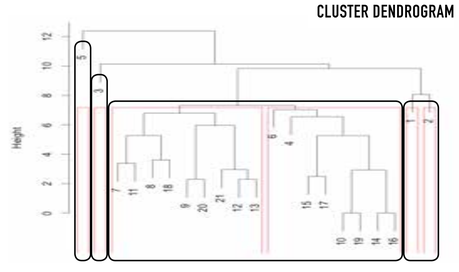
showed trends in the data visualized as a list of values normalized from 0 to 1 in which adacent samples were the most similar to each other. The general trend was for dark colors to cluster together while light colors and clears formed a separate cluster. Sinful Colors Clear had the most peaks and was the most different of all the clears helping it to successfully cluster with **Sinful Red** and **Sinful Blue**.

Cluster 1 OPI Clear	
Cluster 2 Sinful Blue	
Sinful Clear	Clusters 2 and 3 dark pigments and dyes
Sinful Red	
Cluster 3 Wet n Wild Red	
L'Oreal Clear	
OPI Red	
Cluster 4 Sally Hansen Purple	
Cluster 5 Wet n Wild Clear	
Cluster 6 Wet n Wild Peach	Clusters 4 through 6 light colors and clears
INM clear	
Sally Hansen Clear	
Sally Hansen Coral	

EXPERIMENT 9

shows the results of the same six reference clear samples compared to a larger group of unknown colored samples. Additional colored samples changed how the samples clustered. As the number of dark colored samples increased a separate cluster of similar colors formed and diminished the ability of reference clears to identify brands. This result agrees with the theoretical results in Experiment 5.

CLUSTER DENDROGRAM



CLUSTER 1
5 OPI Clear

CLUSTER 2
3 Wet n Wild Clear

CLUSTER 3
7 Wet n Wild Red
11 OPI Red 1
8 Sinful Red
18 OPI Red 2
9 Sally Hansen Purple
20 Sinful Navy
21 Sinful Orange
12 Wet n Wild Peach
13 Sinful Blue

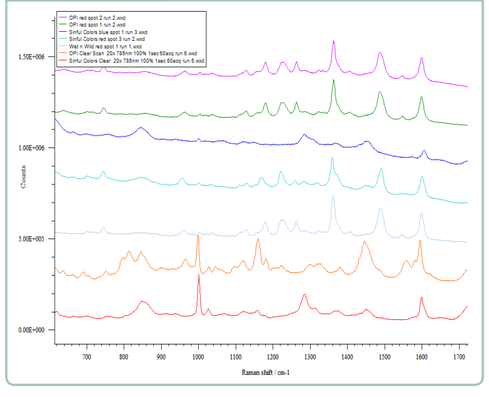
CLUSTER 4
6 Sinful Clear
4 L'Oreal Clear
15 OPI Beige 1

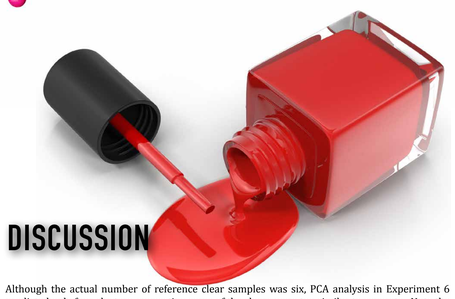
CLUSTER 5
1 Sally Hansen Clear
17 OPI Pink Sparkle
10 Sally Hansen Coral
19 Sinful Green
14 OPI Beige 1
16 OPI Green

CLUSTER 6
2 INM Clear

FIGURE 2

supported the conclusion that red and blue spectra are sufficiently different to cluster in separate clusters from their respective clear polishes. The top five traces are OPI and Sinful Colors red and blue spectra while the bottom two traces are OPI Clear and Sinful Colors Clear respectively.





DISCUSSION

Although the actual number of reference clear samples was six, PCA analysis in Experiment 6 predicted only four clusters, suggesting some of the clears were too similar to separate. Note the similarity of L'Oreal, Wet n Wild, and Sally Hansen spectra in Figure 1. These spectra suggest that L'Oreal and Wet n Wild were the same formulation marketed by different manufacturers, a claim that could not be substantiated by other means. The Sally Hansen spectrum is also very similar except for the much greater magnitude of the peak near 100 wavenumbers. The similarity of these three samples negatively affected accurate clustering.

The most unique clear in Experiment 7 was the Sinful Colors Clear that was able to cluster with its respective blue and red polishes. OPI and Wet n Wild clustered separately in individual clusters, while the remaining clusters were mixed. A clustering trend of the same data was visualized in Experiment 8 by ranking the samples by similarity that revealed dark colored samples clustered together with light colored and clear samples together in a different cluster. Experiment 9 added additional colored samples. Dark colored samples, mostly reds and blues formed cluster 3, while light colored and clears formed cluster 4. OPI Clear, Wet n Wild Clear, Sally Hansen Clear, and INM Clear all formed individual, separate clusters.

Although theoretical modeling suggested clustering followed the similarity of the underlying chemical formulation of the reference clear samples, it also suggested clustering could be driven by the added spectral complexity of new components such as colorants. If the same colorant were used by all manufactures, samples would cluster due the to colorant and not by the manufacturer's underlying proprietary formulation. As a consequence, reference clears would group together due to their spectral simplicity. These trends are seen in this study.

CONCLUSIONS

- Analytical results closely followed theoretical predictions.
- Reference clear samples are go-along markers, not modeled reference standards.
- Clustering is driven by overall spectral similarity.
- Dark samples clustered together because they have similar, stronger absorbing pigments.
- Forming the Truth Table is subjective and can't handle broad, double or triple peaks.
- Very similar spectra cannot be accurately discriminated.

DISCLAIMER

No support or consideration, financial or otherwise, was provided by any commercial entity mentioned in the report.

REFERENCES

1. R: A language and environment for statistical computing. http://www.R-project.org.
2. Renishaw Inc., 5277 Trillium Blvd., Hoffman Estates. IL 60192, United States www.renishaw.com

ACKNOWLEDGEMENTS

1. Kelsey Hartt's travel expenses were kindly sponsored by UVU Office of Engaged Learning - GEL Grant.
2. Dr. Gary H. Naisbitt travel expenses were kindly sponsored by UVU Office of Engaged Learning – GEL Grant, the Department of Criminal Justice and by the College of Aviation and Public Services.
3. The authors thank Mr. Paul Fenske for the graphic design and production of this poster.

CONTACT INFORMATION

Request a Copy

Dr. Gary H. Naisbitt, Associate Professor - Director of Forensic Science
Utah Valley University
Criminal Justice Department MS 126
800 West University Parkway
Orem, UT 84058

Please E-Mail: **naisbiga@uvu.edu** | Phone: **801-863-7230**