Data fusion from spectroscopic techniques for the discrimination of colored automotive paint mixtures

Morgan Carpenter*, B.S.; Patrick Buzzini, Ph.D.

Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

INTRODUCTION

Comparative examinations of paint typically involve the use of an analytical sequence of microscopical examinations and various instrumental analysis methods.

In the event of a non-differentiation, and in order to address the question of a common source between compared sets, the expert should estimate the chance of observing in the population of interest another coated object with the properties detected with method A *and* the properties detected with method B *and* the properties detected with method C *and* so on.

Exploring data fusion methods for this purpose is essential because:

- They offer the potential to use the detected properties simultaneously to address the question of interest.
- They can detect and manage redundancy and dependency of some of the detected features (i.e., color and pigments).

In this study, the interest is in exploring if the way to combine high dimensional data collected from automotive paint samples of known binary pigments compositions and proportions, using two different analytical techniques, impacts the classification accuracy of a classical chemometric method.

The question arises because repeated measurements with both methods are not taken on the same area of the specimen simultaneously. Hence, it is hypothesized that a sequential data acquisition process constitutes a source of variation due to the heterogeneity of automotive paint.

MATERIALS AND METHODS

Samples:

- Known proportions of Standox Standohyd Basecoat Automotive Refinish Paint from the base colors red (C.I. *Pigment Red 254*) and yellow (C.I. *Pigment Yellow 184*), blue (C.I. *Pigment Blue 15*) and yellow and and blue and green (C.I. *Pigment Green 36*) were mixed in the following proportions: 70%-30%; 50%-50% and 30%-70% by weight.

Analytical methods:

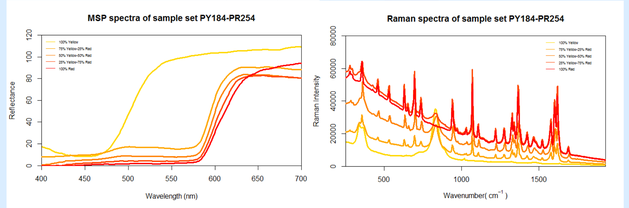
- Visible MSP analysis (400-700nm) in reflectance mode using the Video Spectral Comparator 6000 by Foster & Freeman.
- Raman analyses using an XploRA Raman microscope (50x obj.) from Horiba Scientific using a NIR laser source at 785nm and a spectral range of 2000-250cm^{-1}.
- Seven replicates were taken at random points on each sample.

Data Analysis:

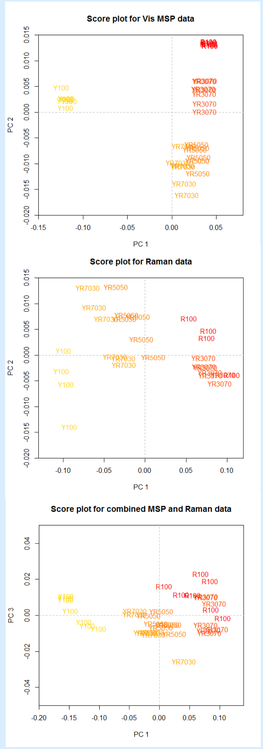
- 5 datasets of MSP and Raman from each sample set were combined as follows:
  - 1 set by side-by-side binding (Comb0)
  - 4 sets by randomly assigning replicates within each proportion (CombR1 through R4)
- Due to differing scales between the MSP and Raman signals the normalization method of area unity was used.
- Principal component analysis (PCA) was used as a data reduction technique and to verify group separations.
- Linear discriminant analysis (LDA) was conducted on the 3 first principal components as new variables; 4 replicates for each mixture set were extracted as a training set and the 3 remaining ones were used as a testing set. Assignments of training sets and testing sets was kept the same for all compared datasets.
- Random assignments for datasets binding, PCA and LDA were computed using R Statistical software (1) with the integrated MASS package (2).

RESULTS

C.I. *Pigment Yellow 184* – C.I. *Pigment Red 254* Sample Sets



Both Vis MSP and Raman spectra show high contribution of the red pigment compared with the yellow



The PC1-PC2 score plots of the Vis MSP data show distinct separation between the mixed sets except for YR 70:30 and 50:50

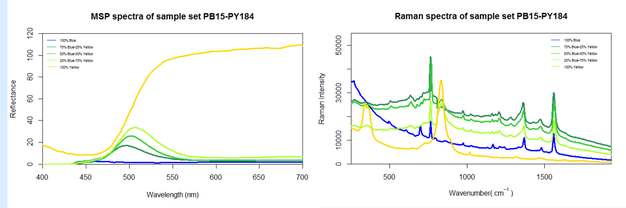
Raman data exhibit more overlapping between the 100% red pigment and the YR 30:70 proportion which is in line with the full spectra above.

Overlaps are present in the PC1-PC3 plot for the side-by-side combined datasets (Comb0)

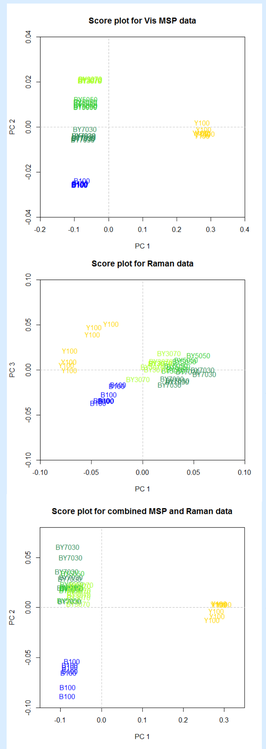
PCA+LDA classification prediction accuracy						
MSP	Raman	Comb0	CombR1	CombR2	CombR3	CombR4
100%	73%	100%	100%	100%	100%	100%

A lower correct classification rate was observed for Raman, both combination techniques yielded accurate classification

C.I. *Pigment Blue 15* – C.I. *Pigment Yellow 184* Sample Sets



Vis MSP spectra show clear differences. The major bands of the blue phthalocyanine pigment dominate Raman spectra at all proportions



The PC1-PC2 score plots of the Vis MSP data show distinct separation between all sets

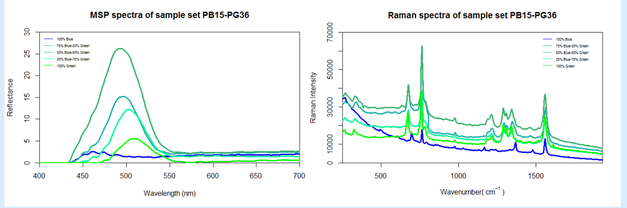
Raman data exhibit closeness between the mixed group in the projected plane, however the groups are still discernible

Overlaps are present in the PC1-PC2 plot for the side-by-side combined datasets (Comb0)

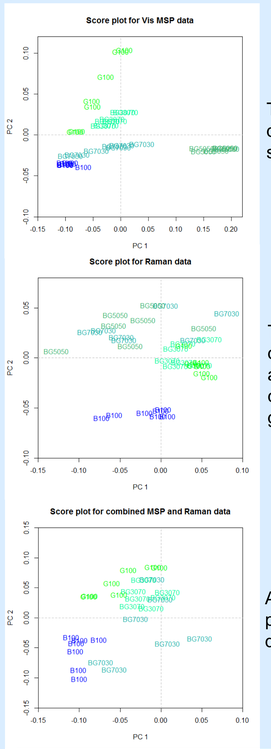
PCA+LDA classification prediction accuracy						
MSP	Raman	Comb0	CombR1	CombR2	CombR3	CombR4
100%	100%	100%	100%	100%	87%	93%

Different accuracy rates are observed for the randomly combined datasets with the side-by-side combined datasets having the highest accuracy rate

C.I. *Pigment Blue 15* – C.I. *Pigment Green 36* Sample Sets



Vis MSP spectra show clear differences. The green pigment is more predominant at lower concentrations



The PC1-PC2 score plots of the Vis MSP data show distinct separation between all sets

The PC1-PC2 score plots of the Raman data show overlaps of BG 50:50 and 70:30 as well as 100% G and BG 30:70 confirming the high contribution of the green pigment

As with the Raman PCA plot, overlaps are present within the side-by-side combined datasets (Combo0)

PCA+LDA classification prediction accuracy						
MSP	Raman	Comb0	CombR1	CombR2	CombR3	CombR4
100%	60%	100%	93%	93%	93%	93%

The same accuracy rates are observed for the 4 randomly combined datasets which are slightly lower than the side-by-side combined datasets.

DISCUSSION

- Analytical techniques were chosen due to the dependencies of the detected analytical information related to the pigment components
- The chemometric approach (PCA+LDA) was not optimized to obtain a fine-tuned accuracy rate (i.e., no crossvalidation); the goal was to avoid the effect of differential random selection of training and testing sets
- Mixture proportions were chosen to avoid challenging the sensitivity of the chosen classifier
- In the present setting, the classifications of MSP data and of side-by-side combined datasets have always resulted in 100% accuracy for all sample sets
- A higher number of randomly assigned combinations could be tested, and different number of replicate measurements could be considered as well

CONCLUSIONS

- This study was concerned with the way that datasets obtained from the same samples using two different analytical techniques should be combined for a simultaneous use
- Out of comparisons involving 5 combined datasets, 1 side-by-side combined dataset and 4 randomly assigned combinations of datasets, differences were observed between the three chosen sample sets

REFERENCES

1) R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
2) Venables WN, Ripley BD. Modern Applied Statistics with S (4th Ed.). Springer, New York, NY (2002).