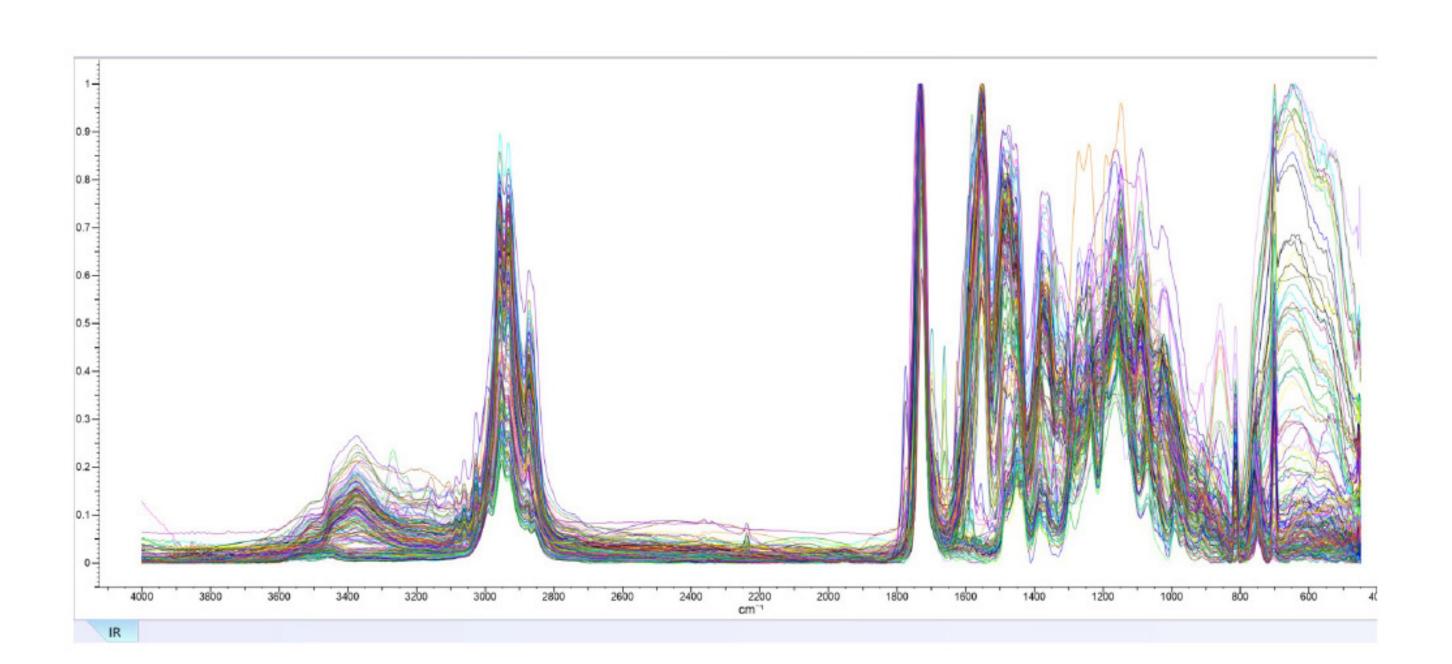


Uncovering Hidden Trends in Analytical Data Using Principal Component Analysis

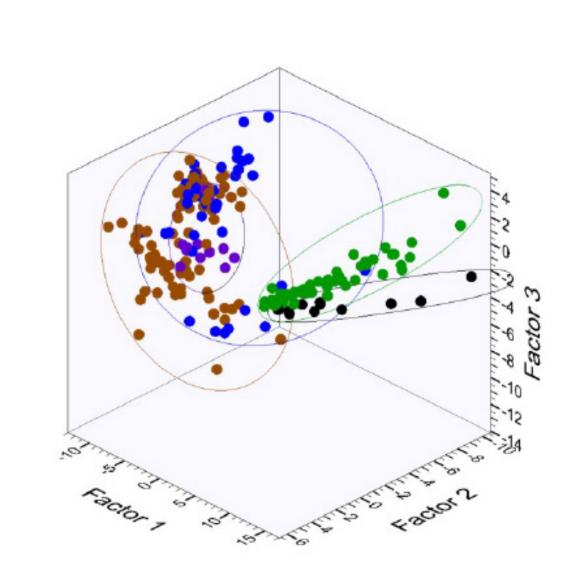
Michelle D'Souza, PH.D., Karl Nedwed, Ty Abshear John Wiley & Sons, Inc.

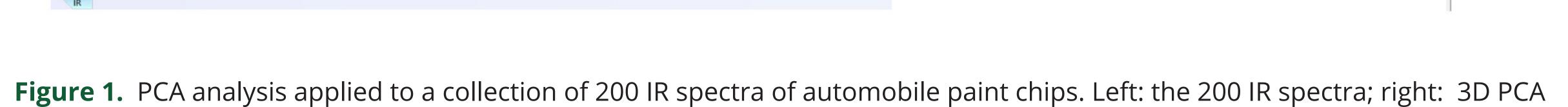
Dataset 1: IR Spectroscopy of Automotive Paint Samples

- **Dataset:** First 200 IR spectra from Wiley's "IR Automobile Paint Chips" database
- PCA Parameters: preprocessing "mean-center"; Y-transform "2nd derivative with number of points 11"
- **Results:** PCA analysis successfully separated paint samples into two distinct clusters corresponding to enamel and lacquer paint types (Figure 1). The 3D scores plot clearly demonstrates this separation, with minimal overlap between paint categories. This clustering enables automated classification of unknown paint samples by projection into the established PCA space.



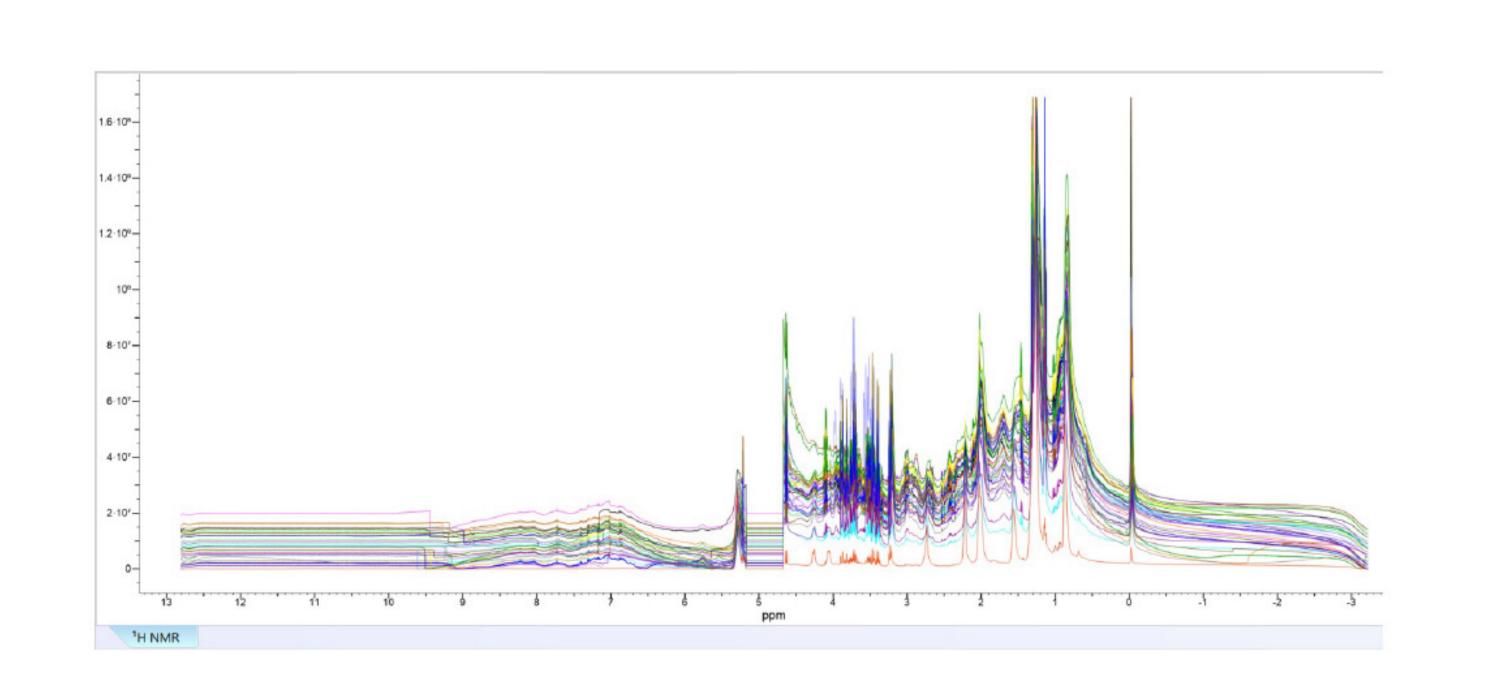
scores plot showing well separated clustering of enamel and lacquer paint types.





Dataset 2: ¹H NMR Analysis of Blood Samples

- **Dataset:** 34 ¹H NMR spectra from diabetic and normal blood samples
- PCA Parameters: preprocessing "mean-center"; Y-transform "divideby Sample-2Norm"; "Intelligent binning" method were used to reduce the resolution of H NMR spectra
- Results: PCA revealed distinct separation between diabetic and normal blood samples in the 2D scores plot (Figure 2). The loadings plot can be used to identify the bio-marker of diabetic conditions.



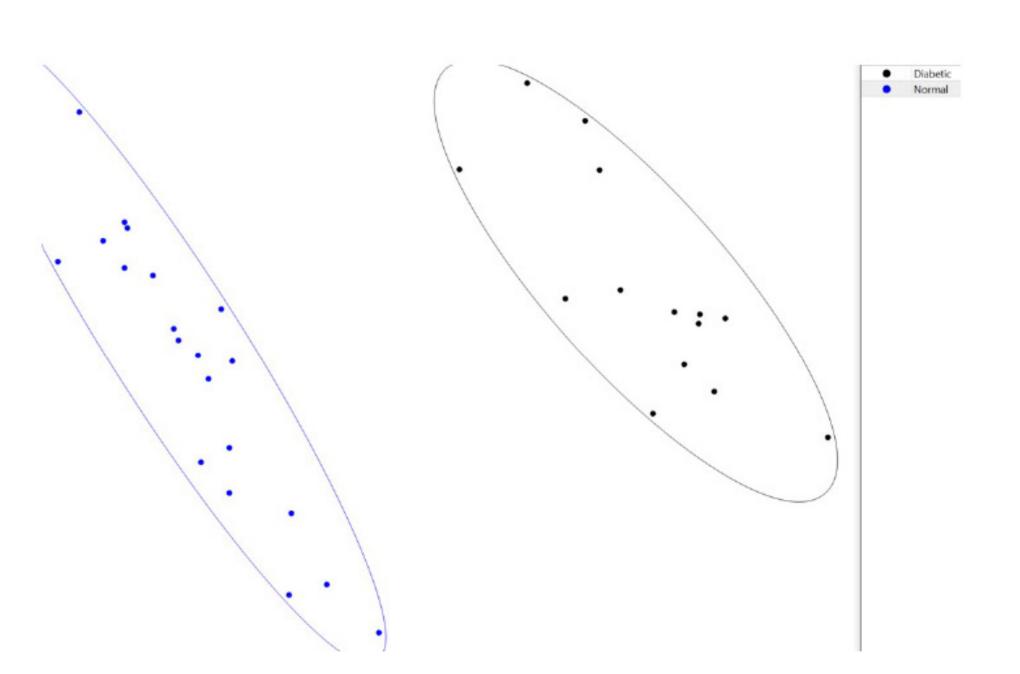
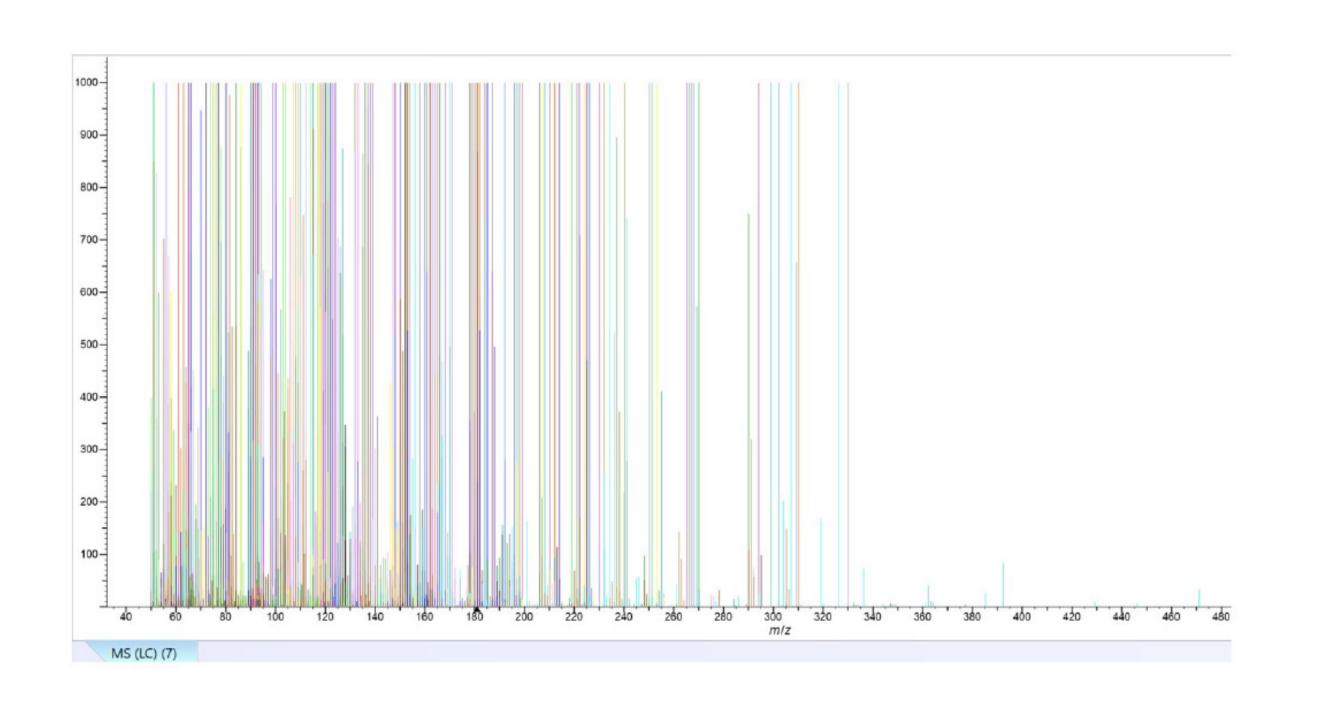


Figure 2. PCA of diabetic vs.normal blood samples. Left: the HMNR spectral data; right: the PCA 2D scores plot demonstrating sample categorization.

Dataset 3: LC-MS Tandem Mass Spectrometry Analysis

- **Dataset:** A similarity structure search hit list of 200 records from "Wiley Registry of Tandem Mass Spectral Data: MSforID" database was used for this research
- **PCA Parameters:** Preprocessing "Mean-center"; Y-transform "divideby Sample Max".
- Results: The 3D PCA scores plot revealed a clear relationship between collision energy and spectral clustering (Figure 3). Higher collision energy spectra showed greater dispersion, while lower collision energy spectra clustered tightly (black dots). This pattern reflects the increased molecular fragmentation at higher collision energies, resulting in more diverse and differentiated mass spectra.



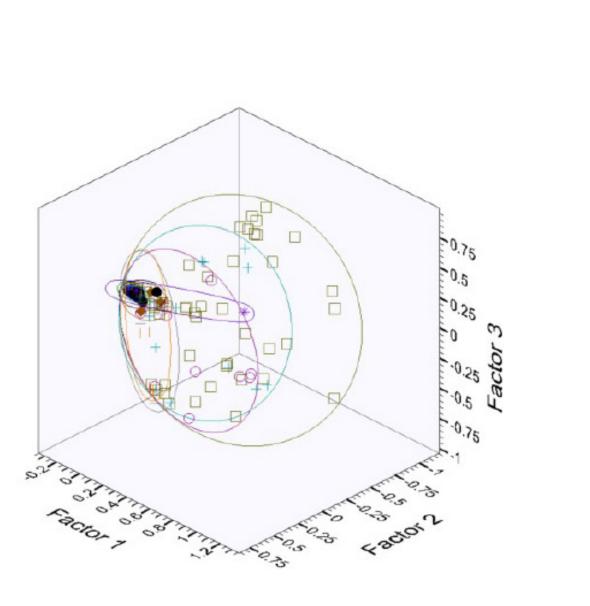


Figure 3. PCA of tandem mass spectrometry data. Left: the 200 hit list MS spectra; right: the PCA 3D scores plot showing energy-dependent clustering patterns.

Abstract

We present a comprehensive analysis of spectral datasets across various techniques using Principal Component Analysis (PCA) that reveals underlying patterns and correlations that remain invisible when examining individual spectral measurements in isolation. This study demonstrates how PCA can enhance scientific discovery by extracting insights from high-dimensional data. Our findings show that PCA successfully identifies spectral clusters and feature correlations, demonstrating its value for improving analytical workflows.

Introduction

Principal Component Analysis (PCA) is a powerful statistical technique used to simplify complex datasets by reducing their dimensionality while retaining most of the original variability. Applied to spectral data, it can review hidden trends not observed by traditional one by one spectral examination. In this poster, we will perform PCA on several database collections including: IR, HNMR, MS though it can also be applied to Raman, UV-Vis, and chromatographic data sets as well.

Method & Results

All analyses were performed using the TrendFinder application in KnowItAll 2026. Three distinct datasets were analyzed to demonstrate PCA's capabilities across different analytical techniques.

Conclusion

PCA is a very powerful tool for spectra processing and analysis of large datasets. It uses mathematical methods to complement traditional spectral analysis by finding trends that are not visible. The results from this kind of approach bring math, computer engineering, and chemistry together.

This study demonstrates PCA's effectiveness as a complementary analytical tool across multiple spectroscopic techniques. Key findings include:

- Classification capability: Successfully distinguished between different sample types without prior knowledge
- Method universality: Proved effective across diverse spectroscopic methods
- Physical insight: Revealed meaningful relationships between experimental parameters and spectral characteristics

These results highlight PCA's broad applicability for enhancing analytical workflows through pattern recognition and sample classification.