



Chemometric Analysis and Visualization Tools for IR Spectral Searching

Marie Scandone¹ and Dana Garcia²

(1) Bio-Rad Laboratories, Inc.

Informatics Division, Two Penn Center, Philadelphia, PA 19102

(2) Arkema Inc.

900 First Avenue, King of Prussia, PA 19406-0936

Outline

- **Introduction**

- Cheminformatics
- Chemometrics: Principal Component Analysis (PCA)
- Overlap Density Heatmaps (ODH)

- **Experimental**

- Materials and Methods

- **Results**

- Simple Search
- PCA
- PCA and ODH Identification of Material Components

- **Conclusions**

Introduction – Study Motivation and Approach

- **IR spectral identification can be a challenging process:**
 - Overlapping peaks
 - Reference spectra unavailability

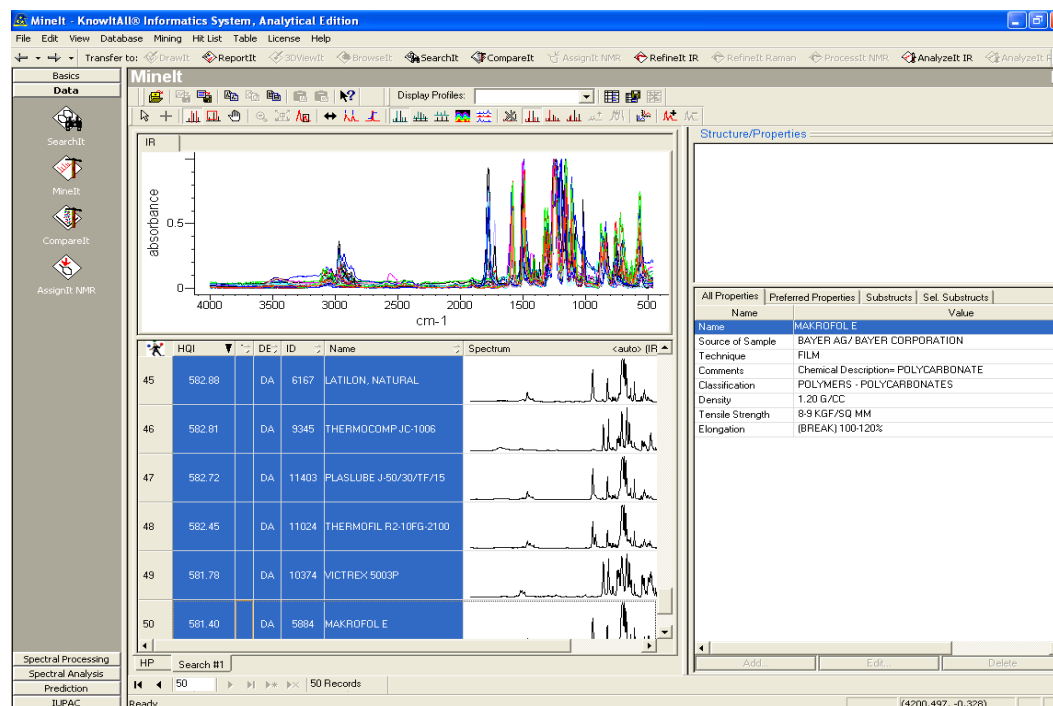
- **Proven approach a combination:**
 - Cheminformatics – spectral and property databases
 - Chemometrics – Principal Component Analysis (PCA)
 - Overlap Density Heatmaps (ODH)

Introduction - Cheminformatics

- **Cheminformatics: IR spectral and property databases.**

Traditional visualization of multiple spectra:

- Overlay
- Offset
- Stacked



These plotting methods obliterate trends when viewing large amounts of data.

Introduction – Chemometrics PCA

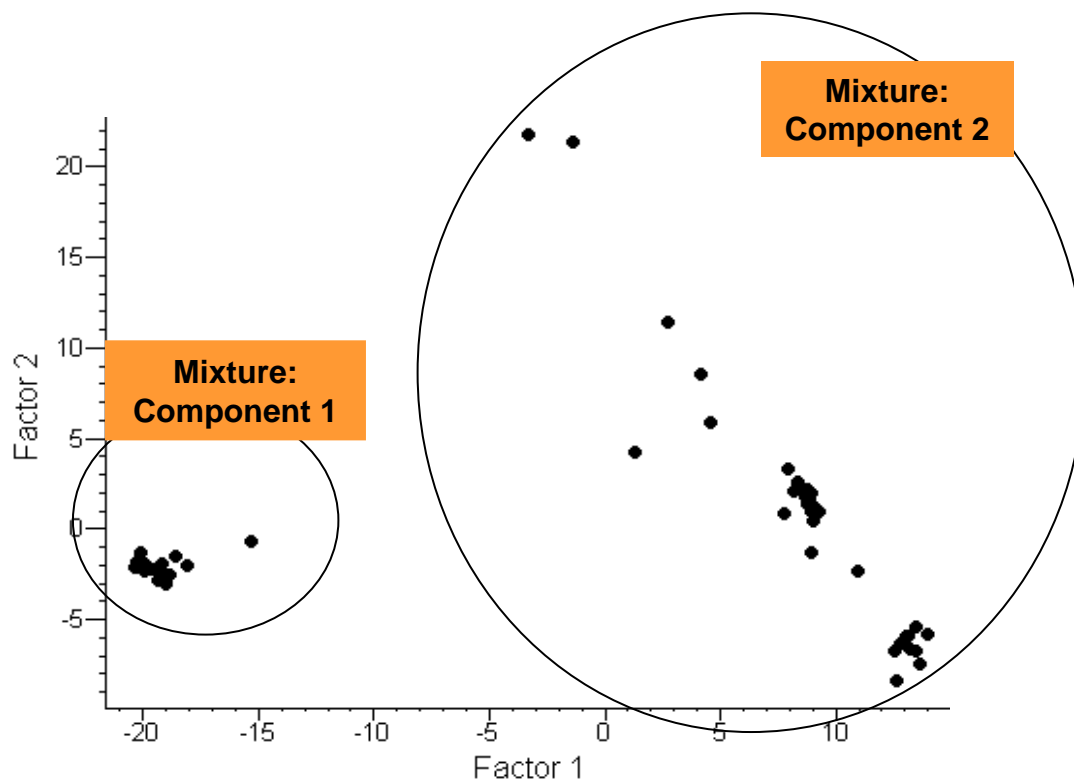
- **Chemometric PCA**: spectral decomposition to create a simplified representation of the data using a set of principal components (PCs).
- PCs describe the phenomena (concentration/property) that generate the spectral variations of interest.
- PCs are used:
 - Explore & Describe Data
 - Characterize & Classify

PCA analysis using Infometrix Pirouette[®]

Hit List Spectra PCA Score Plot

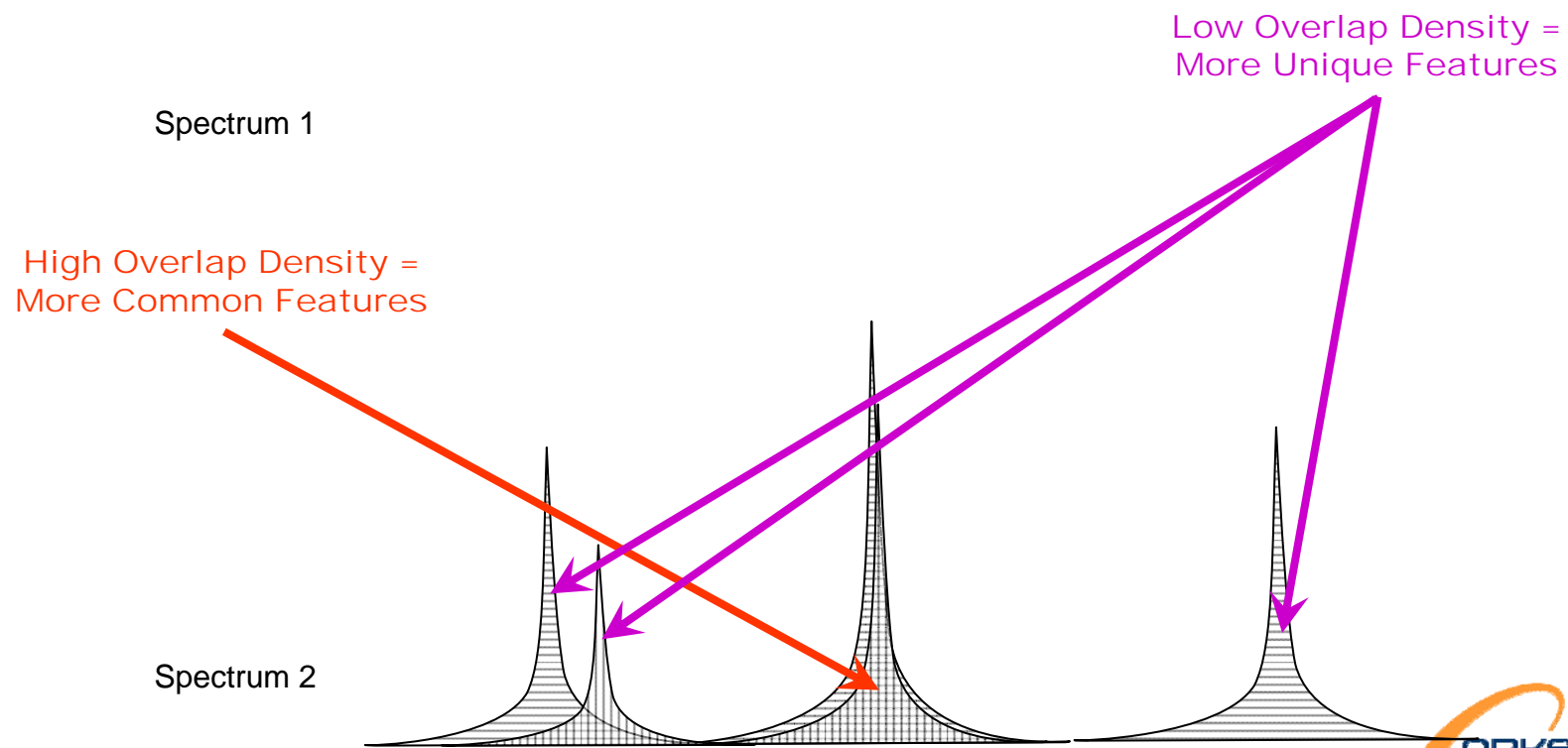
- **2 Component mixture:**

- Two clusters determined by the characteristic IR transitions for each mixture component

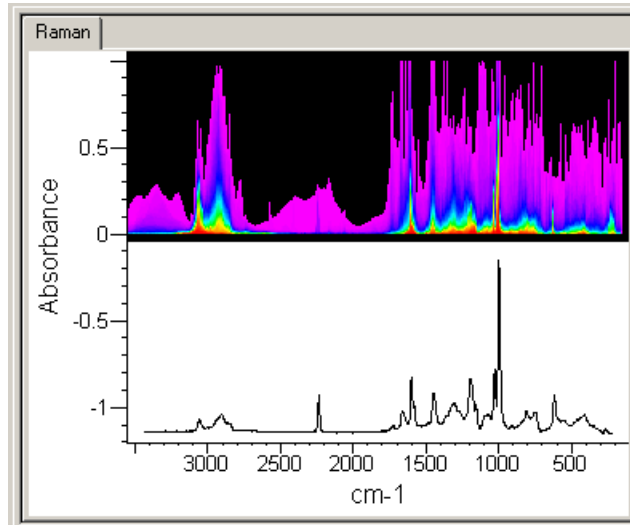


Introduction - Overlap Density Heatmaps (ODH)

- Visual data mining and analysis.
- Assess similarities and dissimilarities.

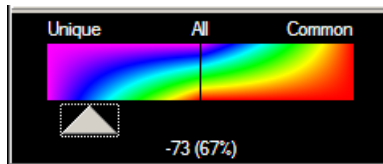


Overlap Density Heatmap (ODH) Display

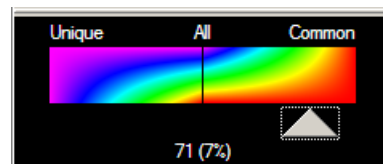


- ODH displays are obtained by selecting Hit List database spectra resulting from a standard search.

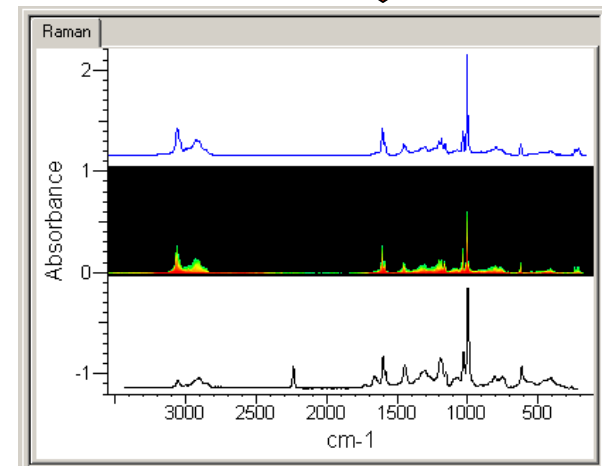
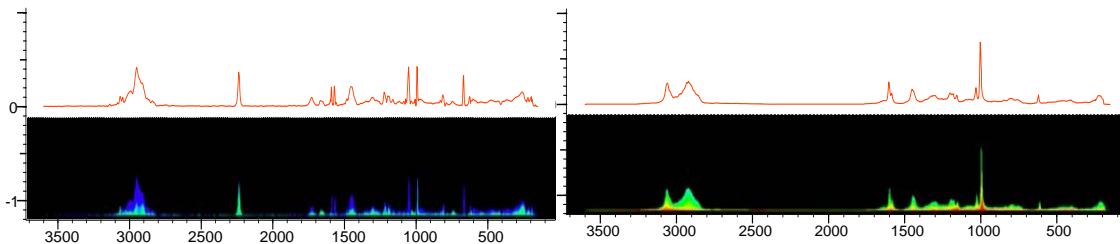
- The positive or negative consensus spectra can be calculated and searched in the databases



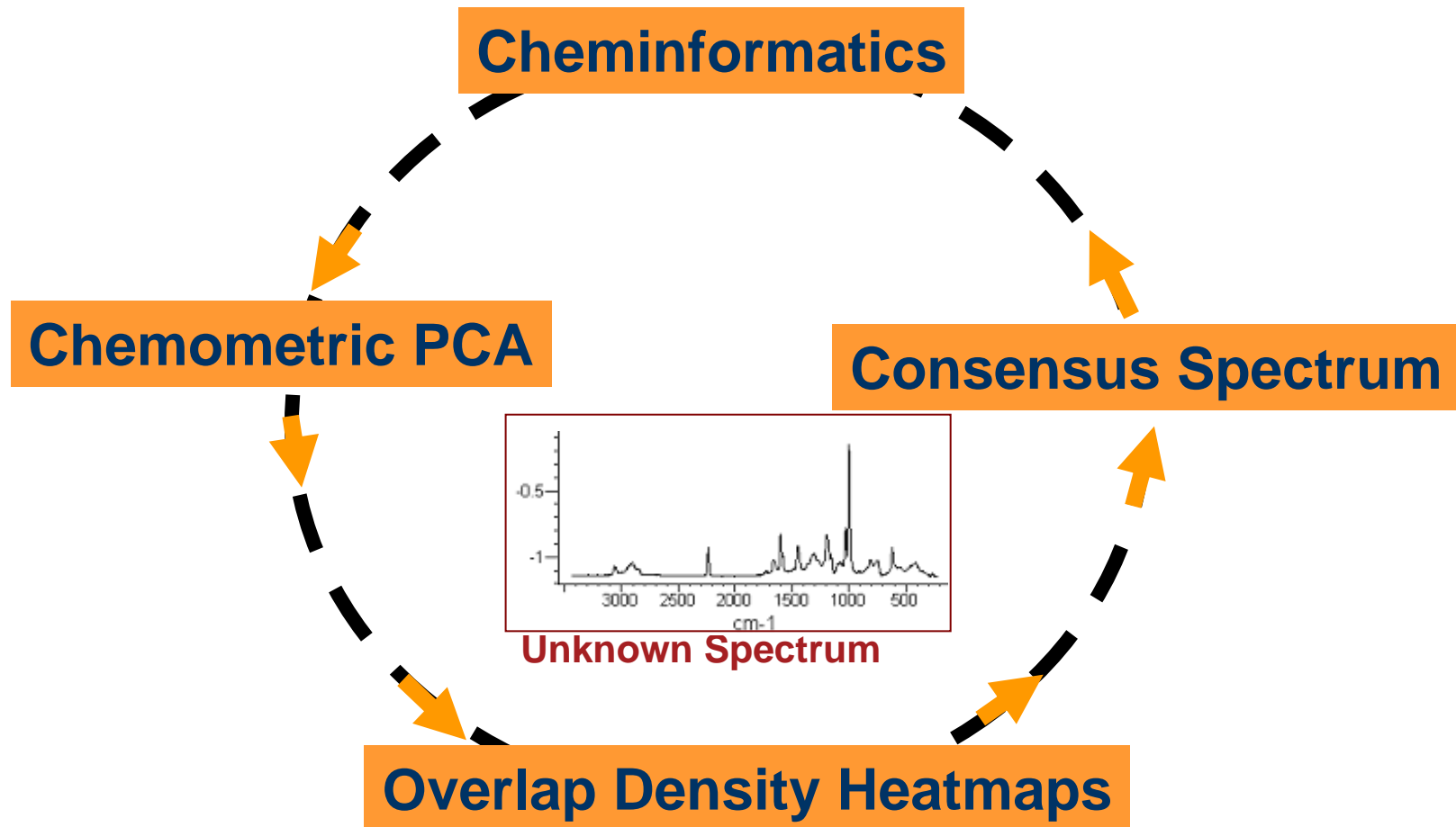
Unique



Similar



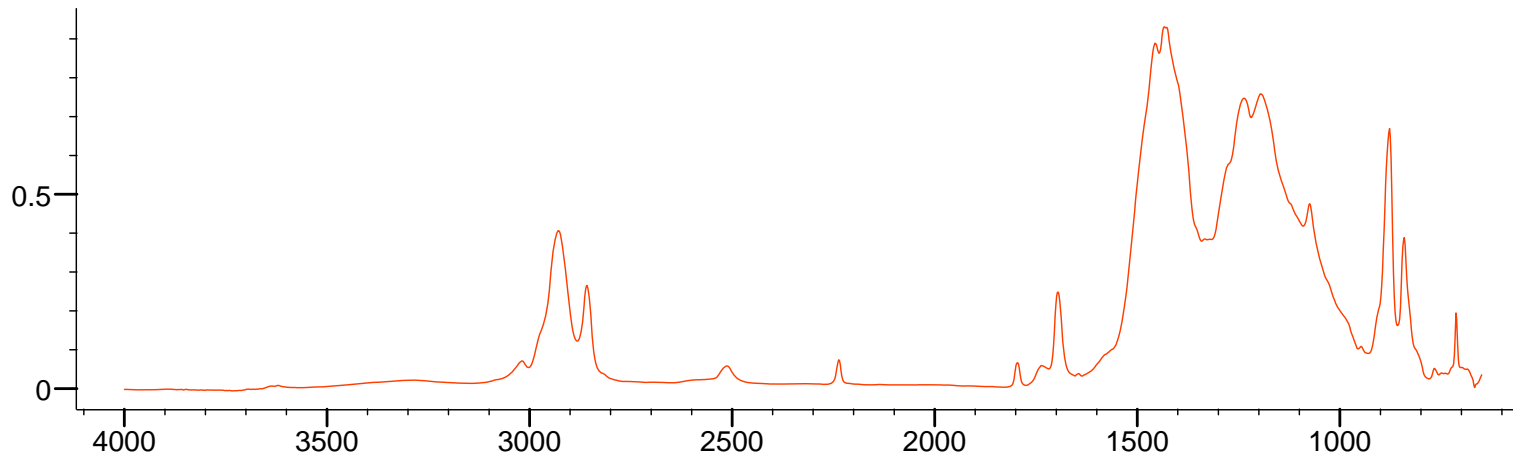
Iterative Process



- Progressive reduction of data complexity.

Experimental: Materials and Spectra

- **Materials:** commercial coating.
- **Spectroscopy:** ThermoFisher Avatar™ FT-IR with a diamond ATR.



- **ATR spectra challenging search:**
 - Subtraction can be difficult and yield inaccurate results
 - ATR correction may not improve results

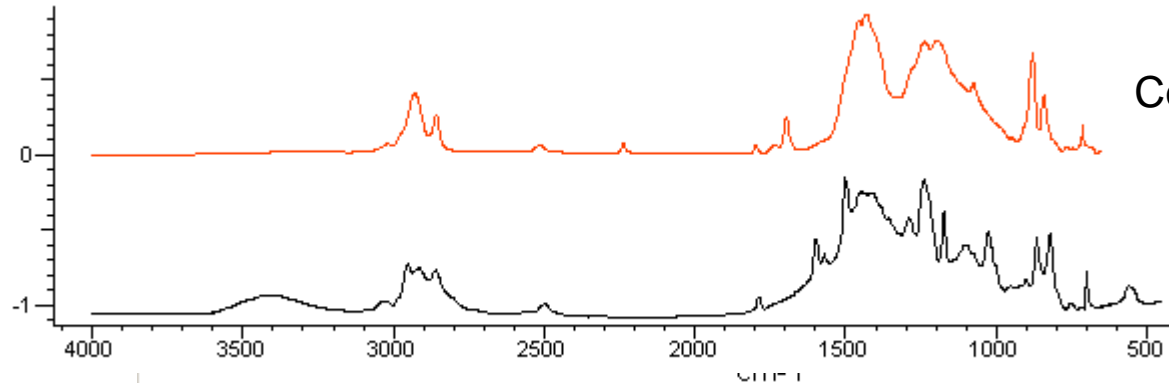
Experimental: Methods

- **Spectrum searched against the Sadtler and Arkema databases using SearchIt™ application.**
 - Euclidean Distance algorithm with 200 hits retrieved
- **Resulting Hit List transferred to Analyzelt™ MVP Chemometrics PCA.**
- **Standard chemometrics tools:**
 - Data pre-processing: smoothing derivative, normalization
 - Model setup: spectral range, variable selection, outlier detection

This study:

- 2nd Derivative, Savitzky-Golay polynomial filter smoothing
- MSC normalization
- Spectral ranges captures major spectral features

Results: Simple Search



Coating spectrum

Hit List – 1st hit

All Properties		Preferred Properties	
Name		Name	Value
Name	CASTALL 402/RT-7 (CURED)		
Technique	FILM		
Comments	Chemical Description= A TWO-COMPONENT, RESIN		
Distortion Temperature	130C		
Mixing Ratio	(RESIN:HARDENER) 100:15 BY WT.		
Source of Sample	CASTALL, INC.		
Tensile Strength	(25C) 7300 PSI		
Thixotropic Index	(25C) 1.12		

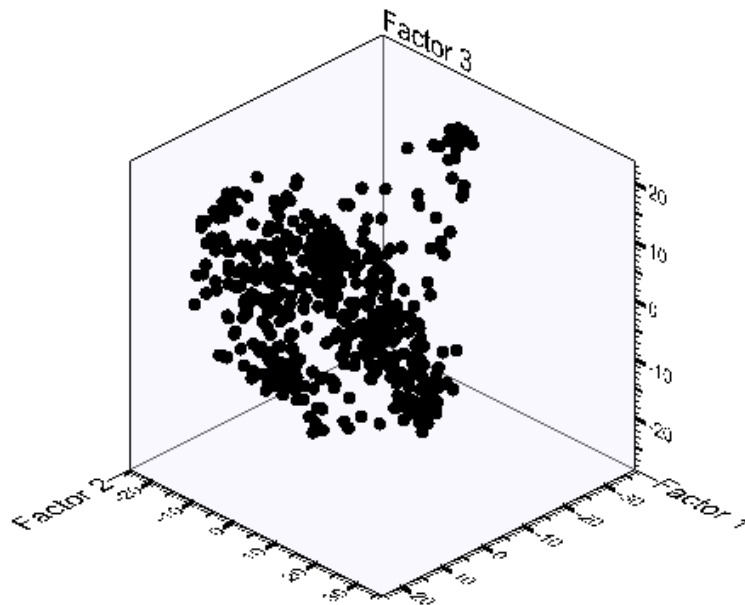
	DE	ID	Name	Spectrum
1	D	4850	CASTALL 402/RT-7 (CURED)	
2	D	2228	SONOLAC	
3	D	4814	AP-2309	
4	D	4848	CASTALL 402 RESIN	

Hit List

Spectral searching did not yield any definite results.

Results: Chemometrics PCA

Transfer Hit List to MVP Chemometrics PCA

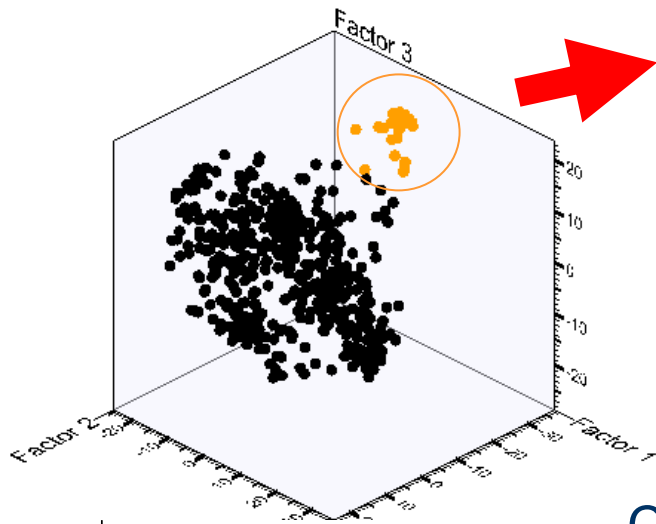


No specific clustering
to separate components

Approach: investigate score plot outer areas
to look for differentiation.

Results: Chemometrics PCA and ODH

Component 1



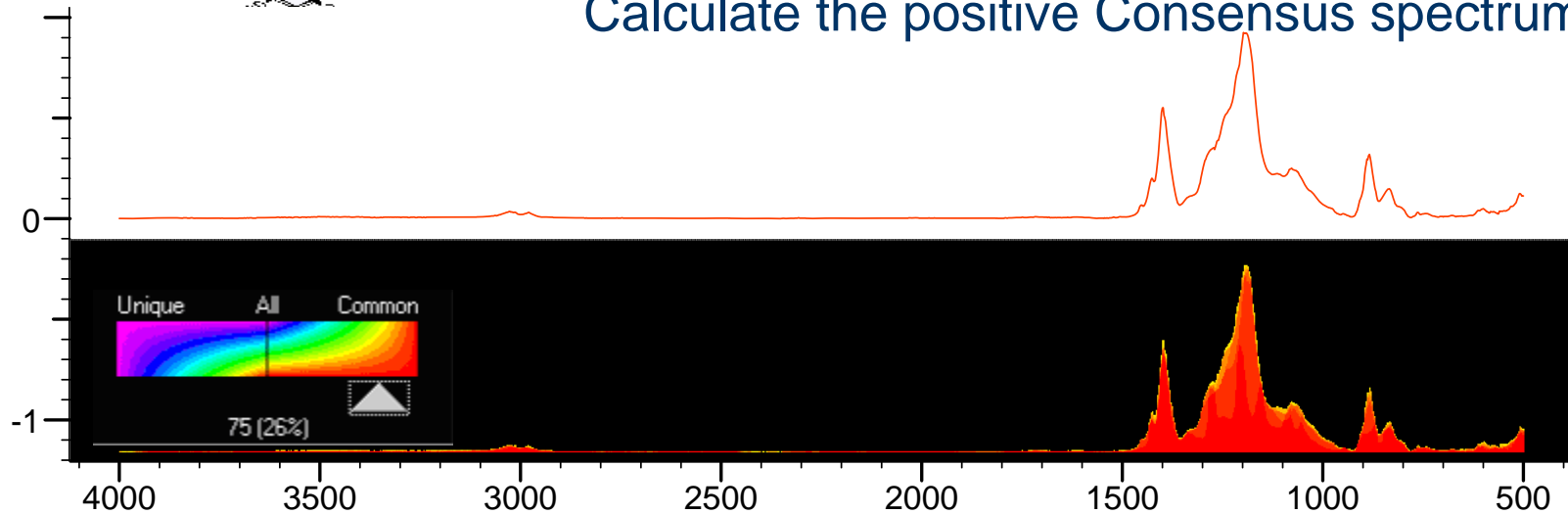
Select samples which are different.



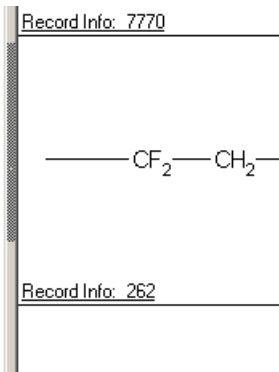
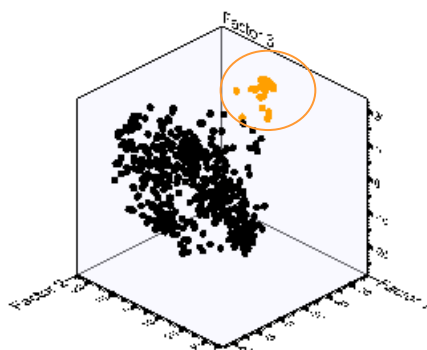
Obtain the Overlap Density Heatmap.



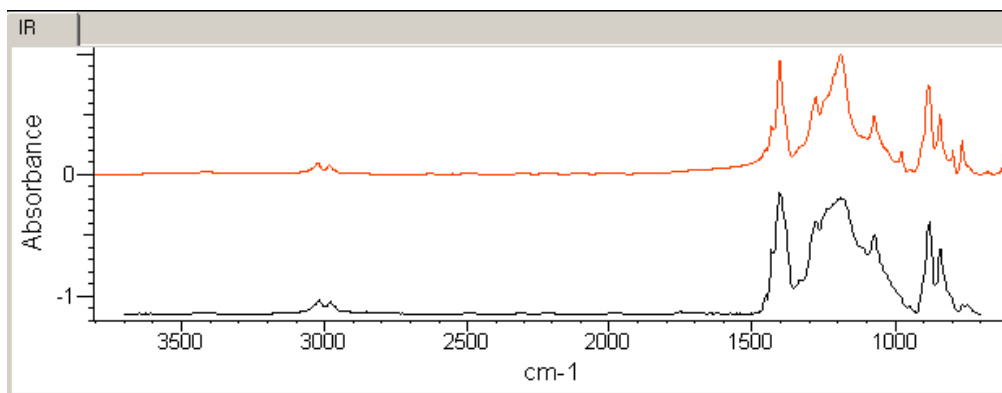
Calculate the positive Consensus spectrum.



Results: Identification Component 1



The common functional group of the chosen scores is —CF₂—CH₂—



Simple search of the Consensus spectrum identifies PVDF.

	HQI	DE	ID	Name	Spectrum
1	848.90	D	10188	POLY(VINYLDENE FLUORIDE)	
2	809.64	D	7770	TECNOFLON FOR 70	
3	797.02	D	10189	POLY(VINYLDENE FLUORIDE)	

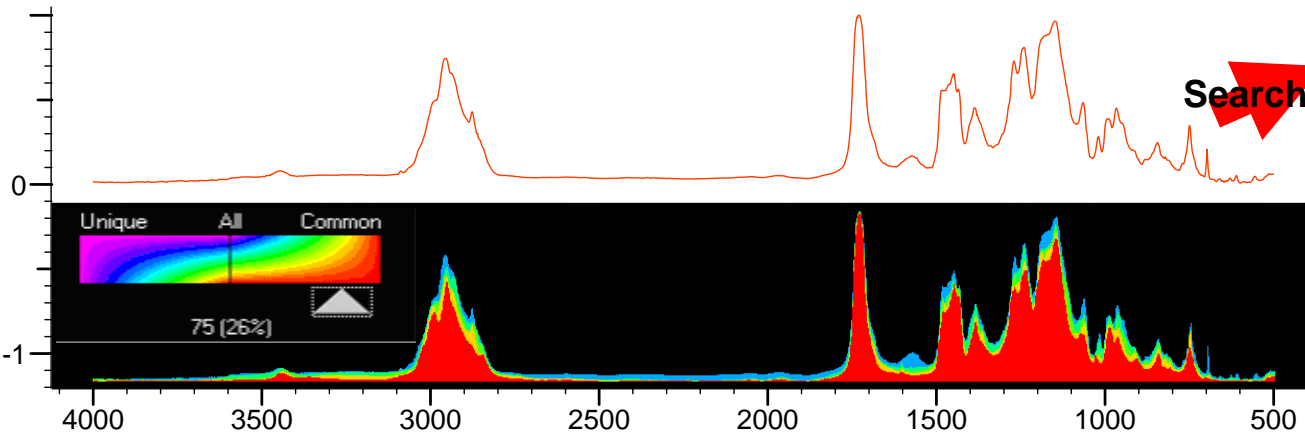
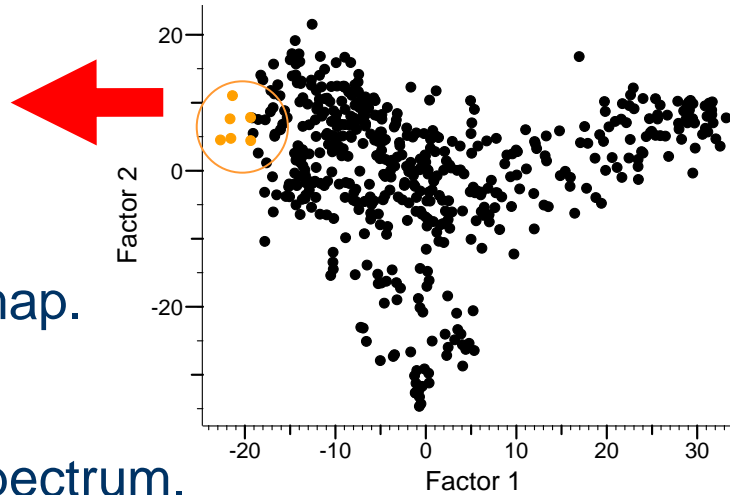
Component 1 = PVDF

Results: Chemometrics PCA and ODH - Component 2

Return back to the PCA analysis and select another set of samples that are different.

Obtain the Overlap Density Heatmap.

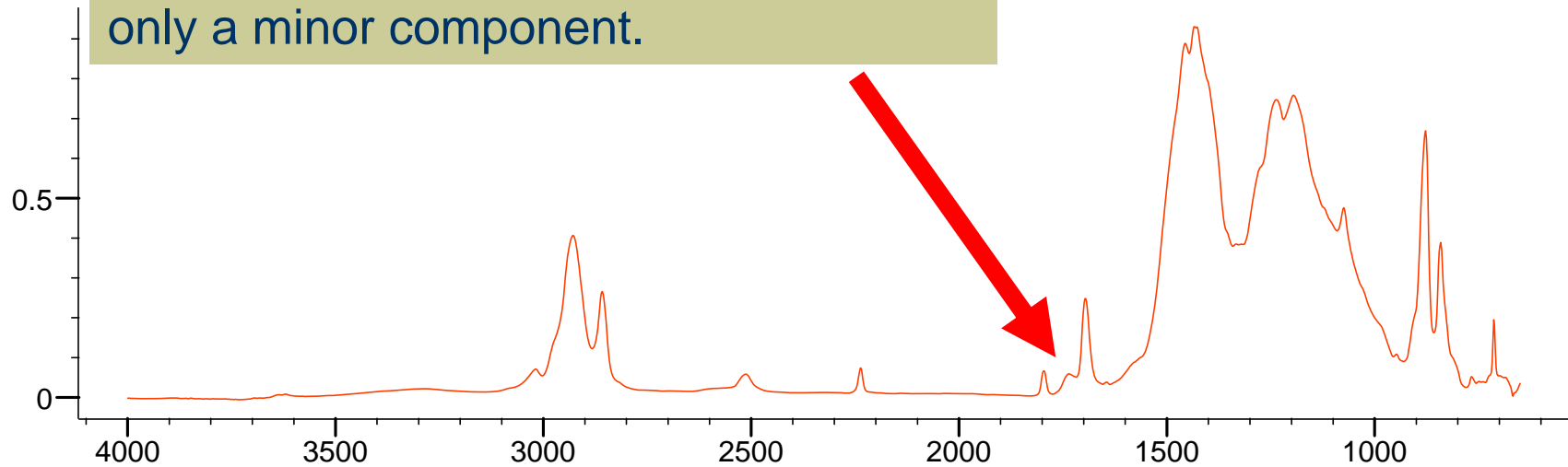
Calculate the positive Consensus spectrum.



Component 2 = Acrylic

Results: Component 2 - Acrylic

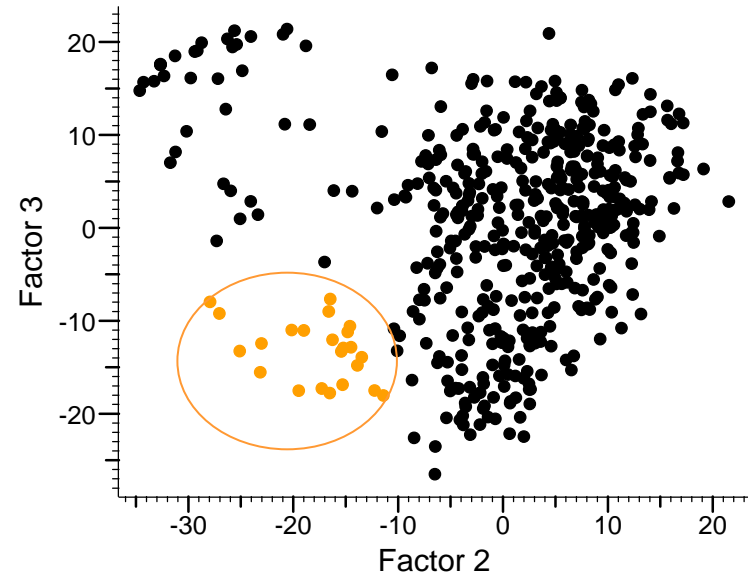
Identification of the acrylic although it is only a minor component.



- **PCA identified the presence of the acrylic.**

Results: Chemometrics PCA and ODH - Component 3

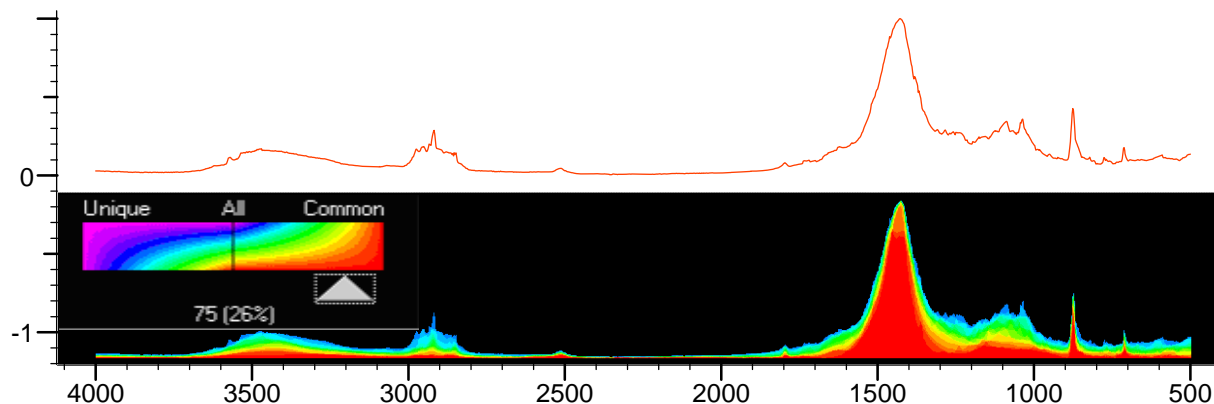
Return back to the PCA analysis and select another set of samples that are different.



Obtain the Overlap Density Heatmap.



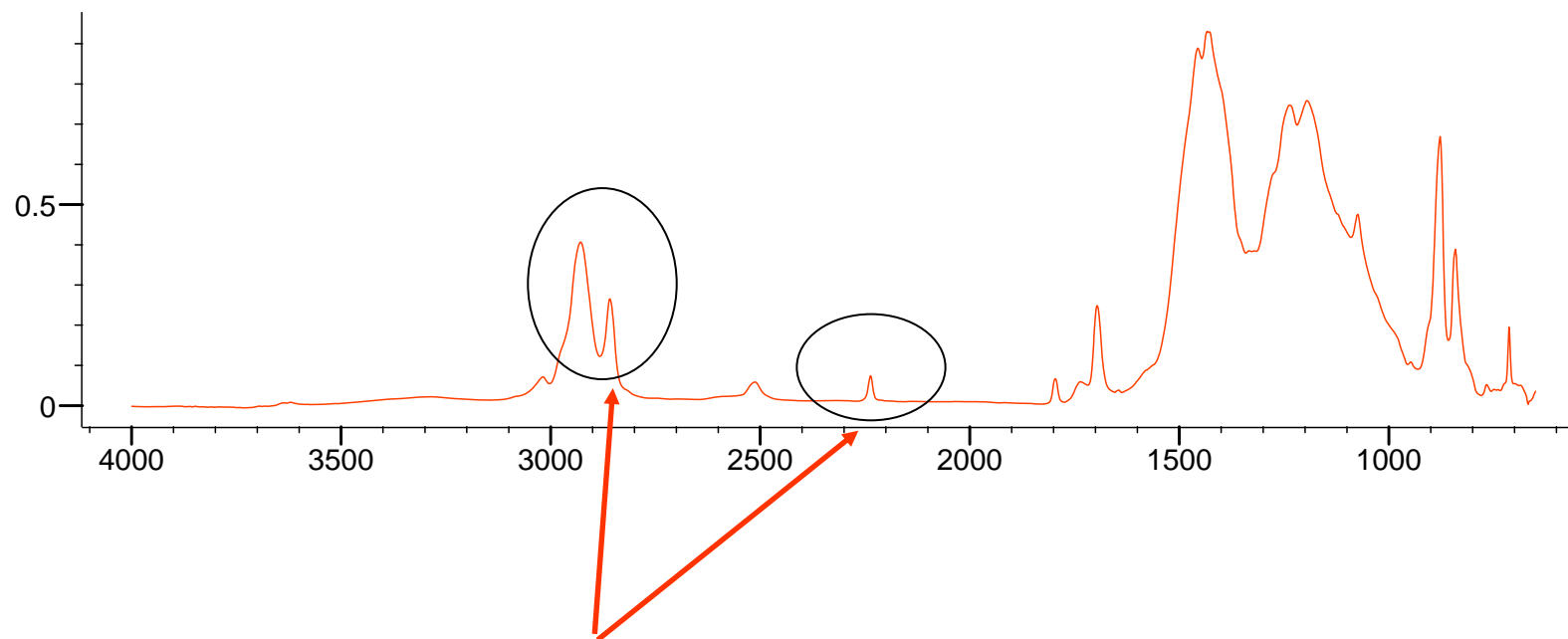
Calculate the positive Consensus spectrum.



Search

Component 3 = CaCO_3

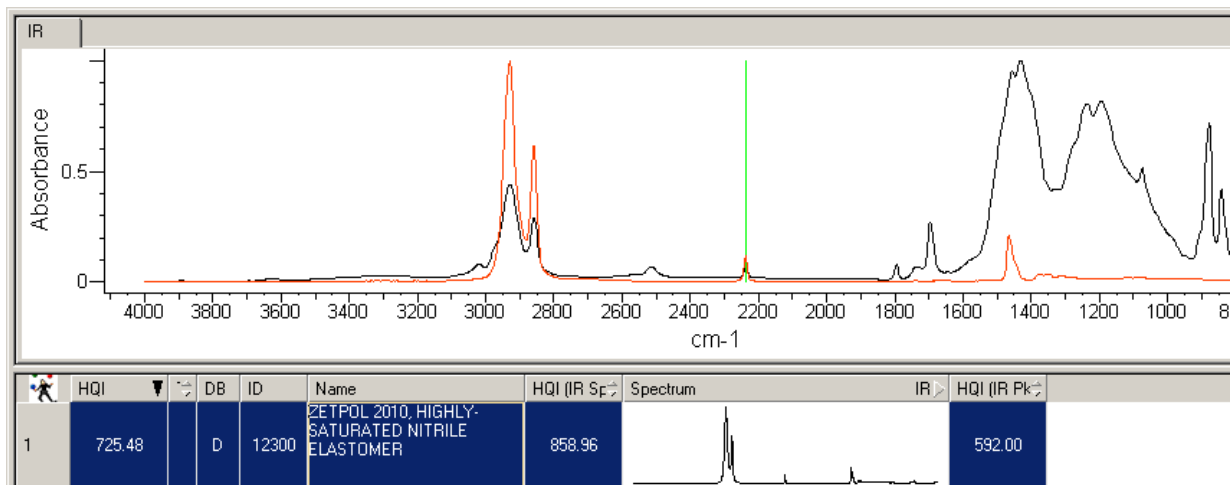
Results: Cheminformatics - Component 4



Remaining unassigned bands.

Results: Cheminformatics - Component 4

- **Simultaneous spectral and peak search.**
 - Sadtler Polymer and Monomer Databases



Search

Component 4 = Nitrile elastomer

Conclusions

- **IR spectral identification of multi-component samples lends itself to a combination approach using Cheminformatics, PCA, ODH and Consensus Spectra.**
- **High degree of module interconnectivity.**
 - Individual module provide partial answers
 - In totality a systematic and efficient analysis
- **Applicable to a wide variety of spectral identification problems.**