Chemometrics in Chromatography

Chromatography is an extremely versatile technique for the analytical laboratory. The chromatographic patterns generated by modern instruments are used in a wide variety of quantitative and qualitative analyses. The techniques are robust enough (and we have assembled experience enough) to allow a rapid development of chromatographic methods and move this experience into routine use in an analytical laboratory, quality control laboratory, or even an in-line process setting.

At least three goals can be identified for projects which use chromatographic instrumentation:

- *quantitation of the components in an analysis mixture*
- *separation of components in the mixture for purposes of fraction collection*
- *matching of the chromatographic patterns to an experience set or library*

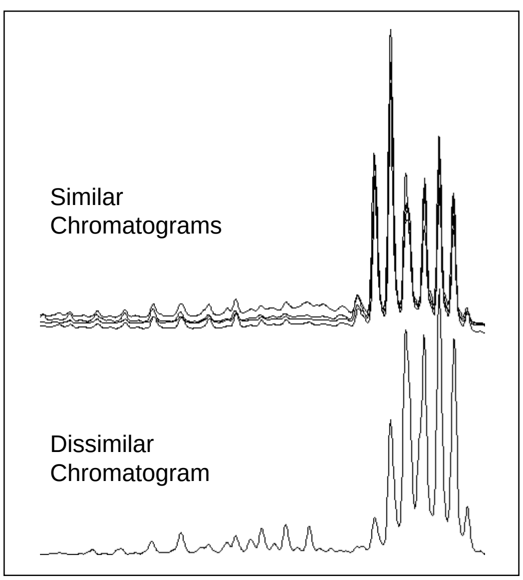

Figure 1
Classification of chromatograms are based on the relative abundance of all the peaks in the mixture.

Although it is not always an expressed goal of a chromatographic analysis, we commonly use human pattern recognition skills to interpret the instrument output. The purpose of this pattern recognition step is usually to classify the sample in some way (*e.g.*, is the sample of acceptable quality or is the sample consistent with a previous run?). Through the methods development process, we often strive to develop a set of rules-of-thumb for interpreting patterns. Often these heuristics involve calculating the ratio of two intensities or developing a simple decision tree based on a series of features in the chromatographic trace.

This overview describes a series of applications in which pattern recognition software has simplified methods development and automated the routine use of robust pattern matching in chromatography. The field of study which encompasses this technology is called chemometrics and the examples cited can be duplicated using Pirouette® multivariate modeling software.

A chromatogram can be thought of as a chemical fingerprint where the pattern emerges from the relative intensities of the sequence of peaks passing by the detector. Chromatographic fingerprinting, whether by human intervention or automated in software, is used in two generic application areas:

- *to infer a property of interest (typically adherence to a performance standard); or*
- *to classify the sample into one of several categories (good versus bad, Type A versus Type B versus Type C ...).*

The following sections contain examples of the use of chemometric technology to problems in chromatographic pattern recognition, with applications drawn from different industries.

Chromatography Applications

Pharmaceutical/Biotech

Protein mapping for product quality control

Grading of raw materials

Drug identification (1)

Much of the research and the quality control effort is aimed at assessing a product's consistency or identifying changes in process parameters that may lead to a degradation of quality standards. In most cases, no single concentration is sufficient to categorize samples for QC purposes. As newly bioengineered forms of products make their way to the market, the lack of standards will drive a further need for pattern recognition technology for batch-to-batch product control.

Medical/Clinical

Identification of microbial species by evaluation of cell wall material (2, 3)

Cancer profiling and classification

Predicting disease state (4-6)

A prime concern of clinical diagnosis is to classify disorders rapidly and accurately. Chemometric techniques can be applied to chromatographic data to develop models allowing clinicians to distinguish among disease states based on the patterns in body fluids or cellular material.

All living systems consist of chemical compounds and the relative distribution of these constituents can be used as a biological fingerprint to type samples. Bacteria, yeast and molds are commonly classified using matching techniques on chromatographic patterns. One example is the identification of the organism causing tuberculosis and related mycobacterial species using HPLC.

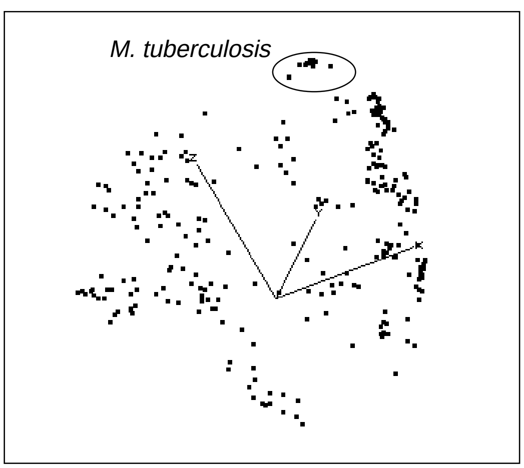

Figure 2
M. tuberculosis can be identified by examining mycolic acid distribution in bacterial cell walls. This figure shows a 3D representation of samples of more than 30 species.

Food/Beverage

Replacing sensory evaluation with instrumented analysis

Geographical/varietal origin

Competitor evaluation (change in process, constituents)

Beer, wine quality control, classification (7-10)

Proving economic fraud (11)

A constant issue in the food industry is the analysis of raw materials and finished products to insure consistency and quality. Chromatographic profiling is useful is detecting changes in a process or in the ingredients and can also be used to monitor plant-to-plant product variations.

A second thrust in the food and beverage industry is to bring analytical instrument techniques to play in sensory evaluation. Traditional sensory panels are expensive to maintain and can lead to inconsistent conclusions. This subjective approach to quality control can be (to some extent) replaced or enhanced by adding the more objective chromatography/chemometrics technique. One example is the profiling of seafood samples to detect where economic fraud (substitution of a lower quality product) has occurred without resorting to a visual inspection by a seafood expert.

Chemical/Petroleum

Oil exploration (oil-oil correlation, oil-source rock correlation) (12)

Refinery QC (product uniformity, raw material variation)

Organic geochemistry often involves the chromatographic analysis of hydrocarbon extracts from geologic formations or oil samples. The patterns reflected in the chromatograms are a combination of biological origin and any geologic alteration. Interpretation of the chromatographic traces can be automated using chemometrics.

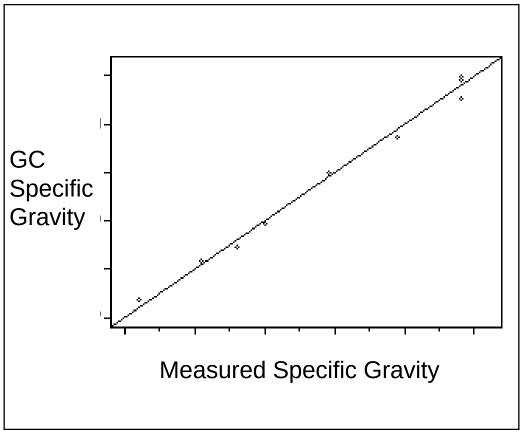

Figure 3
Physical properties, such as the specific gravity of jet fuel, can be determined via calibration of the GC trace to a density measurement.

Environmental

Evaluation of trace organics and pollutants (13,14)

Pollution monitoring where multiple sources are present

Effective extraction of information from large environmental databases

Environmental studies constitute a large portion of the research and monitoring money spent in the world today. This expenditure reflects the concern for the effect chemicals have on the health of the earth's eco system. A typical data set involves the collection of a large amount of data from a diverse set of instrument sources. Chromatography plays a central role in this data assembly because of its sensitivity and specificity for many of the organic compounds of interest.

Chemometric techniques provide the means to extract usable information from the environmental measurements. Through these pattern recognition and modeling techniques, improved descriptions of pollution patterns and their sources are available to the analyst (15).

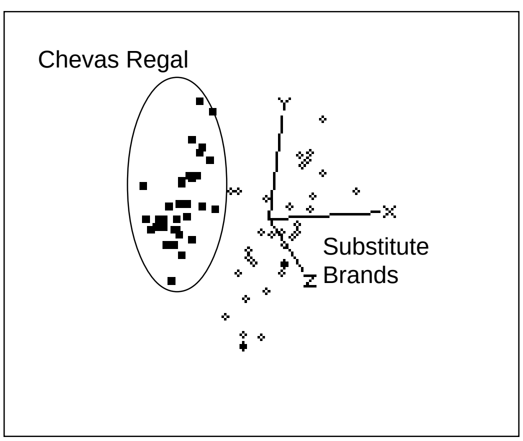

Figure 4
Whisky samples can be classified based on the relative composition of trace constituents.

Forensics

DNA fingerprinting

Arson investigation

Geographical origin of illegal substances

In forensic analysis, the issue is not to determine the concentration of various chemical constituents, but rather to determine if a chromatographic trace is correlated to a known sample. Chemometric pattern matching has been used in a wide variety of applications where the origin of a sample is in question.

Summary

Chemometrics can be used to accomplish a variety of goals in the chromatography laboratory:

- *Speeding of methods development*
- *More effective multivariate calibration*
- *Detection and monitoring of impurities*

Today's chromatographers have jobs to do that extend beyond the act of collecting, analyzing and harvesting reports on individual samples. The true product of the analytical endeavor lies in the consolidation of these individual analyses into an evaluation of the chemical system as a whole. We compare a new sample against a compilation of our prior experience, we try to infer properties of interest with non-specific analytical tools, etc.

Chemometrics can be used to condense large assembly projects into more manageable time frames; the modeling capability allows you to speed methods development and interpretation of complex chromatographic patterns. The multivariate models can be placed in an expert-system context to allow robust implementation of very customized chromatographic systems (16-19).

The Pirouette software is designed to recognize patterns in virtually any type of analytical data. The process can be used to speed methods development and make routine the use of multivariate statistical models. The examples described in this note are easily duplicated or can be used as analogies for custom analyses.

Selected Chromatography References

(1) Musumarra, G.; Scarlata, G.; Romano, G.; Cappello, G.; Clementi, S. and Giulietti, G. "Qualitative Organic Analysis. Part 2. Identification of Drugs by Principal Components Analysis of Standardized TLC Data in Four Eluent Systems and of Retention Indices on SE 30." J. Anal. Toxicology (1987) 11 (Jul./Aug.): 154–163.

(2) Engman, H.; Mayfield, H.T.; Mar, T. and Bertsch, W. "Classification of bacteria by pyrolysis–capillary column gas chromatography–mass spectrometry and pattern recognition." J. Anal. Appl. Pyrolysis (1984) 6 (2): 137–156.

(3) Butler, W.R.; Jost, K.C. and Kilburn, J.O. "Identification of mycobacteria by high performance liquid chromatography." J. Clin. Microbiol. (1991) 29 (11): 2468-2472.

(4) Kowalski, B.R. "Measurement analysis by pattern recognition" Anal. Chem. (1975) 47:1152A-

(5) Marshall, R.J.; Turner, R.; Yu, H. and Cooper, E.H. "Cluster analysis of chromatographic profiles of urine proteins." J. Chromatogr. (1984) 297: 235–244.

(6) Pino, J.A.; McMurry, J.E.; Jurs, P.C. and Lavine, B.K. "Application of pyrolysis/gas chromatography/pattern recognition to the detection of cystic fibrosis heterozygotes." Anal. Chem. (1985) 57 (1): 295–302.

(7) Moret, I.; Scarponi, G. and Cescon, P. "Aroma components as discriminating parameters in the chemometric classification of Venetian white wines." J. Sci. Food Agric. (1984) 35 (9): 1004–1011.

(8) Moret, I.; Scarponi, G.; Capodaglio, G. and Cescon, P. "Characterization Soave wine by determining the aromatic composition and applying the SIMCA chemometric method." Riv. Vitic. Enol. (1985) 38 (4): 254–262.

(9) Stenroos, L.E. and Siebert, K.J. "Application of pattern–recognition techniques to the essential oil of hops." J. Am. Soc. Brew. Chem. (1984) 42 (2): 54–61.

(10) Van Rooyen, P.C.; Marais, J. and Ellis, L.P. "Multivariate analysis of fermentation flavor profiles of selected South African white wines." Dev. Food Sci. (1985) 10 (Prog. Flavour Res.): 359–385.

(11) Saxberg, B.E.H.; Duewer, D.L.; Booker, J.L. and Kowalski, B.R. "Pattern recognition and blind assay techniques applied to forensic separation of whiskies." Anal. Chim. Acta (1978) 103: 201-212.

(12) Zumberge, J.E. "Prediction of source rock characteristics based on terpane biomarkers in crude oils: A multivariate statistical approach." Geochim. Cosmochim. Acta (1987) 51 (6): 1625-1637.

(13) Dunn, W.J.; Stalling, D.L.; Schwartz, T.R.; Hogan, J.W.; Petty, J.D.; Johansson, E. and Wold, S. "Pattern recognition for classification and determination of polychlorinated biphenyls in environmental samples." Anal. Chem. (1984) 56 (8): 1308–1313.

(14) Onuska, F.I.; Mudroch, A. and Davies, S. "Application of chemometrics in homolog–specific analysis of PCBs." HRC & CC (1985) 8: 747–754.

(15) Breen, J.J. and Robinson, P.E., Eds., <u>Environmental Applications of Chemometrics</u> ACS Symposium Series (1985) 292: 286pp.

(16) Chien, M. "Analysis of complex mixtures by gas chromatography/mass spectrometry using a pattern recognition method." Anal. Chem. (1985) 57 (1): 348–352.

(17) Isaszegi–Vass, I.; Fuhrmann, G.; Horvath, C.; Pungor, E. and Veress, G.E. "Application of pattern recognition in chromatography." Anal. Chem. Symp. Ser. (1984) 18 (Mod. Trends Anal. Chem., Pt. B): 109–124.

(18) Smith, A.B.,.; Belcher, A.M.; Epple, G.; Jurs, P.C. and Lavine, B. "Computerized pattern recognition: a new technique for the analysis of chemical communication." Science (1985) 228 (4696): 175–177.

(19) Stepanenko, V.E. "Group analysis and pattern recognition as a basis for chromatographic identification." Zh. Anal. Khim. (1985) 40 (5): 881–886.

Infometrix, Inc., 10634 E. Riverside Dr., Bothell, WA Phone: (425) 402-1450, Fax: (425) 402-1040, email: info@infometrix.com