Chemometrics

Product Note

Automating the Interpretation of Chromatograms

Summary

Chromatography is an extremely versatile technique for the analytical or quality control laboratory. As such, chromatograms are collected for a wide variety of both quantitative and qualitative assessments. In many cases, the interpretation of chromatographic output is subject to manual interpretation. In the past, we would rely on holding two printed chromatograms up to a window, sliding them back and forth to see if a process or product has undergone "significant" change. Chemometrics offers an objective and automated mechanism to replace the human pattern recognition step which can provide reliable answers immediately upon completion of the instrumental analysis. The skilled analyst can then apply their knowledge to the interpretation of the information generated by the automated chemometric engine.

The Infometrix® software packages Pirouette®, InStep™ and LineUp™ allow users to make chromatographic pattern characterization systems without much customization. The current products build on the experience gained from building several custom systems, spanning HPLC, GC and GC/MS for the FDA, the CDC and several food and petroleum companies, automating interpretation for instruments from Agilent, PE, Waters, Shimadzu and others working with GC, HPLC, HPCE, TLC and gel electrophoresis. Thoughts on how to approach multivariate analysis of chromatographic profiles are described in this note.

These steps represent the basics needed to produce an automated multivariate prediction system:

1. *Collect the experience set;*
2. *Build Pirouette models;*
3. *Set up the InStep method and report;*
4. *Configure the instrument software to save files into a folder watched by InStep.*

Use of Peak Table or Whole Chromatogram

In essence, there are two ways to look at the output of a chromatograph: use the peak table generated by the chromatographic software or look at the chromatogram as if it were a spectrum (*i.e.,* use the whole chromatographic profile).

Peak tables have to be constrained in order to work in the automated world. The peaks making up the signature have to be the same peaks from analysis to analysis and suffer from two problems, retention time shifting leading to misidentification and the fact that any unexpected peak would not be processed. From our experience, the primary problem is the unexpected peak.

Using the whole profile solves the unexpected peak problem and handles shoulders and other chromatographic features, but it, like in spectroscopy, can suffer significantly from x-axis (in this case retention time) instability. Pirouette has built-in the ability to correct for retention drift, dramatically improving the quality of whole-profile matching. In addition, LineUp, accessed via the supplied LineUpGUI, provides a stand-alone utility with batch capability for aligning many chromatograms at once.

Applications

Batch Quality Control

The manufacture of synthetic biological materials is treated as a batch process.

Because of the complexity of the product, quality control is often conducted using capillary electrophoresis. The pattern is complex and the majority of the fragment peaks are not identified. There are changes in the chromatogram or electropherogram from batch to batch. Pattern recognition technology provides a simple means to determine what variation is acceptable (no impact on the efficacy or the safety) in the case where the tolerance for variability is not uniform for all peaks.

Other applications -

- *Residual solvent patterns can detect process changes*
- *Product quality can be graded based on more than just the "favorite ratio"*
- *QC runs can be corrected for retention time shift prior to archival*

Product Origin

A state Department of Justice (DOJ) has built a database for comparison of complex chromatograms obtained from clandestine lab (methamphetamine) samples. Their knowledge base consists of GC or GC/MSD data from 100 laboratories in the state. When a new sample is found, it is run against the database to see if there is a match. From the chromatographic pattern, they gain information about the source and the process used to manufacture the illegal drug.

Other applications -

- *Process GCMS to characterize PCB distribution and search against Aroclor patterns*
- *Use GC/FID to identify the source of spills in a refinery*
- *HPLC fingerprinting of herbs, fruits, etc., to determine origin or shipping history*

Mining an Experience Set

A food company has a continual need to check samples for identity and compare samples with standards for quality control or supports of other applications. From a library of 3,000

chromatograms, Pirouette was used to establish an optimal, non-redundant database of good QC samples. Now in routine use, InStep evaluates every new sample; when the software flags the product as being out of class, then Pirouette is used to see why.

Other applications -

- *M. tuberculosis can be identified from a database containing 60+ related species*
- *Source rocks for crude oil can be typed based on GC and GC/MS*
- *Chromatograms collected over a multi-year timeframe can help classify aberrant samples*

Quantitating Change

A major application for this software is the comparison of peptide maps resulting from the enzymatic digestion of recombinant proteins. Peptide mapping is a sensitive technique that biotechnology companies use to assess the quality of their protein products. Chromatographic pattern recognition allows detection of 1% or smaller changes in the composition of the products, thereby helping to meet guidelines for well-characterized biopharmaceuticals.

Other applications -

- *Quantitating the degree of weathering for fuel classes in arson investigation*
- *GC column QC to identify problems before they cause data loss*
- *Identifying seasonal variation in blended products*
- *Differentiate between a chromatographic problem (all of the significant variables are expected to shift) and an impurity problem (only some of the variable shift)*

Mixture Analysis

A spice that is a mixture of various natural ingredients can pose the problem that the product is supposed to stay consistent from year-to-year, but the ingredients can vary. What is the best mixture to reproduce the standard product? Can a new ingredient (source or material) be substituted in the formulation? Note, this technology was commercialized (by Infometrix) first as an add-in to HP, Waters and Beckman HPLC gear starting in 1983.

Other applications -

- *Quantitating relative composition after backblending kerosene into diesel #2*
- *Identifying end members of mixtures (binary, ternary +) based on the GC pattern*
- *GCMS peaks can be unmixed to quantitate impurities, even in the absence of standards*
- *Deconvolute HPLC peaks based on UV spectra, GC peaks based on MS spectra*

How to Start

These applications and more can be readily produced using Pirouette, LineUp and InStep, available on CD or as a download from the Infometrix web site. If you want assistance in building a turn-key system, Infometrix can supply this service usually within a week. Training and custom application development are also available.