Chemometrics in Environmental Science

Scientists involved in environmental studies are faced with many different analytical tasks, such as assembling baseline studies, evaluating the contributing influence of chemical discharge to complex natural systems, and modeling biological response. Industrial scientists are concerned with the mechanics of recycling materials and maintaining process control systems that minimize pollution. Governmental control agencies, such as the EPA, are interested in detecting the presence of specific environmental agents, as well as assessing environmental damage from human sources.

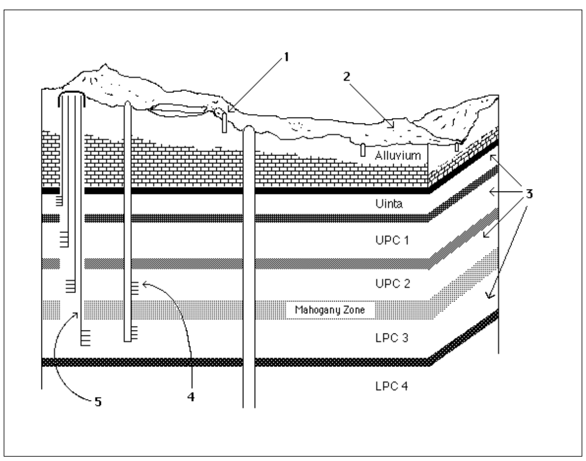

Figure 1 Chemometric fingerprints are used to detect surface discharge, characterize an aquifer's profile and determine the extent of stratigraphic leakage (11).

Sometimes the problem is simply the enumeration of the presence or absence of constituents, whether natural or introduced. Other concerns deal with the influence that environmental factors will have on systemic response. In-field monitors and laboratory instrumentation may not directly measure these influence factors. Therefore, we are forced to make measurements of an indirect set of variables which may be only weakly correlated to the properties of interest in the system. Convenient and powerful multivariate methods have proven useful in managing and analyzing these types of complex problems.

For example, envision a chromatogram or spectral profile of a sediment extract as a fingerprint of the constituents in the sample. The pattern represents the varying amounts of the individual chemicals present. The variation contained in the signature patterns of these samples from multiple sites can reveal chemical relationships which can be characteristic of known natural phenomena or identified pollution sources.

Chemometrics is a multivariate mathematical and statistical approach to the analysis and interpretation of analytical data. Pattern recognition methods have been used in chemometrics to reveal and evaluate complex relationships in a wide variety of environmental applications. These methods have contributed to the systematic understanding of sediment trace metal and organic concentrations arising from natural and anthropogenic sources. Chemometrics is also useful in evaluating biological response to natural or toxic factors, and can identify the source of the contamination. Common uses of this technique are to:

- *identify factors that are combinations of measurable variables;*
- *illustrate groups or cluster associations among samples;*
- *assess spatial distribution of environmental factors or perturbations; and*
- *predict a property of interest (such as biological response to chemical perturbation).*

Environmental Applications

This overview describes a series of applications in which chemometrics software has contributed to the understanding of complex environmental systems. The examples cited can be duplicated using Pirouette® multivariate modeling software and automated for routine analysis using InStep™.

Atmospheric and Sediment Processes

Distribution of natural or toxic chemicals (1-3)

Source identification, regional influence (4-6)

Abatement and control (7, 8)

Pollution modeling is generally pointed toward the identification of man-made sources of chemicals. By understanding the spatial and temporal variation of these pollutants, control measures can be applied to bring levels into compliance with environmental standards. Multivariate chemometric modeling techniques can help in situations where the emission of chemicals is added to natural sources of these same materials. The problem becomes one of discriminating the relative contributions of natural and human influence.

Chemometrics can be used to discern structure in a dataset as a whole, even when individual measurements show only slight degrees of correlation. The most common use of the technology is to apportion the sources of pollution. In atmospheric studies, the relative impact of nature (such as the suspension of sea salt, or impact of forest fires) can be contrasted with suspended road dust, automotive emissions, and specific industrial contributions. Similarly, sediment studies can confirm the presence of chemicals in excess of what would be expected to occur naturally.

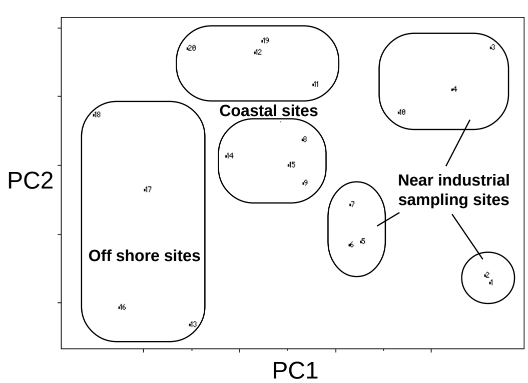

Figure 2
Polluted versus non-polluted environmental samples can be reliably classified along with determining the source of contamination (15).

The idea behind chemometric analysis is that you can effectively attribute a source to an environmental contaminant without the need to find specific marker compounds. By evaluating all of the data at once, complex data can be reduced to a set of interpretable patterns without making *a priori* assumptions about the cause of the perturbation. Not only does chemometrics supply an effective key to interpretation, the analysis yields predictive models that can be used successfully on a routine basis.

Water Management

Pollution assessment and control (9-11)

Nutrient sources and dynamics (12,13)

Trophic studies (14)

Inevitably, the vast majority of the waste generated by man finds its way into surface and subsurface water. Industrial and municipal effluent is pumped into bodies of water directly and the contaminants dispersed in the air or in sediments eventually is partitioned into the water table, lakes, rivers and seas. Tracking the migration of water pollution and assessing enrichment or scavenging ratios is often far more complicated than in atmospheric or sediment studies. The complication stems both from the extreme diversity of sources and from the complexity of effects as the new materials are introduced.

A highly useful aspect of chemometric modeling is that data used to generate patterns is not restricted to a single instrument source. Rather, the technology allows the combination of data from a variety of instrument systems as well as wet chemistry, biology and general descriptive data. An offshoot of the analysis is that, when a model is generated, you have the opportunity to assess the contributing value of individual pieces of your database relative to the inherent information content it brings to the problem. This can lead to more efficient data collection methods in future studies.

Biological Response Modeling

Ecology and toxicity (15-18)

Predicting species growth (19)

Impact on health, tissue analysis (20-22)

While the modeling of pollution assists the evaluation and control of chemical factors in the environment, biological response modeling transforms

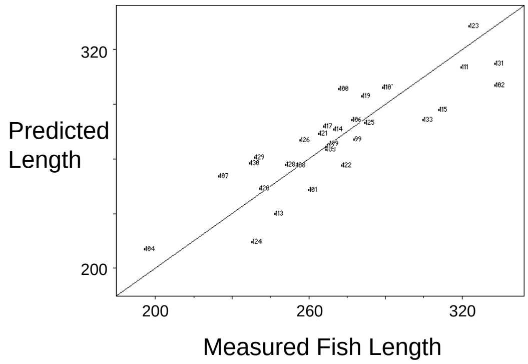

Figure 3
Species growth patterns can be predicted based on the dynamics and the nutrient parameters of lake systems.

these abstracted variables into potential health consequences. Chemometrics can allow a more efficient detailing of the influences of natural and foreign factors on the well being of specific biological systems. Furthermore, vital correlations among secondary factors may exist and be underappreciated without using this approach.

Biological response to environmental contaminants is an extremely complex process. Multivariate classification and calibration methods are particularly well-suited to extracting predictive information from a set of measurements that individually may show only a small correlation to the property of interest. Through the use of multivariate models, we can be more effective in charting relationships between species diversity or growth patterns and climatological or pollution variables. Chemometric assessments have also been used to correlate environmental factors to specific human health concerns.

Industrial Maintenance and Process Control

Sorting of recycled material (23)

Optimizing plant effluent (24, 25)

Assess and control hazardous waste (26, 27)

Industry is often portrayed as the perpetrator of pollution arising from the manufacturing process. In contrast to the popular image, most industrial plants engage in a continuous, serious effort to reduce waste (a non-economic by-product). The benefits are clear for a manufacturer to remain in compliance with regulatory agencies and to provide an increase in manufacturing efficiency.
The primary advantage of the chemometric approach in industrial settings is the relative ease of implementing a highly-focused instrument system for monitoring the quality of a product or raw material. Most instruments sold are general purpose devices designed to generate data, but will not supply the desired information directly. A spectrophotometer can give a spectrum of a piece of plastic, but it does not specify whether it is PVC, PET, etc. Chemometrics software acts as an intermediary, interpreting the spectrum in this case, to provide the exact information desired. There is no need to build (and pay development costs for) a specific sensor system, when a general purpose instrument can be rapidly turned into a source of highly specific quality control information through a chemometric calibration process.

Summary

Environmental scientists are charged with collecting and evaluating complex, inexplicit data to understand and solve concrete problems. Chemometrics is a discipline which utilizes multivariate statistical techniques, directly correlating variations in natural or toxic materials to their environmental response. Patterns in the physical and chemical data are modeled, and the models can be routinely applied to future data in order to predict comparative consequences.

Chemometrics software, such as Pirouette, is designed to recognize patterns in virtually any type of multidimensional analytical data. Chemometrics can be used to speed methods development and make routine the use of statistical models for data analysis. Specifically, the application of chemometrics to environmental analysis can result in:

- *Detection of pollution contributions from a complex mixture of sources;*
- *Assessment of geographical or atmospheric distributions and influences;*
- *Prediction of biological response to perturbation;*
- *Understanding the interplay of influential factors which cannot be directly measured;*
- *Optimization of processes for controlling plant waste or recycling; and*
- *Improvement in the interpretability of analytical instrument data.*

Selected References

(1) Gomez, M.L.S. and Martin, M.C.R. "Application of cluster analysis to identify sources of airborne particles" Atmos. Environ. (1987) 21 (7): 1521-1527.

(2) Hooper, R.P. and Peters, N.E. "Use of multivariate analysis for determining sources of solutes found in wet atmospheric deposition in the United States." Environ. Sci. Technol. (1989) 23:1263-1265.

(3) Wenning, R.J.; Harris, M.A.; Ungs, M.J.; Paustenbach, D.J. and Bedbury, H. "Chemometric comparisons of polychlorinated dibenzo-p-dioxin and dibensofuran residues in surficial sediments from Newark Bay, New Jersey and other industrialized waterways." Arch. Environ. Contam. Toxicol. (1992) 22: 397-413.

(4) Boni, C.; Caruso, E.; Cereda, E.; Lombardo, G.; Braga Marcazzan, G.M. and Redaelli, P., "Particulate matter elemental characterization in urban areas: pollution and source identification." J. Aerosol Sci. (1988) 19 (7):1271-1274.

(5) Cohen, M.A.; Ryan, P.B.; Spengler, J.D.; Ozkaynak, H. and Hayes, C. "Source-receptor study of volatile organic compounds and particulate matter in the Kanawha Valley, WV - II. analysis of factors contributing to VOC and particle exposures." Atmos. Environ. (1991) 25B (1): 95-107.

(6) Winters, G.V. and Buckley, D.E., "Factor analyses as a method of evaluating spatial and temporal variations in sediment environmental quality in Halifax Harbour." Pittsburgh Conference Paper #93-480 (1993) Atlanta, GA.

(7) Kim, D.S.; Hopke, P.K.; Massart, D.L.; Kaufman, L. and Casuccio, G.S., "Multivariate analysis of CCSEM auto emission data." Sci. Total Environ. (1987) 59:141-155.

(8) Vong, R.; Geladi, P.; Wold, S. and Esbensen, K. "Source contributions to ambient aerosol calculated by discriminant partial least squares regression." J. Chemometrics (1988) 2: 281-296.

(9) Bomboi, M.T.; Hernandez, A.; Marino, F. and Hontoria, E. "Application of multivariate analysis for characterization of organic compounds from urban runoff." Sci. Total Environ. (1990) 93: 523-536.

(10) Higashi, K. and Hagiwara, K. "Identification of oil spilled at sea by high performance gel permeation chromatography pattern recognition" Wat. Sci. Tech. (1988) 20 (67): 55-62.

(11) Meglen, R.R. and Erickson, G.A. "Application of pattern recognition to the evaluation of contamination from oil shale retorting" in <u>Environment and Solid Waste, Characterization, Treatment and Disposal</u> C.W. Francis and S.I. Auerbach eds., Butterworth Publishers, Boston (1983): pp369-381.

(12) Grimalt, J.O.; Olive, J. and Gomez-Belinchon, J.I. "Assessment of organic source contributions in coastal waters by principal component and factor analysis of the dissolved and particulate hydrocarbon and fatty acid contents" Intern. J. Environ. Anal. Chem. (1990) 38: 305-320.

(13) Vong, R.J.; Hansson, H.; Ross, H.B.; Covert, D.S. and Charlson, R.J. "Northeastern Pacific submicrometer aerosol and rainwater composition: a multivariate analysis." J. Geophys. Res. (1988) 93 (D2):1625-1637.

(14) Lamparczyk, H.; Ochocka, R.J.; Grzybowski, J.; Halkiewicz, J. and Radecki, A. "Classification of marine environment samples based on chromatographic analysis of hydrocarbons and principal component analysis" Oil & Chem. Pollution (1990) 6:177-193.

(15) Macdonald, C.R.; Norstrom, R.J. and Turle, R. "Application of pattern recognition techniques to assessment of biomagnification and sources of polychlorinated multicomponent pollutants, such as PCBs, PCDDs and PCDFs" Chemosphere (1992) 25 (1-2):129-134.

(16) Simmleit, N. and Schulten, H. "Pattern recognition of spruce trees: an integrated, analytical approach to forest damage" Environ. Sci. Technol. (1989) 23:1000-1006.

(17) Stalling, D.L.; Norstrom, R.J.; Smith, L.M. and Simon, M. "Patterns of PCDD, PCDF, and PCB contamination in Great Lakes fish and birds and their characterization by principal components analysis" Chemosphere (1985) 14 (6-7): 627-643.

(18) Von Rudloff, E.; Lapp, M.S. and Yeh, F.C. "Chemosystematic study of Thuja plicata: multivariate analysis of leaf oil terpene composition", Biochemical Systematics and Ecology (1988)16 (2):119-125.

(19) Moseholm, L. "Analysis of air pollution plant exposure data: the soft independent modelling of class analogy (SIMCA) and partial least squares modelling with latent variable (PLS) approaches", Environ. Pollution (1988) 53: 313-331.

(20) Brakstad, F. "A comprehensive pollution survey of polychlorinated dibenzo-p-dioxins and dibenzofurans by means of principal component analysis and partial least squares regression" Chemosphere (1992) 24 (12):1885-1903.

(21) Diaz-Caneja, N.; Gutierrez, I.; Martinez, A.; Matorras, P. and Villar, E. "Multivariate analysis of the relationship between meteorological and pollutant variables and the number of hospital admissions due to cardio-respiratory diseases" Environ. International (1991) 17: 397-403.

(22) Vogt, N.B. "Polynomial principal component regression: an approach to analysis and interpretation of complex mixture relationships in multivariate environmental data" Chem. Intell. Lab. Sys. (1989) 7: 119-130.

(23) Alam, M.K. and Stanton, S.L. "Sorting of waste plastics using near-infrared spectroscopy and neural networks" Process Control and Quality (1993) 4: 245-257.

(24) Brown, S.D.; Skogerboe, R.K. and Kowalski, B.R., "Pattern recognition assessment of water quality data: coal strip mine drainage" Chemosphere (1980) 9: 265-276.

(25) Gonzalez, M.J.; Fernandez, M.; Jimenez, B. and Hernandez, L.M. "Application of Multivariate Analysis to the Determination of PCBs in Fly Ash From Municipal Incinerators" Chemosphere (1989) 18 (11/12): 2213-2227.

(26) O'Brien, R.; Sinha, B.K. and Smith W.P. "A Statistical Procedure to Evaluate Clean-up Standards" J. Chemometrics (1990) 5 (3): 249-261.

(27) Sarker, M.; Glen, W.G.; Yin, L.; Dunn III, W.J.; Scott, D.R. and Swanson, S. "Comparison of simca pattern recognition and library search identification of hazardous compounds from mass spectra" Anal. Chim. Acta (1990) 257: 229-238.

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