

# HAZWRAP

HAZARDOUS WASTE REMEDIAL ACTIONS PROGRAM

Leading the Way to Environmental Stewardship

## CHEMOMETRICS REVIEW FOR CHEMICAL SENSOR DEVELOPMENT TASK 7 REPORT

May 1994



Sponsored by  
U.S. Department of Energy  
Oak Ridge Field Office

Prepared by  
Advanced Sciences, Inc.  
6739 Academy Road, N.E.  
Albuquerque, NM 87109-3345  
under contract DE-AC05-87OR21706  
Task 9750.K03  
TTP No. OR-121112  
for

HAZARDOUS WASTE REMEDIAL ACTIONS PROGRAM  
Environmental Restoration and Waste Management Programs  
Oak Ridge, Tennessee 37831-7606  
managed by MARTIN MARIETTA ENERGY SYSTEMS, INC.  
for the DEPARTMENT OF ENERGY  
under contract DE-AC05-84OR21400

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RECEIVED  
JUN 20 1995  
OSTI

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## **DISCLAIMER**

**Portions of this document may be illegible  
in electronic image products. Images are  
produced from the best available original  
document.**

**CHEMOMETRICS REVIEW  
FOR CHEMICAL  
SENSOR DEVELOPMENT  
- TASK 7 REPORT -**

**May 1994**

**Prepared by:  
Advanced Sciences, Inc.  
6739 Academy Road, N.E.  
Albuquerque, NM 87109-3345**

**Under Contract DE-AC05-87OR21706  
Task 9750.K03  
TTP No. OR-121112**

**Prepared for:  
Hazardous Waste Remedial Actions Program  
Oak Ridge, TN 37831-7606**

**Managed by:  
Martin Marietta Energy Systems, Inc.  
for the  
U.S. Department of Energy  
Under Contract DE-AC05-84OR21400**

2000  
2000

## CONTENTS

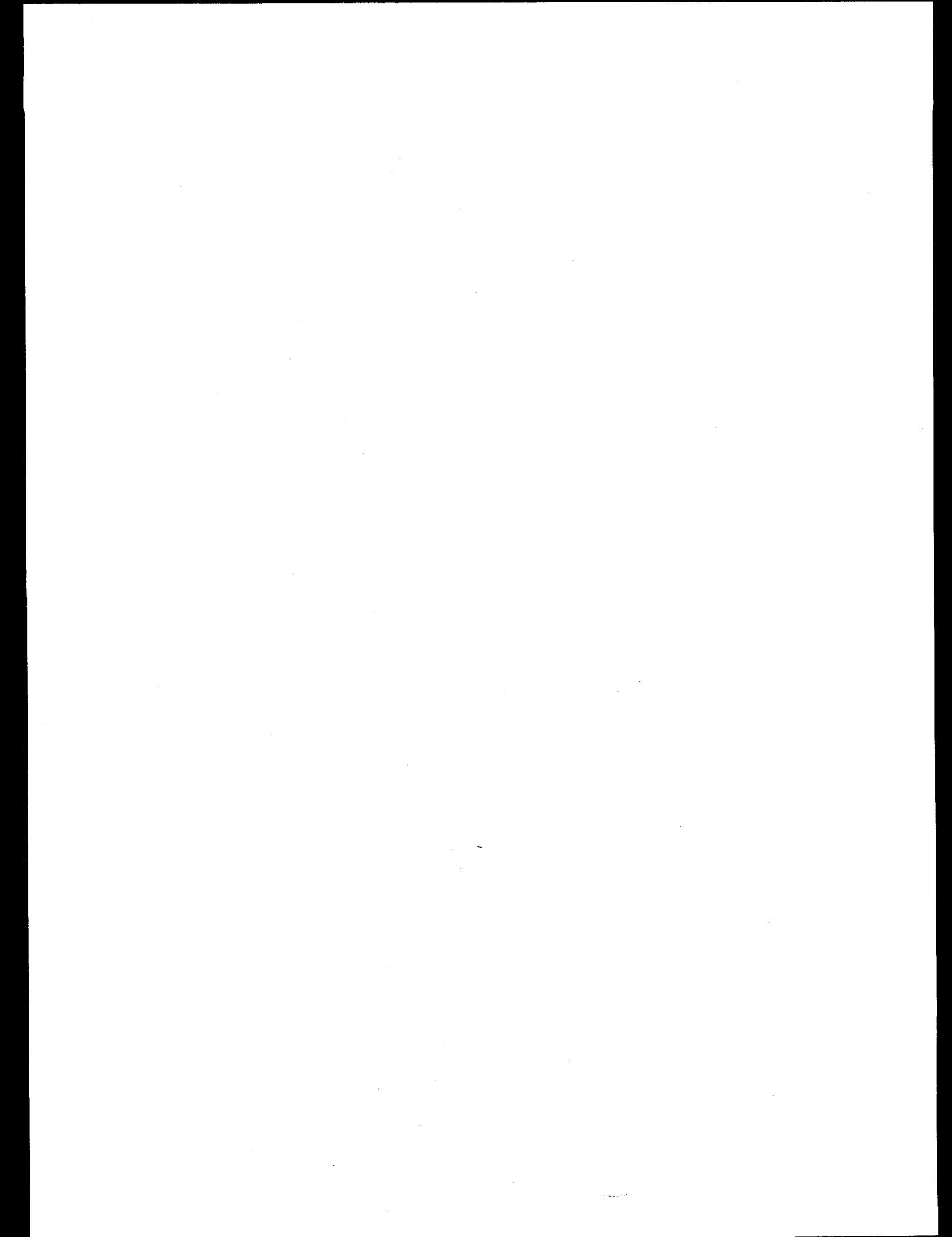
Chapter	Page
PREFACE .....	v
ABSTRACT .....	vii
SUMMARY .....	ix
1. INTRODUCTION .....	1-1
1.1 REFERENCES .....	1-2
2. CHEMOMETRICS SUMMARY .....	2-1
2.1 TERMINOLOGY .....	2-1
2.2 EXPLORATORY DATA ANALYSIS .....	2-2
2.3 CALIBRATION AND PREDICTION .....	2-3
2.3.1 Calibration in Univariate Analysis .....	2-4
2.3.2 First-Order Calibration .....	2-5
2.3.3 Second-Order Calibration .....	2-6
2.4 STATUS OF TOOLS FOR ANALYSIS .....	2-7
2.5 STATUS OF THE FIELD OF CHEMOMETRICS .....	2-8
2.5.1 General Perspective of Chemometrics .....	2-8
2.5.2 Towards Statistical Discipline and Data Stewardship Standards .....	2-10
2.6 SOFTWARE, COMMERCIAL AND OTHER .....	2-12
2.6.1 Commercial Software .....	2-12
2.6.2 Other Software (Shareware, Freeware, Bundled Bookware) .....	2-17
2.7 REFERENCES .....	2-17
3. APPLICATIONS OF CHEMOMETRIC TECHNIQUES TO CHEMICAL SENSOR ARRAYS .....	3-1
3.1 ELECTROCHEMICAL SENSOR ARRAYS .....	3-1
3.2 SEMICONDUCTOR SENSOR ARRAYS .....	3-7
3.3 SAW SENSOR ARRAYS .....	3-12
3.4 GENERAL ARTICLES .....	3-19
3.5 REFERENCES .....	3-22

## CONTENTS (continued)

<b>Chapter</b>		<b>Page</b>
4.	REMARKS/CONCLUSIONS AND RECOMMENDATIONS .....	4-1
4.1	REMARKS/CONCLUSIONS .....	4-1
4.1.1	What Is Known in the Beginning? .....	4-1
4.1.2	Validation and Verification in Applied Chemometrics .....	4-1
4.1.3	Chemometrics Camps .....	4-1
4.1.4	Chemometrics/Modern Computer-Based Applications to Other Aspects of ERWM .....	4-2
4.1.5	Chemometrics for DOE ERWM Problems .....	4-4
4.2	RECOMMENDATIONS .....	4-4
4.3	REFERENCES .....	4-5
 <b>APPENDICES</b>		
A	CHEMOMETRICS BIBLIOGRAPHY .....	A-1
B	SCIENTIFIC AND TECHNICAL INFORMATION NETWORK BIBLIOGRAPHY .....	B-1

## PREFACE

This work was conducted by William H. Roach, Advanced Sciences, Inc. (ASI), Principal Investigator, and H. Milton Peek and Ernest Fickas (consultants). The work was performed under Contract DE-AC05-87OR21706, Task 9750.K03, TTP No. OR-121112, with the Hazardous Waste Remedial Actions Program (HAZWRAP) of Martin Marietta Energy Systems, Inc. Frederick Heacker served as the HAZWRAP Project Manager for this task. The work was sponsored by Caroline Purdy, Office of Technology Development, Office of Environmental Restoration and Waste Management, Department of Energy Headquarters.



## ABSTRACT

This report, the seventh in a series on the evaluation of several chemical sensors for use in the U.S. Department of Energy's (DOE's) site characterization and monitoring programs, concentrates on the potential use of chemometrics techniques in analysis of sensor data.

Chemometrics is the chemical discipline that uses mathematical, statistical, and other methods that employ formal logic to:

- design or select optimal measurement procedures and experiments and
- provide maximum relevant chemical information by analyzing chemical data.

The report emphasizes the latter aspect. In a formal sense, two distinct phases are in chemometrics applications to analytical chemistry problems: (1) the exploratory data analysis phase and (2) the calibration and prediction phase. For use in real-world problems, it is wise to add a third aspect—the independent validation and verification phase. In practical applications, such as the ERWM work, and in order of decreasing difficulties, the most difficult tasks in chemometrics are:

- establishing the necessary infrastructure (to manage sampling records, data handling, and data storage and related aspects),
- exploring data analysis, and
- solving calibration problems, especially for nonlinear models.

Chemometrics techniques are different for what are called zeroth-, first-, and second-order systems, and the details depend on the form of the assumed functional relationship between the measured response and the concentrations of components in mixtures. In general, linear relationships can be handled relatively easily, but nonlinear relationships can be difficult. These difficulties are the subject of a great deal of current chemometrics research.

The applications of chemometrics techniques in the analysis of mixtures using arrays of chemical sensors has met with some success, but almost all of the published work is based on a great deal of *a priori* knowledge about the qualitative and quantitative characteristics of the mixtures being analyzed (such mixture problems are called "white"). By contrast, *a priori* knowledge about most of the real-world DOE ERWM problems is likely to be very meager, approaching what has been called a "black" mixture problem, implying no *a priori* qualitative or quantitative knowledge. Black problems have been tackled using the hyphenated analysis techniques of analytical chemistry.

Chemometrics has significant potential in solving DOE's ERWM analytical chemistry problems, but determination of the validity of this opinion requires that a chemometrics capability be established and used to tackle the real problems (i.e., no set of "chemometrics lore" exists from which one can learn whether the techniques would work for any given problem). A syllogism

expresses the situation: if you want real answers to chemometrics analysis problems, you must pose real problems. If you want to evaluate the methods, you must try them. If you want the benefits of chemometrics, you must have a well-equipped and active chemometrics team.

The authors recommend establishment of advisory panels tasked to plan coordinated programs for sensor array and chemometrics developments. In addition, it is suggested that a research program be established to carry out mixture analysis research for the various types of DOE ERWM problems.

## SUMMARY

The status of chemometrics applications to the Department of Energy's (DOE's) Environmental Restoration and Waste Management Program's (ERWM's) problems can be summarized as follows:

- The multivariate analysis methods in chemometrics, together with all the appropriate statistics, instrumentation, and engineering – including systems engineering and computer technology – are the only choice for which there is any real hope of solving the ERWM-type problem.
- Many analysis methods suited to environmental problems exist and await application. Application requires creation of certain infrastructure, but potentially many advances can be made by tackling the problem.
- The difficulty of judging the efficacy of analysis methods and the lack of "ground truth" in chemometrics is a strong argument for implementing a dedicated chemometrics analysis effort that is tasked to evaluate the methods and validate the answers.
- A dedicated chemometrics effort may be able to determine which currently available analysis methodologies can provide the results needed for particular types of ERWM problems.

The following syllogism says it well.

- If you want real answers to chemometrics analysis problems, you must pose real problems. If you want to evaluate the methods, you must try them. If you want the benefits of chemometrics, you must have a well-equipped and active chemometrics team.

Recommendations include the following items:

- Establish a sensor development advisory panel tasked to recommend a program for development of sensor arrays for DOE's problems,
- Establish a sensor systems chemometrics advisory panel to recommend a chemometrics development program closely coordinated with the sensor array development program,
- Initiate a research program to extend mixture analysis research to consideration of real-world, highest priority ERWM problems.

## 1. INTRODUCTION

This report is the seventh in a series of analytical studies performed by Advanced Sciences, Inc., under a subcontract with the Hazardous Waste Remedial Actions Program, managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy (DOE). The primary purpose of the overall project is to evaluate the various types of chemical sensors that might be useful in the DOE's Environmental Restoration and Waste Management (ERWM) work. This objective is being approached in a series of tasks; this report is the results of Task 7.

The first five task reports (1, 2, 3, 4, and 5) contain reviews of pertinent literature and the results of interviews with researchers and manufacturers of the various types of chemical (and biochemical) sensors: acoustic (previously called piezoelectric), electrochemical, and fiber-optical. In addition, the reports include surveys of radiological analyzers and instrumentation and special detectors used in what we called field-portable gas chromatographs.

This report is the second in the series that is not directly concerned with the sensors themselves. The sixth report (6) contains information about the possible uses of cone penetrometers for DOE ERWM site characterization problems. This seventh report addresses what is a crucial part of analytical chemistry problems (regardless of what sensors or instruments are used in the analyses) — namely, the various techniques that have been developed for analysis of analytical chemistry data. These techniques are part of the field that, according to a world-recognized expert, Massart (7), was named **chemometrics** in 1972. According to Massart,

“ . . . chemometrics, . . . which can be defined as the chemical discipline that uses mathematical, statistical, and other methods employing formal logic (a) to design or select optimal measurement procedures and experiments and (b) to provide maximum relevant chemical information by analyzing chemical data.”

Some remarks in Chapter 4 are about the first of the two listed aspects of chemometrics, but the majority of this report is about the second aspect.

This report takes as a basic programmatic assumption that those DOE personnel who are involved in technology development for ERWM work include as part of their responsibility the development of chemometrics techniques for specific applications to ERWM problems involving chemical sensors and advanced analytical chemistry instruments.

In an attempt to develop a better understanding of the subject, we suggest the following scenario. We imagine a person who has been given responsibility for cleanup of some specified site (or distinct area at a larger site). That person's responsibility is to determine the three-dimensional distribution of the concentrations of some list of contaminants, to plan and then carry out the remediation of the area, and to do whatever is necessary to validate that the site has been “cleaned up” (i.e., that the imposed criteria have been met). Focussing only on the analytical chemistry problem (i.e., a three-dimensional mapping of the concentrations of a list of chemicals), in general three different types of analytical chemistry tools can be used. In a very few instances, highly selective chemical sensors may be used. In more instances, there will be situations in

which arrays of chemical sensors can be used. Finally, there is both the field analytical laboratory and the more extensively equipped central analytical chemistry laboratory. In each of these instances analytical chemistry data are to be analyzed. This report summarizes the applications of chemometrics in such a scenario.

In a general way, chemometrics consists of two phases: the exploratory data analysis phase and the calibration and prediction phase. These aspects of the problem are discussed in Chapter 2.

Chapter 3 includes reviews of several papers devoted to the use of chemical sensor arrays. Three types of arrays are reviewed: electrochemical sensor arrays, semiconductor gas sensor arrays, and surface acoustic wave sensor arrays. In addition, we mention what is a very important consideration for DOE ERWM work, the so-called "white, grey, black" mixture analysis problem (8), where "black" denotes no knowledge of chemical composition or concentrations in a sample, and "white" denotes that all chemical components are known and it is also known that the concentrations of the analytes of interest are within the range covered by a complete set of calibrations that can be carried out. As will be evident, all – or almost all – real samples in DOE ERWM work are either "black" or some shade of "grey."

Chapter 4 includes remarks, conclusions, and recommendations.

At the end of each chapter is a list of references cited in that chapter; however, the lists do not include articles reviewed in the chapter. Bibliographies relevant to Chapters 2 and 3 are presented in Appendix A and B, respectively.

## 1.1 REFERENCES

1. *Literature Search, Review, and Compilation of Data for Chemical and Radiochemical Sensors – Task 1 Report*, DOE/HWP-130, Hazardous Waste Remedial Actions Program, Oak Ridge, TN, January 1993.
2. *Literature Search, Review, and Compilation of Date for Chemical and Radiochemical Sensors – Task 2 Report*, DOE/HWP-133, Hazardous Waste Remedial Actions Program, Oak Ridge, TN, April 1993.
3. *Chemical Sensor R&D Status Based on Research and Industrial Interviews – Task 3 Report*, DOE/HWP-138, Hazardous Waste Remedial Actions Program, Oak Ridge, TN, October 1993.
4. *Literature Search, Review, and Compilation of Data for Biosensors and Thermal Sensors – Task 4 Report*, DOE/HWP-144, Hazardous Waste Remedial Actions Program, Oak Ridge, TN, December 1993.
5. *Literature Search, Review, and Compilation of Data for Gas Chromatography Sensors and Electrochemical Sensors – Task 5 Report*, DOE/HWP-148, submitted to Hazardous Waste Remedial Actions Program, Oak Ridge, TN, January 1994.

6. *Review of Instrumentation and Sensors – Cone Penetrometer Applications – Task 6 Report*, DOE/HWP-149, submitted to Hazardous Waste Remedial Actions Program, Oak Ridge, TN, April 1994.
7. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: A Textbook*, Volume 2 of the series *Data Handling in Science and Technology*, Series Editors B. G. M. Vandeginste and L. Kaufman, Elsevier, New York, 1988.
8. Y.-Z. Liang, O. M. Kvalheim, and R. Manne, "White, Grey and Black Multicomponent Systems. A Classification of Mixture Problems and Methods for Their Quantitative Analysis," *Chemom. Intell. Lab. Syst.* 18 235–250 (1993).

## 2. CHEMOMETRICS SUMMARY

As mentioned in Chapter 1, this chapter is a summary of information about the second aspect of the definition of the field of chemometrics (i.e., "to provide maximum relevant chemical information by analyzing chemical data"). By way of explanation, it is important to remember certain terminology; zeroth-order, first-order, and second-order measurement systems; and univariate and multivariate data analysis techniques.

### 2.1 TERMINOLOGY

Imagine an analytical chemistry experiment in which a pH electrode is used to measure the pH of an aqueous sample. Assuming good laboratory practices are followed, the result of a pH measurement on a sample is a single number. Such a measurement system is called zeroth-order (the result is a scalar quantity), and the data analysis techniques used are called univariate techniques.

Another example of a zeroth-order system would be an optical measurement of sample transmittance or absorptance performed over some relatively narrow wavelength range. In this example, the result of the measurement is either a difference or a ratio of transmitted light intensity measured with and without the sample interposed between light source and light detector. Again, for each sample, a single number is the result of the measurement.

However, if in the optical measurement system we use some instrument that can change the wavelength of light incident on the sample, we imagine producing a set of number pairs, each pair being a number related to transmittance or absorbance and a number that specifies the wavelength. Such a system is called a first-order system (the result of a measurement is a vector), and the data analysis techniques used in this and more complex situations are called multivariate analysis techniques. Note that this same situation would hold regardless of how many different kinds of molecules are present in the sample; the result is still a vector composed of response and wavelength elements.

Again, assuming good laboratory measurement practices, if we imagine using a set of chemical sensors each of which is partially selective (i.e., a given sensor response depends on the concentrations of two or more chemicals) then the result of a measurement for each sensor is a vector, and the set of data for all sensors is a matrix. This is yet another example of a first-order measurement system.

To continue the terminology explanation, if we imagine an optical analysis system in which the chemical composition of the sample whose spectrum is being measured is made to change in some controlled fashion, say in time as the result of passing some sample through a chromatograph, then clearly the spectrum would change with time. [We also suppose that the measurement of a spectrum over whatever wavelength range is of interest can be performed in a time interval that is short compared with the time rate of change of composition (i.e., the efflux

rate from the chromatography tube)]. In this situation, the measurement system is called a second-order system, and multivariate analysis techniques are required in the data analysis.

The example described just above is one of several different techniques called "hyphenated techniques." Other examples of second-order/hyphenated systems are gas chromatograph (GC)-mass spectrometer (MS), GC-ion-mobility spectrometer (IMS), MS-MS, even GC-IMS-MS, [which could instead be a liquid chromatograph (LC)]. Several other well-known examples that involve various types of optical spectroscopy exist.

To summarize, univariate data analysis (implying one independent variable) is used for zeroth-order systems and multivariate analysis (more than one independent variable) is used for first- and higher-order systems. Second-order (and higher-order) systems are the subject of much current chemometrics research, but our reviews do not include systems higher than second-order.

## 2.2 EXPLORATORY DATA ANALYSIS

It seems highly likely that almost all of DOE's ERWM analytical chemistry problems will be first- or second-order. Only very few, if any, problems can be solved by zeroth-order measurements. The principal factor in this conclusion is the belief that there will almost always be some chemical present in a real-world sample of interest that was not included in calibrations, and that this interferent will make some (unknown and unknowable) contribution to the measured response. But even if not invalidated by some interferent, then a great deal of knowledge may be required to know under what conditions of temperature, pressure, . . . and others of a long list of environmental parameters the system used would produce reliable results. Consequently, a problem in the real world may not be amenable to analysis by univariate methods.

For simplicity, we assume an analyst has generated a data set of measurements made using several partially selective sensors, seeking to develop the details to provide determination of the concentrations of some set of chemicals. The techniques of exploratory data analysis (EDA), which are mostly graphical rather than statistical computations, are applied to such data sets. One goal of EDA is to free the data set from any *a priori* assumptions and preconceptions that might be imposed and to attempt to derive unexpected relationships between data elements. Three different but unusually instructive analyses that show the power of graphics in data exploration follow.

1. J. Mandel, "Data Analysis With Minimal Assumptions," *J. Chemom.* 6 247-255 (1992).

According to Mandel, one point of view about graphical exploratory data analysis is that by making properly conceived graphs, one can provide information ". . . that is both detailed and comprehensive about data. They have the further advantage of allowing the data analyst to decide what further analysis, if any, should be carried out. It is our contention that the use of computers in data analysis should consist of making the graphs appropriate to the data."

2. R. R. Meglen, "Examining Large Databases: A Chemometric Approach Using Principal Component Analysis," *J. Chemom.* 5 163-179 (1991).

Meglen discusses another point of view, namely that graphical tools, principal component analysis, and factor analysis can be judiciously combined to allow the objective selection of "information rich" plots from the multitude of possibilities. This approach is suggested as an antidote to the "shotgun approach" and massive efforts to design and assemble instrumentation; emphasis should be on the effort to design the experiment for optimal information extraction. The use of principal component analysis in this paper is "demonstrated in an application that required a multivariate approach to exploratory data analysis." Meglen makes the point that as "the scientific investigations become more complex, it becomes increasingly important to apply techniques that are as interpretationally sophisticated as the measurement instruments" and that "the appropriate instruments of reasoning must be assembled if data are to be fully exploited."

3. C. Weihs, "Multivariate Exploratory Data Analysis and Graphics: A Tutorial," *J. Chemom.* 7 305-340 (1993).

This article is a tour de force: a review article that builds tools from graphical subtools, illustrates their power and their usage on a graphics computer, and uses prediction as well as a demonstration of exploratory data analysis. This article is also a tutorial. The paper includes a quotation from the person considered by many statisticians to be the "father" of exploratory data analysis, J. W. Tukey. Weihs states:

"... (Tukey) claimed that exploratory data analysis, properly understood, 'is based on the principle that it is important to understand what you can do before you learn to measure how well you seem to have done it . . .'"

### 2.3 CALIBRATION AND PREDICTION

We assume that appropriate EDA has been performed for some data set with the conclusion that the planned analysis method has some reasonable chance of achieving the desired results. The next step is to develop a model that expresses the functional relationship between the concentrations of the analytes of interest and the responses of the set of sensors we assume. With that model chosen, the analyst must plan a set of calibration measurements and get the results from those calibrations. Chemometricians refer to this step as deriving the "training data set." Following successful completion of this phase (i.e., the model has been defined), the final measurement step is the measurements on real samples, followed by application of the model derived in the calibration phase. This latter phase is sometimes called the **prediction** phase of a chemometrics problem.

It will frequently be the case that the predictions indicate that the model used is inadequate in some way (e.g., the predicted concentrations lie outside the range of concentrations used in the calibration data for the model) or the predictions suggest unacceptable drift in the sensor response. In such a situation, it may be necessary to repeat the calibration phase and develop a better model. It is not uncommon to perform many iterations through these different phases to achieve best results.

As will be seen in some examples reviewed in Chapter 3 and as discussed at considerable length for many types of sensors by Vaihinger and Göpel (1), the particular form of the functional relationship can have an important effect on the uncertainties associated with predicted concentrations. For example, when using semiconducting metal oxide gas sensors, one can choose to measure either (a) the difference in conductivity (resulting from testing the sample), (b) the ratio of conductivities, or (c) several other quantities. The particular form chosen, one would suppose, would best be chosen from consideration of the basic mechanisms by which the sensor operates. However, for chemometrics purposes, almost any functional form can be assumed and, for mathematical simplification, it is quite common to assume that the response of a given sensor depends on some linear combination of responses from the separate chemicals in the mixture. In addition, the analyst must keep in mind the possible or likely interferents in the analysis. In a logical sense, there are two kinds of interferents to consider: **modeled interferents** are those unknown chemical species present in the calibration sample set to which the sensors responded, and **unmodeled interferents** are chemical species that may not be present in the calibration samples that generate an "interfering" response or in some other way interfere with the measurement result.

The following is a formal discussion of the calibration problem for zeroth- and first-order problems.

### 2.3.1 Calibration in Univariate Analysis

Univariate data analysis is applied when measurement data are obtained from a zeroth-order measurement system, as mentioned above. In this case the response is a function of only one variable, say the concentration of an analyte of interest; for example,

$$r = f(c),$$

where  $r$  denotes the response measured and  $c$  is the analyte concentration. This functional relationship will apply over some range of concentrations – perhaps a relatively narrow range, perhaps a very broad range. Suppose the concentration range of interest is an interval, say  $[c_1, c_2]$ , as determined by other analyses from a certain set of samples. The analyst would first do graphical analysis of a measurement set obtained from calibration samples. It may be that the graph of  $r$  as a function of  $c$  looks linear in the range  $[c_1, c_2]$ , in which case the "model" used might be expressed as

$$r = ac + b.$$

On the other hand, perhaps a non-linear relationship appears to be better, in which case the "model" might be expressed as

$$r = dc^2 + ec + g.$$

Regardless of the functional form, the calibration problem is to find the appropriate unknown parameters for the assumed functional relationship [( $a, b$ ) or ( $d, e, g$ ) in the examples here] to arrive at a way to predict concentrations from measured responses from other samples. The model will

be useful to the extent that it can usefully extend or fill in knowledge of concentrations not included in the calibration data set. For the linear model, the result is trivial (i.e., having determined a and b from calibrations),

$$c^* = (r^* - b)/a, \text{ where the } * \text{ denotes a new sample, not a calibration sample.}$$

### a. Regression and correlation

Clearly, one needs two or more measurements to fit a straight line. If the calibration data set has many points, then a systematic way of finding the parameters (a,b) is desirable: a very common approach is the ancient and efficient least squares regression. It is important to note here that we have found only a few chemometrics papers that discuss procedures for handling the propagation of errors in measurements. An obvious exception is the case of least squares regression for linear functions.

### b. Inherent Limitations of Univariate Analysis

For this type of analysis to be useful, the measuring instrument (sensor or whatever is being used) response must in fact depend on only one variable – we assume concentration in this example, but in most examples of electrochemical sensors the variable would be activity. If the instrument is a chemical sensor, it must be highly selective.

If such is not the case, it may be possible to separate the analyte of interest from interferences in a controlled and quantitative process (i.e., physical resolution may be used to solve the interference problem). Another way might be to use a more selective instrument; for example, in an optical photometric analysis one might use a narrower wavelength range to select a narrow band absorption feature of the molecule of interest. This approach is an example of what is sometimes called improved instrument response resolution.

But if an interferent is present in the calibration samples, there is no way in univariate calibration to detect the error. As a result, the predictions will be in error if the analysis is applied to a sample where either (a) the interferent is missing, or (b) the interferent is present at a significantly different concentration, or (c) other interferences are present.

#### 2.3.2 First-Order Calibration

A first-order system is one that produces a measurement data set consisting of vectors, one vector for each sensor when several sensors are used.

In these situations, multivariate calibration techniques can remove the necessity for complete selectivity or complete resolution, provided that a calibration set of responses is available. We suppose that the calibration data set has the following properties:

- every chemical species of interest is represented in the response data set, and
- the concentrations (or activities, etc., as appropriate to the sensor) of each of the analytes of interest are known for the calibration data set.

Under these conditions, multivariate calibration techniques usually work well and provide excellent results. The various techniques have as their goal the determination of what is called the inverse calibration matrix. For linear functional relationships between response and concentrations, the methods are straightforward. However, often great difficulty and complexity occur when the functional relationship is other than linear. The interested reader should review the chemometrics literature – see Appendix A.

The main first-order calibration methods and the formulas for matrix inverses required are given in the paper by Sanchez and Kowalski (2), along with illuminating comments and comparisons of the methods and their relative strengths and weaknesses. The methods discussed are:

- multiple linear regression by least squares (MLR),
- principal components regression (PCR),
- latent root regression (LRR),
- ridge regression (RR), and
- partial least squares (PLS).

### 2.3.3 Second-Order Calibration

In some instances, second-order data sets can be unfolded (3) so that first-order calibrations can be applied. One example of such a system would be a GC-MS in which the chromatograph column packing and the composition of the sample are such that a clean chromatographic separation is achieved for all components of the mixture before mass spectrometric analysis. In such cases, the analysis is more complicated than direct first-order calibration, but the technique can provide quantitative concentration data for modeled interferents and can detect unmodeled interferents, but concentrations of unmodeled interferents cannot be determined.

Techniques for more complex second-order calibrations – properly called tensorial calibration – depend very strongly on whether or not the data are bilinear. An example of a second-order bilinear data set would be produced by the hyphenated system LC-UV (liquid chromatograph – ultraviolet spectrometry). These techniques can compensate for both modeled and unmodeled interferents under very special conditions described in the papers listed in Appendix A. Difficulties met in trying to solve the second-order bilinear calibration problem have had a damping effect on the earlier optimistic view about the many successes anticipated in applications of chemometrics to complex analytical chemistry problems.

Even further afield from solving real analytical chemistry problems, if the second-order data are not bilinear, the problem is very much more complex and difficult (see Appendix A). According to Sanchez et al. (4), generalized rank annihilation methods (GRAM) have been successfully applied to several bilinear data sets (e.g., LC-UV). However, for application to nonbilinear data sets, GRAM requires modifications.

## 2.4 STATUS OF TOOLS FOR ANALYSIS

The following books contain useful information.

Four textbooks are commonly cited. Each book provides a different orientation to software for learning the techniques involved, which is the recommended way to approach chemometrics.

1. R. G. Brereton, Ed, *Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies*, Elsevier, New York, 1992.

The principal subjects treated are

- multivariate data display and the software SPECTRAMP,
- principal component analysis (PCA),
- PCA cross-validation and the F-test using SIMCA and SIRIUS software, and
- exploratory data analysis and hard modeling using PARVUS software.

2. E. R. Malinowski, *Factor Analysis in Chemistry*, 2nd edition, Wiley-Interscience, New York, 1991.

This book has two distinct parts: (1) a textbook on factor analysis and (2) a software toolkit based on the MATLAB software package. In the textbook, the main subjects are

- eigenvalue analysis with the TARGET 90 program,
- iterative key set factor analysis,
- PCA and binary cross validation,
- rank annihilation factor analysis, and
- uniqueness tests and miscellaneous subjects.

The MATLAB toolkit includes

- significance factor analysis (MATLAB code provided),
- target factor analysis (with MATLAB code),
- loading factor analysis (with MATLAB code), and
- other do-it-yourself MATLAB programs.

3. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: A Textbook*, Vol. 2 of the series *Data Handling in Science and Technology*, Series Editors, B. G. M. Vandeginste and L. Kaufman, Elsevier, New York, 1988.

The main parts of interest in this well-known textbook are

- general statistics package with the BALANCE program,

- principal components and factor analysis,
- hard and soft modeling, and
- supervised pattern recognition.

4. M. Meloun, M. Militky, and M. Forina, *Chemometrics for Analytical Chemistry, Volume I: PC-Aided Statistical Data Analysis*, Translation Editor, Dr. M. Masson, Ellis Horwood, New York, 1992.

Two other even more specialized books are:

5. G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1985.
6. C. R. Rao, *Generalized Inverse for Matrices and Its Applications in Mathematical Statistics*, Research Papers in Statistics Series, Festschrift for J. Neyman, F. N. David, Ed, John Wiley, New York, 1966.

## 2.5 STATUS OF THE FIELD OF CHEMOMETRICS

This section includes reviews of several journal articles that provide a variety of insights into the history, current research activities, and other aspects of chemometrics.

### 2.5.1 General Perspective of Chemometrics

Chemometrics is an emerging multidisciplinary field of research. The disciplines being drawn together, in addition to the dominant chemistry, are mathematics, statistics, and engineering. To these should be added a term to signify the "computing component," but the words computer science, which rarely seem appropriate, seem especially wrong in this context. Through all of these individually, and through chemometrics separately, run the influences, trends, and forces affecting every human enterprise.

According to Geladi and Esbensen, who launched a chemometrics history project (5), the activity of chemometrics is a specific part of chemistry that is based on a large number of successful applications. Since its origin in the early 1970s, it has had a significant impact on academic, industrial, and educational environments in chemistry. To summarize and oversimplify, the key to understanding the chemometrics phenomenon is the applications of mathematics made possible by the digital computer.

The "queen" of the chemometrics imports from mathematics is linear algebra. The theorems and techniques of this elegant part of mathematics are very basic to chemometrics applications. For chemometrics it is probably not overstating the case to draw a parallel between the importance of the concept of multivariate space and the importance of the concept of coordinate systems that was introduced into 17th-century European mathematical analysis by Descartes. Chemometrics has benefitted, too, from applications of mathematical techniques to the social sciences, and

especially from the applications of mathematics and statistics to psychology in the 1950s and 1960s.

Simply put, the status of chemometrics is the status of the techniques used to solve the problems that have so far been addressed. This assessment implies that a new type of problem, such as may arise implicitly in the DOE ERWM work, may not have a ready solution "lying on the shelf" in chemometrics. When this is the case, the clear imperative is to start to work on the problem. Even when there may be an off-the-shelf solution to a problem, it may not be found for a variety of reasons. The most important aspect of tackling a problem is the problem definition, which requires development of a very detailed understanding of the real-world needs and the real-world environment so that the problem can be formulated in a meaningful way.

Another force in shaping chemometrics has been instrumentation development. Put very simply, developments in instrumentation gave applied chemists a task analogous to "drinking out of a fire hydrant" – new instrumentation provided much more data than anyone could analyze or manage. Increasingly sophisticated instruments always follow increased transducer resolution and advances in data acquisition and data processing. Advances in data storage technology, communications, and data compression provide increasingly overwhelming amounts of data from which the analyst must derive insights and solutions to problems.

The tools chosen for the analyses reflect the capabilities and limitations of the human users. Pattern recognition and graphical presentation of data play an essential role in providing geometrical insight to creatures limited to two or three dimensions as they attempt to navigate in multidimensional geometrical spaces whose intrinsic dimensionality exceeds that of the analyst by orders of magnitude.

The term "data reduction" is seldom heard these days: modeling, analysis, and data compression are the methods now applied to reduce the data glut. Multivariate analysis is built on the edifice of matrix computations, which includes many complexities and surprises as well as pitfalls that might well be expected in an environment whose dimension is larger than the usual two, three, and four (with time) dimensions of ordinary human life.

One of this report's authors, after reading quite a bit of the reasonably current literature and interviewing several researchers, has the following observations.

- Among some workers (but not among those interviewed), there is evident a somewhat cavalier attitude towards the essential mathematical facts and rigorous approach that a mathematically oriented, real-world problem solver would like to find. However, there is validity in the proposition that so long as one is firmly grounded in reality, because of the nature of the problem and the luxury of unequivocal tests of validity, one may sacrifice rigor and emphasize intuitive understanding with no sacrifice exacted by the "gods of science."
- To summarize, chemometrics is a very healthy field, so long as one remains problem-oriented and firmly in touch with reality through real-world test and evaluation.
- It may be useful to note that in a young, multidisciplinary field, tradeoffs will always be required. In the case of chemometrics, consider for a moment just four of the disciplines involved: chemistry, mathematics, statistics, and engineering. (One could add computing, but

that would be considered infrastructure like buildings, transportation, and food markets that provide the essentials of life from agriculture.) It would be too much to expect that any one person would become an expert in all four disciplines. Consequently, the imperfections mentioned are inevitable, but proper attention to these aspects is important for those who must, of necessity, delve into the field to learn how to apply chemometrics to real world problems.

### 2.5.2 Towards Statistical Discipline and Data Stewardship Standards

In this section we include reviews of three articles that address, directly or peripherally, issues of rigor, attempts to control or eliminate subjectivity, and preservation of the requisite information when reporting and treating data and their analyses.

1. S. Wold and P. Geladi, "Editorial: The Importance of Raw Data," *Chemom. Intell. Lab. Syst.* 22 1–2 (1994).

The essence of this paper is summarized in the following outline.

- Some of the Problems
  - Lack of adequate documentation and explanation of raw parameters and raw experimental data sets
  - Lack of mention/description of preprocessing of raw data in preparation for analysis
  - Lack of a corpus of reference raw data comprising true test data for verification and validation use
- Some Suggested Solutions
  - Data set count, variables, and any preprocessing should be flagged in the abstract of an article.
  - Data set size, parameters, and preprocessing should be fully described in an article and the count and parameter details repeated a few times to provide a redundant check on their availability and accuracy.
  - Raw data should be published as appendices, whether or not the data sets are candidates for standards.
  - Reference data sets are needed for regression, not just for pattern recognition. These could well include pattern recognition as a subtopic.
  - Most importantly, there needs to be established a data base of large test data sets. Standard data sets from statistics are typically poor for chemometrics because of the traditional emphasis on large samples and few variables. For chemometrics many of each are needed.
- The Need for Standards – The need to capture ground truth in the form of a database of large chemometrics data sets is a complex but pressing requirement for the field. Questions about the management of such a database and the security of the answers to the questions would inevitably raise larger questions, but clearly this would provide a valuable service for those analysts seeking to validate their analysis methods. The economics of the validation exercise might also dictate an approach whereby a user

of information might find it desirable to maintain such a facility for its own use, and one explicitly designed for its own needs. An example is the use of standard tests for quality assurance (QA).

2. T. Permutt, S. D. Eland, M. Moezzi, and S. C. Grosser, "Likelihood Techniques For Interlaboratory Calibration in the National Stream Survey," *J. Chemom.* 5 299-308 (1991).

This paper makes the following points:

- The need for QA and consistency in field survey programs is clear; the analysis methods and measurement interpretations must be consistently applied.
- Statistical methods of testing must be used to detect interlaboratory bias by means of an audit program using field samples.
- There is a need for detection of bias and correction for regional differences (e.g., from the different sites) that are confounded with bias.

Results and conclusions include the following:

- For some analytes, no bias exists.
- For many analytes, a well-fitting interlaboratory bias calibration function was linear over the range of concentrations of the audit samples, but the range was too small.
- For some analytes, a multivariate approach to interlab (two labs) calibration of bias is required, because a linear calibration seemed impossible.
- It is recommended that random samples for audits come from the population of interest and that the experimental design be revised to increase the information per unit cost.

3. V. I. Dvorkin, "Data Processing in the Interlaboratory Test by Analysis of Covariance," *Chemom. Intell. Lab. Syst.* 22 127-146 (1994).

Dvorkin describes a series of statistical tests based on analyses of variance and covariance for data collected from six laboratories. The data included results of both qualitative and quantitative analyses of standard reference materials. The authors conclude that the analysis of variance with scaling, using the ANOVA software, and the analysis of covariance, using the ANCOVA software, complemented each other. Detailed analyses of the errors found made improvements in the analysis procedures possible and suggest that the methods may be useful for detecting errors in both clinical and analytical chemistry laboratories.

## 2.6 SOFTWARE, COMMERCIAL AND OTHER

### 2.6.1 Commercial Software

#### a. MACSYMA

MACSYMA has been one of the two leading mathematical packages for a decade, the other being MATHEMATICA. The two packages share the feature that they not only do mathematics, but they support symbolic mathematics. The programs available cover a wide range of mathematical techniques, including eigenvalue and eigenvector calculations. Neither tries to be a general purpose statistical package, and their overlap with the latter is small.

A free demonstration diskette shows a limited set of capabilities, including a good (advanced) tutorial on eigenvector algebra.

The minimal brochures are primarily promotional; they contain a list of mathematical capability highlights. The program PC MACSYMA 417.125 is a subset of a program for mid-range workstation computers and the VAX VMS or ULTRIX of Digital Equipment Corp.

**VENDOR:** MACSYMA Inc.,  
20 Academy Street  
Arlington, MA 02174-6436  
Telephone: (617) 646-4550 or (800) 622-7962  
Fax: (617) 646-3161 or (800) 622-7962

#### b. MATLAB

Because of the variations in requirements for analysis in the problem-driven field of chemometrics, MATLAB has been a favorite. It is a programming language that targets "technical concepts, design algorithms, and prototype solutions." It is sometimes used to append the code at the end of a technical journal article. A MATLAB advertisement states that it is "an elegant language . . . rich with built-in functionality . . . [with] powerful GUI tools [that] let you build intuitive, interactive displays" with sound and animation.

Reviews of the software are given in the following references:

C. H. Spiegelman, "Software Review, MATLAB and MATLAB Optimization Toolbox (Macintosh Version)," *Chemom. Intell. Lab. Syst.* 19 128 (1993).

T. C. O'Haver, "Teaching and Learning Chemometrics with MatLab," *Chemom. Intell. Lab. Syst.* 6 95-103 (1989).

**VENDOR:** The MathWorks, Inc.  
24 Prime Park Way  
Natick, MA 01760  
Telephone: (508) 653-1415  
Fax: (508) 652-6284  
E-mail: [info@mathworks.com](mailto:info@mathworks.com)

### c. MATLAB Toolboxes

Several MATLAB toolboxes have been developed that allow a potential user to avoid becoming a MATLAB programmer. Some are available through MATLAB: these are "targeted, and include symbolic math, image processing, and statistics. They require the MATLAB environment in which to work."

There is at least one chemometrics-targeted MATLAB toolbox: PLS\_Toolbox, Version 1.3, by Barry Wise is a fairly mature toolbox with a wide variety of MATLAB routines growing out of Wise's research activities. This software, in combination with the necessary MATLAB itself, may represent the best approach to developing familiarity with chemometrics and capability in the techniques because it seems to follow a coherent approach oriented to the mathematics rather than to ad hoc solutions for specific problems.

According to the instruction manual, "PLS\_Toolbox differs somewhat from other MATLAB toolboxes in that it is the mathematical roots of the routines that tie them together, rather than the intended applications."

The manual is a valuable source in itself, including informative discussions of regression methods, dealing with sensor issues, and including references to current literature. It is the author's "continuing education" project, and it is continually updated with new techniques and routines.

**VENDOR:** Barry M. Wise  
4154 Laurel Drive  
West Richland, WA 99352  
E-mail: [bm\\_wise@pnl.gov](mailto:bm_wise@pnl.gov)

### d. PIROUETTE

In contrast to MATLAB, PIROUETTE is a package that presents the user with tools for multivariate data analysis without programming. It is an advanced and powerful capability. The software company, InfoMetrix, considers itself a "leading developer of software products employing pattern recognition techniques for the analysis of multivariate data sets in chemistry and biotechnology." Their goal is software that "will be completely integrated with instrumentation to provide better solutions to complex problems in chemistry, engineering, and other sciences."

A free demonstration diskette of PIROUETTE Version 1.2 for MS-DOS systems includes an initiation by walk-through of a data set to illustrate the capabilities in data exploration, multivariate regression, calibration, and data presentation (plots). The demonstration comes with a 50-page manual.

The brochure comes with high-quality technical notes, with references, chemometrics applications overview, and applications notes.

**VENDOR:** InfoMetrix  
2200 Sixth Avenue, Suite 833  
Seattle, WA 98121  
Telephone: (206) 441-4696  
Fax: (206) 441-0941  
E-mail: infomtrx@halcyon.com

**e. General Purpose Statistical Software**

See the article: S. Canter, "State of the Art," *PC Magazine*, May 11, 1993, pp 227-287.

From time to time, personal computer magazines review commercially available general purpose software packages. While the process of reviewing software is difficult and subjective, the magazines do an admirable job of choosing factors, rating systems, comparing packages, and formatting the results in a user-friendly fashion. Thirteen systems were reviewed in this recent feature article, including some of the packages listed here.

**f. SPSS**

This package has been around a long time and has a large installed base. It won an "editors' choice" in the magazine review mentioned in Item e. An excellent free demonstration diskette is available for SPSS.

The brochure suggests an interesting integration of data presentation with MapInfo, the "desktop mapping software," that could be a powerful tool when combined with the data producing software for managing a DOE cleanup program.

**VENDOR:** SPSS Inc.  
12030 Sunrise Valley Drive, Suite 300  
Reston, VA 22091  
Telephone: (703) 391-6020  
Fax: (703) 391-6002  
E-mail: billr@spss.com

**g. S-PLUS**

This package affords an immediate capability by means of a large number of statistical tools, while it is immediately possible to create custom analyses by using the built-in command language. A detailed review is available in the following reference.

M. C. Denham, "Software Review, S-PLUS," *J. Chemom.* 7 559-566 (1993).

**h. STATGRAPHICS**

This general purpose statistical analysis program consistently rates high for graphical capabilities. It is one that also has needed capabilities that are unusual in general packages. S. Canter states (Item e) that "STATGRAPHICS . . . Version 6, is an interactive program that artfully combines statistics, graphics, and data manipulation. With just a few keystrokes, you can do a regression,

plot residuals, and pop up a screen to create a new variable to be used in the next step of the regression. The program is strong in graphics, quality-control procedures, response surface graphs, time series analysis, and experimental design (an area addressed by few general statistics packages)."

The program provides interactive outlier rejection, which "lets you plot data with a regression line, highlight an outlying point, and see how the line would change with the point removed."

The brochure includes examples of graphics produced.

**VENDOR:** Manugistics, Inc.  
2115 East Jefferson Street  
Rockville, MD 20852  
Telephone: (301) 984-5123 or (800) 592-0050  
Fax: (301) 984-5094

### i. UNSCRAMBLER

This package is closely associated physically and theoretically with the Norwegian school of chemometrics resident at Trondheim. As might be expected from pioneering innovators in chemometrics who are thoroughly connected to real-world problems, the company, CAMO, casually steps up to a convincing attempt to "do it all." In addition to multivariate regression with PLS and PCR, "UNSCRAMBLER has nearly limitless options for data handling, modeling, and assorted graphics capabilities, including pseudo-three-dimensional plots."

The first review cited at the end of this summary (by an academic researcher) complains about "a user interface [that] does not feel like any other program I have used in the last few years." Nevertheless, the reviewer raves about the great manual, the included tutorial examples, and the user's group and its newsletter.

Recent CAMO brochures tout the addition of a neural net module. The reviews in hand are presumably dated since they mention launch of a Version 5. The neural net module, Neural-UNSC, which was to be released in March 1993, is "an add-on module for constructing Neural Network models, which are combined with PCR or PLS models to solve systems with nonlinearities."

It is said that CAMO listens to its users and is committed to continual improvements. This package certainly should be carefully evaluated as a possible tool. Unfortunately, no free demonstration diskette is available. Instead, a 6-week trial installation is available from the CAMO U.S.A. office (see VENDOR) for \$100.00.

The brochure illustrations should calm any concerns about the user interface. The brochure also indicates that Version 5 includes a high-quality capability in support of experimental design. The experimental design description indicates implementation of the "classical approaches of Box, Hunter, and Hunter and includes saturated designs for screening and optimization purposes for up to 8 or 16 experiments, with blocking and the possibility to add center points. The program

suggests low-cost experimental plans [sic] for 3–15 variables in fractional or full factorial designs. . . . For 10–15 variables or more you may also use Plackett-Burman designs."

In discussing response surface capabilities ". . . there is of course no limit on the number of responses. Mixture designs will be implemented in 1993. The program shows you the pitfalls – which main effects and interactions become confounded."

Given the unique addition to UNSCRAMBLER of a neural net add-on module, which is advertised as a measure taken for dealing with nonlinear data sets, the brochure, interestingly, claims that by judicious use of PLS or PCR models as preprocessing, the risk of overfitting is reduced or eliminated, resulting in smaller networks and faster processing.

Copious application notes are available to orient and instruct in the various techniques. Furthermore, as a result of the active interest by Norwegian chemometrists, the user group newsletters are said to be a rich source of current insights and give unusual access to these workers. One reviewer found newsletter discussions, "How to detect and deal with nonlinearities" by Tormod Naes and "Leverage correction and other validation methods" by Tormod Naes and Kim Esbensen, so enlightening that he used the material for a seminar.

The following list of some applications notes accompanied the brochure:

- Prediction of dioxines
- Predictions of concentrations: Spectroscopy
- Finding relationships by projections
- Handling overlap in spectra sets
- Production quality problem
- Design for a reaction

The previous software reviews were forwarded by CAMO along with the brochure and application notes. The review references follow, with the caveat that the earliest is very dated indeed. However, it may be of value in comparisons with other packages or with improvements to the program itself. The track record supports the CAMO claim to be interested in the needs of the users. Reviews are available in the following references.

D. Dahlberg, *Software Review: UNSCRAMBLER II Ex Update*, North American Chapter of the International Chemometrics Society, Newsletter No. 6, April 30, 1993.

"Software Review: UNSCRAMBLER II, Version 4.0," *Chemom. Intell. Lab. Syst.* 19 269–275 (1993).

K. Helland, "Software Review: UNSCRAMBLER II, Version II," *J. Chemom.* 5 413–415 (1991).

**VENDOR: CAMO**

CAMO U.S. (Redwood Shores, CA), contact John Dibble

Telephone: (415) 598-9860

CAMO AS

Olav Tryggvasonsgata 24

N-7011 Trondheim, Norway

Telephone: 011 47 73 51 49 66

Fax: 011 47 73 51 42 57

**2.6.2 Other Software (Shareware, Freeware, Bundled Bookware)****a. SCOUT**

This is an excellent exploratory data analysis package in the public domain developed under contract to the U.S. Environmental Protection Agency by Lockheed Engineering & Sciences Company. SCOUT is the subject and tool of use for the following article.

M. A. Stapanian, F. C. Garner, K. E. Fitzgerald, G. T. Flatman, and J. M. Nocerino, "Finding Suspected Causes of Measurement Error in Multivariate Environmental Data," *J. Chemom.* 7 165-176 (1993).

In the appendix of the article, SCOUT is described as "a small, multivariate, exploratory data analysis package [whose] major statistical features include

- outlier testing and identification,
- estimating prediction intervals of "acceptable" values for the identified discordant observations with standard multiple-regression techniques,
- principal components analysis,
- univariate tests of normality, and
- summary statistics."

SCOUT for IBM compatible machines may be obtained by sending a formatted high-density diskette and a self-addressed, postage-paid mailer to F. C. Garner, Lockheed Environmental Systems and Technologies Co., 980 Kelly Johnson Drive, Las Vegas, NV 89119.

**2.7 REFERENCES**

1. S. Vaihinger and W. Göpel, *Multi-component Analysis in Chemical Sensing*, Chapter 6 in *Chemical and Biochemical Sensors, Part I*, eds. W. Göpel, T. A. Jones, M. Kleitz, J. Lundström, and T. Seiyama, Volume 2 of *Sensors: A Comprehensive Survey*, eds. W. Göpel, J. Hesse, and J. N. Zemel, VCH, New York, 1991.
2. E. Sanchez and B. R. Kowalski, "Tensorial Calibration: I. First-Order Calibration," *J. Chemom.* 2 247-263 (1988).

3. S. Wold, P. Geladi, K. Esbensen, and J. Öhman, "Multi-way Principal Components- And PLS-Analysis," *J. Chemom.* 1 41–56 (1987).
4. E. Sanchez, L. S. Ramos, and B. R. Kowalski, *J. Chromatogr.* 385 151 (1987).
5. P. Geladi and K. Esbensen, "The Start and Early History of Chemometrics: Selected Interviews. Part 1," *J. Chemom.* 4 337–354 (1990).

### 3. APPLICATIONS OF CHEMOMETRIC TECHNIQUES TO CHEMICAL SENSOR ARRAYS

This chapter contains short reviews of journal articles that describe research and development of methods for analysis of mixtures using three different types of chemical sensors: electrochemical sensors, tin oxide semiconductor vapor sensors, and surface acoustic wave devices. Section D includes reviews of a few articles that are general in nature (e.g., mathematical theory of sensor array designs, or cover more than one type of sensor). Reviews are in descending chronological order and alphabetically within the year. A list of other references obtained from the Scientific and Technical Information Network (STN) is presented in Appendix B.

Although there have been a great many studies of the application of chemometrics techniques to the more common analytical methods, such as chromatography and spectroscopy, the great majority of research done using the types of chemical sensors considered in these tasks has been done with the three types of sensors chosen. Readers interested in other applications should consult the several textbooks on chemometrics and the journals devoted to that subject, as well as the commonly available analytical chemistry journals.

#### 3.1 ELECTROCHEMICAL SENSOR ARRAYS

1. M. Esteban, C. Ariño, I. Ruisánchez, M. S. Larrechi, and F. X. Rius, "Expert System for The Voltammetric Determination of Trace Metals, Part V. Methods for Determining Total Iron, Manganese (II), Aluminum and Titanium," *Anal. Chim. Acta* 285 377-389 (1994).

The expert system previously described (see Items 2, 6, 8, and 9 on pages 3-1, 3-2, and 3-3 for Parts I, II, III, and IV by the same authors) has been enlarged and improved by including methods for the determination of total Fe, Mn<sup>+2</sup>, Al, and Ti. The expert system used the commercially available building tool KES.

2. M. Esteban, C. Ariño, I. Ruisánchez, M. S. Larrechi, and F. X. Rius, "Expert System for the Voltammetric Determination of Trace Metals, Part IV. Methods for Speciation of Chromium and Arsenic," *Anal. Chim. Acta* 285 193-208 (1994).

The expert system described in earlier papers (see Items 6, 8, and 9 on pages 3-2 and 3-3) adds Cr and As. Procedures were developed for Cr<sup>+3</sup>, Cr<sup>+6</sup>, As<sup>+3</sup>, and As<sup>+5</sup>. Cr<sup>+3</sup> and As<sup>+5</sup> are determined as the difference between total Cr and Cr<sup>+6</sup> and total As and As<sup>+3</sup>, respectively.

3. F. J. Sáez de Viteri and D. Diamond, "Ammonium Detection Using an Ion-Selective Electrode Array in Flow-Injection Analysis," *Electroanalysis* 6 9-16 (1994).

This paper describes the use of an array of ISEs selective for NH<sub>4</sub><sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, and Ca<sup>+2</sup> as a flow injection analysis sensor. The sensor response was modeled using projection pursuit methods.

4. D. R. Bull, G. J. Harris, and A. B. Ben Rashed, "A Connectionist Approach to Fuel Cell Sensor Array Processing for Gas Discrimination," *Sens. Actuators* B15-16 151-161 (1993).

Fuel cell sensors are widely used for evidential breath testing and environmental monitoring, but problems often arise from lack of specificity because the sensors respond to several alcohols and aldehydes in addition to ethyl alcohol. Specificity can be improved by using discrimination algorithms based on neural networks. Better results are obtained when the neural net techniques are used with arrays of sensors operated under different conditions (e.g., operating temperature).

5. D. Diamond, "Progress in Sensor Array Research, Review Article," *Electroanalysis* 5 795–802 (1993).

Diamond focuses on electrochemical sensors, including arrays of potentiometric ISEs and ion-sensitive field effect transistors (ISFETs), amperometric and voltammetric arrays, and semiconductor oxide sensor arrays. Diamond suggests that "The use of sensor arrays will expand quickly during the 1990s as the cost of computing power continues to plummet, software becomes more user-friendly, statistical tools are developed or refined, and, perhaps most importantly of all, high-quality, replicate arrays become available at low cost. This latter development is the key to future progress and will depend on advances in technology for the fabrication of high-quality microdimensioned substrates, the reproducible deposition of chemically sensitive layers or membranes, and surface analysis techniques for quality control. These are already largely available, and it is now becoming a question of putting all the parts of the jigsaw together."

6. M. Estebán, C. Ariño, I. Ruisánchez, M. S. Larrechi, and F. X. Rius, "Expert System for The Voltammetric Determination of Trace Metals, Part III. Methods for Determining Mercury, Selenium, and Vanadium," *Anal. Chim. Acta* 284 435–443 (1993).

The expert system described in earlier papers for voltammetric determination of Cu, Zn, Cd, Pb, In, Ni, Co, and Tl is enhanced and improved by adding methods for Hg, V, and Se (also Te). Methods used are as before plus cathodic stripping voltammetry and adsorptive stripping voltammetry using dropping Hg electrodes and a gold electrode (for Hg). Data analysis uses the multiple standard addition method. The expert system uses the KES software framework.

7. F. X. Rius, "Expert Systems in Trace Analysis," *Anal. Chim. Acta* 283 518–527 (1993).

This paper reviews the fundamentals of knowledge-based expert systems, mentioning recent applications in trace analysis as well as novel approaches. Applications are distinguished as qualitative or quantitative. Quantitative examples include sample preparation by microwave dissolution, solid phase extraction of drugs from aqueous samples, liquid-liquid extraction of rare earths, sampling, liquid chromatography – lots of examples, atomic absorption spectroscopy, voltammetric analyses, and environmental pollutant analysis advice, among others.

8. M. Esteban, I. Ruisánchez, M. S. Larrechi, and F. X. Rius, "Expert System for the Voltammetric Determination of Trace Metals, Part II. Methods for Determining Nickel, Cobalt, and Thallium at Different Concentration Ratios," *Anal. Chim. Acta* 268 107–114 (1992).

The expert system reported in the preceding paper of the series by the same group (see Item 9) has been extended to include methods for the determination of nickel, cobalt, and thallium. Quantitative data are obtained using the multiple standard addition method. The expert system was developed using the KES framework.

9. M. Esteban, I. Ruisánchez, M. S. Larrechi, and F. X. Rius, "Expert System for The Voltammetric Determination of Trace Metals; Part I. Determination of Copper, Zinc, Lead, and Indium," *Anal. Chim. Acta* 268 95–105 (1992).

This paper gives a detailed description of an expert system for voltammetric determination of Cu, Zn, Cd, Pb, and In. The system guides the user in the choice of sample treatment, the appropriate voltammetric procedure, and identification and determination of the trace metals. Techniques included are differential-pulse polarography and anodic stripping voltammetry using dropping Hg electrodes.

10. R. J. Forster and D. Diamond, "Nonlinear Calibration of Ion-Selective Electrode Arrays for Flow Injection Analysis," *Anal. Chem.* 64 1721–1728 (1992).

This paper describes the application of an array of ion-selective electrodes (ISEs) to the simultaneous determination of  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{+2}$ . The array consists of three highly selective ISEs and a fourth sensor that responds to all three ions but to differing degrees. Calibrations were done using "mixed calibration solutions" and a computational method called simplex optimization. The combination of sensors used gives better results than single electrode measurements and array results obtained using sparingly selective electrodes.

11. I. Ruisánchez, M. S. Larrechi, F. X. Rius, and M. Esteban, "Computer-Aided Voltammetric Method Development Employing a Knowledge-Based Expert System," *Trends in Anal. Chem.* 11 135–142 (1992).

This paper describes an expert system for voltammetric analysis. The system guides the user through the analytical process, offering advice on sample pretreatment, selection of a suitable methodology for voltammetric determination, and appropriate methods for qualitative and quantitative analysis.

12. H. Sundgren, I. Lundström, and H. Vollmer, "Chemical Sensor Arrays and Abductive Networks," *Sens. Actuators* B9 127–131 (1992).

This paper describes the use of abductive induction to evaluate sensor array data. Such techniques have been used in sensor fusion applications, but they have not been used heretofore in chemical sensor array analysis. Research described used a commercially available algorithm called "abductive induction mechanism (AIM)." Data for the analyses were obtained using six FETs for  $\text{H}_2$  and  $\text{NH}_3$ . As explained in the paper, deduction is the process of reasoning from general principles and facts to new facts with certainty. Abduction is the process of reasoning from general principles and facts to new facts under uncertainty. Induction is the process of reasoning from facts to general principles. The term abductive induction is applied to the process of reasoning from facts to abductive principles – also said to be a special class of inductive learning for synthesizing abductive principles from

databases of empirical observations. Abductive functions are numerical functions used to reason under uncertainty.

AIM evolved from research on neural nets and statistical methods; the mathematical background is contained in what is called the "group method of data handling," which originated in the Ukraine about 25 yrs ago.

13. R. J. Forster and D. Diamond, "Multivariate Calibration of Potentiometric Sensor Arrays," *Anal. Proc.* 28 117-122 (1991).

This paper describes nonlinear optimization procedures used to calibrate an array of potentiometric ISEs and applied to the analysis of  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{+2}$ . A fourth sensor that was sparingly sensitive to all three ions was included in the array. The optimization procedures used in the calibration phase of the work were based on "direct search and gradient methods," which included schemes called the Hooke and Jeeves method (1), the modified simplex optimization (2), and the gradient search method of Davidon, Fletcher, and Powell (3).

These procedures and ISEs give better analytical results than traditional single-electrode methods. The methods also avoid the problems associated with traditional methods of defining selectivity coefficients. It is suggested that applications of modern computer technology would lead to the reassessment of the traditional procedures for calibrations of potentiometric ISEs.

14. R. J. Forster, F. Regan, and D. Diamond, "Modeling of Potentiometric Electrode Arrays for Multicomponent Analysis," *Anal. Chem.* 63 876-882 (1991).

The research described in this paper uses the same procedures as those explained in the preceding review. In this paper, the methods were applied to determination of  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{+2}$  in tertiary mixtures of these cations at concentrations found in physiological research. Unlike traditional single-electrode methods, the array and procedures used can determine accurately low levels of the individual cations in the presence of large and widely varying concentrations of the other two ions.

15. H. Sundgren, F. Winquist, I. Lukkari, and I. Lundström, "Artificial Neural Networks and Gas Sensor Arrays: Quantification of Individual Components in a Gas Mixture," *Meas. Sci. Technol.* 2 464-469 (1991).

The research described in this paper used six MOSFETs for analysis of mixtures of  $\text{H}_2$ ,  $\text{NH}_3$ , ethyl alcohol, and ethylene in air and for  $\text{H}_2$  and acetone in air. Data analyses were done in three ways: conventional multivariate analysis, PLS, and neural nets.  $\text{H}_2$  and  $\text{NH}_3$  can be done by PLS, but better results are obtained with neural nets. Ethyl alcohol and acetone results were poor by both of those methods. In two-component mixtures,  $\text{H}_2$  and acetone are both done best by neural nets.

16. M. Bos, A. Bos, and W. E. Van Der Linden, "Processing of Signals from an Ion-Selective Electrode Array by a Neural Network," *Anal. Chim. Acta* 233 31-39 (1990).

This paper describes neural net software for processing signals from ISE arrays. The software was tested in determinations of  $\text{Ca}^{+2}$  and  $\text{Cu}^{+2}$  in binary mixtures of copper nitrate and calcium chloride, and in simultaneous determinations of  $\text{K}^+$ ,  $\text{Ca}^{+2}$ ,  $\text{NO}_3^-$ , and  $\text{Cl}^-$  in mixtures of  $\text{KCl}$ ,  $\text{CaCl}_2$ , and  $\text{NH}_4\text{NO}_3$ .  $\text{Ca}^{+2}$  and  $\text{Cu}^{+2}$  measurements used pH-glass electrodes and  $\text{Ca}^{+2}$  and  $\text{Cu}^{+2}$  ISEs and gave results accurate to  $\pm 8\%$ . For other mixtures, ISEs for the analyte ions were used with glass electrodes, resulting in relative errors of  $\pm 6\%$ , but worst case errors were as large as  $\pm 20\%$ . Many problems are encountered in such work – see Items 20, 21, and 23 on page 3-6. Neural networks seem ideally suited to handling ISE array data because they can treat interactions and nonlinear behavior.

17. H. V. Shurmer, "Basic Limitations for an Electronic Nose," *Sens. Actuators* B1 48–53 (1990).

The model nose was first described by Persaud and Dodd (4), who showed that a system using three  $\text{SnO}_2$  sensors could provide discrimination between chemically similar odors. "Many laboratories are now investigating alternative types of sensors. The particular form of development that appears most attractive is the creation of integrated arrays of sensing elements, each of which is tailored to have a broad-tuned response that is offset from that of the other elements. Developments in related test procedures and information processing techniques are still in their infancy. In due course, however, each will significantly influence the form of instrumentation for the wide range of applications awaiting exploitation."

Sensor types being investigated include, among others,  $\text{SnO}_2$ , conducting polymers such as substituted polypyrroles and polyanilines, and Langmuir-Blodgett films (including Cu phthalocyanine and certain porphyrins) in CHEMFET designs.

18. H. Sundgren, I. Lundström, F. Winquist, I. Lukkari, R. Carlsson, and S. Wold, "Evaluation of a Multiple Gas Mixture with a Simple MOSFET Gas Sensor Array and Pattern Recognition," *Sens. Actuators* B2 115–123 (1990).

Analyses reported in this paper used PLS. Good results were obtained for  $\text{H}_2$  in the presence of three other interfering gases.

19. M. Gall and R. Müller, "Investigation of Gas Mixtures with Different MOS Gas Sensors With Regard to Pattern Recognition," *Sens. Actuators* 17 583–586 (1989).

Work described in this paper used metal oxide semiconductor (MOS) sensors sensitive to different gases; sensors use different catalytic metals. Data evaluation includes a model of sensor response. Cross sensitivity of a single MOS sensor is compared with cross sensitivity of an MOS sensor array using pattern recognition methods.

Mixture analysis can be done with nonselective sensors if responses are linear. This paper describes what is called the Transformed Least Squares method.

20. K. R. Beebe and B. R. Kowalski, "Nonlinear Calibration Using Projection Pursuit Regression: Application to an Array of Ion-Selective Electrodes," *Anal. Chem.* 60 2273–2278 (1988).

Beebe and Kowalski describe a new method of regression analysis that performs both linear and nonlinear regression and can determine the form of the model as well as the model parameters. The method is applied to data obtained using ion-selective electrodes (ISEs) in solutions containing mixtures of  $\text{Na}^+$  and  $\text{K}^+$ . The sensor response is assumed to be Nernstian. Other analysis techniques discussed include (a) multiple linear regression, (b) partial least squares, (c) nonlinear regression based on a simplex algorithm, and (d) multiple linear regression. Electrodes used include Corning general purpose cation glass electrode, Orion sodium glass electrode with nonselective behavior, and three plastic membrane electrodes.

21. K. Beebe, D. Uerz, J. Sandifer, and B. Kowalski, "Sparingly Selective Ion-Selective Electrode Arrays for Multicomponent Analysis," *Anal. Chem.* 60 66–71 (1988).

This paper describes experiments and a theoretical basis that demonstrate the superior analytical results that can be obtained by the use of sparingly selective ISEs. The principal requirements for successful analyses are stable electrodes and reproducible characterization/calibration of the array.

22. R. Müller and E. Lange, "Multidimensional Sensor for Gas Analysis," *Sens. Actuators* 9 39–48 (1986).

Sensors used for the work described in this paper are four MOS gas sensors coated with layers of different types of zeolite filters. Data analysis used pattern recognition. The paper points out that with MOS sensors, one can measure two characteristics of the signal change resulting when gas is introduced, namely, the slope at the beginning of the transient, and the stationary value. (Although not discussed by the authors, their analysis assumes that the response of the sensor is slow compared to the rate at which the sample is introduced. Otherwise, the analysis discussed would depend on the response time of the sensor and the speed of sample introduction.)

23. M. Otto and J. D. R. Thomas, "Chemometrics in Ion-Selective Electrode Potentiometry," *Ion-Selective Electrode Rev.* 8 55–84 (1986).

Section 3.1, "Selectivity and Multicomponent Analysis" discusses arrays of ISEs. The schemes described are based on the Nikolskii extension of the Nernst equation, which takes into account selectivities of ISEs for different ions.

Techniques used in data analysis include Response Surface Methodology (RSM), which is used to judge the adequacy of the model and for drawing conclusions about the performance of the analytical method. Other necessary techniques include singular value decomposition, principal components analysis, partial least squares, rank annihilation, and eigenvector projection or factor analysis – to correct for drift and problems of low selectivity.

24. J. R. Stetter, P. C. Jurs, and S. L. Rose, "Detection of Hazardous Gases and Vapors; Pattern Recognition Analysis of Data from an Electrochemical Sensor Array," *Anal. Chem.* 58 860–866 (1986).

Work reported in this paper used four different electrochemical sensors operated in four different modes. Results were analyzed using pattern recognition techniques.

25. M. Otto and J. D. R. Thomas, "Model Studies on Multiple Channel Analysis of Free Magnesium, Ca, Na, and K at Physiological Concentration Levels with Ion-Selective Electrodes," *Anal. Chem.* 57 2647-2651 (1985).

Work described in this paper used PLS analysis of ISE data to analyze mixtures at concentrations typical of human intracellular, urine, and serum fluids.

26. S. Zaromb and J. R. Stetter, "Theoretical Basis For Identification and Measurement of Air Contamination Using An Array of Sensors Having Partly Overlapping Selectivities," *Sens. Actuators* 6 225-243 (1984).

This is an early, general paper on the use of electrochemical sensor arrays for mixture analysis. Sensors were operated in different modes, which is said to mean using, for example, different pre-catalysts, different temperatures, chemically selective filters, or any combination of circumstances that affect sensor output in a predictable manner. The analysis techniques used were capable of recognizing nonlinear character.

### 3.2 SEMICONDUCTOR SENSOR ARRAYS

1. H. V. Shurmer, "The Electronic Nose," *Anal. Proc.* 31 39-40 (1994).

Shurmer reviews the work done at the University of Warwick (UK) on artificial olfactory sensing systems. While most of the work has involved tin oxide semiconductor vapor sensor arrays, some work is also done using conducting polymer sensor arrays. Data analysis work has been focused on the use of artificial neural networks. Considerable effort has been devoted to identifying the optimum sensor output and developing the best equipment for making those measurements (conductance measured by a conductivity bridge). Effort is now focused on development of standardized odors.

2. J. A. de Agapito, L. de Agapito, M. Schneider, R. Garcia Rosa, and T. de Pedro, "Fuzzy Logic Applied to Gas Sensors," *Sens. Actuators* B15-16 105-109 (1993).

Fuzzy logic in this paper refers to a combination of knowledge-based expert systems and fuzzy logic. The technique is applied to measurement of the concentration of NO in air that also contains CO and water vapor. Three sensors are used.

3. S. W. Moore, J. W. Gardner, E. L. Hines, W. Göpel, and U. Weimar, "A Modified Multilayer Perceptron Model for Gas Mixture Analysis," *Sens. Actuators* B15-16 344-348 (1993).

This paper describes the application of neural networks to analysis of gas mixtures in air using  $\text{SnO}_2$  sensors. The aim was to design a neural net paradigm for analysis of  $\text{H}_2$ ,  $\text{CH}_4$ , CO, and  $\text{CO}_2$ . Sensors used included Figaro sensors types 813, 815, and 822, along with type 812 doped with a microgram of Pd, Au, and Rh, respectively - six sensors all told. Individual sensor responses were nonlinear and  $\text{CO}_2$  data were in the noise. Previous

research on electronic noses (Items 8, 10, 11, and 14 on pages 3-9 and 3-10) shows that neural nets perform well; work on quartz resonator arrays and on a piezoelectric gas sensor array shows that neural nets hold considerable promise in gas mixture analysis. The paper uses what is called a multilayer perceptron model of a neural net together with a back-propagation algorithm for the learning paradigm as most suitable for classification of combustible gas mixtures.

4. M. S. Nayak, R. Dwivedi, and S. K. Srivastava, "Transformed Cluster Analysis: An Approach to the Identification of Gases/Odours Using an Integrated Gas-Sensor Array," *Sens. Actuators* B12 103-110 (1993).

This paper presents another data analysis method, this one tailored to analysis of data from four different types of semiconductor sensor ( $\text{SnO}_2$ ,  $\text{ZnO}$ ,  $\text{MoO}$ , and  $\text{CdS}$ ). Vapors used were acetone, methyl ethyl ketone, carbon tetrachloride, and xylene. The authors state, "It has been observed that while the technique is useful for the analysis of one or a mixture of two gases, much work has still to be carried out to develop optimized gas sensors, the data from which can be reliably used for analysis using the TCA method."

5. X. Wang, J. Fang, P. Carey, and S. Yee, "Mixture Analysis of Organic Solvents Using Nonselective and Nonlinear Taguchi Gas Sensors with Artificial Neural Networks," *Sens. Actuators* B13-14 455-457 (1993).

The prototype gas analysis system described in this paper uses eight nonselective, nonlinear  $\text{SnO}_2$  sensors to analyze two- and three-component mixtures of benzene, toluene, acetone, and trichloroethylene. Sensors included a temperature sensor. Data analysis used multivariate calibration algorithms, including neural nets.

6. X. Wang, S. Yee, and P. Carey, "An Integrated Array of Multiple Thin-Film Metal Oxide Sensors for Quantification of Individual Components in Organic Vapor Mixtures," *Sens. Actuators* B13-14 458-461 (1993).

The research described in this paper used eight thin film metal oxide sensors –  $\text{SnO}_2$ ,  $\text{ZnO}$ ,  $\text{TiO}_2$ , and  $\text{WO}_3$ , with and without a heated Pd catalyst. Sensor response data were processed by a neural net system for calibrating the sensor array and predicting vapor concentrations. Mixtures used were methanol and acetone. Results show quantitative discrimination.

7. W. P. Carey and S. S. Yee, "Calibration of Nonlinear Solid-State Sensor Arrays Using Multivariate Regression Techniques," *Sens. Actuators* B9 113-122 (1992).

This paper describes the use of eight semiconductor sensors to analyze two- and three-component mixtures of toluene, benzene, acetone, and trichloroethylene. Calibrations were performed using two linear-based parametric modeling techniques, PLS and NLPLS, and two nonparametric modeling methods – multivariate adaptive regression splines and projection pursuit regression. Nonparametric techniques for calibration of nonlinear sensors produces much better calibration results than linear parametric methods.

8. J. W. Gardner, E. L. Hines, and H. C. Tang, "Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition Techniques, Part 2. Artificial Neural Networks," *Sens. Actuators* B9 9–15 (1992).

Performance of pattern recognition techniques depends on the choice of parametric expression used to define the array output (see Item 11 below). At present, there is no generally agreed choice of this parameter for either individual semiconducting oxide sensors or arrays of those sensors. Data used in this paper consist of responses of 12  $\text{SnO}_2$  sensors to five alcohols and three beers. Five different parametric expressions are used for sensor response: fractional conductance change, relative conductance, log of conductance change, and normalized versions of the last two. Data were processed using neural nets. Nearly all nets can correctly identify all alcohols used, so the paper's results use the total sum of squared network errors to determine relative performance. Best results were obtained using fractional change in conductance as the data parameter; results using relative conductance gave "errors" about twice as large. The neural net outperforms principal component and cluster analysis (discussed in the paper reviewed in Item 11), primarily because it has significant capability to cope with nonlinear and highly correlated data.

9. J. W. Gardner, H. V. Shurmer, and T. T. Tan, "Application of an Electronic Nose to the Discrimination of Coffees," *Sens. Actuators* B6 71–75 (1992).

Twelve  $\text{SnO}_2$  sensors were used to sense the headspace of three different coffee packs that contained two different blends and two roasts and a coffee that had been roasted in six different ways. Results were analyzed in several ways to show that  $\text{SnO}_2$  sensors can discriminate between blend and roasting time in coffee.

10. H. V. Shurmer and J. W. Gardner, "Odour Discrimination with an Electronic Nose," *Sens. Actuators* B8 1–11 (1992).

This paper is a review of research conducted at the University of Warwick (UK). Summaries are given of current work on various sensor developments and on implementation of pattern recognition techniques that include correlation, principal component analysis, cluster analysis, and artificial neural networks. Application areas most likely to be of interest in the next decade include the development of neural network computer chips that can be used for such work.

11. J. W. Gardner, "Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition, Part 1. Principal Component and Cluster Analysis," *Sens. Actuators* B4 109–115 (1991).

Mathematical models for the response of individual sensors and arrays of  $\text{SnO}_2$  sensors are based on the barrier-limited electron mobility model. The fractional change in conductance is the optimal response parameter to characterize array performance instead of the more usual relative conductance (see Item 8 above). The data used for this research were obtained from 12  $\text{SnO}_2$  sensors exposed to five alcohols and six beers and spirits. The data were analyzed using pattern recognition methods. Results show that theoretically derived normalization of data improves the analysis results. Improvements suggested are (a) use only

five sensors, and perhaps SAWs as well, or (b) use more generalized analysis techniques such as neural networks.

12. R. Müller, "High Electronic Selectivity Obtainable with Nonselective Chemosensors," *Sens. Actuators* B4 35-39 (1991).

This short review paper on different methods of pattern recognition for chemosensors shows that an essential first step is identification of substances. Two examples are used to illustrate the needed definitions of electronic selectivity and sensitivity. Usually the real number of distinguishable substances (patterns) is much lower than the theoretical limit. Two methods to improve this situation are tailoring the cross-sensitivities of the sensor elements and reducing the number of sensor elements.

For linearly responding sensors, PLS has been used with great success. If responses are nonlinear, different approaches must be used. Piecewise linear regression is possible when substances are identified first and when their concentrations are known roughly. In many cases it is possible to transform signals into an "artificial superposition plane" – this transformed least squares (TLS) method works best when the substances are identified first. Partial model building (see Item 15 on page 3-11) uses two steps sequentially; first is identification of substances or components of the mixture, and next is application of a substance-specific PLS or TLS method. The paper stresses that identification is essential for nonlinear sensor systems. The example given is analysis of a mixture of CO, n-butane, and methane using four semiconductor sensors. Another example is a mixture of  $\text{NF}_3$ ,  $\text{CF}_2\text{Cl}_2$ ,  $\text{C}_2\text{HCl}_3$ , CO,  $\text{H}_2$ , and  $\text{C}_2\text{H}_2$  using four electrochemical cells ( $\text{H}_2$ ,  $\text{H}_2\text{S}$ , CO, and  $\text{Cl}_2$ ) with thermal catalysts.

13. A. R. Newman, "Electronic Noses," *Anal. Chem.* 63 585A-588A (1991).

This *Analytical Chemistry* "Focus" article gives a summary of papers presented at the 1991 Pittsburgh Conference in Chicago on electronic chemical analysis of air. The majority of the article is a discussion of work by Gardner et al. on artificial noses (see Items 1, 8-11, 14, 17, and 19 on pages 3-7 thru 3-12) and the very careful work by Grate et al (see the reviews in Section C of this chapter) on the design of coatings for organic vapor analysis using acoustic sensors.

14. J. W. Gardner, E. L. Hines, and M. Wilkinson, "Application of Artificial Neural Networks to an Electronic Olfactory System," *Meas. Sci. Technol.*, 1 446-451 (1990).

This paper summarizes the development of the  $\text{SnO}_2$  electronic nose and introduces data analysis using neural networks: "Results of the classification of the signal spectra measured from several alcohols are reported and they show considerable promise for the future application of Artificial Neural Networks within the field of sensor array processing."

15. G. Horner and C. Hierold, "Gas Analysis by Partial Model Building," *Sens. Actuators* B2 173-184 (1990).

Analysis of data from nonselective sensors requires two steps: classification and quantification. This paper describes a nonlinear pattern recognition model that combines a

nonlinear transformation of the data with a linear regression model. This method, which the authors call a TLS model, was applied to a set of calibration data from  $\text{SnO}_2$  sensors used for two-component mixture analysis. To overcome certain difficulties, the authors use a "preclassification" of the several classes of possible gas mixtures, enabling a more specific TLS model to be applied in quantification. This procedure is called partial model building.

Data used were obtained from up to eight  $\text{SnO}_2$  sensors, and sometimes four additional sensors (for temperature, pressure, etc.). Gases used included methane, n-butane, CO, and their mixtures in synthetic air. Partial modeling results for binary mixtures of three different gases are significantly better than those without the classification procedure.

16. K. D. Schierbaum, U. Weimar, and W. Göpel, "Multicomponent Gas Analysis: An Analytical Chemistry Approach Applied to Modified  $\text{SnO}_2$  Sensors," *Sens. Actuators* B2 71–78 (1990).

The essential part of this complex paper is the methodology for chemically changing the selectivity of  $\text{SnO}_2$ -type sensors in a systematic way to achieve reversible sensor signals in the analysis of mixtures of CO and methane in moist air.

17. H. V. Shurmer, J. W. Gardner, and P. Corcoran, "Intelligent Vapour Discrimination Using a Composite 12-Element Sensor Array," *Sens. Actuators* B1 256–260 (1990).

This paper describes the use of 12  $\text{SnO}_2$  sensors for discrimination of odors. Each sensor has significant sensitivity to a broad spectrum of gases. Mixtures used in the research were of two types: (a) several alcohol compounds and (b) beverages and alcoholic spirits. Data analyses used a weighted, fault-tolerant, least-squares method. Results show that such systems successfully discriminate between methanol, ethanol, propan-2-ol, and butan-1-ol, but are not so good at discriminating between beverages and spirits and mixtures of same.

18. U. Weimar, K. D. Schierbaum, W. Göpel, and R. Kowalkowski, "Pattern Recognition Methods for Gas Mixture Analysis: Application to Sensor Arrays Based upon  $\text{SnO}_2$ ," *Sens. Actuators* B1 93–96 (1990).

This paper reports research on determination of single gases in air ( $\text{H}_2$ ,  $\text{CH}_4$ , and CO) from specific response patterns of chemically modified  $\text{SnO}_2$  sensors. Data were analyzed in two different multicomponent approaches; one in which the conductivity of several chemically modified sensors is measured, and the other in which measurements are made of different parameters from a single sensor. The former approach gives somewhat better results than the latter for the mixtures and sensors used.

19. H. V. Shurmer, J. W. Gardner, and H. T. Chan, "The Application of Discrimination Techniques to Alcohols and Tobaccos Using Tin-Oxide Sensors," *Sens. Actuators* 18 361–371 (1989).

This early paper discussed the use of  $\text{SnO}_2$  sensors for odor discrimination research. The techniques used have been superseded by later work.

20. C. Hierold and R. Müller, "Quantitative Analysis of Gas Mixtures With Non-Selective Gas Sensors," *Sens. Actuators* 17 587-592 (1989).

Hierold and Müller discuss the use of nonselective, nonlinear sensors with data transformation and pattern recognition analysis. To illustrate the method, data were obtained from commercially available  $\text{SnO}_2$  sensors exposed to mixtures of  $\text{CO}$  and  $\text{CH}_4$ . Data transformation used is a combination of data transformation and linear regression techniques called transformed least squares.

21. H. Abe, S. Kanaya, Y. Takahashi, and S.-I. Sasaki, "Extended Studies of the Automated Odor-Sensing System Based on Plural Semiconductor Gas Sensors With Computerized Pattern Recognition Techniques," *Anal. Chim. Acta* 215 155-168 (1988).

Research described in this paper used an odor sensing system described in the first paper of the series (see Item 22). In this work, that system was used to study 47 compounds. Analysis techniques used included three different pattern recognition schemes; k-NN, simplex, and potential function methods. Best results were obtained using the potential function method.

22. H. Abe, T. Yoshimura, S. Kanaya, Y. Takahashi, Y. Miyashita, and S.-I. Sasaki, "Automated Odor-Sensing System Based on Plural Semiconductor Gas Sensors and Computerized Pattern Recognition Techniques," *Anal. Chim. Acta* 194 1-9 (1987).

This paper presents a scheme for classification and identification of odors. Data were obtained from eight semiconductor gas sensors that have different gas sensitivities. Data for 30 substances were analyzed by cluster analysis. Sensors used were generally sensitive to  $\text{NH}_3$ ;  $\text{CO}$ ; alcohols; and mixtures of alcohols,  $\text{CO}$ , chlorofluoro-carbons, in addition to three that were generally sensitive to combustible gases. Substances used for testing included the following: methyl acetate, diethyl ether, acetone, ethyl acetate, acetyl acetone, methyl formate, ethyl formate, acetaldehyde, allyl acetate, isobutyl methyl ketone, pentyl formate, isopropyl acetate, tertiary pentanol, 2-methylcyclohexanone, 3-methylcyclohexanone, 4-methylcyclohexanone, methyl butyrate, isobutylaldehyde, menthone, acetic acid, methyl valerate, furfural, acrylic acid, phenetole, propionic acid, dioxane, chloroform, pyridine, 2-methyl-3-butanone, and pyrrole.

### 3.3 SAW SENSOR ARRAYS

The most comprehensive work on the use of SAW sensor arrays for chemical mixture analysis seems to be that reported since 1986 by J. W. Grate and coworkers, first at the U.S. Naval Research Laboratory and now at the DOE/Battelle Pacific Northwest Laboratory. Their work is aimed at the development of models that can be used to choose the number of SAWs for an array and to select SAW coatings for those sensors so as to achieve optimum selectivity and sensitivity for a specified mixture analysis problem. Six papers on that subject are reviewed here. The most comprehensive paper, a review paper published in 1991, is discussed first. The other papers follow in descending chronological order.

The basic idea in Grate's work is that the appropriate number of SAWs and the types of coatings to be used on those SAWs can be derived from a detailed understanding of the chemical

interactions that occur between the analyte molecules and the coatings. The particular focus is on the use of solubility parameters to indicate the relative magnitudes of basic types of chemical interactions (i.e., dipole-dipole, hydrogen bonding, induced-dipole/induced-dipole, etc.).

1. J. W. Grate and M. H. Abraham, "Solubility Interactions and the Design of Chemically Selective Sorbent Coatings for Chemical Sensors and Arrays: Review Paper," *Sens. Actuators B3* 85-111 (1991).

Grate and Abraham explain in great detail the chemical basis for the methods being developed.

By understanding sorption phenomena and solubility interactions, "interactive materials can be chosen for vapor detection that will collect and concentrate analyte molecules selectively at the sensor's surface. . . .

" . . . when it is necessary to detect an analyte at trace concentrations in field environments where potential interferences (such as humidity) may be present at orders of magnitude higher concentrations, . . . the use of pattern recognition techniques in combination with an array of chemical sensors offers improved selectivity. . . .

"The success of the array detector approach depends on the amount and quality of information being provided by the sensors. To be useful, each sensor must give unique information" (i.e., sensors and sensor coatings must be carefully chosen). "Each sensor in the array must be selective for different classes of vapor. . . . However, . . . one must also design sensors that will be selective for known and possibly unknown potential interferences." Thus the need to take a "comprehensive approach to chemical selectivity so that an array can gather as much information as possible about any vapor which may sorb and produce sensor signals."

The paper focuses "on a solubility model for absorption" in SAWs exposed to "organic vapors." Emphasis is "on bulk absorption (as opposed to surface adsorption) of neutral organic vapors into soft coating materials containing neutral organic functional groups. In addition, the discussion (is) limited to reversible sensors. The issue of reversibility is closely related to the issue of chemical selectivity. Weak interactions between the vapor and sensor coating will produce sensors with good reversibility and little hysteresis. However, such sensors may not have sufficient sensitivity and selectivity to be widely useful. Very strong interactions may improve the sensitivity and selectivity, but can result in sensors which are irreversible or only slowly reversible." Necessary "balance can be achieved by using sensor coatings which interact with vapors via solubility interactions."

The method explained in the paper uses solvation parameters that have been developed to describe the types of solubility interactions occurring between non-ionic, organic functional groups. Solubility interactions considered are:

- induced-dipole/induced-dipole (dispersion) interactions,
- dipole/induced-dipole (dipole induction) interactions,

- dipole/dipole (dipole orientation) interactions, and
- hydrogen-bonding interactions.

As explained in the paper, the characterization of the total absorption resulting from these interactions can be made in terms of the partition coefficient, K, which measures the equilibrium distribution of solute molecules between the gas phase and the solvent phase. K is defined as the ratio of concentration of the solute in the stationary phase (sensor coating – assuming thin film),  $C_s$ , to the concentration of the solute in the vapor phase,  $C_v$ . For SAW delay lines, the frequency shift resulting from absorption of vapor is given by

$$\Delta f_v = \Delta f_s C_v K / \rho,$$

where  $\Delta f_s$  is the frequency shift observed when the sorbent coating was applied to the bare SAW device, and  $\rho$  is the density of the coating material.

K can also be expressed in terms of certain solute parameters (i.e., hydrogen bond parameters, a dispersion parameter, and parameters for dipolarity and polarizability).

Much of the work by Grate et al. has focused on the necessary evaluation of many coatings exposed to many organic vapors so as to derive a self-consistent set of solubility parameters for the coatings/vapors interactions. This work is discussed in great detail in the paper, and references are given for reports of the applications of such results to analysis of organic vapors using this design technique and various types of pattern recognition techniques. Several of those papers are reviewed following this review.

The review paper describes the process of developing an array. Optimization of an array for an application depends on boundary conditions [i.e., the analyte to be detected (compound or class of compounds) and associated questions – how fast, reversibly, sensitivity required, are there other compounds that could cause sensor signals]. Selection of coating materials to use starts with a large set of carefully chosen materials, to include materials in all of the solubility interaction classes, as well as materials optimized for the analyte and materials optimized for known interferents. Once the materials are chosen and sensors made, a “training set” of data is collected from which the pattern recognition algorithm is developed. The design of the training set is not a simple task. The next step is the use of statistical pattern recognition techniques to decrease the number of coatings used in the array. These procedures reveal sensors that are providing needlessly redundant information.

When vapor mixtures are involved, two types of problems may occur. One is simply the detection of the analyte in the environment (i.e., it is not important to identify all components of the mixture). If the array must detect more than one analyte, then a separate algorithm must be developed for each analyte. The other type of problem is simultaneous analysis of all mixture components.

The paper stresses an important point: pattern recognition techniques do not operate by checking the observed pattern against a library of vapor signatures. Instead, pattern recognition techniques plot the “pattern of response” as a point in n-dimensional space,

where  $n$  is the number of sensors and each axis represents the response of the sensor for that axis. Any vapor that is dissimilar to the target analyte will be plotted in a different region of the  $n$ -dimensional space. Consequently, it is important to choose coating materials carefully – a diverse set of sensors with strong, selective, and uncorrelated responses will more effectively spread the different vapors into different regions of  $n$ -space, facilitating discrimination.

This paper makes it clear that successful results with SAW arrays require that the analyte (or analytes) be identified and a reasonable idea of the concentrations of all analytes and possible interferents must be available. With this information, one can then build sensors that have appropriate coatings, train them, derive the optimum set of sensors, and finally get meaningful test results that can be analyzed. It will not work to assume you can develop SAW array sensors that will analyze for compound A in the presence of unknown compounds.

2. S. L. Rose-Pehrsson, J. W. Grate, and M. Klusty, *Chemical Sensors, Proc. Chemical Sensors Symposium, Honolulu, 1993*, The Electrochemical Society, Pennington, NJ, 1993, pp. 609–620.

Four temperature-controlled SAWs with preconcentrators for sampling were used to analyze two- and three-component mixtures of several organophosphorus and organosulfur vapors (chemical warfare agents).

3. J. W. Grate, S. L. Rose-Pehrsson, D. L. Venezky, M. Klusty, and H. Wohltjen, “Smart Sensor System for Trace Organophosphorus and Organosulfur Vapor Detection Employing a Temperature-Controlled Array of Surface Acoustic Wave Sensors, Automated Sample Preconcentration, and Pattern Recognition,” *Anal. Chem.* **65** 1868–1881 (1993).

Four temperature-controlled SAWs and pattern recognition analysis, together with an automated sampling system (including thermally desorbed preconcentrator tubes), were used to detect organophosphorus and organosulfur vapors.

4. S. L. Rose-Pehrsson, J. W. Grate, D. S. Ballantine, Jr., and P. C. Jurs, “Detection of Hazardous Vapors Including Mixtures Using Pattern Recognition Analysis of Responses from Surface Acoustic Wave Devices,” *Anal. Chem.* **60** 2801–2811 (1988).

Ten SAWs coated with polymers were used to sense several hazardous vapors and two-component mixtures of the vapors.

5. J. W. Grate, A. Snow, D. S. Ballantine, Jr., H. Wohltjen, M. H. Abraham, R. A. McGill, and P. Sasson, “Determination of Partition Coefficients from Surface Acoustic Wave Vapor Sensor Responses and Correlation with Gas-Liquid Chromatographic Partition,” *Anal. Chem.* **60** 869–875 (1988).

This paper compares partition coefficients from gas chromatography and SAW measurements using 158 MHz SAWs and fluoro-polyol coatings. Results give the same trends, but absolute values differ for some vapors.

6. D. S. Ballantine, Jr., S. L. Rose, J. W. Grate, and H. Wohltjen, "Correlation of Surface Acoustic Wave Device Coating Responses with Solubility Properties and Chemical Structure Using Pattern Recognition," *Anal. Chem.* 58 3058–3066 (1986).

Twelve different SAWs/coatings were used to sense 11 different chemicals, and the SAW responses were then correlated with the solubility properties and coating structures. Data analysis used included principal component analysis and hierarchical clustering.

Another research group that has performed significant work on the use of acoustic sensor arrays for multicomponent analysis is the University of Washington (UW) group spearheaded by W. P. Carey. Reviews of four UW papers follow.

7. W. P. Carey and B. R. Kowalski, "Monitoring a Dryer Operation Using an Array of Piezoelectric Crystals," *Anal. Chem.* 60 541–544 (1988).

An array of six SAWs was used to analyze mixtures of (a) water vapor and acetone and (b) water vapor, hexane, and methylene chloride. Multicomponent prediction of solvent concentrations in exhaust gases was used as the indicator of degree of dryness. Tests made used a laboratory simulator, not a commercial dryer. There is no mention of what is being dried, other than "a wet product" so it is unclear whether a real case would have interferences. Analysis was done using a principal component regression technique – partial least squares couldn't be used because of insufficient calibration data.

8. W. P. Carey, K. R. Beebe, and B. R. Kowalski, "Multicomponent Analysis Using an Array of Piezoelectric Crystal Sensors," *Anal. Chem.* 59 1529–1534 (1987).

Nine SAWs were used for multicomponent organic vapor analysis. Analyses of two- and three-component mixtures used multiple linear regression and PLS techniques.

9. W. P. Carey, K. R. Beebe, B. R. Kowalski, D. L. Illman, and T. Hirschfeld, "Selection of Adsorbates for Chemical Sensor Arrays by Pattern Recognition," *Anal. Chem.* 58 149–153 (1986).

Twenty-seven SAWs were used to sense 14 analytes to determine whether some number less than 27 sensors would produce essentially the same quality results as the full set. Results showed that approximately seven sensors gave 95% of the available information. A selection procedure for use in other systems would clearly depend on sorption interactions, but the discussion also mentions iterative use of analysis techniques together with perhaps some chemical interaction insights.

10. W. P. Carey and B. R. Kowalski, "Chemical Piezoelectric Sensor and Sensor Array Characterization," *Anal. Chem.* 58 3077–3084 (1986).

This paper describes a model for the response of SAWs. Data obtained from two sets of arrays for three- and seven-component mixtures were analyzed using principle components analysis.

A series of four papers published by a Japanese group, the first in 1989, describes research on the use of acoustic devices as artificial noses. The papers are similar to those published by the

University of Warwick (UK) group (reviewed in Sections A and B), but the Warwick papers give more detailed explanations of the work than the Japanese papers.

11. T. Nakamoto, A. Fukuda, and T. Moriizumi, "Perfume and Flavour Identification by Odour-Sensing System Using Quartz-Resonator Sensor Array and Neural-Network Pattern Recognition," *Sens. Actuators* B10 85-90 (1993).

An array of six quartz-resonator sensors (the same array as used for earlier work – see Items 12, 13, and 14 below) was used to sense perfumes and food flavors. Perhaps the most interesting result is that when the ratio of foreign substances to orange flavor is more than about 0.5%, the sample could be distinguished from the pure orange sample, with a detection limit much like normal people's limit.

12. T. Nakamoto, A. Fukuda, T. Moriizumi, and Y. Asakura, "Improvement of Identification Capability in an Odor-Sensing System," *Sens. Actuators* B3 221-226 (1991).

The same six-resonator array as in the above review was used to sense alcoholic drinks.

13. T. Nakamoto, K. Fukunishi, and T. Moriizumi, "Identification Capability of Odor Sensor Using Quartz-Resonator Array and Neural-Network Pattern Recognition," *Sens. Actuators* B1 473-476 (1990).

This paper is similar to the other papers by the Japanese group, except that 18 sensors were used to sense five Japanese whiskies and five other alcoholic spirits.

14. K. Ema, M. Yokoyama, T. Nakamoto, and T. Moriizumi, "Odour-Sensing System Using a Quartz-Resonator Sensor Array and Neural-Network Pattern Recognition," *Sens. Actuators* 18 291-296 (1989).

Six quartz resonators were used to sense 11 kinds of alcoholic spirits. Data were analyzed using neural network pattern recognition techniques.

Reviews of five papers on SAW arrays follow.

15. A. J. Ricco, S. J. Martin, R. M. Crooks, C. Xu, and R. E. Allred, *Chemically Sensitive Interfaces on SAW Devices*, in *Chemically Sensitive Interfaces*, T. E. Mallouk and D. J. Harrison, eds., American Chemical Society Symposium Series xxx, American Chemical Society, Washington, D.C. 1994, Chapter xx (to be published in 1994).

This paper describes three approaches to using nonselective interfaces for chemical sensing: one is time dependence of the response of a single SAW, which depends on details of chemical interactions between vapor and coating (authors dub this "time-resolved permeation transients"); another is multifrequency SAW devices; and the third is an array of nonselective SAWs. The first approach is not useful for simultaneous analysis of multicomponent mixtures. It is unclear how far the multifrequency approach could be taken (i.e., how many components could conceivably be analyzed, and how many might be possible realistically).

16. S. J. Patrash and E. T. Zellers, "Characterization of Polymeric Surface Acoustic Wave Sensor Coatings and Semiempirical Models of Sensor Responses to Organic Vapors," *Anal. Chem.* 65 2055–2066 (1993).

The approach Patrash and Zellers describe is very similar to the Grate approach, but it is probably less comprehensive. The experimental work seems to be limited by using only four SAWs.

17. A. J. Ricco and S. J. Martin, "Multiple-Frequency SAW Devices for Chemical Sensing and Materials Characterization," *Sens. Actuators* B10 123–131 (1993).

This research report describes the use of two multifrequency devices: a four-frequency device and a seven-frequency device. The measured responses were wave velocity and attenuation. Results show that multifrequency SAW devices can function as selective chemical sensors without selective coatings. In addition, for three organic vapors studied, the data also determine the concentrations of three vapors as well as that there are three vapors. Furthermore, these techniques enhance the ability of acoustic sensors to distinguish changes in other perturbations (e.g., temperature and pressure, from changes in mass, viscoelastic properties, and thin-film conductivity).

18. E. T. Zellers, T.-S. Pan, S. J. Patrash, M. Han, and S. A. Batterman, "Extended Disjoint Principal-Components Regression Analysis of SAW Vapor Sensor-Array Responses," *Sens. Actuators* B12 123–133 (1993).

This paper describes the use of what is called "extended disjoint principal components regression" (EDPCR) for analysis of data from 10 SAWs exposed to 9 organic vapors. Data for mixtures – for all 36 possible binary mixtures and all 84 possible ternary mixtures – were calculated from measured responses for the single vapors with the assumption that the sensor responses were additive, but there is little discussion of the probable validity of the assumption.

The mixture data were analyzed using several different multivariate techniques. The conclusion reached is that the EDPCR technique is an alternative to other methods that have been used and is well-suited to SAW arrays where responses are linear and additive. Regarding the assumption of additivity of response, the paper states that it should be possible to develop methods for nonadditive mixture responses.

From an experimental point of view, it seems likely that this paper is only a numerical exercise whose real value is obscure at the best. Perhaps that opinion is only a reflection of the inadequate experience of the reviewer.

19. A. Schmautz, "Application-Specific Design of a Piezoelectric Chemosensor Array," *Sens. Actuators* B6 38–44 (1992).

The approach Schmautz describes is similar to Grate's approach, but uses partition coefficients and heats of adsorption to derive criteria for selecting coatings. Data used were from four bulk acoustic wave resonators (10 MHz) and one uncoated crystal, but there is no

mention of temperature control of the sensors. The scheme was applied to the analysis of anesthetic vapors using pattern recognition data processing.

### 3.4 GENERAL ARTICLES

In this section we have included short reviews of several papers that describe sensor array research using more than one type of sensor as well as other papers of a more general nature than those devoted to one type of sensor. One paper that seems highly significant to us for its application to the DOE ERWM programs describes work on analysis of all types of mixtures (see Item 1 below).

1. Y.-Z. Liang, O. M. Kvalheim, and R. Manne, "White, Grey and Black Multicomponent Systems. A Classification of Mixture Problems and Methods for Their Quantitative Analysis," *Chemom. Intell. Lab. Syst.* 18 235-250 (1993).

This is one of the series of "Tutorial" papers that periodically appears in this journal. The authors discuss multivariate calibration and resolution methods for handling samples of chemical mixtures from the point of view of the analytical chemist. Advantages and limitations of available multivariate calibration and resolution methods are discussed for three types of mixtures, called – following Chinese usage – "white," "grey," and "black" multicomponent systems.

White systems are those in which – in spectroscopic terminology – spectra of the chemical species present in the sample, as well as spectra from samples that contain possible interferents, are available, and the concentrations of all the desired analytes are known for the "training set" samples. "Methods for this kind of analysis are relatively mature and do almost always provide excellent results (5). Exceptions are ill-conditioned and seriously nonlinear systems (5,6)."

Black systems are those for which there is no *a priori* information regarding the chemical composition of the samples. "Several methods for resolving this type of analytical system have been developed. Most of them are based on instrumental techniques giving data of matrix form (i.e., hyphenated chromatography). In the quantitation step, one needs samples of known concentrations of the resolved chemical components. Effectively, this strategy transforms the black system into a white or grey system, depending on whether the system is fully or partially resolved."

Grey systems lie between the white and black, their basic characteristic being incomplete knowledge about the qualitative chemical composition. The aim of the analysis is to determine quantitatively the concentrations of certain analytes in the presence of unknown coexisting interferents. "Grey systems are frequently encountered in the analysis of real-world samples. Several powerful multivariate calibration methods have been developed to cope with vectorial data (each sample being represented by a spectrum, for instance) and with bilinear data of matrix form (each sample being represented, for example, by a matrix of chromatographic elution profile-time spectra)."

The citation given by the authors at the start of their paper is of historical interest:

“Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. . . . If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would cause a rapid and widespread degeneration of that science.” Auguste Comte, *Philosophie Positive* (1830).

This is the only paper we have found that treats what will often be the situation faced by the DOE in site characterization (i.e., sometimes grey, sometimes black mixtures/systems). We also note that we found no mention of chemical sensors in this paper. The authors conclude:

“Quantification of multicomponent systems by means of multivariate calibration and resolution methods has released the chemist from constraints imposed by classical analytical methods. On the other hand, the application of these methods has placed new burdens upon the shoulders of the cautious analytical chemist. He must not only be a knowledgeable chemist, but also needs competence in applied mathematics, computer science, and statistics in order to be at the forefront of his profession. These requirements reflect the development of chemical instrumentation during the last decade and the penetration of chemical measurement techniques, to an ever-increasing extent, into new areas.”

“Despite the prophesy of Auguste Comte, analytical chemistry has not deteriorated or lost its spirit because of this. On the contrary, analytical chemistry has revived and expanded to become the science of measurements and will surely continue to do so in the future.”

2. W. P. Carey, *Multicomponent Vapor Monitoring Using Arrays of Chemical Sensors*, in *Pollution Prevention in Industrial Processes*, J. J. Breen and M. J. Dellarco, eds., American Chemical Society Symposium Series 508, American Chemical Society, Washington, D.C., 1992, Chapter 21, pp. 258–269.

This paper describes the process of analyzing multicomponent vapor mixtures using arrays of partially selective sensors and multivariate statistics. Examples discussed include data obtained using both quartz crystal acoustic sensors and metal oxide semiconductor sensors. Applications discussed include process analytical chemistry and pollution prevention. Calibration techniques used include both linear and nonparametric regression techniques.

3. F. A. M. Davide and A. D'Amico, “Pattern Recognition from Sensor Arrays: Theoretical Considerations,” *Sens. Actuators* A32 507–518 (1992).

This paper describes a general approach (mathematical) to the design and evaluation of multisensor measurements based on nonselective reproducible sensors.

4. J. W. Gardner, P. N. Bartlett, G. H. Doss, and H. V. Shurmer, *The Design of an Olfactory System*, in *Chemosensory Information Processing*, NATO ASI Series, Volume H-39, D. Schild, Ed, Springer-Verlag, Berlin, 1990.

This article by researchers at the University of Warwick (UK) reviews several aspects of sensor array applications in the development of artificial olfactory systems. The paper gives an excellent summary of the sense of smell in humans, including current models of the interactions of odorants with human sensory membranes, a short section on the history and evolution of electronic noses, and a thorough discussion of "artificial odour transducers." The types of transducers described include (a) inorganic semiconductor gas sensors/transducers – MOSFET transistors, capacitors, Schottky barrier diodes, hot wire resistors, SAW devices, and metal oxide resistors, (b) conducting polymers – pyrrole and several of its derivatives, furan, thiophene and several of its derivatives, and indole and a few of its derivatives, (c) semiconducting metal phthalocyanines, and (d) Langmuir-Blodgett film devices. The most widely used sensor for artificial odor sensing is the semiconducting tin oxide (or tin oxide doped with a wide variety of catalysts) device.

Parts of the final major section of the article may be helpful:

"Several developments during the last fifty years have provided a tremendous impetus for the design of algorithms for processing signals from distributed sensor arrays. These include the discovery of RADAR and SONAR which generated the need for complex processing algorithms; the development of semiconductor technology for the fabrication of  $1,024 \times 1,024$  element CCD arrays for optical imaging; the study of image analysis in such diverse fields as medical tomography; and nondestructive testing. These have produced a plethora of processing algorithms based upon statistical techniques such as Correlation, Maximum Likelihood, Minimum Entropy, etc. However, these sophisticated algorithms are not generally applied to chemical sensor arrays, presumably due to the relatively slow response time of a chemical sensor (1–100 s) compared to an optical sensor (1 ns–1  $\mu$ s).

"The use of integrated sensor arrays in olfaction was postulated by Persaud and Dodd in the early 1980s (4), and designed to overcome problems encountered by selectivity. Its arrival acknowledged the partial failure to identify specific odours or vapours by a single sensing material. The earliest work involved the investigation of different types of semiconducting material in the array, such as metal oxides (7, 8, 9) and Langmuir-Blodgett films (10). Simultaneously, statistical methods were applied to the detection of flammable and toxic vapours by others (11, 12, 13). At Warwick we have applied a variety of algorithms . . . ."

"By far the majority of pattern recognition techniques that have been applied to chemical sensor arrays are based upon standard statistical techniques. However, more powerful techniques are also being applied, such as Principal Components Analysis and Factor Analysis. . . . Artificial Neural Networks offer several advantages over classical algorithms such as fault-tolerance, flexibility, and adaptability. The inherent parallelism leads to a high speed of operation which could be an advantage when processing large arrays. The method is nonparametric and heuristic, this makes it difficult to interpret results but is closer to the real olfactory system and may yet prove to be the most beneficial approach."

5. W. P. Carey, K. R. Beebe, E. Sanchez, P. Geladi, and B. R. Kowalski, "Chemometric Analysis of Multisensor Arrays," *Sens. Actuators* **9** 223–234 (1986).

This paper is a review of multivariate calibration techniques and their possible application to data from partially selective sensor arrays. Analysis techniques discussed include linear and multiple linear regression, principal component regression (which couples factor analysis or principal component analysis with multilinear regression), partial least squares, and rank annihilation factor analysis.

6. B. R. Kowalski and C. F. Bender, "Pattern Recognition. A Powerful Approach to Interpreting Chemical Data," *J. Am. Chem. Soc.* 94 5632–5639 (1972).

S. Wold, one of the founders of the field of chemometrics, says of this paper (14), "The first [paper] was Kowalski's 1972 pattern recognition in the *Journal of the American Chemical Society*. That set the world on fire. That was before the Chemometrics Society started and even before the word chemometrics was invented."

One interesting aspect of the paper is the general statement of the problem treated: "can an obscure property of a collection of objects (elements, compounds, mixtures, etc.) be detected and/or predicted using indirect measurements made on the objects?" The principal example describes the use of spectroscopic data for detection of molecular structural units.

### 3.5 REFERENCES

1. R. Hooke and T.A. Jeeves, *J. Assoc. Comput. Mach.* 8 212 (1961).
2. J. A. Nelder and R. Mead, *Comput. J.* 7 308 (1965).
3. R. Fletcher and M. J. D. Powell, *Comput. J.* 6 163 (1963).
4. K. Persaud and G. H. Dodd, *Nature (London)* 299 352–355 (1982).
5. H. Martens and T. Næs, *Multivariate Calibration*, John Wiley & Sons, New York, 1989.
6. H. Mark, *Appl. Spectros.* 42 832–844 (1988).
7. A. Ikegami and M. Kaneyasu, *Proc. SICE. Japan*, 673–674 (1984).
8. A. Ikegami and M. Kaneyasu, *IEEE* 9 136–139 (1985).
9. M. Kaneyasu, A. Ikegami, A. Arima, and S. Iwanaga, *Trans. Comp. Hybrids Manu. Technol.* 10 267–273 (1987).
10. M. Fard, M.Sc. Thesis, University of Warwick, UK, 1985.
11. S. Zaromb and J. R. Stetter, *Sens. Actuators* 6 225–243 (1984).
12. J. R. Stetter, P. C. Jurs, and S. L. Rose, *Anal. Chem.* 58 860–866 (1986).

13. R. Müller and G. Horner, *Siemens Forsch.-u. Entwickl.-Ber.* 13(3), Springer-Verlag, Berlin, 1986.
14. S. Wold interview in P. Geladi and K. Esbensen, "The Start and Early History of Chemometrics: Selected Interviews. Part 1," *J. Chemom.* 4 337-354 (1990).

## 4. REMARKS/CONCLUSIONS AND RECOMMENDATIONS

### 4.1 REMARKS/CONCLUSIONS

#### 4.1.1 What Is Known in the Beginning?

One noteworthy point that was mentioned only briefly in Chapter 2 is that there are very few papers on the analysis of mixtures **for which the composition and likely concentrations were unknown**. The one notable paper that is markedly different is the paper by Liang et al. (1) on "white, grey, and black" mixture analysis problems.

As demonstrated very clearly by the research of Grate et al. (2), determination of the optimal set of SAWs **requires** detailed knowledge of the chemical interactions between SAW coating and analyte(s) of interest. Knowledge of the relative concentrations of various chemicals in a mixture is a strong secondary need.

To the authors of this report this situation seems to be a significant aspect of the DOE ERWM problem in that, to the best of our knowledge, it must be expected that real world samples of interest to DOE will contain both unknown chemicals and unknown concentrations. Consequently, unless future research uncovers new methods, it will probably be necessary always to perform initial qualitative analyses to identify all significant chemicals, lest some unrecognized component is a strong interferent. The need for information about concentrations is perhaps secondary, but only real-world practice can provide reliable information.

#### 4.1.2 Validation and Verification in Applied Chemometrics

Several papers (not reviewed in Sect. 2) make a simple point: because chemometrics analyses rely very heavily on computer programs and because analysis of mixtures is complicated in its own right, it becomes very important always to perform validation and verification of results. As many years of experience show in almost any aspect of complicated science, this is best done completely independently (clearly, we don't imply that the samples are independent!). The procedures for performing such independent validation and verification (IVV) will depend on many factors, but the crucial aspect is that both chemometrics "teams" be experienced. A wide variety of statistical tests can be applied to monitor IVV work.

To put it another way – chemometrics in the real world consists of three phases: exploratory data analysis (EDA), calibration and prediction (CP), and IVV.

#### 4.1.3 Chemometrics Camps

Several different groups of chemometrists around the world have developed their own sets of outlooks about their work. Some of that is evident in the following notes about the principal groups.

- The Swedish group at the University of Umeå seems to have important expertise and insights into tackling real-world problems, yet they are also very much aware of theory and maintain a practical and pragmatic interest. The principal researchers are S. Wold, P. Geladi, J Öhman, et al.
- The Norwegian group at Bergen seems smaller (than the Swedish group) but is quite applied. The Bergen group and the Umeå group have shared scientific responsibilities for the development of what may be the best commercially available chemometrics software – UNSCRAMBLER II, which is discussed in Chapter 2.
- The Belgian/Dutch group – the Belgian Massart being a world leader – represents breadth, with the Belgian group emphasizing multivariate analysis, statistical methodology, and signal processing while the Dutch are more interested in neural networks and expert systems.
- We are aware that the U.S. EPA has some significant capability in their contractor, Lockheed, but we have not found sufficient information to comment on their work.
- The group at the University of Washington under Kowalski is, of course, a major player globally. Essentially, Kowalski is one of the two or three founders of the field and is the principal author of what is the classic, original paper(s) in the field. However, their interest seems to be chemical process-oriented and is much more strongly focused in academic pursuits, especially in theory rather than applications. (Perhaps this is not surprising in view of the fact that Kowalski is also a principal in the Center for Process Analytical Chemistry, a university/industry cooperative research group.) The UW/CPAC capability is almost certainly the largest, strongest group in the U.S.
- The UK effort is notable (witness the work by the University of Warwick team to develop artificial olfactory systems), but we have not found enough of their work to make significant comparisons with the work of other groups. It may be of interest to note that relative to U.S. scientists, British scientists are much more likely to tackle problems analytically rather than use computer modeling.

#### **4.1.4 Chemometrics/Modern Computer-Based Applications to Other Aspects of ERWM**

It may be helpful to point out the potentially effective application of chemometrics and chemometrics-like analysis tools in site characterization, particularly in the so-called expedited site characterization (ESC) team concept developed by Argonne National Laboratory (3). As noted in Chapter 2, the experienced chemometrician is essential in judging the validity of any chemical analyses that might be performed using sensors and field laboratory instruments. But in addition, very similar techniques – if not already being used as part of the data analysis capability of the ESC team – could make valuable improvements in treating and interpreting data from geology, hydrology, stratigraphy, etc.

The following sources illustrate the application of chemometrics techniques in designing the sampling for chemical analyses:

a. R. A. Olivero, S. Seshadri, and S. N. Deming, "Development of an Expert System for Selection of Experimental Designs," *Anal. Chim. Acta* 277 441-453 (1993).

"An expert system has been developed to assist chemists in the selection of experimental designs for research projects. The system ranks thirteen types of experimental designs. . . . Design categories included are factorial, response surface, sequential simplex optimization, simplex mixture, and statistical testing. . . . uses mathematical concepts to mimic features of human intuition and decision making. . . . A validation test was conducted with the participation of four other experts in the field."

b. K. Hitchcock, J. H. Kalivas, and J. M. Sutter, "Computer-Generated Multicomponent Calibration Designs For Optimal Analysis Sample Predictions," *J. Chemom.* 6 85-96 (1992).

"This paper utilizes variable step size generalized simulated annealing (VSGSA) to design multicomponent calibration samples for spectroscopic data. VSGSA is an optimization procedure which is capable of converging to exact positions of global optima located on multidimensional continuous functions. On the basis of analysis of sample response vectors, optimally designed calibration concentration matrices are obtained assuming knowledge of components present. . . ."

c. S. N. Deming, J. A. Palasota, and J. M. Palasota, "Experimental Design in Chemometrics," *J. Chemom.* 5 181-192 (1991).

"Chemometrics is defined as the application of mathematical and statistical methods to chemical systems. Systems theory is seen to be useful for organizing and categorizing the inputs to and outputs from chemical systems. Advances in measurement science in the 1950s and 1960s, particularly in analytical chemistry, created a need for a multivariate approach to data analysis. Early chemometrics emphasized the use of structure-finding methods for existing data sets. In many instances, data sets can be obtained from designed experiments. Such data sets are more likely to contain the desired information and the data can usually be acquired at less cost. Renewed interest in statistical process control will provide many new, more robust data sets in the future."

d. P. Minkkinen, "SAMPEX - A Computer Program for Solving Sampling Problems," *Chemom. Intell. Lab. Sys.* 7 189-194 (1989).

"Chemical analysis always involves sampling. With heterogeneous solids the sampling error may be the largest source of error which determines the overall reliability of the analysis. To help analysts to evaluate the reliability of sampling and sample preparation procedures, a computer program, SAMPEX, was written. It is based on the theory developed by Pierre Gy which allows the estimation of the fundamental sampling error, i.e., the error due to the intrinsic heterogeneity of particulate material, from some basic material properties. . . . subroutines included . . . (1) estimation of the sampling error for a given sample size, (2) minimum sample size for a given sampling precision, (3) maximum particle size when both the sampling size and the sampling precision are given, (4) analysis of multistep sampling

and sample preparation procedures, (5) estimation of sampling constants from experimental data, and (6) sampling error in particle size distribution measurement."

#### 4.1.5 Chemometrics for DOE ERWM Problems

Application of chemometrics techniques to DOE ERWM problems can be summarized as follows:

- The multivariate analysis methods in chemometrics, together with all the appropriate statistics, instrumentation, engineering (including systems engineering), and computer technology, appear to be the only choice for which there is any real hope of solving the ERWM-type problem.
- Many analysis methods suited to environmental data analysis exist and await application. Application requires creation of certain infrastructure, but potentially many advances can be made by tackling the problem.
- The difficulty of judging the efficacy of analysis methods and the lack of "ground truth" in chemometrics is a strong argument for implementing a dedicated chemometrics analysis effort, tasked to evaluate the methods and validate the answers.
- A dedicated chemometrics effort may be able to determine which currently available analysis methodologies can provide the results needed for particular types of ERWM problems.

The syllogism suggests itself.

- If you want real answers to chemometrics analysis problems, you must pose real problems. If you want to evaluate the methods, you must try them. If you want the benefits of chemometrics, you must have a well-equipped and active chemometrics team.

## 4.2 RECOMMENDATIONS

The information collected and reviewed suggests the following:

- Establish a sensor development advisory panel tasked to recommend a program for development of sensor arrays for DOE's problems.
- Establish a sensor systems chemometrics advisory panel to recommend a chemometrics development program closely coordinated with the sensor array development program.
- Initiate a research program to extend mixture analysis research to consideration of real-world, highest priority ERWM problems.

### 4.3 REFERENCES

1. Y.-Z. Liang, O. M. Kvalheim, and R. Manne, "White, Grey, and Black Multicomponent Systems. A Classification of Mixture Problems and Methods for Their Quantitative Analysis," *Chemom. Intell. Lab. Syst.* 18 235-250 (1993).
2. J. W. Grate and M. H. Abraham, "Solubility Interactions and the Design of Chemically Selective Sorbent Coatings for Chemical Sensors and Arrays: Review Paper," *Sens. Actuators* B3 85-111 (1991).
3. J. C. Burton et al., preprint of Superfund XIV Conference Presentation, Argonne National Laboratory, Argonne, IL, December 1993.

## APPENDIX A

### CHEMOMETRICS BIBLIOGRAPHY

Several other references found that may be of interest were not reviewed in Chapter 2. These references are presented in descending chronological order.

1. S. N. Deming, J. A. Palasota, and J. M. Nocerino, "The Geometry of Multivariate Object Preprocessing," *J. Chemom.* 7 393-425 (1993).
2. W. J. Dunn III and D. Swain, "Computer-Assisted Interpretation of Mass Spectrometry-Mass Spectrometry Data of Potentially Hazardous Environmental Compounds," *Chemom. Intell. Lab. Syst.* 19 175-179 (1993).
3. P. K. Hopke and D. L. Massart, "Reference Data Sets for Chemometrical Methods Testing," *Chemom. Intell. Lab. Syst.* 19 35-41 (1993).
4. S. Li and P. Gemperline, "Eliminating Complex Eigenvectors and Eigenvalues in Multiway Analyses Using The Direct Trilinear Decomposition," *J. Chemom.* 7 77-88 (1993).
5. F. X. Rius, "Expert Systems in Trace Analysis," *Anal. Chim. Acta* 283 518-527 (1993).
6. S. Sekulic, M. B. Seasholtz, Z. Wang, and B. R. Kowalski, "Nonlinear Multivariate Calibration Methods in Analytical Chemistry," *Anal. Chem.* 65 835A-845A (1993).
7. M. Valcárcel and A. Rios, "The Hierarchy and Relationships of Analytical Properties," *Anal. Chem.* 65 781A-787A (1993).
8. B. Walczak and W. Wegscheider, "Non-Linear Modelling of Chemical Data by Combinations of Linear and Neural Net Methods," *Anal. Chim. Acta* 283 508-517 (1993).
9. Y. Wang, O. S. Borgen, B. R. Kowalski, M. Gu, and F. Turecek, "Advances in Second-Order Calibration," *J. Chemom.* 7 117-130 (1993).
10. M. Otto, T. George, C. Schierle, and W. Wegscheider, "Fuzzy Logic and Neural Networks - Applications to Analytical Chemistry," *Pure Appl. Chem.* 64 497-502 (1992).
11. A. Ruiz, D. Guinea, L. J. Barrios, P. Bustos, and F. Betancourt, "Data Structures for Multi-Sensor Integration," *Sens. Actuators* A32 491-498 (1992).
12. Y. Zeng and P. K. Hopke, "A New Receptor Model: A Direct Trilinear Decomposition Followed by a Matrix Reconstruction," *J. Chemom.* 6 65-83 (1992).
13. B. R. Kowalski and M. B. Seasholtz, "Recent Developments in Multivariate Calibration," *J. Chemom.* 5 129-145 (1991).

14. R. O'Brien, B. K. Sinha, and W. P. Smith, "A Statistical Procedure To Evaluate Clean-Up Standards," *J. Chemom.* 5 249–261 (1991).
15. M. A. Stapanian, "Some Issues in Chemometrics With Environmental Applications," *J. Chemom.* 5 121–128 (1991).
16. J. Zupan and J. Gasteiger, "Neural Networks: A New Method for Solving Chemical Problems or Just a Passing Phase?," *Anal. Chim. Acta* 248 1–30 (1991).
17. E. Sanchez and B.R. Kowalski, "Tensorial Resolution: A Direct Trilinear Decomposition," *J. Chemom.* 4 29–45 (1990).
18. W. G. Glen, W. J. Dunn III, and D. R. Scott, "UNIPALS, Software for Principal Components and Partial Least Squares Analysis," *Tetrahedron Computer Technology* 2 377–390 (1989).
19. A. Lorber and B. Kowalski, "Estimation of Prediction Error for Multivariate Calibrations," *J. Chemom.* 2 93–109 (1988).
20. M. Marsili, E. Marengo, and H. Saller, "SPECTRE: An Approach to an Expert System in Chemometrics," *Anal. Chim. Acta* 210 33–50 (1988).
21. E. Sanchez and B. R. Kowalski, "Tensorial Calibration: II. Second-Order Calibration," *J. Chemom.* 2 265–280 (1988).
22. A. Lorber, L. E. Wangen, and B. R. Kowalski, "A Theoretical Foundation For The PLS Algorithm," *J. Chemom.* 1 19–31 (1987).
23. W. Lindberg, J. Öhman, and S. Wold, "Multivariate Resolution of Overlapped Peaks in Liquid Chromatography Using Diode Array Detection," *Anal. Chem.* 58 299–303 (1986).
24. S. Wold, M. Sjöström, R. Carlson, T. Lundstedt, S. Hellberg, B. Skagerberg, C. Wikström, and J. Öhman, "Multivariate Design," *Anal. Chim. Acta* 191 17–32 (1986).
25. T. Brown and S. Brown, "Resolution of Overlapped Electrochemical Peaks with the Use of the Kalman Filter," *Anal. Chem.* 53 1410–1417 (1981).
26. H. Martens, "On the Calibration of a Multivariate Instrument for Quantitative Estimation of Individual Components in a Mixture," in *Symposium on Applied Statistics, Copenhagen, 23–25 January 1980*, ed. A. Höskuldsson, K. Conradsen, B. Sloth-Jensen, and K. Esbensen, Danish Universities Computing Center, Copenhagen, 1980, pp 393–414.
27. H. Martens, "Factor Analysis of Chemical Mixtures," *Anal. Chim. Acta* 112 423–442 (1979).

28. D. Massart, A. Dijkstra, and L. Kaufman, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures: A Survey of Statistical and Mathematical Techniques*, Elsevier, Amsterdam, 1978.
29. S. Deming, "Optimization of Experimental Parameters in Chemical Analysis," in *Validation of the Measurement Process*, ed. J. Devoe, American Chemical Society Symposium Series 63, American Chemical Society, Washington, DC, 1977.
30. B. Kowalski, Ed., *Chemometrics: Theory and Applications*, American Chemical Society Symposium Series 52, American Chemical Society, Washington, DC, 1977.
31. S. Wold, "Pattern Recognition by Means of Disjoint Principal Components Models," *Pattern Recognition* 8 127–139 (1976).
32. B. Kowalski and C. Bender, "Pattern Recognition II: Linear and Nonlinear Methods for Displaying Chemical Data," *J. Am. Chem. Soc.* 95 686–693 (1973).
33. B. Kowalski and C. Bender, "Pattern Recognition: A Powerful Approach to Interpreting Chemical Data," *J. Am. Chem. Soc.* 94 5632–5639 (1972).
34. B. Kowalski, P. Jurs, T. Isenhour, and C. Reilly, "Computerized Learning Machines Applied to Chemical Problems: Interpretation of Infrared Spectrometry Data," *Anal. Chem.* 41 1945–1949 (1969).
35. B. Kowalski, P. Jurs, T. Isenhour, and C. Reilly, "Computerized Learning Machines Applied to Chemical Problems; Multicategory Pattern Classification by Least Squares," *Anal. Chem.* 41 695–700 (1969).
36. P. Jurs, B. Kowalski, and T. Isenhour, "Computerized Learning Machines Applied to Chemical Problems; Molecular Formula Determination from Low Resolution Mass Spectrometry," *Anal. Chem.* 41 21–27 (1969).
37. I. Blifford and G. Meeker, "A Factor Analysis Model of Large Scale Pollution," *Atmos. Environ.* 1 147–157 (1967).

## APPENDIX B

### SCIENTIFIC AND TECHNICAL INFORMATION

### NETWORK BIBLIOGRAPHY

The following listings were obtained from STN database searches.

1. A. Bos, M. Bos, and W. E. Van der Linden, "Artificial Neural Networks as a Multivariate Calibration Tool: Modeling the Iron-Chromium-Nickel System in X-Ray Fluorescence Spectroscopy," *Anal. Chim. Acta* 277 289-295 (1993).
2. M. Bos, A. Bos, and W. E. Van der Linden, "Data Processing by Neural Networks in Quantitative Chemical Analysis," *Analyst* 118 323-328 (1993).
3. F. A. M. Davide, C. Di Natale, and A. D'Amico, "Resolution of Sensor Array: Noise Considerations and Design Implications," *Sens. Actuators* A37-38 296-300 (1993).
4. F. A. M. Davide, C. Di Natale, and A. D'Amico, "Sensor Array Figures of Merit: Definitions and Properties," *Sens. Actuators* B13-14 327-332 (1993).
5. C. Di Natale, A. D'Amico, and F. A. M. Davide, "Redundancy in Sensor Arrays," *Sens. Actuators* A37-38 612-617 (1993).
6. J. A. Ramola, M. S. Larrechi, and F. X. Rius, "Chemometric Characterization of 5th Century A.D. Amphora-Producing Centers in the Mediterranean" (Ed. note: using X-ray fluorescence data), *Talanta* 40 1749-1757 (1993).
7. J. M. Slater, J. Paynter, and E. J. Watt, "Multi-Layer Conducting Polymer Gas Sensor Arrays for Olfactory Sensing," *Analyst* 118 379-384 (1993).
8. G. Rauret, M. T. Galceran, R. Rubio, F. X. Rius, and M. S. Larrechi, "Factor Analysis for Assigning Sources of Groundwater Pollution," *Int. J. Environ. Anal. Chem.* 38 389-397 (1990).
9. F. X. Rius, M. S. Larrechi, C. Benet, E. Subias, D. L. Massart, and A. Thielemans, "The Application of Multivariate Techniques to Data from Spanish Glass-Making Objects from the Roman Era" (Ed. note: using X-ray fluorescence data), *Anal. Chim. Acta* 225 69-81 (1989).
10. W. E. Van der Linden, M. Bos, and A. Bos, "Arrays of Electrodes for Multicomponent Analysis," *Anal. Proc. (London)* 26 329-331 (1989).
11. G. Rauret, E. Casassas, M. Baucells, M. S. Larrechi, and F. X. Rius, "Assigning the Origin of Catalan Medieval Stained Glasses by Pattern Recognition Methods of X-Ray Fluorescence Data," *J. Chemom.* 3 (Suppl.A) 163-174 (1988).
12. G. Rauret, R. Rubio, F. X. Rius, and M. S. Larrechi, "Cluster Analysis as a Tool in the Study of Groundwater Quality," *Int. J. Environ. Anal. Chem.* 32 255-268 (1988).