AD-A100 433 WASHINGTON UNIV SEATTLE LAB FOR CHEMOMETRICS F/G 7/2
ANALYTICAL CHEMISTRY AS AN INFORMATION SCIENCE.(U)
JUN 81 B R KOWALSKI
UNCLASSIFIED 20 N00014-75-C-0538 NL

END
DATE FILMED
7 81
DTIC

AD A10 0433

OFFICE OF NAVAL RESEARCH

Contract N00014-75-C-0536

Task No. NR 051-565

TECHNICAL REPORT NO. 20

ANALYTICAL CHEMISTRY AS AN INFORMATION SCIENCE

by

B. R. Kowalski

Prepared for Publication

in

Trends in Analytical Chemistry

University of Washington
Department of Chemistry
Seattle, Washington 98195

DTIC
JUN 2 2 1981
A

June 1981

Reproduction in whole or in part is permitted for any purpose of the United States Government

This document has been approved for public release and sale; its distribution is unlimited

DTIC FILE COPY

81 6 17 021

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 14 29	2. GOVT ACCESSION NO.: AD-A100 433
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): 6 ANALYTICAL CHEMISTRY AS AN INFORMATION SCIENCE,	5. TYPE OF REPORT & PERIOD COVERED: Technical Report - Interim 2/1981 - 6/1981
	6. PERFORMING ORG. REPORT NUMBER: --
7. AUTHOR(s): 10 Bruce R. Kowalski 15	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0536
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Laboratory for Chemometrics, Dept. of Chemistry, University of Washington, Seattle, WA 98195	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 051-565
11. CONTROLLING OFFICE NAME AND ADDRESS: Materials Sciences Division, Office of Naval Research, Arlington, Virginia 22217 11	12. REPORT DATE: June 1981 12
	13. NUMBER OF PAGES: 13
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 9 Interim rept. Feb-Jun 81	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

--

18. SUPPLEMENTARY NOTES

Prepared for publication in Trends in Analytical Chemistry

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Chemometrics
Analytical Chemistry
Intelligent Instrumentation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper resulted from a presentation with the same title at the Symposium on New Directions in Analytical Chemistry held at the Pittsburg Conference in Analytical Chemistry and Spectroscopy, Atlanta City, March, 1981. This paper attempts a look into the future of Analytical Chemistry discussing such possibilities as intelligent analytical instrumentation and the generation and analysis of chemical pictures.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED 408726
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ANALYTICAL CHEMISTRY AS AN INFORMATION SCIENCE

The decade of the 1980's has the potential of being the most exciting period that analytical chemistry has yet seen. Or it can amount to ten years of "business as usual". The choice is strongly dependent on whether or not analytical chemists become aware of, and participate in, a number of major changes currently taking place in science and society. At the heart of these changes is a generation of computer technology that can put the world's supply of data at our fingertips with the added potential of converting it to useful information and perhaps even knowledge. This article is based on the author's contribution to the Symposium on New Directions in Analytical Chemistry at the 1981 Pittsburg Conference entitled "New Directions in Information Science".

It is indeed difficult to ignore the effect that computers now have on our lives. It will be almost impossible to do so in the future. A consensus report prepared for the U.S. National Science Foundation (1) states that the "U.S. is rapidly transforming from an economy based on industrial production to one based on the transfer of information. Computers are now used in all aspects of daily life to improve the fuel economy of cars and homes and to provide more efficient services. These changes, along with the merging of communications and computers, are causing the emergence of an information-based economy -- the transition from an industrial to an information society". The report goes on to say that "Information is clearly the dominant national commodity, with approximately one-half the labor force holding information-

related jobs, and earning over one-half the labor income".

Analytical chemistry is now, and has always been, an information science. In fact, the type of information provided by analytical chemistry may be the most reliable, informative and desperately needed information that any science can offer to society. Herein lies the first of two paths of opportunity for analytical chemistry opened by the computer revolution.

If analytical chemists concern themselves solely with their abilities to fill notebooks and disk files with quantitative chemical measurements, then the next ten years will see business as usual for the analytical chemist. The sample input rates seen by analytical laboratories will be exceeded only by the data generation rates that can be achieved with modern analytical instruments. On the other hand, some analytical chemists are becoming keenly aware of powerful data analysis methods that can guarantee efficient conversion of raw data to useful information and knowledge. For example, the environmental analytical chemist uses calibration mathematics to convert, say, electroanalytical measured currents and voltages (data) to concentrations of various chemical species (information) in water samples collected from a watershed. Now, there is little doubt that the chemistry of a natural watershed is complex, involving variations in a multitude of chemical components. If the chemist has a knowledge of the power of multivariate statistics, then by using these tools to analyze the concentrations of several species measured on

several samples, the concentration information can be combined to provide a knowledge of the complex water chemistry of the watershed.

A number of tutorials have been offered to analytical chemists in recent years expounding the virtues of statistical experimental design (2), factor analysis (3), pattern recognition (4) and several other tools from statistics and applied mathematics. The transition from the analytical chemist as simply a data generator to the analytical chemist as an effective problem solver has been aided by the development of chemometrics (5,6,7). As defined by the Chemometrics Society (8),

> "Chemometrics is the chemical discipline that uses mathematical and statistical methods
> a) to design or select optimal measurement procedures and experiments, and
> b) to provide maximum chemical information by analyzing chemical data.
> In the field of analytical chemistry, chemometrics is the chemical discipline which uses mathematical and statistical methods to achieve the aim of analytical chemistry namely the obtention in the optimal way of relevant information about material systems".

The use of modern applied mathematics by analytical chemometricians to extract useful chemical information from information rich analytical measurements is certain to elevate the status of

analytical chemistry in science and society by efficiently providing solutions to complex problems. In this way, full advantage of the computer will be taken both for its chemical information storage and retrival capabilities and its computational abilities for the combination of information for the purpose of acquiring knowledge.

There is yet another avenue opened by computers that is potentially even more rewarding to the analytical chemist. By taking full advantage of multivariate mathematics, and the computational and logical decision making abilities of computers, it will be possible to completely alter standard analytical procedures and methods to provide vastly improved analytical measurements. However, this can only come about when analytical chemists choose to alter the historical trend of analytical method development.

In the past the development of a new analytical method usually involved the exploitation of a newly discovered chemical or physical phenomenon. More recently, the combination of analytical methods (e.g. LC/MS) to form hyphenated methods (9) has been responsible for some very powerful tools. Computers have certainly become invaluable components of modern analytical instrumentation for data acquisition, storage, display and processing. However, as the remainder of the article will attempt to show, the proper combination of computer hardware, software and chemometrics can yield analytical measurements normally thought to be impossible to acquire. Here is the

second avenue opened by computer technology that will allow analytical chemists to alter the future.

In 1974, Rogers and Coworkers (10) showed that the number of components eluting from a gas chromatograph as an unresolved peak could be determined by applying principal component analysis to mass spectra sampled as a function of time. This approach has been extended to the recovery of the mass spectra of the eluting components (11). This curve resolution problem could only be solved by a mathematical approach that had the capability of detecting covariation patterns in a table (matrix) of mass spectral intensities; a multivariate statistical method. The combining of two instruments, GC/MS, coupled the mixture resolution power of the chromatograph with the identification power of the mass spectrometer. It also made possible the generation of two-dimensional spectra (times vs. mass/change) thereby allowing the use of a multivariate data analysis tool. The burden of complete chromatographic separation, is no longer necessary as unresolved peaks can now be further resolved by the computer. This may be seen as just one example of a new philosophy of analytical method development. The philosophy includes the exploitation of all of the tools available to the analytical chemists, including those from chemometrics, in order to achieve a proper balance between chemical or physical resolution and mathematical resolution.

Activity in hyphenated instrumentation development will continue to yield fertile research ground for the analytical chemometrician. If the current development philosophy persists,

new instrument combinations will generate new data analysis problems to be solved. However, if teams, comprise of analytical chemists, analytical chemometricians and supporting engineers, mathematicians and statisticians, are formed to develop new methods, such topics as statistical error propagation, optimal control and error amplification can guide the development of balanced analytical measurement systems.

By approaching analytical method development as an information science problem, the theoretical limitations can guide the analyst to an optimal product. As a simple example, consider the selection of wavelengths for a multi-analyte atomic emission spectrometer. Since interfering wavelengths should be avoided, compromise wavelengths, free from spectral overlap and other interference problems, are usually selected. The price for fewer interferences is a lower sensitivity. A useful balance between the extremes of optimal sensitivity with severe interferences and lower sensitivity with little or no interferences is usually sought. Now, the generalized standard addition method (GSAM) (12,13) can be used to eliminate interference effects and matrix effects thereby allowing the selection of more sensitive wavelengths. The price paid for analysis in the presence of interferences is a potential error amplification. (The propagation of measurement error to the estimated concentrations may be accompanied by a magnification of the error.) Fortunately, the theory behind the GSAM provides a means of minimizing error amplification. Using this theory, an analytical chemist can select wavelengths for analysis that

provide the optimum balance between sensitivity, precision and accuracy instead of simply sacrificing sensitivity to avoid interferences.

The theory of multi-analyte resolution and calibration has enormous potential for analytical method development in the future. The proper experimental design makes possible a complete characterization the "analysis power" of an analytical method (14). More over, the computer can even monitor the general health of an analytical instrument; detecting and correcting for such problems as measurement drift or temperature instability (14). It is this combination of measurement theory, information science and the computer that is certain to lead to some very exciting analytical chemistry; intelligent analytical instrumentation. Figure 1 is a dialogue (slightly tongue in cheek) between a chemist and an intelligent analytical/computer network of the future.

Now, more than ever, analytical chemists are curious about the reach and limitations of the methods they employ. This new interest has the potential of giving birth to something that has been needed for decades; a theory of chemical analysis. Since we use, or should use, mathematics for calibration and resolution and statistics for expressing the uncertainty in our measurements, a theory will evolve from mathematical limitations and statistical constructs. For example, the condition number of a matrix of linear response constants can guide the analyst in the development of an optimal method for multi-component analysis (13). There is little doubt in the author's mind that a useful and extensive

theory of chemical analysis will evolve in the years to come. Every good science deserves some theory.

Finally, a few words about technology transfer are in order. Each type of analytical instrumentation had its beginning as a jumble of aluminum foil, mismatched plumbing and a potentially lethal mass of uninsulated wires. As development continued, the analysis power of the instrument increased and the world became more interested in the new tool. However, it can be argued that the real potential of any tool can only be realized when it can be used by many. Historically, commercial instrument manufacturers have been responsible for the transfer of new technology from the research laboratory to the application laboratory. In recent years, the long lag between the demonstration of feasibility and commercial development has been shortened, perhaps due to the increased importance of analytical measurements in society. How then will the new developments of chemometrics be transferred from the research to the applications environment? To the author's knowledge, there is at least one commercial organization (15) committed to this important interface at this time. The company is actively following new developments in analytical chemometrics and, when feasibility has been clearly demonstrated, computer programs are either written for general use or written for a specific instrument and distributed O.E.M. in much the same way that computer hardware is distributed as a part of an analytical instrument. Various estimates predict that software costs will exceed hardware costs in the future. The days of cheap or free

software are coming to a rightful end.

For the past decade, the author and others have called for more application of applied mathematics and statistics in analytical chemistry. There has been some resistance. However, as society itself is hurled into the age of information, analytical chemists, a rather pragmatic group, will gradually learn more about computers, statistics and applied mathematics (perhaps from our children) ensuring a key role for analytical chemistry as an information science in an information society.

Figure 1

CHEMIST Connect me with the Separations Lab Computer.

COMPUTER I know what it knows; it is part of the network.

CHEMIST Were my 7422 river samples analyzed last night?

COMPUTER NATURLICH, (I speak 12 languages you know!) would you like to see complete organic analysis results? You'll need two hours to read the listing.

CHEMIST No, did any sample contain components from the EPA list?

COMPUTER Yes, the sediment cores from sector SS-1 had PCB's above 1 ppm. Those results are on your printer. Have you checked?

CHEMIST It's early.

COMPUTER Before we continue, you should know that I've shortened GC runs by an extra 32.4% resulting in a greater sample throughput at a lower analysis cost. Your new GC/MS resolution algorithm can completely resolve eluent peaks so there is no need to separate further. We should publish this!

CHEMIST We?

COMPUTER Next command, please.

CHEMIST I'll need your help with several new water samples for trace metals analysis. Matrix effects and interferences suspected. Run GSAM.

COMPUTER How many analytes?

CHEMIST 19

COMPUTER Enter channel sensitivity estimates and names of analytes.

CHEMIST Sensitivities unknown, names on file X46.

COMPUTER Enter interference estimates.

CHEMIST Unknown.

COMPUTER Your samples will be run automatically, all interferences will be characterized, results in one hour. Incidentally, the photo multiplier on Channel 14 is drifting badly. I'll correct the analyte concentrations this time but it should be repaired soon. Anything else?

CHEMIST -Tell my home computer that I'll be home by 7:00 p.m..

-Search the commercial computer network for the best price for 4 snow tires for my truck. Verify an order at the best price, transfer money from my bank for payment and schedule an appointment for installation.

-Make complete reservations with my travel agent's computer network for a ski trip to Crystal Mountain.

That's all for now.

COMPUTER Lucky human!

References

1. Irving Shain, A consensus Statement: The Panel on Computing and Higher Education, March 30, 1981, National Science Foundation.

2. R.F. Hirsch (1977), Analytical Chemistry, 49,691A.

3. E.R. Malinowski and D.G. Howery, Factor Analysis in Chemistry, (1980), Wiley-Interscience.

4. K. Varmuza, Pattern Recognition in Chemistry (1980), Springer-Verlag, New York.

5. B.R. Kowalski, Chemometrics: Views and Propositions (1975), J. Chem. Infor. & Comp. Sci., 15,201.

6. B.R. Kowalski, Chemometrics (1978), Chemistry and Industry, 22,882.

7. B.R. Kowalski, Fundamental Review on Chemometrics (1980), Analytical Chemistry 52,112R.

8. Chemometrics Society: Prof. D.L. Massart
   Secretary of Analytical Section
   Chemometrics Society
   Vrije Universiteit Brussel
   Farmaceutisch Instituut
   Laarbeeklaan 103
   B-1090 Brussel
   Belgium

9. T. Hirschfeld (1980), Analytical Chemistry, 52,299A.

10. J.E. Davis, A. Shepard, N. Stanford and L.B. Rogers (1974) Analytical Chemistry 46,821.

11. M.A. Sharaf and B.R. Kowalski, Extraction of Individual Mass Spectra from GC-MS Data of Unseparated Mixtures (1981), 53, 518-522.

12. B.E.H. Saxberg and B.R. Kowalski, The Generalized Standard Addition Method (1979) Analytical Chemistry, 51,1031.

13. C. Jochum, P. Jochum and B.R. Kowalski, Error Propogation and Optimal Performance in Multicomponent Analysis (1981), 53, 85-92.

14. J.H. Kalivas and B.R. Kowalski, paper in review.

15. Informetrix, Inc. P.O. Box 25808, Seattle, WA. 98125.

DAT
ILM