DTIC FILE COPY

(4)

AD-A197 534

ate Entered)

N PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: TR40	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Multi-Order Calibration	5. TYPE OF REPORT & PERIOD COVERED: Technical Report - Interim
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Eugenio Sanchez and Bruce R. Kowalski	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0536
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Laboratory for Chemometrics, Department of Chemistry BG-10, University of Washington	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	12. REPORT DATE: July 1, 1988 13. NUMBER OF PAGES: 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) 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

Submitted and accepted for publication in IUPAC special report on chemometrics.

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

Bilinear	Multidimensional Arrays	PCR	PARAFAC
Calibration	Multiorder	GRAM	
Dyadic	Multilinear	Rank Annihilation	
Multivariate	PLS	Tensor	

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

Please see reverse side.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE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

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Instruments that generate two-dimensional arrays of data are now commonplace in the analytical laboratory. Time decay and emission-excitation fluorescence, chromatography-spectroscopy combinations, MS-MS and 2D-NMR, are a few of the many so-called "hyphenated methods" that generate such data. These instruments have become very important for the analyst mainly because of their higher selectivity and resolution of signals, allowing for analysis of mixtures. The main similarity between all these instruments is that each sample analyzed produces a two-dimensional array of data (second order tensor). The amount of information produced by such an instrument is overwhelming; for quantitative analysis usually only a small portion of the data is actually used and the rest discarded.

The situation is even worse if several samples have to be compared, because the accumulated data would be a three-dimensional array (third order tensor). It is obvious that with the standard statistical tools (e.g., univariate linear regression) the chemist is seriously under-prepared to analyze these data. Even multivariate statistical techniques are hard pressed to analyze higher order data, and in the best case, they cannot fully extract all the information available.

This paper will summarize a multi-order, tensorial approach to calibration, that takes advantage of all the information from instruments that produce data arrays of any order, for prediction of unknown properties such as analyte concentrations. The types of instruments are classified here according to the kind of array of data produced per sample: some instruments generate a single number (signal) per analysis or sample. Others generate two or more signals (first order instruments), i.e., a vector of signals. Yet, other instruments can generate a matrix of data per sample, or a 3-dimensional array (second and third-order instruments). This distinction will be called the "order of the instrument", in analogy with the order of a tensor.

The "order of the instrument" has special importance beyond the simple fact of the form of the data. There are possibilities of analysis with some higher order instruments which are not available for lower order instruments -the multi-order advantage. The simplest example is provided by the difference between a single sensor instrument (zero order) and an array of sensors (first order): first order permits quantitation of multicomponent mixtures and detection of outlier unknowns, which are not possible with zero order data. LC/DA-UV data will be used to illustrate how some second order Instruments permit multicomponent analysis with a single calibration sample.

OFFICE OF NAVAL RESEARCH

Contract N00014-75-C-0536

Task No. NR 051-565

TECHNICAL REPORT NO. 40

Multi-Order Calibration

by

Eugenio Sanchez and Bruce R. Kowalski

Prepared for Publication
in
IUPAC Special Report on Chemometrics

Department of Chemistry, BG-10
University of Washington
Seattle, Washington 98195

July 1, 1988

MULTI-ORDER CALIBRATION

Eugenio Sanchez and Bruce R. Kowalski
Laboratory for Chemometrics, Department of Chemistry BG-10
University of Washington
Seattle, Washington 98195, U.S.A.

Instruments that generate two-dimensional arrays of data are now commonplace in the analytical laboratory. Time decay and emission-excitation fluorescence, chromatography-spectroscopy combinations, MS-MS and 2D-NMR, are a few of the many so-called "hyphenated methods" that generate such data.[1] These instruments have become very important for the analyst mainly because of their higher selectivity and resolution of signals, allowing for analysis of mixtures. The main similarity between all these instruments is that each sample analyzed produces a two-dimensional array of data (second order tensor). The amount of information produced by a second order instrument is overwhelming; for quantitative analysis usually only a small portion of the data is actually used and the rest discarded.

The situation is even worse if several samples have to be compared, because the accumulated data can be represented by a three-dimensional array (third order tensor). It is obvious that with the standard statistical tools (e.g., univariate linear regression) the chemist is seriously under-prepared to analyze these data. Even multivariate statistical techniques are hard pressed to analyze higher order data, and even in the best case, they cannot fully extract all the information available.

This paper summarizes a multi-order, tensorial approach to calibration, that takes full advantage of all the information from instruments that produce data arrays of any order, for prediction of unknown properties such as analyte concentrations. The types of instruments are classified here according to the kind of array of data produced per sample: some instruments generate a single number (signal) per analysis or sample. Others generate two or more signals (first order instruments), i.e., a vector of signals. Yet, other instruments generate a matrix of data per sample, or a 3-dimensional array (second

and third-order instruments). This distinction will be called the *order of the instrument*, in analogy with the order of a tensor.

The order of the instrument has special importance beyond the simple fact of the form of the data. There are possibilities of analysis with some higher order instruments which are not available for lower order instruments. The simplest example is provided by the difference between a single sensor instrument (zero order) and an array of sensors (first order): first order permits quantitation of multicomponent mixtures and detection of outlier unknowns, which are not possible with zero order data.

Zero Order Instruments: Univariate Calibration.

Zero order calibration is by far the most common kind of instruments in analytical chemistry. Simple sensors or detectors are included in this category. Fig. 1a shows a typical linear calibration experiment. Several samples of known concentration are used to build a model, and then the model is used to predict the concentration, c, of an unknown from its response r. Fig. 1b illustrates the same model applied to a sample that has an interferent. Clearly, not only the estimation of the actual concentration of the analyte is impossible, but it is also impossible to detect the interferent. The importance of this simple fact cannot be over-emphasized; much of an analytical chemist's time is spent ensuring that the sample is pure, without interferences, or finding a measurement technique that only responds to the analyte of interest (fully selective).

First Order Instruments: Multivariate Calibration.

A multichannel spectrometer or an array of sensors constitutes a first order instrument. The numerical values of the responses of a sensor array with p sensors can be arranged as the components, r_i, of a vector $\mathbf{r}$. Therefore, this sensor array response "spectrum" can be considered a vector in a multidimensional vectorial space, where the base vectors of the

space are the unitary responses for each sensor, and the components are the actual numbers that the sensor array has provided.

The problem of linear multivariate calibration consists on finding a linear combination of the instrument responses optimal for prediction of the analyte concentration in the sample. To estimate this optimal weighting of the responses, the p responses, r_i, from a set of samples of known concentration are recorded and used for the prediction of analyte concentrations, $\hat{c}_u$, for future, unknown samples,

$$\hat{c}_u = \sum_{i=1}^{p} r_i \, x_i^* \tag{1}$$

It can shown that the optimal set of weights for prediction are a vector $\mathbf{x}^*$ which must be perpendicular to the spectra of all other analytes and interferences present in the unknown sample. In other words, it is the *contravariant* component of the analyte spectrum,[2] or *net analyte signal* vector.[3] A general solution to the problem of finding $\mathbf{x}^*$ given a calibration set of response vector - concentration pairs $\{\mathbf{r}_i, c_i\}$, ordered into a matrix $\mathbf{R}$ and a vector $\mathbf{c}$, is

$$\hat{\mathbf{x}}^* = \hat{\mathbf{R}}^\dagger \, \mathbf{c} \tag{2}$$

where $\hat{\mathbf{R}}^\dagger$ is an estimated pseudoinverse of $\mathbf{R}$. Many different methods have been developed to estimate the optimal pseudoinverse for prediction, among them principal components regression[4] (PCR) and partial least squares[5] (PLS) calibrations have been extensively studied in the recent chemometrics literature.

Fig. 2 illustrates an example of multivariate calibration. Two kinds of interferences are possible with a first order instrument:

(1) Fig. 2a. Interferences present both in the calibration set and in the unknown sample: component B was present in the calibration samples, even though its concentration was unknown. Multivariate calibration can still accurately predict the concentration of component A in the presence of B.

(2) Fig. 2b. Interferences or background constituents present only in the unknown sample: component C was not present in the calibration samples, therefore, no good estimate of the concentration of component A is possible. But it is possible to detect this sample as an outlier to our model, because C is not in the space spanned by A and B.

Second Order Instruments: *Second Order Calibration.*

Instruments that generate two-dimensional arrays of data (second order instruments) are now commonplace in the analytical laboratory. In chemistry, the normal way to handle this kind of data has been to choose from the array a single element which is unique for the analyte of interest, discarding or perhaps not collecting the rest of the data. For example, in MS-MS, is often possible to find daughter ions which are completely unique for one analyte of a mixture. For an emission-excitation matrix (EEM), it is sometimes possible to find a combination of excitation and emission wavelengths for which only the analyte of interest has a significant signal. This is a valid approach when the analyst knows that the signal being used is unique, just like in univariate calibration, however it does not take advantage of all the information available.

It is also possible to unfold the data into a vector as it has been suggested by Wold and coworkers,[6] and then use multivariate techniques such as PLS for calibration and data analysis. Unfortunately, when breaking a second order array of data into a first order array, e.g., by separating the columns of the matrix into a long column vector, the relationship of the rows is lost in the process. For some kinds of data, this may not cause problems because that relation may be unimportant, but for bilinear data arrays, such as LC/UV or EEM, unfolding produces a drastic lost of information.

Assuming a linear model, the response of a second order instrument to a multicomponent sample **M** should be approximately equal to a linear combination of the responses of all the individual analytes present in the sample, N_i, plus error, **E**:

$$\mathbf{M} = \sum_{i=1}^{q} c_i \mathbf{N}_i + \mathbf{E} \qquad (3)$$

where the matrices $\mathbf{N}_i$ have been scaled such that the coefficients c_i are the corresponding concentrations. If the matrices $\mathbf{N}_i$ and $\mathbf{M}$ have rank approximately equal to *1* and *q* respectively (Bilinear data), and the data matrix $\mathbf{N}_k$ for a particular analyte is known, it is possible to show that c_k can be estimated in equation 3 for an unknown $\mathbf{M}$, by using[7,8]

$$1/c_k = \sum_{i=1}^{I} \sum_{j=1}^{J} N_{ij} \, (M^{\dagger})_{ij} \qquad (4)$$

where $(M^{\dagger})_{ij}$ is the element of a pseudoinverse of $\mathbf{M}$, corresponding to the i^{th} row and the j^{th} column of $\mathbf{M}^{\dagger}$, and the subscript k has been dropped from N_{ij} for simplicity. Equation 4 implies that a single calibration sample with the analyte of interest is sufficient for estimating the concentration of the analyte in an unknown sample.[9]

If a multicomponent calibration sample is used, with some analytes at known concentration, e.g., $\mathbf{N}$, it is possible to determine the concentration of all those analytes in an unknown sample, together with their individual spectra. Fig. 3 illustrates this with an example: The bilinear LC/UV data from (1) a two-component calibration sample, and (2) a test ("unknown") sample with 3 constituents, are collected. Applying second order bilinear calibration methods,[10] (i) the UV spectra, (ii) the chromatographic concentration profiles, and (iii) the ratios of concentration (calibration/unknown) are obtained. This is possible using the non-symmetrical eigenvalue-eigenvector equation[10]

$$(\mathbf{N}\,\mathbf{M}^{\dagger})\,\mathbf{X} = \mathbf{X}\,\lambda \qquad (5)$$

where $\mathbf{M}^{\dagger}$ is the pseudoinverse of the unknown sample data matrix, $\mathbf{X}$ are the eigenvectors (pure spectra in one of the orders) and λ is a diagonal matrix of eigenvalues, which are the ratio of concentrations, between the calibration sample and the unknown sample, of the analytes in common. More complete details are given elsewhere.[9,10,11]

It is also possible to use several calibration samples instead of one, and in many cases desirable, i.e., for covering a wider dynamic range, reduce the effect of collinearities or increase the precision of the predicted concentrations.[8]

Conclusion.

Many aspects of calibration have been omitted from this discussion to focus in the order of the instrument. Among them, non-linear responses, outlier detection, precision and accuracy, sampling, sample selection, variable selection, experimental design, and time dependence of the responses. These factors are well studied only for the case of zero order calibration. The multivariate regression literature provides a good starting point for first order calibration, but few papers have been published that address these issues for calibration[12]. For second and higher order calibration, no work has been done, with very few exceptions.[13] These issues represent an important challenge for the chemometrician, and we expect important developments in these areas in the near future.

Acknowledgement

The authors thank Bruce Wilson for proof-reading this paper. This work was supported in part by the Office of Naval Research.

Bibliography

1 Hirschfeld, T. (1980) "The Hyphenated Methods". Anal. Chem. 52, 297A-312A.

2 Sanchez, E. and Kowalski, B. R. (1987) "Tensorial Calibration II. Second Order Calibration". *J. Chemometrics* (Submitted for publication).

[3] Lorber, A. (1986) "Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations". *Anal. Chem.* **58**, 1167-1172.

[4] Beebe, K. R. and Kowalski, B. R. (1987) "An Introduction to Multivariate Calibration and Analysis". *Anal. Chem.* **59**, 1007A-1017A.

[5] Martens, H.; Karstang, T. and Næs, T. (1987) "Improved Selectivity in Spectroscopy by Multivariate Calibration"

[6] Wold, S.; Geladi, P.; Esbensen, K. and Öhman, J. (1987) "Multi-way Principal Components and PLS-Analysis". *J. Chemometrics* **1**, 41-56.

[7] Lorber, A. (1985) "Features of quantifying chemical composition from two-dimensional data array by the rank annihilation factor analysis method". *Anal. Chem.* **57**, 2395-2397.

[8] Sanchez, E. and Kowalski, B. R. (1987) "Tensorial Calibration II. Second Order Calibration". *J. Chemometrics* (Submitted for publication).

[9] Ho, C-N.; Christian, G. D. and Davidson, E. R. (1981) "Simultaneous multicomponent rank annihilation and applications to multicomponent fluorescent data acquired by the video fluorometer". *Anal. Chem.* **53**, 92-98.

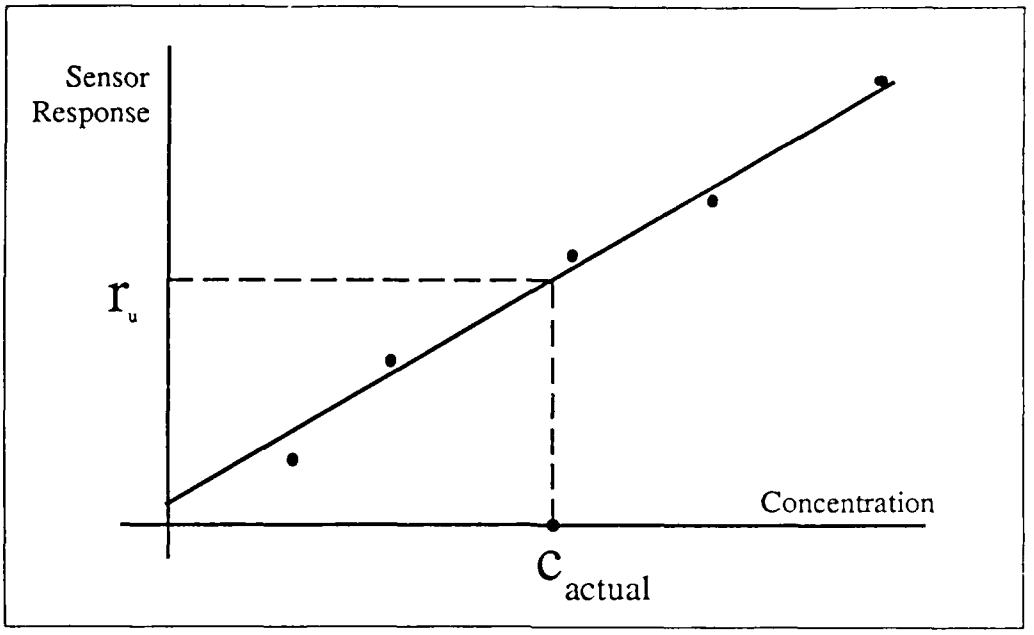
[10] Sanchez, E. and Kowalski, B. R. (1986) "Generalized Rank Annihilation Factor Analysis". *Anal. Chem.* **58**, 496-499.

[11] Sanchez, E.; Ramos, L. S. and Kowalski, B. R. (1987) "Generalized Rank Annihilation Method. I. Application to Liquid Chromatography / Diode Array UV Data". *J. Chromatogr.* **385**, 151-164.

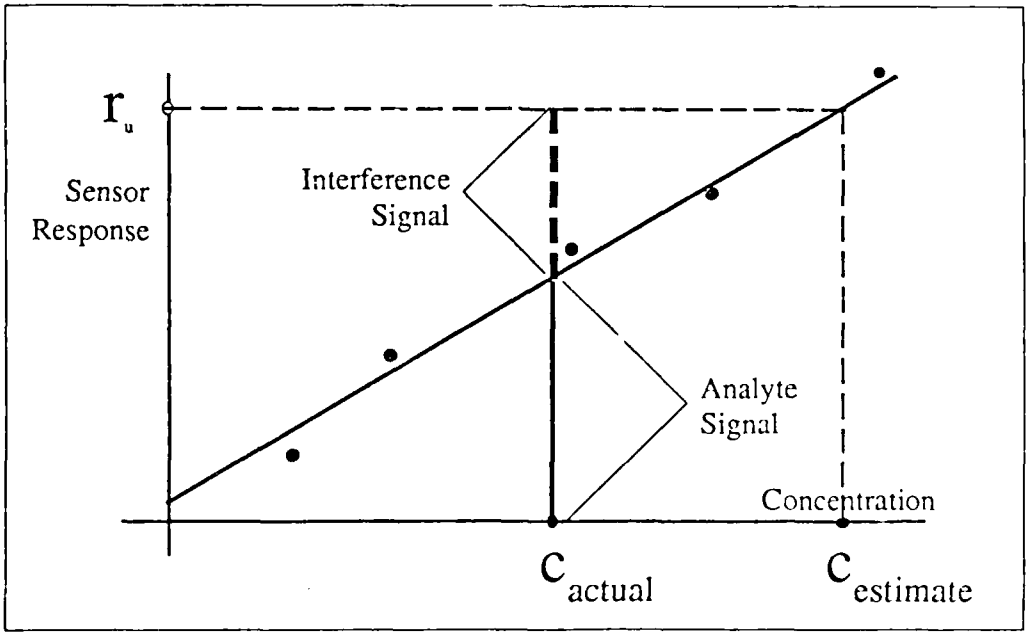
[12] Ramos, L. S.; Beebe, K. R.; Carey, W. P.; Sanchez M. E.; Erickson, B. C.; Wilson, B. E.; Wangen, L. E. and Kowalski, B. R. (1986) "Chemometrics". *Anal. Chem.* **58**, 294R-315R.

[13] Sanchez, E. (1987) "Tensorial Calibration: The Generalized Rank Annihilation Method". *Ph.D. Dissertation*, University of Washington.

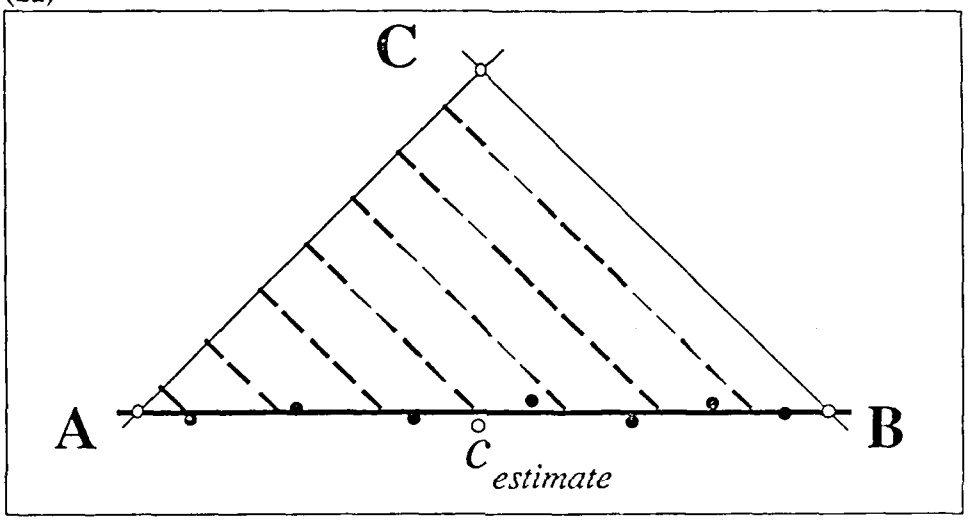
(1a)


Sensor
Response
r_u
Concentration
C_{actual}

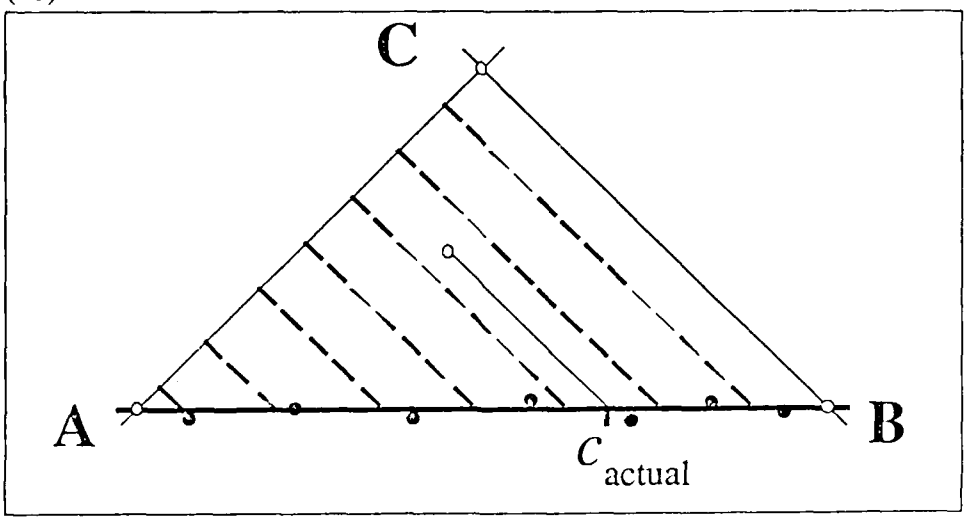
(1b)

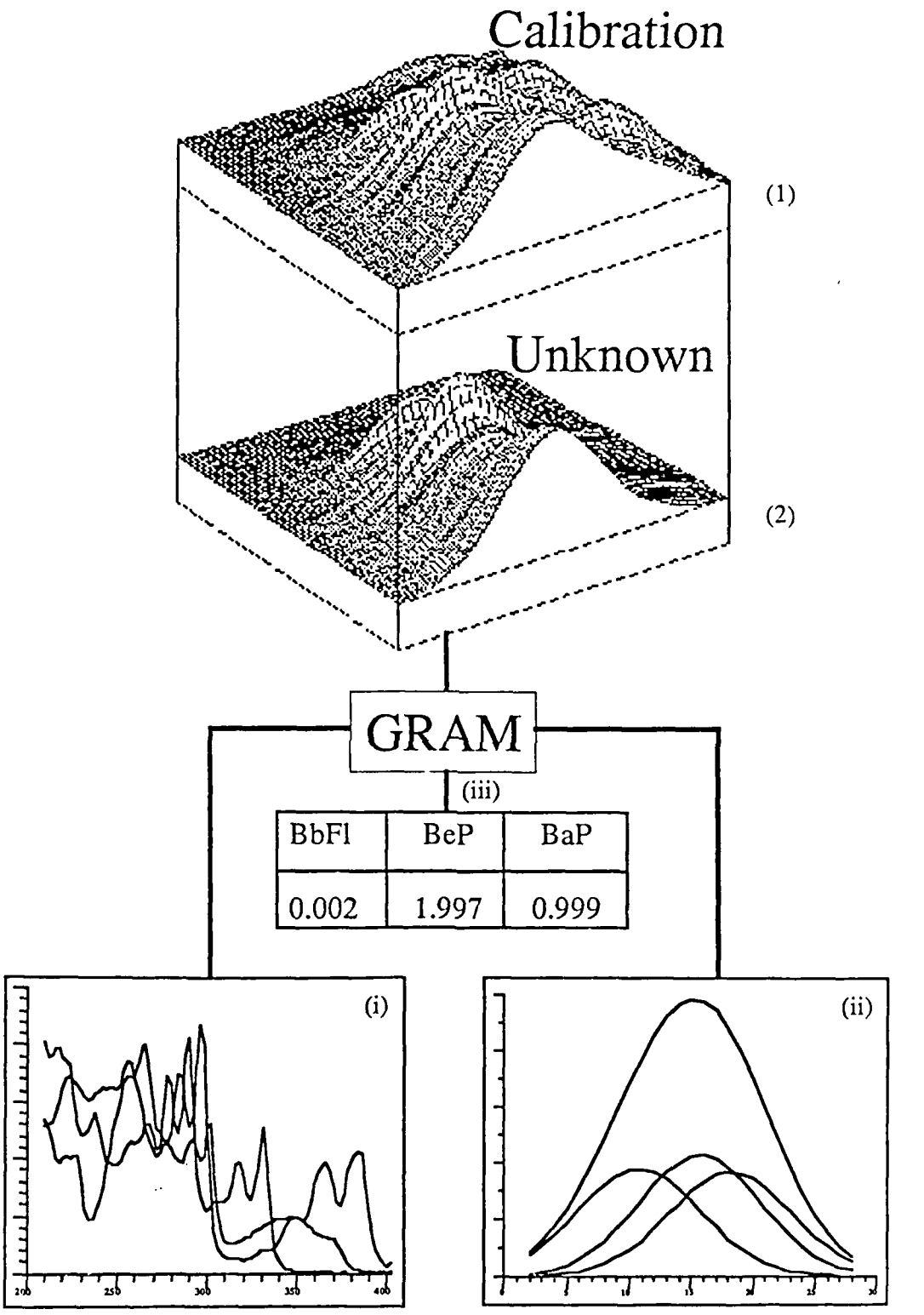

r_u
Sensor
Response
Interference
Signal
Analyte
Signal
Concentration
C_{actual}
$C_{estimate}$

(2a)



(2b)




Calibration
(1)
Unknown
(2)
GRAM
(iii)
BbFl
BeP
BaP
0.002
1.997
0.999
(i)
(ii)

DL/413/83/01
GEN/413-2

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Office of Naval Research Attn: Code 413 800 N. Quincy Street Arlington, Virginia 22217	2
Dr. Bernard Douda Naval Weapons Support Center Code 5042 Crane, Indiana 47522	1
Commander, Naval Air Systems Command Attn: Code 310C (H. Rosenwasser) Washington, D.C. 20360	1
Naval Civil Engineering Laboratory Attn: Dr. R. W. Drisko Port Hueneme, California 93401	1
Defense Technical Information Center Building 5, Cameron Station Alexandria, Virginia 22314	12
DTNSRDC Attn: Dr. G. Bosmajian Applied Chemistry Division Annapolis, Maryland 21401	1
Dr. William Tolles Superintendent Chemistry Division, Code 6100 Naval Research Laboratory Washington, D.C. 20375	1
Dr. David Young Code 334 NORDA NSTL, Mississippi 39529	1
Naval Weapons Center Attn: Dr. A. B. Amster Chemistry Division China Lake, California 93555	1
Scientific Advisor Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380	1
U.S. Army Research Office Attn: CRD-AA-IP P.O. Box 12211 Research Triangle Park, NC 27709	1
Mr. John Boyle Materials Branch Naval Ship Engineering Center Philadelphia, Pennsylvania 19112	1
Naval Ocean Systems Center Attn: Dr. S. Yamamoto Marine Sciences Division San Diego, California 91232	1

DL/413/83/01
051B/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Dr. R. A. Osteryoung
Department of Chemistry
State University of New York
Buffalo, New York 14214

Dr. J. Osteryoung
Department of Chemistry
State University of New York
Buffalo, New York 14214

Dr. H. Chernoff
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dr. A. Zirino
Naval Undersea Center
San Diego, California 92132

Dr. George H. Morrison
Department of Chemistry
Cornell University
Ithaca, New York 14853

Dr. Alan Bewick
Department of Chemistry
Southampton University
Southampton, Hampshire
ENGLAND SO9 5NH

Dr. M. B. Denton
Department of Chemistry
University of Arizona
Tucson, Arizona 85721

Dr. S. P. Perone
Lawrence Livermore National
Laboratory L-370
P.O. Box 808
Livermore, California 94550

Dr. G. M. Hieftje
Department of Chemistry
Indiana University
Bloomington, Indiana 47401

Dr. Christie G. Enke
Department of Chemistry
Michigan State University
East Lansing, Michigan 48824

Walter G. Cox, Code 3632
Naval Underwater Systems Center
Building 148
Newport, Rhode Island 02840

Professor Isiah M. Warner
Department of Chemistry
Emory University
Atlanta, Georgia 30322

Dr. Kent Eisentraut
Air Force Materials Laboratory
Wright-Patterson AFB, Ohio 45433

Dr. Adolph B. Amster
Chemistry Division
Naval Weapons Center
China Lake, California 93555

Dr. B. E. Douda
Chemical Sciences Branch
Code 50 C
Naval Weapons Support Center
Crane, Indiana 47322

Dr. John Eyler
Department of Chemistry
University of Florida
Gainesville, Florida 32611

DL/413/83/01
0518/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Professor J. Janata
Department of Bioengineering
University of Utah
Salt Lake City, Utah 84112

Dr. J. DeCorpo
NAVSEA
Code 05R14
Washington, D.C. 20362

Dr. Charles Anderson
Analytical Chemistry Division
Athens Environmental Laboratory
College Station Road
Athens, Georgia 30613

Dr. Ron Flemming
B 108 Reactor
National Bureau of Standards
Washington, D.C. 20234

Dr. Frank Herr
Office of Naval Research
Code 422CB
800 N. Quincy Street
Arlington, Virginia 22217

Professor E. Keating
Department of Mechanical Engineering
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. M. H. Miller
1133 Hampton Road
Route 4
U.S. Naval Academy
Annapolis, Maryland 21401

Dr. Clifford Spiegelman
National Bureau of Standards
Room A337 Bldg. 101
Washington, D.C. 20234

Dr. Denton Elliott
AFOSR/NC
Bolling AFB
Washington, D.C. 20362

Dr. B. E. Spielvogel
Inorganic and Analytical Branch
P.O. Box 12211
Research Triangle Park, NC 27709

Ms. Ann De Witt
Material Science Department
160 Fieldcrest Avenue
Raritan Center
Edison, New Jersey 08818

Dr. A. Harvey
Code 6110
Naval Research Laboratory
Washington, D.C. 20375

Mr. S. M. Hurley
Naval Facilities Engineering Command
Code 032P
200 Stovall Street
Alexandria, Virginia 22331

Ms. W. Parkhurst
Naval Surface Weapons Center
Code R33
Silver Spring, Maryland 20910

Dr. M. Robertson
Electrochemical Power Sources Division
Code 305
Naval Weapons Support Center
Crane, Indiana 47522

Dr. Andrew T. Zander PI204
Perkin-Elmer Corporation
901 Ethan Allen Highway/MS905
Ridgefield, Connecticut 06877

DL/413/83/01
051B/413-2

ABSTRACTS DISTRIBUTION LIST, 051B

Dr. Marvin Wilkerson
Naval Weapons Support Center
Code 30511
Crane, Indiana 47522

Dr. J. Wyatt
Naval Research Laboratory
Code 6110
Washington, D.C. 20375

Dr. J. MacDonald
Code 6110
Naval Research Laboratory
Washington, D.C. 20375

Dr. H. Wohltjen
Naval Research Laboratory
Code 6170
Washington, D.C. 20375

Dr. John Hoffsommer
Naval Surface Weapons Center
Building 30 Room 208
Silver Spring, Maryland 20910

Dr. Robert W. Shaw
U.S. Army Research Office
Box 12211
Research Triangle Park, NC 27709