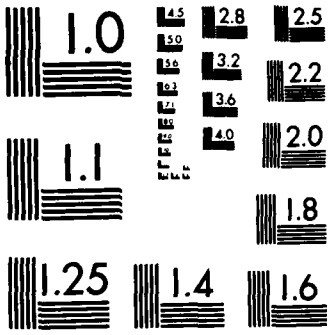
AD-A153 805 CHEMOMETRIC ANALYSIS OF MULTISENSOR ARRAYS(U) 1/1
WASHINGTON UNIV SEATTLE LAB FOR CHEMOMETRICS
W P CAREY ET AL. 15 APR 85 TR-33 N00014-75-C-0536

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A153 805

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR33
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Chemometric Analysis of Multisensor Arrays
5. TYPE OF REPORT & PERIOD COVERED: Technical Report - Interim
6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): W. P. Carey, K. R. Beebe, E. Sanchez, P. Geladi and B. 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, 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

12. REPORT DATE: April 15, 1985
13. NUMBER OF PAGES: 17

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

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.

DTIC ELECTE MAY 16 1985 S D A

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

18. SUPPLEMENTARY NOTES:
Prepared for publication in the ACS Symposium Series book entitled "Chemical Sensors"; accepted.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Multivariate Analysis
Multisensor Arrays
Partial Least Squares (PLS)
Multiple Linear Regression (MLR)
Principal Component Regression (PCR)
Rank Annihilation Factor Analysis (RAFA)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):
A discussion of multivariate calibration techniques and their possible application to sensor array data is reviewed. The progression from multiple linear regression, and principal components regression, to partial least squares can be described by the analysis of their respective algorithms and the data analysis problems each of them solve. Additional analysis models are introduced to detect underlying background problems for sensor arrays and the ability to quantitate analytes directly by rank annihilation factor analysis when two dimensional data is obtained.

DTIC FILE COPY

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)

85 4 2 077

OFFICE OF NAVAL RESEARCH

Contract N00014-75-C-0536

Task No. NR 051-565

TECHNICAL REPORT NO. 33

Chemometric Analysis of Multisensor Arrays

by

W. P. Carey, K. R. Beebe, E. Sanchez, P. Geladi and B. R. Kowalski

Prepared for Publication

in

ACS Symposium Seris book "Chemical Sensors"

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

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

April 15, 1985

DDC
QUALITY INSPECTED
1

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

Chemometric Analysis of Multisensor Arrays

W. Patrick Carey, Kenneth R. Beebe,
Eugenio Sanchez, Paul Geladi, Bruce R. Kowalski.

Laboratory for Chemometrics
Department of Chemistry, BG-10
University of Washington
Seattle, WA 98195

ABSTRACT

A discussion of multivariate calibration techniques and their possible application to sensor array data is reviewed. The progression from multiple linear regression, and principal components regression, to partial least squares can be described by the analysis of their respective algorithms and the data analysis problems each of them solve. Additional analysis models are introduced to detect underlying background problems for sensor arrays and the ability to quantitate analytes directly by rank annihilation factor analysis when two dimensional data is obtained.

With the arrival of new and novel types of chemical sensors, the chemist will have to meet a new challenge in data acquisition and analysis. The abundant information content that was once only common in areas of chemistry such as spectroscopy, where hundreds of wavelengths are scanned, is now available for microsensors. The progress of chemical sensors from the pH electrode to the CHEMFET has been remarkable. With microlithography, hundreds of solid state sensors can be combined on a single chip to form a data gathering array equivalent to a spectrometer. The next problem facing the user of sensor arrays is that of data analysis. The complexities of multisensor array responses or outputs have to be sorted into useful information since solid state technology does not eliminate such fundamental problems as matrix effects, interferences, and backgrounds. The variety of techniques used in multivariate calibration and quantitation

and how they might apply to the chemical sensor array will be discussed in this paper.

Linear Regression.

The single-component, single-sensor model used in most analyses is a linear relationship between the response of a sensor and the concentration of an analyte:

$$r = b_0 + b_1 c + \epsilon \qquad (1)$$

This model requires that a calibration step be performed to find the sensitivity of the sensor, b_1 (slope of the regression line), and the r-axis intercept, b_0. The residual error ϵ is the portion of the response not described by the model. The second step in the analysis is then performed by measuring the responses of unknown samples and then estimating their concentrations. The assumptions made using least squares regression are that the calibration plot is linear and the sensor is fully specific for the analyte of interest (this also assumes that b_0 is zero). The most commonly encountered problems using this method are matrix effects and the presence of a background response. Matrix effects are due to a change in the sensitivity coefficient b_1 between calibration and quantitation. This type of effect is most readily treated by using the method of standard additions. The addition of a background or interference correction cannot be directly included in this model, especially if the interferent is unknown. This suggests the use of sensor arrays which are necessary for the analysis of mixtures. The advantages include reduction in analysis time and the allowance of the use of non-specific sensors.

The data obtained from an array of sensors does not differ from that obtained from more classical analytical instrumentation. For this reason, the types of data analysis used for sensor arrays does not differ from those

commonly used in, for example, emission spectroscopy. In emission spectroscopy, the spectrum of an analyte is defined as the emission intensities of the analyte at a predetermined number of wavelengths. A more unique or distinguishing spectrum is one in which more wavelengths are employed. An important characteristic that determines the amount of information contained in a spectrum is the degree of orthogonality or independence of the wavelengths. This becomes clear if one considers the hypothetical case of a spectrum defined by two wavelengths. If one adds a third wavelength whose intensity is always the sum of the first two, no new information is obtained by its inclusion. For an array of sensors, the "spectrum" or signature of an analyte is the composite of responses of each sensor to the analyte. It is not a true spectrum in that the ordering of the sensors, and therefore the shape of the spectrum, is arbitrary. In true spectra, either time or a progression of wavelengths serves to define their shapes, although adjacent wavelengths are often highly correlated. Such a correlation arises since adjacent wavelengths are nearly equal in energy. When this occurs, many wavelengths must be used to obtain information that could be contained in a smaller, more independent array of sensors. To improve the uniqueness of an array signature, more sensors can be used. Therefore, one powerful aspect of arrays is that sensors whose responses are independent can be chosen. The mechanisms which determine the responses of the individual sensors can be chosen to be nearly orthogonal.

Multiple Linear Regression.

One of the most widespread uses of chemometrics is in the estimation of a model used for calibration. This is often done for a multicomponent mixture where the concentration of p analytes in a mixture is sought. Perhaps the simplest method to understand is the method of multiple linear regression

(MLR) (1). With an array of sensors, a typical problem would be to determine the concentrations of p analytes in n samples where an array of m sensors is employed, Figure 1. To solve this problem, two assumptions must first be made. First, it is assumed that the sensors respond linearly. This states that doubling the concentration of an analyte will double that analyte's response from each sensor. Second, it is assumed that the response of a sensor to two or more analytes is additive. For example, if a sensor has a response of two and three units to one molar solutions of analytes A and B, respectively, then its response to a mixture of one molar A and one molar B will be five units. With these assumptions, the problem can be solved using MLR.

The response matrix **X** is a n x m matrix whose rows are the n samples and columns are the responses of the m sensors:

$$\mathbf{XK} = \mathbf{Y} \qquad (2)$$

Y is a matrix of unknown concentrations with n samples as rows and p analytes as columns, and **K** is the calibration matrix of regression coefficients. To solve the problem, one must first use a calibration set ($\mathbf{X}_0$ and $\mathbf{Y}_0$ are the matrices of calibration responses and concentrations) where the analyte concentrations for the n samples are known. Knowing $\mathbf{X}_0$ and $\mathbf{Y}_0$, **K** can be determined by linear algebra. If the number of samples does not equal the number of sensors, then $\mathbf{X}_0$ is not a square matrix, and its inverse cannot be determined. Thus both sides of the equation must be multiplied by $\mathbf{X}_0^T$ before the inverse is found.

$$\mathbf{K} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{Y}_0 \qquad (3)$$

Once this **K** is determined, the initial $\mathbf{X}_{mix}$ obtained from the unknown mixtures can be used to find the concentrations of the mixtures.

$$\mathbf{Y}_{mix} = \mathbf{X}_{mix}\,\mathbf{K} \qquad (4)$$

Note that this **K** matrix does not contain information about the sensitivities of the sensors to the analytes. Since this information is often important, the matrix of sensitivities $\mathbf{K}^{+}$ can be derived as follows:

$$\mathbf{K}^{+} = \mathbf{K}^{T}\,(\mathbf{K}\mathbf{K}^{T})^{-1} \qquad (5)$$

This equation calculates the best least square fit describing the data points for each analyte and each sensor using the criteria. It is an adequate method in many cases, but in some instances its application is either not appropriate or impossible. In the under-determined case where the number of analytes (p) is greater than the number of sensors (m), MLR gives an infinite number of solutions to the problem and is therefore useless. When p equals m, there is one unique solution, and when p is less than m, more information is available and therefore, a better statistical fit is achieved.

Another common problem is collinearity, which occurs when the $\mathbf{X}_0$ matrix does not have full rank, i.e. some number z of the columns in $\mathbf{X}_0$ are nearly linear combinations of the remaining m-z columns. When this occurs, the inversion procedure becomes sensitive to errors when calculating **K**, and small errors in an observed sensor response value in **X** will result in large errors in the corresponding concentration estimates in **Y**.

Background Identification.

It is assumed in the MLR calibration step that the calibration mixture contains the same or greater number of analytes as in the unknown mixture. The unknown presence of a background, an uncalibrated analyte, if not eliminated, will cause erroneous results in the measurement of a mixture. A method proposed by Osten and Kowalski for background identification can be used to test whether further calibration or sample purification is necessary (2). This technique uses a new maxtrix **X** where the first N rows of **X** are identical to the **K** calibration matrix (actually each row contains the responses of each sensor to a pure analyte) and the N+1st row is the measured responses of a mixture of analytes (e.g. a real unknown sample).

The method involves the normalization of **X** by rows (the sum of each row element equals one) and then mean-centering the data by subtraction of the mean response of each sensor for the N analytes from each entry for that sensor in the matrix. The second moment matrix $\mathbf{X}'\mathbf{X}/N$ is calculated using only the first N rows of the normalized, mean-centered **X** matrix. Diagonalization of this moment matrix gives rise to two matrices, one being N-1 eigenvalues, E, and the other being N-1 eigenvectors, **V**.

Now the scaled and centered mixture response vector, $\mathbf{x}_{N+1}$ (the N+1 row of **X**), can be rotated by the eigenvector matrix **V**.

$$\mathbf{s}_{N+1} = \mathbf{x}_{N+1}\mathbf{V} \qquad (6)$$

The mixture response vector can be estimated by the eigenvectors and the above scores vector, $\mathbf{s}_{N+1}$, by

$$\mathbf{x}''_{N+1} = \mathbf{V}\mathbf{s}_{N+1} \qquad (7)$$

The difference between the actual and predicted response vectors results in the residuals vector h.

$$\mathbf{h} = \mathbf{x}_{N+1} - \mathbf{x}''_{N+1} \qquad (8)$$

The sum of the squares of each element of $\mathbf{x}_{N+1}$ and $\mathbf{x}''_{N+1}$ are calculated and compared. If the residual vector h contains more than $1 \times 10^{-3}\%$ of the original variance in the mixture response, then a background component is present. The major problem with sensor array data is that background detection is possible but not correctable without knowing what the components of the background are in order to include them in the calibration step.

Principal Component Regression.

An alternative to the linear calibration problem, principal component regression (PCR), couples factor analysis or principal component analysis (PCA) with multi-linear regression (3). This method is not as sensitive to the collinearity problem and aids in the determination of the best solution in the under-determined case. PCA is a method in which more descriptive variables (which correspond to columns) are calculated. These new variables, which are called principal components or eigenvectors, are linear combinations of the original columns. They are more descriptive because they are chosen to describe the maximum amount of variance in a data matrix. To illustrate this, the entries in each column can be viewed as defining a corresponding vector's orientation in m-dimensional space. The first eigenvector is the vector whose direction describes the maximum amount of variance out of all of the possible directions. The second eigenvector is by definition orthogonal to the first and is the second most descriptive direction. Since the eigenvectors are in m-space, the maximum number of eigenvectors equals the number of columns or m. Often times r eigenvectors or columns, where $r<p$, can be used to describe all or nearly all of the variance in a data matrix. When this is true, the m x p matrix can be reduced to a m x r matrix where the new columns in the reduced matrix are

independent and are the coordinates of the samples in the new coordinate system defined by the eigenvectors (as opposed to the original columns). The first step in PCR is to perform PCA on **X** and calculate a new smaller matrix **S** (n x r) using the first r eigenvectors (r<m). This step is identical to that used in equation 6. A similar procedure called cross-validation (4) can also be used to optimally choose a number of components when reduction of dimensionality is necessary.

The next step is to perform MLR of **S** onto **Y**.

$$\mathbf{SK} = \mathbf{Y} \qquad (9)$$

Because the columns of **S** are orthogonal, there is no collinearity problem, and the number of eigenvectors used can be chosen to accommodate the number of rows or samples present. Of course the best approach is to always be over-determined (m>p), but where this is not possible, the best result is that in which the maximum amount of information is used to derive the solution.

When using an array of sensors, PCA can have another important use. Performing PCA on the **X** matrix will yield new columns that are linear combinations of the original columns as stated before. The loadings or contributions of the original columns to the new columns can be examined to determine the information content of the original columns. Since these columns correspond to the responses of individual sensors, an analyst can use the results of PCA to determine whether or not a sensor is useful (5). Informative sensors will load heavily into the first r eigenvectors where r is chosen such that the first r eigenvectors describe a predetermined amount of the variance in the data set. In this way many sensors can be tested simultaneously, and the sensors that contribute the most to the first r eigenvectors are selected to form the array. The best approach is to

select the major contributor to the first eigenvector as the first sensor, the major contributor to the second eigenvector as the second sensor, and so on. One would not chose all of the sensors that loaded into the first eigenvector, even though it is the most descriptive, because columns that load into the same eigenvectors are usually highly correlated and therefore would cause collinearity problems. The successive eigenvectors are orthogonal; therefore the major contributors to the successive eigenvectors are also more nearly orthogonal.

Partial Least Squares

One of the latest regression procedures to be developed is that of partial least squares (PLS). PLS was first described in the middle of the 1960's by Herman Wold (6). It was used moderately in the fields of econometrics, sociology and psychology during the seventies (7). The first use in chemistry was reported by H. Wold, B. Gerlach and B. Kowalski in the late seventies (7, Vol. 2, chapt.9). Since then, the groups of Svante Wold at Umeå University (Umeå, Sweden) and Harald Martens at the Norwegian Food Research Institute (Ås, Norway) have been refining and specializing the method for chemical applications. References 8, 9 and 10 give a description of that work.

The PLS method of regression is based on the properties of multiple linear regression (MLR) and of principal component analysis (PCA). It is considered the best of both methods. The important aspects of the PLS method are:

-Model building

-Prediction (Figure 2)

-Parsimony

One important field of use of PLS is in multivariate calibration in analytical chemistry. Applications in this field can easily be extended to the calibration of sensors. Up to now, almost all chemical applications have used the 2-block PLS model, where the response and concentration matrices, Figure 1, are considered blocks of data. One advantage of PLS is the use of more than one response block to be regressed with the concentration block. This aspect of PLS has not been applied in a chemical experiment as of this time.

The 2-block PLS uses the same data structure as MLR, except the regression algorithm decomposes both blocks into sums of simpler matrices. Principal Component Regression (PCR) is the case where the X-block is decomposed and the regression is modeled from the X-block scores against the Y-block, equation 9. In PLS, rotated factors called latent variables are used for the regression part instead of principal components. The variables are rotated for optimizing the correlation between the scores of both blocks as in Canonical Correlation. Although it is impossible to describe completely how PLS works in this paper, interested readers are referred to a tutorial on PLS written by Geladi and Kowalski (11). Table I gives a comparison between MLR, PCR and PLS.

Table I.
A comparison of MLR, PCR and PLS

	MLR	PCR	PLS
1.	#samples > #sensors	no requirement	no requirement
2.	works best with orthogonal sensors	accepts collinear sensors	accepts collinear sensors
3.	matrix inversion is difficult	matrix inversion is easy	no matrix inversion

4.	no data on matrix condition	matrix condition data available	matrix condition data available
5.	block data is not analyzed	factor analysis part allows classification and pattern recognition	factor analysis part allows classification and pattern recognition
6.	sensitive to noise	sensitive to noise	separates noise from relevant information
7.	multiple dependent variables are treated independently	multiple dependent variables are treated independently	makes meaningful linear combinations in dependent block

PLS as a regression method can be used no matter how many variables (sensors) there are in the X and Y blocks and collinearity problems can be avoided. PLS and PCR, by their nature give data on the condition of the X-block matrix. For MLR this would require an extra calculation step that is almost never carried out by "black-box" MLR users. One of the main advantages of PLS over PCR and its applicability to sensor arrays is that it can separate noise from useful information.

Rank Annihilation Factor Analysis.

The analytical chemist is frequently confronted with the problem of analyzing complex mixtures in which he is only interested in the concentration of a few components. It would be convenient if quantitative information could be obtained for the analytes of interest without worrying about the rest of the sample components. Second order bilinear sensors, i.e. sensors that give a two dimensional data matrix of the form $M_{ij}=\sum_k \beta_k x_{ik} y_{jk}$,

are specially suited for this purpose, and the technique for quantitation is known as rank annihilation([12-13]). So far this method has been applied to excitation-emission fluorescence([12-14]) and to LC/UV ([15]) with excellent results. One problem in the calculation has been that an iterative solution requiring many matrix diagonalizations was necessary. Lorber ([16]) has reported a noniterative solution, rank annihilation factor analysis (RAFA), presenting the problem as a generalized eigenvalue problem in which a direct solution is found by using the singular value decomposition.

In practice, $\mathbf{N}_k$, the bilinear spectrum of a pure compound **k**, is known, and **M**, the bilinear spectrum of a mixture sample where the **k** compound is present, is measured. This data matrix **M** can be expressed as a linear combination of the n pure components bilinear spectra $\mathbf{N}_k$:

$$\mathbf{M} = \sum_k \beta_k \mathbf{N}_k \qquad \text{where } \mathbf{N}_k = \mathbf{x}_k \mathbf{y}_k^T \ ; \ (N_{ij})_k = x_{ik}\, y_{jk} \qquad (10)$$

The $\mathbf{x}_k$ are column vectors with information in one order, e.g. excitation spectra, and the $\mathbf{y}_k^T$ are row vectors with information in the second order, e.g. emission spectra. If $\mathbf{N}_k$ is defined as <u>unitary</u> concentration pure component bilinear spectra, then β_k is the <u>concentration</u> of the $\mathbf{k}^{th}$ compound.

If the data matrix **M** has rank p, subtracting from **M** the right amount of $\mathbf{N}_k$, i.e. $\beta_k \mathbf{N}_k$, the resultant matrix will have rank p-1, or in a equivalent manner

$$\det(\mathbf{M} - \beta_k \mathbf{N}_k) = 0 \qquad (11)$$

To solve this equation for β_k, the generalized eigenvalue problem is applied. The matrices **M** and $\mathbf{N}_k$ are normally rectangular, so a transformation is necessary.

Equation (11) can be rewritten as:

$$\det(\mathbf{N}_k - \lambda_k \mathbf{M}) \qquad \lambda_k = 1/\beta_k \qquad (12a)$$

$$\mathbf{N}_k \mathbf{z} = \lambda_k \mathbf{M} \mathbf{z} \tag{12b}$$

Next the singular value decomposition of the **M** matrix is obtained.

$$\mathbf{M} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T \tag{13}$$

where

$$\mathbf{M}\,\mathbf{V} = \mathbf{S}\,\mathbf{U} \tag{14}$$

$$\mathbf{M}^T\mathbf{U} = \mathbf{S}\,\mathbf{V} \tag{15}$$

$$\mathbf{M}^T\mathbf{M}\mathbf{V} = \mathbf{S}^2\mathbf{V} \qquad \text{eigen-equations in } \mathbf{V} \text{ space.} \tag{16}$$

$$\mathbf{M}\mathbf{M}^T\mathbf{U} = \mathbf{S}^2\mathbf{U} \qquad \text{eigen-equations in } \mathbf{U} \text{ space.} \tag{17}$$

The second step is to determine the number of significant eigenvalues p (equations 16-17) using abstract factor analysis (15). To transform equation 12 to the normal eigenvalue equation, a new matrix $\underline{\mathbf{M}}$ which is obtained from **U**,**V** and **S** is generated by taking the first p columns,

$$\underline{\mathbf{M}} = \underline{\mathbf{U}}\,\underline{\mathbf{S}}\,\underline{\mathbf{V}}^T \tag{18}$$

substituting $\underline{\mathbf{M}}$ *for* **M** *in equation 12b.*

$$\mathbf{N}_k \mathbf{z} = \lambda_k \underline{\mathbf{M}}\, \mathbf{z} \tag{19}$$

$$\mathbf{N}_k \mathbf{z} = \lambda_k \underline{\mathbf{U}}\,\underline{\mathbf{S}}\,\underline{\mathbf{V}}^T \mathbf{z} \tag{20}$$

The eigenvector **z** is replaced by $\mathbf{z} = \underline{\mathbf{V}}\,\underline{\mathbf{S}}^{-1}\mathbf{z}'$, where $\mathbf{z}' = \underline{\mathbf{S}}\,\underline{\mathbf{V}}^T\,\mathbf{z}$; therefore

$$\mathbf{N}_k\,\underline{\mathbf{V}}\,\underline{\mathbf{S}}^{-1}\mathbf{z}' = \lambda_k \underline{\mathbf{U}}\,\mathbf{z}' \tag{21}$$

Left multiplying by $\underline{\mathbf{U}}^T$ results in

$$(\underline{\mathbf{U}}^T\,\mathbf{N}_k\,\underline{\mathbf{V}})\,\underline{\mathbf{S}}^{-1}\mathbf{z}' = (\underline{\mathbf{U}}^T\lambda_k\underline{\mathbf{U}})\,\mathbf{z}' = \lambda_k\,\mathbf{z}'$$

$$(\underline{\mathbf{U}}^T\,\mathbf{N}_k\,\underline{\mathbf{V}}\,\underline{\mathbf{S}}^{-1})\,\mathbf{z}' = \lambda_k\,\mathbf{z}' \tag{22}$$

This is the normal eigenvalue equation, with matrix $(\underline{\mathbf{U}}^T\,\mathbf{N}_k\,\underline{\mathbf{V}}\,\underline{\mathbf{S}}^{-1})$ being square. Because the rank of $\mathbf{N}_k$ is one, there will be p-1 zero solutions for the eigenvalues. Therefore, the only non-zero solution will be equal to the trace of matrix $(\underline{\mathbf{U}}^T\,\mathbf{N}_k\,\underline{\mathbf{V}}\,\underline{\mathbf{S}}^{-1})$. By calculating the trace of the above matrix, the concentration of that component (inverse of λ) is solved directly. Other analytes can then be analyzed sequentially.

It has been shown that a wide variety of data analysis techniques can be applied to sensor array data. Since each method offers different advantages, it is essential that the proper technique used should maximize information received while minimizing error at the cost of programming complexity. A direct result of the above comparison of multivariate techniques shows that in moving toward sensor arrays, problems such as interferences, if known, can be calibrated into a model for quantitation. Futhermore, the more complex the response data of a sensor array, time responses for example, the more useful information can be extracted as seen in RAFA. A direct result of this type of data analysis could provide a sensor array on a single silicon chip which could directly measure 10 to 20 blood constituents intravenously in minutes. A number of studies are underway in our laboratory aimed at extending the calibration mathematics described above and applying them to a variety of sensor arrays for process monitoring and control.

Acknowledgments

This work is supported by a grant from the Office of Naval Research.

Literature Cited

1. Kowalski, B.; Proc. of the NATO Advanced Study Institute on Chemometrics-Mathematics and Statistics in Chemistry, 1983, p. 82.
2. Osten, D.; Kowalski, B. Anal. Chem. 1985, 57, 908.
3. Draper, N.; Smith, H. In "Applied Regression Analysis"; J. Wiley: New York, 1981; p. 327.
4. Eastment, H.T.; Krzanowski, W.J. Technometrics 1982, 24, 73.
5. Carey, W.P.; Beebe, K.R.; Kowalski, B.R.; Illman, D.L.; Hirschfeld, T. Anal. Chem. 1985, submitted for publication.
6. Wold, H.; In "Nonlinear Estimation by Iterative Least Squares Procedures"; *Festschrift Jerzy Neyman*; Wiley: New York, 1966; p. 411-444.
7. Jöreskog, K.; Wold, H. eds.; "Systems under Indirect Observation: Causality - Structure - Prediction, Part II"; North Holland: Amsterdam, 1982.
8. Wold, S.; et. al. In "Pattern Recognition: finding and using Regularities Multivariate Data", Martens, H.; Russwurm, H. eds.; *Food research and Data Analysis*, Applied Sience Publishers: London, 1983; p.147-188.
9. Wold, S.; et. al. In "Multivariate Data Analysis in Chemistry"; Kowalski, B., ed.; *Chemometrics: Mathematics and Statistics in Chemistry*, Reidel Publishers: Dordrecht, 1984; p. 17-95.
10. Wold, S.; Martens, H.; Wold, H. In "The Multivariate Calibration Problem in Chemistry Solved by the PLS method"; Kågström; Ruhe eds.; *Matrix Pencils, Lecture Notes in Mathematics*, Springer: Heidelberg, 1983; p. 286-293.
11. Geladi, P.; Kowalski, B. 1985, *Partial Least Squares Regression (PLS): a Tutorial*, in press.
12. Ho, C-N.; Christian, G.D.; Davidson, E.R. Anal. Chem. 1978, 50, 1108-1113.
13. Ho, C-N.; Christian, G.D.; Davidson, E.R. Anal. Chem. 1980, 52, 1071-1079.
14. Ho, C-N.; Christian, G.D.; Davidson, E.R. Anal. Chem. 1981, 53, 92-98.
15. McCue, M.; Malinowski, E.R.; J. Chromatogr. Sci. 1983, 21, 229.
16. Lorber, A.; Anal. Chim. Acta 1985, in press.

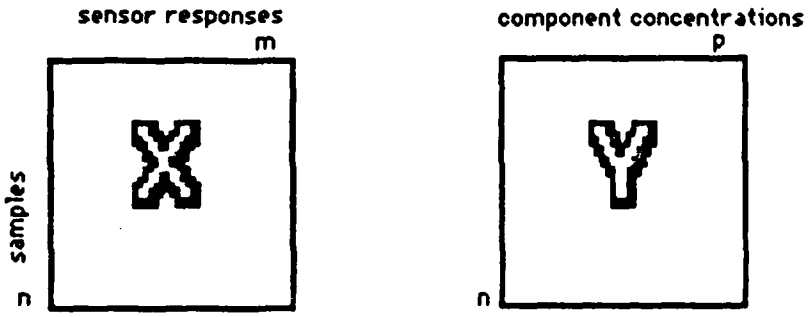


FIGURE 1. Data blocks or matrices using multivariate regression.

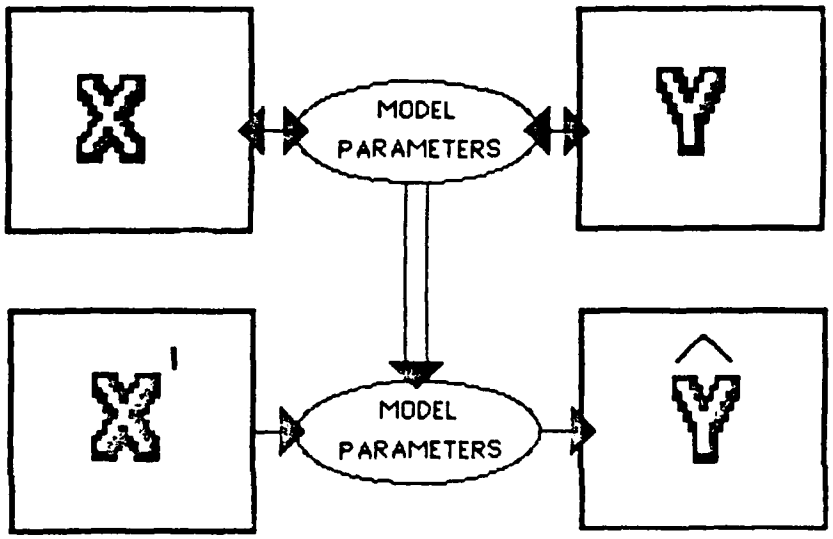


FIGURE 2. The predictive aspect of PLS. The top half of the diagram is used for the calibration data, while the bottom half is for the estimation for test data. X and Y are calibration data, X' is the unknown response data, and $\hat{Y}$ is the unknown concentration data

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
051B/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

END

FILMED

6-85

DTIC