Outline

- Beginnings
- Multi-way models
- Second order analytical techniques
- The Videofluorometer
- Applications in Analytical Chemistry
- Applications in Chemical Engineering

Beginnings (for me!)

- Started graduate school in Chemical Engineering at University of Washington in fall of 1985

- Took Chemometrics (Chem 526?) from Bruce R. Kowalski

- Course included lectures by his grad students on their current work

- Eugenio Sanchez talked about Rank Annihilation Factor Analysis (RAFA)

Sources of multi-way chemical data

- Many sources of multi-way data in chemistry

- By far the most common are
  - Fluorescence spectroscopy
  - Hyphenated techniques
    - Separation w/multi-channel detector
    - GC-MS, LC-UV, etc.
  - Batch chemical processes

Examples

Sensory analysis
- Score as a function of (Food sample, Judge, Attribute)

Process analysis
- Measurement as a function of (Batch, Variable, time)
- Measurement as a function of (Variable, Lag, Location)

Image analysis
- Pixelvalue as a function of (Sample, Image pixel, Variable)

Experimental design
- Response as a function of (factor 1, factor2, factor3,..)

Spectroscopy
- Intensity as a function of (Wavelength, Retention, Sample, Time, Location , Treatment)
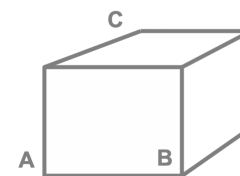
Environmental analysis
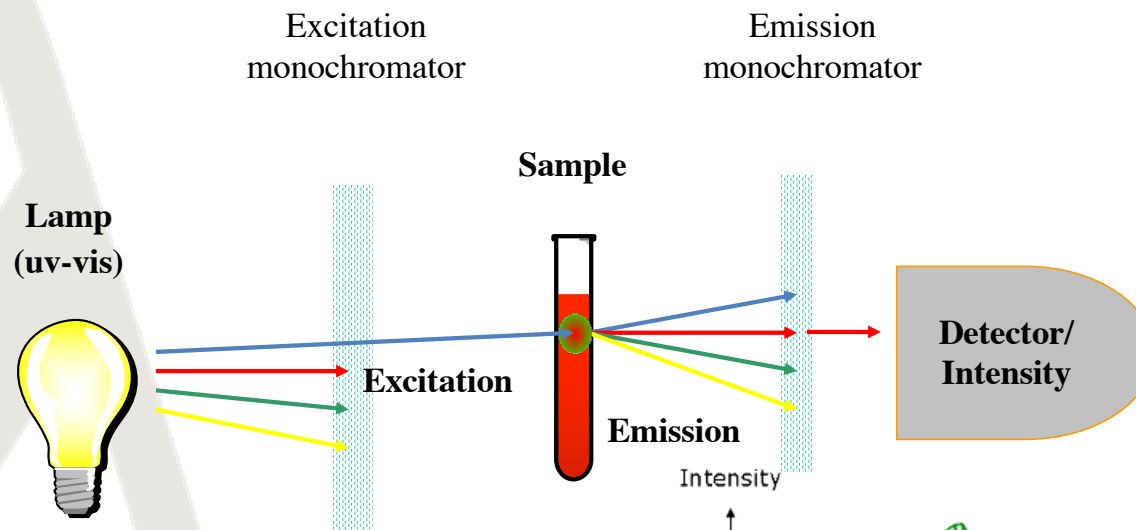- Measurement as a function of (Location, Time, Variable)

Chromatography
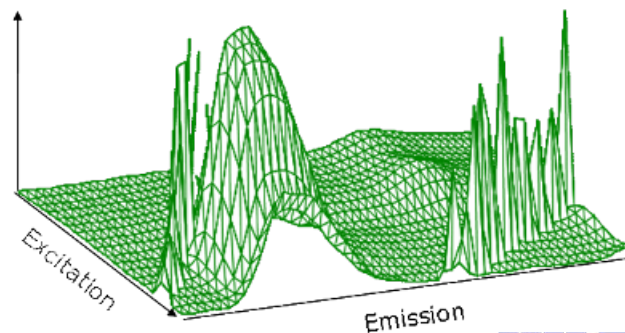- Measurement as a function of (Sample, Retention time, Variable)

Excitation-Emission Fluorescence

Excitation monochromator

Emission monochromator

Sample

Lamp (uv-vis)

Excitation

Emission

Detector/ Intensity

Intensity

Excitation-emission matrix – a chemical fingerprint

Excitation

Emission

Bilinear: Single analytes produce matrices of rank 1! Also linearly additive.

EIGENVECTOR RESEARCH INCORPORATED

Gas Chromatography-Mass Spectrometry

- Sample separated by gas or liquid chromatography
- Effluent measured by multi-channel detector, commonly mass spectrometry or UV/Vis

Also bilinear!

time

mass

EIGENVECTOR
RESEARCH INCORPORATED
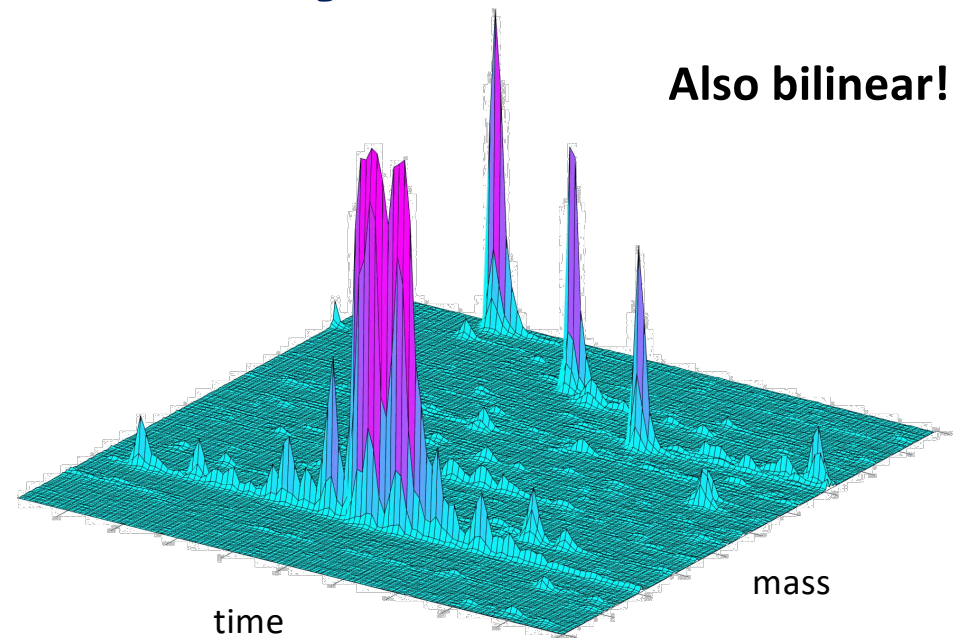
Batch Chemical Process Data

- Reactants loaded into single vessel
- Allowed to react for some specified time or until some end point is reached
- Typically controlled to some temperature
- Often have intermediate steps where other things are added or temperature is cooled, etc.
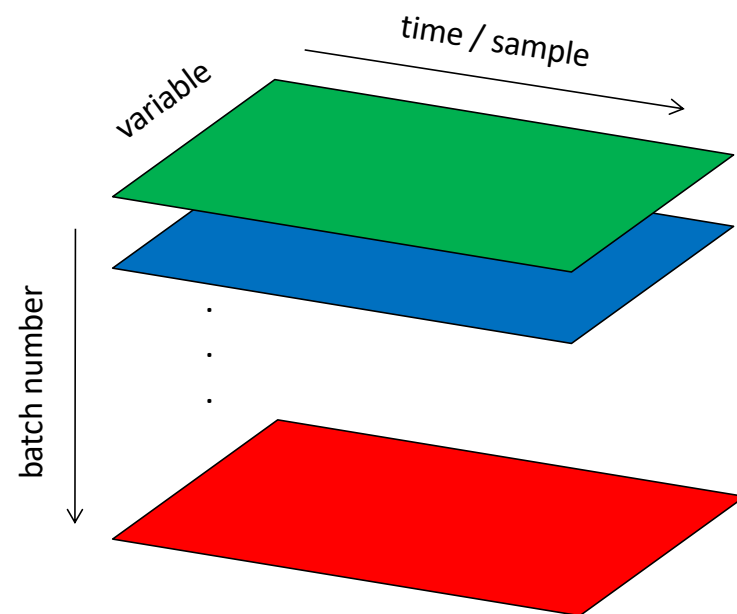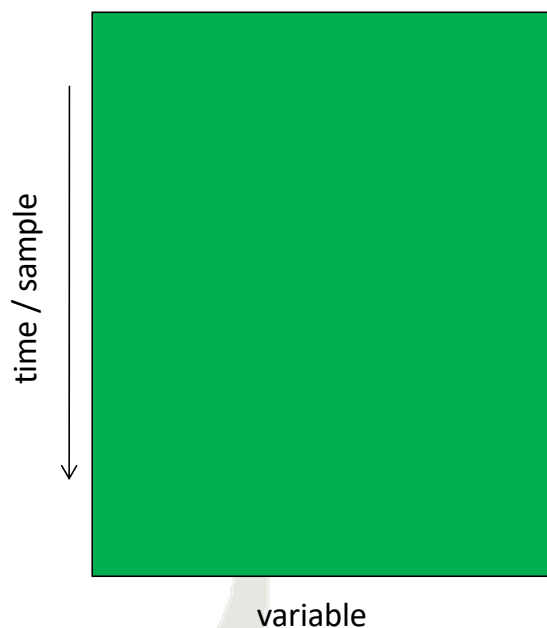
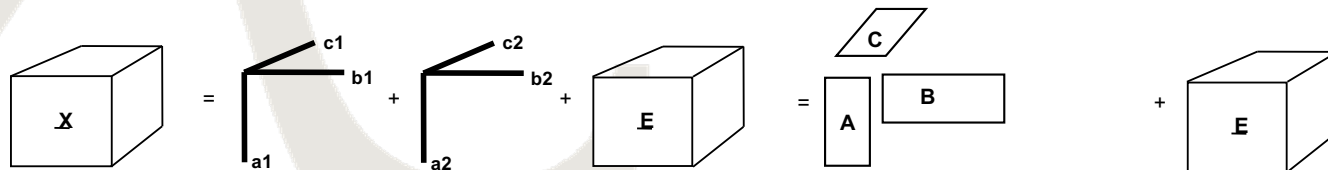Not generally bilinear!

PARAFAC

- PCA - bilinear model,

$$x_{ij} = \sum_{f=1}^{F} a_{if} b_{jf} + e_{ij}$$

- PARAFAC - trilinear model,

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

- Decomposition is unique under fairly mild conditions
- Form of the model identical to fluorescence and GC/MS etc.
- Provides good estimates of true underlying components (if rank is chosen correctly)

PARAFAC

…. or

Canonical Decomposition

=

## CanDecomp

…or combined

CP

R. A. Harshman, Foundations of the PARAFAC procedure: model and conditions for an 'explanatory' multi-mode factor analysis, *UCLA Working Papers in phonetics*, **16** (1970) 1.

J. D. Carroll, J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, *Psychometrika*, **35** (1970) 283.

The Tucker3 model

- For three-way data, three orthogonal bases, **A**, **B**, and **C**; one for each mode

- Tucker3 is $\mathbf{X} = \mathbf{AG(C \otimes B)}' + \mathbf{E}$

- Loadings are truncated bases and **G** the representation of **X** in these reduced spaces

- Number of factors in each mode can be different!

- Not unique, only subspace unique
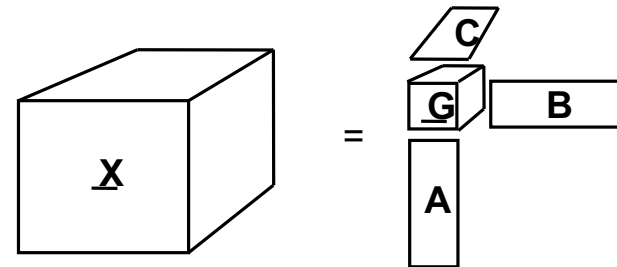
L. R. Tucker. The extension of factor analysis to three-dimensional matrices. In: *Contributions to Math. Psychology*, Eds. Frederiksen, Gulliksen, New York:Holt, Rinehart & Winston, 1964

L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31:279-311, 1966

EIGENVECTOR RESEARCH INCORPORATED

Sanchez and Kowalski

Generalized Rank Annihilation Factor Analysis

Sir: The analytical chemist is frequently confronted with the problem of analyzing complex mixtures of which only concentrations of a few components are of interest. In these cases, it is desirable to be able to obtain quantitative information for the analytes of interest without concern for the rest of the components in the sample. Second-order bilinear sensors, i.e., sensors that yield a two-dimensional data matrix is found by using the singular value decomposition. With his method, to obtain the concentrations of the p analytes of interest in the sample, its bilinear spectrum and the p calibration spectra for each pure analyte must be recorded to obtain the concentrations. Analysis for each analyte requires a separate calculation. This letter presents the generalized rank annihilation method, of which Lorber's noniterative

Anal. Chem. **1986**, *58*, 496–499

(1) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1978**, *50*, 1108–1113.
(2) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1980**, *52*, 1071–1079.
(3) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1981**, *53*, 92–98.

The Videofluorometer

- Acquired 241 fluorescence spectra at 241 excitation wavelengths
- Displayed them in three dimensions
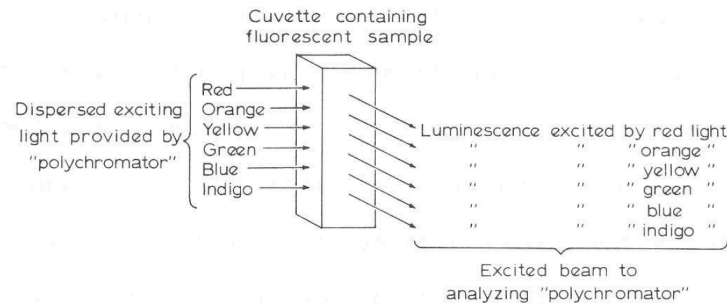- Data captured on a television camera element

FLUORESCENCE ANALYSIS: A NEW APPROACH

KEY WORDS: fluorimetry, videofluorometer, tv detector, simultaneous excitation and emission spectra

I. M. Warner, J. B. Callis, E. R. Davidson, M. Gouterman and G. D. Christian

Department of Chemistry
University of Washington
Seattle, Washington 98195

ABSTRACT

The Video Fluorometer is a new instrument for fluorescence analysis which offers significant advantages for the study of complex systems. This instrument can acquire 241 fluorescence spectra, taken at 241 different exciting wavelengths in 16.7 milliseconds. These results are displayed in a three-dimensional graphical format in real time. Coupled with a laboratory computer, this instrument should provide new algorithms for the analysis of luminescence spectra of multicomponent systems.

Cuvette containing fluorescent sample

Dispersed exciting light provided by "polychromator" — Red, Orange, Yellow, Green, Blue, Indigo

Luminescence excited by red light
" " " orange "
" " " yellow "
" " " green "
" " " blue "
" " " indigo "

Excited beam to analyzing "polychromator"

FIG. 1

Geometry of excitation and emission beams.

EIGENVECTOR RESEARCH INCORPORATED
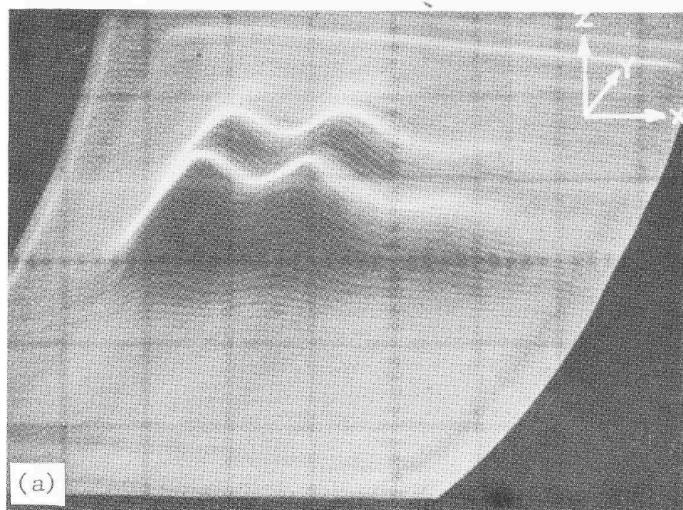
Fluorescence Landscapes

FIG. 3A

Three-dimensional display of the fluorescence intensity of perylene as a function of excitation and emission wavelengths.
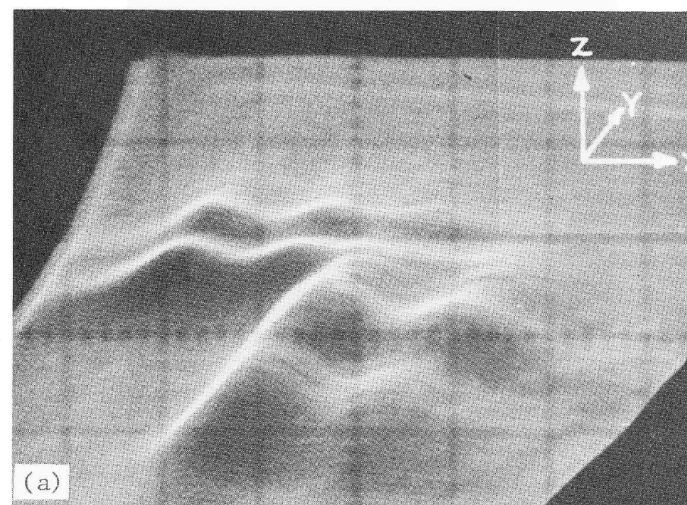

FIG. 4A

Three-dimensional display of the fluorescence intensity of a mixture of 2×10^{-6} M perylene and 2×10^{-5} M anthracene in ethanol.

EIGENVECTOR RESEARCH INCORPORATED

Warner et. al.

- First analysis of videofluorometer data
- Least squares fits
- Rotated eigenvectors of single EEM to non-negative values

Multicomponent Analysis in Clinical Chemistry by Use of Rapid Scanning Fluorescence Spectroscopy

Isiah M. Warner,[1] James B. Callis,[2] Ernest R. Davidson,[1] and Gary D. Christian[1]

To be useful in the clinical laboratory, multicomponent fluorescence analysis requires both the rapid measurement of the fluorescence intensity at a variety of excitation and emission wavelengths and the unambiguous reduction of the data by efficient algorithms. The Video Fluorometer, which exploits the multi-channel capability of a low-light-level television sensor to simultaneously acquire excitation and emission spectra, can meet the first requirement. For example, a complete set of emission and ...

multi-component analysis has invol... of differences in excitation and properties of the components (1, 2... of this approach is the increased an... for doing spectral scans with a conv... Other novel techniques involving resolution (4) and derivative spect... cently been developed.

Analysis of Multicomponent Fluorescence Data

I. M. Warner, G. D. Christian,* and E. R. Davidson

Department of Chemistry, University of Washington, Seattle, Wash. 98195

J. B. Callis

Department of Pathology, University of Washington, Seattle, Wash. 98195

A method for the analysis of a fluorescent sample containing multiple components is presented which utilizes the experimental "Emission–Excitation Matrix", M. The elements M_{ij} of this matrix represent the fluorescent intensity measured at wavelength λ_j for excitation at λ_i. The number of independently emitting components may be obtained by a determination of the number of non-zero eigenvalues of the matrices MM^T or M^TM. If only two components are present a range of possible excitation and emission spectra for each may be obtained through a linear transformation of the eigenvectors of MM^T and M^TM. Analysis of real and simulated multicomponent data using the above strategies illustrates the potential of the techniques. Also, examples are given of how noise affects the outcome and how spectral overlaps lead to partial ambiguities in the derived spectra.

emitting species, we sha... emission–excitation ma... proximation (9) by:

$$M_{ij} = 2.303 \Phi$$

where $I_0(\lambda_i)$ is the inten... light incident on the sam... sion $2.303\epsilon(\lambda_i)bc$ represe... and is the product of the... ficient $\epsilon(\lambda_i)$, the pathlen... emitting species c; Φ_f is... rescence; $\gamma(\lambda_j)$ is the frac... at wavelength λ_j; and $\kappa(\lambda$... dence of the sensitivity... geometrical factors, tra... grating quantum efficie...

EIGENVECTOR RESEARCH INCORPORATED

Ho, Christian and Davidson

(1) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1978**, *50*, 1108–1113.
(2) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1980**, *52*, 1071–1079.
(3) Ho, C-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1981**, *53*, 92–98.

Series of papers from 1978 to 1981, mostly in *Analytical Chemistry*

Application of the Method of Rank Annihilation to Quantitative Analyses of Multicomponent Fluorescence Data from the Video Fluorometer

C.-N. Ho, G. D. Christian, and E. R. Davidson*

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

A new scheme for quantitative analysis of a multicomponent fluorescent mixture using the Excitation-Emission Matrix (EEM) acquired by the video fluorometer has been developed and tested. This method, rank annihilation, is capable of quantifying a particular component known to be present in the mixture without having also to know the identity of the rest of the components. The theory of the method and the results of its application to a mixture matrix are presented and comparison is made with the method of least squares.

summing 512 video frames of the fluorescence, followed by subtraction of the same number of frames of dark current. A scattered light EEM for the pure solvent was similarly acquired. This scattered light EEM was subtracted from each EEM prior to mathematical analysis.

Data Analysis. The data produced by the video fluorometer were written onto floppy disks of the mini-computer system. The data were then transmitted to the CDC-6400 and mathematical reduction performed by a set of FORTRAN-IV programs.

Reagents. The chemicals used in this study were zone refined perylene and anthracene (both from James Hinton, Valparaise

$$E = M - \beta N \qquad (16)$$

where N is the standard matrix of some test compound which may or may not be present. Then minimization of $\mathrm{Tr}\ E^T E$

Now if Equations 21 and 22 are satisfied, then the matrix E defined by Equation 16 will have rank $r - 1$ when evaluated at $\beta = c/c^0$, where c is the concentration of the test compound in the mixture. Further, the matrix $E^T E$ or $E E^T$ with ei-
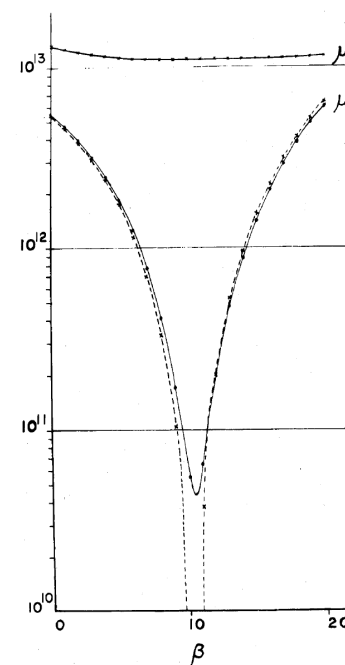

Figure 1. First and second eigenvalues of e as a function of β. Solutions and standards are the same as the data in Table II for a solution of relative concentration 10.0. The solid line is for a 2 × 2 matrix e given by Equation 27 and the dotted line is for a similar calculation using normalized a', b' coefficients

EIGENVECTOR RESEARCH INCORPORATED

Ho et. al. Cont.

The usefulness of the method of rank annihilation for quantitative analyses of multicomponent fluorescence data as acquired in the form of an excitation–emission matrix (EEM) by the videofluorometer is exemplified using ten different samples of a six-component polynuclear aromatic hydrocarbon solution whose constituents have a wide range of fluorescence quantum efficiencies and spectral overlaps. The optimal number of basis vectors to adequately represent the mixture EEMs, the best eigenvalue to minimize, and other criteria for increasing the reliability of the calculated results are explored.

EXPERIMENTAL

Reagents. Perylene, fluoranthene, dimethylanthracene, chrysene, and anthracene were zone refined (James Hinton, Valparaiso, Fla.). These and tetracene (Eastman Kodak) were dissolved in spectrograde cyclohexane (MCB).

Data Collection and Handling. Each EEM obtained by the videofluorometer consisted of 512 frames of fluorescence video signals followed by 512 frames of dark current subtraction actuated by closing a software controlled shutter placed in front of the field lens at the entrance slit of the emission monochromator. The scattered light EEM was similarly acquired using a solvent blank.

A procedure based upon the Fletcher–Powell algorithm for minimization of functions of more than one variable has been incorporated into the method of rank annihilation to allow simultaneous calculation of the concentrations of several known species from appropriate multicomponent data. This new method, called simultaneous multicomponent rank annihilation, provides the flexibility of analyzing for some constituents without requiring complete knowledge of the qualitative composition of the sample. The method was applied to a set of six-component polynuclear aromatic hydrocarbon solutions by use of data acquired by the video fluorometer in the form of an excitation–emission matrix.

efficient and yields satisfactory results.

THEORY

The theory of the method of rank annihilation for the determination of one known species present in a mixture of more than one component was presented in previous work (11, 12). To extend the method for simultaneous multicomponent analysis using the Fletcher–Powell algorithm (14) requires only minor modification of the previous algorithm.

For a single component when absorbances at all wavelengths are kept low the EEM is ideally of the form

$$M_{ij} = \alpha x_i y_j \tag{1}$$

Increasing levels of sophistication in the analysis, ending with using optimization methods to find the where rank is minimized for multiple components

The data generated by the videofluorometer were written and stored on floppy disk (Data System Design 210) by the minicomputer (PDP-1104 DEC). These data were then transmitted via telephone to the campus CDC-6400 where they were analyzed by a set of FORTRAN IV programs.

Appellof and Davidson

Strategies for Analyzing Data from Video Fluorometric Monitoring of Liquid Chromatographic Effluents

C. J. Appellof and E. R. Davidson*

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

A method for qualitative analysis of a multicomponent fluorescent mixture is developed. This method analyzes a three-dimensional matrix obtained by passing the effluent of a high-performance liquid chromatograph (HPLC) through the video fluorometer and provides an estimate of the number of components in the mixture as well as the emission and excitation spectra and HPLC profile for each component. The theory of this method, and its application to several sets of synthetic and real data, is presented.

Fluorescence emission spectrometry is a technique for quantitative analysis of multicomponent mixtures of aromatic hydrocarbons. A common problem with this type of analysis is spectral overlap. That is, the fluorescence emission spectra of two components in a solution may be very similar, leading to difficulties in quantifying these components. Warner et al. (1) showed that by measuring a two-dimensional emission–excitation matrix (EEM) some of this difficulty could be overcome. There still may exist high spectral overlaps, even with this method of measurement, so Johnson et al. (2) have proposed adding a third dimension (e.g., pH) to the data matrix to provide more separation between the different components. Recently, Hershberger et al. (3) developed a flow-cell curvette which can be attached to a high-performance liquid chromatograph (HPLC) and can be placed in the sample chamber of a video fluorometer (4, 5). This allows the

$$\mathbf{M}_{ijk} = \sum_{s=1}^{N} c_s x_{is} y_{js} z_{ks} \qquad (3)$$

The analysis consists of determining N and c, x, y, and z for all N components. For convenience, c is absorbed into x, y,

These can be expressed as three matrix equations

$$\mathbf{X}\mathbf{Q}(yz) = \mathbf{P}(yz) \qquad (9)$$
$$\mathbf{Y}\mathbf{Q}(xz) = \mathbf{P}(xz) \qquad (10)$$
$$\mathbf{Z}\mathbf{Q}(xy) = \mathbf{P}(xy) \qquad (11)$$

where $\mathbf{Q}(yz)$ corresponds to a matrix whose stth element is the quantity in brackets in eq 6, and the itth element of $\mathbf{P}(yz)$ is the right hand side of (6). The other $\mathbf{Q}$ and $\mathbf{P}$ matrices are similarly defined by eq 7 and 8.

Following Carrol and Chang (8), we use Wold's procedure (9) of nonlinear iterative least squares (NILES) to solve this system of equations. A guess is made for the number of components N in the solution. Then a guess is made for the $\mathbf{Y}$ and $\mathbf{Z}$ matrices, eq 9 is solved for $\mathbf{X}$.

$$\mathbf{X} = \mathbf{P}(yz)\mathbf{Q}^{-1}(yz) \qquad (12)$$

giving a least-squares estimate of $\mathbf{X}$ based on the current estimates for $\mathbf{Y}$ and $\mathbf{Z}$. Then the new $\mathbf{X}$ and the estimate of $\mathbf{Z}$ are used to solve eq 10 for a new estimate of $\mathbf{Y}$. Similarly, eq 11 is solved for $\mathbf{Z}$. This procedure is repeated until the spectral matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ converge.

Referenced Carrol & Chang, re-invented Harshman's PARAFAC!

EIGENVECTOR RESEARCH INCORPORATED

De Ligney et. al.

THREE-MODE FACTOR ANALYSIS OF DATA ON RETENTION IN NORMAL-PHASE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

C. L. DE LIGNY* and M. C. SPANJER
Laboratory for Analytical Chemistry, University of Utrecht, Croesestraat 77A, 3522 AD Utrecht (The Netherlands)
and
J. C. VAN HOUWELINGEN and H. M. WEESIE
Institute for Mathematical Statistics, University of Utrecht, Budapestlaan 6, 3584 CD Utrecht (The Netherlands)
(First received December 23rd, 1983; revised manuscript received May 15th, 1984)

SUMMARY

It is shown that the Snyder equation is not quite satisfactory for fitting retention data in normal-phase high-performance liquid chromatography (HPLC) on chemically bonded phases. This equation is a special case of the mathematical–statistical three-mode factor analysis model. This model, in its general form, has been used to fit two sets of literature data on the retention in normal-phase HPLC for 19 solutes on six adsorbents with two eluents, and for 39 solutes on three adsorbents with two eluents, respectively. This study represents the first application of three-mode factor analysis with missing data, and also the first application of three-mode factor analysis in the field of the natural sciences. The accuracy of the fit of the observations and of the prediction of the missing data, for various numbers of extracted factors, is discussed.

Used a Tucker model, referenced Kroonenberg, didn't reference Appellof!

EIGENVECTOR RESEARCH INCORPORATED

Avi Lorber

Short Communication

Quantifying chemical composition from two-dimensional data arrays

Avraham Lorber

⊞ Show more

Get rights and content

Abstract

Many analytical instruments (e.g., videofluorimeter and liquid chromatographs with spectrometric detectors) can produce two-dimensional data arrays. Rank-annihilation factor analysis enables chemical composition to be quantified from such an array. Solution by this method requires a search for a minimum value of an eigenvector which cannot be calculated even iteratively. The problem is presented here as a generalized eigenvalue problem and a direct solution is found by using singular value decomposition.

- Showed RAFA could be cast as a generalized eigenvector problem and solved non-iteratively
- Solution required separate samples for each standard

Sanchez and Kowalski

Generalized Rank Annihilation Factor Analysis

Sir: The analytical chemist is frequently confronted with the problem of analyzing complex mixtures of which only concentrations of a few components are of interest. In these cases, it is desirable to be able to obtain quantitative information for the analytes of interest without concern for the rest of the components in the sample. Second-order bilinear sensors, i.e., sensors that yield a two-dimensional data matrix

is found by using the singular value decomposition. With his method, to obtain the concentrations of the p analytes of interest in the sample, its bilinear spectrum and the p calibration spectra for each pure analyte must be recorded to obtain the concentrations. Analysis for each analyte requires a separate calculation. This letter presents the generalized rank annihilation method, of which Lorber's noniterative

A solution to this problem can be obtained by using the principal components of the sum of the matrices $\mathbf{M}$ and $\mathbf{N}$, defining $\mathbf{W} \equiv \mathbf{M} + \mathbf{N}$

$$\mathbf{W} = \bar{\mathbf{U}}_W \tilde{\mathbf{S}}_W \bar{\mathbf{V}}_W{}^T \tag{34}$$

$$(\bar{\mathbf{U}}_W{}^T \mathbf{M} \bar{\mathbf{V}}_W \tilde{\mathbf{S}}_W{}^{-1}) \mathbf{Z}_W{}^* = \mathbf{Z}_W{}^* \lambda_W \tag{35}$$

$$\mathbf{Y}^T = (\bar{\mathbf{V}}_W \tilde{\mathbf{S}}_W{}^{-1} \mathbf{Z}_W{}^*)^+ \tag{36}$$

$$\mathbf{X}\beta = \bar{\mathbf{U}}_W \mathbf{Z}_W{}^* \tag{37}$$

The eigenvalues λ_k are the ratio of concentrations $\beta_k/(\xi_k + \beta_k)$. For all the components present in both mixtures, the concentration in the unknown is $\beta_k = \lambda_k \xi_k/(1 - \lambda_k)$. When one component is not present in the calibration sample, $\xi_k = 0$ and $\lambda_k = 1$.

The solution presented for this case can be applied to all the previous cases, and no testing with target factor analysis is necessary. An artifical matrix $\mathbf{W}$ is generated to perform

A limitation of rank annihilation as originally formulated is that an iterative solution requiring many matrix diagonalizations is necessary (*1*). Lorber (*6*) has reported a non-iterative solution presenting the problem as a generalized eigenvalue–eigenvector equation for which a direct solution is found by using the singular value decomposition. With his

Did not reference Appellof and Davidson! Kowalski didn't like PARAFAC!

GRAM

CHROMSYMP. 982

GENERALIZED RANK ANNIHILATION METHOD

I. APPLICATION TO LIQUID CHROMATOGRAPHY–DIODE ARRAY ULTRAVIOLET DETECTION DATA

EUGENIO SANCHEZ*, L. SCOTT RAMOS and BRUCE R. KOWALSKI

Laboratory for Chemometrics, Department of Chemistry BG-10, University of Washington, Seattle, WA 98195 (U.S.A.)

For quantitative analysis, one sample is defined as a calibration mixture. From the two data matrices corresponding to the two samples, GRAM will generate the following information about each shared component, *i.e.*, compounds present in both samples: an extracted spectrum, a resolved chromatogram, and the relative concentration (unknown:calibration).

- Problem: Generalized eigenvector problem results in imaginary solutions.
- Many follow on papers to address this

Trilinear Decomposition (TLD)

Tensorial resolution: A direct trilinear decomposition

Eugenio Sanchez, Bruce R. Kowalski

PDF   TOOLS   SHARE

- Based on GRAM
- Problem is to find two representative samples containing all components in different ratios

Abstract

Modern instrumentation in chemistry routinely generates two-dimensional (second-order) arrays of data. Considering that most analyses need to compare several samples, the analyst ends up with a three-dimensional (third-order) array which is difficult to visualize or interpret with the conventional statistical tools.

Some of these data arrays follow the so-called *trilinear* model,

$$R_{ijk} = \sum_{r=1}^{N} X_{ir} Y_{jr} Z_{kr} + \text{Error}_{ijk}$$

EIGENVECTOR RESEARCH INCORPORATED

Restricted Tucker Models

THEORY OF MEDIUM-RANK SECOND-ORDER CALIBRATION WITH RESTRICTED-TUCKER MODELS

AGE K. SMILDE,* YONGDONG WANG† AND BRUCE R. KOWALSKI‡
Laboratory for Chemometrics, Department of Chemistry, BG-10, University of Washington, Seattle, WA 98195, U.S.A.

SUMMARY

If an analytical instrument or instrumental method gives a response matrix when analyzing a pure analyte, the instrument or instrumental method is called a second-order method. Second-order methods that generate a response matrix for a pure analyte of rank one are called rank-one second-order methods. If the response matrix of a pure analyte is not rank one, essentially two cases exist: medium rank (between two and five) and high rank (greater than five). Subsequently, medium- and high-rank second-order calibration tries to use medium- and high-rank second-order methods to analyze for analytes of interest in a mixture. A particular advantage of second-order methods is the ability to analyze for analytes of interest in a mixture which contains unknown interferences. Keeping this advantage is the challenge on moving away from rank-one second-order calibration methods. In this paper a medium-rank second-order calibration method is proposed based on least-squares restricted Tucker models. With this method the second-order advantage is retained.

Multi-Linear PLS: N-PLS

Multiway calibration. Multilinear PLS
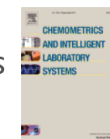
Rasmus Bro

PDF   TOOLS   SHARE

Abstract

A new multiway regression method called *N*-way partial least squares (N-PLS) is presented. The emphasis is on the three-way PLS version (tri-PLS), but it is shown how to extend the algorithm to higher orders. The developed algorithm is superior to unfolding methods, primarily owing to a stabilization of the decomposition. This stabilization potentially gives increased interpretability and better predictions. The algorithm is fast compared with e.g. PARAFAC, because it consists of solving eigenvalue problems.

EIGENVECTOR
RESEARCH INCORPORATED

PARAFAC Tutorial

Tutorial

PARAFAC. Tutorial and applications

Rasmus Bro ✉

⊞ Show more

Get rights and content

Abstract

This paper explains the multi-way decomposition method PARAFAC and its use in chemometrics. PARAFAC is a generalization of PCA to higher order arrays, but some of the characteristics of the method are quite different from the ordinary two-way case. There is no rotation problem in PARAFAC, and e.g., pure spectra can be recovered from multi-way spectral data. One cannot as in PCA estimate components

EIGENVECTOR RESEARCH INCORPORATED

Bro Thesis, 1998
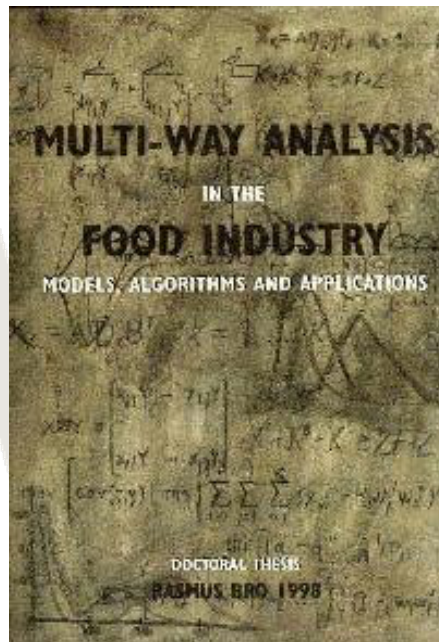
Application to Batch Data

Monitoring batch processes using multiway principal component analysis

Paul Nomikos, John F. MacGregor

Tucker-1, aka unfold data and do PCA

TLD and PARAFAC on Process Data

Research Article

A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process

Barry M. Wise ✉, Neal B. Gallagher, Stephanie Watts Butler, Daniel D. White Jr, Gabriel G. Barna

    🔴 PDF   🔧 TOOLS   < SHARE

Abstract

Multivariate statistical process control (MSPC) tools have been developed for monitoring a Lam 9600 TCP metal etcher at Texas Instruments. These tools are used to determine if the etch process is operating normally or if a system fault has occurred. Application of these methods is complicated because the etch process data exhibit a large amount of

Research Article

Application of PARAFAC2 to fault detection and diagnosis in semiconductor etch

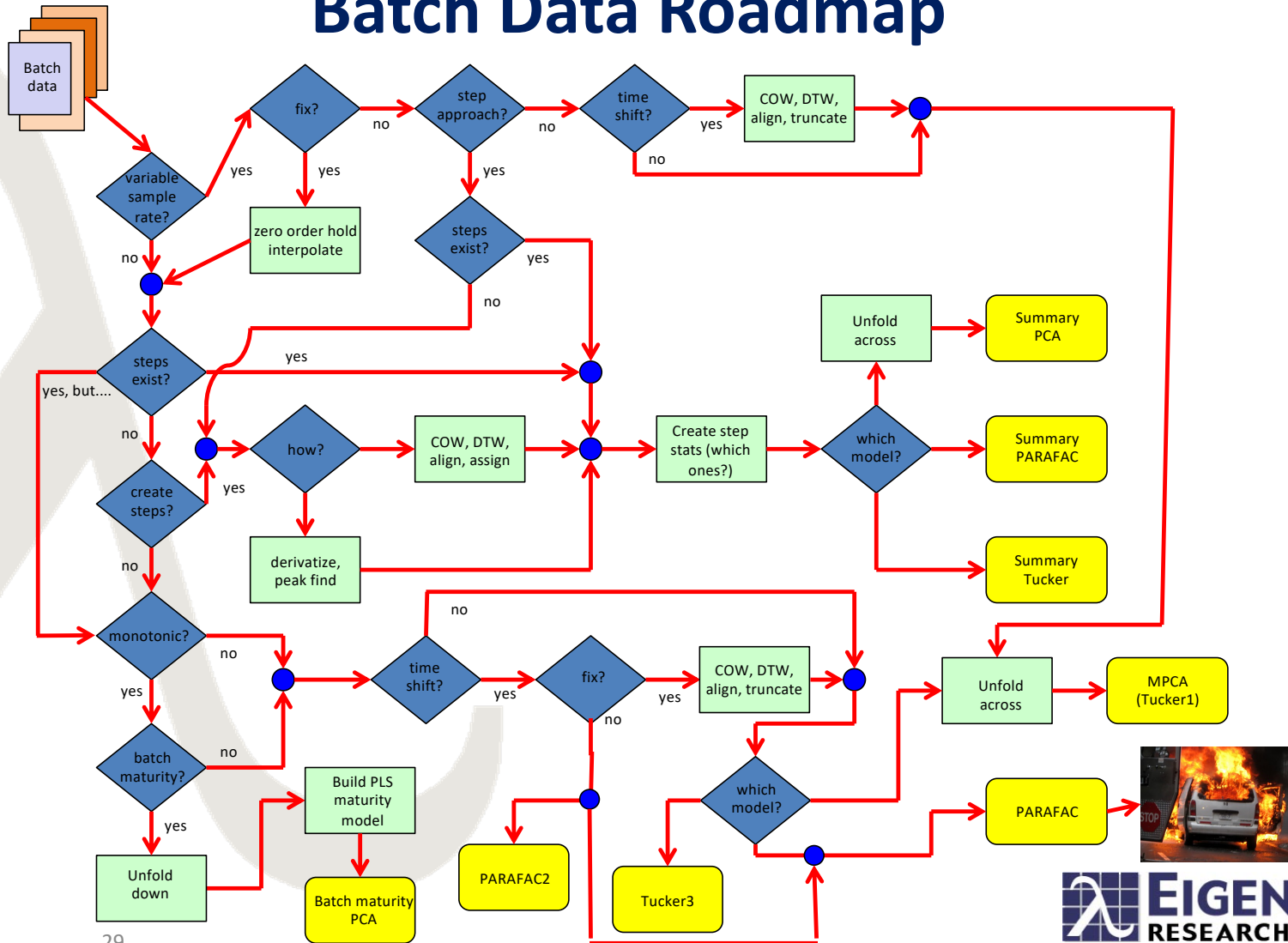Barry M. Wise ✉, Neal B. Gallagher, Elaine B. Martin

    🔴 PDF   🔧 TOOLS   < SHARE

Abstract

Monitoring and fault detection of batch chemical processes are complicated by stretching of the time axis, resulting in batches of different length. This paper offers an approach to the unequal time axis problem using the parallel factor analysis 2 (PARAFAC2) model. Unlike PARAFAC, the PARAFAC2 model does not assume parallel proportional profiles, but only that the matrix of profiles preserves its 'inner product structure' from sample to sample. PARAFAC2 also allows each matrix in the multiway array to have a different number of rows. It has previously been demonstrated how the PARAFAC2 model can be

EIGENVECTOR
RESEARCH INCORPORATED

Batch Data Roadmap

Batch data

variable sample rate? — no → steps exist? — yes, but.... / yes / no

fix? — yes → zero order hold interpolate
fix? — no → step approach?

step approach? — no → time shift?
step approach? — yes → steps exist?

time shift? — yes → COW, DTW, align, truncate
time shift? — no

steps exist? — yes
steps exist? — no

create steps? — yes → how?
create steps? — no

how? → COW, DTW, align, assign
how? → derivatize, peak find

Create step stats (which ones?) → which model?

which model? → Summary PCA (via Unfold across)
which model? → Summary PARAFAC
which model? → Summary Tucker

monotonic? — no → time shift?
monotonic? — yes → batch maturity?

batch maturity? — no
batch maturity? — yes → Unfold down → Build PLS maturity model → Batch maturity PCA

time shift? — no
time shift? — yes → fix?

fix? — yes → COW, DTW, align, truncate
fix? — no → PARAFAC2

which model? → Tucker3
which model? → Unfold across → MPCA (Tucker1)
which model? → PARAFAC

EIGENVECTOR RESEARCH INCORPORATED

Software

3WayPack

PLS_Toolbox 2.0

N-way Toolbox

All available in ~1998

TRICAP'97

- Three-way Methods in Chemistry and Psychology
- May 4-9, 1997
- Manson, WA
- http://web.archive.org/web/20010223003407/http://www.eigenvector.com/Old/TRICAP_schedule.html

Where are we now?

- Multi-way analysis in chemistry is more common but still not yet routine
  - Compare to PLS which is ubiquitous
- Mostly in fluorescence
  - PARAFAC in water quality analysis
- True multi-way models still uncommon in process analysis
  - Some use of the unfold approaches
- University of Washington was the nexus of multi-way in chemistry!

EIGENVECTOR
RESEARCH INCORPORATED