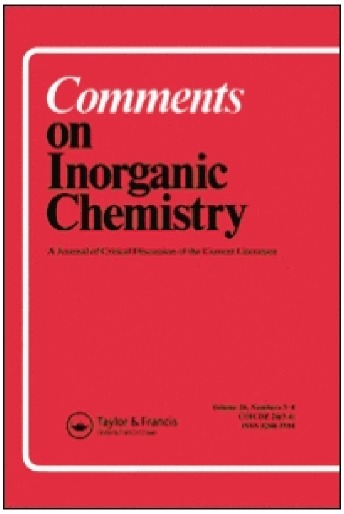
This article was downloaded by:
On: *15 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Comments on Inorganic Chemistry

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455155

Evolving Factor Analysis

Harald Gampp[a]; Marcel Maeder[a]; Charles J. Meyer[a]; Andreas D. Zuberbuehler[a]
[a] Institut für Anorganische Chemie der Universität, Basel, Switzerland

To cite this Article Gampp, Harald , Maeder, Marcel , Meyer, Charles J. and Zuberbuehler, Andreas D.(1987) 'Evolving Factor Analysis', Comments on Inorganic Chemistry, 6: 1, 41 — 60

To link to this Article: DOI: 10.1080/02603598708081852

URL: http://dx.doi.org/10.1080/02603598708081852

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Evolving Factor Analysis

Modern instrumentation allows the easy accumulation of high-quality bilinear data matrices. The new algorithm of Evolving Factor Analysis makes full use of all information contained directly in the data. Essentially perfect deconvolution, yielding the concentrations and absorption spectra of all species, becomes possible on a completely model-free basis.

INTRODUCTION

Equilibrium studies of spectroscopic titrations and peak resolution in chromatography are two typical problems which are appropriately tackled by using multiwavelength (more generally, multichannel) detection systems, i.e., by collecting an $M \times W$ data matrix $\mathbf{Y}$ from M measurements at W wavelengths (channels). In all such experiments it is the primary goal of data reduction to decompose the original matrix $\mathbf{Y}$ into two smaller matrices containing the concentration profiles $\mathbf{C}$ and the absorptivities $\mathbf{A}$ of the individual species. Provided there is linear response of the measuring equipment to the individual concentrations (validity of Beer's law), this task can be described by the simple matrix equation (1):

$$\mathbf{Y} = \mathbf{CA} \tag{1}$$

During the last decade, progress in general instrumentation, and especially microprocessor technology, has led to various rapid data collecting systems and also to a remarkable improvement of data quality,[1,2] together with the development of further hyphenated

Comments Inorg. Chem.
1987, Vol. 6, No. 1, pp. 41–60
0260-3594/87/0601-0041/$20.00/0

techniques such as GC-MS/IR.[3] This progress has been paralleled by an equal upsurge in practical applications and interest in more efficient algorithms for numerical treatment of the data: Chemometrics is a rapidly expanding part of analytical chemistry and several efficient programs are now available for the analysis of complicated equilibrium systems based on spectroscopic data.[4,5]

Unfortunately, decomposition of **Y** according to Eq. (1) is not unique. In the case of equilibrium studies the analysis is normally based on a chemical model, i.e., on assumptions with regard to the number and identity of the chemical species and the law of mass action. Selection of the correct chemical model is by no means trivial in many cases, however, and the law of mass action cannot be applied to problems such as peak resolution in chromatography or the analysis of equilibrium systems based on measurements with strongly variable ionic strength and/or solvent composition. Therefore, a completely model-free approach to the problem described by Eq. (1) would be most welcome. No matter if a reasonable model might theoretically be available, a completely model-free approach has its intrinsic advantages: "In the former case you get back what you assumed, in the latter you get what you get."[6]

It is the purpose of this Comment to show that the newly developed algorithm of Evolving Factor Analysis (EFA)[7-10] indeed yields a successful, quantitative and model-free deconvolution of the original data into concentration profiles and spectra of the individual absorbing species. EFA is not the first approach in this direction. Principal component analysis (PCA) or singular value decomposition accomplishes a related task.[11] The resulting orthogonal matrices are, however, completely abstract, containing negative "concentrations" and "absorptivities" with no physical interpretation. Numerous attempts have been described recently to transform the abstract PCA result into the chemically correct picture. We propose to show that none of them has made full use of the information contained directly in the original data and therefore could not generally be expected to give a picture as accurate as that of EFA.

The following points will be discussed subsequently: (i) the mathematical concept behind EFA; (ii) the scope of other model-free attempts to decompose **Y** according to Eq. (1), as compared to EFA; (iii) model-free calculation of concentration profiles, absorption spectra and equilibrium constants based on spectroscopic titrations; (iv) peak resolution in chromatography; (v) analysis of equilibrium mixtures which cannot be treated by the law of mass action; (vi) detection of impurities; (vii) semiquantitative analysis of complicated systems.

THE CONCEPT OF EVOLVING FACTOR ANALYSIS

For a system of S absorbing species, deconvolution of the original data matrix $\mathbf{Y}$ according to Eq. (1) gives two smaller matrices $\mathbf{C}$ ($M \times S$) of the concentrations and $\mathbf{A}$ ($S \times W$) of the molar absorptivities. Similarly, $\mathbf{Y}$ can be decomposed into the product of three matrices $\mathbf{U}$ ($M \times S$), $\mathbf{V}$ ($S \times W$), and $\mathbf{S}$ ($S \times S$) by singular value decomposition, Eq. (2)[12]:

$$\mathbf{Y} = \mathbf{CA} = \mathbf{USV} \tag{2}$$

$\mathbf{U}$ is formed by the significant eigenvectors of $\mathbf{YY}'$ and $\mathbf{V}$ by those of $\mathbf{Y}'\mathbf{Y}$ ($\mathbf{Y}'$ is the transpose of $\mathbf{Y}$). The columns of $\mathbf{U}$ as well as the rows of $\mathbf{V}$ are orthonormal, $\mathbf{U}'\mathbf{U} = \mathbf{VV}' = \mathbf{I}$ (identity matrix). $\mathbf{S}$ is a diagonal matrix, its nonzero elements are the square roots of the significant eigenvalues of $\mathbf{Y}'\mathbf{Y}$ in descending order.

EFA consists of two more or less independent parts. The basic idea behind the first of them (primary EFA)[7] is to follow the change of rank of $\mathbf{Y}$ and the concomitant evolution of eigenvalues of $\mathbf{Y}'\mathbf{Y}$ with progressing titration. This is done by rank analysis of the submatrices $\mathbf{Y}_f$ formed by the first f spectra (rows) of $\mathbf{Y}$ (forward EFA). The appearance of an additional component j in the course of the progressing titration will be indicated by the rise of the rank of $\mathbf{Y}_f$ from $j - 1$ to j, i.e., a significant increase of the concomitant jth eigenvalue. For this statement to be true, neither the concentration profile nor the absorption spectrum of component j may be a linear combination of the previous species, and both the concentrations and the absorptions must be sufficiently different from zero. Rank estimation of data matrices critically depends on the experimental noise level and several methods of rank estimation have been published,[13-15] but this problem is not discussed here in detail.

Using primary EFA, instructive pictures may be obtained by plotting the eigenvalues Λ_f of $\mathbf{Y}_f'\mathbf{Y}_f$ in the function of the progressing experiment, e.g., the pH or the elution time. This is shown in Fig. 1 for model data for three strongly overlapping chromatographic peaks. Because of the numerical disparity of successive eigenvalues, the logarithms of Λ_f are used for graphical representation in Fig. 1(a).

The rank analysis can be repeated in the backward direction with the submatrices $\mathbf{Y}_b$ formed by the last b spectra of $\mathbf{Y}$. The resulting backward EFA plot is given in Fig. 1(c) and shows the disappearance of components with the elution time. Proper combination of the appearance of a given component with its disappearance (the results of the forward and backward EFA plots) yields the ranges of existence of all species or their "concentration windows" as indicated in Fig. 1(b). Combination of the corresponding eigenvalues finally gives a semiquantitative description of the concentration profiles of the absorbing species.

The second part of EFA performs a quantitative calculation of the concentration matrix **C** and of the absorption spectra **A** by a variant of target factor analysis (iterative EFA).[8-10] **C** and **A** are refined successively by linear least-squares methods, considering only those species which are acceptable according to the "concentration windows" of primary EFA and, in addition, setting all negative concentrations to zero. No further assumptions are made with respect to either concentration profiles or absorption spectra. Additional details making primary as well as iterative EFA fast and reliable algorithms are described elsewhere.[9]

ALTERNATIVE MODEL-FREE APPROACHES

Most of the methods for the model-free estimation of spectra and concentration profiles are based on the pioneering work of Lawton and Sylvestre.[16] Their self-modeling curve resolution is based only on the non-negativity of the absorption spectra; its main disadvantage is the intrinsic limitation to two-component systems. Self-modeling curve resolution has been successfully applied in different fields, such as chromatography[17-19] and x-ray photoelectron spectroscopy.[20]

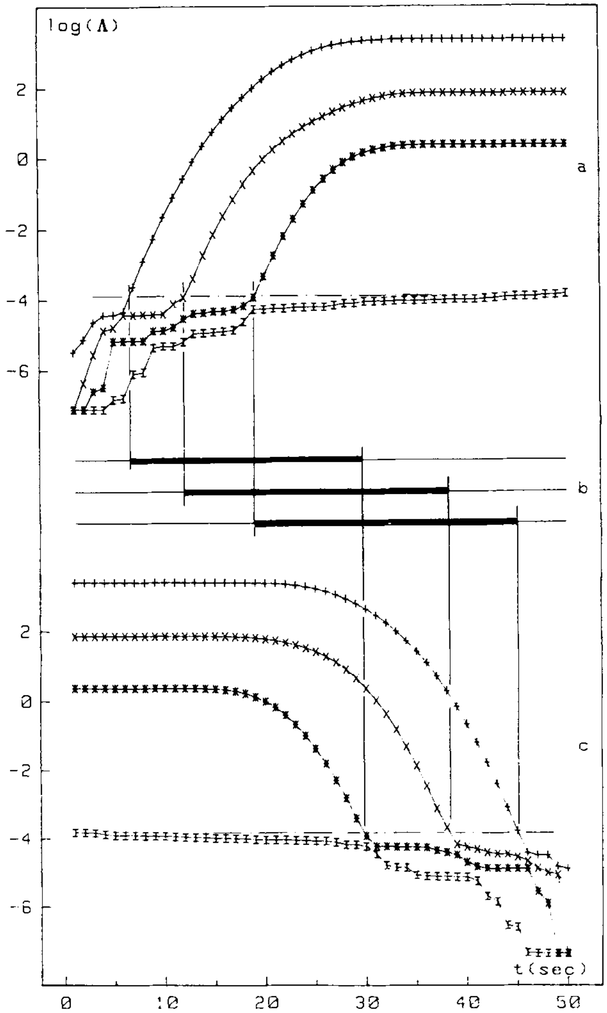


FIGURE 1 EFA results for model data for three strongly overlapping chromatographic peaks. Logarithms of the eigenvalues Λ_f ((a) forward EFA) and Λ_b ((c) backward EFA) together with the resulting time windows (b) defined by the significant increase of the pertinent eigenvalues.

Extension of the method to three or more components has been described subsequently.[21] Different problems have to be overcome in these cases: graphical display of the polydimensional space spanned by the eigenvectors **U** or **V** (cf. Eq. (2)) is not trivial. One possibility is to switch to polar coordinates,[22,23] obscuring the information available with regard to the concentrations, or projection onto hyperplanes.[24,25] In addition, the results obtained for such multicomponent systems are usually much less well-defined, as long as the acceptable range of solutions is defined only by the non-negativity of the physical parameters (concentrations and absorptivities).[25] Several further restrictions, such as maximum dissimilarity[24] or minimum envelope of all spectra or concentration profiles,[23] have been suggested to obtain unique solutions, but obviously such *ad hoc* assumptions are not in general backed by physical reality.

Recently, several methods for the iterative refinement of initial estimates of either absorption spectra or concentration profiles have been developed for the model-free deconvolution.[26–29] For these methods there is no intrinsic limit to the maximum number of components. Most of these algorithms are based on iterative target factor analysis, ITFA.[26,27,29] The strategy of ITFA can be divided into two parts. First, a set of approximations for the concentration profiles has to be selected. These are iteratively refined in the second step by repeated target transformation:

$$\mathbf{C}_{\text{new}} = \mathbf{U}\mathbf{U}^{t}\mathbf{C}_{\text{old}} \tag{3}$$

Different strategies have been designed to modify the resultant concentration profiles $\mathbf{C}_{\text{new}}$ (this is necessary to keep the iterative process going). All such methods are setting negative values in **C** to zero in each cycle. In Ref. 26 this is the only correction. The result is a poorly controlled spreading out of the concentration profiles as seen in Fig. 2 (dotted lines) for model data used previously in our testing of EFA.[9] A method to prevent this spreading has been proposed by setting "small" positive values in the concentration matrix **C** to zero as well.[27] This algorithm has been successfully applied to resolve chromatographic peaks with up to six overlapping components. It is obviously not trivial to choose the appropriate level of "smallness" and no general procedure has been proposed. Using a positive cutoff level certainly helps ensure against undue spreading. However, the choice of this level seems

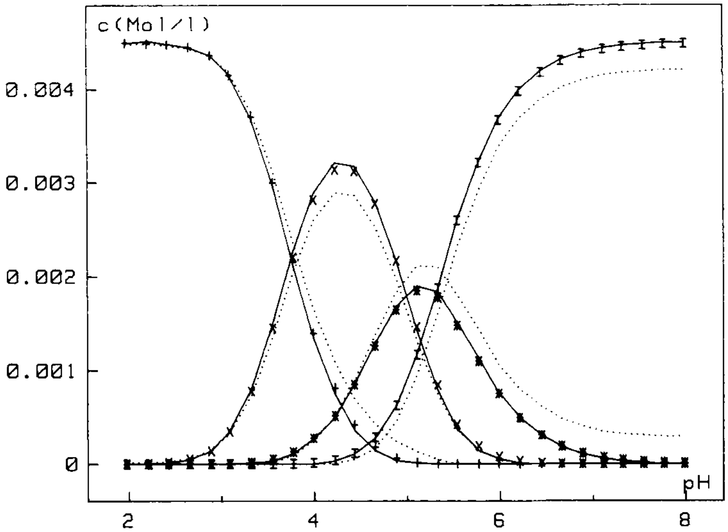


FIGURE 2 Concentration profiles of model data using c_{Cu} = 0.0045 M, log β_{110} = 10.00, log β_{11-1} = 5.00, and log β_{11-2} = −0.32. Results of iterative EFA (———) and iterative target transformation (Ref. 26) (· · ·) compared with the original data (+: Cu^{2+}, ×: CuL^{2+}, *: $CuLH^{+}_{-1}$, I: $CuLH_{-2}$).

to be a matter of trial and error, and the algorithm is still not inherently fail-proof with respect to spreading out or to contraction. An analogous approach is given in Ref. 29, where concentrations with a relatively high level of statistical uncertainty are reset to zero. This problem is solved using EFA (solid lines in Fig. 2), where the ranges of existence for all species are uniquely defined. We therefore feel that the main difference between EFA and other new model-free approaches does not lie in any details of the iterative refinement, but in the availability of the "concentrations windows" through primary EFA.

CALCULATION OF CONCENTRATION PROFILES, SPECTRA AND EQUILIBRIUM CONSTANTS

As in other equilibrium studies, data reduction of spectrophotometric titrations normally consists of a least-squares refinement of equilibrium constants based on a chemical model.[5] Concentration profiles are calculated using the selected model estimates of the corresponding equilibrium constants, and the analytical concentrations of the components; they are not in any way directly extracted from the experimental data. The molar absorptivities can even be eliminated from the algorithm and are calculated noniteratively at the end of the refinement.[30,31] Significant progress has been made in this direction during the last decade and powerful programs are available to deal with spectrophotometric data almost as routinely as with potentiometric ones, even in complicated systems.[5,32,33] This is not the place to discuss the relative merits of potentiometric and spectrophotometric titrations in detail, but the present state of the art seems to indicate that spectrophotometric titrations now match their potentiometric counterpart in reliability and accuracy,[34] while being generally superior in their power to discriminate between different chemical models.[35–37]

Evolving Factor Analysis opens a radically different approach to the calculation of concentration profiles, absorption spectra, and equilibrium constants.[7–9] Here, the primary step is the resolution of the measured absorbances **Y** into the concentration profiles **C** and the molar absorptivities **A** according to Eq. (1). Least-squares refinement is done directly on **C** and **A**, minimizing the squares of the residuals **R**, Eq. (4):

$$\mathbf{R} = \mathbf{Y} - \mathbf{CA}; \qquad SQ \overset{\text{min!}}{=} \sum_{i=1}^{M} \sum_{j=1}^{W} R(i, j) \tag{4}$$

Only after completion of the numerical treatment are chemical species associated with the concentration profiles, based on the corresponding absorption spectra and/or by chemical reasoning. Equilibrium constants are obtained at the very end; they are read directly from the respective intersections of the concentration profiles after assigning stoichiometric coefficients to the previously abstract species. Thus, using EFA, the normal order of dealing with spectroscopic titrations is reversed: First the data are deconvoluted according to Eq. (1), then a model is selected and the law of mass action is introduced last.

This application of EFA has been extensively tested using spectrophotometric, ESR, and semi-synthetic model data. For details we refer to the original literature.[7–9] In general, the results were very satisfactory, indeed, even for four- and five-component systems. In most cases, the standard deviation between concentrations calculated with EFA and the law of mass action were between 0.5 and 2%, absorption spectra were essentially identical and the final equilibrium constants were the same within experimental error. Differences between calculated and experimental data are even smaller using EFA rather than the law of mass action (LMA). This is very satisfactory, but not really surprising for a successful analysis, since EFA, of course, is subject to less stringent restrictions and thus may much more easily adapt to experimental errors.

The power of EFA is demonstrated for one typical example in Fig. 3. Data were generated based on a system of four strongly overlapping species with the spectral characteristics of Cu^{2+} and the complexes CuL^{2+}, $CuLH^{+}_{-1}$ and $CuLH_{-2}$ with L = 3,7-diazanonane diamide, $H_2NCOCH_2NH(CH_2)_3NHCH_2CONH_2$. A random noise of 2×10^{-4} absorbance as well as 0.008 pH units, corresponding to our experimental setup, was superimposed.[9]

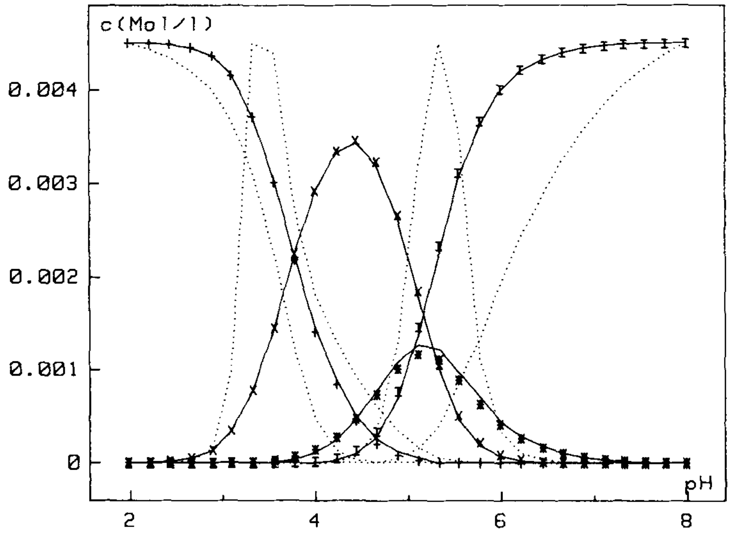


FIGURE 3 Concentration profiles of model data using c_{Cu} = 0.0045 M, log β_{110} = 10.00, log β_{11-1} = 4.70, and log β_{11-2} = −0.32. Results of primary EFA (···) and iterative EFA (———) compared with the analysis based on the law of mass action (+: Cu^{2+}, ×: CuL^{2+}, *: $CuLH^{+}_{-1}$, I: $CuLH_{-2}$).

Three concentration profiles are shown: (a) the results of primary EFA (dotted lines); (b) the EFA result after refinement (solid lines), and (c) the analysis based on the LMA (symbols). As is seen in Fig. 3 the two latter results are virtually identical, as are the absorption spectra (not shown). Not surprisingly, the correct equilibrium constants may be read off the EFA plots: log K^{Cu}_{CuL} = 9.99 (9.99),[9] log K^{H}_{CuL} = 5.29 (5.33),[9] log $K^{H}_{CuLH_{-1}}$ = 5.02 (5.01)[9] (results based on the LMA in brackets). EFA results were obtained by linear interpolation of the respective intersections near pH 3.8, 5.3, and 5.0, using log K^{H}_{LH} = 8.40 and log $K^{H}_{LH_2}$ = 6.55 for the protonation constants of the free ligand.[38]

Needless to say, such a close correspondence of the results of these two completely independent types of analysis is most rewarding and not only elucidates the power of EFA but also helps a lot in trusting the selected chemical model.

PEAK RESOLUTION IN CHROMATOGRAPHY

For equilibrium studies based on spectroscopic titrations, model-free data reduction normally is but an interesting alternative to numerical treatment based on the law of mass action (vide infra for exceptions). The situation is completely different in the analysis of chromatographic data. Clearly, the reason is that no satisfactory general model-based description is available. Approximation of elution profiles, e.g., by Gaussian or Lorentzian type functions or combinations thereof, is of course possible,[39,40] but always remains to a certain degree arbitrary and error prone. Therefore, a completely model-free approach would be most important.[6] As already discussed, numerous attempts have been made to solve this problem, but none has made use of the information available by means of the "concentration windows" obtained through primary EFA. In a recent paper EFA has been successfully applied to synthetic data mimicking HPLC chromatography.[10] In the case of three overlapping peaks with a resolution as small as 0.2 and rather unequal contributions to the total absorbance, this method accurately produces the number of components, their elution profiles and their absorption spectra. The data matrix (50 spectra at different elution times) was first subjected to primary EFA, the result of which is shown in Fig. 1. While the original data do not reveal any overlapping peaks, the presence of three species is easily detected by the evolution of three significant eigenvalues at three different elution times, and combination of forward with backward EFA yields a "time window" for each species. Based on these windows, concentration profiles and absorption spectra are obtained as shown in Fig. 4 through least-squares refinement. The strongly overlapping original data **Y** which are displayed in three dimensions (time, wavelength, absorbance) are deconvoluted into concentration profiles **C** (left) and spectra **A** (right) of the individual species and projected onto the rear "walls" of the imaginary cube containing the data. The residuals in absorbance have a standard deviation of 1.26×10^{-3} absorbance units, as compared to the random noise of 1.00×10^{-3} imposed on the model data. The standard deviation between theoretical and calculated concentra-

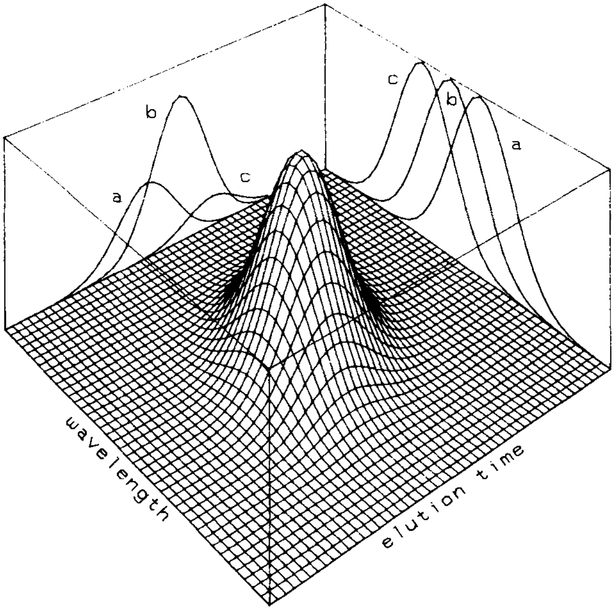


FIGURE 4 Model chromatogram with three components. Three-dimensional representation of the original data. The deconvolution into concentration profiles and absorption spectra by iterative EFA is shown on the rear walls.

tion profiles is 1.2%. EFA has also been applied to cases with four and five strongly overlapping species, with practically equal success.

We thus conclude that EFA is generally applicable to peak resolution in chromatography and that the challenge expressed recently by Delaney[6] is fully met: "The ultimate goal of curve resolution would be to be able to determine the number of components in an overlapping chromatographic peak as well as the spectrum and concentration profile of each compound without assumptions regarding peak shape, location or identity." There is one cautioning remark which has to be maintained with respect to this statement: The heights of the concentration profiles and of the corresponding absorption spectra are strictly coupled through an inverse proportionality relation. If neither is known, there will always necessarily remain this final ambiguity, no matter which algorithm is applied. In the present calculation, all absorption spectra were therefore normalized to equal a height of unity, in contrast to analyses of spectroscopic titrations, here absolute values may be obtained by normalization with respect to the known analytical (total) concentrations.

ANALYSIS OF EQUILIBRIUM MIXTURES WITH NONIDEAL BEHAVIOR

In many equilibrium systems, models based on the law of mass action cannot be applied in a simple, reliable way. This is true whenever activity-concentration relations are not well established and the essential concentrations cannot be measured directly. Examples are titrations at low and variable ionic strength, experiments in mixed solvents of variable composition, and studies in strongly acidic or basic media where no reliable pH measurement is possible. Obviously, EFA can be applied unproblematically to all such cases, since it does not depend at all on the validity of a model such as the LMA. Calculated EFA concentrations are identical regardless of whether the individual experiments are identified by the respective pH values, by the amount of titrant added, or simply by any numbers in ascending or descending order.

As one relatively simple example we discuss the protonation of 1,4,7-triazacycl[3.3.3]azine.[41] Besides taking up a first proton at pH 3.46, this molecule can be further protonated to LH_2^{2+} and LH_3^{3+} in strongly acidic media near pH −2 and pH −7, respectively.[41] Absorption spectra had to be obtained in up to 85% H_2SO_4 in order to observe both protonation steps. Of course, no reliable pH measurements are possible under such conditions and acidities had to be estimated using Hammett ρ-values.[41] Based on such values, ligand protonation constants, and thus concentration profiles, as well as absorption spectra could of course be calculated, but as expected the overall fit was not optimal, $\sigma(Y) = 7.5 \times 10^{-3}$ a.u.[41] As is shown in Fig. 5, absorption spectra (and concentration profiles) could be obtained without any problems using EFA, and the overall fit to the data is considerably improved, $\sigma(Y) = 4.1 \times 10^{-3}$ a.u., without bothering about pH values or the validity of the LMA. In the present case, the spectra calculated on the basis of the LMA still are essentially correct because of only weak overlap in the concentration profiles, but there is no reason to doubt that the EFA result is at least as reliable and that in more complicated cases EFA could be the only acceptable method.

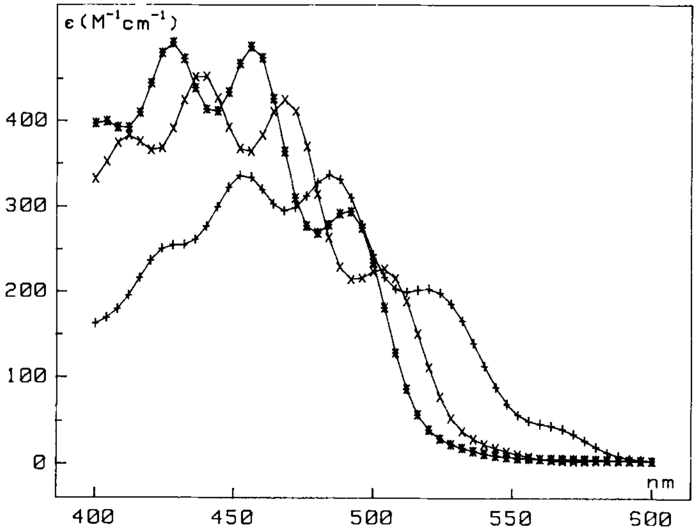


FIGURE 5 Absorption spectra of protonated 1,4,7-triazacycl[3.3.3]azine (L). +: LH^+, ×: LH_2^{2+}, *: LH_3^{3+}.

PRODUCT QUALITY CONTROL

Although not a primary goal in our applications, EFA may well be used in various fields of product quality control and detection of impurities. Due to our main field of research, this is again demonstrated on an example from coordination chemistry, the complexation of Cu^{2+} by 1,4,7-triazacyclodecane, [10]aneN_3.[42] This system has been studied previously by potentiometric[43,44] and by spectrophotometric[34] titration; it was explained with the species CuL^{2+}, CuL_2^{2+}. In the course of a broad study on the complexing properties of a series of triaza-macrocycles, the interaction of Cu^{2+} with [10]aneN_3 was restudied and the data subjected to EFA.[42] As is shown by the dotted lines in Fig. 6, EFA of a titration with 40 mol% copper per ligand clearly shows four absorbing species at low pH, while only three such species, Cu^{2+}, CuL^{2+}, and CuL_2^{2+}, would be expected. The dimer $Cu_2L_2H_{-2}^{2+}$ is only formed above pH 8 in solutions containing more than 50 mol% of Cu^{2+}.

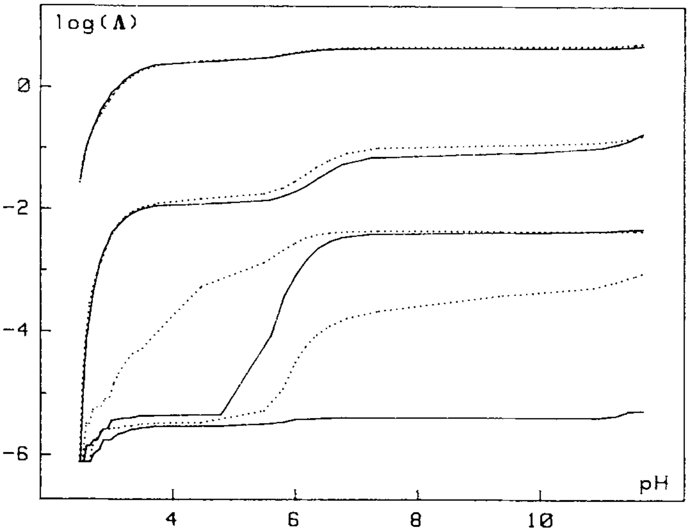


FIGURE 6 Forward EFA plots for the spectroscopic titration of [10]aneN_3 with Cu^{2+}. ···: Results obtained with material contaminated by dimeric impurity, four significant eigenvalues. ———: Results with purified ligand, three significant eigenvalues.

Careful inspection of individual titration curves (absorbance vs. pH) at a given wavelength indeed revealed some minor deviations at low pH. Various alternative models featuring additional species such as $CuLH^{3+}$, CuL_2H^{3+} or $Cu(LH)_2^{4+}$ were tried without much success. Despite the fact that the ligand had a perfect elemental analysis and did not show inhomogeneity based on TLC, the synthesis was repeated starting from cyclic tritosylate which had been carefully purified by column chromatography. Using this new material, we obtained a clearcut solution of the problem already based on forward EFA, as indicated by the solid lines in Fig. 6. The one superfluous eigenvalue between those corresponding to CuL^{2+} and CuL_2^{2+} now is completely absent, the others remaining essentially unchanged. While the nature of the impurity in the older product has not been studied in detail, we assume it to be the dimer of [10]aneN_3 which would of course have an identical elemental composition. Such dimers have indeed been observed for other triaza macrocycles[45–47] and seem to be a relevant problem in syntheses based on the tosylester method of Richman and Atkins.[48]

SEMIQUANTITATIVE ANALYSIS AND MODEL SELECTION OF SPECTROSCOPIC DATA

We have shown that very satisfactory quantitative analyses are possible through EFA in many cases. Nevertheless, the use of models based on a specific system of chemical equilibria will continue and the calculation of the corresponding stability constants remains the final goal of spectroscopic titrations, at least in cases where such experiments can be performed under strictly controlled conditions, e.g., of temperature, solvent composition, and ionic strength. Even in these fortunate cases, EFA is an important tool for several reasons. The most obvious of them is the availability of a completely independent test of the result based on the law of mass action, as has been discussed above. If the model-free and model-based concentration profiles and spectra turn out to be the same, this provides suggestive and nontrivial support for the model selected to explain the data.

Other reasons pertain more specifically to complicated systems. Here, the selection of the correct chemical model may become ambiguous and also the least-squares refinement based on EFA may run into complications. Semiquantitative considerations based on primary EFA still are most helpful in such cases and this is discussed in the following.

As one example we choose the complexation of Cu^{2+} by 4,7,10-triazatridecane dioic acid, which has been studied potentiometrically and spectrophotometrically along with a series of other pentadentate chelators.[35] Based on coordination chemical considerations one would expect two complexes, $CuLH^{+}$ and CuL, to be the most stable species, and the potentiometric data could be explained satisfactorily based on that model. Spectrophotometric data, however, clearly indicated that this model is incomplete and additional species had to be assumed. This finding is quite in line with the general observation that potentiometry, while quite remarkable in the reproducibility of the data, is inferior to spectrophotometry as far as discrimination between different models is concerned. Additional species which might be expected for the system under consideration are $CuLH_2^{2+}$, possibly $CuLH_3^{3+}$, and CuL $(OH)^{-}$. The first two species would be expected at low pH, the last above pH 9. As shown in Fig. 7, the result of primary EFA is quite unambiguous[7]: a total of four absorbing species is present, all at low pH. The fifth eigenvalue remains insignificant over the whole experimental range. The correct model thus consists of the species Cu^{2+}, $CuLH_2^{2+}$ $CuLH^{+}$, and CuL, while $CuLH_3^{3+}$ and $CuL(OH)^{-}$ are not detected. Indeed, the spectrophotometric titrations could be explained very well, based on that model and using the law of mass action. It should be obvious that the result of EFA not only provided the basis for selecting the appropriate model but also yielded excellent starting values for the respective equilibrium constants. Application of EFA prior to the classical least-squares analysis thus obviates the time-consuming testing of many different models and any preliminary, e.g., graphical, methods for obtaining estimates of the unknown equilibrium constants.

The analysis of the system discussed above still was rather straightforward and the correct number of absorbing species (although, of course, not their ranges of existence) would also have been revealed by ordinary principal component analysis (PCA). In still more complicated systems containing more than perhaps

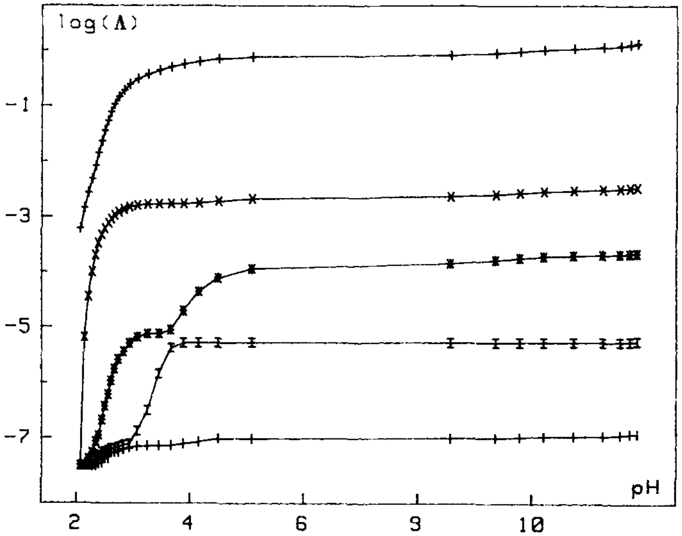


FIGURE 7 Forward EFA plot for the spectroscopic titration of 4,7,10-triazatridecane dioic acid with Cu^{2+}, showing four components at pH < 4 and no further change up to pH 12.

five or six absorbing species, this is no longer always true. This has been observed, e.g., for the complexation of Cu^{2+} by L-alanine amide,[7] glycine ethylamide[37] and N,N'-diglycyl-1,2-ethanediamine,[37] which has to be described by a model based on seven absorbing species in each case. In none of these cases could this number have been deduced correctly from the number of significant eigenvalues based on PCA because of the broad unstructured d-d* transitions in Cu^{2+} complexes. It also must be admitted that iterative EFA would not yield reliable concentration profiles and spectra in these cases. EFA has its scope and limits like any other method. Nevertheless, semiquantitative application of primary EFA still is possible and helpful in the model selection process in such complicated situations. Even if the number of absorbing species cannot be determined directly from PCA, the formation of a new species at a given pH still is noticed through EFA by its influence on the actual eigenvalues in the forward and backward analyses. This application of EFA is not discussed any further here; it has been dealt with in detail in the first paper on EFA.[7] It may suffice to state that a sixth or seventh species, even if not actually leading to an additional significant eigenvalue, still may be clearly noticed and its range of existence estimated by a strong upsurge of the already detected more significant eigenvalues.

CONCLUSIONS

Evolving Factor Analysis is a new and promising algorithm for the completely model-free *and* quantitative analysis of a variety of bilinear data. Applicability of EFA is based on only two general conditions, linear response of the dectection system and contiguous ranges of species existence. Due to our persistent interest in complexation equilibria, EFA has been first developed and used in this field of coordination chemistry and used for the analysis of spectrophotometric as well as ESR titrations. As is obvious from the mathematical concept of EFA, its scope is much broader, however; it can be applied to almost any kind of ordered data matrices. The most important field of application, for which EFA also has been tested successfully, could well be the field of peak resolution in chromatography using multichannel detection systems. It is in this field of chemometrics where an impressive effort on model-free analysis has been going on for the past few years, all the more understandable since no satisfactory general model is likely to emerge for peak shapes in chromatography in the near future.

Acknowledgment

This work has been supported by the Swiss National Science Foundation (Grant No. 2.213-0.81).

HARALD GAMPP, MARCEL MAEDER, CHARLES J.
MEYER and ANDREAS D. ZUBERBUEHLER
Institut für Anorganische Chemie der Universität,
Spitalstrasse 51, CH-4056 Basel,
Switzerland

References

1. J. A. Howell and L. G. Hargis, Anal. Chem. **58**, 108R (1986).
2. Chem. Eng. News, "Instrumentation 86," **64** (12), 21 (1986).
3. C. L. Wilkins, Science **222**, 291 (1983).
4. L. S. Ramos, K. R. Beebe, W. P. Carey, E. Sanchez M., B. C. Erickson, B. E. Wilson, L. E. Wangen and B. R. Kowalski, Anal. Chem. **58**, 294R (1986).
5. *Computational Methods for the Determination of Formation Constants*, ed. D. J. Leggett (Plenum Press, New York, 1985).
6. M. F. Delaney, Anal. Chem. **56**, 261R (1984).
7. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, Talanta **32**, 1133 (1985).
8. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, Chimia **39**, 315 (1985).
9. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, Talanta (in press).
10. M. Maeder and A. D. Zuberbühler, Anal. Chim. Acta **181**, 287 (1986).
11. E. R. Malinowski and D. G. Howery, in *Factor Analysis in Chemistry* (Wiley-Interscience, New York, 1980).
12. G. H. Golub and C. Reinsch, Numer. Math. **14**, 403 (1970).
13. E. R. Malinowski, Anal. Chem. **49**, 612 (1977).
14. S. Wold, Technometrics **20**, 397 (1978).
15. T. M. Rossi and I. M. Warner, Anal. Chem. **58**, 810 (1986).
16. W. H. Lawton and E. A. Sylvestre, Technometrics **13**, 617 (1971).
17. D. Macnaughtan, L. B. Rogers and G. Wernimont, Anal. Chem. **44**, 1421 (1972).
18. M. A. Sharaf and B. R. Kowalski, Anal. Chem. **54**, 1291 (1982).
19. D. W. Osten and B. R. Kowalski, Anal. Chem. **56**, 991 (1984).
20. R. A. Gilbert, J. A. Llevellin, W. E. Surark and J. W. Palmer, Appl. Spectrosc. **36**, 428 (1982).
21. N. Ohta, Anal. Chem. **45**, 553 (1973).
22. J.-H. Chen and L.-P. Hwang, Anal. Chim. Acta **133**, 271 (1981).
23. B. G. M. Vandeginste, R. Essers, T. Bosman, J. Reijnen and G. Kateman, Anal. Chem. **57**, 971 (1985).
24. A. Meister, Anal. Chim. Acta **161**, 149 (1984).
25. O. S. Borgen and B. R. Kowalski, Anal. Chim. Acta **174**, 1 (1985).
26. P. J. Gemperline, J. Chem. Inf. Comput. Sci. **24**, 206 (1984).
27. B. G. M. Vandeginste, W. Derks and G. Kateman, Anal. Chim. Acta **173**, 253 (1985).
28. W. Lindberg, J. Oehman and S. Wold, Anal. Chem. **58**, 299 (1986).
29. R. F. Lacey, Anal. Chem. **58**, 1404 (1986).
30. H. Gampp, M. Maeder and A. D. Zuberbühler, Talanta **27**, 1037 (1980).
31. M. Maeder and H. Gampp, Anal. Chim. Acta **122**, 303 (1980).
32. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, Talanta **32**, 95 (1985).
33. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, Talanta **32**, 257 (1985).
34. A. D. Zuberbühler and Th. A. Kaden, Talanta **26**, 1111 (1979).
35. H. Gampp, D. Haspra, M. Maeder and A. D. Zuberbühler, Inorg. Chem. **23**, 3274 (1984).

36. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, in *Biological and Inorganic Copper Chemistry*, Vol. 1, eds. K. D. Karlin and J. Zubieta, (Adenine Press, Guilderland, 1985), p. 255.
37. M. Briellmann and A. D. Zuberbühler, Helv. Chim. Acta **65**, 46 (1982).
38. Th. A. Kaden and A. D. Zuberbühler, Helv. Chim. Acta **57**, 286 (1974).
39. J. Grimalt, H. Iturriaga and X. Tomas, Anal. Chim. Acta **139**, 155 (1982).
40. M. D. King and G. S. King, Anal. Chem. **57**, 1049 (1985).
41. W. Leupin, D. Magde, G. Persy and J. Wirz, J. Am. Chem. Soc. **108**, 17 (1986).
42. M. Briellmann, Th. A. Kaden, C. J. Meyer and A. D. Zuberbühler (in preparation).
43. T. Riedo and Th. A. Kaden, Chimia **31**, 220 (1977).
44. H. Gampp, M. Maeder, A. D. Zuberbühler and Th. A. Kaden, Talanta **27**, 513 (1980).
45. T. N. Margulis and L. J. Zompa, J. Chem. Soc. Chem Commun. 430 (1979).
46. M. E. Angley, M. Dwyer, S. F Lincoln, G. H. Searle, R. J. Geue and F. R. Keene, Inorg. Chim. Acta **45**, L91 (1980).
47. G. H. Searle and R. J. Geue, Austr. J. Chem. **37**, 959 (1984).
48. J. E. Richman and T. J. Atkins, J. Am. Chem. Soc. **96**, 2268 (1974).