ANEXO 3

“Evaluation of classical and three-way multivariate calibration procedures in kinetic-spectrophotometric analysis.”

Stanley R. Crouch, Jordi Coello, Santiago Maspoch, Marta Porcel

Analytica Chimica Acta, **424**, 115-126, **2000**

ELSEVIER

Analytica Chimica Acta 424 (2000) 115–126

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Evaluation of classical and three-way multivariate calibration procedures in kinetic-spectrophotometric analysis

S.R. Crouch [a], J. Coello [b], S. Maspoch [b,*], M. Porcel [b]

[a] *Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA*
[b] *Departamento de Quimica, Unidad de Quimica Analitica, Facultad de Ciencias, Universidad Autónoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*

Received 18 January 2000; received in revised form 9 May 2000; accepted 31 July 2000

Abstract

The bidimensional multivariate regression procedures: multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLS) and continuum regression (CR), and several N-way methods such as N-way PLS (nPLS) and parallel factor analysis (PARAFAC) are tested as calibration methods for the kinetic-spectrophotometric determination of ternary mixtures in a pseudo first-order kinetic system. The different calibration procedures were first applied to computer simulated kinetic-spectrophotometric data where the effect of spectral overlap and the differences in the kinetic constants were evaluated at a low level of experimental noise. Later they were applied to the stopped-flow kinetic-spectrophotometric simultaneous resolution of Co(II), Ni(II) and Ga(III) using 4-(2-pyridylazo)resorcinol (PAR) as a chromogenic reagent. Accurate estimations of concentrations with relative standard errors of prediction of about 8% were obtained even though a high degree of spectral overlap and similar rate constants were present. The study of the influence of experimental noise on the 3-component system justifies the difference between the simulations and the experimental results for the different calibration procedures. PARAFAC and MLR did not allow the resolution of the proposed 3-component system. CR provided slightly better results than those obtained by PLS, PCR and nPLS.

Keywords: Simultaneous kinetic determination; Multivariate calibration; N-way methods; 4-(2-pyridylazo)resorcinol; Metal ions determination; Stopped-flow photodiode array spectrophotometry

1. Introduction

Applications of kinetic-spectrophotometric methods of analysis to simultaneous determinations have grown recently as the result of the incorporation of computerized data acquisition systems based on multiwavelength detectors (e.g. diode arrays) and the development of powerful mathematical treatments for processing the recorded information (e.g. multivariate calibration) [1,2].

Principal component regression (PCR) and partial least squares regression (PLS) are well known multivariate calibration procedures [3] widely used in recent years for the simultaneous determination of analytes in mixtures by means of kinetic-spectrophotometric procedures. Less widely used in this field is continuum regression which has unified under one approach the regression techniques of multiple linear regression (MLR), PCR and PLS [3]. These calibration procedures do not require a prior knowledge of the kinetic

* Corresponding author. Tel.: +34-93-581-1011; fax: +34-93-581-2477
E-mail address: iqan7@blues.uab.es (S. Maspoch).

0003-2670/00/$ – see front matter
PII: S0003-2670(00)01108-9

model and are relatively robust in the presence of experimental noise. In spite of their flexibility, multivariate calibration presents certain limitations when the chemical system in study shows a non-linear behaviour. For highly non-linear systems artificial neural networks (ANN) [4,5] with a previous decomposition of the data into principal components (PCA) have proved successful.

Kinetic-spectrophotometric data has an intrinsical three-dimensional structure (sample, wavelength, time). To be able to apply the above methods to kinetic-spectrophotometric data (three-way data), it is necessary to unfold the data in two ways obtaining tables similar to bidimensional tables [3,6]. If this approach is taken, it means that the scans recorded at various times are sequentially linked together to form a single row per sample in the $\boldsymbol{X}$ data matrix ($\lambda_1 t_1, \ldots \lambda_2 t_1 \ldots \lambda_n t_1 \ldots \lambda_i t_j \ldots \lambda_1 t_m, \lambda_2 t_m \ldots \lambda_n t_m$). It is obvious that for a sample, the measured absorbance at λ_i and t_j is related to the measured absorbance at λ_{i+1} and t_{j+1}. This information is lost when the data are unfolded, resulting in matrices with very correlated variables which give rise to models difficult to interpret. These problems can be surpassed using multi-way procedures which keep the tridimensional structure of the data. Recently the multi-way data analysis methods: N-way partial least squares (nPLS) [7,8], parallel factor analysis (PARAFAC) [9], trilinear decomposition (TLD) and multivariate curve resolution based on alternating least-squares (ALS) [10] have been applied to kinetic-spectrophotometric data. In these methods, the data for each sample are arranged in a matrix with the rows containing the spectra measured at preselected successive times during the evolution of the reaction, and the columns the kinetic profiles measured at preselected successive wavelengths. For the analysis of three-way data sets, using methods such as PARAFAC and nPLS, trilinearity is an essential property (i.e. every chemical species must be defined in the different data matrices by the same spectral and concentration profiles). Under such conditions, these methods can recover the concentration and pure spectral profiles of the analytes. Also, with only the additional input of the analyte spectrum, they can determine the analyte concentration in the presence of unknown interferences by using a reduced number of standards (second-order advantage). Kinetic-spectrophotometric data have been found to produce rank deficiency and rank overlap [11]. A data matrix is rank deficient when the number of significant contributions to the data variation estimated is lower than the real number of chemical components presents in the system. Such a situation occurs when at the beginning of a chemical reaction, more than one component already exists (absorption of the reagent or analytes). When two chemical species are characterized by the same profile in any of the two orders (i.e. the same kinetic profile or complete spectral overlap for two analytes) there is rank overlap.

Saurina et al. analyzed the resolution of simulated and experimental first-order kinetic systems in the presence of rank deficiencies caused by rank overlap in any of the two orders (time or spectral) using a curve resolution method based on alternating least squares [11]. Thus, if trilinearity is present and the system is full rank, the estimation of the correct profiles and concentrations is possible and it results in simple and robust models which preserve all the sources of variability of the data. In spite of the potential advantages that can be provided by three-way data analysis methods, it has not been shown yet that they provide a better predictive capacity than the conventional unfolded methods with kinetic systems under the same experimental conditions. Also, it has yet to be shown how spectral overlap, similarity in the rate constants and experimental noise affect the results.

In this work, simulations were first carried out to learn how spectral overlap and reaction rate differences of three analytes influence the predictive ability of the multivariate calibration procedures mentioned above. Subsequently, the methods were applied to the stopped-flow kinetic-spectrophotometric resolution of the mixtures of Ni(II), Co(II) and Ga(III) complexes with the chromogenic agent 4-(2-pyridylazo)resorcinol, (PAR). Preliminary studies showed slight differences in UV–VIS spectra of metal–PAR complexes. Moreover, some differences were also observed in the reaction rates of each metal, allowing some kinetic discrimination. Thus, the three analytes present different behaviour in the two orders of measurement (i.e. the spectral order and the kinetic order). Before proceeding to the resolution of the 3-component system, this reaction was applied to mixtures of Co(II) and Ni(II) by direct reaction with PAR, which has not been described previously in the

literature. Finally, using the spectra of the reaction products for each metal–PAR complex and the experimentally estimated rate constants, the effects of increasing instrumental and rate constant noise were studied.

1.1. Theory

The multivariate calibration algorithms PLS, PCR and MLR have been extensively described in the literature [3,6] so, only a brief description of the less known methods used in this work (CR and N-way procedures) is given here.

1.1.1. Continuum regression (CR)

First, the algorithm carries out a singular value decomposition SVD (or PCA) on the centered data matrix, $\boldsymbol{X}$ [3,6]:

$$\boldsymbol{X} = \boldsymbol{USV}^{\mathrm{T}} \tag{1}$$

where $\boldsymbol{US} = \boldsymbol{T}$ is the matrix of scores and $\boldsymbol{V}$ is the loading matrix. The elements on the diagonal of $\boldsymbol{S}$ are positive values (singular values), while the values in the off-diagonal positions are zero. The matrix is then modified to:

$$\boldsymbol{X}^{(\gamma)} = \boldsymbol{US}^{(\gamma)}\boldsymbol{V}^{\mathrm{T}} \tag{2}$$

i.e. the singular values are raised to a certain power γ and a modified predictor matrix is constructed. Then one applies PLS regression and the results are back-transformed to the original $\boldsymbol{X}$ matrix. Changing the power from 0 to ∞ the model changes from MLR ($\gamma = 0$) via PLS ($\gamma = 1$) to PCR ($\gamma = \infty$).

1.1.2. Parallel factor analysis (PARAFAC)

PARAFAC is a decomposition method, in which a three-way table $\underline{\boldsymbol{X}}(I \times J \times K)$ is decomposed into three, two-way loading matrices (one for each mode or way) $\boldsymbol{A}$ ($I \times F$), $\boldsymbol{B}$ ($J \times F$) and C($K \times F$) such that, [3,6]

$$x_{ijk} = \sum_{f}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{3}$$

where e_{ijk} represents a residual error term and F the number of factors. I, J and K may be regarded as the number of samples, wavelengths and times, respectively. This model can also be written as,

$$\boldsymbol{X} = \sum_{f}^{F} a_{if} \otimes c_{kf} + \boldsymbol{E} \tag{4}$$

where $\otimes$ represents the ternary tensor product of the three vectors and E represents the three-way array of residuals. The elements of the loading matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ are computed by alternating least squares.

When PARAFAC is used for a calibration with kinetic-spectrophotometric data, the three two-way matrices correspond to the concentration versus sample, the concentration versus time, and the absorbance versus wavelength dimensions of the data. The number of factors in each two-way matrix ($\boldsymbol{A}$, $\boldsymbol{B}$ or $\boldsymbol{C}$) corresponds to the number of species that contributes to the data; usually this is equal to the number of analytes. Since the concentration of the analytes in the calibration set are known, it is possible to compare them to each factor in the matrix of concentration versus sample and determine which factor corresponds to which analyte. In a second step, it is possible to do a simple least-square regression between the concentration of the analyte in each sample and the concentration versus sample dimension of that analyte's factor.

The important difference between PCA and PARAFAC is that in PARAFAC there is no need for orthogonality between the three loading matrices ($\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$) to identify the model, and the solution is unique.

1.1.3. Multilinear partial least squares regression (nPLS)

The main feature of the nPLS algorithm is that it produces score vectors that in a trilinear sense have maximum covariance with the unexplained part of the dependent variable [6,12]. The goal of the algorithm is to make a decomposition of the data matrix $\underline{\boldsymbol{X}}$ into a set of triads. A triad consists of one score vector $\boldsymbol{t}$ and two weight vectors, one in the second order called $\boldsymbol{w_j}$ and one in the third order called $\boldsymbol{w_k}$ and the model is given by the equation:

$$x_{ijk} = \sum_{f}^{F} t_{if} w_{jf} w_{kf} + e_{ijk} \tag{5}$$

where e_{ijk} represents a residual error term and F the number of components. It assumes that both the dependent and the independent data are three-way. Let $\underline{\boldsymbol{X}}$ be the $I \times J \times K$ array of independent and $\boldsymbol{X}$ the $I \times JK$ unfolded array. Let $\underline{\boldsymbol{Y}}$ be the $I \times L \times M$ array of dependent data and $\boldsymbol{Y}$ the $I \times LM$ unfolded array. Thus, the nPLS models decompose $\underline{\boldsymbol{X}}$ as:

$$\boldsymbol{X} = T(W^K / \otimes / W^J)^{\mathrm{T}} + E_x \tag{6}$$

i.e. a trilinear model similar to the PARAFAC model and $\underline{\boldsymbol{Y}}$ as:

$$\boldsymbol{Y} = U(Q^M / \otimes / Q^L)^{\mathrm{T}} + E_y \tag{7}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are the scores matrices, $\boldsymbol{W}$ and $\boldsymbol{Q}$ are the loading matrices and $\boldsymbol{E}_x$, $\boldsymbol{E}_y$ the error matrices for the independent and dependent variables, respectively. A superscript *J, K, L* or *M* respectively, is used to specify which mode the vectors refer to. The decomposition models of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{Y}}$ and the expression $\boldsymbol{U} = \boldsymbol{TB} + \boldsymbol{E}_u$ relating these models, together constitute the nPLS regression model.

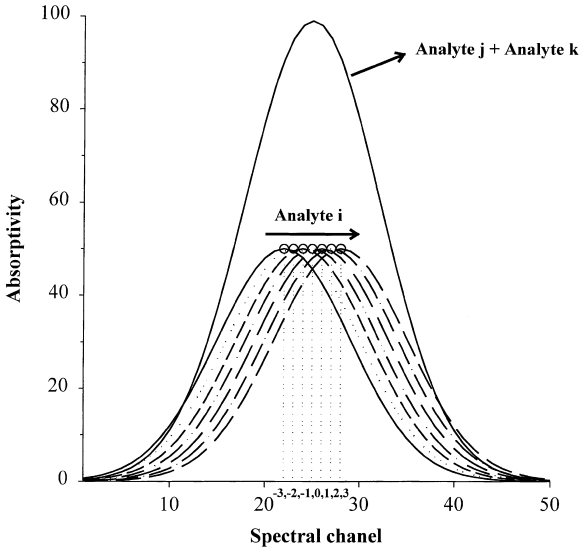

Fig. 1. Gaussian spectra used for simulations. Displacement of analyte *i* from the coded position −3 to 3 keeping constant the position of the analyte *j* and *k*,which are centered in the position −1 and 1, respectively; (−spectrum centered in position 0 is the sum of spectrum analyte *j* plus spectrum analyte *k*).

2. Experimental section

2.1. Simulations

Kinetic-spectrophotometric data were simulated using a program written in MATLAB (The Math Works, Natick, Mass.). The algorithm generates kinetic-spectra by solving differential equations and assuming that only the reaction products absorb with Gaussian spectral bands. The Gaussian bands were built with the same width ($\sigma = 20$ nm) and the same absorptivity coefficients every 1 nm over a wavelength range of 50 nm. In all cases, adherence to Beer's law was presumed for each component, and the total absorbance at each wavelength was assumed to be the sum of the absorbances of the components. The analyte concentrations were varied between 0.5–1.5 $\times 10^{-6}$, 0.75–1.75 $\times 10^{-6}$ and 1.0–2.0 $\times 10^{-6}$ M for the first, second and third analyte, respectively. Data were generated for 27 standard calibration mixtures and for eight unknown mixtures. In order to ensure pseudo first-order kinetics in relation to the PAR reagent, its concentration was 5.00 $\times 10^{-4}$ M. Under these conditions, 100 times were used in calculation (a kinetic-spectra of 100 × 50 = 5000 variables per sample), simulating that the observed fraction as reaction of the slower reacting analyte at the end of data collection was 90%. In order to study only the effect of spectral overlap and rate constant variations, the instrumental and rate constant noise contributions were kept initially constant at 1% of their values.

In order to reproduce different spectral overlaps and different reaction rates, the position and rate constant of one analyte are modified whereas the other analytes are kept constant. The analytes that are not moved remain in the positions coded as −1 and 1 (Fig. 1) and the values of the rate constants are 60, 40 and 20 for analyte 1, 2 and 3, respectively. The analyte that moves goes from position −3 to 3, producing seven spectral overlaps. In each case, seven rate constant values (10, 20, 30,40, 50, 60 and 70) were studied. The total number of simulations was 147.

2.2. Reagents

All solutions were prepared in a Borax medium consisting in 0.1 M $Na_2B_4O_7 \cdot 10H_2O$ adjusted to pH = 7 with nitric acid in distilled water.

Table 1
Concentration ranges (mg/l) for each analyte in binary and ternay mixtures used for calibration

Metal	2-component system (mg/l)	3-component system (mg/l)
Cobalt(II)	0.2–1.0	0.2–0.6
Nickel(II)	0.4–2	0.4–1.2
Gallium(III)	–	0.4–2.0

Analytical-reagent grade $Ga(NO_3)_3{\cdot}H_2O$, $NiCl_2{\cdot}6H_2O$, $Co(NO_3)_2{\cdot}6H_2O$ and 4-(2-pyridylazo) resorcinol monosodium salt hydrate (PAR) were used throughout.

A 1×10^{-3} M stock solution of PAR was prepared by weighing the appropriate amount of the reagent. Different stock solutions for each metal were prepared and diluted to obtain the working mixtures with the concentrations shown in Table 1. Concentration levels were chosen based on the linear ranges obtained with single analyte experiments. Calibration models were constructed from 25 and 47 mixtures for the 2- and 3-component system, respectively. The predictive capacity of the different models tested was assessed by using a prediction set of 15 mixtures in both cases, which contained analyte concentrations within the calibration range (Table 1). To introduce experimental variability, calibration samples, were prepared and measured on different days.

2.3. Apparatus

A stopped-flow apparatus [13–14] interfaced to a thermoelectrically cooled Tracor Northern (Model TN-6123) 512 element intensified diode array (Tracor Northern, Philadelphia, PA) configured to acquire spectra in the 400–800 nm range was used. All measurements were performed at room temperature with no additional thermostating in the stopped-flow apparatus.

2.4. Procedure

The metal solutions were mixed in a 1:1 ratio with the PAR solution in the stopped-flow mixing system. The reaction progress was followed spectrophotometrically in the 400–800 nm range. In the wavelength range of interest, 520–560 nm, absorbances were measured at 52 equally-spaced wavelengths. Kinetic information was obtained by acquiring 43 (2-components system) and 57(3-components system) scans at a rate of 7.0 scans/s for a total acquisition time of 6 and 8 s, respectively.

2.5. Data processing

Kinetic-spectrophotometric data were collected in duplicate and averaged. Data were mean-centered before being input to the appropriate algorithms. Multivariate calibration algorithms provided in the PLS_TOOLBOX 2.0 (Eigenvector Technologies, Manson, WA) [6] and run in MATLAB were used to perform determinations. Cross-validation of the calibration set was used with all unfolded methods and nPLS. The number of factors giving the minimum error was chosen to quantify the prediction set. With continuum regression, in a first step, γ values of 0, 0.06, 0.12, 0.25, 0.5, 1, 1.5, 2, 3, 4, 6, 8 and ∞ were tested. For each value of γ, the optimal number of factors was found by cross-validation. Finally, the y value was adjusted in 0.02 steps around the minimum found in the first step. The number of factors and y value found with the calibration set were used to quantify the prediction set. In PARAFAC, we decomposed a data set that consisted of both calibration and unknown data. This makes sure that the decomposition is the same for both standards and unknowns. The least-squares regressions were performed between the factors and the known concentrations for the calibration data, and then the resulting calibration model (slope and intercept) and the factors for the unknowns were used to calculate the concentrations of the analytes in the unknowns, one analyte at a time.

In order to facilitate the comparison of the different calibration methods evaluated, the relative standard error of prediction per analyte $RSEP(\%)_a$, and per mixture $RSEP(\%)_m$ are used. This error per mixture is defined as:

$$RSEP(\%)_m = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij} - \hat{c}_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij})^2}} \times 100 \quad (8)$$

where c_{ij} is the concentration added of the jth component to the jth sample, $\hat{c}_{ij}$ the estimated (mean of two runs) concentration in the test set. $RSEP(\%)_a$ in the equation only refers to one analyte ($m = 1$).

3. Results and discussion

3.1. Simulations

A problem that arises when making simulations in a kinetic system of more than two components is the interpretation and visualization of results. In systems of two components, the representation of results is simplified by using the ratio of rate constants k_x/k_y and the spectral resolution or the distance between absorption maxima [15–17].

Resolution of mixtures can be achieved by both spectral and kinetic differences. A way to express the difference between spectra is by computing the correlation coefficient (ρ^{spectral}). This parameter has the advantage of a very easy interpretation, taking into account not only the differences in the maximum, but also the differences in the spectral shape. A value of 1 means complete correlation and consequently the impossibility of resolving these mixtures using only the spectral information. That idea can be extended to the kinetic profile computing the kinetic difference as the corresponding correlation coefficient (ρ^{kinetic}). For a complex system, a measure of the discriminating information can be acquired by the discrimination index [18],

$$\text{Discrimination index} = \left[\prod_{x=1,y=2}^{x=m-1,y=m} (1 - \rho_{xy}^{\text{spectral}} \cdot \rho_{xy}^{\text{kinetic}}) \right]^{1/n} \quad ; \quad x < y \tag{9}$$

where ρ_{xy} is the spectral or kinetic correlation between x and y, y and m the number of analytes. The parameter n is the number of terms and it normalizes the expression in order to compare systems with a different number of analytes. A discrimination index approaches 1 when a large difference (kinetic or spectral) exists between analytes, so the mixture should be resolved accurately. A discrimination index approaches 0 when a very similar behaviour exists between analytes and a poorer resolution should be expected. A disadvantage that the use of this equation presents is that it does not make any differentiation between the kinetic and spectral influence, only one value is obtained that includes both variables.

Fig. 2 illustrates the accuracy of the predicted concentrations, expressed as the $\text{RSEP}(\%)_{\text{m}}$ Eq. (8), for each algorithm as a function of the discrimination index. The trend of decreasing error when the discrimination index increases appears in all unfolded methods. Determinations using CR globally produced the most accurate predictions, followed by PLS and PCR results. Despite the fact that MLR provides poor results, they still follow the expected trend of lower error when the differentiation among analytes increases. The high degree of overlap in spectral or time orders with simulated data makes the system lack of trilinear structure in many of the cases studied, so it is not possible to observe any global tendency in the errors. nPLS provides errors of the same order as the other methods, but its behaviour is not predictable with respect to the discrimination index, whereas PARAFAC always produced the worst prediction errors.

3.2. Chemical system

The reaction between some metal ions and PAR has been described in the literature [19] and applied in analytical chemistry [17,20,21]. At equilibrium, with excess PAR, Ni(II), Co(II) and Ga(III), form complexes where two PAR molecules react with one metal ion [22].

Other papers describe several methods for the determination of some metals based on differences in the rates of substitution reactions of EDTA or EGTA by PAR [23–26].

The reaction rate depends on the pH and strongly on the nature of the buffer used. At pH = 7 the UV/V is spectrum of 1×10^{-5} M of PAR shows one absorption band centered at 420 nm. The spectra of the three metal–PAR complexes and the spectrum of PAR are shown in Fig. 3.

Although multivariate techniques do not need prior knowledge of the rate constant values to resolve the mixtures, an estimation of them is convenient to facilitate a suitable selection of the concentration ranges of the calibration samples. At pH = 7 and a PAR concentration of 1×10^{-3} M, the reaction can be assumed to be pseudo first-order in the metal concentration, $v = k_{\text{apparent}}[\text{Me}]$. The apparent rate constant for each metal was calculated from the three or four reagent mixtures containing a variable concentration of metal and a constant

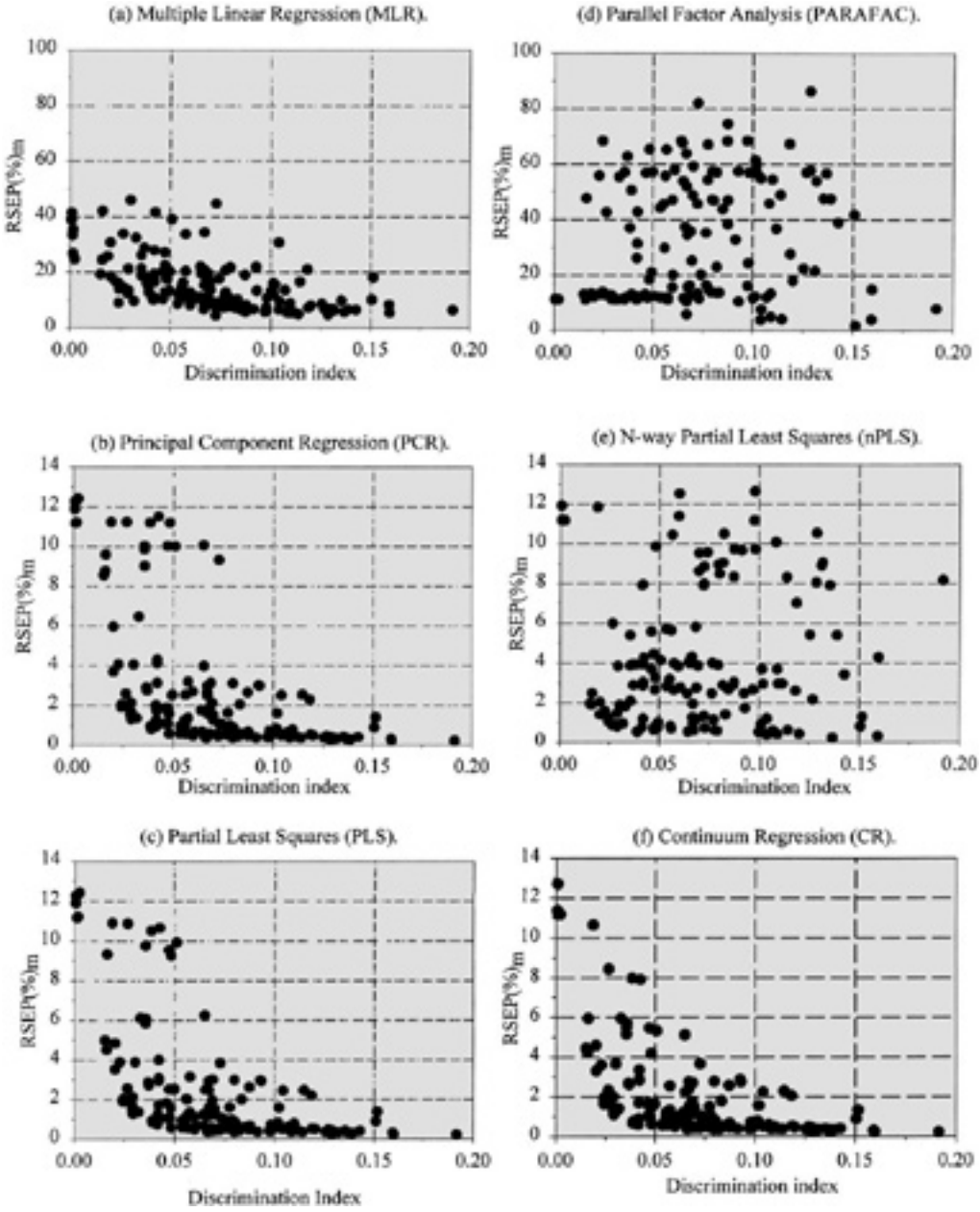

Fig. 2. RSEP(%)$_m$ Eq. (8) versus discrimination index Eq. (9). (a) MLR, (b) PCR, (c) PLS, (d) PARAFAC, (e) nPLS and (f) CR.

concentration of PAR in excess. Each kinetic measurement was made in duplicate. The following values of the rate constants were obtained ($k_{\text{apparent}} = k_{\text{Me}}[\text{PAR}]$), where uncertainties are quoted as the 95% confidence intervals: $k_{\text{apparent}}(\text{Co}^{2+}) = (2.212 \pm 0.066)\text{s}^{-1}$, $k_{\text{apparent}}(\text{Ni}^{2+}) = (0.558 \pm 0.007)\text{s}^{-1}$ and $k_{\text{apparent}}(\text{Ga}^{3+}) = (0.425 \pm 0.008)\text{s}^{-1}$.

From the spectra obtained in these individual studies for each Me–PAR complex, their respective absorptivities ($\varepsilon_{\text{Me–PAR}} - \varepsilon_{\text{PAR}}$) in the spectral range of

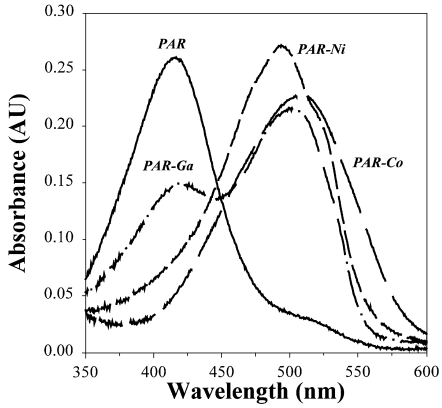

Fig. 3. Absorbance spectra for 1×10^{-4} M of reagent PAR and 2 mg/l complex solutions of Co(II)–PAR, Ni(II)–PAR and Ga(III)–PAR at pH = 7 and room temperature.

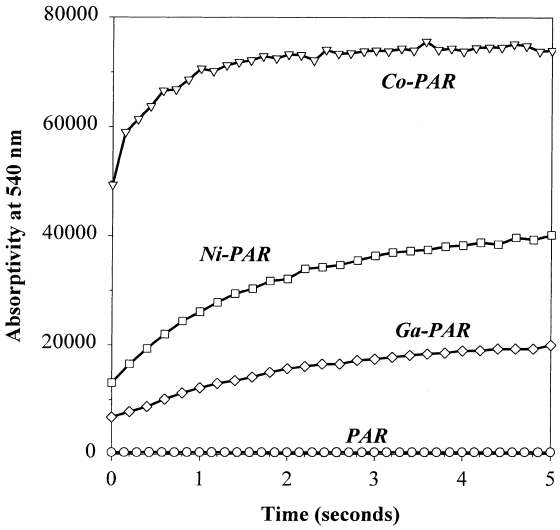

Fig. 4. Absorbance divided by molar concentration of analyte at 540 nm vs. time for Co(II), Ni(II) and Ga(III) complexes. Signal evolution of PAR.

520–560 nm have been used to calculate the kinetic and spectral correlation coefficients which are shown in Table 2 with their corresponding discrimination indexes.

Kinetic differentiation is the major source of selectivity; it is thus expected that a higher kinetic correlation coefficient will result in a higher error of prediction. Indeed, this is what is observed. Fig. 4 shows the variation of the absorptivities at 540 nm with time for solutions containing Ga(III), Ni(II) and Co(II) and a 1×10^{-3} M stock solution of PAR. In observing the kinetic profile at 540 nm the reaction for Co(II) is nearly complete before the first measurement, i.e. in the time delay of 75 ms between the flow stopping and the first data point.

3.3. Kinetic-spectrophotometric resolution of 2- and 3-component systems

An excess of ligand and pH control produce pseudo first-order conditions for both cases studied. Thus, the concentration of the different species formed during the reaction of the analytes is always linearly related to the initial concentration of the analyte.

As can be seen in Fig. 3 the spectral profiles show a high spectral overlap, confirmed by the values of the spectral correlation coefficient in Table 2, so the determination of 3-component mixtures is the most difficult task.

Despite the fact that three-way methods do not require a large number of samples for a calibration, in order to compare the results from different procedures, the same calibration and prediction sets were used in all calibration models tested.

3.3.1. Kinetic-spectrophotometric determination of Co(II) and Ni(II)

In spite of the high reaction rate of Co(II) with PAR, it was possible to determine binary mixtures of Co(II)

Table 2
Correlation coefficients (kinetic and spectral) and discrimination index Eq. (9) between analytes

Metals	Kinetic correlation coefficient($\rho_{1,2}^{\text{kinetic}}$)	Spectral correlation coefficient $\rho_{1,2}^{\text{spectral}}$	Discrimination index
Co–Ni	0.8100	0.9816	0.205
Co–Ga	0.7604	0.9829	0.252
Ni–Ga	0.9932	0.9972	0.010
Co–Ni–Ga	–	–	0.079

Table 3
RSEP%$_m$ and RSEP%$_a$ Eq. (8) in the resolution of the binary mixture of Co(II) and Ni(II) for the different calibration methods

Method	Model	RSEP(%)$_{Co}$	RSEP(%)$_{Ni}$	RSEP(%)$_m$
Unfolded	MLR	14.8	14.9	14.9
	PCR[a]	5.3	4.7	4.8
	PLS[a]	5.3	4.6	4.7
	CR[b]	4.4	4.4	4.4
Three-way	PARA[c]	26.6	40.5	38.2
	nPLS[d]	4.9	4.8	4.8

[a] Models defined with 4 factors.
[b] Models defined with 4 factors ($\gamma = 0.48$) and ($\gamma = 0.74$) for Co and Ni, respectively.
[c] The decomposition of three-way matrix was performed using two factors.
[d] Models defined with 3 factors for Co and 4 factors for Ni.

and Ni(II) using both kinetic and the spectral differentiation. The spectral region between 520–560 nm was found to contain the majority of spectral differentiation between the analytes although other wavelength ranges have been tried. Increasing the spectral range increased the data processing time, but did not improve the error of prediction. The spectral region below 520 nm is not useful due to the strong absorbance of PAR. Above 560 nm, the metal–PAR complexes exhibit low absorbance, except for Co–PAR and there is little absorbance change throughout the course of the reaction. Different time intervals were used, but the best results for both analytes were obtained using a 6 s interval. Decreasing the time used did not effect the fastest analyte [Co(II)], but the error of prediction of the slowest analyte, [Ni(II)] was higher. It is clear that the selection of wavelength range and time interval are an important consideration and must be done carefully. The results of these determinations are presented in Table 3. As expected, according to the high index of discrimination computed (Table 2), the determination of binary mixtures of Co(II) and Ni(II) was achieved with a high accuracy and an RSEP(%) on the order of 4%. For binary mixtures of Ni(II) and Ga(III), with a smaller discrimination index, the prediction errors were higher than 10% under similar experimental conditions as previously reported [17]. In comparing the multivariate calibration techniques, CR produced slightly better results, especially for Co, probably because the Co–PAR complex has the highest absorptivity in the wavelength range selected. The techniques MLR and PARAFAC were clearly inferior. PLS, PCR and nPLS provided similar results.

3.3.2. Kinetic-spectrophotometric determination of Co(II), Ni(II) and Ga(III)

The results for 3-component mixtures obtained using the same spectral range and an overall measurement time of 8 s are presented in Table 4. Under our experimental conditions, CR also produced the most accurate predictions. The techniques MLR and PARAFAC were clearly inferior and did not allow the determination of the 3-component system. PLS, PCR and nPLS provided similar results, but worse than CR.

The trend in the results is similar to what is expected based on the simulation studies at a very low noise level, where nPLS was often worse than the unfolded techniques PLS, PCR and CR. Clearly, PARAFAC and MLR cannot be used to resolve a 3-component system with these characteristics. The prediction errors are as expected according to the discrimination indexes, but higher than the values found in the simulations. The reason for this disagreement could be the difference in the noise level which was forced to be very low in the simulations in order to study the predictive capacity and behaviour of the calibration procedures. Thus, a new set of simulations was performed using the data derived from spectra of the reaction products and the calculated rate constants for the 3-component

Table 4
RSEP%$_m$ and RSEP%$_a$ Eq. (8) in the resolution of the ternary mixture of Co(II), Ni(II) and Ga(III) for the different calibration methods

Method	Model	RSEP (%)$_{Co}$	RSEP (%)$_{Ni}$	RSEP (%)$_{Ga}$	RSEP (%)$_m$
Unfolded	MLR	31.4	25.2	36.0	32.8
	PCR[a]	8.8	7.4	8.2	8.0
	PLS[a]	8.6	7.8	8.8	8.5
	CR[b]	5.3	6.8	7.9	7.5
Three-way	PARA[c]	24.0	28.9	35.5	33.0
	nPLS[d]	8.1	6.6	8.6	8.0

[a] Models defined with 4 factors for Co and 5 factors for Ni and Ga.
[b] Models defined with 4 factors ($\gamma = 0.60$), ($\gamma = 0.80$) and ($\gamma = 0.90$) for Co, Ni and Ga, respectively.
[c] The decomposition of three-way matrix was performed using three factors.
[d] Models defined with 4 factors for Co and Ni and 5 factors for Ga.

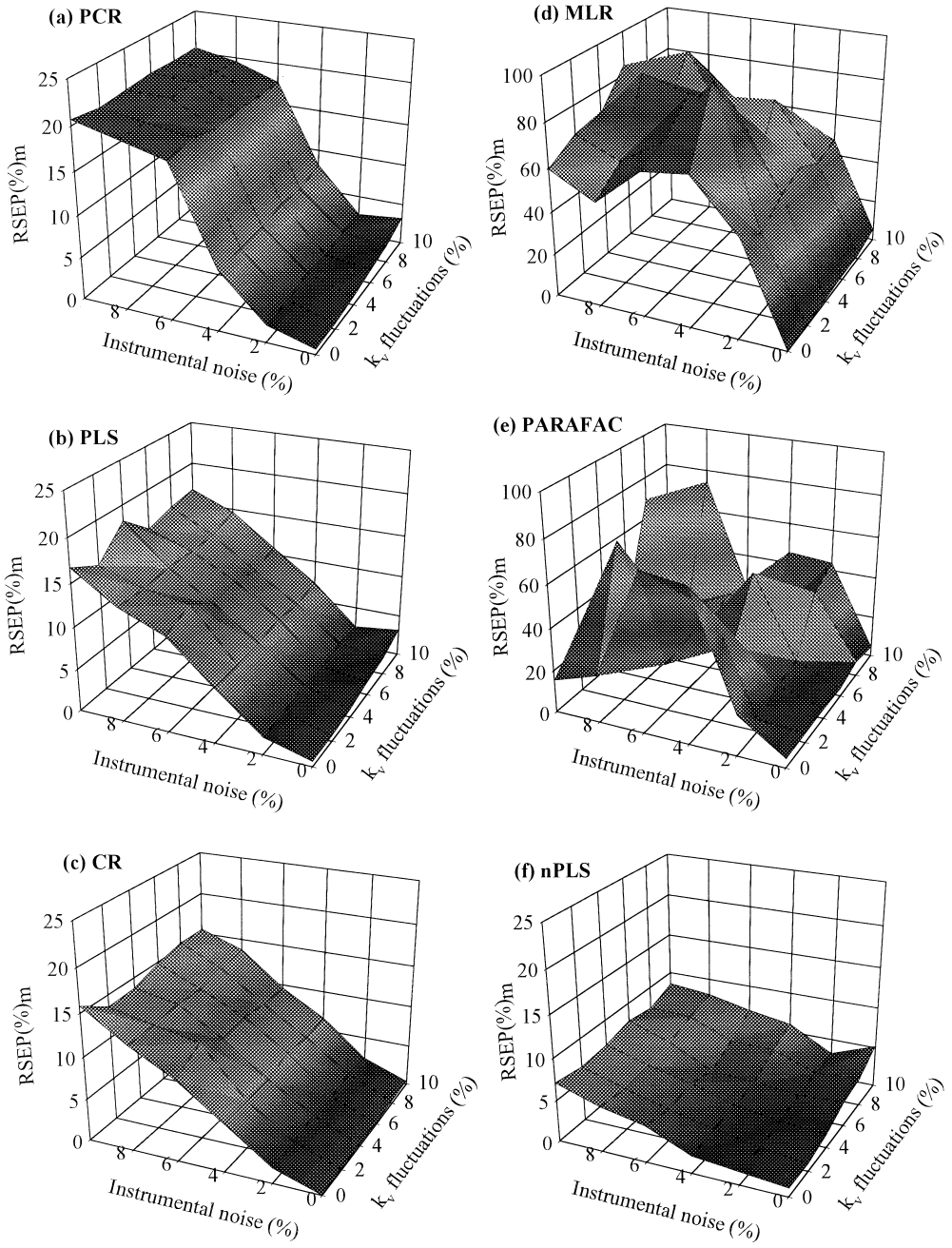

Fig. 5. RSEP(%)$_m$ as a function of constant rate (k_v) fluctuations and instrumental noise. (a) PCR, (b) PLS, (c) CR, (d) MLR, (e) PARAFAC and (f) nPLS.

mixtures. A study of the influence of instrumental noise and rate constant fluctuations was carried out. The rate constants were varied with a Gaussian distribution centered on the true value and having a standard deviation proportional to a percentage of the rate constant between 0–10%. Instrumental noise proportional to the absorbance at each wavelength was added to all simulated data taking values between 0–10%. So again, as can be seen in Fig. 5, all the unfolded multivariate calibration techniques followed the same trend. The results expressed as $RSEP(\%)_m$ have been influenced by instrumental noise and practically unaltered by the constant rate fluctuations. This can be explained because similarity between the spectra is higher than similarity in reaction rate constants (Table 2) and uncontrolled changes in rate constants will have less effect in prediction errors than changes in spectra. Moreover, nPLS is not affected by instrumental noise as much as the unfolded models, but it is more affected by rate constant fluctuations. Thus, the nPLS technique is expected to be more sensititive than the unfolded models to such uncontrolled experimental conditions as temperature fluctuations, pH variations and ionic strength changes that affect the rate constants. The results obtained by PARAFAC did not follow a clear trend when the two different sources of variation are added.

4. Conclusions

The discrimination index is an adequate parameter to define the kinetic-spectral features of a kinetic system. This parameter allows the comparison of systems with different numbers of analytes. As is shown in the simulations, at low noise level the classical techniques PLS and PCR provide similar or slightly better results than the three-way models. Continuum regression, which combines the characteristics of PLS, PCR and MLR, provides the smallest errors in the prediction samples.

The error found in the resolution of the 2- and 3-component real systems studied here and the value found previously [17] for Ga(III) and Ni(II) mixtures are correlated with the discrimination index of the system.

Under experimental conditions with high spectral overlap and similarity in the kinetic model of the components, PARAFAC and MLR do not present a good alternative for the resolution of mixtures; nPLS and unfolded methods allow accurate estimations of the concentrations.

In the case studied, with different constant rate and similar spectra, PLS, PCR and CR are more sensitive to instrumental noise than nPLS. However, nPLS is more sensitive to fluctuations in the rate constant values than the unfolded methods.

Acknowledgements

The authors are grateful to the Spain's DGICyT for funding this research within the framework of projects PB97–0213 and PB96–1 180. M. Porcel acknowledges additional support by the Spanish Ministry of Education and Sciences in the form of an FPI grant. The authors are also grateful to Dr. Thomas F. Cullen for his expert assistance with the chemometrics software, with the stopped-flow kinetics apparatus and with the chemical system.

References

[1] S.R. Crouch, T.F. Cullen, Anal. Chem. 70 (1998) 53R.
[2] T.F. Cullen, S.R. Crouch, Mickrochim. Acta 126 (1997) 1.
[3] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics. Elsevier, Amsterdam, 1998.
[4] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Redón, Anal. Chem. 67 (1995) 4477.
[5] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Porcel, Anal. Chim. Acta 398 (1999) 83.
[6] B.M. Wise, N.B. Gallagher, PLS-Toolbox 2.0 User's Guide, Eigenvector Technologies, Manson, WA, 1998.
[7] M. Azubel, M.F. Fernández, B.M. Tudino, E.O. Troccoli, Anal. Chim. Acta 398 (1999) 93.
[8] Å.K. Petterson, B. Karlberg, Anal. Chim. Acta 354 (1997) 241.
[9] J.C.G. Esteves de Silva, C.J.S. Oliveira, Talanta 49 (1999) 889.
[10] J. Saurina, S. Hernández-Cassou, R. Tauler, Anal. Chem. 69 (1997) 2329.
[11] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, J. Chemom. 12 (1998) 183.
[12] R. Bro, J. Chemom. 10 (1996) 47.
[13] P.M. Beckwith, S.R. Crouch, Anal Chem. 44 (1972) 221.
[14] T.F. Cullen, Ph.D. Dissertation, Michigan State University, East Lansing, MI, 1999.
[15] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Redón, Anal. Chim. Acta 303 (1995) 309.

[16] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, M. Redón, N. Villegas, Analyst 121 (1996) 395.
[17] T.F. Cullen, S.R. Crouch, Anal. Chim. Acta 407 (2000) 135.
[18] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, N. Villegas, 9as Jornadas de Análisis Instrumental, Barcelona, 1999.
[19] E. Mentasti, C. Baiocchi, L.J. Kirschenbaum, J. Chem. Soc., Dalton Trans. 12 (1985) 2615.
[20] B.M. Quencer, S.R. Crouch, Anal. Chem. 66 (1994) 458.
[21] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Riba, E. Rovira, Talanta 40 (1993) 261.
[22] S. Shibata, in: H.A. Flaschka, A.J. Barnard Jr. (Eds.), Chelates in Analytical Chemistry, Vol. 4, Marcel Dekker, New York, 1972.
[23] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Riba, Anal. Chem. 66 (1994) 2905.
[24] A. Cladera, E. Gómez, J.M. Estela, V. Cerda, Anal. Chem. 65 (1993) 707.
[25] O. Abollino, E. Mentasti, C. Sarzanini, V. Porta, L.J. Kirschenbaum, Analyst 116 (1991) 1167.
[26] A. Izquierdo, G. López-Cueto, J.F. Rodríguez-Medina, C. Ubide, Química Analítica 17 (1998) 67.