ORIGINAL PAPER

Multivariate standard addition method solved by net analyte signal calculation and rank annihilation factor analysis

Bahram Hemmateenejad · Saeed Yousefinejad

Abstract This article describes the use of the net analyte signal (NAS) concept and rank annihilation factor analysis (RAFA) for building two different multivariate standard addition models called "SANAS" and "SARAF." In the former, by the definition of a new subspace, the NAS vector of the analyte of interest in an unknown sample as well as the NAS vectors of samples spiked with various amounts of the standard solutions are calculated and then their Euclidean norms are plotted against the concentration of added standard. In this way, a simple linear standard addition graph similar to that in univariate calibration is obtained, from which the concentration of the analyte in the unknown sample and the analytical figures of merit are readily calculated. In the SARAF method, the concentration of the analyte in the unknown sample is varied iteratively until the contribution of the analyte in the response data matrix is completely annihilated. The proposed methods were evaluated by analyzing simulated absorbance data as well as by the analysis of two indicators in synthetic matrices as experimental data. The resultant predicted concentrations of unknown samples showed that the SANAS and SARAF methods both produced accurate results with relative errors of prediction lower than 5% in most cases.

Introduction

A big role of chemometrics in analytical chemistry is instrumental specialization. In other words, multivariate calibration models are built to provide selectivity for a multivariate analytical instrument in the presence of direct interference. Direct or spectral interferences in chemical analysis based on spectroscopic methods are those which arise when a sensor is not completely specific for the analyte and are quite common in most spectroscopic methods of analysis. So, the development of multivariate calibration methods such as principal component regression and partial least squares (PLS) [1–3] reduced the problem of direct interferences.

Calibration is the method of relating the known state of a system to measured data collected from the system. Generally, analytical signals can be classified as zero order (e.g., include absorbance at a single wavelength, a single pH measurement, or a single temperature measurement), first order (e.g., absorption spectra, emission spectra, chromatograms, and kinetic profiles), and second or higher order (e.g., many hyphenated chemical analysis techniques such as fluorescence excitation–emission spectra, chromatograms with diode-array detection, and kinetic profiles collected with array detectors) [4]. Multivariate calibration, then, is the process of relating the known state of a system to a series of measured variables describing the system. So, it is clear that multivariate calibration is applicable to first- or higher-order data. Nowadays, since many of the data collected by analytical scientists are at least first order, the use of multivariate calibration is becoming routine.

B. Hemmateenejad (✉) · S. Yousefinejad
Department of Chemistry, Shiraz University,
Shiraz, Iran
e-mail: hemmatb@sums.ac.ir

B. Hemmateenejad · S. Yousefinejad
Medicinal & Natural Products Chemistry Research Center,
Shiraz University of Medical Sciences,
Shiraz, Iran

For the compositional analysis of an unknown mixture, another very serious problems that appears while using a first-order multivariate model is the presence of an unexpected interference that may render a chemical analysis invalid [5–7]. This problem is named "indirect interferences" (or "matrix effect"), which affect the signal produced by the analyte of interest by alteration of, e.g., viscosity, surface tension, vapor pressure of the sample solution, or pH and the other interactions between the analyte and the other substances present in the sample solution. The standard addition method (SAM) is a well-known way to solve the problem of the matrix effect [8, 9].

In the presence of both direct and indirect interferences, the analysis is accompanied by some complexity, and specific approaches (e.g., combining different methods such as matrix simulation, SAM with different multivariate calibration methods) should be employed. In the case of very complex samples, the usual way is to build the model from samples that have a nature similar to that of unknown samples. The analytes of interest are determined in these samples using standard methods and their concentration is related to the spectrum of the sample in the calibration step. In this way, urea and uric acid in human blood plasma [10], glucose in blood [11], protein, fat, and water in meat [12], and reducing sugars, humidity and acidity in honey [13] were determined. This method requires the collection of a large number of calibration samples to get a representative set that properly spans all known sources of variation.

Another well-known way to solve the matrix effect, or indirect interferences, is the generalized SAM (GSAM) proposed by Saxberg and Kowalski [14]. It was the generalization of the conventional (univariate) SAM in multicomponent analysis. The GSAM can be applied in the presence of both direct and indirect interferences. The GSAM provides a means of detecting interference effects, quantifying the magnitude of the interferences, and simultaneously determining analyte concentrations in different branches of analytical chemistry [15, 16]. However, if there was an unexpected and uncalibrated interference in the standard addition set, the GSAM as a first-order SAM could not eliminate its influence. Osten and Kowalski [17] proposed a method based on curve resolution and a calibration technique for background detection and correction in multicomponent analysis, but unfortunately this method cannot provide a unique solution, because it is applied to one-dimensional analytical signals. So, the GSAM was extended to two-dimensional bilinear data to reduce the problem [18].

Currently, the best choice to cover all indirect, direct, and uncalibrated interferences is the second-order SAM (SOSAM) [19]. The generalized rank annihilation method (GRAM) as a second-order calibration method has been combined with SAM for high-performance liquid chromatography–diode-array detection data [20]. But since the GRAM is restricted by one mode having maximally dimension two (i.e., two samples), this application has been limited to only one addition for each sample. Booksh et al. [21] have extended the SAM to second-order instrumentation using trilinear decomposition. Other methods, such as alternative trilinear decomposition and parallel factor analysis, are also applied as the decomposition tool in SOSAM [19, 22]. Although SOSAMs have a high power in resolving matrix effect problems, the first-order SAMs are still employed for different analytical purposes. The first-order instruments are simpler and cheaper than second-order apparatuses. In addition, SOSAMs need specific types of instruments, which are not ready available in all laboratories. So, in the absence of unexpected direct interferences, a first-order SAM, e.g., the GSAM, could be applied well.

Net analyte signal (NAS) calculation [23] is a well-known and efficient method introduced as a powerful tool in the development of new multivariate calibration methods [24–27], wavelength selection and outlier detection [28, 29], multivariate figures of merit calculation [23, 24, 30], and the development of new spectral preprocessing methods [31–36]. All of the NAS-based methods share a similar idea, which is extracting a part of the signal that is directly related to the concentration of the analyte of interest, and which is hence useful for prediction purposes [4, 24]. Applications to complex samples of biomedical and pharmaceutical origin have recently been described [27, 29, 31, 32, 35], including a case in which NAS calculations enabled a significant reduction in the number of calibration samples required [29]. By tacking the Euclidean norm of the multivariate NAS vector, one obtains the analogue of the univariate scalar signal, which is directly related to the analyte concentration. In this way, the multivariate model is represented as a univariate calibration graph [37].

In this work, we used the concept of NAS calculations to represent a first-order multivariate SAM as a univariate standard addition graph. In this method, called "SANAS," the NAS vectors of the absorption spectra of an unknown solution, spiked with various amounts of the standard analytes, were converted to a scalar signal and then plotted against the concentration of the added standards. By using the resulting plot, we also calculated the figures of merit. The method was applied to the spectrophotometric simulated data as well as analysis of two indicators in synthetic matrices. For comparison, the standard addition spectrophotometric data were also analyzed by the rank annihilation factor analysis (RAFA) method [38]. In this method, called "SARAF," the concentration of the analyte in the unknown sample was iteratively changed until the rank of the original data matrix was annihilated by one.

Theory

Net analyte signal

In the current presentation, the data follow the Beer–Lambert law for spectroscopic (linear and additive) signals. NAS for an analyte is defined as the part of its spectrum that is orthogonal to the spectral space of the other sample components [23]. The theory of NAS calculations has been described in detail elsewhere. The NAS vector of the kth analyte in a multicomponent mixture ($\mathbf{r}_k^*$) can be computed by finding the orthogonal part of its spectrum ($\mathbf{r}_k$) with respect to the contribution of all other coexisting constituents:

$$\mathbf{r}_k^* = \left(\mathbf{I} - \mathbf{R}_{-k}\mathbf{R}_{-k}^+\right)\mathbf{r}_k, \tag{1}$$

where $\mathbf{R}_{-k}$ represents the matrix of spectra with all components in the mixture except the kth component, $\mathbf{R}_{-k}^+$ represents the Moore–Penrose pseudo inverse of $\mathbf{R}_{-k}$, and $\mathbf{I}$ is an identity matrix having the same dimension as $\mathbf{R}_{-k}\mathbf{R}_{-k}^+$. The spectral variation caused by fluctuation of instrumental and environmental conditions is also included in the $\mathbf{R}_{-k}$ matrix. The matrix $(\mathbf{I} - \mathbf{R}_{-k}\mathbf{R}_{-k}^+)$ is a projection matrix that projects $\mathbf{r}_k$ onto the null space of the rows of $\mathbf{R}_{-k}$, which is the orthogonal complement of the column space of $\mathbf{R}_{-k}$. In the classical calibration, $\mathbf{R}_{-k}$ is simply considered as pure spectra of all coexisting components, except the analyte [23]. Several approaches have also been proposed to obtain $\mathbf{R}_{-k}$ and to construct the projection matrix to use in a practical inverse calibration. Lorber et al. [24], Berger et al. [25], Xu and Schechter [26], and Goicoechea and Olivieri [27] used different algorithms to find the $\mathbf{R}_{-k}$ matrix. Here, we use the inverse approach of Lorber et al. [24] based on RAFA for computation of $\mathbf{R}_{-k}$. In this method, the absorbance data matrix of calibration mixtures ($\mathbf{R}$) is reproduced by principal components analysis (PCA), or PLS using f significant principal components, yielding $\mathbf{R}_{\text{reb}}$. Then a rank annihilation step in the f-dimensional space is used to find the part of the original matrix spanned by the interferences:

$$\mathbf{R}_{-k} = \mathbf{R}_{\text{reb}} - \alpha\widehat{\mathbf{c}}_k\mathbf{r}^{\text{T}}, \tag{2}$$

where $\widehat{\mathbf{c}}_k$ is the projection of the vector of analyte concentration $\mathbf{c}_k$ onto the f-dimensional subspace and is calculated by $\widehat{\mathbf{c}}_k = \mathbf{R}_{\text{reb}}\mathbf{R}_{\text{reb}}^+\mathbf{c}_k$. The vector $\mathbf{r}$ is related to the pure spectrum of the kth analyte or a linear combination of the rows of $\mathbf{R}$, which is chosen to include a contribution from the spectrum of the kth analyte. Although any reasonable spectrum can be used for this purpose, it is recommended to use a spectrum that contains maximal information on the analyte; therefore, the pure spectrum of the kth analyte is the best choice. The scalar α can be calculated as

$$\alpha = 1/\mathbf{r}^{\text{T}}\mathbf{R}_{\text{reb}}^+\widehat{\mathbf{c}}_k. \tag{3}$$

Standard addition by NAS (SANAS)

Consider that the analysis of an unknown sample containing p different analytes is requested. To do standard addition, a calibration set of standard mixtures of the analyte, the same as those used in simple multivariate calibration, is provided. If n standard solutions are used, the concentration data are collected in a matrix $\mathbf{C}_s$ ($n \times p$). These solutions are prepared in a series of volumetric flasks containing the same amounts of the unknown sample. The absorption spectra of the resulting solutions are recorded, digitized, and then collected in a standard addition data matrix, $\mathbf{R}_{\text{sa}}$ ($n \times m$), where m is the number of absorbance readings per spectrum. A solution containing only the unknown sample, without added standards, is also provided and its absorption spectrum is collected in a row vector, $\mathbf{r}_{\text{un}}$. To calculate $\mathbf{R}_{-k}$ by Eq. 2, the concentration of the analyte of interest in the solutions is required. The concentration of the analyte of interest in each standard addition solution (c_k) is equal to the concentration of the analyte in the unknown sample (c_{uk}) plus the concentration of the added standard (c_{sk}), so $c_k = c_{sk} + c_{uk}$. Since c_{uk} is not available, calculation of $\mathbf{R}_{-k}$ by Eq. 2 is not straightforward. The second way to calculated $\mathbf{R}_{-k}$ is using the pure spectrum ($\mathbf{s}_k$) of the kth analyte; however, because of the matrix effect, $\mathbf{s}_k$ cannot be calculated correctly and therefore this method is also not applicable.

In attempts to find a way to calculate $\mathbf{R}_{-k}$, the following invention was made for this article. Each row vector of $\mathbf{R}_{\text{sa}}$ [i.e., $\mathbf{R}_{\text{sa}}(i)$] can be considered as the sum of the absorption spectrum of the unknown sample and that of the standard mixture added to the unknown sample:

$$\mathbf{R}_{\text{sa}}(i) = \mathbf{r}_{\text{u}} + \mathbf{R}_{\text{sm}}(i), \tag{4}$$

where $\mathbf{R}_{\text{sm}}$ contains the absorption spectra of the calibration set standards in the presence of the matrix effect. This is shown graphically in Fig. 1. Therefore, by subtracting the unknown absorption ($\mathbf{r}_{\text{u}}$) from $\mathbf{R}_{\text{sa}}$, we can obtain the absorption matrix of the standards only ($\mathbf{R}_{\text{sm}}$) that are affected by the unknown matrix. The subscript "sm" is used to denote both the standard solution and the matrix effect. $\mathbf{R}_{\text{sm}}$ can be decomposed to the concentration and pure spectrum data matrices according to the Beer–Lambert law:

$$\mathbf{R}_{\text{sm}} = \mathbf{C}_s \mathbf{S}_m, \tag{5}$$

where the rows of $\mathbf{S}_m$ are the pure spectra of the analytes affected by the unknown's matrix. Since the concentrations

of standards producing $\mathbf{R}_{sm}$ are known, this matrix can be used for calculation of $\mathbf{R}_{-k}$. To do so, $\mathbf{R}_{sm}$ is subjected to PCA or PLS analysis and then is reproduced by using the first f significant principal components to produce $\mathbf{R}_{sa,reb}$, and then $\mathbf{R}_{-k}$ is simply calculated using Eq. 2.

Consequently, the NAS vectors of the kth analyte in the standard addition samples are calculated by replacing $\mathbf{r}_k$ in Eq. 1 by $\mathbf{R}_{sa,reb}$:

$$\mathbf{R}_{sa,k}^* = \left(\mathbf{I} - \mathbf{R}_{-k}\mathbf{R}_{-k}^+\right)\mathbf{R}_{sa,reb}, \tag{6}$$

where $\mathbf{R}_{sa,reb}$ is the $\mathbf{R}_{sa}$ (matrix of standard addition data) that is rebuilt with the appropriate number of factors. The row vectors of $\mathbf{R}_{sa,k}^*$ represent a direct relationship with the concentration of the kth analyte in standard addition samples (the summation of unknown and added standard concentrations). By plotting the Euclidean norm of these vectors $\left(\left\|\mathbf{R}_{sa,k}^*\right\|\right)$ against the concentration of the added standards of the kth analyte, one obtains the standard addition graph (Fig. 1).

Standard addition by RAFA (SARAF)

Generally, the rank of a nonsquare matrix can be reduced using Weddeburn's formula [39]. RAFA is another rank reduction method [38], and enables one to determine the analyte of interest in the presence of unknown interferences using a single calibration sample. SARAF and SANAS have similar theory and therefore they have many common steps. Equation 2 for calculating $\mathbf{R}_{-k}$ was derived by Lorber et al. [24] on the basis of a noniterative inverse RAFA. In this equation, if $\mathbf{c}_k$ has been selected correctly, $\mathbf{R}_{-k}$ will have one rank lower than that of the original data matrix. In the SAM, owing to the unknown value of the analyte's

concentration in the unknown sample, $\mathbf{c}_k$ is not available for $\mathbf{R}_{sa}$. As noted previously, $c_k = c_{sk} + c_{uk}$. Thus, to determine the correct value of $\mathbf{c}_k$, c_{uk} is iteratively varied and in each iteration $\mathbf{c}_k$ and then $\mathbf{R}_{-k}$ are calculated:

$$\mathbf{R}_{-k} = \mathbf{R}_{reb} - \alpha(c_{uk} + \mathbf{c}_{sk})\mathbf{r}^T. \tag{7}$$

In classical RAFA, the contribution of the analyte of interest in the absorbance data matrix ($\mathbf{R}_k$) is annihilated directly. According to the Beer–Lambert law, $\mathbf{R}_k$ can be calculated as

$$\mathbf{R}_k = \mathbf{c}_k \mathbf{s}_{km}^T = \left(c_{uk} + \mathbf{c}_{sk}\right)\mathbf{s}_{km}^T, \tag{8}$$

where $\mathbf{s}_{km}$ is the sensitivity vector of the kth analyte affected by the unknown's matrix, which can be calculated by least-squares regression of Eq. 5:

$$\mathbf{S}_m = \mathbf{C}_s^+ \mathbf{R}_{sm}. \tag{9}$$

It should be noted that the calculated $\mathbf{S}_m$ is the same for $\mathbf{R}_{sm}$ and $\mathbf{R}_{sa}$. Therefore, the annihilation of the contribution of the kth analyte from the standard addition absorbance data matrix can be easily achieved:

$$\mathbf{R}_{-k} = \mathbf{R}_{sa} - \left[(c_{uk} + \mathbf{c}_{sk})\mathbf{s}_{km}^T\right]. \tag{10}$$

Only at one specified value of c_{uk} the rank of $\mathbf{R}_{-k}$ is annihilated by one, which is the correct value of the analyte concentration in the unknown sample. PCA can be used to determine at what value of c_{uk} rank annihilation has occurred. The use of relative standard deviation (RSD) is a common method applicable for determination of the number of principal components [40]:

$$\mathrm{RSD}(f) = \sqrt{\left(\sum_{i=f+1}^{n} e_i\right) \bigg/ [m(n-f)]}, \tag{11}$$

where e_i is the ith eigenvalue, and n, m, and f are the same as defined previously. At the correct value of c_{uk}, the rank of $\mathbf{R}_{-k}$ is $f-1$ and therefore its fth RSD is at the minimum value.

Experimental

Instrumentation

Electronic absorption measurements were carried out using an Ultraspec 4000 spectrophotometer (Phamacia Biotech, UK), equipped with 10.0-mm quartz cells. Absorbance data were collected using the Swift(II) software program of the instrument, and were then transferred to a Pentium IV personal computer with the Windows XP operating system for subsequent manipulation. All necessary programs were written in a MATLAB (version 7, The MathWorks)
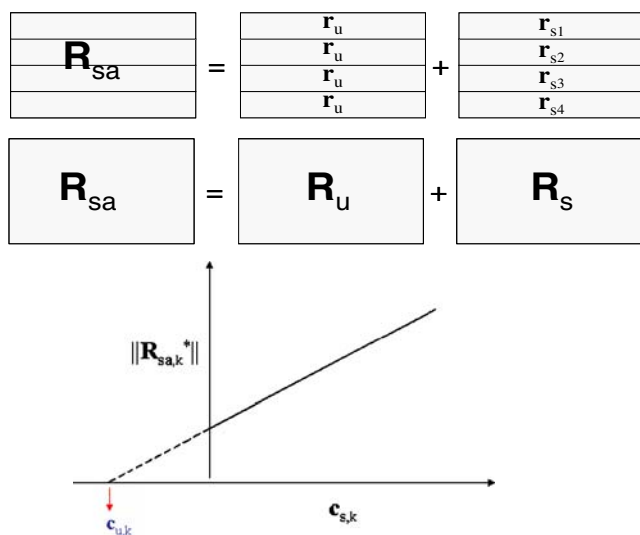

Fig. 1 The SANAS method

environment. pH measurements were performed with a Metrohm 707 (Switzerland) pH meter equipped with a combined glass electrode.

Reagents

All experiments were performed with analytical reagent grade chemicals. Stock solutions of alizarin red S (ARS) (0.01 mol L^{-1}), Bengal rose B (BRB) (0.01 mol L^{-1}), cetyltrimethyl ammonium bromide (CTAB) (0.01 mol L^{-1}), and phosphate buffer of pH 7 (0.05 mol L^{-1}) were prepared by dissolving suitable amounts of the compounds in doubly distilled water. The standard solutions were obtained daily from these solutions by appropriate dilutions.

Simulated data set

Spectra were simulated as a sum of Gaussian functions, calculated at constant intervals (200 channels). Ternary mixtures of three analytes (A1, A2, and A3) were assumed. Figure 2 shows the generated pure absorption spectra of the analytes in pure form and in the presence of the matrix effect. Many unknown samples with randomly generated concentration combinations, which were evenly distributed in the range 0.0–1.0 with a mean of 0.5 and unit standard deviation, were used. In addition, a calibration set composed of nine standard samples was also generated. The response data were calculated by postmultiplication of the

concentration by pure spectra considering the signal additive principle. Normally distributed homoscedastic noise at a given level (standard deviation of absorbance unit) was added to the exact data.

Experimental data set

The unknown sample was a synthetic mixture of two indicators, ARS and BRB, in a complex matrix composed of high concentrations of CTAB (0.01 M) and phosphate salts (0.05 M, pH 7). Some binary mixtures of ARS and BRB with different concentrations were prepared in the synthesized matrix (Table 1). The linear dynamic range of each analyte was determined separately and the amounts of analytes in unknown samples and standards were selected in their corresponding linear ranges.

The calibration standard set was composed of nine randomly designed binary mixtures of indicators, where in the first sample the concentrations of both analytes were taken as zero (Table 2). These solutions were used for standard addition experiments. In the case of each unknown sample, to a series of nine volumetric flasks (10.0 mL) containing a constant amount of unknown sample (i.e., 1.0 mL) were added desired volumes of 0.01 M indicators (to reach to the final concentrations indicated in Table 2) and the mixtures were then diluted to the mark with doubly distilled water. The absorption spectra of the resulting solutions were recorded in the wavelength range from 300

Fig. 2 Pure spectra of three analytes in the simulated data set in pure form (*left*) and in the presence of a matrix effect (*right*)
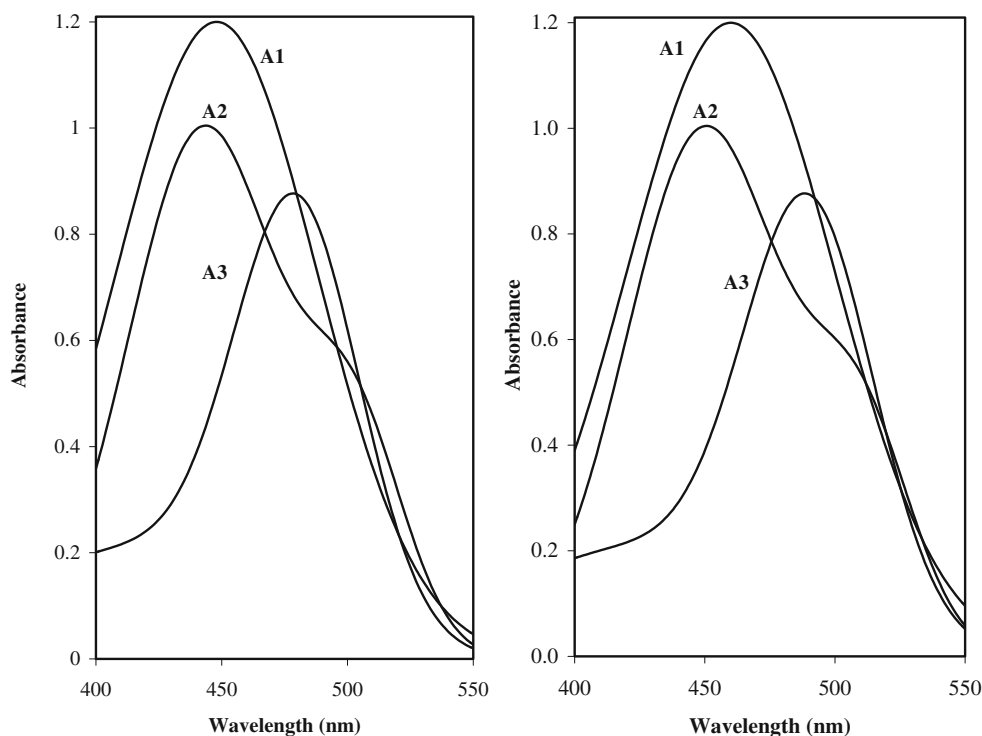
Table 1 Composition of unknown experimental samples

Sample	ARS concentration ($\times 10^5$ M)	BRB concentration ($\times 10^6$ M)
1	1.0	2.0
2	2.0	3.0
3	4.5	1.5
4	7.0	5.0
5	5.0	3.5
6	4.0	6.0

ARS alizarin red S, *BRB* Bengal rose B

to 700 nm at 2,500 nm/min scan rate. The absorption spectrum of each sample, digitized in 1-nm intervals, was a row of the $\mathbf{R}_{sa}$ matrix, whereas the first solution, containing no added standard, was selected as the unknown sample ($\mathbf{r}_u$). By subtracting $\mathbf{r}_u$ from every row of $\mathbf{R}_{sa}$, we obtained $\mathbf{R}_{sm}$.

The 450–650-nm wavelength interval was selected from the total recorded spectra to continue the calculations and analysis. The dimensions of available matrices for analysis were 9×201 ($\mathbf{R}_{sa}$), 9×201 ($\mathbf{R}_s$), 1×201 ($\mathbf{r}_u$), and 9×2 ($\mathbf{C}_s$).

Figures of merit

The selectivity, sensitivity, and limit of detection (LOD) in multivariate calibration can be calculated to study the quality of a given analytical method using the NAS concept [23]. If $\mathbf{R}_{sa,reb}$ in Eq. 6 is replaced by $\mathbf{R}_{sm,reb}$, the result will be the NAS of the calibration standards affected by the unknown matrix $\left(\mathbf{R}_{sm}^*\right)$. The variances of this matrix are directly related to the concentration of the analyte of interest in the calibration set:

$$\mathbf{R}_{sm}^* = \mathbf{c}_k \mathbf{s}_k^*, \tag{12}$$

where $\mathbf{s}_k^*$ is the net sensitivity vector of the kth analyte. The ratio of the norm of this vector to the norm of the pure spectrum gives a measure of multivariate selectivity:

$$\mathrm{SEL}_k = \frac{\|s_k^*\|}{\|s_k\|}. \tag{13}$$

The sensitivity for the kth analyte was calculated using the following equation [23]:

$$\mathrm{SEN}_k = \|s_k^*\|. \tag{14}$$

We can also calculate the LOD, as another important figure of merit, using the following equation [4]:

$$\mathrm{LOD}_k = \frac{3\|\varepsilon\|}{\|s_k^*\|}, \tag{15}$$

where ε is a measure of the instrumental noise at different wavelengths. The value of $\|\varepsilon\|$ was estimated, in turn, by registering spectra for several blank samples and calculating the norm of the NAS for each sample and the corresponding standard deviation. The latter was taken as an approximation to $\|\varepsilon\|$.
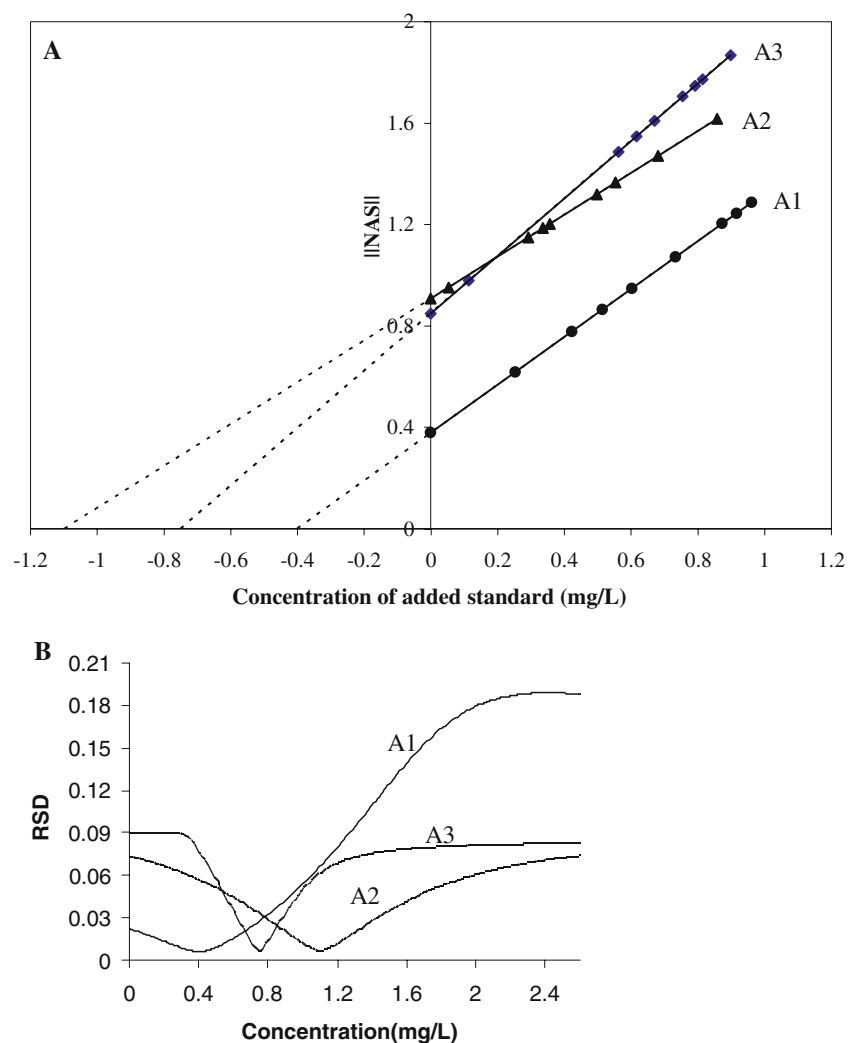
Results and discussion

Simulation data

Many synthetic unknown samples containing three analytes were examined. The hypothetical spectra of the compounds in pure form and in the presence of matrix effects are shown in Fig. 2. Obviously, the absorption spectra were assumed to be highly overlapped so that those of analytes A2 and A3 are overlaid under the spectrum of A1. To study the noise effect, different amounts of homoscedastic noise (random error with standard deviation of absorbance unit)

Table 2 Composition of calibration standards added to each of the unknown experimental samples

Sample	ARS concentration ($\times 10^5$ M)	BRB concentration ($\times 10^6$ M)
1	0.0	0.0
2	1.0	10.0
3	4.0	7.0
4	9.0	4.0
5	2.0	5.0
6	5.0	6.0
7	7.0	1.0
8	10.0	2.0
9	3.0	8.0

were added to both $\mathbf{r}_u$ and $\mathbf{R}_s$ before constructing $\mathbf{R}_{sa}$. The resultant standard addition plot obtained by the SANAS method in the presence of noise (0.02 standard deviation of the signal) for one of the hypothetical unknown mixtures (with the composition A1 0.40, A2 1.10, and A3 0.75) is represented in Fig. 3a. Obviously, linear standard addition plots similar to those of univariate standard addition were obtained. The intersection of the linear calibration graph and the concentration axes gives accurately the negative sign of the analyte's concentration in the unknown mixture. Table 3 lists the results for analysis of the hypothetical unknown mixture in the presence of different added noise levels. .The simulation was done 1,000 times with the same conditions (noise level, peak positions of analyte concentration). The mean of 1,000 prediction values is reported together with their standard deviation in parentheses. As observed, approximately accurate and precise results were obtained even in the presence of noise, with a standard deviation of 0.005 to 0.02. Similar results were obtained for analysis of other hypothetical unknown mixtures.

The simulated standard addition data sets were also analyzed by the SARAF method. The variation of the RSD against the analyte concentration for the unknown sample mentioned above is represented in Fig. 3b. As observed, the RSD passes through a distinct minimum at the actual concentration of the analyte in the unknown sample. The results obtained at different noise levels (Table 3) confirm the good accuracy and precision of the SARAF since even in the presence of 0.02 noise levels, the analyzed and actual values are completely matched for all analytes. The same results were obtained for analyses of other unknown samples.

Experimental data

In this section the result of applying two proposed multivariate SAMs to the spectrophotometric data of the binary mixtures of two indicators, i.e., ARS and BRB, in the presence of a significant matrix effect will be discussed. As observed from Fig. 4, the absorption spectra of the

Fig. 3 **a** SANAS and **b** SARAF plots for the analysis of unknown mixture simulated data with actual concentrations of A1 (0.40), A2 (1.10), and A3 (0.75). *NAS* net analyte signal, *RSD* relative standard deviation

Table 3 Mean value and standard deviation of the predicted concentrations obtained by applying SANAS and SARAF for the 1,000 repeatedly simulated data containing three analytes A1, A2, and A3 with concentrations of 0.40, 1.10, and 0.75, respectively

Standard deviation of random noise	SANAS			SARAF		
	A1	A2	A3	A1	A2	A3
0	0.400 (±0.000)	1.100 (±0.000)	0.750 (±0.000)	0.400 (±0.000)	1.100 (±0.000)	0.750 (±0.000)
0.005	0.411 (±0.017)	1.083 (± 0.022)	0.747 (±0.007)	0.421 (±0.030)	1.032 (±0.027)	0.741 (±0.006)
0.01	0.431 (±0.029)	1.054 (±0.039)	0.746 (±0.012)	0.436 (±0.012)	1.140 (±0.078)	0.760 (±0.025)
0.02	0.438 (±0.042)	1.024 (±0.056)	0.768 (±0.020)	0.441 (±0.072)	1.190 (±0.069)	0.771 (±0.028)

The values in *parentheses* are the standard deviations of the 1,000-times simulations

indicators represent a clear shift in going from pure water to the synthetic unknown matrix. The unknown matrix not only changed the peak position of the indicator's spectrum but it also affected the peak intensity. As shown in Fig. 4, the unknown matrix shifted the absorption spectra of both indicators to higher wavelengths. An increase in absorbance of ARS and a decrease in absorbance of BRB are observed in the presence of the unknown matrix. Owing to high overlap between the absorption spectra of the indicators on one hand and significant matrix effects on the other hand, the simultaneous spectrophotometric determination of these indicators in this matrix is not feasible by the conventional univariate SAM. On the other hand, such analysis will be feasible by using the SAM combined with multivariate calibration, i.e., SANAS and SARAF methods. To examine the performances of the proposed methods, some binary synthetic unknown mixtures of the indicators (Table 1) were analyzed. For the SAM, in the case of each unknown sample, nine randomly designed standard solutions (Table 2) containing the same amount of unknown sample were prepared and their absorption spectra were recorded.
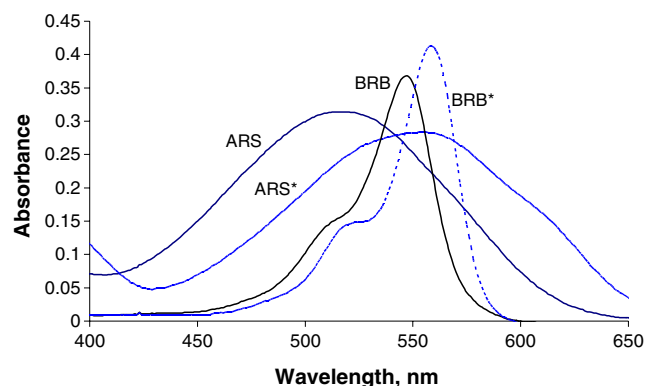

Fig. 4 Electronic absorption spectra of 5×10^{-5} M alizarin red S (*ARS*) and 5×10^{-6} M Bengal rose B (*BRB*) in pure form and in the presence of a matrix effect. The latter are indicated by *asterisks*

The number of principal components used for calculating $\mathbf{R}_{sa,reb}$ from $\mathbf{R}_{sa}$ was obtained using leave-one-out cross-validation. For all unknown samples, optimum results were obtained by using two principal components for both analytes. The resultant SANAS plot is represented in Fig. 5 for one unknown sample (i.e., sample 2 in Table 1), and the predicted concentrations of the indicators in all unknown samples by the SANAS method are listed in Table 4. The plots shown in Fig. 5 clearly indicate the success of the proposed SANAS method in converting second-order data of experimental samples to univariate order so the data can be easily seen and interpreted. On the other hand, a close agreement between the actual and predicted values of the concentrations of both analytes in unknown samples is found. The relative errors are in a reasonable range. All the predicted ARS concentrations have a relative error with an absolute value almost smaller than 9% and in most cases lower than 4%. The predicted concentrations of BRB represent relative errors smaller than 4.4%, except for samples 1 and 3, which have relative errors of 6.78 and 9.07%, respectively.

The results of the analysis of the multivariate standard addition data by the SARAF method are also represented in Table 4. The corresponding SARAF plots for unknown sample 4 (Tables 1 and 4) are shown in Fig. 6. As observed, the SARAF plots reached a distinct minimum at a concentration equal to that in the unknown sample. Similar observations were obtained for other unknown samples. The relative errors of prediction obtained by the SARAF method (Table 4) confirm that this proposed method has a very good prediction ability for unknown concentrations especially for BRB. The relative errors are lower than 5% in most cases. A comparison between the results obtained using SANAS and SARAF shows that both methods gave comparable results, whereas the latter produced better results in some cases.

A comparison was also made between the methods proposed in this article and the well-known GSAM proposed by Saxberg and Kowalski [14]. In this case, PLS regression was used to model the relationship between

Fig. 5 SANAS plots of unknown sample 2 with actual concentrations of 2×10^{-5} M ARS and 3×10^{-6} M BRB
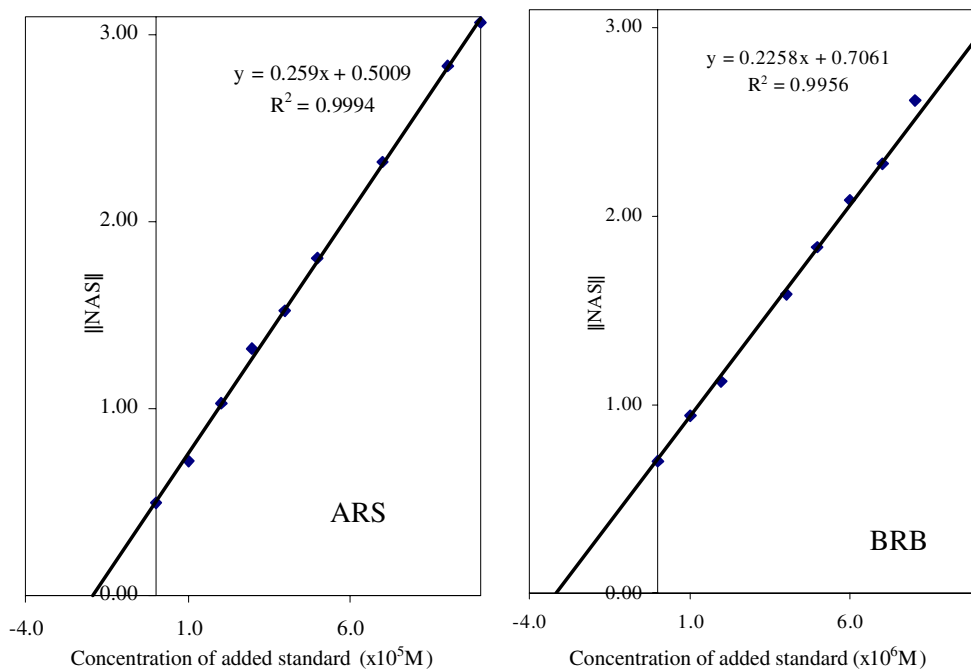
R_{sm} and c_k and then the calculated regression coefficients were premultiplied by R_{sa} to calculate the total analyte concentrations in the standard addition samples. The analyte's concentration in the unknown sample was calculated by subtracting the standard concentrations from the calculated concentrations in the previous step. The results are represented in the last four columns of Table 4. It is clearly observed that the GSAM produced underestimated values in comparison with SARAF in most cases, whereas comparable relative prediction errors were obtained by SANAS and the GSAM in most cases and better results were obtained by SANAS in some cases.

Thus, it could be said that SARAF resulted in almost better predictions. However, it is difficult to prefer SANAS over the GSAM for prediction error, but SANAS is a simple method creating a visual standard addition plot.

The lower prediction errors obtained using SARAF can be attributed to the limited numbers of calculation steps without the need to obtain a projection matrix as is used in SARAF (Eqs. 1 and 6). In addition, the objective function in SARAF is matrix rank and the rank annihilated matrix (R_{-k}) is simply calculated by matrix subtraction (Eq. 10). Moreover, random noises cannot significantly affect the rank determination by PCA. The RSD value calculated by

Table 4 The predicted concentration of indicators in unknown samples and the percent relative errors obtained by three different multivariate standard addition methods

Sample	Actual values		SANAS				SARAF				GSAM			
			Predicted		Relative error		Predicted		Relative error		Predicted		Relative error	
	ARS	BRB	ARS	BRB	ARS	BRB	ARS	BRB	ARS	BRB	ARS	BRB	ARS	BRB
1	1.0	2.0	0.90	2.14	−9.77	6.78	1.01	2.11	1.0	5.5	0.86	2.10	−14.2	5.0
2	2.0	3.0	1.93	3.12	−3.69	4.09	2.00	3.08	0.2	2.7	1.92	3.09	−4.0	2.8
3	4.5	1.5	4.40	1.64	−2.26	9.07	4.21	1.53	−6.4	2.0	4.37	1.61	−2.8	7.7
4	7.0	5.0	7.22	5.01	3.14	0.16	7.11	4.88	1.6	−2.4	7.17	4.97	2.4	−0.7
5	5.0	3.5	5.24	3.59	4.86	2.67	5.16	3.52	3.2	0.6	5.24	3.55	4.7	1.6
6	4.5	8.0	4.42	7.95	−1.88	−0.57	4.11	7.95	−8.7	−0.6	4.07	8.00	−9.5	0.1

The concentration unit of ARS is $\times10^5$ M and that of BRB is $\times10^6$ M
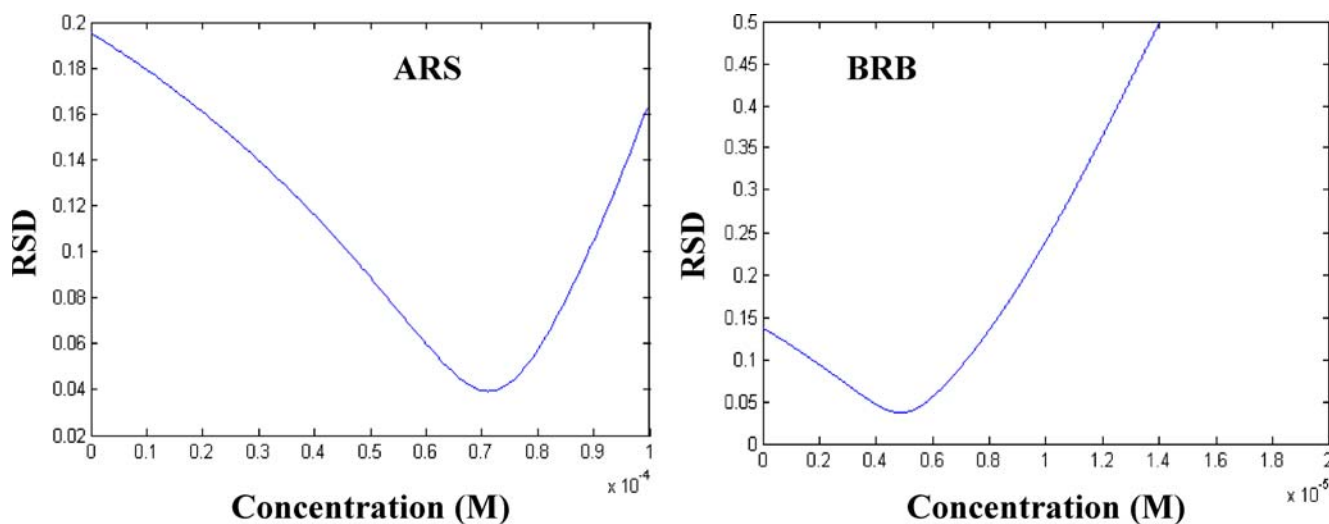
GSAM generalized standard addition method

Fig. 6 SARAF plots of unknown sample 4 with concentrations of 7×10^{-5}M ARS and 5×10^{-6}M BRB

Eq. 11 is a measure of random error in the data and in rank annihilation analysis it reaches a minimum value (at the correct value of c_{un}) irrespective of the random noise level [38]. In SANAS, errors in $\mathbf{R}_{sa}$ are propagated to the NAS vectors of analyte k $\left(\mathbf{R}_{sa,k}^{*}\right)$, which are subsequently propagated to the NAS standard addition plot. In other words, the slope and intercept of the SANAS plot are sensitive to the random errors or the presence of outliers. Hence, the values predicted by SANAS are more affected by random (or even systematic) errors with respect to SARAF. A similar discussion can be given for the lower prediction ability of the GSAM compared with SARAF. Actually, the random (or systematic) errors in $\mathbf{R}_{sm}$ are propagated to the regression coefficients obtained by the least-squares methods (such as principal component regression and PLS), which are then propagated to the predicted concentration of unknown samples using the calculated regression coefficients. As a conclusion, since the concentration of the unknown sample determined by SARAF is not calculated by regression coefficients or regression lines, it is less affected by the error propagation from the original data. However, this should be examined more by application of these methods to different samples using different analytical data (e.g., fluorescence, voltammetry, chromatography).

One of the well-known applications of NAS in multivariate calibration is the calculation of figures of merit.

Table 5 Figures of merit for analysis of indicators

	Selectivity	Sensitivity (M^{-1})
ARS	0.624	2.62×10^{4}
BRB	0.623	2.33×10^{5}

After calculating $\mathbf{R}_{sm}$, which contained spectral information on nine calibration mixtures of both indicators in the presence of matrix effects, we calculated the NAS vectors of these mixtures $\left(\mathbf{R}_{sm}^{*}\right)$ for each analyte by replacing $\mathbf{R}_{sa,reb}$ in Eq. 6 by $\mathbf{R}_{sm,reb}$. Then analytical figures of merits (i.e., selectivity, sensitivity) were calculated using Eqs. 12–15. The results are listed in Table 5. As observed, for both indicators similar selectivity was obtained.

Conclusion

By application of NAS calculations to the multivariate standard addition data, the two-dimensional array of absorbance data was successfully converted to a one-dimensional array, similar to the univariate SAM. The resultant SANAS plot allowed us to calculate the concentration of interfering analytes in the unknown samples with a significant indirect matrix effect. Also, the proposed method was used to calculate the analytical figure of merit in multivariate calibration. The prediction results for both simulated and experimental data confirmed the power of the SANAS method in analyzing multivariate standard addition data. On the other hand, better results were obtained when the data were analyzed by the SARAF method, a combination of the SAM and the RAFA method. In spite of the higher accuracy of the SARAF method, it cannot produce a visualized standard addition graph. Although the proposed methods produced results comparable with those of the well-known GSAM, they have the advantages of simplicity and data visualization ability.

Owing to the order-reduction ability of the NAS calculations and some limitations of the three-way SAMs, the proposed SANAS method may also be used for analyzing second-order data and converting them to simple

univariate calibration graphs. Study of this topic is in progress in our research group.

References

1. Wold S, Martens H, Wold H (1983) Lect Notes Math 973:286–293
2. Beebe KR, Kowalski BR (1987) Anal Chem 59:1007A–1017A
3. Martens H, Naes T (1992) Multivariate calibration. Wiley, Chichester
4. Booksh KS, Kowalski BR (1994) Anal Chem 66:782A–791A
5. Galuska AA, Morrison GH (1984) Anal Chem 56:74–77
6. Powley CR, George SW, Ryan TW, Buck RC (2005) Anal Chem 77:6353–6358
7. Zhang TM, Yuan B, Liang YZ, Cao J, Pan CX, Ying B, Lu DY (2007) Anal Sci 23:581–587
8. Larsen NR, Hansen EH (1976) Anal Chim Acta 84:31–36
9. Franke JP, Zeeuw RAD, Hakkert R (1978) Anal Chem 50:1374–1380
10. Janatsch G, Kruse-Jarres JD, Marbach R, Heise HM (1989) Anal Chem 61:2016–2023
11. Haaland DM, Robinson MR, Koepp GW, Thomas EV, Eaton RP (1992) Appl Spectrosc 46:1575–1578
12. Isaksson T, Miller CE, Næs T (1992) Appl Spectrosc 46:1685–1694
13. Pataca LCM, Neto WB, Marcucci MC, Poppi RJ (2007) Talanta 71:1926–1931
14. Saxberg BEH, Kowalski BR (1979) Anal Chem 51:1031–1038
15. Gerlach RW, Kowalski BR (1982) Anal Chim Acta 134:119–127
16. Bechmann IE, Norgaard L, Ridder C (1995) Anal Chim Acta 304:229–236
17. Osten DW, Kowalski BR (1985) Anal Chem 57:908–917
18. Xie Y, Wang J, Liang Y, Ge K, Yu R (1993) Anal Chim Acta 281:207–218
19. Sena MM, Trevisan MG, Poppi RJ (2006) Talanta 68:1707–1712
20. Comas E, Gimeno RA, Ferre J, Marce RM, Borrull F, Rius FX (2003) J Chromatogr A 988:277–284
21. Booksh K, Henshaw JM, Burgess LW, Kowalski BR (1995) J Chemom 9:263–282
22. Wu HL, Yu RQ, Shibukawa M, Oguma K (2000) Anal Sci 16:217–220
23. Lorber A (1986) Anal Chem 58:1167–1172
24. Lorber A, Faber K, Kowalski BR (1997) Anal Chem 69:1620–1626
25. Berger J, Koo TW, Itzkan I, Feld MS (1998) Anal Chem 70:623–627
26. Xu L, Schechter I (1997) Anal Chem 69:3722–3730
27. Goicoechea HC, Olivieri AC (1999) Anal Chem 71:4361–4368
28. Ferre J, Rius FX (1998) Anal Chem 70:1999–2007
29. Goicoechea HC, Olivieri AC (1999) Analyst 124:725–731
30. Faber NM, Ferre J, Boque R, Kalivas JH (2003) Trends Anal Chem 22:352–361
31. Goicoechea HC, Olivieri AC (2001) Chemom Intell Lab Syst 56:73–81
32. Marsili NR, Sobrero MS, Goicoechea HC (2003) Anal Bioanal Chem 376:126–133
33. Truyols GV, Torres-Lapasio JR, Garcya-Alvarez-Coque MC (2003) J Chromatogr A 991:47–59
34. Hemmateenejad B, Ghavami R, Miri R, Shamsipur M (2006) Talanta 68:1222–1229
35. Dela Pena AM, Mansilla AE, Valenzuela MIAA, Goicoechea HC, Olivieri AC (2002) Anal Chim Acta 463:75–88
36. Hemmateenejad B, Safarpour MA, Mehranpour A (2005) Anal Chim Acta 535:275–285
37. Faber NM (2000) Chemom Intell Lab Syst 50:107–114
38. Ho CN, Christian CD, Davidson ER (1978) Anal Chem 50:1108–1113
39. Ferre J, Faber NM (2003) J Chemom 17:603–607
40. Elbergali J, Kubista M (1999) Anal Chim Acta 379:143–158