Available online at www.sciencedirect.com
ScienceDirect
Talanta 74 (2008) 1201–1210
Talanta
www.elsevier.com/locate/talanta

Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples

T. Azzouz, R. Tauler*

Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18, Barcelona 08034, Spain

Received 13 April 2007; received in revised form 13 August 2007; accepted 24 August 2007
Available online 30 August 2007

Abstract

Application of multivariate curve resolution alternating least squares (MCR-ALS), for the resolution and quantification of different analytes in different type of pharmaceutical and agricultural samples is shown. In particular, MCR-ALS is applied first to the UV spectrophotometric quantitative analysis of mixtures of commercial steroid drugs, and second to the near-infrared (NIR) spectrophotometric quantitative analysis of humidity and protein contents in forage cereal samples. Quantitative results obtained by MCR-ALS are compared to those obtained using the well established partial least squares regression (PLSR) multivariate calibration method.

Keywords: Forage analysis; NIR spectroscopy; MCR-ALS; PLSR

1. Introduction

In this work, direct quantitative spectrophotometric determination of mixtures of analytes in two types of samples is investigated. First, mixtures of pharmaceutical products in drug samples are analyzed by UV–vis spectrophotometry. This technique is a rapid and inexpensive analytical technique and as such is highly suitable for control analyses of pharmaceutical preparations. However, the lack of selectivity of UV–vis absorption measurements hinders its general application in the presence of strongly overlapped absorption bands of the different sample components. Pharmaceutical preparations are usually mixtures of the active principles and of various excipients absorbing in the same spectral region. The development of multivariate calibration methods based on the mathematical resolution of multivariate signals can allow their rapid resolution and quantification. The second type of samples studied in this work are forage (cereal) agricultural samples analyzed by near-infrared (NIR) spectroscopy. This technique has gained wider acceptance in different fields, due to its advantages over other analytical spectroscopic techniques. One of the best advantages of NIR spectroscopy is the possibility of working in reflectance mode and the use of fibre optical probe modules easily coupled to the spectrophotometer. This allows measurements of solid and liquid samples with little sample pretreatment, the implementation of continuous methodologies, the fast acquisition of spectra and the prediction of both physical and chemical parameters from the same sample. Near-infrared spectrophotometry (NIRS) is a non-destructive technique, very fast and easy to implement, without needing reagents and without wastes produced. Once calibrated, NIRS is simple to operate and it is well suited for the determination of the major components in many types of samples, such as protein and water contents in food and forage samples. Multivariate calibration methods like partial least squares regression (PLSR) have been frequently used to extract analytical information from UV–vis and NIR spectra.

Since in general it is very difficult to have completely selective analytical signals for every analyte of interest in a multicomponent sample, their physical separation by chromatographic methods or by any other analytical separation methods, or their mathematical resolution using chemometric methods is a preliminary step necessary for their quantitative determination, especially in the analysis of complex natural samples. Apart from the well known multivariate calibration methods

* Corresponding author.
E-mail address: rtaqam@iiqab.csic.es (R. Tauler).

0039-9140/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.talanta.2007.08.024

like PLSR [1,2] other chemometric methods exist that allow for the direct mathematical analysis of the different components in evolving mixture systems (*Evolving Factor Analysis*, EFA) [3], for the detection of the more selective variables (SIMPLISMA [4–6]) or for the resolution of the components simultaneously present in a particular data window (*Window Factor Analysis method* [7]). The multivariate curve resolution alternating least squares (MCR-ALS) method proposed in this work [8–13] has been shown to provide an improved resolution compared to other methods and to allow quantitative determinations in the analysis of complex mixtures using spectroscopic means. MCR-ALS has been applied to the study of complex industrial evolving processes [14], to the investigation of multiequilibria systems using spectroscopic titrations (fluorescence, UV–vis absorption, etc.) [15], to the resolution of multiple coeluted peaks in chromatography [10], to the resolution and quantification of mixtures in flow injection analysis [16], to the resolution of the different components in kinetic reactions and processes [17], to the resolution of spectroscopic images [18], to multidimensional spectroscopy [19], to electrophoretic studies of amino acids [20], to voltammetric studies of metal complexes [21–23], to studies of conformational changes of polynucleotide [24] and protein folding processes [25], and to the resolution and apportionment of environmental sources of contamination [26]. In this work, the use of MCR-ALS is proposed for the quantitative determination of mixtures of analytes using first order spectrophotometric data (UV–vis and NIR absorption spectrophotometric data). A *correlation constraint* introduced in a previous work for the analysis of mixtures of metal ions analyzed by voltammetry [22] is extended in this work to establish alternating least squares (ALS) multivariate calibration models for the quantitative determination of analyte mixtures using UV and NIR spectrophotometric data. The results obtained using MCR-ALS with this new correlation constraint are then compared to those obtained using the nowadays well established PLSR multivariate calibration [37].

2. Experimental

2.1. Reagents and solutions

The following reagents and solutions were used:

- Acetonitril (Carlo Erba) for HPLC.
- Methanol (Panreac) was used for the synthetic preparation of standard, synthetic mixture and drug samples.
- Etinilestradiol (Sigma), minimum 98% (HPLC).
- Levonorgestrel (Sigma), minimum 98% (HPLC).

2.2. Pharmaceutical products

Concentrated stock solutions of etinilestradiol and levonorgestrel were prepared in methanol. From these stock solutions, 25 synthetic mixtures were prepared from which 15 samples were used as a calibration data set, and the remaining samples were used as external validation data set. The concentration range of etinilestradiol was between 3 and 31 mg/l, and for levonorgestrel between 3 and 20 mg/l. Etinilestradiol and levonorgestrel were analyzed in commercial drugs: Microgynon, Neogynona and Triagynon (ochre and brown color). From 30 original drug samples, 20 were used as a calibration data set, and the remaining samples were used as external validation data set.

The procedure used to prepare the drug samples was the following: for each drug around 20 tablets were weighted, grinded and homogenized. Methanol was used as dissolvent. Drug samples were placed in an ultrasounds bath and then centrifuged. UV–vis spectra were recorded, using methanol as a blank. Concentrations of the two analytes in these drug samples were estimated by high performance of liquid chromatography (HPLC). Chromatographic determinations were performed using diode array UV–vis, Hewlett Packard detector. The column used is a reversed phase Spherisorb ODS-2 C18 column (15 cm long × 0.4 cm i.d., 5 μm particle size) with a C18 precolumn. The mobile phase composition used for the chromatographic determinations was acetonitril/H_2O (40/60) (v/v). A flow rate of 1.2 ml/min was used.

2.3. Forage samples

Different analytes in Ray-Grass forage samples were determined. The experimental procedure used before analyzing the Ray-Grass samples by means of NIR spectrophotometry was rather simple: samples were grinded, milled, homogenized, put in a capsule and directly measured by NIR spectrophotometry. Spectra from 125 samples were selected randomly for calibration, and spectra from other 46 samples were used for validation. Calibration samples were selected randomly trying to cover all the observed spectral data variance. If one of the selected validation samples resulted to be outside of the range covered by the calibration samples, it was exchanged by a calibration sample within the calibration range.

Humidity reference concentrations were estimated from the sample weight loss after oven drying at 103 °C for 4 h [27]. Some volatile compounds apart from water could evaporate decreasing sample total weight and causing excess errors, while other compounds may be oxidized, increasing the sample total weight and causing defect errors. Weight errors will depend on the compensation of these two effects. Protein reference concentrations have been estimated from nitrogen content analyzed using Kjeldhal method [28,29] and multiplied by a factor equals to 6.25 (which is derived from the fact that proteins of forages have an average content of nitrogen equals approximately to 15%). The humidity concentration range was between 4.86 and 13.33% (w/w), and for protein between 6.53 and 21.70% (w/w).

2.4. Instrumentation

UV–vis spectrophotometric determinations were performed by a Hewlett-Packard (Waldbronn, Germany) HP8452A diode array spectrophotometer. The instrument's bundled software HP 89530 MS-DOS UV–vis includes facilities for controlling, acquiring and processing spectra. In Fig. 1, the normalized

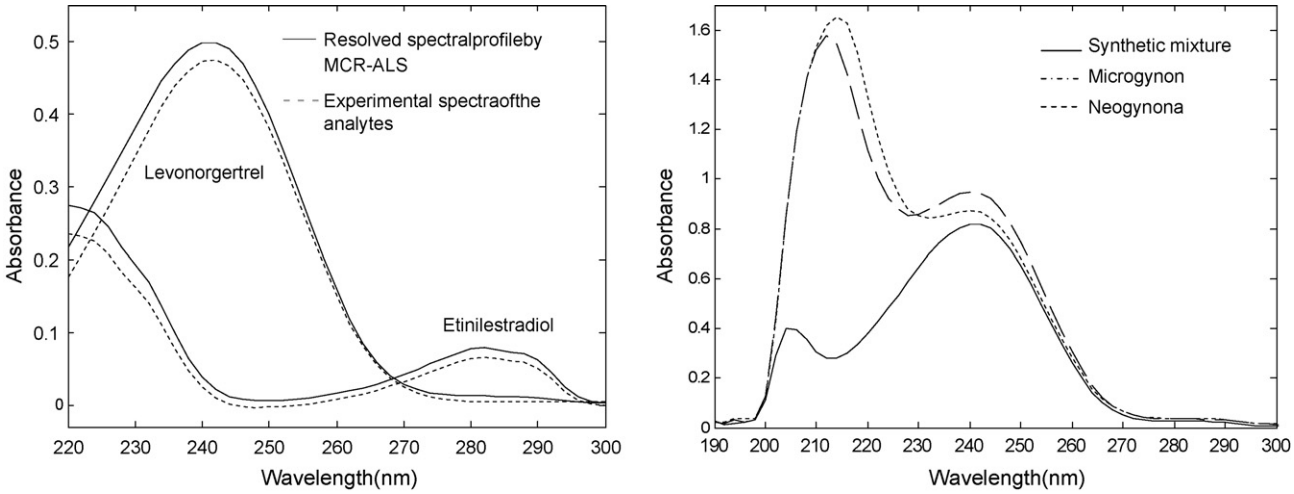

Fig. 1. Spectra of etinilestradiol and levonorgestrel analytes (left) and spectra of a synthetic mixture of them and of Microgynon and Neogynona drugs (right).

experimental spectra of etinilestradiol and levonorgestrel are presented (left), as well as spectra of the synthetic and commercial drug mixtures of them (right).

- Ultrasounds Bath, Selecta 0.61.
- Centrifuge, Arlesa model Digicen.
- NIR spectrophotometric determinations were performed using NIRSystems 6500 FOSS spectrophotometer. Each of the spectra finally considered is an average of 32 diffuse reflectance spectra. Sample containers were rectangular cups. The wavelength interval was 1100–2500 nm with 2 nm resolution. In Fig. 2, the obtained spectra of Ray-Grass samples are shown. Laboratory temperatures were always kept between 20 and 25 °C and relative humidity was always between 45 and 65%.

3. Chemometric methods

3.1. Multivariate curve resolution alternating least squares (MCR-ALS)

The first step of this chemometric data analysis procedure is to build up the data matrix, **D**. In the rows of this data matrix are the different individual spectra measured for the different analyzed samples and in the columns the absorbance (UV–vis) or Log 1/*R* (NIRS) measured values at each spectral wavelength. First a rough estimation of the possible number of components is obtained using different methods like principal component analysis (PCA) [2,5,13]. A bilinear relation between the experimental data, the concentrations and the pure spectra of the components is assumed, of analogous structure to the generalized law of Lambert–Beer [30], where the individual responses of each analyte or component are additive. In matrix form, this bilinear model is expressed in the following way:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \quad (1)$$

where $\mathbf{D}(I,J)$ is the matrix of experimental data, of dimensions I samples (spectra) by J wavelengths; $\mathbf{C}(I,K)$ is the matrix of concentration profiles of the different K analytes presents in the samples; $\mathbf{S}^{\mathrm{T}}(K,J)$ is the spectra matrix, whose K rows contain the pure spectra associate with the K species present in the samples; $\mathbf{E}(I,J)$ is the matrix associated to the experimental error. The resolution of experimental spectral data matrix **D** consists of the following steps, which are summarized in Fig. 3.

To initiate the iterative ALS procedure, an initial estimation is needed for the spectral or concentration profiles for each species. Different methods are used for this purpose like evolving factor analysis [1–3] or the determination of the purest variables [4–6]. In this work, initial estimations based on purest variables were preferred. If the initial estimations are the spectral profiles, the unconstrained least squares solution for the concentration profiles can be calculated from the expression:

$$\mathbf{C} = \mathbf{D}(\mathbf{S}^{\mathrm{T}})^{+} \quad (2)$$

where $(\mathbf{S}^{\mathrm{T}})^{+}$ is the pseudoinverse matrix of the spectra matrix $\mathbf{S}^{\mathrm{T}}$, which is equal to $\mathbf{S}(\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}$, when $\mathbf{S}^{\mathrm{T}}$ is of full rank [31]. If the initial estimations are the concentration profiles, the unconstrained least squares solution for the spectra can be calculated

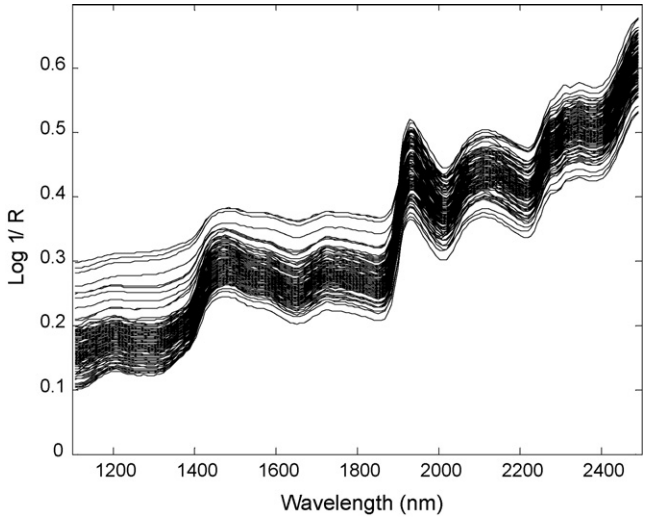

Fig. 2. NIR spectra of Ray-Grass calibration samples.

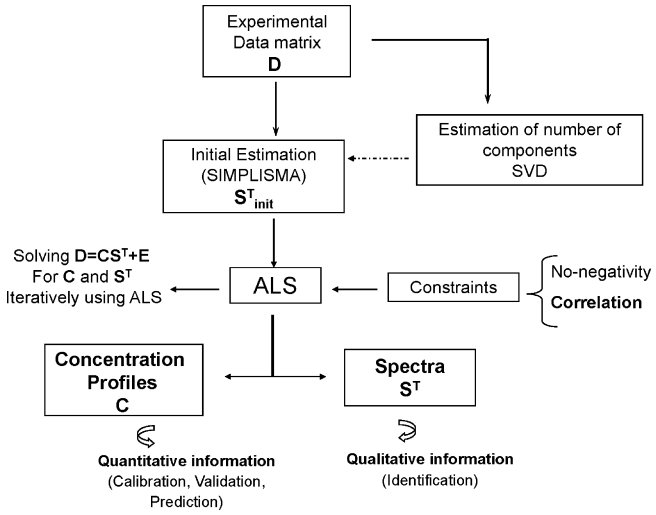

Fig. 3. Scheme of step of the resolution process in MCR-ALS method. See Section 3.1.

from the expression:

$$\mathbf{S}^{\mathrm{T}} = \mathbf{C}^{+}\mathbf{D} \tag{3}$$

where $\mathbf{C}^{+}$ is pseudoinverse of matrix $\mathbf{C}$ ($\mathbf{C}^{+} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}$, when $\mathbf{C}$ is of full rank) [31]. Both steps can be implemented in an alternating least squares cycle so that in each iteration new matrices of $\mathbf{C}$ and $\mathbf{S}^{\mathrm{T}}$ are then obtained. However, during these iterative calculations, a series of constraints with the purpose of giving solutions with physical meaning and of limiting their possible number are applied [11,12]. Iterations continue until an optimal solution is obtained that fulfils the constraints postulated and the established convergence criteria. Constraints applied in this work are only described briefly.

3.1.1. Nonnegativity concentration constraint

This is a general constraint used in curve resolution methods [32–36]. It is applied to the concentration profiles, due to the fact that the concentrations of the chemical species are always positive values or zero.

3.1.2. Nonnegativity spectra constraint

The application of this constraint depends on what instrumental technique is used for detection. In the case of UV–vis or NIR spectra, the intensity of the radiation absorbed or reflected by the sample never takes negative values.

3.1.3. Correlation constraint

This constraint has been introduced for the simultaneous quantitative analysis of mixtures of metal ions using voltammetric analysis and it implies [22] the establishment of calibration models for MCR-ALS, to be used for the quantitative determination of the analytes in the presence of unknown interferences. In this work, this constraint is extended to quantitative analysis

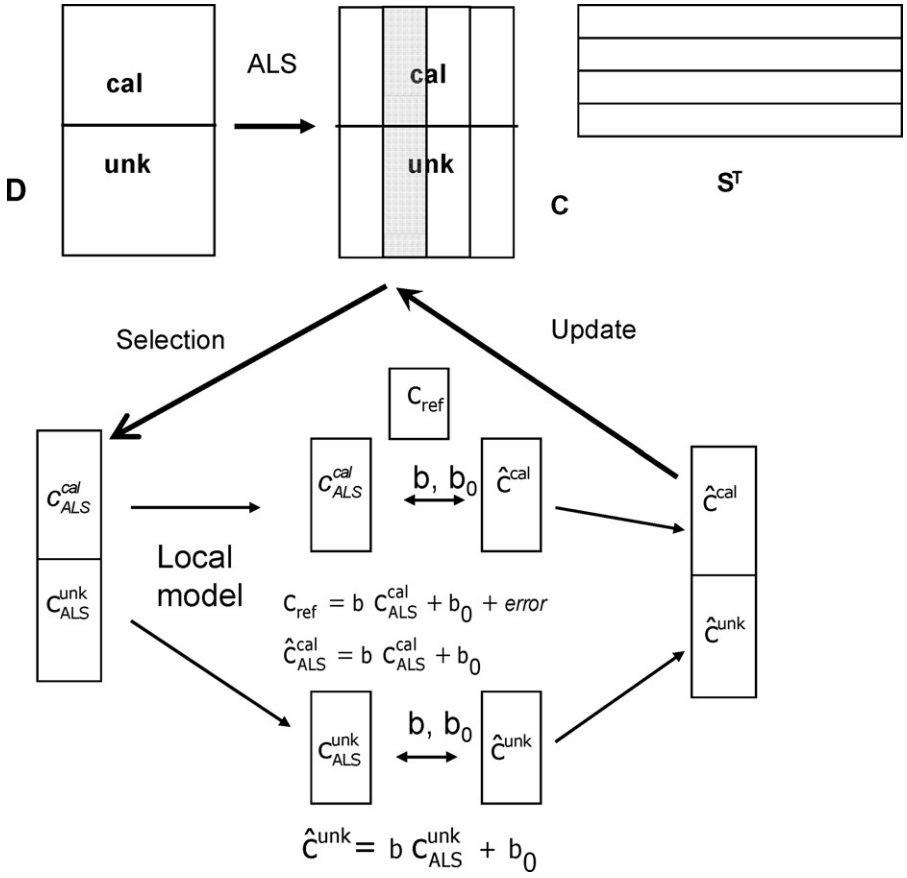

Fig. 4. Detailed description of the correlation constraint. See Section 3.1.

of spectral data in the simultaneous analysis of different analytes in samples of increasing complexity, including forage samples. This correlation constraint is explained in detail in Fig. 4, and it consists, of a series of steps performed during each iteration of the ALS optimization. Concentrations of a particular analyte in calibration samples, $\mathbf{c}_{\mathrm{ALS}}^{\mathrm{cal}}$, obtained by ALS at each iteration are correlated with previously known reference concentration values of the analyte $\mathbf{c}^{\mathrm{ref}}$ in these samples. A local linear model between the values $\mathbf{c}_{\mathrm{ALS}}^{\mathrm{cal}}$ and $\mathbf{c}^{\mathrm{ref}}$, is then built up so that:

$$\mathbf{c}^{\mathrm{ref}} = b\mathbf{c}_{\mathrm{ALS}}^{\mathrm{cal}} + b_0 + \mathbf{e}^{\mathrm{ref}} \tag{4}$$

where, b and b_0 are then the slope and offset values which better fit $\mathbf{c}_{\mathrm{ALS}}^{\mathrm{cal}}$ to $\mathbf{c}^{\mathrm{ref}}$, obtained by least squares linear regression, and $\mathbf{e}^{\mathrm{ref}}$ is the error in the reference concentrations (not modeled) The corresponding concentration values of these calibration samples calculated using this local model are:

$$\hat{\mathbf{c}}^{\mathrm{cal}} = b\mathbf{c}_{\mathrm{ALS}}^{\mathrm{cal}} + b_0 \tag{5}$$

And in order to predict the unknown concentration of the analyte in the new prediction samples $\hat{\mathbf{c}}^{\mathrm{unk}}$, the equation used is:

$$\hat{\mathbf{c}}^{\mathrm{unk}} = b\mathbf{c}_{\mathrm{ALS}}^{\mathrm{unk}} + b_0 \tag{6}$$

where b and b_0 are the values obtained previously in the calibration step from $\mathbf{c}^{\mathrm{ref}}$, and $\mathbf{c}_{\mathrm{ALS}}^{\mathrm{unk}}$ are the concentrations of the samples predicted by ALS. Each ALS iteration is then completed after updating the obtained values of prediction (i.e., by substitution of $\mathbf{c}_{\mathrm{ALS}}^{\mathrm{unk}}$ by $\hat{\mathbf{c}}^{\mathrm{unk}}$).

3.2. Partial least squares regression (PLSR)

PLSR method has been widely used in chemometrics to regression problems with highly correlated variables as it is often encountered in spectroscopy [37,38]. This regression method is based on a prediction model for the analyte concentration in the samples using efficiently the information contained in both data blocks, the spectroscopic data block (**D** matrix) and the concentrations data block (**c** vector). **D** and **c** were mean-centered prior to decomposition in factors. The PLSR algorithm selects successive orthogonal factors that maximize the covariance between spectra (**D** matrix) and analyte concentration (**c** vector). The objective of fitting a PLSR model, is to find a few number of PLSR factors that explain most of the covariation between both data blocks. Briefly, PLSR decomposes **D** and **c** into factor scores (**T**) and factor loadings (**P** and **q**) according to:

$$\mathbf{D} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{7}$$

$$\mathbf{c} = \mathbf{T}\mathbf{q} + \mathbf{f} \tag{8}$$

where **T** is the scores matrix, $\mathbf{P}^{\mathrm{T}}$ and **q** are the matrix and vector loadings describing the variance in **D** and **c**, respectively, and **E** and **f** are the residuals in **D** and **c**, respectively. This decomposition is performed simultaneously and in such a way that the first few factors should explain most of the covariation between **D** and **c**. The remaining factors resemble noise and can thus be ignored, hence the addition of residuals **E** and **f**.

3.3. Validation of results

In order to asses the quality of multivariate calibration models (from PLSR and MCR-ALS), it is convenient to do their validation using new samples not used during the calibration step. In this work, external validation was performed using a set of independent samples, whose spectra were not used to build the calibration model. From the whole original data set, a number of representative samples were selected for the calibration set. The remaining samples were then only used to validate the model.

The following expressions were used to express the validation results:

Root mean square error of prediction (RMSEP)

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{c}_i - \hat{\mathbf{c}}_i)^2}{n}} \tag{9}$$

Standard error of prediction (SEP)

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{c}_i - \hat{\mathbf{c}}_i - \mathrm{Bias})^2}{n-1}} \tag{10}$$

Bias (is a meaning of systematic error)

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n}(\mathbf{c}_i - \hat{\mathbf{c}}_i)}{n} \tag{11}$$

In all these expressions, $\mathbf{c}_i$ and $\hat{\mathbf{c}}_i$ are, respectively, the known and calculated analyte concentration in sample i, and n is the total number of samples considered in the validation. Also in order to evaluate the quality of the obtained results of the concentrations predicted by the application of the MCR-ALS or PLS models, for a particular analyte using n samples, the relative error in the predicted concentrations, in percentage (RE%), was calculated as:

$$\mathrm{RE}\,(\%) = 100\sqrt{\frac{\sum_{i=1}^{n}(\mathbf{c}_i - \hat{\mathbf{c}}_i)^2}{\sum_{i=1}^{n}\mathbf{c}_i^2}} \tag{12}$$

3.4. Chemometrics software

Data processing and PLS calibration calculations were carried out using commercial software packages: PLS Toolbox software version 2.1 (Eigenvector research, WA, USA) in MATLAB computer and visualization environment (The Mathworks, MA, USA), and UNSCRAMBLER software version 6.11 (CAMO A/S, Trondheim, Norway, 1986–1997). Multivariate curve resolution (MCR-ALS) has been implemented in MATLAB and it is available in Internet. See Ref. [13] for further details.

4. Results and discussion

4.1. MCR-ALS resolution and quantification of etinilestradiol and levonorgestrel steroids on commercial drugs analyzed by UV spectrophotometry

MCR-ALS has been applied first to synthetic experimental mixture samples for the resolution and quantification of steroids. Results were compared to those obtained by application of PLSR. Different wavelength intervals were investigated.

In Fig. 1, spectral profiles obtained by MCR-ALS are given. They are in agreement with the spectra of etinilestradiol and levonorgestrel pure standards. In contrast to MCR-ALS, PLS regression does not provide direct estimation of the pure spectra of the components of the mixture, although PLS loadings and weights may be interpreted in relation to the more relevant spectral features of the components present in the analyzed mixtures, specially for those for which the quantitative analytic information is available during the calibration step.

Quantitative results obtained by ALS and PLS methods for the different wavelength intervals are compared in Table 1.

Errors in Table 1 are obtained for external validation samples of synthetic mixtures. Number of components used in the calibration model was two in both cases, either for etinilestradiol or levonorgestrel. Constraints used in ALS optimization were non-negativity (for concentration and spectra profiles) and the new correlation constraint proposed in this work. Quantification errors obtained by MCR-ALS are in all the cases of the same order of magnitude than those obtained by application of PLSR. Bias in the case of etinilestradiol is higher than the bias obtained for levonorgestrel, which might be due to the lower concentrations used for this analyte in the mixtures. In the case of etinilestradiol the 230–300 nm interval gave the optimal quantification results and for levonorgestrel the best interval was 220–300 nm. Once the wavelength interval was chosen for the analysis of the two analytes in their synthetic mixtures, commercial drugs were analyzed using the same conditions.

Table 1
Figures of merit in the quantitative analysis of etinilestradiol and levonorgestrel analytes in synthetic mixtures using UV spectrophotometry (at different wavelength ranges) and PLS and MCR-ALS multivariate calibration methods

	RMSEP	SEP	Bias	RE (%)	r^2
Etinilestradiol					
220–300 nm					
ALS	0.738	0.486	0.581	3.316	0.9996
PLS	0.721	0.480	0.564	3.242	0.9996
230–300 nm					
ALS	0.616	0.395	0.493	2.769	0.9997
PLS	0.610	0.399	0.483	2.742	0.9997
250–300 nm					
ALS	0.591	0.371	0.478	2.655	0.9994
PLS	0.645	0.383	0.536	2.897	0.9994
Levonorgestrel					
220–300 nm					
ALS	0.098	0.073	0.070	0.784	0.9999
PLS	0.093	0.072	0.065	0.749	0.9999

RMSEP is root mean square error of prediction (Eq. (9)); SEP is standard error of prediction (Eq. (10)); Bias is a systematic error (Eq. (11)); r^2 is coefficient of correlation between calculated and actual concentration values of the analyzed compounds; RE% is the relative error in the predicted concentrations, in percentage (Eq. (12)). See text.

In Table 2 (upper part), results of the quantification of etinilestradiol and levonorgestrel steroids in Microgynon, commercial drug are given. Quantification of etinilestradiol in Microgynon resulted to be better in the 250–300 nm wavelength interval than in 220–300 nm range (see below), using three components, either for PLSR or MCR-ALS. Obtained errors were of the same order for MCR-ALS and PLSR. These errors were slightly higher than those obtained for synthetic mixtures, which is reasonable, since in synthetic mixtures no excipient interferences were present. Also since etinilestradiol is a minor component, it was more affected by the presence of these interferences. For levonorgestrel, rather good quantification results were obtained at the same wavelength intervals than with syn-

Table 2
Figures of merit in the quantitative analysis of etinilestradiol and levonorgestrel analytes, in Microgynon, Neogynona and Triagynon (brown and ochre) commercial drugs, using UV spectrophotometry (at different wavelength ranges) and PLS and MCR-ALS methods

	Etinilestradiol (250–300 nm)					Levonorgestrel (220–300 nm)				
	RMSEP	SEP	Bias	RE (%)	r^2	RMSEP	SEP	Bias	RE (%)	r^2
Microgynon										
ALS	0.149	0.086	0.129	4.534	0.9919	0.284	0.279	0.150	1.723	0.9998
PLS	0.143	0.081	0.125	4.366	0.9931	0.282	0.278	0.146	1.706	0.9998
Neogynona										
ALS	0.062	0.073	0.017	1.856	0.9963	0.184	0.226	0.002	1.007	0.9979
PLS	0.088	0.075	0.063	2.642	0.9964	0.200	0.229	−0.071	1.094	0.9980
Triagynon (brown)										
ALS	0.439	0.304	−0.361	4.457	0.9907	0.685	0.522	−0.551	3.747	0.9947
PLS	0.422	0.289	−0.349	4.287	0.9906	0.704	0.528	−0.557	3.856	0.9936
Triagynon (ochre)										
ALS	0.092	0.064	6.38é$^{-2}$	2.909	0.9921	0.165	0.206	−3.21e^{-3}	1.226	0.9970
PLS	0.124	0.071	1.09e^{-1}	3.927	0.9873	0.159	0.181	−6.0ê$^{-2}$	1.183	0.9986

See Table 1 for the meaning of RMSEP, SEP, bias, r^2 and RE (%).

thetic mixtures (220–300 nm), either using MCR-ALS or PLSR, using three components in each case.

In Table 2 (middle part), results in the quantification of the same analytes in Neogynona drug are given. The optimal interval for the quantification of etinilestradiol in Neogynona commercial drug was the same than for the quantification of Microgynon (250–300 nm). The use of the more restricted 250–300 nm spectral range for the analysis of these two commercial drugs instead of the 220–300 nm spectral range used during the analysis of the synthetic mixtures was due to the presence of drug interferences (excipient) that also absorb in the 220–250 nm spectral range. The inclusion of the 220–250 range would affect negatively the quantification of etinilestradiol in the commercial drugs. So, finally the 250–300 nm spectral range was considered to be the best one for the quantification of this analyte. Obtained errors were of the same order for MCR-ALS and for PLSR, and a little higher than for the synthetic mixtures. For levonorgestrel also a good quantification was obtained in the same wavelengths interval than for synthetic mixtures (220–300 nm), either for MCR-ALS or for PLSR. The number of components used to explain the model for each analyte, were in this case (like for Microgynon) three components, either for PLSR or MCR-ALS.

And finally, also in Table 2 (lower part), obtained results for the quantification of the two steroids in Triagynon (brown and ochre color), using MCR-ALS and PLS are also given. Optimal wavelength interval for the quantification of etinilestradiol in Triagynon (ochre and brown color pills) was at 250–300 nm, and obtained errors were of the same order for MCR-ALS and for PLSR. For levonorgestrel, the optimal quantification was obtained in the wavelength interval of 230–300 nm, either for MCR-ALS or for PLSR. The number of components used to explain the model, for each analyte, was again three components, either for PLSR or for MCR-ALS.

Differences observed in the results obtained in all cases by application of MCR-ALS or PLSR were considered to be little significant. A deeper interpretation of these small differences would require the study of a larger number of samples with a better control of the factors that can influence these small differences. Nevertheless, it is possible to conclude that at least in the analysis of synthetic mixtures and in the analysis of the investigated commercial drugs, MCR-ALS provided quantitative results of similar quality to those provided by the application of PLSR. The obvious advantage of MCR-ALS compared to PLSR is, however, that MCR-ALS recovers the qualitative information as well, including the pure spectra of the components (Fig. 1), and also of the interferents, allowing their possible identification/confirmation.

4.2. MCR-ALS resolution and quantification of humidity and protein content on natural samples (Ray-Grass) analyzed by NIR spectrophotometry

NIR spectroscopy has been widely applied as an analytical technique in the agricultural food sector, using partial least squares (PLS) to develop calibration equations for the determination of the humidity and protein content [39–41]. In this work, MCR-ALS and PLS methods have been applied and compared in the analysis of natural Ray-Grass forage samples using NIR spectrophotometry, with the purpose of obtaining both qualitative and quantitative information of the humidity and protein present in these samples. Obviously, in this case, difficulties for a proper calibration will be more important because of the larger contribution of unknown physical contributions and chemical interferents in the measured NIR spectra of the analyzed forage samples. This example will probably illustrate the limits of the use of the proposed MCR-ALS method for quantitative determinations of natural samples using first order spectrophotometric data. Comparison of MCR-ALS results with PLSR results is pertinent since this is a much extended method used for calibration of NIR spectrophotometric data [37].

In Table 3, obtained results in the quantification of humidity and protein in Ray-Grass forage samples using NIR spectrophotometric data and MCR-ALS and PLS are given. A summary of prediction errors for these two analytes, using different number

Table 3
Figures of merit in the quantitative analysis of humidity and protein analytes in Ray-Grass samples using NIR spectrophotometry and PLS and MCR-ALS methods

Number of factors	RMSEP		SEP		Bias		r^2		RE (%)	
	ALS	PLS	ALS	PLS	ALS	PLS	ALS	PLS	ALS	PLS
Humidity										
5	0.391	0.313	0.370	0.316	0.045	0.008	0.962	0.973	4.383	3.721
6	0.369	0.301	0.370	0.305	0.046	0.009	0.962	0.975	4.387	3.585
7	0.358	0.307	0.360	0.311	0.043	0.006	0.964	0.974	4.260	3.654
8	0.361	0.289	0.363	0.292	0.030	0.008	0.963	0.977	4.287	3.432
9	0.285	0.313	0.289	0.313	0.004	0.004	0.977	0.973	3.394	3.724
10	0.286	0.269	0.290	0.268	−0.005	0.004	0.977	0.980	3.406	3.199
Protein										
5	1.286	0.789	1.300	0.797	0.013	0.026	0.957	0.984	7.656	4.695
6	0.808	0.724	0.816	0.730	0.046	0.061	0.983	0.986	4.813	4.314
7	0.841	0.623	0.850	0.628	0.002	0.049	0.982	0.990	5.008	3.711
8	0.871	0.560	0.880	0.555	0.003	0.011	0.980	0.992	5.184	3.334
9	0.973	0.559	0.981	0.559	0.074	0.083	0.976	0.992	5.792	3.327
10	0.748	0.559	0.755	0.560	0.050	0.070	0.986	0.992	4.45	3.327

See Table 1 for the meaning of RMSEP, SEP, Bias, r^2 and RE (%).

of factors for PLS and MCR-ALS are given. The optimal number of components was selected in each case by considering the minimal RMSEP values. Errors in Table 3 are calculated for external validation samples. In the case of the PLSR method, results shown in Table 3 correspond to the application of the model to mean centered spectra. And in the case of the MCR-ALS, results shown in Table 3 correspond to data without any data pretreatment. Constraints applied during the ALS optimization were non-negativity (for concentration and spectra profiles) and the new correlation constraint discussed in this work.

According to results shown in Table 3, it is difficult to decide about the optimal number of components for the determination of humidity. Whereas in the determination of humidity by PLS, six components gave a first minimum of RMSEP (0.301) and of relative error RE (%) (3.58) in the case of MCR-ALS, seven components were needed from RMSEP (0.358) and RE% (4.26%) values. This first selection of components should be considered rather parsimonious since lower RMSEP and RE% values could still be obtained (Table 3) if a larger number of components were considered, for both PLS and ALS. However, since these differences were not large, this first selection estimation was considered good enough for the purposes of this comparative work. For protein, the best number of components was eight factors for PLS and six factors for MCR-ALS. Prediction errors (RMSEP) were 0.560 and 0.808 for PLS and MCR-ALS, respectively. In this case, PLSR clearly outperformed MCR-ALS. This improvement was probably due to the possibility to incorporate a larger number of components in PLSR models compared to MCR-ALS models. MCR-ALS could not resolve more components as PLSR for a better quantification of protein because of the intrinsic difficulties resolving minor components by MCR-ALS, whereas they could still have some effect improving quantitative estimations in PLSR. On the other hand, if RMSEP values are compared using the same number of factors, for PLS and MCR-ALS, some differences in the prediction error results are always encountered which can be due to the fact that PLS, maximizes relevant information in the first factors, and fits better calibration and validation data. In Fig. 5, regression of humidity and protein contents predicted by MCR-ALS and PLS versus the concentration reference values, using the optimal number of factors in each case, are given. In caption of Fig. 5 results of the elliptic joint confidence region F test [42] for the slope and the intercept of these regressions are given. This test considers that if no systematic errors are present, the theoretical point intercept should be zero and the theoretical slope should be equal to one and that their uncertainties should be located inside the corresponding

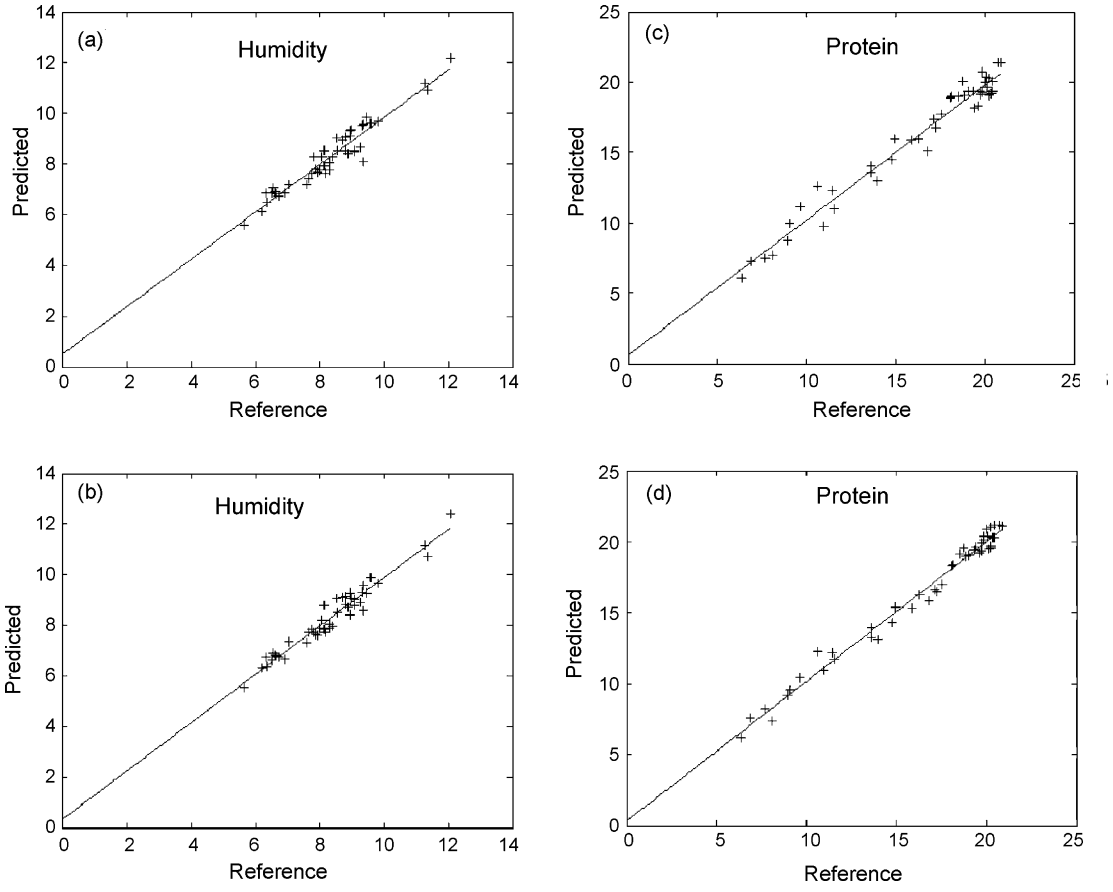

Fig. 5. Humidity and protein concentrations values predicted by MCR-ALS and PLS models vs. concentration reference values in validation samples. (a) Humidity values predicted by MCR-ALS vs. their reference values ($F = 0.18$, $\alpha = 0.91$), (b) humidity predicted values by PLSR vs. their reference values ($F = 0.24$, $\alpha = 0.87$), (c) protein predicted values by MCR-ALS vs. reference values of protein ($F = 0.43$, α=0.73), (d) protein predicted values by PLSR versus reference values of protein ($F = 2.06$, $\alpha = 0.12$). In parenthesis, calculated F values and significance levels for the regression slope and offset confidence region test (see Ref. [42]) are given. Tabulated F value at the same degrees of freedom ($\nu_1 = 2 + 1 = 3$ and $\nu_2 = 46 - 2 - 1 = 43$) and $\alpha = 0.05$ significance level is $F = 2.82$.

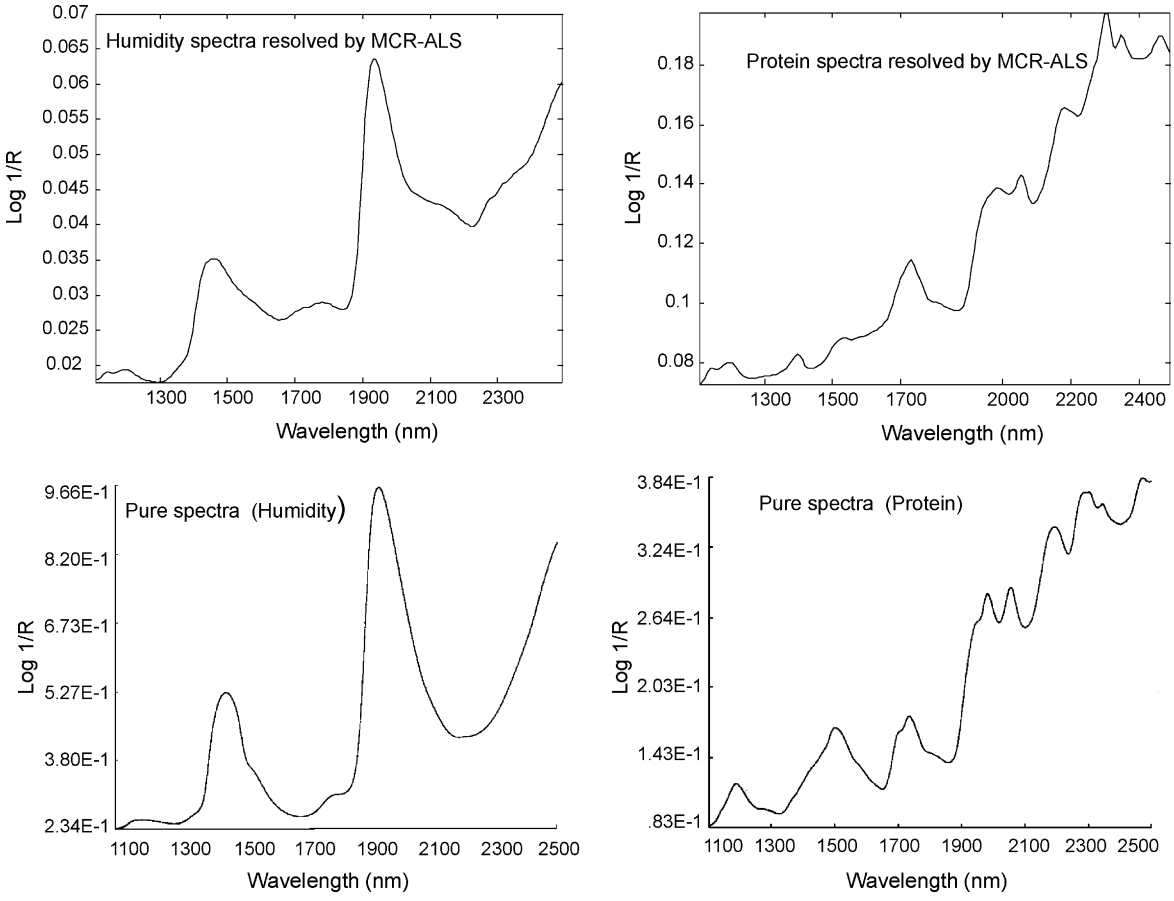

Fig. 6. Pure spectra of moisture and protein estimated by MCR-ALS (two top plots) and the same pure spectra of moisture and protein taken from the literature [43].

elliptic joint confidence region. In all cases, F test confirmed the adequacy of the postulated models.

Nevertheless and in general, the determination of humidity and protein in Ray-Grass forage samples using MCR-ALS were also rather good and close to the optimal ones obtained by PLSR. This is especially relevant if it is taken into account the intrinsic difficulties inherent to MCR-ALS to properly resolve and quantificate components contributing very little to the measured spectroscopic signal and also to the fact that neither sample nor spectra pretreatment was performed using MCR-ALS. NIR spectra resolved for humidity and protein using this method are given in Fig. 6. Humidity spectra show clearly the characteristic water bands at 1450 and 1940 nm. Protein spectra gave the characteristic bands at 2050 and 2180 nm. In general, the variations in the content of N in forages caused a change in the form of the bands located between 2000 and 2200 nm. In both cases (humidity and protein), spectra resolved by MCR-ALS were very similar to NIR spectra previously reported in the literature for humidity and protein in forages [43] (see lower part of Fig. 6).

5. Conclusions

The predictive capability of MCR-ALS using the new correlation constraint for a particular analyte in presence of interferences in unknown mixtures and natural samples, was in general comparable to the results obtained using PLSR calibration approaches. The main advantage of using MCR-ALS instead of PLSR is, however, the simultaneous recovery of qualitative information (spectra confirmation) about the analyte and possible unknown intereferents. In this work, we have presented a preliminary contribution to this problem and further work is needed to confirm the results here obtained.

References

[1] H. Gampp, M. Maeder, C. Meyer, A.D. Zuberbühler, Talanta 33 (1986) 943.
[2] M. Maeder, A. Zilian, Chemometr. Intell. Lab. 3 (1988) 205.
[3] M. Maeder, Anal. Chem. 59 (1987) 527.
[4] W. Windig, J. Guilment, Anal. Chem. 63 (1991) 1425.
[5] W. Windig, D.A. Stephenson, Anal. Chem. 64 (1992) 2735.
[6] W. Windig, C.E. Heckler, F.A. Agblevor, R.J. Evans, Chemometr. Intell. Lab. 14 (1992) 195.
[7] E.R. Malinowski, J. Chemometr. 6 (1992) 29.
[8] R. Tauler, A. Izquierdo-Ridorsa, E. Cassasas, Chemometr. Intell. Lab. 18 (1993) 293.
[9] R. Tauer, A. Izquierdo-Ridorsa, R. Gargallo, E. Casassas, Chemometr. Intell. Lab. 27 (1995) 163.
[10] R. Tauler, D. Barceló, Trends Anal. Chem. 12 (1993) 319.
[11] R. Tauler, A. Smilde, B.R. Kowalski, J. Chemometr. 9 (1995) 31.
[12] R. Tauler, J. Chemometr. 16 (2002) 117.
[13] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, Chemometr. Intell. Lab. 76 (2005) 101.
[14] R. Tauler, B. Kowalski, S. Fleming, Anal. Chem. 65 (1993) 2040.

[15] J. Saurina, S. Hernández-Cassou, R. Tauler, Anal. Chem. 67 (1995) 3722.
[16] J. Saurina, S. Hernandez-Cassou, Anal. Chim. Acta 438 (1/2) (2001) 335.
[17] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemometr. Intell. Lab. 54 (2000) 123.
[18] A. de Juan, R. Dysson, C. Marcolli, M. Rault, R. Tauler, M. Maeder, Trends Anal. Chem. 23 (2004) 70.
[19] J. Jaumot, R. Gargallo, N. Escaja, C. Gonzalez, E. Pedroso, R. Tauler, Nucleic Acid Res. 30. (2002).
[20] R.M. Latorre, J. Saurina, S. Hernández-Cassou, Electrophoresis 21 (2000) 563.
[21] M.S. Díaz-Cruz, J. Mendieta, R. Tauler, M. Esteban, Anal. Chem. 71 (1999) 4629.
[22] M.C. Antunes, J.E.J. Simao, A.C. Duarte, R. Tauler, Analyst 127 (6) (2002) 809.
[23] M.J. López, C. Ariño, S. Diaz-Cruz, J.M. Díaz-Cruz, R. Tauler, M. Esteban, Environ. Sci. Technol. 37 (2003) 5609.
[24] M. Vives, R. Tauler, V. Moreno, R. Gargallo, Anal. Chim. Acta 446 (2001) 439.
[25] S. Navea, A. de Juan, R. Tauler, Anal. Chem. 75 (2003) 5592.
[26] E. Peré-Trepat, M. Petrovic, D. Barceló, R. Tauler, Anal. Bioanal. Chem. 378 (2004) 642.
[27] AFNOR association Francaise de Normalization 82411. NF V18-109 (1982). "Determination de la teneur en eau".
[28] A. Vogel, A Textbook of Quantitative Inorganic Analysis including elementary instrumental analysis, fourth ed., Longman Group Ltd., London, 1978.
[29] AOAC Official Method, 11, Crude Protein in Animal Feed, Forage, Grain, and Oilseeds, Block Digestion using Copper Catalyst, Steam Distillation into Boric Acid, AOAC International, Gaithersburg, MD, 2001.
[30] D.A. Skoog, D.M. West, F.J. Holler, Fundamentals of Analytical Chemistry, 7th Sanders Coll. Pub., 1995.
[31] G.H. Golub, C.F. Van Loan, Matrix computation, second ed., The Johns Hopkins University Press, Baltimore, 1989, ISBN 0-8018-3739-1.
[32] H.W. Lawton, F.A. Sylvestre, Technometrics 13 (1971) 617.
[33] O.S. Borgen, B.R. Kowalski, Anal. Chim. Acta 174 (1985) 1.
[34] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173 (1985) 253.
[35] E. Spjotvoll, H. Martens, R. Volden, Technometrics 3 (1982) 173.
[36] D. Macnaughtan, L.B. Rogers, G. Wernimont, Anal. Chem. 44 (1972) 1421.
[37] H. Martens, T. Naes, Multivariate Calibration, Wiley, New York, 1991.
[38] R. de Carvalho, M. del Nogal Sĭanchez, J. Wattoom, R.G. Brereton, Talanta 68 (2006) 1190.
[39] J.S. Shenk, J. Workman, M.O. Westerhaus, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near Infrared Analysis, Marcel Dekker, New York, 1992.
[40] D. Cozzolino, I. Delucchi, M. Kholi, D. Vázquez, Agricultura Técnica 66 (2006) 370.
[41] P.L. Bras, A.S. Bernandino, J.A. Lopez, J.C. Menezes, Chemometr. Intell. Lab. 75 (2005) 91.
[42] N. Drapper, H. Smith, Applied Regression Analysis, second ed., Wiley, New York, 1981.
[43] I. Murray, P.C. Williams, Near Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Minnesota, 1987 (Chapter 2).