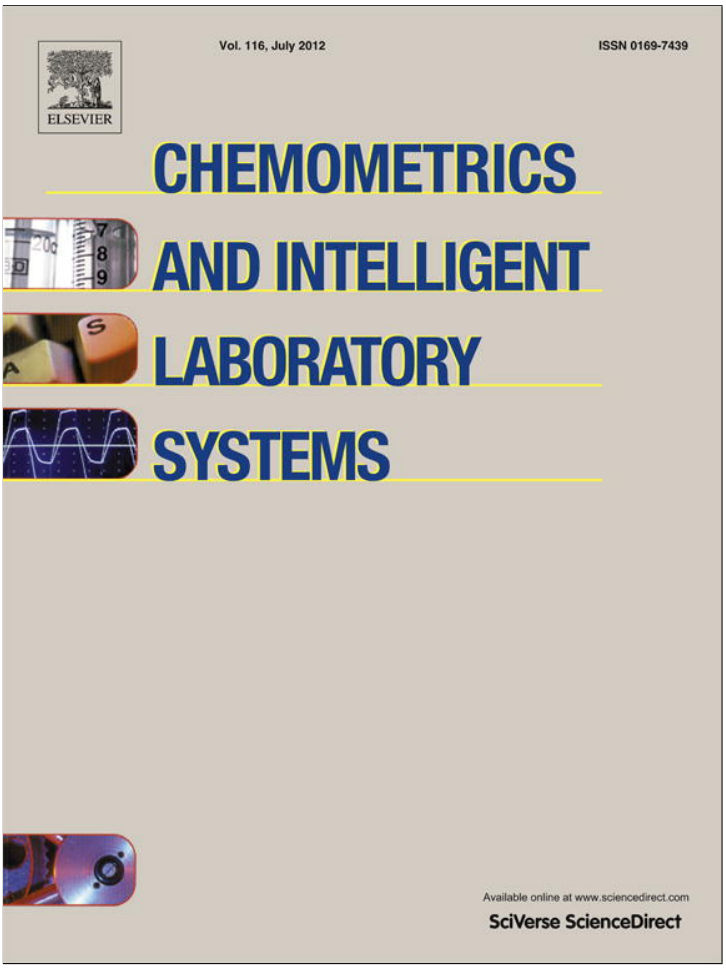
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Chemometrics and Intelligent Laboratory Systems 116 (2012) 9–16

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

MVC3: A MATLAB graphical interface toolbox for third-order multivariate calibration

Alejandro C. Olivieri [a,b,*], Hai-Long Wu [c], Ru-Qin Yu [c]

[a] Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Argentina
[b] Instituto de Química Rosario (IQUIR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Suipacha 531, Rosario S2002LRK, Argentina
[c] State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, People's Republic of China

ARTICLE INFO

Article history:
Received 21 December 2011
Received in revised form 5 March 2012
Accepted 6 March 2012
Available online 2 May 2012

Keywords:
Third-order multivariate calibration
MATLAB program
Graphical interface
Figures of merit

ABSTRACT

A new MATLAB graphical interface toolbox for implementing third-order multivariate calibration methodologies is discussed. Multivariate calibration 3 (MVC3) is a sequel of the already described first-order (MVC1) and second-order (MVC2) toolboxes. MVC3 accepts a variety of ASCII data for input, depending on whether the third-order data are vectorized or matricized. If required, data for sample sets are arranged into four-way arrays for processing with several quadrilinear and non-quadrilinear algorithms. Quadrilinear decomposition techniques and latent structured models based on partial least-squares regression and residual trilinearization are included in the software. Appropriate working sensor regions in the three data dimensions can be selected. Model development and its subsequent application to unknown samples are straightforward from the interface. Prediction results are provided along with analytical figures of merit and standard concentration errors, as calculated by modern concepts of uncertainty propagation.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Higher-order multivariate calibration is increasingly used by the analytical community, as instrumental hyphenation becomes popular for increasing sensitivity and selectivity of the determinations. Several reviews have appeared on the subject, all highlighting the progress in both instrumentation and data processing which has taken place in recent years [1–3]. In the field of third-order data recording for quantitative analytical purposes, several recent works can be mentioned. They are based on: (1) two-dimensional gas chromatography (GC–GC) with mass spectrometric (MS) detection [4–6] or two-dimensional liquid chromatography (LC–LC) with UV–visible diode array detection (DAD) [7,8], and (2) time evolution of excitation–emission luminescence matrices, either following the kinetics of a reaction with fluorescence matrices [9–16] or the time decay of phosphorescence matrices [17].

Programs for first- and second-order multivariate analysis are freely available on the internet (Table 1), with MATLAB [18] as the preferred programming environment. However, although the latter shows a number of advantages, implementing multi-way analysis in MATLAB requires some level of programming skill. MATLAB graphical user interfaces (GUI) are useful in bridging the gap between pure chemometricians and end users [19–22], though considerable work is still required in developing easy-to-use software for routine applications in most analytical laboratories. Commercial software is available for implementing first- and second-order calibration (Table 1), but in the latter case the gap between commercial and free academic software is significant.

MVC3 was developed as an integrated chemometric MATLAB toolbox in order to manage several different third-order multivariate calibration algorithms in an easy-to-use graphical interface environment. The toolbox can be applied to any type of data which are structured as three-dimensional arrays for each sample. It is a sequel of the already described MVC1 MATLAB interface [19] and the MULTIVAR Visual Basic program [23], both developed for handling first-order multivariate calibration methods, and the MVC2 MATLAB interface for second-order multivariate calibration [22].

The third-order multivariate calibration techniques included in MVC3 can be divided in two relevant groups, namely those based on: (1) quadrilinear decomposition (QLD) or (2) residual trilinearization (RTL). The former group includes parallel factor analysis (PARAFAC) [24], alternating penalty QLD (APQLD) [25] and alternating weighted residual constraint QLD (AWRCQLD) [26].

The second group of methods, based on residual trilinearization, comprise: 1) trilinear least-squares followed by RTL (TLLS/RTL) [27], 2) unfolded partial least-squares/RTL (U-PLS/RTL) [27], 3) multidimensional partial least-squares/RTL (N-PLS/RTL) [14], and unfolded principal component analysis/RTL (U-PCA/RTL) [16]. The latter methodology has been developed to produce test sample scores from nonlinear instrumental data, which are free from the contribution of interferences, for further analysis using artificial neural networks [16]. In the U-PLS and N-PLS methods, the calibration models can be optimized by leave-one out cross-validation [28].

For discussions concerning the properties and applicability of the different algorithms, see the reviews [1–3].

* Corresponding author. Tel./fax: +54 341 4372704.
E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2012.03.018

Author's personal copy

Table 1
Free and commercial software for multivariate calibration.

Free software			
Algorithm			Website
Parallel factor analysis, N-way partial least-squares and other multi-way methods			http://www.models.kvl.dk/algorithms
Multivariate curve resolution			http://www.ub.es/gesq/mcr/mcr.htm
Multivariate curve resolution			http://personal.ecu.edu/gemperlinep
Several first-order and multi-way methods			www.chemometry.com
Generalized rank annihilation and direct trilinear decomposition			http://www.cpac.washington.edu
Commercial software			
	Software	Company	Website
First-order partial least-squares	GRAMS IQ	Thermo Scientific	www.thermoscientific.com
First-order partial least-squares	PLS Toolbox	Eigenvector.com	www.eigenvector.com
First-order partial least-squares	PLS Toolbox	The MathWorks	www.mathworks.com
First-order and N-way partial least-squares, and multivariate curve resolution	UNSCRAMBLER	Camo	www.camo.com
First-order partial least-squares and multivariate curve resolution	PIROUETTE	Infometrix Software	www.infometrix.com
First-order partial least-squares	EZINFO	Umetrics	www.umetrics.com
PARAFAC and other multi-way methods	3 Way Pack	The Three-mode Company	http://three-mode.leidenuiv.nl

Concerning the estimation of figures of merit, the subject is now well-established in the field of second-order multivariate calibration [29–31]. However, expressions for assessing the third-order sensitivity, which is a key ingredient for most figures of merit, have been elusive for some time [32]. Very recently, a closed expression has been developed for PARAFAC [33]. The latter can conceivably be extended for the RTL-based methodologies, but considerable theoretical work is still needed to firmly establish its applicability. Following a somewhat provisional definition, the following figures of merit are estimated and reported in MVC3: sensitivity (SEN), selectivity (SEL), analytical sensitivity (γ), limit of detection (LOD) and standard error in concentration for each predicted concentration (SD).

Calculations and graphical outputs are conveniently managed in MVC3 through graphical user interface (GUI) shells. The software does not require a highly experienced user, but a basic knowledge on the underlying methods is advisable in order to successfully interpret the results.

In the present report, the toolbox versatility and performance are illustrated through the analysis of a simulated and an experimental example.

2. Software

The software runs under MATLAB version 7 [18]. The files do only need to be copied into a folder declared in the MATLAB path. Please refer to the document named 'MVC3_manual.doc' which is provided with the software. The codes are freely available, along with examples and manual, and will be provided on request via e-mail. Please contact the corresponding author or download the program (including manual and examples) from: http://www.chemometry.com/Index/Links%20and%20downloads/Programs/Olivieri/mvc3_zipped.zip.

In order to perform PARAFAC and N-PLS modeling, the interface employs the routines available in R. Bro's webpage (http://www.models.kvl.dk/algorithms, accessed November 2011). The APQLD and AWRCQLD routines were developed by H.L. Wu and R.Q. Yu (Hunan University, China), while all RTL routines were written by A.C. Olivieri (University of Rosario, Argentina).

3. Simulated data

A set of simulated data is provided along with the program and manual. It involves simulated time-evolving fluorescence excitation–emission third-order data for a set of calibration and test samples containing two calibrated analytes and a single interferent. Fig. 1 shows the noiseless profiles from which the simulated data were built. All of them are normalized to unit length. Notice that a strong overlapping occurs between the profiles of analyte 2 and the interferent, particularly in the first data dimension. This implies that, although the total signals for both analytes are the same (i.e., 1, because the profiles are normalized), the final sensitivity toward them will be different: analyte 2 will show lower sensitivity. Nine calibration samples with a central composite design for the concentrations of both analytes (in the range of 0–1 concentration units) are provided, along with ten test samples with the three components in random concentrations. The concentration of the interferent in the test samples was kept high, in the range from 0.5 to 1 concentration units, in order to ensure that all samples present the challenge of achieving the second-order advantage to the available algorithms. Random Gaussian noise with 0.002 units of standard deviation was added to all signals, and with 0.01 units was added to all nominal concentrations. This implies a ca. 1% of noise with respect to both maximum signal and concentration. This simulated data set will be discussed in detail below.

4. Experimental data

In the selected experimental system, malondialdehyde was measured in olive oils treated with methylamine, which led the analyte to develop a strongly fluorescent product [16]. Calibration was performed with fourteen samples with the analyte in the range of 0.00–2.16 mg L^{-1}, and 8 spiked olive oils were examined. When analyzing each test sample, the size of the four-dimensional array was 15 samples × 14 excitation wavelengths × 11 emission wavelengths × 21 times. See Ref. [16] for further details. These data are available from the authors on request.

5. Program description

MVC3 has a single main window (Fig. 2), from which all steps required to implement the different third-order multivariate calibration strategies can be carried out. The first step is to select the desired multivariate methodology and the corresponding number of responsive components (this can be the total number of components for QLD algorithms, the number of unexpected sample components for TLLS/RTL, or the numbers of calibrated and unexpected sample components for U-PLS/RTL, N-PLS/RTL and U-PCA/RTL).

The working sensor regions should be provided so that the program is able to reconstruct the third-order data arrays if unfolded data are used as input. Sensor regions can be selected from the screen, if part of the data arrays need to be discarded (for example, when unwanted phenomena occur, such as scattering in fluorescence spectroscopy).

Several different data types are admissible, all contained in ASCII files, whose names should be provided to the graphical interface. They may be arranged in any of the following formats: 1) X_vectors means one-column unfolded vectors, with each $J \times K \times L$ data array

Author's personal copy

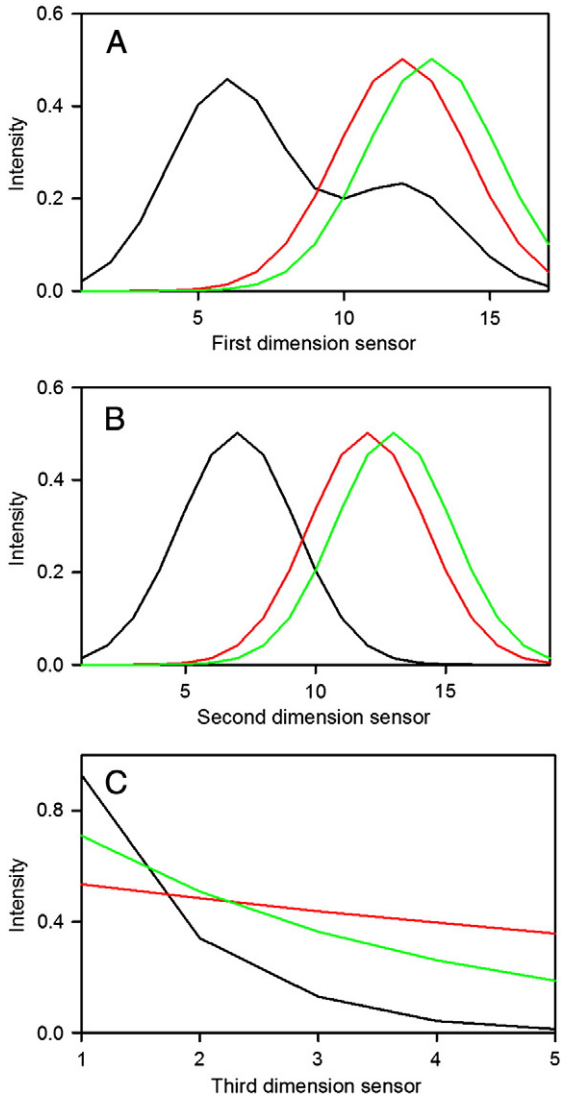


Fig. 1. Noiseless profiles employed to build the synthetic data set analyzed by the present MVC3 software. A), B) and C) First, second and third dimension profiles for the analytes respectively. The color codes are: black line, analyte 1, red line, analyte 2, green line, interferent. The first dimension mimics fluorescence excitation spectra, the second the corresponding emission spectra and the third the kinetic time evolution.

unfolded and saved as a $JKL \times 1$ vector, 2) X,Y_vectors implies two-column data with the first column indicating, for example, the wavelength and the second the unfolded signal, 3) X,Y_matrices corresponds to data in the form of $J \times 2K$ matrices repeated L times, with a column of wavelengths separating successive data matrices recorded at different times (the latter format is produced by some spectrofluorimeters).

Specific calibration samples can be excluded from the model, for example, if they have been found to have wrong nominal concentrations.

Uncertainties in both signals and calibration concentrations will subsequently be used in the estimation of certain analytical figures of merit (AFOM), such as standard error in concentrations, limit of detection, etc. These uncertainties will be taken as: (1) the fitting residuals of the test sample to the corresponding model for signals and (2) the average uncertainty in predicting the calibration samples for concentrations.

6. QLD models for simulated data

All implemented methods require a certain number of responsive components to be preset for building the calibration models. In the case of the QLD models, this number can be estimated by analyzing how the core consistency varies as a function of an increasing number of trial components. The optimum number of components is reached when the core consistency significantly drops below 50 [34]. The residual standard deviation of the least-squares fit of the four-way data array to the quadrilinear model is another important parameter, which stabilizes when the correct number of components has been reached.

PARAFAC specific results obtained for the processing of a typical test sample of the simulated data are as follows: (1) core consistency values for 1 to 5 trial components are 100, 99.2, 31.9, −0.2 and 0.12, and (2) residual fit values (in signal units, to be compared with the degree of signal uncertainty of 0.002, see above) are 0.01, 0.004, 0.002, 0.002 and 0.002 respectively. As can be seen, the core consistency parameter and the analysis of the residual fit leads to the conclusion that three responsive components are present in these samples. Similar results are found with the remaining QLD models (APQLD and AWRCQLD). It should be noticed that an additional indicator that the correct number of components has been surpassed is that repeated profiles are obtained.

Once the number of components is set to three, the 'Predict' button allows one to perform quadrilinear decomposition of the four-way data set formed by joining calibration and test sample data (this activity is conducted sample by sample). A plot of separate component profiles in each dimension is then produced (Fig. 3), allowing the user to select the relevant component number which best correlates with the known properties of the analyte. In our case, comparison with Fig. 1 shows that analyte 1 is component 2 and analyte 2 is component 1, while component 3 is the interferent (this component numbering may vary from sample to sample). Finally, a pseudo-univariate scores-concentration plot for the selected analyte is produced (Fig. 4 shows the results for analyte 1), and the analyte concentration is estimated by interpolation in the pseudo-univariate graph. Concentrations predicted in this manner for both analytes are collected in Table 2. As can be seen, for this particular example, PARAFAC, APQLD and AWRCQLD furnish good prediction results, which are incidentally better for analyte 1 than for analyte 2. This can be judged from the root mean square errors of prediction (RMSEP) quoted in Table 2, and is due to the lower intrinsic sensitivity toward analyte 2 (see above). Also noticeable in Table 2 is the homogeneity of standard deviations in predicted concentrations (lower for analyte 1 than for analyte 2), a result which directly follows from the capability of these QLD models in efficiently decomposing the signals for the analyte from the mixture signal.

Table 3 shows the estimated figures of merit for the three QLD algorithms. There is a good correlation between the higher sensitivity for analyte 1 and the lower RMSEP values for this analyte in Table 2. Conversely, analyte 2 shows a correspondingly larger RMSEP, consistent with the lower sensitivity toward this constituent.

Notice that in certain specific analytical cases (not in the presently discussed one), QLD models need to be properly initialized and constrained [1–3]. This can be done in MVC3 using any of the several initializing and constraining options which come with Bro's PARAFAC package.

7. RTL models for simulated data

RTL models require two separate numbers of components: the calibrated ones and the unexpected ones or interferents, respectively. In the case of the TLLS method, the number of calibrated components is automatically set according to the number of columns of the concentration matrix used for calibration, which is provided in a separate concentration file. In the latent-structured RTL cases (i.e., U-PLS and N-PLS), the number of calibration factors can be tuned through leave-one-out cross-validation ('Cross-validation' button in the MVC3 screen). This involves the systematic removal of one of the training samples in turn, and use of the remaining ones for building PLS models. The predicted concentrations of the left-out samples are compared with the nominal values for

Author's personal copy

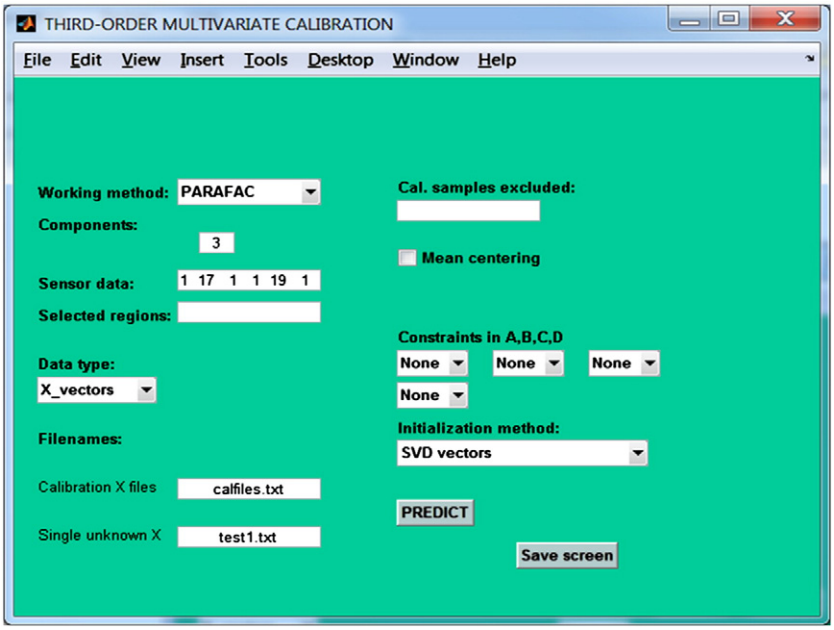


Fig. 2. Main screen of the graphical interface MVC3, prepared to process an unknown sample with PARAFAC using three responsive components.

each calibration sample, and the predicted error sum of squares [PRESS = $\Sigma\,(c_{act} - c_{pred})^2$] is calculated. After finishing the cross-validation procedure, the optimum number of factors is assessed using Haaland and Thomas' criterion [28]. The following statistical parameters are displayed as a function of the number of factors: the PRESS, the root mean square error of cross-validation (RMSECV), the *F* ratio between successive PRESS values to the minimum PRESS, and the associated probability. This information is given in tabular format (see Table 4 for U-PLS when calibrating for analyte 1), and also by activating the button 'CV Plots'. Similar results were obtained for analyte 2 and for N-PLS calibration. Additionally, outliers can be detected using the criteria suggested in Ref. [28].

Once the number of calibration factors is set, the number of RTL components can be estimated by examining the changes in residual fit as this number is increased. Table 4 provides such results for U-PLS/RTL when calibrating for analyte 1, which allows one to set 1 as the number of RTL components to be included in subsequent analyses, because at this number the residual fit achieves the value of the instrumental noise (0.002, see above) and does not significantly decrease on increasing the number of RTL components. The same result was obtained for analyte 2 and for N-PLS/RTL.

For the TLLS/RTL case, the number of calibrated constituents must be known from the start (two in the present case), while the number

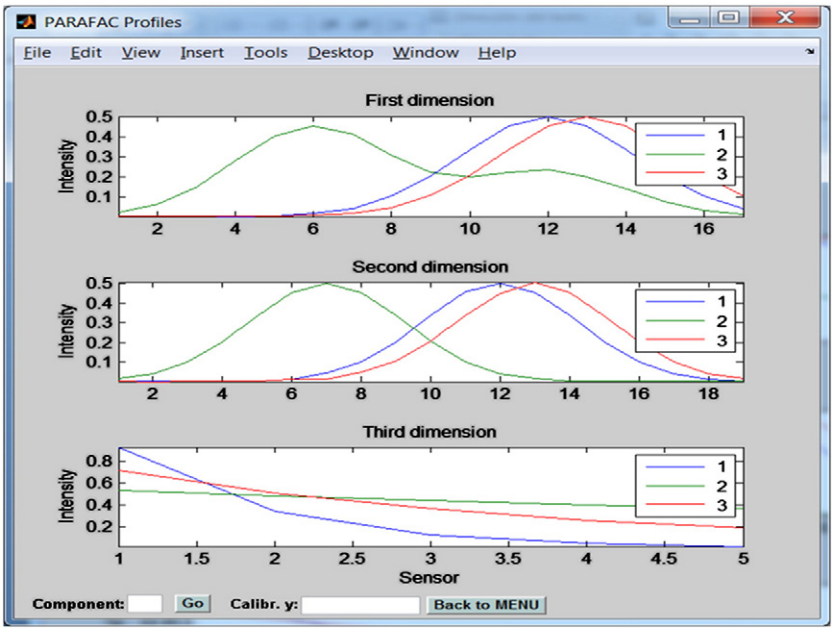


Fig. 3. Profiles retrieved by PARAFAC for the three responsive components, labeled according to their contribution to the overall spectral variance. Compare with the component profiles in Fig. 1. At the bottom of this figure, the component number corresponding to the calibrated analyte and the file name containing the calibration concentrations should be provided to the program.

Author's personal copy

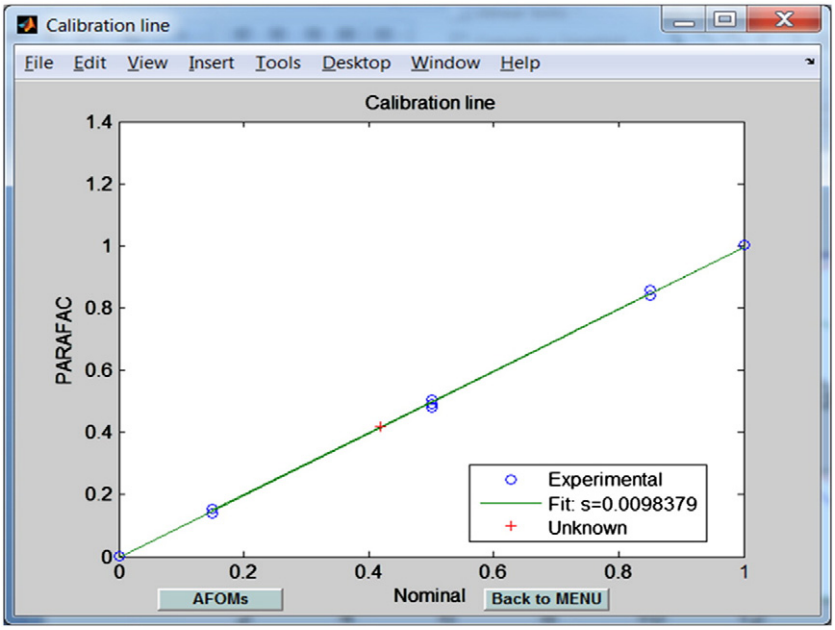


Fig. 4. Pseudo-univariate plot of PARAFAC scores for the selected component vs. nominal analyte 1 calibration concentrations. Blue circles, calibration samples, red cross, unknown sample interpolated into the calibration graph. The regression fit for the calibration graph is provided by the program.

of RTL or unexpected components can be set using a procedure similar to that described in relation with U-PLS/RTL and N-PLS/RTL. Table 4 shows that this number is also 1 for TLLS/RTL.

With both number of components established, the 'Predict' button produces a plot of profiles for the RTL components (Fig. 5). This latter figure shows that the retrieved profiles coincide with those for the interferent included in the test samples (compare with Fig. 1).

The estimated analyte concentrations are then provided by the program, which are collected in Table 2 for the RTL methods (N-PLS/RTL gave very similar results to U-PLS/RTL). They are of similar analytical quality to those obtained by employing the QLD models. As regards the standard deviations in predicted concentrations, they appear as more dispersed in comparison with the QLD models, due to the fact that for RTL models this parameter depends on all calibrated analytes. Notice that the cross-validation and prediction procedure should be repeated for each calibrate analyte for U-PLS/RTL and N-PLS/RTL, while TLLS/RTL provides the concentrations of all calibrated analytes at the same time.

Finally, the comparison of figures of merit shows similar values to those furnished by the QLD models (Table 3). They allow for similar correlations with the RMSEP parameters collected in Table 2 for the predicted analyte concentrations.

The U-PCA/RTL model is unique in that no concentration information is obtained, because the aim of this model is to produce calibration and test sample scores to be employed in the training of a suitable non-linear multivariate model, such as neural networks. The idea is to free the test sample data from the effect of unexpected sample components. When the button 'Unfolded principal component analysis' is activated, a variety of statistical indicators are provided, in order to select the correct number of principal components to model the calibration sample set.

Table 2
Predicted concentrations for both analytes using the QLD and RTL methods in the ten samples of the test set.

Test sample	Nominal	PARAFAC, APQLD and AWRCQLD	TLLS/RTL	U-PLS/RTL
Analyte 1[a]				
1	0.408	0.419 (3)	0.417 (6)	0.421 (5)
2	0.392	0.395 (3)	0.394 (4)	0.394 (4)
3	0.334	0.328 (3)	0.327 (4)	0.331 (6)
4	0.370	0.369 (3)	0.368 (4)	0.370 (1)
5	0.600	0.610 (3)	0.608 (4)	0.611 (6)
6	0.544	0.558 (3)	0.556 (5)	0.558 (5)
7	0.410	0.394 (3)	0.393 (5)	0.395 (4)
8	0.366	0.375 (3)	0.373 (4)	0.374 (4)
9	0.564	0.570 (3)	0.568 (6)	0.571 (5)
10	0.216	0.222 (3)	0.222 (8)	0.226 (6)
RMSEP		0.01	0.01	0.01
Analyte 2[a]				
1	0.714	0.68 (1)	0.69 (2)	0.65 (2)
2	0.206	0.23 (1)	0.24 (1)	0.24 (1)
3	0.782	0.77 (1)	0.78 (2)	0.72 (2)
4	0.402	0.41 (1)	0.42 (1)	0.40 (1)
5	0.716	0.71 (1)	0.72 (1)	0.69 (2)
6	0.590	0.56 (1)	0.58 (1)	0.55 (2)
7	0.512	0.53 (1)	0.53 (2)	0.51 (1)
8	0.276	0.26 (1)	0.27 (1)	0.27 (1)
9	0.710	0.69 (1)	0.71 (2)	0.67 (2)
10	0.790	0.78 (1)	0.79 (2)	0.72 (2)
RMSEP		0.02	0.02	0.04

[a] Standard deviation in concentration between parenthesis.

Table 3
Analytical figures of merit for all methods in typical samples of the simulated test set.[a].

AFOM	PARAFAC, APQLD and AWRCQLD		TLLS/RTL[a,b]		U-PLS/RTL[a,b]	
	Analyte		Analyte		Analyte	
	1	2	1	2	1	2
SEN	0.89	0.21	0.89	0.22	0.86	0.20
SEL	0.89	0.21	0.89	0.22	–	–
γ	440	110	450	110	400	92
LOD	0.007	0.03	0.008	0.02	0.008	0.04

[a] Figures of merit have been computed for the test sample No. 1. To estimate γ and LOD, the standard deviation of signal residuals was 0.01.
[b] LOD estimated from the sample of lowest analyte concentration.

Author's personal copy

Table 4
Leave-one-out cross-validation (U-PLS) and RTL results (TLLS/RTL and U-PLS/RTL) for analyte 1 when processing the test sample No. 1.

Cross-validation results			
Latent variables	U-PLS[a]		
	PRESS	*F*	*p*
1	0.49041	505	0.999
2	0.00103	1.06	0.533
3	0.00097	1	0.5
4	0.00097	–	–
5	0.00097	–	–

RTL results		
RTL components	TLLS/RTL	U-PLS/RTL[b]
	SD	SD
0	0.011	0.011
1	0.002	0.002
2	0.002	0.002
Calibration residue	0.001	0.002

[a] Calibrating for analyte 1.
[b] Using 2 calibration latent variables and calibrating for analyte 1.

8. QLD and RTL models for experimental data

The experimental data set was analyzed using all models, following the general prescriptions detailed above in connection with the simulated data set. In this case a single analyte occurs, whose response is embedded in the background signal of the test samples. The three QLD models (PARAFAC, APQLD and AWRCQLD) all required two components for successful decomposition (typical emission, excitation and time profiles are shown in Fig. 6). Once selected the component ascribed to the analyte, the pseudo-univariate plot of scores vs. concentrations is produced (see Fig. 7 for a typical plot produced by PARAFAC for a certain test sample). As can be seen, the relationship between scores and concentrations is not linear, but can be adequately described by a simple second-degree polynomial expression (a short MATLAB code for performing this calibration is provided along with the data set). This permits quantitation of the analyte concentration in the eight test samples, whose values are quoted in Table 5. In this case the results from the three QLD methods are almost identical, with good success in analyte quantitation, even due to the non-linear nature of the signal–concentration relationship.

In the case of TLLS/RTL, a single analyte occurs, and a single RTL component is needed to model the signal of the interferent in all samples. This is in agreement with the above QLD results, which required two components (one analyte and one interferent). U-PLS/RTL, on the other hand, requires a single latent variable for modeling the calibration data (as dictated by cross-validation analysis). Since TLLS and U-PLS are basically linear, analyte quantitation proceeds with moderate accuracy (see Table 5).

These experimental data illustrate the subtleties and complications which may arise in real-world third-order multivariate applications to experimental analytical systems. It should be noticed that the above discussed experimental system can also be tackled by non-linear calibration models based on artificial neural networks, which are universal non-linear approximators, combined with RTL. The reader may found details on this subject in Ref. [16].

9. Conclusions

MVC3, a new flexible and free MATLAB toolbox is described for the implementation of many third-order multivariate calibration methods. Calibration and prediction of unknown samples are performed from a single window without requiring any extra efforts. Outlier detection, prediction confidence intervals and analytical figures of merit are readily obtained. These characteristics make the new toolbox useful for less experienced users which would like to adequately process third-order instrumental data with analytical purposes.

10. Validations

Two independent researchers have tested the software, and their comments are provided below.

Prof. Olivieri's Group has developed the MVC3 toolbox for third-order multivariate calibration, which consists of up to 7 different third-order multivariate calibration methodologies through easily managed graphical user interfaces. After preparing the dataset following the specifications in MVC3's manual, calibration, validation,

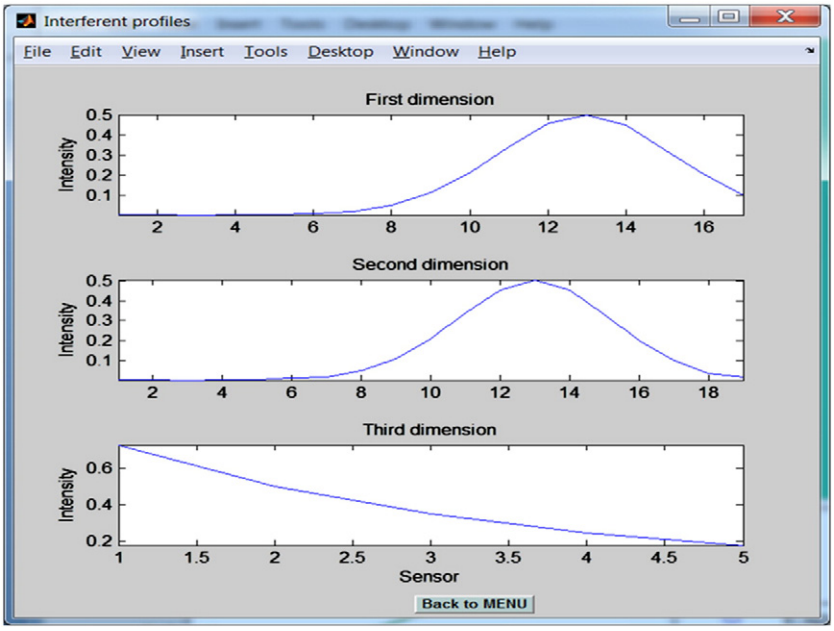


Fig. 5. Profiles in both dimensions for the RTL component, as retrieved by the U-PLS/RTL method. Compare with the black interferent profiles in Fig. 1.

Author's personal copy

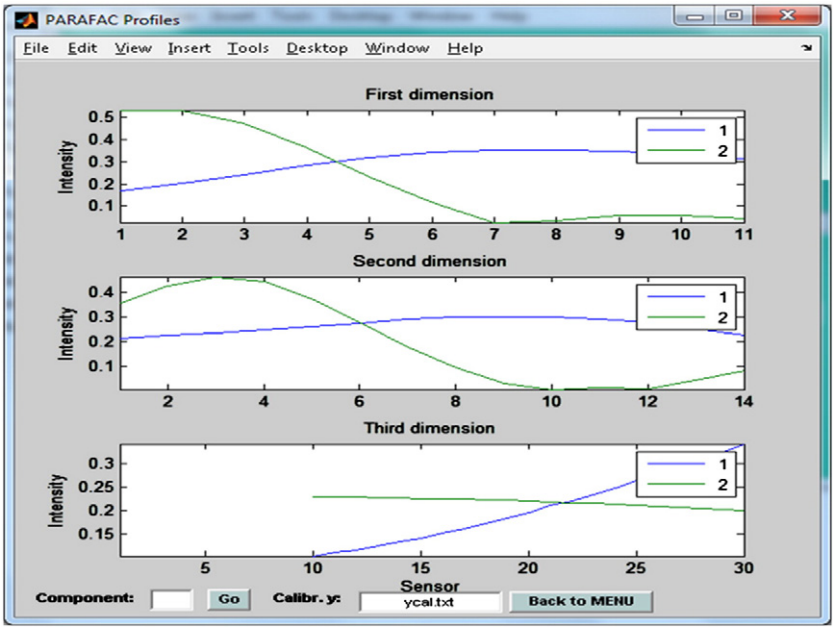


Fig. 6. Profiles recovered by PARAFAC for both components of the experimental data set.

prediction and model selection can be performed within the main window intuitively. This work will bridge the gap between pure chemometricians and end users. There are a few issues which I would like to see addressed in the future: the first is MVC3 depends on the PARAFAC and N-PLS packages, it seems necessary for the authors to integrate the used functions of PARAFAC and N-PLS into MVC3 to make easier in usage and installation. The other is the dataset given in the toolbox seems rather simple, the authors should add more examples in the manual and datasets into the toolbox. Then, the readers and users can easily know the limitation and application domain of each method in the toolbox.

The MVC3 graphical interface software can be employed for the processing of four-way analytical data, and implements several third-order multivariate calibration algorithms. These algorithms include those based on quadrilinear decomposition (QLD): PARAFAC, APQLD and AWRCQLD, and those based on residual trilinearization (RTL): TLLS, U-PLS, N-PLS, and U-PCA, all in combination with the RTL algorithm. The toolbox has been already used in several analytical applications since 2004, although including only several of the selected prediction algorithms in older versions of the MVC3 interface. I had the opportunity of using the MVC3 graphical interface and I can confirm that the software is working fine in different analytical

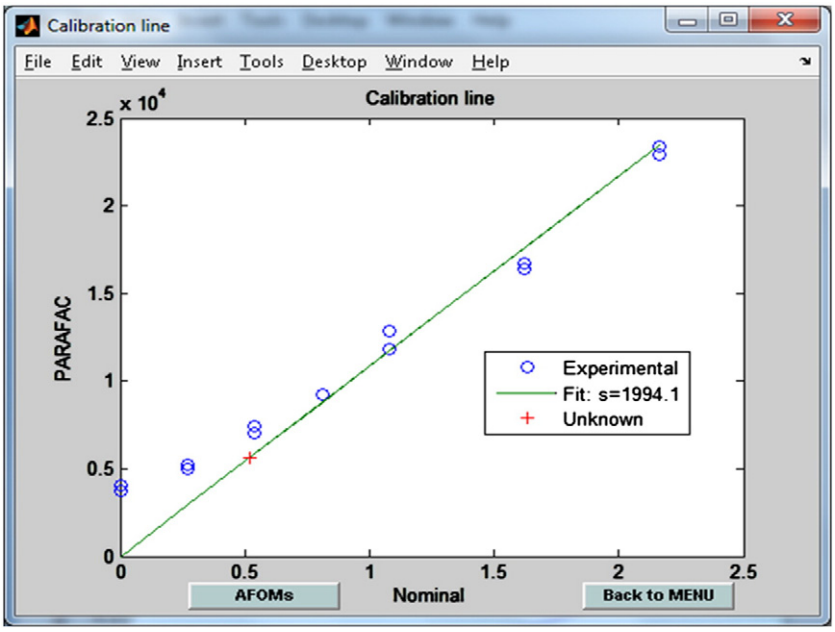


Fig. 7. Pseudo-univariate PARAFAC plot of scores for the selected component vs. nominal analyte calibration concentrations in the experimental data set. Blue circles, calibration samples, red cross, unknown sample. Notice that pseudo-univariate calibration should be implemented in this case by fitting the scores to the nominal concentrations using a polynomial rather than a straight line.

Author's personal copy

Table 5
Predicted analyte concentrations using the QLD and RTL methods in the eight samples of the experimental test set. All concentration values are given in mg L^{-1}.

Test sample	Nominal	PARAFAC, APQLD and AWRCQLD[a]	TLLS/RTL[b]	U-PLS/RTL[c]
1	0.00	0.05	0.14	0.19
2	0.00	0.04	0.14	0.18
3	0.54	0.47	0.46	0.49
4	0.54	0.49	0.48	0.51
5	1.08	0.99	0.90	0.91
6	1.08	1.13	1.04	1.03
7	1.62	1.59	1.55	1.50
8	1.62	1.51	1.46	1.40
RMSEP		0.07	0.12	0.14
REP%		8.6	15	17

[a] Using a two-component quadrilinear model. Mean centering was applied to remove background signals.
[b] Using mean centering and a single RTL component to model the interferent.
[c] Using one calibration latent variable, mean centering and a single RTL component.

situations with data of different complexities. The new version of the MVC3 toolbox is now including a higher number of prediction algorithms, as those based on APQLD and AWRCQLD have been added. The MVC3 graphical interface toolbox is of interest to scientists that need to perform multivariate calibration with any class of four-way data.

Acknowledgments

Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project no. PIP 1950) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project no. PICT-2010-0084) are gratefully acknowledged for financial supports. The providers of the function subroutines APQLD and AWRCQLD (Prof. Hai-Long Wu & Prof. Ru-Qin Yu, Hunan University, China) gratefully acknowledge the National Natural Science Foundation of China (grant no. 21175041), the National Basic Research Program (grant no. 2012CB910602) and the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) for financial supports.

References

[1] G.M. Escandar, N.M. Faber, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, R.J. Poppi, Second and third-order multivariate calibration: data, algorithms and applications, Trends in Analytical Chemistry 26 (2007) 752–765.
[2] A.C. Olivieri, Analytical advantages of multivariate data processing. One, two, three, infinity? Analytical Chemistry 80 (2008) 5713–5720.
[3] A.C. Olivieri, G.M. Escandar, A. Muñoz de la Peña, Second- and higher-order multivariate calibration methods applied to non multi-linear data. Advantages and limitations of the different algorithms, Trends in Analytical Chemistry 30 (2011) 607–617.
[4] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Analytical Chemistry 78 (2006) 2700–2709.
[5] H. Parastar, J.R. Radovic, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and quantification of complex mixtures of polycyclic aromatic hydrocarbons in heavy fuel oil sample by means of GC×GC-TOFMS combined to multivariate curve resolution, Analytical Chemistry 83 (2011) 9289–9297.
[6] K.M. Pierce, S.P. Schale, Predicting percent composition of blends of biodiesel and conventional diesel using gas chromatography–mass spectrometry, comprehensive two-dimensional gas chromatography–mass spectrometry, and partial least squares analysis, Talanta 83 (2011) 1254–1259.
[7] S.E.G. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Analysis of four-way two-dimensional liquid chromatography-diode array data: application to metabolomics, Analytical Chemistry 78 (2006) 5559–5569.
[8] H.P. Bailey, S.C. Rutan, Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC–DAD analysis of human urine, Chemometrics and Intelligent Laboratory Systems 106 (2011) 131–141.
[9] A.C. Olivieri, J.A. Arancibia, A. Muñoz de la Peña, I. Durán-Merás, A. Espinosa Mansilla, Second-order advantage achieved with four-way fluorescence excitation–emission-kinetic data processed by parallel factor analysis and trilinear least-squares. Determination of methotrexate and leucovorin in human urine, Analytical Chemistry 76 (2004) 5657–5666.
[10] A. Muñoz de la Peña, I. Durán-Merás, A. Jimenez Girón, H.C. Goicoechea, Evaluation of unfolded-partial least-squares coupled to residual trilinearization for four-way calibration of folic acid and methotrexate in human serum samples, Talanta 72 (2007) 1261–1268.
[11] A. Muñoz de la Peña, I. Durán Merás, A. Jiménez Girón, Four-way calibration applied to the simultaneous determination of folic acid and methotrexate in urine samples, Analytical and Bioanalytical Chemistry 385 (2006) 1289–1297.
[12] S.H. Zhu, H.L. Wu, A.L. Xia, J.F. Nie, Y.C. Bian, C.B. Cai, R.Q. Yu, Excitation–emission-kinetic fluorescence coupled with third-order calibration for quantifying carbaryl and investigating the hydrolysis in effluent water, Talanta 77 (2009) 1640–1646.
[13] R.M. Maggio, P.C. Damiani, A.C. Olivieri, Four-way kinetic-excitation-emission fluorescence data processed by multi-way algorithms. Determination of carbaryl and 1-naphthol in water samples in the presence of fluorescent interferents, Analytica Chimica Acta 677 (2010) 97–107.
[14] P.C. Damiani, I. Durán Merás, A.G. García Reiriz, A. Jiménez Girón, A. Muñoz de la Peña, A.C. Olivieri, Multiway partial least-squares coupled to residual trilinearization: a genuine multidimensional tool for the study of third-order data. Simultaneous analysis of procaine and its metabolite p-aminobenzoic acid in equine serum, Analytical Chemistry 76 (2007) 6949–6958.
[15] A. Jiménez Girón, I. Durán Merás, A. Espinosa Mansilla, A. Muñoz de la Peña, F. Cañada Cañada, A.C. Olivieri, On line photochemically induced excitation-emission-kinetic four-way data. Analytical application for the determination of folic acid and its two main metabolites in serum by U-PLS and N-PLS/residual trilinearization (RTL) calibration, Analytica Chimica Acta 622 (2008) 94–103.
[16] A.G. García Reiriz, P.C. Damiani, A.C. Olivieri, F. Cañada Cañada, A. Muñoz de la Peña, Nonlinear four-way kinetic-excitation-emission fluorescence data processed by a variant of parallel factor analysis and by a neural network model achieving the second-order advantage: malonaldehyde determination in olive oil samples, Analytical Chemistry 80 (2008) 7248–7256.
[17] H.C. Goicoechea, S. Yu, A.C. Olivieri, A.D. Campiglia, Four-way data coupled to parallel factor model applied to environmental analysis: determination of 2,3,7,8-tetrachloro-dibenzo-para-dioxin in highly contaminated waters by solid-liquid extraction laser-excited time-resolved Shpol'skii spectroscopy, Analytical Chemistry 77 (2005) 2608–2616.
[18] MATLAB, The Mathworks Inc., Natick, Massachusetts, USA, 2010.
[19] A.C. Olivieri, H.C. Goicoechea, F.A. Iñón, MVC1: an integrated Matlab toolbox for first-order multivariate calibration, Chemometrics and Intelligent Laboratory Systems 73 (2004) 189–197.
[20] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemometrics and Intelligent Laboratory Systems 76 (2005) 101–110.
[21] P.J. Gemperline, E. Cash, Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with Penalty Functions, Analytical Chemistry 75 (2003) 4236–4243.
[22] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC2: a MATLAB graphical interface toolbox for second-order multivariate calibration, Chemometrics and Intelligent Laboratory Systems 96 (2009) 246–251.
[23] H.C. Goicoechea, A.C. Olivieri, MULTIVAR. A program for multivariate calibration incorporating net analyte signal calculations, Trends in Analytical Chemistry 19 (2000) 599–605.
[24] R. Bro, PARAFAC: tutorial and applications, Chemometrics and Intelligent Laboratory Systems 38 (1997) 149–171.
[25] A.L. Xia, H.L. Wu, S.F. Li, S.H. Zhu, L.Q. Hu, R.Q. Yu, Alternating penalty quadrilinear decomposition algorithm for an analysis of four-way data arrays, Journal of Chemometrics 21 (2007) 133–144.
[26] H.Y. Fu, H.L. Wu, Y.J. Yu, L.L. Yu, S.R. Zhang, J.F. Nie, S.F. Li, R.Q. Yu, A new third-order calibration method with application for analysis of four-way data arrays, Journal of Chemometrics 25 (2011) 408–429.
[27] J.A. Arancibia, A.C. Olivieri, D. Bohoyo Gil, A. Muñoz de la Peña, I. Durán-Merás, A. Espinosa Mansilla, Trilinear least-squares and unfolded-PLS coupled to residual trilinearization: new chemometric tools for the analysis of four-way instrumental data, Chemometrics and Intelligent Laboratory Systems 80 (2006) 77–86.
[28] D.M. Haaland, E.V. Thomas, Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data, Analytical Chemistry 60 (1988) 1193–1202.
[29] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Uncertainty estimation in spectroscopic multivariate calibration, Pure and Applied Chemistry 78 (2006) 633–661.
[30] A.C. Olivieri, N.M. Faber, A closed-form expression for computing the sensitivity in second-order bilinear calibration, Journal of Chemometrics 19 (2005) 583–592.
[31] A.C. Olivieri, N.M. Faber, in: S. Brown, R. Tauler, B. Walczak (Eds.), Validation and error, Comprehensive Chemometrics, 3, Elsevier, Amsterdam, 2009, pp. 91–120.
[32] A.C. Olivieri, Computing sensitivity and selectivity in parallel factor analysis and related multi-way techniques: the need for further developments in net analyte signal theory, Analytical Chemistry 77 (2005) 4936–4946.
[33] A.C. Olivieri, N.M. Faber, New developments for the sensitivity estimation in third-order multivariate calibration with parallel factor analysis. Analytical Chemistry. (in press, http://dx.doi.org/10.1021/ac202268k).
[34] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, Journal of Chemometrics 17 (2003) 274–286.