Chemometrics and Intelligent Laboratory Systems 173 (2018) 21–29

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Short communication

MVC3_GUI: A MATLAB graphical user interface for third-order multivariate calibration. An upgrade including new multi-way models

Sarmento J. Mazivila [a], Santiago A. Bortolato [b], Alejandro C. Olivieri [b,*]

[a] Centro de Investigação em Química (CIQ), Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, 4169-007, Portugal
[b] Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario e Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, S2002LRK Rosario, Argentina

ARTICLE INFO

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

ABSTRACT

An upgrade is presented of a MATLAB graphical user interface toolbox for implementing third-order/four-way multivariate calibration models. The new Multivariate Calibration 3 (MVC3_GUI) incorporates new models and features that make it a very versatile tool for four-way data processing. In addition to the quadrilinear decomposition (QLD) and latent structure models based on partial least-squares regression and residual trilinearization, included in the earlier software version, non-QLD models are now available. The latter include extended multivariate curve resolution-alternating least-squares (MCR-ALS), augmented parallel factor analysis (Augmented PARAFAC) and PARAFAC2. The software is presented as both a set of MATLAB codes and as a standalone program. MVC3_GUI accepts a variety of ASCII data for input. Appropriate working sensor regions in the different data modes 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. Different examples of use of this updated interface are given in this work.

1. Introduction

The first version of a graphical user interface (GUI) for third-order multivariate calibration (MVC3) was published in 2012 [1]. It was conceived to manage several different models and algorithms in an easy-to-use graphical interface environment, and can be applied to any type of data which are structured as three-dimensional arrays for each sample.

Higher-order/multi-way multivariate calibration has become ubiquitous in the field of analytical chemistry, as instrumental hyphenation is implemented for increasing sensitivity and selectivity of the determinations. Particularly, in the case of third-order calibration, a book [2] and recent reviews have highlighted the progress in both instrumentation and data processing which has taken place [3–7]. Montemurro et al. [4] reported an interesting review related to the generation of third-order liquid chromatography (LC) data with detection by excitation-emission fluorescence matrices (EEFM). The potential of different third-order multivariate calibration models was shown; some of them are implemented in the presently discussed interface.

Among the most recent contributions, those based on chromatographic approaches can be mentioned: (1) two-dimensional gas chromatography (GC–GC) with mass spectrometric (MS) detection [8,9] or two-dimensional liquid chromatography (LC–LC) with UV–visible diode array detection (DAD) [10,11], and (2) LC with EEFM [12–16]. Some of these previous works have employed the new MVC3_GUI software [12–16].

The new version offers the same models included in the old interface, namely those based on quadrilinear decomposition (QLD) and residual trilinearization (RTL) [1]. In addition, data processing models deviating from multilinearity, such as multivariate curve resolution with alternating least-squares (MCR-ALS), Augmented parallel factor analysis (Augmented PARAFAC) and PARAFAC2 and are now included.

Briefly, all multivariate calibration models included in the new MVC3 can be divided in three relevant groups, (1) the QLD approaches: PARAFAC [17], alternating penalty QLD (APQLD) [18] and alternating weighted residual constraint QLD (AWRCQLD) [18], (2) the RTL-based approaches: trilinear least-squares followed by RTL (TLLS/RTL) [19], unfolded partial least-squares/RTL (U-PLS/RTL) [19], multidimensional

* Corresponding author. Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario e Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, S2002LRK Rosario, Argentina.
E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri).

https://doi.org/10.1016/j.chemolab.2017.12.012
Received 23 October 2017; Received in revised form 22 November 2017; Accepted 19 December 2017
Available online 24 December 2017

Table 1
Summary of new options included in MVC3_GUI.

Innovation	Details
Two software alternatives	MATLAB codes or standalone program
New plotting options	Landscapes and contour plots
Non-quadrilinear decomposition models	MCR-ALS [22], Augmented PARAFAC [14], PARAFAC2 [23]
Constraints	Area correlation constraint in MCR-ALS [32]
New figures of merit	According to the recent developments [28]

partial least-squares/RTL (N-PLS/RTL) [20], and unfolded principal component analysis/RTL (U-PCA/RTL) [21], and (3) non-QLD approaches: MCR-ALS [22], Augmented PARAFAC [12] and PARAFAC2 [23]. An important incorporation to the GUI is the inclusion of constraints in the non-quadrilinear models.

A summary of the features included in the new version of the MVC3_GUI is listed in Table 1. The software does not require a highly experienced user, but a basic knowledge on the underlying models is advisable in order to successfully interpret the results. Appropriate examples are discussed to show the new capabilities of MVC3_GUI.

In relation to the estimation of analytical figures of merits (AFOMs), the sensitivity (SEN) is estimated on the basis of uncertainty propagation analysis [24–28]. The employed expression is a specific version of a general mathematical equation encompassing all possible degrees of data complexity and multi-way models [28]. As will be explained in more detail, all appropriate recommendations were followed for reporting additional figures of merit in the new MVC3_GUI, such as selectivity (SEL), analytical sensitivity (γ), limit of detection (LOD) and standard deviation (SD) in concentration for each predicted concentration.

Additional software is available for conducting multi-way data processing: MCR-ALS (https://mcrals.wordpress.com/, freely available) and PARAFAC, N-PLS and MCR (http://www.eigenvector.com/software/pls_toolbox.htm, commercial software). The presently described software is more specifically directed to analytical calibration with third-order data.

2. Software

The software runs under MATLAB version 2012 [29]. The files need to be copied into a folder declared in the MATLAB path as described in the document named 'MVC3_manual.pdf' which is provided with the software. The codes are freely available, along with examples and manual, from www.iquir-conicet.gov.ar/descargas/mvc3.rar.

The MATLAB codes for APQLD and AWRCQLD were written by Hai-Long Wu and Ru-Qin Yu, State Key Laboratory of Chemo/Biosensing & Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China, and were already incorporated into the old MVC3 version [1]. The N-way toolbox has been generously made public by Rasmus Bro and Claus Andersson, Copenhagen University, DK-1958 Frederiksberg, Denmark, and freely available in http://www.models.life.ku.dk/algorithms (accessed September 2017). This has been properly acknowledged in the corresponding program codes. The codes for U-PLS, RTL, MCR-ALS and Augmented PARAFAC were written by the authors. In addition, a compiled version is available at the following site:

https://www.dropbox.com/sh/nruf3lp0ge1gbww/AAAj6r97UBMIhgQmukRGYFPKa?dl=0, including detailed instructions for installation.

3. Data examples

Four sets of simulated data are provided along with the program and manual: (1) KinEEFM, (2) KinEEFM_IF, (3) LCEEFM and (4) LCLCDAD. The first two sets mimic the kinetic evolution of EEFM measurements, in the absence and presence of inner filter effects, and can be successfully processed with the old MVC3 program [1]. The new simulated third-order data sets involve, respectively: (1) liquid chromatography with excitation-emission fluorescence matrix detection (LCEEFM) and (2) two-dimensional liquid chromatography – diode array detection (LCLCDAD).

The LCEEFM and LCLCDAD cases include a set of calibration samples containing two analytes, and test samples containing the analytes and a single interferent. Nine calibration samples with a central composite design for the concentrations of both constituents (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 models.

Fig. 1 shows the noiseless profiles from which the simulated data were built. All of them are normalized to unit 2-norm. Notice that a strong overlapping occurs between the profiles of analyte 1 and the interferent. Random Gaussian noise with 0.1 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.

The systems are non-quadrilinear because chromatographic shifts have been considered for LCEEFM and LCLCDAD [2]. For LCEEFM, small and random chromatographic shifts exist in the calibration set and in test samples from 1 to 5. However, large chromatographic shifts appear in test samples 6 to 10. In LCLCDAD, small and random chromatographic shifts occur in the calibration set, but in all samples of the test set the peaks are shifted by a constant value, larger than the calibration shifts.

4. General description

As in the old interface, MVC3_GUI has a single main window (Fig. 2), from which all steps required to implement the different third-order multivariate calibration strategies can be conducted. The starting point is to select the desired multivariate calibration model and the corresponding number of responsive components (namely, total number of components for QLD and non-QLD models, or the numbers of calibrated and unexpected sample components for UPLS/RTL, N-PLS/RTL and U-PCA/RTL).

Once the model is selected, it is necessary to define the structure of the data in order to continue. The same layout options of the old MVC3 interface have been kept: 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.

Several different data types are admissible, all contained in ASCII files, whose names should be provided to the graphical interface, or selected from a suitable browser. 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 unfolded and saved as a $JKL \times 1$ vector, where J, K and L are number of data points in each mode, 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).

As a novelty, the new MVC3_GUI allows one to plot the sample signals, both as landscape and contour plots (Fig. 3). This option allows for a better interpretation of the data set, as well as for a preliminary checkpoint of proper program execution, even for untrained users. After pressing the 'PLOT' button in the main window, a new window provides the landscape and contour plots of the data according to the instrumental sequence from which it was generated.

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

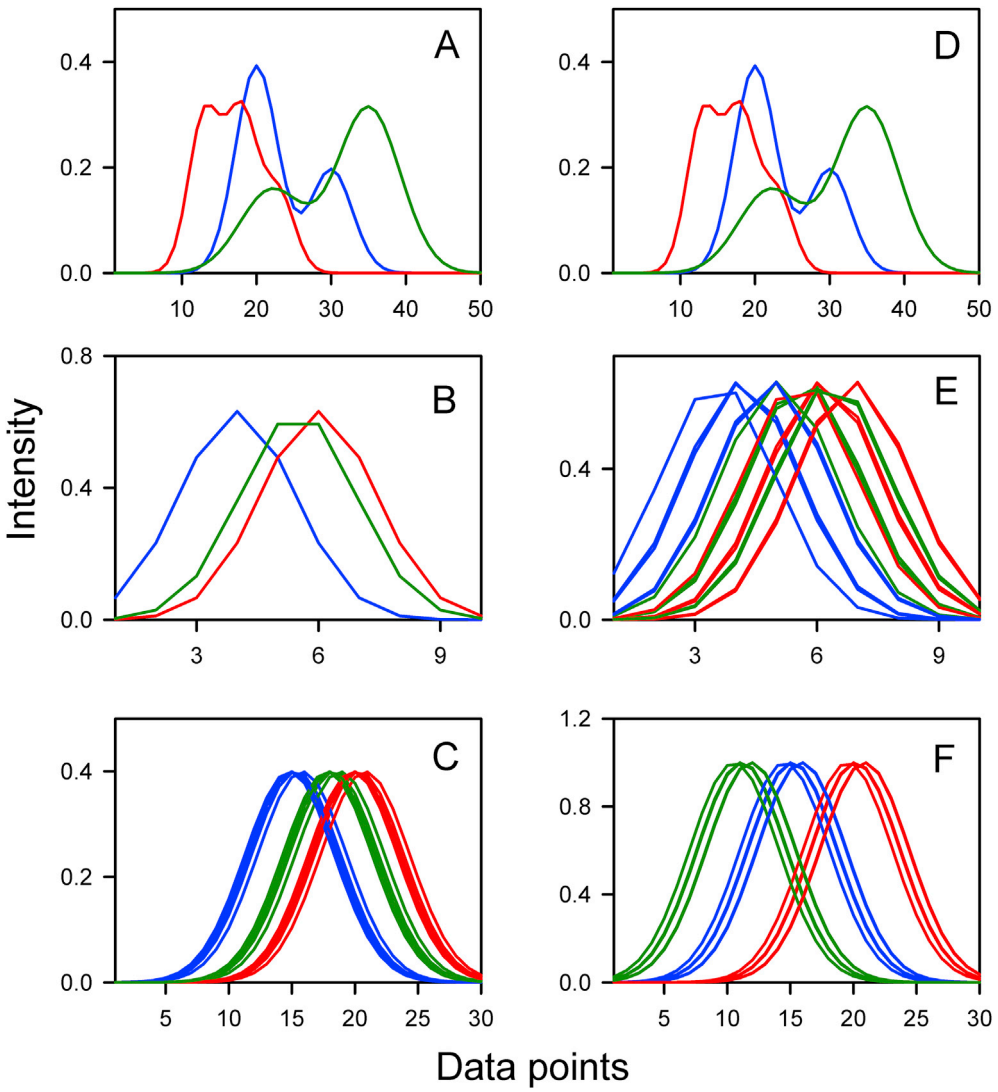

Fig. 1. Pure component profiles used to build the simulated data sets. In all cases, the blue lines identify analyte 1, the green lines analyte 2 and the red line the potential interferent which is only present in the test samples. A), B) and C), emission, excitation and elution time profiles for the LCEEFM data set respectively. Notice in C) the various profiles corresponding to the random shifts which occur in the first five test samples of the set. D), E) and F), absorption, elution time 1 and elution time 2 profiles for the LCLCDAD data set respectively. Random shifts in peak positions occur in both time modes E) and F). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

subsequently be used in the estimation of certain AFOMs, such as SD in concentrations, LOD, 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.

The four-way data modes are identified with letters: A corresponds to the sample mode, and B, C and D for the instrumental modes. Augmentation modes for MCR-ALS and PARAFAC2 can be selected from six different possibilities, i.e., along B, C, D, and along the combinations B&C, B&D and C&D (if the data arrays need to be unfolded along combinations of two modes), whereas for Augmented PARAFAC the available augmentation modes are the instrumental ones, i.e., B, C or D. This lettering identification is also important for additional activities such as the application of constraints (see below).

Table 2 presents a comprehensive snapshot of all program functions. The program can be thought as a sequence of six steps, which are articulated from the following simple guidelines: (1) define the model to be used (quadrilinear or non-quadrilinear), (2) select the data type and sensor regions, (3) load the data as text files with specific names, (4) introduce the appropriate constraints and initialization procedure, if applicable, (5) define the number of components and predict the concentration of the analyte with the corresponding figures of merit; and (6) save the relevant information in a file.

Fig. 4 shows a flowchart which compactly describes the software organization.

5. Simulated data

5.1. General

This section discusses the performance of the models included in the MVC3_ GUI over two simulated systems. Independently of the model considered (QLD, RTL or non-QLD), the starting point is to define the number of responsive components. For 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 is reached when the core consistency significantly drops below 50 [30]. More generally, the residual standard deviation of the least-squares fit of the four-way data array to the quadrilinear model is very useful for this purpose, because this value stabilizes when the correct number of components has been reached [2].

For non-QLD models such as Augmented PARAFAC or PARAFAC2,

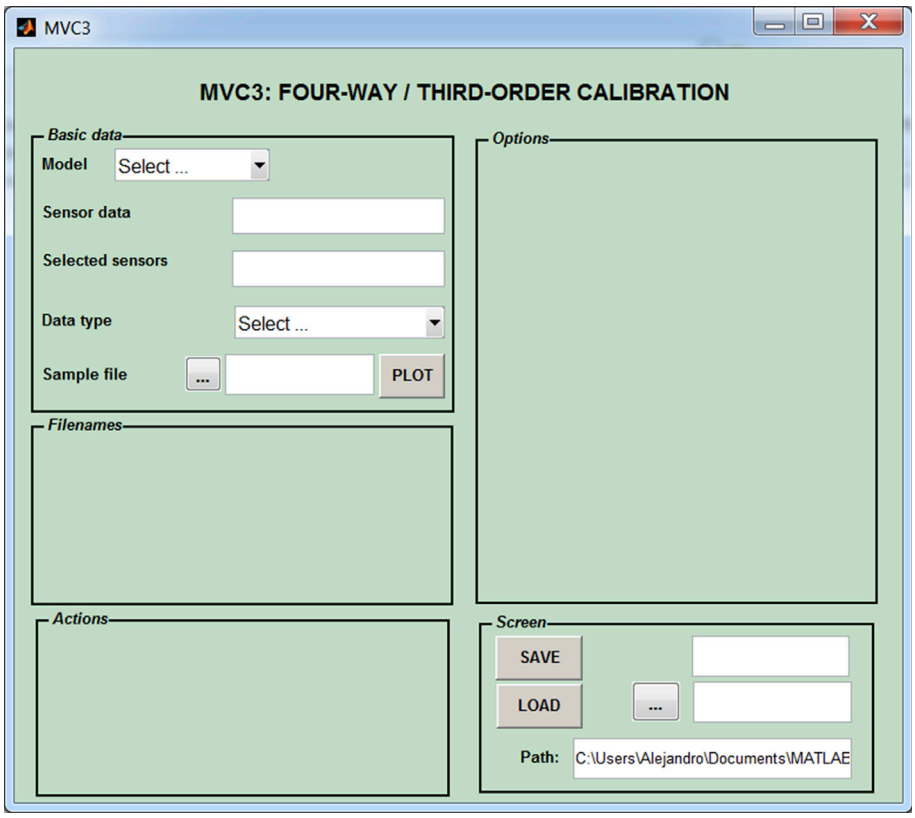

Fig. 2. Main screen of the new MVC3_GUI interface, showing the selection of the model by means of a so-called popup menu (upper left), and the screen load/save panel (bottom right).

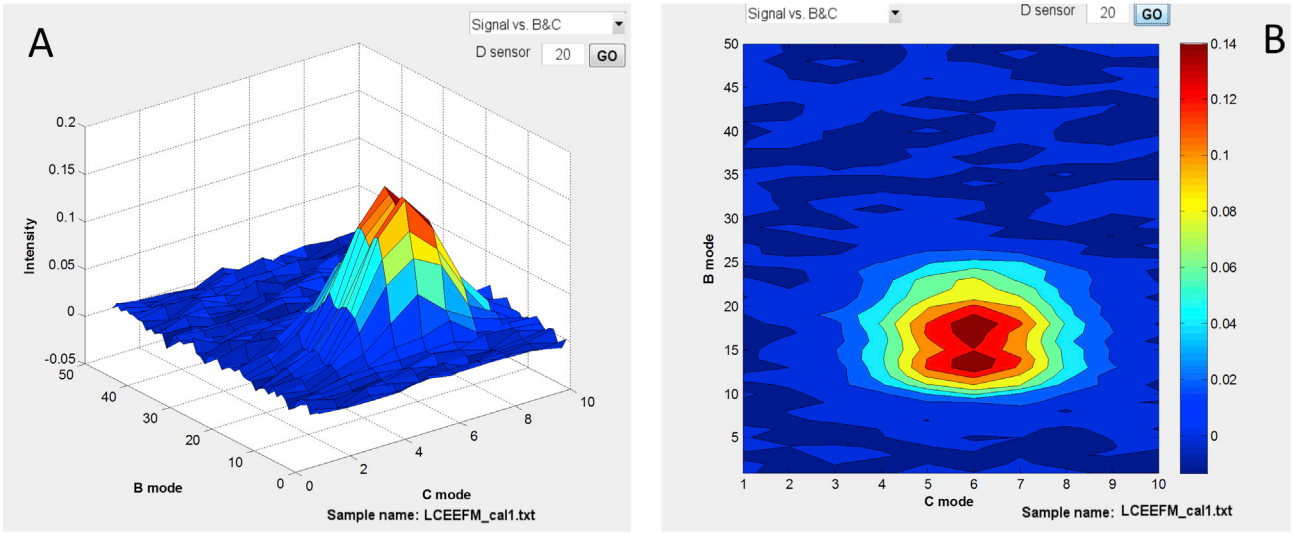

Fig. 3. A) Landscape and B) contour maps for the signals of the first calibration sample of the LCEEFM data set. The user can select the two modes for plotting and the sensor value for the third mode.

the procedure described for classical PARAFAC can be followed, i.e., core consistency diagnostic and/or analysis of the residual standard deviation of the least-squares fit. In MCR-ALS, however, clicking in 'ESTIMATE COMPONENTS' leads to principal component analysis of the super-augmented matrix to estimate the correct number of components, considering the explained variance and residual fit as a function of the number of principal components [31].

With respect to initialization, approximations to pure profiles can be supplied to MCR-ALS through the estimation of the so-called purest variables, implemented through an auxiliary display, or as known experimental profiles [31]. On the other hand, Augmented PARAFAC is usually initialized by direct trilinear decomposition (DTLD) or the best fit of a reduced set of trial runs. This latter option is also preferred for PARAFAC2 (singular value decomposition vectors are also available).

Constraints such as non-negativity and unimodality can be imposed from the main MVC3_GUI screen in the new non-QLD models. This will force the profiles to be non-negative and/or to have a single maximum. However, for MCR-ALS and Augmented PARAFAC, additional options are available after pressing in 'PREDICT': the user can impose these restrictions for specific components in each mode independently ('1' and

Table 2
Comprehensive snapshot of all program functions.

		Quadrilinear decomposition (QLD)	Residual trilinearization (RTL)	Non-quadrilinear decomposition (Non-QLD)		
Step 1 Model type	Model	PARAFAC, APQLD, AWRCQLD	U-PLS, N-PLS, TLLS, U-PCA	Augmented PARAFAC	MCR-ALS	PARAFAC2
Step 2 Basic Data	Data structure	Sensors & Data type				
Step 3 Filenames	Filenames	Calibration sample signals Test sample signals	Analyte calibration concentrations Calibration sample signals Test sample signals	Calibration sample signals Test sample signals		
Step 4 Options	Pre-processing	Mean centering (optional)	Mean centering	None		
	Augmentation			B, C, D	B, C, D, B&C, B&D, C&D	
	Constraints	Non-negativity Unimodality Orthogonality	Particle swarm optimization (PSO) for RBL	Non-negativity Unimodality Orthogonality Interferent correspondence	Non-negativity Unimodality Correspondence Area correlation	Non-negativity Unimodality Orthogonality
	Initialization	SVD vectors Best fit of several trials	None	DTLD Random orthogonal Best fit of several trials	Purest signals Pure experimental signals	SVD vectors Random orthogonal Best fit of several trials
Step 5 Actions	Components	Set number of components	Set number of component through leave-one-out cross-validation	Set number of components		
	Predict	Predict concentration with AFOMs				
Step 6 Save		Save system information				

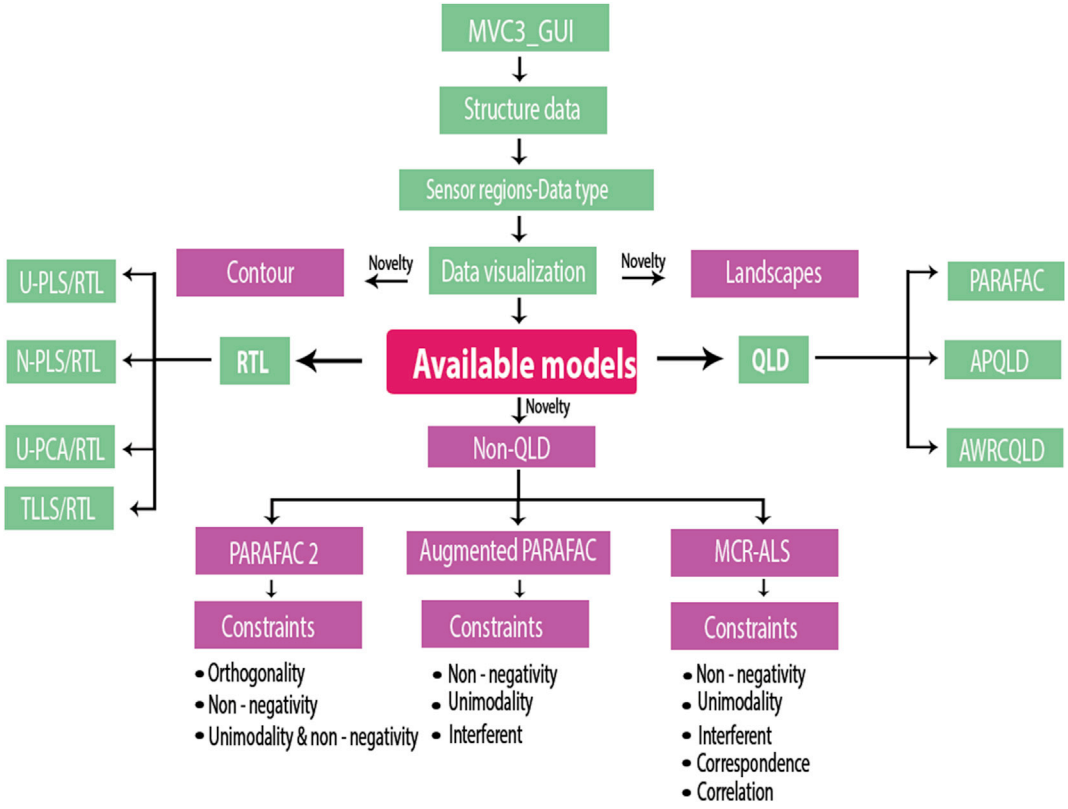

Fig. 4. Flow sheet showing the software organization.

'0' indicate that constraints are on or off). Additionally, the interferents can be identified, so as to force the model to remove them from the calibration samples. Notice that the interferent indexes may change according to the selected model and test sample under study. In any case, Augmented PARAFAC models are often unique and do not require constraints.

For MCR-ALS, the correspondence and area correlation constraints can also be applied for calibrated analytes during the optimization phase [32]. The user has to select the vector (for a single analyte) or matrix (for more than one analyte) containing nominal known calibrations, supplying the analyte indexes. These constraints remove analyte signals from the calibration samples where they are absent, and linearly regress the

areas of the resolved profiles vs. nominal analyte concentrations during the alternating least-squares optimization.

5.2. LCEEFM data set

This data set mimics liquid chromatography with matrix excitation-emission detection. A single breaking mode occurs (the elution time mode) and therefore the data are classified as non-quadrilinear type 1 [3], for which the recommended model is Augmented PARAFAC [3]. Results from PARAFAC and MCR-ALS will also be discussed for comparison purposes.

When applying Augmented PARAFAC, once the number of components is set to three (the known number of chemical constituents), the 'PREDICT' button allows one to perform the decomposition of the augmented three-way data set formed by joining calibration and test sample data (this activity is conducted sample by sample) in the same way as with the old interface [1]. No constraints were needed in this case, and a plot of separate component profiles in each mode is then produced (Fig. 5A). The latter can be satisfactorily compared with the pure synthetic ones shown in Fig. 1. Subsequently, the component number that best correlates with the known properties of the analyte is selected. Finally, a pseudo-univariate scores-concentrations plot for the selected analyte is produced and its concentration is estimated by interpolation. Fig. 6 shows the results for analyte 1 in the test sample number 1 of the LCEEFM data set. Notice that in Fig. 6 (bottom), calibration concentration errors are plotted against sample number, which is useful to spot outlying calibration samples.

In the case of MCR-ALS, matrix super-augmentation is performed in the following way: the spectral modes B and C (emission and excitation) are first unfolded into a single combined mode, generating matrix data from the original third-order data. Then the matrices are appended along the time direction (D mode). The selected super-augmentation mode from the MVC3_GUI interface is thus D. Spectral profiles were restricted to be non-negative, while non-negativity and unimodality were imposed to chromatographic profiles. Interferent and calibration correspondence, as well as area correlation constraints were also applied. Fig. 5B shows the retrieved profiles in the three modes (B and C modes were refolded after MCR-ALS decomposition). Notice the change in component indexes (compare Fig. 5A with Fig. 5B), due to the fact that each model defines the indexes in a different way. The selection of the analyte of interest and the prediction of its concentration proceed similarly to Augmented PARAFAC.

For comparison purposes, concentrations predicted for both analytes are collected in Table 3 for four-way PARAFAC (applied as in the old interface), MCR-ALS and Augmented PARAFAC, along with the estimated figures of merit. As can be seen, Augmented PARAFAC furnishes the best prediction results. This can be judged from the root mean square errors of prediction (RMSEP) and the relative error prediction (REP, in %): 0.006 concentration units (1.2%) for analyte 1 in the test samples 1 to 5 (with small and random chromatographic shifts) and 0.024 (4.8%) for analyte 2. The Augmented PARAFAC performance is also excellent for test samples 6 to 10 showing large chromatographic shifts, with RMSEP and REP values of 0.011 (2.2%) for analyte 1 and 0.004 (0.8%) for analyte 2. These RMSEP values seem reasonably low taking into account the level of noise introduced in the system, i.e., ca 1% both in concentrations and signals.

The reported figures of merit (Table 3) for Augmented PARAFAC are also satisfactory: in particular, detection and quantitation limits are on the order of the concentration uncertainty of 0.01 units.

5.3. LCLCDAD data set

This data set illustrates two-dimensional liquid chromatography with diode array detection. Here two breaking modes occur, namely both elution times, and the data are thus classified as non-quadrilinear type 2 [3]. The recommended procedure is to unfold the four-way array into a bilinear super-augmented matrix, appropriate to be modeled by extended MCR-ALS [33]. The super-augmented matrix is built along a direction given by the concatenation of two breaking modes (C&D). The additional instrumental mode is the spectral one (B).

Bilinear matrix decomposition by MCR-ALS proceeds under similar constraints, i.e., non-negativity in all modes, unimodality in each elution time mode, correspondence and area correlation. After obtaining satisfactory profiles (Fig. 7), the remaining activities are analogous to those described above.

Table 4 shows the concentrations predicted for both analytes and estimated figures of merit for PARAFAC, MCR-ALS and Augmented PARAFAC. Evaluation of the prediction performance leads to the conclusion that the best results are achieved by MCR-ALS, with RMSEP and REP values of 0.027 (5.4%) for analyte 1 and 0.033 (6.6%). The analytical performance of MCR-ALS in this case is satisfactory, in view of the overall uncertainty of ca. 1% in concentrations and signals employed to simulate the data.

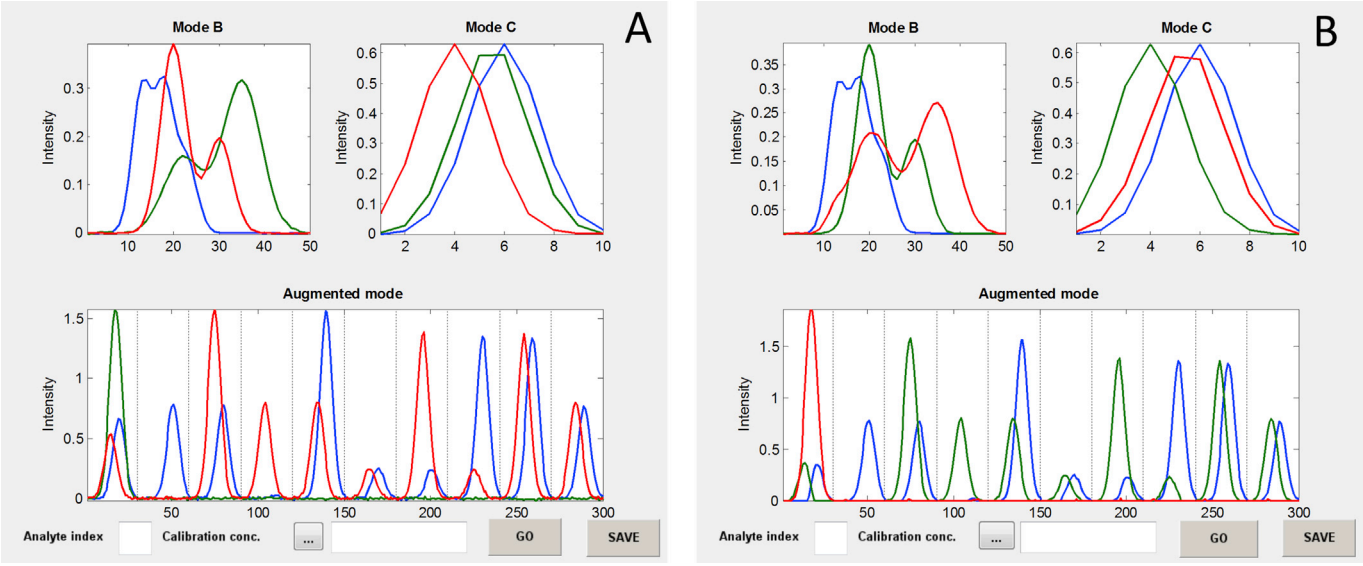

Fig. 5. A) Profiles retrieved by Augmented PARAFAC in the LCEEFM data set. Modes B and C represent the emission and excitation spectra, while the augmented mode D the sequence of elution time sub-profiles along the augmentation direction. B) Analogous profiles retrieved by MCR-ALS decomposition of the super-augmented matrix along the time mode, by first concatenating the spectral modes. After decomposition, the spectral modes have been refolded.

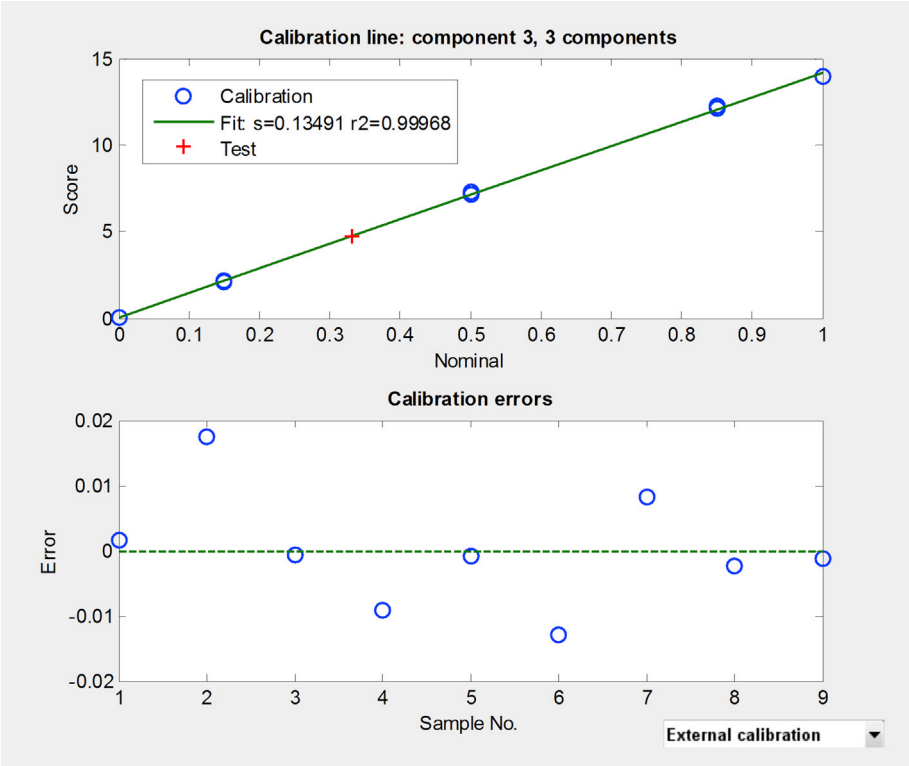

Fig. 6. Top, pseudo-univariate plot of Augmented PARAFAC scores for the analyte of interest (component 3 in the present case) as a function on nominal calibration concentrations. The red cross is the interpolation of the test sample score. Bottom, prediction errors as a function of sample number. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Predicted analyte concentrations in the test samples using third-order multivariate calibration models on LCEEFM data.[a]

Analyte	PARAFAC	MCR-ALS	Augmented PARAFAC	Analyte	PARAFAC	MCR-ALS	Augmented PARAFAC
Analyte 1 (test1-5)[b]				Analyte 2 (test 1-5)[b]			
0.33	0.33	0.19	0.33	0.42	0.39	0.18	0.42
0.42	0.46	0.54	0.43	0.06	0.07	0.12	0.10
0.71	0.72	1.03	0.71	0.28	0.29	0.32	0.30
0.21	0.23	0.30	0.22	0.22	0.22	0.26	0.23
0.46	0.45	0.30	0.46	0.83	0.89	0.37	0.80
RMSEP	0.021	0.184	0.006	RMSEP	0.031	0.235	0.024
REP	4.2	36.8	1.2	REP	6.2	47.0	4.8
LOD	0.02	0.02	0.02	LOD	0.02	0.02	0.02
LOQ	0.07	0.07	0.07	LOQ	0.07	0.07	0.07
SEN	2.40	2.34	2.00	SEN	1.72	2.34	2.10
SEL	0.59	0.83	0.58	SEL	0.44	0.86	0.71
γ	478	468	390	γ	342	470	428
Analyte 1 (test6-10)[b]				Analyte 2 (test 6-10)[b,c]			
0.65	0.60	0.72	0.64	0.70	–	0.74	0.70
0.92	0.87	1.17	0.91	0.51	–	0.60	0.51
0.89	0.91	0.88	0.89	0.59	–	0.37	0.58
0.79	0.81	0.78	0.79	0.65	–	0.45	0.65
0.41	0.27	0.40	0.39	0.29	–	0.19	0.29
RMSEP	0.071	0.116	0.011	RMSEP	–	0.147	0.004
REP	14.2	23.2	2.2	REP	–	29.4	0.8
LOD	0.02	0.02	0.02	LOD	–	0.02	0.02
LOQ	0.06	0.07	0.07	LOQ	–	0.07	0.07
SEN	2.74	2.34	2.00	SEN	–	2.36	2.10
SEL	0.70	0.84	0.58	SEL	–	0.86	0.71
γ	550	474	390	γ	–	472	430

[a] RMSEP, root mean square error of prediction, REP, relative error prediction in %, LOD, limit of detection, LOQ, limit of quantification, SEN, sensitivity, SEL, selectivity, γ, analytical sensitivity.
[b] First five entries indicate the nominal analyte concentrations.
[c] In these samples PARAFAC cannot predict the concentrations of analyte 2.

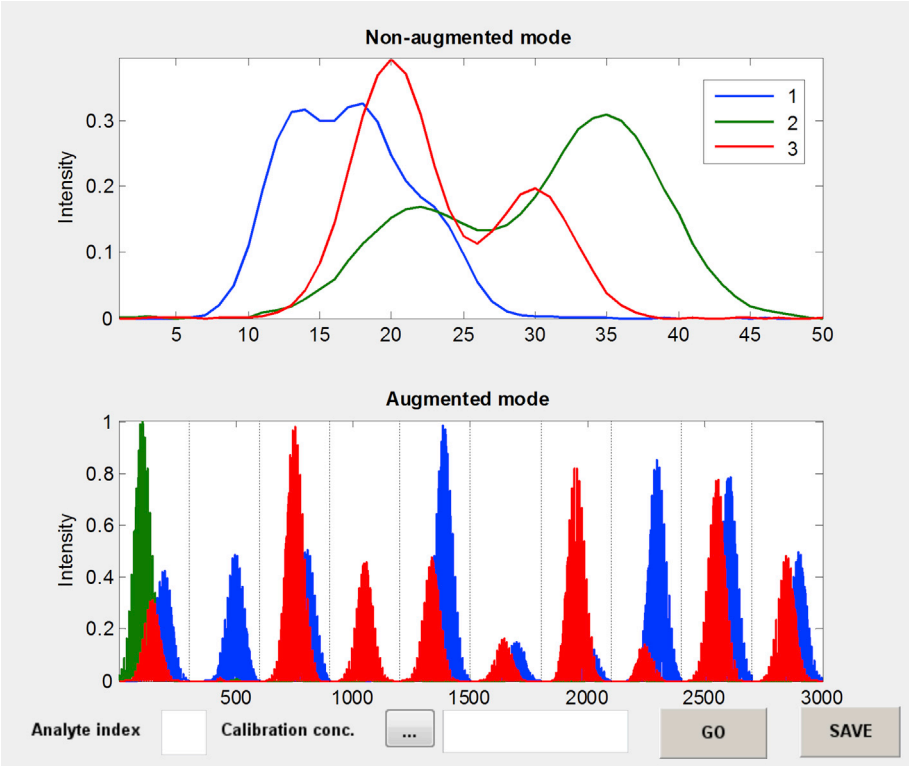

Fig. 7. Profiles retrieved by MCR-ALS in the LCLCDAD data set, after decomposition of the super-augmented matrix. Top, spectral (DAD) mode, bottom, augmented profiles corresponding to the concatenation of the C and D modes (elution time modes).

Table 4
Predicted concentrations in the test samples using third-order multivariate calibration models on LCLCDAD data.[a]

Analyte	PARAFAC	MCR-ALS	Augmented PARAFAC	Analyte	PARAFAC	MCR-ALS	Augmented PARAFAC
Analyte 1[b]				Analyte 2[b]			
0.33	0.38	0.31	0.32	0.42	0.38	0.42	0.41
0.42	0.42	0.39	0.38	0.06	–	0.03	0.02
0.71	0.73	0.69	0.69	0.28	0.25	0.28	0.28
0.21	0.24	0.22	0.20	0.22	0.15	0.22	0.20
0.46	0.57	0.43	0.45	0.83	0.80	0.83	0.83
0.65	–	0.61	0.05	0.70	0.12	0.67	0.38
0.92	0.19	0.92	0.15	0.51	0.24	0.50	0.16
0.89	0.15	0.86	0.21	0.59	0.31	0.58	0.20
0.79	0.13	0.75	0.31	0.65	0.35	0.62	0.21
0.41	0.12	0.39	0.28	0.29	0.09	0.28	0.07
RMSEP	0.402	0.027	0.408	RMSEP	0.250	0.033	0.249
REP	80.4	5.4	81.6	REP	50.0	6.6	49.8
LOD	0.12	0.02	0.06	LOD	0.14	0.02	0.07
LOQ	0.35	0.07	0.19	LOQ	0.42	0.07	0.22
SEN	1.99	1.39	1.84	SEN	2.75	1.50	2.27
SEL	0.52	0.57	0.47	SEL	0.75	0.71	0.72
γ	255	270	307	γ	341	300	371

[a] See Table 3 for the meaning of acronyms and symbols.
[b] First five entries indicate the nominal analyte concentrations.

6. Conclusion

MVC3, a new flexible and freely available MATLAB toolbox is described for the implementation of several third-order multivariate calibration models. Analyte calibration and prediction in unknown samples are performed from a single window without requiring any extra efforts. Prediction confidence intervals and AFOMs are readily obtained. These characteristics make the new toolbox useful for less experienced users which would like to adequately process third-order instrumental data for analytical purposes.

7. Validations

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

Héctor C. Goicoechea, Professor, Laboratorio de Desarrollo Analítico y Quimiometría-LADAQ, Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, CONICET, Ciudad Universitaria (3000) Santa Fe, Argentina.

The authors have notably improved the current MVC3 software presented in the paper “MVC3: A MATLAB graphical interface toolbox for third-order multivariate calibration”, A.C. Olivieri, H-L. Wu, R-Q. Yu,

Chemom. Intell. Lab. Syst. 116 (2012) 9-16. The new version of the software presents the following innovations: a) the alternative of running it under MATLAB environment or as a standalone program, b) new plotting options, c) three non-QLD models (MCR-ALS, Augmented PARAFAC and PARAFAC2) were included as modeling options, d) a new constraint (area correlation constraint) is available for MCR-ALS modeling, and e) new figures of merit can be computed following the recent ideas in this field.

I installed and used both options not only with the data provided by the authors, but also with data generated in our lab. The program works as described in the user manual in a user friendly way.

Arsenio Muñoz de la Peña, Professor, Department of Analytical Chemistry, Faculty of Sciences, University of Extremadura, 06006 Badajoz, Spain.

The authors have developed a new MATLAB based toolbox "MVC3_GUI" which allows one to carry out various multivariate calibration models with third-order, four-way, analytical data, and implements several third-order algorithms. These models include those based on QLD: PARAFAC, APQLD and AWRCQLD, those based on RTL: U-PLS/RTL, N-PLS/RTL, U-PCA/RTL and TLLS/RTL, and non-QLD: MCR-ALS, Augmented PARAFAC and PARAFAC2. The graphical interface is an extension of a previous one (MVC3) (Chem. Lab. Intell. Lab. System, 116 (2012) 9-16), including several innovations: a) Two software alternatives: MATLAB environment or standalone program, b) New plotting options: landscapes and contour plots, c) A higher number of prediction algorithms, as those based on MCR-ALS, including the recently developed area correlation constraints, Augmented PARAFAC and PARAFAC2, and e) New figures of merit according to recent developments. MVC3_GUI accepts a variety of ASCII data for input, depending on whether third-order data are vectorized or matricized. The new toolbox has been already used in our research group in several analytical applications since 2015, and I can confirm that the software is working fine in different analytical situations with data of different complexity. There are a few issues which I would like to see addressed in the future: the authors should clarify the explanations given in the user manual about the different available data types: X_vectors, X,Y_vectors, X,Y,Z_vectors and X,Y_matrices, and the authors should add more examples and data sets into the manual and toolbox, for a better understanding of the readers.

Acknowledgements

S. J. Mazivila thanks the Camões Instituto de Cooperação e da Língua, for the scholarship granted to his Ph.D. A. C. Olivieri thanks Universidad Nacional de Rosario (Project 19/B487), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 0163) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2016-1122) for financial support. We also thank one of the reviewers for thoroughly testing the compiled software and finding a number of bugs.

References

[1] A.C. Olivieri, H.L. Wu, R.Q. Yu, MVC3: a MATLAB graphical interface toolbox for third-order multivariate calibration, Chemometr. Intell. Lab. Syst. 116 (2012) 9–16.
[2] A.C. Olivieri, G.M. Escandar, Practical Three-way Calibration, Elsevier, Waltham, USA, 2014.
[3] G.M. Escandar, A.C. Olivieri, A road map for multi-way calibration models, Analyst 142 (2017) 2862–2873.
[4] M. Montemurro, G.G. Siano, M.R. Alcaráz, H.C. Goicoechea, Third order chromatographic-excitation-emission fluorescence data: advances, challenges and prospects in analytical applications, Trends Anal. Chem. 93 (2017) 119–133.
[5] A.R. Jalalvand, H.C. Goicoechea, D.N. Rutledge, Applications and challenges of multi-way calibration in electrochemical analysis, Trends Anal. Chem. 87 (2017) 32–48.
[6] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, Second- and higher-order data generation and calibration: a tutorial, Anal. Chim. Acta 806 (2014) 8–26.
[7] J.A. Arancibia, P.C. Damiani, G.M. Escandar, G.A. Ibañez, A.C. Olivieri, A review on second- and third-order multivariate calibration applied to chromatographic data, J. Chromatogr. B 910 (2012) 22–30.
[8] H. Parastar, J.R. Radović, J.M. Bayona, R. Tauler, Solving chromatographic challenges in comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry using multivariate curve resolution–alternating least squares, Anal. Bioanal. Chem. 405 (2013) 6235–6249.
[9] H. Parastar, J.R. Radović, 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, Anal. Chem. 83 (2011) 9289–9297.
[10] E.D. Larson, S.R. Groskreutz, D.C. Harmes, I.C. Gibbs-Hall, S.P. Trudo, R.C. Allen, S.C. Rutan, D.R. Stoll, Development of selective comprehensive two-dimensional liquid chromatography with parallel first-dimension sampling and second-dimension separation–application to the quantitative analysis of furanocoumarins in apiaceous vegetables, Anal. Bioanal. Chem. 405 (2013) 4639–4653.
[11] R.C. Allen, S.C. Rutan, Semi-automated alignment and quantification of peaks using parallel factor analysis for comprehensive two-dimensional liquid chromatography–diode array detector data sets, Anal. Chim. Acta 723 (2012) 7–17.
[12] V.A. Lozano, A.M. de la Peña, I. Durán-Merás, A.E. Mansilla, G.M. Escandar, Four-way multivariate calibration using ultra-fast high-performance liquid chromatography with fluorescence excitation–emission detection. Application to the direct analysis of chlorophylls a and b and pheophytins a and b in olive oils, Chemometr. Intell. Lab. Syst. 125 (2013) 121–131.
[13] M.R. Alcaráz, G.G. Siano, M.J. Culzoni, A.M. de la Peña, H.C. Goicoechea, Modeling four and three-way fast high-performance liquid chromatography with fluorescence detection data for quantitation of fluoroquinolones in water samples, Anal. Chim. Acta 809 (2014) 37–46.
[14] S.A. Bortolato, V.A. Lozano, A.M. de la Peña, A.C. Olivieri, Novel augmented parallel factor model for four-way calibration of high-performance liquid chromatography–fluorescence excitation–emission data, Chemometr. Intell. Lab. Syst. 141 (2015) 1–11.
[15] M.D. Carabajal, J.A. Arancibia, G.M. Escandar, On-line generation of third-order liquid chromatography-excitation-emission fluorescence matrix data. Quantitation of heavy-polycyclic aromatic hydrocarbons, J. Chromatogr. A 1527 (2017) 61–69.
[16] M.R. Alcaráz, S.A. Bortolato, H.C. Goicoechea, A.C. Olivieri, A new modeling strategy for third-order fast high-performance liquid chromatographic data with fluorescence detection. Quantitation of fluoroquinolones in water samples, Anal. Bioanal. Chem. 407 (2015) 1999–2011.
[17] R. Bro, PARAFAC. Tutorial and applications, Chemometr. Intell. Lab. Syst. 38 (1997) 149–171.
[18] H.Y. Fua, H.L. Wua, Y.J. Yua, L.L. Yua, S.R. Zhanga, J.F. Niea, S.F. Lia, R.Q. Yu, A new third-order calibration method with application for analysis of four-way data arrays, J. Chemom 25 (2011) 408–429.
[19] J.A. Arancibia, A.C. Olivieri, D.B. Gil, A.E. Mansilla, I. Durán-Merás, A.M. de la Peña, Trilinear least-squares and unfolded-PLS coupled to residual trilinearization: new chemometric tools for the analysis of four-way instrumental data, Chemometr. Intell. Lab. Syst. 80 (2006) 77–86.
[20] P.C. Damiani, I. Durán-Merás, A. García-Reiriz, A. Jiménez-Girón, A.M. 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, Anal. Chem. 79 (2007) 6949–6958.
[21] A.C. Olivieri, A combined artificial neural network/residual bilinearization approach for obtaining the second-order advantage from three-way non-linear data, J. Chemom 19 (2005) 615–624.
[22] R. Tauler, M. Maeder, A. de Juan, Multiset data analysis: extended multivariate curve resolution, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Elsevier, Oxford, UK, 2009, pp. 473–505.
[23] H.A.L. Kiers, J.M.F. Ten Berge, R. Bro, PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemom 13 (1999) 275–294.
[24] A.C. Olivieri, K. Faber, New developments for the sensitivity estimation in four way calibration with the quadrilinear parallel factor model, Anal. Chem. 84 (2012) 186–193.
[25] M.C. Bauza, G.A. Ibanez, R. Tauler, A.C. Olivieri, Sensitivity equation for quantitative analysis with multivariate curve resolution-alternating least-squares: theoretical and experimental approach, Anal. Chem. 84 (2012) 8697–8706.
[26] F. Allegrini, A.C. Olivieri, Analytical figures of merit for partial least-squares coupled to residual multilinearization, Anal. Chem. 84 (2012) 10823–10830.
[27] A.C. Olivieri, Computing sensitivity and selectivity in parallel factor analysis and related multiway techniques: the need for further developments in net analyte signal theory, Anal. Chem. 77 (2005) 4936–4946.
[28] A.C. Olivieri, Analytical figures of merit: from univariate to multiway calibration, Chem. Rev. 114 (2014) 5358–5378.
[29] MATLAB, The Mathworks Inc, 2012. Natick, Massachusetts, USA.
[30] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom 17 (2003) 274–286.
[31] R. Tauler, Multivariate curve resolution applied to second order data, Chemometr. Intell. Lab. Syst. 30 (1995) 133–146.
[32] A.C.O. Neves, R. Tauler, K.M.G. de Lima, Area correlation constraint for the MCR-ALS quantification of cholesterol using EEM fluorescence data: a new approach, Anal. Chim. Acta 937 (2016) 21–28.
[33] H. Parastar, R. Tauler, Multivariate curve resolution of hyphenated and multidimensional chromatographic measurements: a new insight to address current chromatographic challenges, Anal. Chem. 86 (2014) 286–297.