Analytica Chimica Acta 859 (2015) 20–28

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Modeling nonbilinear total synchronous fluorescence data matrices with a novel adapted partial least squares method

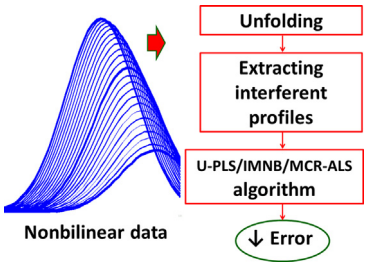
Agustina V. Schenone [a], Adriano de Araújo Gomes [b], María J. Culzoni [a], Andrés D. Campiglia [c], Mário Cesar Ugulino de Araújo [b], Héctor C. Goicoechea [a,*]

[a] 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, Santa Fe S3000ZAA, Argentina
[b] Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA) Universidade Federal da Paraíba, CCEN, Departamento de Química, Caixa Postal 5093, CEP 58051-970, João Pessoa, PB, Brazil
[c] Department of Chemistry, University of Central Florida, Orlando, FL 32816, USA

HIGHLIGHTS

- A new residual modeling for nonbilinear data is presented.
- The problem is addressed by extracting the interference profile with MCR-ALS.
- Total synchronous fluorescence spectroscopy are conveniently modeled.
- The novel algorithm perform better than known second-order algorithms.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 October 2014
Received in revised form 3 December 2014
Accepted 5 December 2014
Available online 10 December 2014

Keywords:
Second-order advantage
Synchronous fluorescence
Residual modeling
Ciprofloxacin

ABSTRACT

A new residual modeling algorithm for nonbilinear data is presented, namely unfolded partial least squares with interference modeling of non bilinear data by multivariate curve resolution by alternating least squares (U-PLS/IMNB/MCR-ALS). Nonbilinearity represents a challenging data structure problem to achieve analyte quantitation from second-order data in the presence of uncalibrated components. Total synchronous fluorescence spectroscopy (TSFS) generates matrices which constitute a typical example of this kind of data. Although the nonbilinear profile of the interferent can be achieved by modeling TSFS data with unfolded partial least squares with residual bilinearization (U-PLS/RBL), an extremely large number of RBL factors has to be considered. Simulated data show that the new model can conveniently handle the studied analytical problem with better performance than PARAFAC, U-PLS/RBL and MCR-ALS, the latter modeling the unfolded data. Besides, one example involving TSFS real matrices illustrates the ability of the new method to handle experimental data, which consists in the determination of ciprofloxacin in the presence of norfloxacin as interferent in water samples.

* Corresponding author. Tel.: +54 3424575206x190.
E-mail address: hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

http://dx.doi.org/10.1016/j.aca.2014.12.014
0003-2670/

1. Introduction

Nowadays, there is a widespread acceptance in the analytical community of the growing need for the study of increasingly complex samples by improving the available analytical methods. In the field of chemometrics, this fact is linked to the intensive research of new multi-way data generation and its consequent modeling with suitable algorithms [1–3]. Multi-way instrumental data carry the intrinsic potential to achieve the second-order advantage, which in principle permits analyte quantitation in samples containing unexpected constituents (i.e., sample concomitants not included in the calibration set). This property is especially interesting in the field of complex sample analysis owing to the fact that it allows building calibration models with a limited number of standards, maintaining quality of analyte prediction ability regardless of the presence of interferences, and without the need of previous chromatographic or electrophoretic separation.

A literature search on the available second-order algorithms capable of coping with the second-order advantage reveals an interesting fact: most of them call for the property of trilinearity. According to Olivieri et al. [1], the following requirements are necessary to fulfill the multilinearity property: (I) constancy of profiles across the different samples for each component; (II) signal linearly related to the analyte concentration; and, (III) bilinear signal for a given sample. The property of bilinearity implies that a single component data matrix can be decomposed into the product of two vectors, each containing the component profile in one of the two data dimensions (a well-known example of this kind of data is an excitation–emission fluorescence matrix, EEM). This can be condensed in the concept of rank of a matrix, i.e., the number of bilinear terms needed to reproduce the data matrix. If the rank is close to the number of responsive constituents, the matrix is bilinear. When the three-way array built with standards and test data matrices follow the trilinear structure, parallel factor analysis (PARAFAC) [4] seems to be the appropriate algorithm, because of its inner trilinear structure. A three-way array can then be built with a set of second-order **X** data matrices. Each data element in the array can be modeled by the following equation:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk} \quad (1)$$

where N is the total number of chemical constituents generating the measured signal, a_{in} is the relative concentration or score of component n in the i-th sample, and b_{jn} and c_{kn} are the intensities in the instrumental channels (or data dimensions) J and K, respectively. The values of e_{ijk} are the elements of the three-dimensional array **E**, representing the residual error, and having the same dimensions as **X**. Summation in Eq. (1) implies that the individual signals of the matrix constituents are additive. Usually the loadings are normalized to unit length, and collected into the loading matrices **B** and **C**, of size $J \times N$ and $K \times N$ respectively. The scores are collected into the score matrix **A** (size $I \times N$), which reflects the relative concentration values of the various constituents in all samples. This information is used in the context of calibration to build a univariate plot. This plot allows the prediction of analyte concentrations in unknown samples by projecting its score onto the fitted line.

Very recently, Olivieri and Escandar discussed the different kind of second-order data that can be gathered in the lab; their characteristics and the most convenient algorithm to be employed in each case [5]. Especial attention is paid on what they called “the Cinderella type 3 data”, nontrilinear data which involve nonbilinear data matrices, i.e., different spectral profiles for a component in a single sample. This is one of the most difficult problems to be solved. A reduced number of articles have been published dealing with nonbilinear data from single sample components [6–11] generated by (1) tandem mass-spectrometry (MS^2) [6], (2) total synchronous fluorescence spectroscopy matrix (TSFSM) [7–9], (3) ultrafast time-resolved spectroscopy [10], and (4) two-dimensional nuclear magnetic resonance spectroscopy [11]. TSFSM data sets are not bilinear due to the fact that synchronous spectral profiles vary with the offset between excitation and emission wavelengths [7–15]. MS^2 data sets are not bilinear because each fragment of a single compound has a specific MS pattern in the second MS dimension, making it impossible to describe the MS^2 data in terms of one MS profile in each dimension [6].

To the best of our knowledge, there are only two publications, which have been reported by our group, exploiting the second-order advantage on TSFSM data for quantitation purposes [8,9]. Essentially, synchronous spectroscopy consists in simultaneously scanning both the excitation and the emission monochromators while maintaining a constant wavelength interval ($\Delta\lambda$) between them. The judicious choice of Δl values along with the reduction of excitation and emission bandwidths often provides the spectral simplification needed to overcome severe overlapping in the analysis of multi-component samples.

Spectral simplification is always possible under proper instrumental parameters [16]. Unfortunately, in some cases synchronous fluorescence spectroscopy needs to be combined with a multivariate calibration method. This is particularly true for the direct analysis of complex samples of unknown composition having multiple fluorescence species [14]. In addition to resolving spectral overlapping, the chosen algorithm should be able to handle nonbilinear signals. Although U-PLS/RBL has been proposed for this task [5], the RBL procedure only works properly when data from unexpected components are bilinear. This is due to the fact that RBL models the signals of unexpected components via singular value decomposition (SVD) [17]. Therefore, the number of components needed to model the data may be extremely large, precluding the possibility of modeling the full variability of the data by only using a few of the first components. Thus, RBL will probably fail to model the matrix, not because of inability to model a nonbilinear matrix, but because of the fact that the large series of bilinear components that would be required to do so is truncated in some arbitrary way, i.e., if fifteen RBL components are necessary to model the variability of one interferent, this number should be truncated because it is not practical, or even feasible, especially when more than one interferent should be considered.

In an earlier work [8], the second-order advantage was gained from TSFSM data modeling with both U-PLS/RBL and MCR-ALS, but in the latter case the second-order data were previously unfolded assuming that the unfolded-TSFSM data set have a bilinear structure [12], i.e., the strategy consisted in modeling first-order responses with MCR-ALS, and exploiting the potentiality of the correlation constraint [18].

Herein, we present a novel method that comprises nonbilinear partial least squares followed by the modeling based on MCR-ALS with the correlation constraint. During the interferent modeling step, profiles corresponding to interfering species are extracted by MCR-ALS processing of unfolded data and incorporated into a step which updates the PLS scores in order to minimize the residual of the model. The new algorithm has been named unfolded partial least squares with interference modeling of non bilinear data by multivariate curve resolution by alternating least squares (U-PLS/IMNB/MCR-ALS). Both simulated and experimental data sets are used to compare the prediction ability of the new method to conventional U-PLS/RBL, PARAFAC and the previously mentioned unfolded MCR-ALS.