Available online at www.sciencedirect.com
ScienceDirect
Analytica Chimica Acta 584 (2007) 397–402
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

A novel strategy for solving matrix effect in three-way data using parallel profiles with linear dependencies

Morteza Bahram [a,*], Rasmus Bro [b]

[a] *Department of Chemistry, Faculty of Sciences, Urmia University, Urmia, Iran*
[b] *Chemometrics Group, Department of Food Science Royal Veterinary & Agricultural University Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark*

Received 9 May 2006; received in revised form 12 October 2006; accepted 28 November 2006
Available online 3 December 2006

Abstract

This work presents a novel strategy for solving matrix effects using the second-order advantage and a new method called PARAllel profiles with LINear Dependencies (PARALIND). PARALIND is a generalization of parallel factor analysis (PARAFAC) and was developed to extend its use to problems with linearly dependent factors where normal PARAFAC analysis will fail to provide meaningful results. Such linearly dependent factors occur in standard addition with second-order data such as fluorescence excitation emission matrices (EEM). By successive standard addition of an analyte, the concentrations of the remaining components (interferences) remain constant and introduce linear dependency between interference concentrations in the samples. This theoretically leads to rank deficiency in the score matrix holding the relative concentrations when using PARAFAC for modeling. In practice, PARAFAC models of such data will mostly provide solutions where the score matrix is not rank deficient but a function of the noise in the data. This problem is shown to be solved by using PARALIND. In order to evaluate the applicability of the method a simulated as well as an experimental data set is tested. The results from experimental data relate to the direct determination of salicylic acid (SA), the main product of aspirin degradation, in undiluted human plasma by spectrofluorimetry.

Keywords: Parallel profiles with linear dependencies; Standard addition; Three-way data; Fluorescence excitation emission matrices

1. Introduction

When a multivariate calibration model is used it is usually required that there are no new constituent(s) in the samples being analyzed. If there are new constituents, a recalibration including this new constituent will be necessary in order to be able to predict accurately, but this will be possible only if the interference(s) can be identified. In case of multi-way data, it is possible to handle unknown interferences as part of the calibration. Several methods for doing so have been developed; most notably generalized rank annihilation methods [1] and parallel factor analysis (PARAFAC).

Chemical analysis can be further complicated by matrix effects [2]. When the sensitivity of the response depends on the matrix composition, quantitative predictions based on pure standards may be affected by differences in the sensitivity of the response of the analyte in the presence and in the absence of chemical matrix of the sample. The standard addition method can be used to compensate for such matrix effects. Standard addition can compensate for non-spectral interferences and certain types of spectral interferences (e.g. non-analyte absorption) which enhance or depress the analytical signal of the analyte concentration [2].

As stated above, certain second-order calibration methods are able to resolve and recover the pure analyte response even in the presence of new interferences. In these cases pure analyte standards are commonly used for quantifying unknown samples even though matrix effects may degrade the quality of the resulting predictions.

Recently several methods were presented based on combining the second-order advantage and standard addition. Saurina and Tauler [3] proposed a new strategy using multivariate curve-resolution based on alternating least squares (MCR-ALS). The potential of the proposed strategy for solving matrix effects has been studied using an example of determination of triphenyltin in sea water samples for excitation emission matrices (EEM)

* Corresponding author. Tel.: +98 441 2780952; fax: +98 441 2776707.
E-mail address: m.bahram@urmia.ac.ir (M. Bahram).

0003-2670/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2006.11.070

fluorescence data. Also several results using PARAFAC has been published in the literature [4–7].The main problem using a curve-resolution method such as PARAFAC is that the model will not reflect what is known about the data. For example, it is a fact that the concentrations of the unknown interferences will be constant in all the samples that are varying only by different amounts of added analyte. Each chemical species will ideally require one PARAFAC component but the scores associated with the interferences would then ideally be identical in shape which is not consistent with the requirement for obtaining meaningful uniqueness in PARAFAC [7]. Due to the properties of the PARAFAC algorithm, however, each estimated component will typically have different estimated scores even though they should theoretically be identical. Another related problem is that the spectral loadings will be mathematically unique due to noise in the data even though they are in fact unidentified. Sidiropoulos and Bro [7] proposed a new method called 'parallel profiles with linear dependencies' (PARALIND) and in this paper PARALIND is proposed as a general tool for handling second-order standard addition.

2. Theoretical background

Assume that a three-way data array $\mathbf{X}$ ($I \times J \times K$) is given for which an S-component PARAFAC model holds. Hence,

$$\mathbf{X}_k = \bar{\mathbf{A}}\mathbf{D}_k\mathbf{B}^{\mathrm{T}} + \mathbf{E}_k, \quad k = 1 \ldots K. \tag{1}$$

The first mode loading matrix is denoted by $\bar{\mathbf{A}}$ ($I \times S$). Assume that the data come from a standard addition experiment with I measurements. For example, the first of the I samples could be a sample to be analyzed and the remaining $I-1$ samples are the same actual sample but with different amounts of analyte added. Then the score vector corresponding to the analyte should be increasing corresponding to the amount of added analyte whereas the scores for all interferences will *ideally* be constant as their concentrations do not vary.

Even though the score matrix is theoretically rank-deficient, this constraint is not actively enforced if PARAFAC is used to model the data. Noise may therefore lead to actual PARAFAC models, which are not rank-deficient as they should be. The factor matrices that should physically be rank-deficient will obtain full rank by fitting the noise part of the data [7].

A *dependency* matrix, $\mathbf{H}$, is used in PARALIND to explicitly incorporate the rank-deficiency into the model. The rank-deficient $\bar{\mathbf{A}}$ may be written

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{H} \tag{2}$$

where $\mathbf{A}$ is an $I \times R$ matrix and $\mathbf{H}$ is $R \times S$ matrix. If there are, e.g. four different components in the above example then $S = 4$. Assuming that the first component corresponds to the analyte, then the three last columns in $\bar{\mathbf{A}}$ must be identical. This can be achieved by defning $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2]$ and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}. \tag{3}$$

It directly follows that

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{H} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_2 \ \mathbf{a}_2] = [\overline{\mathbf{a}_1} \ \overline{\mathbf{a}_2} \ \overline{\mathbf{a}_3} \ \overline{\mathbf{a}_4}] \tag{4}$$

and thus the second, third and fourth column of $\bar{\mathbf{A}}$ are forced to be identical. This example of a dependency or interaction matrix is typical for the use of PARALIND for a standard addition problem.

3. Data and models

3.1. Simulated data

In order to evaluate the applicability of PARALIND for resolving standard addition data, eight different EEM fluorescence samples were simulated. Each sample contained four chemical species of which one was considered the analyte of interest. For every sample, five successive additions of the analyte were done and a 6 (addition mode) × 91 (emission) × 21 (excitation) array for each sample obtained. Each sample was analyzed with PARALIND and PARAFAC. Non-negativity constraints were enforced on the parameters of the model. Furthermore for PARALIND, the interaction matrix ($\mathbf{H}$) was chosen as

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}. \tag{5}$$

3.2. Salicylate determination in plasma using standard addition method

A 300 μg mL^{-1} stock solution of sodium salicylate (Merck) was prepared from deionized water (Milli-Q) and from these two dilute solutions of concentrations 120 and 75 μg mL^{-1} were also prepared. Human plasma samples (fresh frozen plasma) from 10 healthy volunteers were obtained at the Hemocentro of State University of Campinas, Brazil and were kept in the freezer at −8 °C. It was assumed that the salicylate concentration of all the plasma samples is zero [4].

A number of undiluted plasma samples were prepared in the range from 1.5 to 30.0 μg mL^{-1} by spiking appropriate amounts of 75 or 300 μg mL^{-1} salicylate solutions and plasma from different individuals was added. Another set of undiluted samples were prepared in the range from 1.5 to 24.0 μg mL^{-1}. Plasma from different individuals was used for each concentration level. For each measurement, 2.5 mL of sample was added in the cuvette and four successive additions of 50 μL of a 120 μg mL^{-1} salicylate solution were performed and measured. All of the determinations were carried out in duplicate or triplicate.

The excitation-emission data were measured with excitation from 280 to 340 nm (5 nm steps) and emission from 360 to 580 nm (0.5 nm steps). Slit widths were 4 nm and scanning rate 1200 nm min^{-1} [4]. For each sample a 5 × 13 × 442 three-way array was obtained. For each three-way array three to four components was indicated by using singular value decomposition for each slab of excitation × emission matrix. For example, a