Journal of Advanced Research (2016) xxx, xxx–xxx

Cairo University

Journal of Advanced Research

ORIGINAL ARTICLE

Independent components analysis as a means to have initial estimates for multivariate curve resolution-alternating least squares

Leonardo Valderrama [a], Rayanna P. Gonçalves [a], Paulo H. Março [a], Douglas N. Rutledge [b], Patrícia Valderrama [a,*]

[a] Universidade Tecnológica Federal do Paraná (UTFPR), C.P. 271, 87301-899 Campo Mourão, Paraná, Brazil
[b] AgroParisTech/INRA, UMR1145 Ingénierie Procédés Aliments, 75005 Paris, France

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 5 December 2015
Accepted 6 December 2015
Available online xxxx

Keywords:
UV–Vis
Signal overlapping
Olive oil
PURE
Curve resolution
Independent Components Analysis

ABSTRACT

Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) is a curve resolution method based on a bilinear model which assumes that the observed spectra are a linear combination of the spectra of the pure components in the system. The algorithm steps include the determination of the number of components by rank analysis methods, initial estimates for the concentrations and/or spectra and an iterative optimization. Sometimes, suitable results may not be achieved when MCR-ALS is applied. One reason for this is the importance of the initial estimates of the spectral profiles. In that case, the MCR-ALS algorithm may reach a local minimum instead of a global minimum and this can result in ineffective curve resolution. The most popular algorithm used to find the initial estimates (PURE derived from SIMPLISMA) suffers from an essential drawback, which is the necessity to have "pure" variables related to a single spectral component, which cannot be expected in all cases because of the strong signal overlapping as in the Ultraviolet–Visible (UV–Vis) spectroscopy. This work summarizes this problem, presenting a case study based on UV–Vis spectroscopy of heated olive oil. To solve the problems of the need for "pure" variables and to avoid local minima with MCR-ALS, Independent Components Analysis (ICA) was used to calculate initial estimates for MCR-ALS. The results from this study suggest that this use of ICA prior to MCR-ALS improves the resolution for UV–Vis data and provides acceptable resolution results when compared to the most used method, PURE.

* Corresponding author. Tel.: +55 44 3518 1400.
E-mail addresses: patriciav@utfpr.edu.br, pativalderrama@gmail.com (P. Valderrama).

Peer review under responsibility of Cairo University.

Production and hosting by Elsevier

Introduction

Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) is a curve resolution method based on a bilinear model which assumes that the observed spectra are a linear combination of the spectra of the pure components in the

http://dx.doi.org/10.1016/j.jare.2015.12.001
2090-1232 © 2015 Production and hosting by Elsevier B.V. on behalf of Cairo University.

system [1]. The algorithm steps include the determination of the number of components by rank analysis methods, such as percentage of explained variance from Principal Component Analysis (PCA) [2], PCA-Loadings or Durbin-Watson (DW) criterion [3] and Singular Value Decomposition (SVD) [4]. PCA, PCA-Loadings and SVD are similar methods, while DW criterion to rank analysis has been proposed as a measure of the signal/noise ratio of the PCA loadings and regression vectors obtained by multivariate analysis of signals [3]. Then, an initial estimation for concentration and/or spectra with as many profiles as the number of components estimated from the rank analysis is constructed to start the iterative curve resolution process. Once the initial estimate is generated, the iterative optimization step can be started [5].

The initial estimate is found, normally, based on methods of finding the purest variables, as the PURE method that is derived from the simple-to-use interactive self modeling analysis (SIMPLISMA) [6], or on evolving factor analysis (EFA) [7]. However, these algorithms suffer from an essential drawback, which consists in the need for "pure" variables, which cannot be expected in all cases because of the possibility of strong signal overlapping [8]. This work offers an alternative to obtain initial estimates based on Independent Components Analysis (ICA) [9]. Here it is shown that the ICA improved the MCR-ALS resolution results when "pure" variables are not present due to the strong signal overlapping, as in the case of analyzing heated olive oil using Ultraviolet–Visible (UV–Vis) spectroscopy. The main objective is to verify the modifications that occur in olive oil samples when it is heated from room to high temperatures, such as happens during frying, without the need for physical separation, only by using curve resolution methods. This data set was used in order to show the problem of methods based on SIMPLISMA as initial estimates for MCR-ALS when "pure" variables are not present due to the strong signal overlapping, as occur at UV–Vis spectroscopy due the lack of selectivity in this technique.

Experimental

Samples of Portuguese olive oil (two samples from two different batches) were analyzed in triplicate. The samples were heated from 30 °C until 170 °C, increasing it by steps of 10 °C, and a first spectrum being taken at room temperature (25 °C). UV–Vis spectra were acquired in the range from 300 to 540 nm (steps of 2 nm) in a 1 mm quartz cuvette. Data were analyzed using MATLAB version R2007b (The Mathworks Inc., MA, USA) where curve resolution was performed by Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS). The MCR-ALS algorithm code and Graphical User Interface for MATLAB [10] are freely available from the home page of MCR at http://www.mcrals.info/. By this interface, there are two options to estimate the matrix rank: one is based on the percentage of variance captured by singular values decomposition (SVD) analysis and the other is the manual decision of it. Sometimes it is hard to decide the rank based only on these percentages and the decision can be improved by the graphical visualization of the PCA loadings [3]. ICA was performed using Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [11], that involves matrix diagonalization. The MATLAB code of this algorithm can be found on the website <http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m>.

Chemometric methods

MCR-ALS

The usual assumption in multivariate curve resolution methods is that the experimental data follow a linear model similar to Lambert–Beer's law in absorption spectroscopy. In matrix form this model can be written as [12]:

$$\mathbf{D}_{(i\times j)} = \mathbf{C}_{(i\times k)}\mathbf{S}^{T}_{(k\times j)} + \mathbf{E}_{(i\times j)} \quad (1)$$

where $\mathbf{D}_{(i\times j)}$ is the UV–Vis data matrix with "i" rows and "j" columns (where spectra are on the rows of **D** and absorbance at different wavelength is on the columns of **D**, $\mathbf{C}_{(i\times k)}$ is the matrix of the relative amounts or concentrations with "i" rows and "k" different species (in this case the relative concentration profile for each sample is placed in the rows of **C**, while in the **C** columns are the information concerning the different species), $\mathbf{S}^{T}_{(k\times j)}$ contains the pure spectra with "k" different species and "j" columns (recovered spectra are located in the columns of **S** and the information about different species is in the rows of **S**), and $\mathbf{E}_{(i\times j)}$ is the matrix associated with noise or experimental error with "i" rows and "j" columns.

To start the algorithm, the number of chemical species present in a particular system is determined based on the chemical rank associated with the data matrix **D**. Here, the chemical rank was determined using PCA-Loadings [3], and confirmed by the Leverage analysis [13].

The main goal of curve resolution methods is the determination of the true **C** and **S** matrices only from the analysis of matrix **D**. Initial estimates of the **C** or **S** matrices can be obtained using methods based on the detection of "purest" variables as methods based on SIMPLISMA [6], or from techniques based on evolving factor analysis [7].

Methods based on the choice of "purest" variables aim to the determination of the most representative series of different components in the experimental data set. If the choice is successful, the majority of similar algorithms offers a chance for qualitatively estimating the number of components in the system and also the relative concentrations and spectra of individual compounds. In the group of methods based on the choice of "purest" variable, the PURE algorithm is the most commonly used. However, algorithms of this group suffer from an essential drawback, which consists in the necessity of the presence of "pure" variables, which cannot be expected in all cases because of the strong signal overlapping [8]. To solve this problem, the use of ICA scores or signals is proposed as initial estimative to MCR-ALS. In this work, the scores and signals were tested as initial estimative for concentration (**C**) and spectra (**S**), respectively, and the results are identical.

These initial estimations of **C** or **S** are optimized solving Eq. (1) iteratively by alternating least squares optimization [12]. At each iteration of the optimization, a new estimation of the **C** and **S** matrices is calculated under the constraints of two least-squares steps [14]:

$$\mathbf{C} = \mathbf{DS}(\mathbf{S}^{T}\mathbf{S})^{-1} \quad (2)$$

$$\mathbf{S}^{T} = (\mathbf{C}^{T}\mathbf{C})^{-1}\mathbf{C}^{T}\mathbf{D} \quad (3)$$

Please cite this article in press as: Valderrama L et al., Independent components analysis as a means to have initial estimates for multivariate curve resolution-alternating least squares, J Adv Res (2016), http://dx.doi.org/10.1016/j.jare.2015.12.001