Analytica Chimica Acta 816 (2014) 18–27

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements

Hadi Parastar*, Nadia Akvan

Department of Chemistry, University of Isfahan, Isfahan 81746-73441, Iran

HIGHLIGHTS

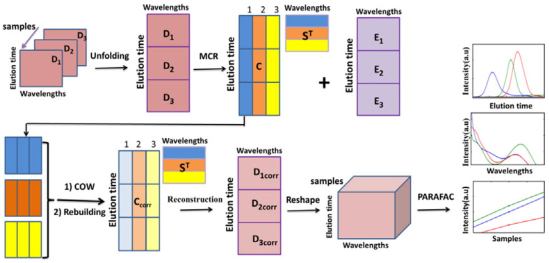
- MCR followed by COW is proposed for chromatographic peak alignment.
- The aligned data is decomposed using PARAFAC to exploit second-order advantage.
- The performance of MCR-COW-PARAFAC is tested using simulated and real HPLC-DAD data.
- The proposed method is compared with PARAFAC, COW-PARAFAC, MCR-ALS and MCR-COW-MCR.

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form 3 January 2014
Accepted 27 January 2014
Available online 6 February 2014

Keywords:
Multivariate curve resolution
Correlation optimized warping
Parallel factor analysis
Chromatography
Second-order advantage

GRAPHICAL ABSTRACT



ABSTRACT

In the present contribution, a new combination of multivariate curve resolution-correlation optimized warping (MCR-COW) with trilinear parallel factor analysis (PARAFAC) is developed to exploit second-order advantage in complex chromatographic measurements. In MCR-COW, the complexity of the chromatographic data is reduced by arranging the data in a column-wise augmented matrix, analyzing using MCR bilinear model and aligning the resolved elution profiles using COW in a component-wise manner. The aligned chromatographic data is then decomposed using trilinear model of PARAFAC in order to exploit pure chromatographic and spectroscopic information. The performance of this strategy is evaluated using simulated and real high-performance liquid chromatography-diode array detection (HPLC-DAD) datasets. The obtained results showed that the MCR-COW can efficiently correct elution time shifts of target compounds that are completely overlapped by coeluted interferences in complex chromatographic data. In addition, the PARAFAC analysis of aligned chromatographic data has the advantage of unique decomposition of overlapped chromatographic peaks to identify and quantify the target compounds in the presence of interferences. Finally, to confirm the reliability of the proposed strategy, the performance of the MCR-COW-PARAFAC is compared with the frequently used methods of PARAFAC, COW-PARAFAC, multivariate curve resolution-alternating least squares (MCR-ALS), and MCR-COW-MCR. In general, in most of the cases the MCR-COW-PARAFAC showed an improvement in terms of lack of fit (LOF), relative error (RE) and spectral correlation coefficients in comparison to the PARAFAC, COW-PARAFAC, MCR-ALS and MCR-COW-MCR results.

1. Introduction

Getting useful chemical information from complex mixtures with a large number of chemical components has long been a challenging task to chemists [1–3]. Chromatographic systems coupled to sophisticated detection systems, such as high-performance

---

* Corresponding author. Tel.: +98 311 7932737; fax: +98 311 6689732.
E-mail addresses: h.parastar@sci.ui.ac.ir, h.parastar@gmail.com (H. Parastar).

0003-2670/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2014.01.051

liquid chromatography-diode array detection (HPLC-DAD) and gas chromatography–mass spectrometry (GC–MS) have been used for many years in the analysis of complex mixtures which generate a large amount of data [4–7]. In spite of the recent technological advances, the complexity of many naturally occurring mixtures still exceeds the capacity of any single method, even when optimized to resolve them. Therefore, chromatographic analyses suffer from fundamental problems such as baseline/background contributions, low signal-to-noise ratio (S/N), elution time shifts and peak overlap which can directly affect final qualitative and quantitative chromatographic results [1,2,8].

Multivariate resolution methods (i.e., multi-set and multi-way methods) have been proposed in recent decades to overcome fundamental chromatographic challenges occurred during analysis of complex mixtures [9–13]. Multivariate resolution methods exploit “second-order advantage” which implies the calibration and determination of analytes in the presence of any uncalibrated and unknown interferences in the mixture. Among different multi-way methods, parallel factor analysis (PARAFAC) [14] has attracted great attention in chemistry in recent years due to its unique properties in the solutions and wide variety of applications [15–17]. PARAFAC calibration models have been used to model second-, third- and fourth-order data obtained by spectrofluorimetry [18,19], UV–Vis spectroscopy [20], kinetic spectroscopy [21], hyphenated chromatography [2] and multi-dimensional chromatography [22,23]. However, PARAFAC model relies on trilinear model assumption, a prerequisite seldom met due to variations in the chromatographic conditions affecting both peak position and peak shape. In general, there are two approaches to handle problems associated with changes in peak shapes and peak positions. The first approach is based on modeling elution time shifts without need for prior correction using chemometric methods, such as PARAFAC2 [24] and multivariate curve resolution-alternating least squares (MCR-ALS) [25]. PARAFAC2 is another version of PARAFAC model in which strict trilinearity fulfillment needed in PARAFAC is not required and it allows for the data analysis where there are elution time shifts [17,26]. However, PARAFAC2 is computationally more complex and expensive, and it does not allow for the application of constraints like non-negativity or unimodality in one of the data mode and therefore, unreasonable negative values and multimodal peaks appear in the results [27]. In addition, application of bilinear based models such as MCR-ALS has been proposed to handle chromatographic challenges during the analysis of complex samples [8,13,27–30]. However, bilinear based models suffer from the presence of rotational ambiguities in their solutions that greatly affect their qualitative and quantitative results [31]. Additionally, this method is very sensitive to the initial estimates of elution and spectral profiles.

Another approach to tackle elution time shift problem is to preprocess the data by some kind of alignment algorithms. Several applications of alignment algorithms, such as dynamic time warping (DTW) [32] and correlation optimized warping (COW) [33] have been proposed by different research groups to preserve trilinear model assumption and to obtain reliable qualitative and quantitative chromatographic information. However, alignment algorithms such as COW cannot correctly work in the case of complex chromatographic data with heavy overlap between components and coeluted interferences. Very recently, multivariate curve resolution based alignment methods have been proposed for chromatographic peak alignment in complex mixtures to take advantage of the second-order structure of data [34–36]. The most important feature of these methods is the efficient peak alignment of target compounds regardless of the presence of the overlapped interferences. As a consequence, combination of these effective alignment algorithms with trilinear decomposition methodologies can be used in order to exploit the well-known second-order advantage. However, in many practical applications the presence of seriously overlapping interferences, especially when the target compounds are seriously overlapped with unknown compounds, still poses great challenges to the aforementioned alignment algorithms [35].

In the present contribution, a new combination of MCR-COW with trilinear PARAFAC model was proposed for efficient correction of elution time shifts among chromatographic measurements and for obtaining pure qualitative and quantitative chromatographic information. MCR-COW is based on the bilinear decomposition of column-wise augmented chromatographic data and subsequent alignment of resolved elution profiles using COW. The aligned chromatographic data was arranged in a three-way data array and was analyzed by PARAFAC to obtain pure elution and spectral profiles as well as calibration curves for target compounds. For this purpose, simulated and real HPLC-DAD data sets with different number of chemical components, given level of noise (i.e., homoscedastic and heteroscedastic), elution time shifts and various concentration levels were evaluated. The PARAFAC results showed a significant improvement after MCR-COW in terms of lack of fit (LOF), relative error (RE) and spectral correlation coefficients compared to PARAFAC, COW-PARAFAC, MCR-ALS and MCR-COW-MCR results.

2. Methodology

2.1. Correlation optimized warping

The COW algorithm was introduced by Nielsen et al. [33] as a method to correct elution time shifts in chromatographic signals. It is a piecewise or segmented data preprocessing technique that uses Dynamic Programming (DP) to align a sample chromatogram toward a reference chromatogram. In the present contribution, a slightly modified version of the COW algorithm developed by Skov et al. [37] is used, which uses the summed correlation coefficient as optimization criterion for determining the optimal path for alignment (largest value of the summed correlation coefficients).

In COW, two parameters require to be set, namely the segment length, s and a slack size, l. Given the segment length s, the reference (y) and sample (x) signals are subdivided into local regions which are iteratively stretched and compressed for maximally l data points via interpolation so as to maximize the correlation between the sample and reference signals. This correlation is calculated for all possible combinations of warped segments as allowed by the warping parameters of segment length and slack size according to Eq. (1):

$$\rho(\mathbf{n}) = \frac{cov[y(\mathbf{n}), x'(\mathbf{n})]}{\sqrt{var[y(\mathbf{n})]var[x'(\mathbf{n})]}} \tag{1}$$

where $\mathbf{n}$ is a vector containing the time points selected from the original reference signal y and the warped signal x'. Also, the *cov* and *var* functions stand for covariance and variance of the desired vectors, respectively.

The selection of segment length and slack size can be performed by an optimization scheme, such as simplex non-linear optimization algorithm [37]. Once correlations are obtained for all possible segments under optimum segment length and slack size, the warping function can be determined for a global optimum where the theory of DP is used. The optimal path of optimal border positions of the segments is then applied and the points in between the boundaries are interpolated resulting in the best matching preprocessed sample signal for the predetermined slack and segment sizes.