Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 205 (2018) 381–390

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Comparison of several third-order correction algorithms applied to fluorescence excitation–emission-sample data array: Interference-free determination of polycyclic aromatic hydrocarbons in water pollution

Zhe Yang [a], Tingting Liu [a,b,*], Yutian Wang [a], Yuanyuan Yuan [a], Fengkai Shang [a]

[a] Measurement Technology and Instrument Key Lab of Hebei Province, Yanshan University, Qinhuangdao 066004, China
[b] Vocational and Technical College of Liuzhou, Liuzhou 545000, China

ARTICLE INFO

Article history:
Received 3 April 2018
Received in revised form 12 July 2018
Accepted 15 July 2018
Available online 18 July 2018

Keywords:
Four-way spectral data
Polycyclic aromatic hydrocarbons
Three-dimension fluorescence
Spectra
Alternating weighted residue constraint quadrilinear decomposition

ABSTRACT

Interference-free determination of polycyclic aromatic hydrocarbons (PAHs) in water pollution is proposed based on third-order correction algorithms with quadrilinear component modeling applied to the constructed four way fluorescence excitation–emission-sample data array with higher accuracy and better predictive ability than second-order (three-dimension) correction. Alternating weighted residue constraint quadrilinear decomposition (AWRCQLD), quadrilinear parallel factor analysis (4-PARAFAC), alternate penalty quadrilinear decomposition (APQLD) and alternate penalty trilinear decomposition (APTLD) are applied to acenaphthene (ANA), naphthalene (NAP) and fluorene (FLU) respectively. Fulvic acid affects PAHs determination seriously in real-world situation, so it is simulated as an interfering agent. Excitation-emission fluorescence matrixes (EEMs) of PAHs are measured at different volumes of fulvic acid simulated different interference conditions, to construct a four-way data array. After the four-way spectra data is analyzed by AWRCQLD, 4-PARAFAC, and APQLD, three-way EEMs analyzed by APTLD, results show that, on the one hand, PAHs can be measured more accurately with four-way data combined with third-order calibration than lower-order. On the other hand, AWRCQLD algorithm can reflect the superiority of third-order advantage better with higher recovery rate and smaller root mean square error, than other third-order or second-order correction algorithms.

1. Introduction

The basic point of classical analytical chemistry is based on a single point of data (scalar), such as a physical or chemical signal that has a corresponding quantitative relationship (usually linear relationship) with a certain analyte in the system, which requires full selectivity for the analyte of interest for qualitative and quantitative analysis of the chemical. First-order data from which called multivariate calibration, processed by first-order calibration algorithm, typically principal component regression (PCR) and partial least-squares (PLS) analyses with central idea to compensate for the lack of selectivity, requires construction of a sufficiently representative calibration set, that includes all the variability expected in unknown samples [1]. With the developments of modern instrument and hyphenated techniques, second-order date and corresponding second-order calibration analysis is gaining widespread acceptance by the analytical community with flourishing literatures about algorithms of second-order calibration such as the generalized rank annihilation method (GRAM) [2], direct trilinear decomposition (DTLD) [3], parallel factor analysis (PARAFAC) [4] and multivariate curve resolution-alternating least squares (MCR-ALS) [5], owning to the available mathematics, statistics and computer science. Because of the natural basis of the unique decomposition of second-order data with trilinearity, it may allow for quantitation of analytes even in the presence of unexpected constituents, a property known as the second-order advantage [6]. This avoids a major drawback of lower-order analytical methods, namely the requirement for a large, various calibration set, a practical benefit that can hardly be overstated. It solves problems that are considered impossible or even unimaginable in scalar analysis, opening up a new field of research.

In analytical chemistry, due to the increasing demand of real-world analytical situation, the application of higher order data was required. Since more information of samples is collected by introducing additional information as a fourth dimension, the accuracy and selectivity of analysis will be improved, for example, fast high-performance liquid chromatography with fluorescence excitation-emission detection [7]. Because of the complexity of actual data, various second-order correction algorithms with their own characteristics have different analytical problems [8–17]. The third-order correction [16,17] is more advantageous in sensitivity, selectivity and ability of effective information extraction, besides the uniqueness of the trilinear decomposition and the "second-order advantage", which are shared by APTLD [18], PARAFAC

---

* Corresponding author.
E-mail address: 1040879372@qq.com (T. Liu).

https://doi.org/10.1016/j.saa.2018.07.045
1386-1425/© 2018 Published by Elsevier B.V.

[19,20] and other algorithms [21,22]. Kim et al. [23] compared the results of the tri-linear parallel factor analysis and the four-linear parallel factor analysis, which concluded that the latter achieved higher selectivity than the former. Olivieri A.C. [24] believed, that third-order correction can achieve better computing sensitivity and selectivity in Parallel Factor Analysis and Related Multiway Techniques. Shufang Li et al. [25] used a four-way chemometrics algorithm to determine procaine hydrochloride in human plasma samples, and results showed that the third-order calibration algorithm was more advantageous. At present, several algorithms appear for four-way data calibration, such as 4-PARAFAC, APQLD, AWRCQLD, unfolded partial least-squares and multipoint least squares Linearization (U-PLS/RTL, N-PLS/RTL) [26]. A toolbox, named multivariate calibration 3 (MVC3), has recently been developed by A.C. Olivieri et al. [27] for implementing third-order multivariate calibration methodologies. AWRCQLD [28] not only has the "third-order advantage" with the more stable decomposition results, but also is less sensitive to the number of components than 4-PARAFAC [7,12,29] and has ability of more tolerant of noise than APQLD [25,28,30], which makes "mathematical separation" replace the separation of physics and chemistry ideally in the unknown interference or background.

Polycyclic aromatic hydrocarbons (PAHs) [31–41], a carbohydrate composed of two or more benzene rings in the thickening pattern, is a class of trace organic pollutants with an inert and carcinogenic nature. They are widely distributed in the air, sea and soil, mostly from the leakage or refining of petroleum products, incomplete combustion of organic matter, volcanic eruptions, forest fires, etc. With the development of modern industrialization, water pollution is becoming more and more serious and PAHs, although low in water, are priority objects of water pollution determination, because of their high toxicity, and have become a hot concern of environmental protection. In 1990, there were 7 kinds of PAHs in 68 kinds of water pollution controlled preferentially in China [42]. The European Union Environmental Protection Agency (EU EPA) regards 6 PAHs as the target pollutants and 16 kinds by the US EPA, some of which are considered human carcinogens. Therefore, it is of great significance to study a method for rapid determination of PAHs for real-world application.

At present, the determination methods for PAHs in water are mainly by chromatography and spectrophotometry. However, due to the possible corrosion and volatility of chromatography, the equipment are relatively expensive, and the solvent needs to be strictly purified. Therefore, in recent years, in the detection of organic pollutants in water, spectrophotometry, especially three-dimensional fluorescence spectrometry is becoming more and more popular. Three-dimensional fluorescence spectroscopy [43,44] with rich information, striking feature and good selectivity, can be used for multi-components mixture analysis.

PAHs determination in real-world water bodies are usually affected by higher concentration fulvic acid, so in present work simulate fulvic acid as a sample of interference. In this work, three kinds of third-order [45] correction algorithms are used to analyze the four-way data array, and the recoveries [46] of PAHs obtained under different volumes of interference were compared respectively; experimental results show that PAHs can be measured more accurately by the higher-order data obtained from the constructed four way excitation-emission fluorescence spectra array under different volumes of fulvic acid (0 μL, 100 μL, 200 μL, 300 μL) and the recovery rate is higher. PAHs are similar in chemical structure and cannot be directly measured by traditional fluorescence method, the combination of AWRCQLD algorithm and third-order data can obtain good results in qualitative and quantitative analysis of trace PAHs in water.

2. Quadrilinear Model

For a quadrilinear component model, the expression of each element x_{ijkl} of the four-way data array $\underline{X}_q$ can be deduced from the multi-linear component model, scalar representation as shown in Eq. (1).

$$
\begin{aligned}
&x_{ijkl} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} d_{ln} + e_{ijkl} \\
&i = 1, 2, \ldots, I, j = 1, 2 \ldots, J, k = 1, 2, \ldots, K, l = 1, 2, \ldots L
\end{aligned}
\tag{1}
$$

When there are N fluorophores in the samples and a tensor of experimental data, it will be compatible with the structure of Eq. (1), where, the data are quadrilinear; a, b, c are the elements of A, B, C matrixes respectively; e_{ijkl} is the element of the four-way residual array E_q; N is the total number of constituents of the quadrilinear model (the actual contribution to the fluorescence including the interference and background). This process of quadrilinear model was pictorially illustrated in Fig. 1.

2.1. AWRCQLD Algorithm

On the basis of the least squares principle [47], the weighted residual function is used as the constraint term, combined with the loss function of the quadrilinear model in the incomplete extended matrix to construct the final objective function. Four objective functions are shown in Eqs. (2)–(5).

$$
\begin{aligned}
\sigma_{(A)} = {} & \sum_{l=1}^{L} \left\| X_{\ldots l} - A diag\left(d_{(l)}\right) (\mathrm{C} \odot \mathrm{B})^T \right\|_F^2 + \\
& \lambda_A \sum_{l=1}^{L} \left\| \left(X_{\ldots l} \left((\mathrm{C} \odot \mathrm{B})^T \right)^+ - A diag\left(d_{(l)}\right) \right) W_D \right\|_F^2
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\sigma_{(B)} = {} & \sum_{i=1}^{I} \left\| X_{i\ldots} - B diag\left(a_{(i)}\right) (\mathrm{D} \odot \mathrm{C})^T \right\|_F^2 + \\
& \lambda_B \sum_{i=1}^{I} \left\| \left(X_{i\ldots} \left((\mathrm{D} \odot \mathrm{C})^T \right)^+ - B diag\left(a_{(i)}\right) \right) W_A \right\|_F^2
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\sigma_{(C)} = {} & \sum_{j=1}^{J} \left\| X_{.j..} - C diag\left(b_{(j)}\right) (\mathrm{A} \odot \mathrm{D})^T \right\|_F^2 + \\
& \lambda_C \sum_{j=1}^{J} \left\| \left(X_{.j..} \left((\mathrm{A} \odot \mathrm{D})^T \right)^+ - C diag\left(b_{(j)}\right) \right) W_B \right\|_F^2
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
\sigma_{(D)} = {} & \sum_{k=1}^{K} \left\| X_{..k.} - D diag\left(c_{(k)}\right) (\mathrm{B} \odot \mathrm{A})^T \right\|_F^2 + \\
& \lambda_D \sum_{k=1}^{K} \left\| \left(X_{..k.} \left((\mathrm{B} \odot \mathrm{A})^T \right)^+ - D diag\left(c_{(k)}\right) \right) W_C \right\|_F^2
\end{aligned}
\tag{5}
$$

where, $W_A = diag\ (sqrt(1.\ /diagm(A^TA)))$, $W_B = diag\ (sqrt(1.\ /diagm\ (B^TB)))$, $W_C = diag\ (sqrt(1.\ /diagm(C^TC)))$, $W_D = diag\ (sqrt(1.\ /diagm\ (D^TD)))$; λ_A, λ_B, λ_C, λ_D, respectively, are constraint coefficients, which are used to balance weighted residual function and loss function, two parts of the quadrilinear model in matrix of four incomplete extensions. According to the principle of alternating least squares [48], A, B, C and D can be obtained by alternately optimizing the four objective functions Eqs. (2)–(5), results shown in Eqs. (6)–(9).

$$
\begin{aligned}
A = {} & \left(\sum_{l=1}^{L} X_{\ldots l} (\mathrm{C} \odot \mathrm{B}) + \lambda_A \left((\mathrm{C} \odot \mathrm{B})^T \right)^+ W_D W_D \right) diag\ \left(d_{(l)}\right) \\
& \odot \left(\sum_{l=1}^{\mathrm{L}} diag\left(d_{(l)}\right) \left(\left(C^T C\right) * \left(B^T B\right) + \lambda_A W_D W_D \right) \mathrm{diag}\left(\mathrm{d}_{(l)}\right) \right)^+
\end{aligned}
\tag{6}
$$