Four-dimensional data coupled to alternating weighted residue constraint quadrilinear decomposition model applied to environmental analysis: Determination of polycyclic aromatic hydrocarbons

Tingting Liu [a], Ling Zhang [a,b], Shutao Wang [a], Yaoyao Cui [a,*], Yutian Wang [a], Lingfei Liu [a], Zhe Yang [a]

[a] Measurement Technology and Instrument Key Lab of Hebei Province, Yanshan University of China, Hebei Street West438#, Qinhuangdao, Hebei 066004, China
[b] Hebei University of Engineering, Handan 056038, China

ARTICLE INFO

ABSTRACT

Qualitative and quantitative analysis of polycyclic aromatic hydrocarbons (PAHs) was carried out by three-dimensional fluorescence spectroscopy combining with Alternating Weighted Residue Constraint Quadrilinear Decomposition (AWRCQLD). The experimental subjects were acenaphthene (ANA) and naphthalene (NAP). Firstly, in order to solve the redundant information of the three-dimensional fluorescence spectral data, the wavelet transform was used to compress data in preprocessing. Then, the four-dimensional data was constructed by using the excitation-emission fluorescence spectra of different concentration PAHs. The sample data was obtained from three solvents that are methanol, ethanol and Ultra-pure water. The four-dimensional spectral data was analyzed by AWRCQLD, then the recovery rate of PAHs was obtained from the three solvents and compared respectively. On one hand, the results showed that PAHs can be measured more accurately by the high-order data, and the recovery rate was higher. On the other hand, the results presented that AWRCQLD can better reflect the superiority of four-dimensional algorithm than the second-order calibration and other third-order calibration algorithms. The recovery rate of ANA was 96.5%~103.3% and the root mean square error of prediction was 0.04 μgL^{-1}. The recovery rate of NAP was 96.7%~115.7% and the root mean square error of prediction was 0.06 μgL^{-1}.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) [1–11], a class of trace organic pollutants with inert and carcinogenic nature, are carbohydrate composed of two or more benzene ring in the pattern of thickening ring. PAHs are numerous and find more than 200 kinds so far. With the development of modern industrialization, water pollution is becoming more and more serious. PAHs and other organic pollution detection technology have become a hot concern of environmental protection workers. PAHs are low in water, but they are still selected as priority detectors for water pollution because of their high risk. The European Union (EU EPA) has regarded 6 PAHs as the target pollutants. In 1990, there were 7 kinds of PAHs in 68 kinds of water pollution to control preferentially in China [12].

At present, the detection methods of PAHs in water are mainly by chromatography and spectrophotometry. However, due to the corrosion and volatility of chromatography, the equipment are relatively expensive and the solvent needs to be strictly purified. In recent years, in the detection of organic pollutants, spectrophotometry is becoming more and more popular, especially three- dimensional fluorescence spectrometry. Three-dimensional fluorescence spectroscopy [13, 14] with rich information, striking feature and good selectivity, can be used for multi-components mixture analysis.

In order to solve the redundant data in the three- dimensional fluorescence spectral data, the wavelet transform [15–17] (the db3 wavelet function) was used to compress [18] the experimental data in this work. Wavelet transform has the characteristics of strong local resolution and large compression ratio. In recent years, it has been successfully used to eliminate the redundant information in the infrared spectrum data without losing the spectral strength information.

With the development of a various disciplines and technologies, the application of higher-order data has been gotten wide attention in analytical chemistry. By introducing another additional information to the samples, the accuracy and selectivity of the analysis have been improved. When introducing additional information as a fourth dimension, for example, fast high-performance liquid chromatography with fluorescence excitation- emission detection for each sample can be

* Corresponding author at: Department of Testing and Measurement Technology and Instruments, School of Electrical Engineering, Yanshan University of China, Hebei Street West438#, Qinhuangdao, Hebei 066004, China.
E-mail address: tingtingliu@stumail.ysu.edu.cn (Y. Cui).

used. In fact, several emission wavelength-elution time matrices were recorded as a function of the excitation wavelength in; while four-dimensional data were acquired by excitation- emission matrices at different elution times in [19]. In many multidimensional data decomposition algorithms, the second-order calibration algorithm [20–27] can achieve the same result when dealing with ideal data, but due to the complexity of the actual data, the results are not ideal when processing the actual data. The third-order calibration [28,29] has lots of advantages, such as the uniqueness of the tri-linear decomposition and the "second-order advantage" which is also shared by Alternating Penalty Three-linear Decomposition (APTLD) [30], Parallel Factor Analysis (PARAFAC) [31,32] and other algorithms [33,34]. Kim et al. [35] concluded that the four-linear parallel factor analysis had higher selectivity than the tri-linear parallel factor analysis. Olivieri A C [36] believed that, computing sensitivity and selectivity in Parallel Factor Analysis and Related Multi-way Techniques, third-order calibration could achieve better sensitivity. Shufang Li et al. [37] used a four-dimensional chemometrics algorithm to determine procaine hydrochloride in human plasma samples. The results showed that the third-order calibration algorithm was more advantageous. At present, several algorithms for four-dimensional data calibration, such as Quadrilinear Parallel Factor Analysis (4-PARAFAC), Alternating Penalty Quadrilinear Decomposition (APQLD), Alternating Weighted Residue Constraint Quadrilinear Decomposition (AWRCQLD), unfolded least squares and multipoint least squares Linearization (U-PLS/RTL, N-PLS/RTL) [38], are available for the analysis of four-dimensional data tensors. A toolbox, namely, multivariate calibration3 (MVC3), for implementing third-order multivariate calibration methodologies, has recently been developed by Olivieri et al. [39]. The AWRCQLD method has not only "third-order advantage", which is more stable decomposition results, less susceptible to the number of components than 4-PARAFAC [40–42] and stronger anti-interference ability than APQLD [37,43]. In this work, we analyzed the four-dimensional data matrix using three kinds of third-order [44] calibration algorithms and compared respectively the recovery rate [45] of PAHs in three solvents. Experimental results showed that PAHs can be measured more accurately by the high-order data obtained from the combination of three excitation-emission fluorescence spectra of different concentration PAHs which are in methanol, ethanol and Ultra-pure water. Because of the similar chemical structure of PAHs, it cannot be directly measured by traditional fluorescence method. Therefore the AWRCQLD algorithm was used to achieve qualitative and quantitative analysis of PAHs trace and obtained good results.

2. Theoretical Part

2.1. Quadrilinear Model

For a quadrilateral composition model, the expression of each element x_{ijkl} of the four-dimensional data matrix $\underline{X}_q$ can be deduced from the multi-linear component model by scalar representation as shown in Eq. (1):
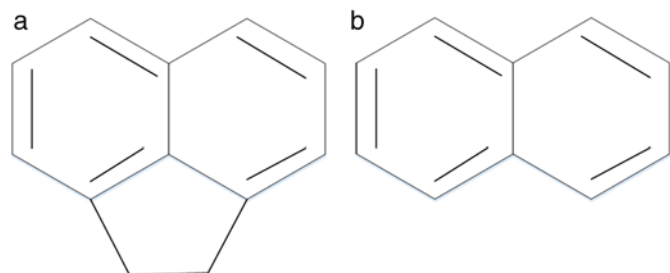
$$x_{ijkl} = \sum_{n=1}^{N} a_{in}b_{jn}c_{kn}d_{ln} + e_{ijkl} \tag{1}$$
$$i = 1, 2, \cdots, I, j = 1, 2 \cdots, J, k = 1, 2, \cdots, K, l = 1, 2, \cdots L$$

When there are N fluorophores in the samples, a tensor of experimental data is compatible with the structure of Eq. (1). That is, the data are quadrilinear. Where a, b and c are the elements of the matrix A, B and C, respectively. E_{ijkl} is a constituent element corresponding to the four-dimensional residual array E_q. N is the total number of components of the quadrilinear model (the actual contribution to the fluorescence and the background of the main interference).

According to cyclical symmetry properties of the linear component model, the four-linear component model can also be used by four completely equivalent slice-wise matrices and four fully equivalent stretched matrices.

One of slice-wise matrices is represented as Eq. (2), and so is the others.

$$X_{i \cdot l} = Bdiag(a_{(i)}) \; diag(d_{(l)})C^T + E_{i \cdot l} \quad l = 1, 2, \cdots, L; i = 1, 2, \cdots I \tag{2}$$

In the formula, diag represents the vector diagonalization, the vector into a vector element for the diagonal matrix.

One of stretched matrices expressions is shown in Eq.(3), and so is the others.

$$X_{I \times JKL} = A(D \odot C \odot B)^T + E_{I \times JKL} \tag{3}$$

There, '$\odot$' represents product of 'Khatri-Rao'.

2.2. AWRCQLD Algorithm

On the basis of the least squares principle [46], the weighted residual function is used as the constraint term of the loss function of

Fig. 1. Chemical structure of ANA and NAP.

Table 1
Concentrations of two PAHs samples (μgL^{-1}).

Sample no.	In different solvents	ANA	NAP	Sample No.	In different solvents	ANA	NAP
C1	C1^e	0.5	0	C9	C9^e	4.2	0.8
	C1^m	0.5	0		C9^m	4.2	0.8
	C1^U	0.5	0		C9^U	4.2	0.8
C2	C2^e	0	0.8	T1	T1^e	0.6	0.9
	C2^m	0	0.8		T1^m	0.6	0.9
	C2^U	0	0.8		T1^U	0.6	0.9
C3	C3^e	0.5	3.8	T2	T2^e	2.1	0.7
	C3^m	0.5	3.8		T2^m	2.1	0.7
	C3^U	0.5	3.8		T2^U	2.1	0.7
C4	C4^e	1.5	3.5	T3	T3^e	2.5	2.0
	C4^m	1.5	3.5		T3^m	2.5	2.0
	C4^U	1.5	3.5		T3^U	2.5	2.0
C5	C5^e	2.2	3.1	T4	T4^e	2.7	2.6
	C5^m	2.2	3.1		T4^m	2.7	2.6
	C5^U	2.2	3.1		T4^U	2.7	2.6
C6	C6^e	2.6	2.9	T5	T5^e	2.3	2.4
	C6^m	2.6	2.9		T5^m	2.3	2.4
	C6^U	2.6	2.9		T5^U	2.3	2.4
C7	C7^e	3.0	3.3	T6	T6^e	3.3	3.3
	C7^m	3.0	3.3		T6^m	3.3	3.3
	C7^U	3.0	3.3		T6^U	3.3	3.3
C8	C8^e	3.5	2.2	T7	T7^e	1.6	1.8
	C8^m	3.5	2.2		T7^m	1.6	1.8
	C8^U	3.5	2.2		T7^U	1.6	1.8

C1^e ~ C9^e are calibration samples in ethanol solvent. C1^m ~ C9^m are calibration samples in methanol solvent. C1^U ~ C9^U are calibration samples in Ultra-pure water. T1^e ~ T9^e are test samples in ethanol solvent. T1^m ~ T9^m are test samples in methanol solvent. T1^U ~ T9^U are test samples in Ultra-pure water.