

# Application of Three-Dimensional Fluorescence Spectra Combined with Algorithm Combination Methodology in Environmental Pollution Monitoring: Oil Identification and Quantitative Analysis

CHEN Zhi-kun<sup>1</sup>, HUANG Wei<sup>1\*</sup>, CHENG Peng-fei<sup>1</sup>, SHEN Xiao-wei<sup>1</sup>,  
WANG Fu-bin<sup>1</sup>, WANG Yu-tian<sup>2</sup>

1. College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China

2. Measurement Technology and Instrument Key Lab of Hebei Province, Yanshan University, Qinhuangdao 066004, China

**Abstract** In order to solve the problem that the composition of oil pollutants is complex and the spectrum overlap is difficult to identify, qualitative and quantitative analysis of oil pollutants was carried out by three-dimensional fluorescence spectroscopy combined with Algorithm Combination Methodology (ACM). Rayleigh scattering in fluorescence spectra has a great influence on the detection of three-dimensional fluorescence spectrum. In this paper, the missing data recovery-principal component analysis (MDR-PCA) method was proposed. The principle is that the single fluorescence spectrum excitation emission matrix conforms to bilinearity and can be analyzed by principal component analysis (PCA). The scattering interference data were first deducted completely, and then the deducted part was repaired by using the remaining effective signal data in the iteration process. This method not only eliminates the scattering interference, but also makes full use of the effective information in the fluorescence spectrum matrix. The three-dimensional data were constructed by using the excitation-emission fluorescence spectra of different concentration mineral oil. The sample data were obtained from carbon tetrachloride solutions of 0<sup>#</sup> diesel, 95<sup>#</sup> gasoline and ordinary kerosene solutes. The trilinear decomposition algorithms commonly used for three-dimensional fluorescence spectral data analysis include parallel factor analysis (PARAFAC), alternating trilinear decomposition (ATLD), and self-weighted alternating trilinear decomposition algorithm (SWATLD). PARAFAC is based on the strict principle of least squares and has strong anti-noise ability. Its model is the stablest and the error is expected to be the smallest. It can provide the best fit of 3D data array, but the convergence speed of PARAFAC algorithm is slow and correct. The estimated number of components is more sensitive. The ATLD algorithm is based on the Moore-penrose generalized inverse of singular value decomposition to realize the trilinear model decomposition. By using the inverse diagonal element and the tangent singular value to solve the generalized inverse, the convergence speed of the method is greatly improved, and the sensitivity of the algorithm to the component number is reduced, but the operation of the diagonal element makes the ATLD method more sensitive to noise. SWATLD inherits the advantages of ATLD, which is insensitive to the number of components and fast convergence, and has the characteristics of being insensitive to noise levels. However, the SWATLD algorithm has a slightly lower ability to resist collinearity than ATLD. This paper divides the iteration process according to the change of loss function in the iteration process of trilinear decomposition algorithm, and proposes the algorithm combination method (ACM)-combining ATLD, SWATLD and PARAFAC, giving full play to the advantages of each algorithm, and realizing the complementary advantages of the second-order correction algorithm. The

**Received:** 2018-09-11; **accepted:** 2019-01-23

**Foundation item:** the National Natural Science Foundation of China (61471312, 61771419) and the Hebei Provincial Natural Science Foundation Project (F2015203240)

**Biography:** CHEN Zhi-kun, (1961—), doctor, professor in North China University of Science and Technology  
e-mail: zkchen@ncst.edu.cn \* Corresponding author e-mail: huangwei\_ncst@163.com

three-dimensional fluorescence spectra of two-component and three-component mineral oil samples were analyzed by ACM algorithm, and the recovery rates of three mineral oil samples were calculated. The recovery rate of diesel was 97.08%, the recovery rate of gasoline was 97.34%, and the recovery rate of kerosene was 97.25%. The analytical spectrum and the recovery rate show that the ACM algorithm can realize species identification and concentration measurement of oil pollutants.

**Keywords** Three-dimensional fluorescence spectrum; Algorithm combination methodology; Missing data recovery-principal component analysis; Component number; Recovery rate

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3313-08

## Introduction

With the development of modern industrialization, water pollution is becoming more and more serious. Petroleum pollution is widespread and becomes one of the most important pollutants. In recent years, in the detection of oil pollutants, spectrophotometry is becoming more and more popular, especially 3D fluorescence spectrometry. 3D fluorescence spectroscopy<sup>[1-3]</sup> with high sensitivity, good selectivity and fast analysis speed, can be used for multi-components mixture analysis.

In the three-dimensional fluorescence spectrum, Rayleigh scattering appears in the region where the excitation wavelength is close to the emission wavelength. The Rayleigh scattering signal has a high intensity, sometimes several orders of magnitude higher than the fluorescence signal, and its intensity is inversely proportional to the fourth power of the wavelength. Since the Rayleigh scattering region does not conform to bilinear and trilinear, it often causes great damage to the fluorescence spectrum analysis results, so it should be eliminated as much as possible in the fluorescence spectrum analysis. The missing data recovery-principal component analysis was used to pretreatment the experimental data in this work. In this method, the scattering interference data were firstly deducted and then the subtracted part was repaired by using the fluorescence data in the fluorescence material spectrum matrix, thus avoiding the interference of the scattering region to the analysis.

At present, the second-order correction algorithm in chemometrics has been widely used in the analysis of three-dimensional fluorescence spectrum data<sup>[4-8]</sup>. The main reason is that this algorithm can realize the "second-order advantage", that is, in the case of unknown interference coexistence, the analyte of interest can also be qualitatively and quantitatively resolved. In many multidimensional data decomposition algorithms, the second-order calibration algorithm<sup>[9-11]</sup> can achieve the same result when dealing with ideal data. However, due to the complexity of the actual data, the results are not ideal when processing the actual data. In this paper, a new trilinear

decomposition algorithm was proposed, which combines ATLD, SWATLD and PARAFAC to give full play to the advantages of each algorithm to realize the complementary advantages of the second-order correction algorithm<sup>[12]</sup>. The ACM algorithm was used to analyze the three-dimensional fluorescence spectrum data of two-component and three-component mineral oil samples, and the effectiveness of ACM for the detection of mixed oils with spectral overlap was verified.

## 1 Theoretical part

### 1.1 Missing data recovery-principal component analysis

The single fluorescence excitation emission matrix (EEM) is bilinear and can be analyzed by principal component analysis (PCA). The MDR-PCA method firstly deducts the scattered interference data, and then uses the remaining fluorescence data to recover the deducted parts through the defect data reconstruction algorithm during the iterative process of the principal component analysis, completely avoiding the interference of the scattered signal to the fluorescence analysis. The MDR-PCA method can be divided into the following main steps:

(1) The scattering region in the EEM spectral matrix  $Z$  is identified. For Rayleigh scattering, it is generally the diagonal region of  $|EM - EX| \leq 10 \sim 15$  nm;

(2) Set the weighting matrix  $W$  of the same size as  $Z$ . The part of scattering data in corresponding matrix  $Z$  is set to 0 and the rest is set to 1;

(3) The data of spectral scattering region can be set as defective data, such as 0 or nan;

(4) In the iterative process of principal component analysis (PCA) of EEM matrix  $Z$ , the missing data in the matrix is reconstructed using the score vector  $t$  and the load vector  $p$  calculated in each step;

$$p^T = t^T \times Z \quad (1)$$

$$p = p / |p| \quad (2)$$

$$Z = Z \times W + t \times p^T \odot (1 - W) \quad (3)$$

$$t = Z \times p \quad (4)$$

$$Z = Z \odot W + t \times p^T (1 - W) \quad (5)$$

(5) Iterative calculation to convergence, convergence loss

function does not take into account the missing data part;

(6) If you need to calculate more principal components, make  $Z_{k+1} = Z_k - t_k \times p_k^T$ , continue the calculation of step (4)–(5);

(7) Replace the scattering region with the principal component reconstructed defect region to obtain a repaired spectral matrix  $X$ :

$$X = Z \odot W + T \times p^T \odot (1 - W) \quad (6)$$

This method not only eliminates the scattering interference, but also makes full use of the effective information in the fluorescence spectrum matrix to repair the scattering region, which is beneficial to the non-destructive analysis of the three-dimensional fluorescence spectrum.

### 1.2 Algorithm combination methodology

The expression of the trilinear model is:

$$x_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + e_{ijk}$$

$$i = 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K \quad (7)$$

There,  $x_{ijk}$  is an element of the three dimensional data matrix  $\mathbf{X}$ ;  $e_{ijk}$  is a constituent element corresponding to the error matrix  $\mathbf{E}$ ;  $a_{in}$ ,  $b_{jn}$  and  $c_{kn}$  are the elements of the matrix  $A$ ,  $B$  and  $C$ , respectively.

According to the change of loss function

$$\left( \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \sum_{n=1}^N a_{in} b_{jn} c_{kn})^2 \right)$$

the iterative optimization process of trilinear decomposition algorithm is decomposed. The loss function increases with the number of iterations as shown in Figure 1. From the figure we can see that from the beginning of the random initial value of the algorithm to the convergence of the analytical results with physical meaning, this process can be roughly divided into three parts: Initial value optimization process, algorithm optimization process and algorithm convergence process.

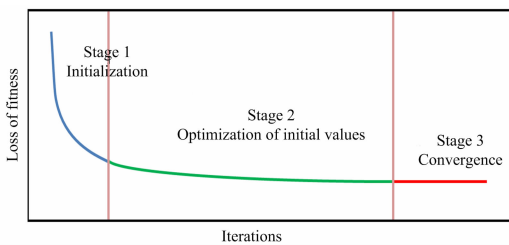


Fig. 1 Loss function change process

Comparing the performance of the ATLD, SWATLD, and PARAFAC algorithms with the division of the iterative process, it can be seen that the three algorithms are more important than optimizing the different parts of the iterative process. Therefore, we can try to use the above three algorithms in the corresponding part to realize the complementary advantages of each algorithm. The combination of the three algorithms described above is implemented in the following

manner: Firstly, the random initial value is optimized by ATLD, then the convergence result of ATLD is further optimized by SWATLD. Finally, the result of SWATLD is optimized by using PARAFAC to realize the trilinear decomposition of data. This algorithm is called the algorithm combination methodology (ACM).

When the data structure is simple, there is no obvious difference in the results obtained by each algorithm. Therefore, ACM can be further optimized to make it more efficient in data parsing. From the experience of application, the loss function of SWATLD will decrease monotonously until it converges when the data structure is relatively simple. This phenomenon is incorporated into ACM. If the loss function of SWATLD converges monotonously, the result will be further optimized without using PARAFAC. Otherwise, you need to introduce PARAFAC. The whole process of ACM is shown in Fig. 2.

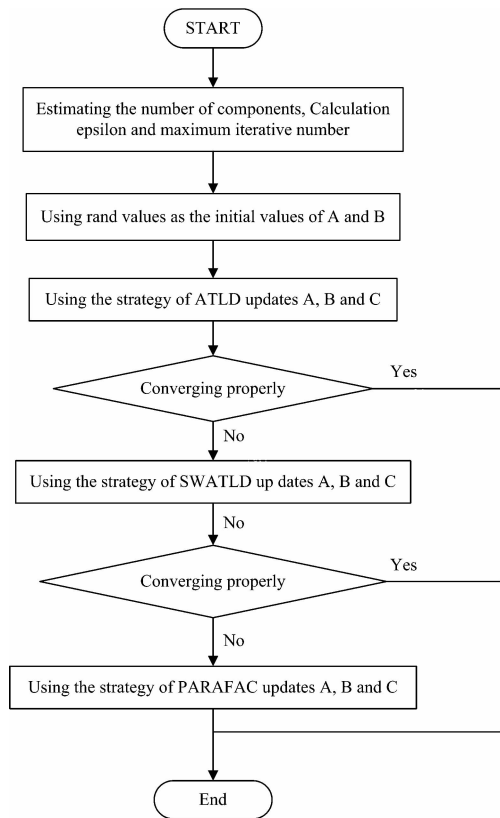


Fig. 2 Algorithm combination methodology data analysis flow chart

## 2 Experimental

### 2.1 Experimental instruments and equipment

Hitachi F-7000 fluorescence spectrometer was used for the experiment. Parameter settings: excitation wavelength range 250 ~ 430 nm, emission wavelength range 310 ~ 520

nm, step length is 5 nm— slit width is 10 nm; scanning speed is  $12\,000\text{ nm} \cdot \text{min}^{-1}$ , PMT voltage is 400 V. Each sample was measured three times in parallel, and the average value was taken as the fluorescence spectrum of the sample.

## 2.2 Experimental materials and sample preparation

Three product oils (0<sup>#</sup> diesel, 95<sup>#</sup> gasoline, and ordinary kerosene) were selected as representatives of petroleum-based materials, and carbon tetrachloride as solvent. Separately take 1g of each of 0<sup>#</sup> diesel, 95<sup>#</sup> gasoline, and ordinary

kerosene in a 100 mL volumetric flask with an electronic balance, and add  $\text{CCl}_4$  to fully dissolve to obtain a standard solution of  $10\,000\text{ mg} \cdot \text{L}^{-1}$  of three oils. Separately take three different standard solutions of different volumes in Eighteen 50 mL volumetric flasks and add  $\text{CCl}_4$  to volume to prepare samples of different concentrations. Number them; 1<sup>#</sup>—11<sup>#</sup> are calibration samples, and 12<sup>#</sup>—26<sup>#</sup> are prediction samples. The concentration of mineral oil in each sample is shown in Table 1.

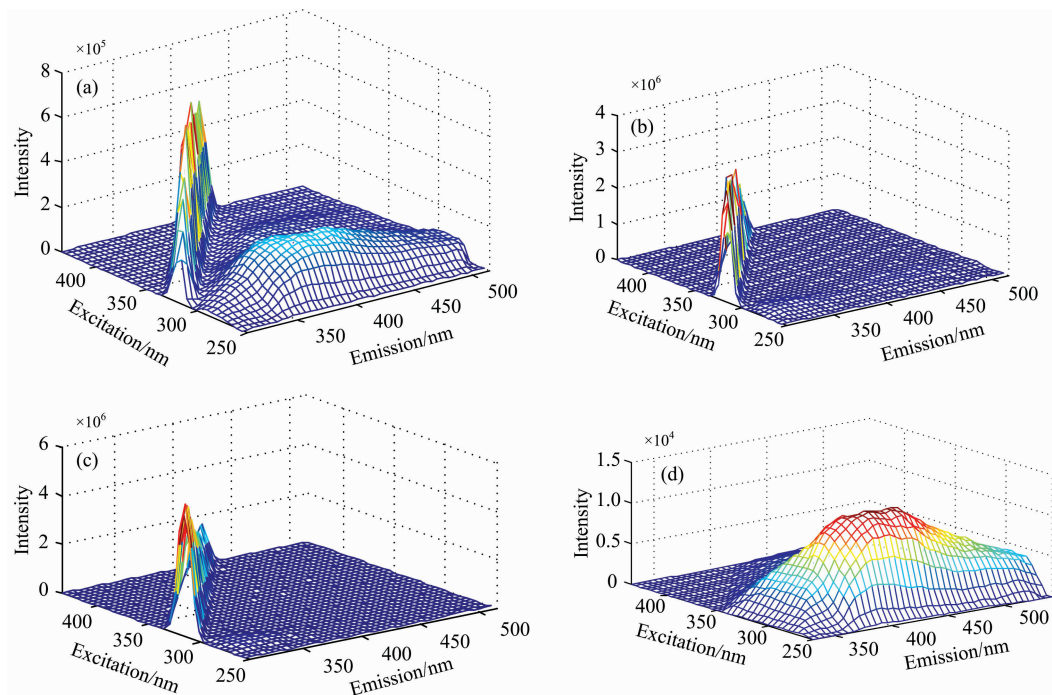
Table 1 The concentration of oil in the sample ( $\text{mg} \cdot \text{L}^{-1}$ )

Samples	0 <sup>#</sup> Diesel	95 <sup>#</sup> gasoline	kerosene	Samples	0 <sup>#</sup> Diesel	95 <sup>#</sup> gasoline	kerosene
1	100	0	0	14	30	70	0
2	0	100	0	15	20	80	0
3	0	0	100	16	40	28	0
4	80	10	0	17	60	17	0
5	60	20	0	18	80	10	0
6	50	30	0	19	50	40	65
7	70	25	5	20	25	50	50
8	60	10	25	21	60	25	25
9	30	20	50	22	80	10	10
10	10	25	50	23	30	35	13
11	5	50	80	24	50	22	26
12	50	25	0	25	70	14	39
13	75	50	0	26	40	30	20

## 2.3 Data preprocessing

In the excitation-emission fluorescence spectral matrix (EEM) of diesel, gasoline and kerosene carbon tetrachloride solution, Rayleigh scattering occurs in the region where the excitation wavelength is similar to the emission wavelength,

as shown in Fig. 3(a), (b) and (c). The Rayleigh scattering signal has a high intensity, sometimes up to several orders of magnitude higher than the fluorescence signal, and its intensity is inversely proportional to the fourth power of the wavelength. Because the Rayleigh scattering region does not accord



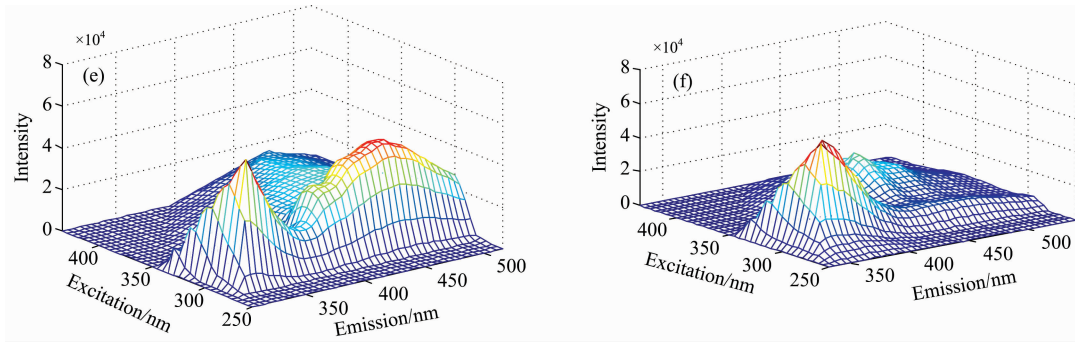


Fig. 3 3-D fluorescence spectrum of three standard samples

(a): 3-D fluorescence spectrum of 0# diesel without calibration; (b): 3-D fluorescence spectrum of 95# gasoline without calibration; (c): 3-D fluorescence spectrum of kerosene without calibration; (d): 3-D fluorescence spectrum of 0# diesel after eliminating scattering; (e): 3-D fluorescence spectrum of 95# gasoline after eliminating scattering; (f): 3-D fluorescence spectrum of kerosene after eliminating scattering

with bilinear and trilinear, it often causes great damage to the results of fluorescence spectrum analysis, so it should be eliminated as far as possible in fluorescence spectrum analysis. Using the MDR-PCA method, the region  $|EM-EX| \leq 8$  nm is set as the data defect area and the scattering region in the EEM spectrum is corrected. The corrected EEM spectrum is shown in Fig. 3(d), (e) and (f). Compared with Fig. 3(a), (b) and (c), it can be seen that the EEM spectrum has a smooth transition in the scattering region. The MDR-PCA method can effectively utilize the normal fluorescence signal and eliminate the influence of the scattering interference signal.

### 3 Analysis of mixed oil by ACM algorithm

#### 3.1 Diesel and gasoline mixed solution

Scanning calibration samples 1#—6# and prediction samples 12#—18#, and after data preprocessing to remove the scattering, a  $13 \times 37 \times 43$  three-dimensional matrix  $X_1$  is constructed. Before using the ACM algorithm, it is important to determine the number of factors in the analysis object. In this paper, the factor number of three-dimensional data matrix of experimentally measured is estimated by using core consistency diagnostics method and residual analysis method<sup>[13-14]</sup>. The changes of the core consistency value and residual sum of squares with the number of factors are shown in Fig. 4.

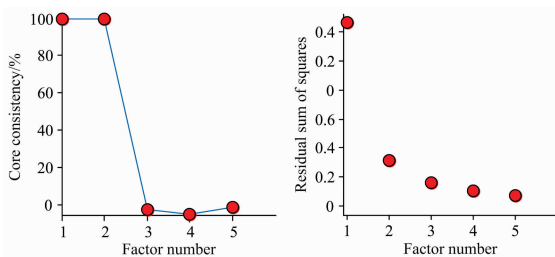


Fig. 4  $X_1$  nuclear consistent diagnosis results and residual square sum analysis results

When the factor number is 1 or 2, the kernel concordance is 100%, and when the factor number is 3, 4, 5, the nuclear concordance coefficient decreases significantly. When the number of factors increases, the sum of squared residuals decreases monotonously, that is, the decomposition error can be reduced by selecting more factors. According to the two indexes, the selected factor number is 2, which is consistent with the actual situation of sample preparation. For the 2-factor analysis, the excitation and emission spectra obtained by the discrimination are shown in Fig. 5. It can be seen that the spectrum after the resolution of diesel and gasoline has high spectral similarity and good reproducibility compared with the actual measurement, which qualitatively shows that ACM decomposition has good resolving power for mixed oils.

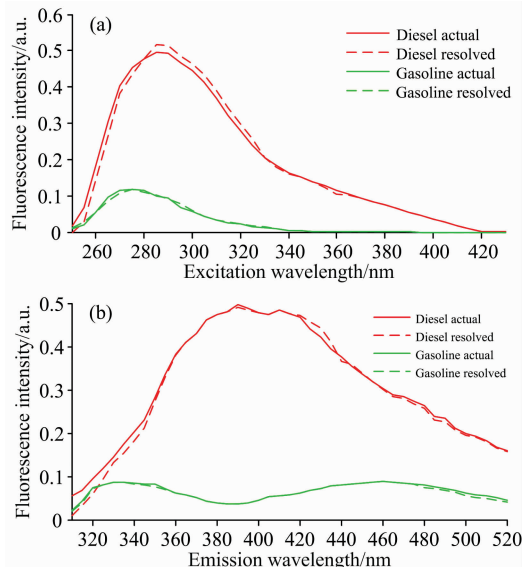


Fig. 5 Real spectra and ACM decomposition spectra of diesel and gasoline

The concentration of each oil component in the predicted sample was predicted, and the result was expressed as recov-

ery rate. The expression of the recovery rate is

$$R = (A/B) \times 100\% \quad (8)$$

There,  $A$  is the predicted concentration;  $B$  is the real concentration.

The predicted concentrations of diesel and gasoline in the 12<sup>#</sup>—18<sup>#</sup> predicted samples and the sample recovery rate are

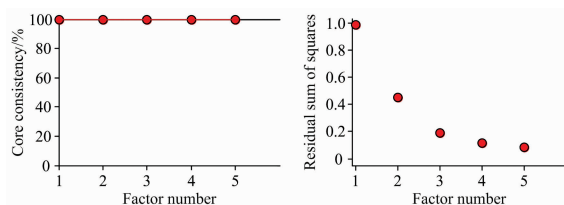
shown in Table 2. It can be seen from the table that the predicted concentration of diesel and gasoline is very close to the true concentration, the average recovery rate is 97.36% and 97.26% respectively. The quantitative analysis indicates that the ACM decomposition algorithm has a good resolution for diesel and gasoline.

**Table 2** Predicted concentration and recovery of ACM decomposition of diesel and gasoline

Samples	Predicted concentration/(mg · L <sup>-1</sup> )		Recovery/%		Average recovery/%	
	Diesel	gasoline	Diesel	gasoline	Diesel	gasoline
12	51.82	20.62	103.6	82.5		
13	73.90	45.23	98.5	90.5		
14	26.33	76.44	87.8	109.2		
15	19.66	83.92	98.3	104.9	97.36	97.26
16	38.92	27.63	97.3	98.7		
17	58.61	16.77	97.7	98.6		
18	78.65	9.64	98.3	96.4		

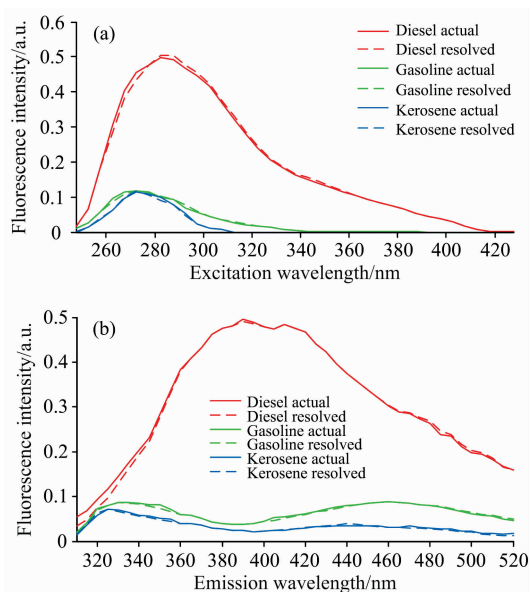
### 3.2 Mixed solution of diesel, gasoline and kerosene

Scanning calibration samples 7<sup>#</sup>—11<sup>#</sup> and prediction samples 19<sup>#</sup>—26<sup>#</sup>, and after data preprocessing to remove the scattering, a 13 × 37 × 43 three-dimensional matrix  $\mathbf{X}_2$  is constructed. The changes of the core consistency value and residual sum of squares with the number of factors are shown in Fig. 6.



**Fig. 6**  $\mathbf{X}_2$  nuclear consistent diagnosis results and residual square sum analysis results

Assuming that there are five factors, the degree of core consistency is always close to 100%, that is, from the point of view of core consistency, it is acceptable to use one to five factors for ACM analysis. Analysis of the sum of squared residuals found that the sum of the squared residuals of the 3-factor analysis was significantly lower than that of the 2-factor analysis, but it was less different from the sum of the squared residuals of the 4-factor analysis. The known samples are made up of three substances; diesel, gasoline and kerosene. If the ACM analysis of  $F=3$  is used according to the actual condition of sample preparation, the excitation and emission spectra can be resolved as shown in Fig. 7. It can be seen that the resolving spectra of diesel, gasoline, and kerosene have higher similarity with the actual spectrum and good repeatability, which qualitatively shows that ACM decomposition has good resolution of oil mixtures.



**Fig. 7** Real spectra and ACM decomposition spectra of diesel, gasoline, and kerosene

Table 3 lists the predicted concentrations and sample recoveries of diesel, gasoline and kerosene from ACM decomposition in the predicted sample of 19<sup>#</sup>—26<sup>#</sup>. The average recovery rates of diesel, gasoline, and kerosene were 97.08%, 97.34%, and 97.25% respectively. The fitting accuracy is high and the concentration prediction deviation is small. It can be seen that ACM decomposition algorithm also has good resolution ability for three or more kinds of mixed oil.

## 4 Conclusion

(1) In the process of spectral data preprocessing, the missing data recovery-principal component analysis is used to

**Table 3 Predicted concentration and recovery of ACM decomposition of diesel, gasoline and kerosene**

Samples	Predicted concentration/(mg · L <sup>-1</sup> )			Recovery/%			Average recovery/%		
	Diesel	gasoline	kerosene	Diesel	gasoline	kerosene	Diesel	gasoline	kerosene
19	47.81	38.68	64.30	95.6	96.7	98.92			
20	23.64	48.94	47.83	94.6	97.9	95.66			
21	59.05	24.51	24.02	98.4	98.0	96.08			
22	79.92	9.62	9.75	98.7	96.2	97.5			
23	28.98	33.90	12.73	96.6	96.9	97.9	97.08	97.34	97.25
24	48.77	21.68	25.54	97.5	98.5	98.2			
25	68.67	13.73	37.54	98.1	98.1	96.3			
26	38.85	28.91	19.47	97.1	96.4	97.4			

deal with experimental spectral data, which effectively solves the interference problem of scattering signal to fluorescence analysis.

(2) Due to the similarity of chemical components in each petroleum product, there are some overlapping regions or even similarities in the spectra of each petroleum product, which makes it difficult to distinguish the spectra of mixed solutions. Combining the respective advantages of ATLD, SWATLD and PARAFAC algorithm, an algorithm combina-

tion strategy is proposed.

(3) In this paper, the method of combining the three-dimensional fluorescence spectroscopy with ACM algorithm is proposed. The number of factors is determined by using core consistent diagnosis and residual error analysis method. The combined algorithm is used for the “mathematical separation” of the mixed solution. The experimental results show that the combination algorithm has a good separation effect for the mixed solution with overlapping spectra.

## References

- [ 1 ] Zhou Yanlei, Zhou Feifei, Jiang Congcong, et al. Spectroscopy and Spectral Analysis, 2018, 38(2): 475.
- [ 2 ] Wang Yutian, Zhang Lijuan, Zhao Xu, et al. Spectroscopy and Spectral Analysis, 2017, 37(12): 3771.
- [ 3 ] Xiao Changjiang, Zhang Jingchao. Chinese Journal of Luminescence, 2017, 38(10): 1391.
- [ 4 ] Azcarate S M, Teglia C M, Karp F, et al. Microchemical Journal, 2017, 133.
- [ 5 ] Wang Juan, Zhang Fei, Wang Xiaoping, et al. Acta Optica Sinica, 2017, 37(7): 357.
- [ 6 ] Wang Yutian, Liu Tingting, Liu Lingfei, et al. Spectroscopy and Spectral Analysis, 2018, 38(4): 1171.
- [ 7 ] Xu Buyi, Ye Yi, Ruan Ruoyun, et al. Chemical Journal of Chinese Universities, 2015, 36(9): 1667.
- [ 8 ] Wu Hailong, Li Yong, Kang Chao, et al. Chinese Journal of Analytical Chemistry, 2015, 43(11): 1629.
- [ 9 ] Wang Li, Wu Hailong, Yin Xiaoli. Spectrochimica Acta Part A: Molecular and Biomolecular, 2017, 170: 104.
- [ 10 ] S Maryam Sajjadi, Hamid Abdollahi, Reza Rahmanian. Spectrochimica Acta Part A: Molecular and Biomolecular, 2016, 156: 63.
- [ 11 ] Maryam Vosough, Sara Noroozi Eshlaghi, Reza Zadmand. Spectrochimica Acta Part A: Molecular and Biomolecular, 2015, 136: 655.
- [ 12 ] Yu Yongjie. Hunan University, 2012.
- [ 13 ] Li Run, Chen Guoqing, Zhu Chun, et al. Spectroscopy and Spectral Analysis, 2014, 34(1): 111.
- [ 14 ] Maggio R M, Peña A M D L, Olivieri A C. Chemometrics & Intelligent Laboratory Systems, 2011, 109(2): 178.

# 三维荧光光谱结合组合算法在环境污染监测中的应用： 油种鉴别与定量分析

陈至坤<sup>1</sup>, 黄 微<sup>1\*</sup>, 程朋飞<sup>1</sup>, 沈小伟<sup>1</sup>, 王福斌<sup>1</sup>, 王玉田<sup>2</sup>

1. 华北理工大学电气工程学院, 河北 唐山 063210

2. 燕山大学河北省测试计量技术与仪器重点实验室, 河北 秦皇岛 066004

**摘 要** 针对油类污染物成分复杂, 光谱重叠难以识别的问题, 提出采用三维荧光光谱结合组合算法对油类污染物进行了定性和定量分析。荧光光谱中存在的瑞利散射对三维荧光光谱检测有较大影响, 提出了缺损数据修复-主成分分析(MDR-PCA)方法对矿物油三维荧光光谱的瑞利散射进行处理, 原理是单个荧光光谱激发发射矩阵符合双线性, 可用主成分分析(PCA)法来解析。MDR-PCA法首先将荧光数据中的散射干扰数据全部扣除, 之后利用主成分分析(PCA)迭代过程对扣除数据进行重构修复后补全数据。该方法在消除散射干扰的同时充分利用了荧光物质光谱矩阵中的有效信息。利用不同浓度的矿物油的激发-发射荧光光谱构建了三维数据。样品数据来源于柴油、汽油和煤油三种溶质的四氯化碳溶液。常用于三维荧光光谱数据分析的三线性分解算法有平行因子分析(PARAFAC)、交替三线性分解(ATLD)和自加权交替三线性分解算法(SWATLD)等。PARAFAC基于严格意义上的最小二乘原则, 具有抗噪声强、模型稳定、微小预期误差等优点, 可以实现三维数据阵列的最佳拟合, 但该算法收敛速度较慢, 对组分数敏感。ATLD算法通过提取对角主元和切尾奇异值求解广义逆, 极大提高了收敛速度并降低了对组分数的敏感度, 从而实现三线性分解。然而, 取对角元时易使ATLD方法对噪声敏感。SWATLD算法既继承了对组分数不敏感、收敛速度快等优点, 又降低了噪声水平的影响。但是在抗共线性程度方面, SWATLD算法在抵抗共线性程度方面的能力较ATLD略有降低。基于此, 论文根据三线性分解算法迭代过程中损失函数的变化, 对迭代过程进行划分, 提出了三线性迭代方法的组合算法(algorithm combination methodology, ACM)—将ATLD, SWATLD与PARAFAC组合在一起, 充分发挥各算法的优点, 实现二阶校正算法的优势互补。采用ACM算法对两组分及三组分矿物油样品的三维荧光光谱数据进行解析, 并对三种矿物油的回收率进行了计算。柴油的回收率为97.08%, 汽油的回收率为97.34%, 煤油的回收率为97.25%。解析光谱和回收率表明, ACM算法能够实现油类污染物的种类识别及浓度测量。

**关键词** 三维荧光光谱; 组合算法; 缺损数据修复-主成分分析; 组分数; 回收率

(收稿日期: 2018-09-11, 修订日期: 2019-01-23)

\* 通讯联系人