

基于组合算法的油类污染物三维荧光光谱分析

陈至坤, 黄微*, 程鹏飞, 沈小伟, 王福斌

华北理工大学电气工程学院, 河北 唐山 063210

摘要 针对油类污染物成分复杂、光谱重叠难以识别的问题, 提出三维荧光光谱结合组合算法(ACM)。将交替三线性分解(ATLD)、自加权交替三线性分解(SWATLD)与平行因子分析(PARAFAC)算法组合, 实现 3 种算法的优势互补。通过配制以四氯化碳为溶剂的不同质量浓度的柴油、汽油和煤油的混合溶液, 利用 F-7000 荧光光谱仪测量混合溶液的三维荧光光谱, 采用空白扣除法与缺损数据修复——主成分分析法进行预处理消除散射干扰, 对三维光谱数据矩阵进行分解, 并与以上 3 种算法解析结果进行对比。结果表明, ACM 对组分数不敏感, 且解析结果更准确, 样本中对柴油、汽油和煤油的平均回收率分别为 96.68%、97.83%、97.11%。实现了混合油类物质的定性、定量分析, 具有一定的普适性。

关键词 光谱学; 荧光分析; 组合算法; 油类污染物; 三线性分解

中图分类号 O433.4; O657.3

文献标识码 A

doi: 10.3788/LOP56.033002

Three-Dimensional Fluorescence Spectra Analysis of Oil Contaminants Based on Algorithm Combination Methodology

Chen Zhikun, Huang Wei*, Cheng Pengfei, Shen Xiaowei, Wang Fubin

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

Abstract This study proposes a new technique that combines three-dimensional fluorescence spectra with algorithm combination methodology (ACM) to address issues associated with complex components of oil pollutants and the difficulty in identifying their overlapping spectra. By combining alternating trilinear decomposition (ATLD), self-weighted alternating trilinear decomposition (SWATLD), and parallel factor analysis (PARAFAC), ACM realizes the complementary advantages of using three algorithms. First, using carbon tetrachloride as the target contaminant, a three-component mixed solution of diesel, gasoline, and kerosene with different concentrations are prepared. Then, the three-dimensional fluorescence spectra of the mixed solution are measured using a F-7000 fluorescence spectrometer. Blank deduction and missing data recovery-principal component analysis are then employed as pretreatment methods to eliminate the scattering interference. Finally, ACM is used to decompose the three-dimensional spectral data matrix. Results are compared with the three separate algorithms for component analysis, revealing that ACM is insensitive to component concentration. The average recoveries for diesel, gasoline, and kerosene are 96.68%, 97.83% and 97.11%, respectively, which further indicated that this method is more universal and can be used for the qualitative and quantitative analyses of contaminants in oil mixtures.

Key words spectroscopy; fluorescence analysis; algorithm combination methodology; oil contamination; trilinear decomposition

OCIS codes 300.2530; 300.6280

1 引言

二阶校正算法可在未知干扰存在的情况下分辨和测量出目标分析物, 并保证其具有较高的精度, 从

而实现“二阶优势”, 因此广泛应用于三维荧光光谱数据的分析中^[1-7]。目前, 研究人员已开发了许多分析性能出色的二阶校正算法。Bro^[8]提出的平行因子分析(PARAFAC)算法基于严格意义上的最小二乘原

收稿日期: 2018-06-15; 修回日期: 2018-08-29; 录用日期: 2018-08-31

基金项目: 国家自然科学基金(61471312, 61771419)、河北省自然科学基金(F2015203240)

* E-mail: huangwei_ncst@163.com

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,应用范围较广。但该算法收敛速度较慢,对组分数敏感。基于此,Wu等^[9]提出交替三线性分解(ATLD)算法。而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^[10]。然而,在取对角元时 ATLD 方法对噪声敏感,分析结果的准确性降低。Chen 等^[11]在 ATLD 的基础上提出了自加权交替三线性分解(SWATLD)算法。即通过创造性的增加额外权重项,达到了平衡各目标函数的目的,该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。但是在抗共线程度方面,SWATLD 算法仍有待提高。

基于三线性分解算法迭代过程中损失函数的变化,本文将迭代过程进行划分,提出了三线性迭代法的组合算法(ACM),即将 ATLD、SWATLD 与 PARAFAC 算法有机组合在一起,充分发挥每种算法的优势,实现二阶校正算法的优势互补^[12]。采用 ACM 对三组分矿物油样品的三维荧光光谱数据进行解析,验证了 ACM 用于光谱重叠的混合油类物质检测的有效性。

2 基本原理

2.1 算法组合策略

三线性模型表达式为

$$x_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + e_{ijk}, i = 1, 2, \dots, I; \\ j = 1, 2, \dots, J; k = 1, 2, \dots, K, \quad (1)$$

式中 x_{ijk} 为三维数据矩阵 \mathbf{X} 中的元素; e_{ijk} 为误差矩阵 \mathbf{E} 中的对应元素; a_{in} 、 b_{jn} 、 c_{kn} 分别为载荷 \mathbf{A} 、 \mathbf{B} 、 \mathbf{C} 矩阵中的元素。

依据损失函数 $\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \sum_{n=1}^N a_{in} b_{jn} c_{kn})^2$ 的变化,对三线性分解算法的迭代优化过程进行了分解,图 1 所示为损失函数随迭代次数增加的变化情况。从算法的随机初始值到具有物理意义的分析结果的收敛,这一过程大致可以分为初始化过程、初值

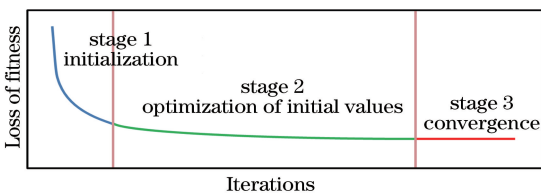


图 1 损失函数的变化过程

Fig. 1 Change process of loss function

优化过程及算法收敛过程 3 个部分。

比较 ATLD, SWATLD 和 PARAFAC 算法的性能与迭代过程的划分,不难发现 3 种算法更侧重于优化迭代过程的不同。因此,为了更好地利用各算法的优势,可以采用以下方式对以上三种算法进行组合: 1) 采用 ATLD 算法来对随机初始值进行优化; 2) 采用 SWATLD 算法对 ATLD 算法收敛结果进行进一步优化; 3) 将 SWATLD 算法所得结果用 PARAFAC 算法进行优化,这样便可以实现对数据的三线性分解。这一算法被称为组合算法 ACM。

当数据结构简单时,每种算法得到的结果无显著差异。因此,为了使算法在处理数据时更高效,可以进一步优化 ACM。在 SWATLD 算法过程中,当数据结构相对简单时,其损失函数单调下降直至收敛,将这种现象整合到 ACM 中,如果 SWATLD 算法的损失函数单调收敛,则所得结果无需采用 PARAFAC 算法优化,否则,需要引入 PARAFAC 算法。ACM 整个流程如图 2 所示。

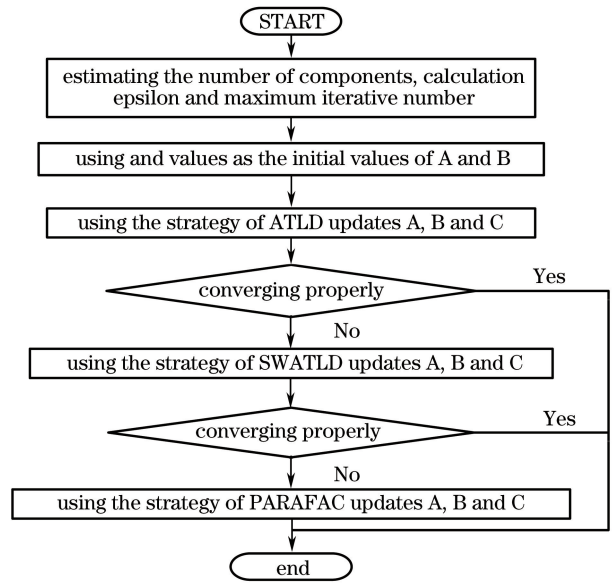


图 2 ACM 的数据分析流程图

Fig. 2 Data analysis flow chart of ACM

2.2 核一致诊断法

核一致诊断法(CORCONDIA)是由 Bro 等^[13]提出的用于估计三线性模型组分数的方法,通过比较模型中两个相关矩阵(最小二乘拟合阵 \mathbf{G} 和超对角阵 \mathbf{T})之间的相似度确定体系的因子数为

$$c = 100 \times \left[1 - \frac{\sum_{i=1}^F \sum_{j=1}^F \sum_{k=1}^F (g_{ijk} - t_{ijk})^2}{\sum_{i=1}^F \sum_{j=1}^F \sum_{k=1}^F t_{ijk}^2} \right], \quad (2)$$

式中 F 为模型的因子数; g_{ijk} 为最小二乘拟合阵 G 的元素; t_{ijk} 为超对角阵 T 的元素。

当模型选取的因子数 F 小于或等于正确的因子数时, $c \approx 100\%$; 否则, $c \approx 0$ 。通常当 $c \geq 60\%$ 时, 认为因子数选取合理, 三线性分解结果具有较高精度。

2.3 残差分析法

在复杂体系中, 核一致值并不一定随因子数增加而单调递减。可以结合另一种确定因子数的有效方法即残差分析法确定最优组分数, 依据模型残差平方和(RSS)的变化为

$$s = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \epsilon_{ijk}^2 \quad (3)$$

若因子数 F 由 1 递增, 则 RSS 会随之减小。随着 F 值逐渐逼近实际因子的数量, 其变化幅度会越来越小。若 RSS 保持稳定, 则对应 F 值可认定为最佳因子数^[14]。

表 1 样品中油类物质的质量浓度

Table 1 Concentration of oil in the sample

mg/L

Sample	0# Diesel	95# Gasoline	Kerosene	Sample	0# Diesel	95# Gasoline	Kerosene
1	1.0	5.2	2.6	9	0.9	1.2	0.8
2	0.2	4.8	2.4	10	1.0	1.0	1.0
3	1.2	3.6	1.2	11	0.3	3.5	1.3
4	0.4	3.2	1.8	12	0.4	2.8	2.0
5	2.5	2.6	3.5	13	0.5	2.2	2.6
6	0.6	2.4	5.0	14	0.6	1.7	3.3
7	5.0	1.2	4.5	15	0.7	1.4	3.9
8	0.8	1.0	2.0	16	0.8	1.0	4.5

实验参数设置: 激发波长为 250~430 nm, 步长为 5 nm, 狭缝宽度为 10 nm; 发射波长为 310~520 nm, 步长为 5 nm, 狭缝宽度为 10 nm; 扫描速率为 12000 nm/min。标准条件下测量 3 次, 取平均值。

4 结果与讨论

采用 ACM 进行荧光分析的基本步骤如图 3 所示, 通过 F-7000 荧光光谱仪分别测量实验样品, 获得激发-发射光谱矩阵(EEM); 对荧光数据进行预处理去散射; 并将所有样本的 EEM 组成三维数据矩阵(EEMs); 利用 ACM 对 EEMs 进行三线性分解, 得到样品中各组分的特征光谱与相对质量浓度。

4.1 数据预处理

以柴油、汽油、煤油的标准溶液为例, 如图 4(a)~(c)所示, 当 $E_x = E_m$ 时, 散射峰强度远远大于荧光峰且其数据不符合三线性, 不利于三线性分解。鉴于此, 可采用空白扣除法消除拉曼散射干

3 实验分析

3.1 仪器与试剂

实验采用 Hitachi F-7000 型荧光光谱仪。选择三种成品油(0#柴油、95#汽油、普通煤油)作为油类污染物的代表, 四氯化碳作为溶剂。实验中, 四氯化碳的拉曼线和激发光的波长是十分接近的, 因此对荧光光谱干扰极其微弱。

3.2 实验方法

分别称量各 1 g 的 0#柴油、95#汽油和普通煤油并置于 100 mL 容量瓶中, 加入四氯化碳溶解稀释后得到三种油的 100 mg/L 标准溶液; 分别量取不同体积的上述溶液并于 16 支 50 mL 的容量瓶中混合。加入四氯化碳溶剂, 定容配制成不同质量浓度的样品, 并对其进行编号: 1#~10#为校正样品, 11#~16#为预测样品。各样品中矿物油的质量浓度如表 1 所示。

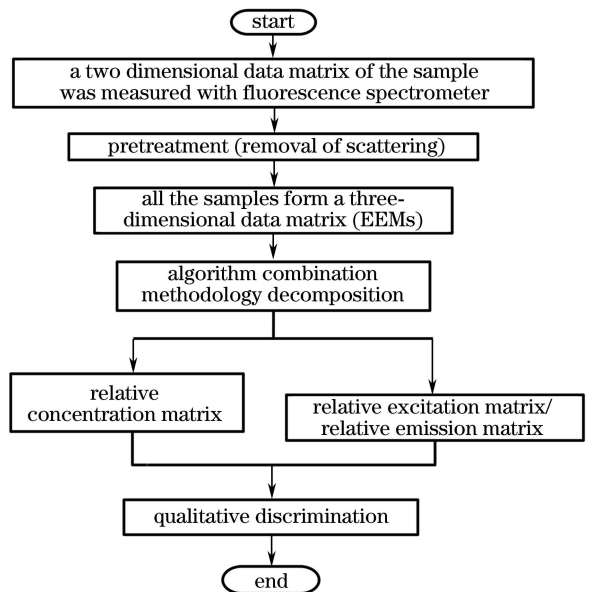


图 3 基于 ACM 的三线性分解

Fig. 3 Trilinear decomposition based on ACM

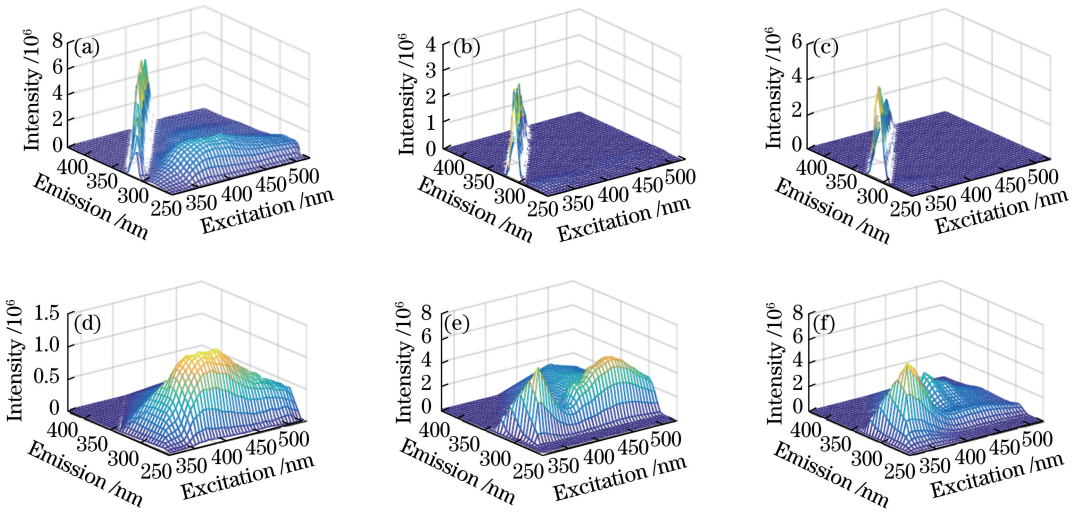


图4 标准溶液未校正的三维荧光光谱和预处理后的三维荧光光谱。
Fig. 4 Three-dimensional fluorescence spectra of uncorrected standard solution and after pretreatment.

(a) (d) 0# Diesel; (b) (e) 95# gasoline; (c) (f) ordinary kerosene

扰,采用缺损数据修复——主成分分析(MDR-PCA)法消除瑞利散射的影响^[15]。处理后荧光光谱如图4(d)~(f)所示。

4.2 ACM算法分辨柴油、汽油和煤油混合溶液

扫描校正样品1#~10#和预测样品11#~16#,经过数据预处理去除散射,并构造 $16 \times 37 \times 43$ 的三维数矩阵 X_1 。采用核一致诊断法与残差分析法结合样本的实际情况估计实验测量的三维数据阵的因子数,得到其变化情况,如图5所示。

假设体系中含5个因子,发现其核一致性总是接近100%。而3因子分析的残差平方和显著低

于2因子分析的,其与4因子分析的差别较小,表明因子数为3或4时可能较为合理。已知样本是由柴油、汽油和煤油3种物质混合而成,若依照样本配制实际情况选择因子数为3,则ATLD解析所得激发、发射光谱如图6(a)、(b)所示,而ACM解析所得激发、发射光谱如图6(c)、(d)所示。实线为真实光谱,虚线为解析光谱。可以看出,两种算法的解析光谱与真实光谱吻合度较好,表明ATLD、ACM能够成功地实现混合油类的定性光谱解析。相比之下,ACM较之ATLD分辨所得光谱更接近真实值。

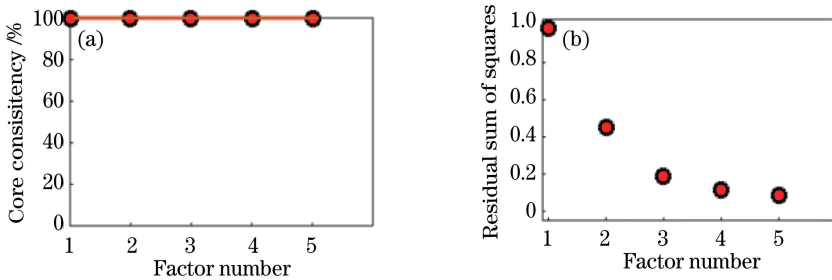


图5 X_1 的核一致诊断结果及残差平方和分析结果

Fig. 5 X_1 nuclear consensus diagnosis results and residual square sum analysis results

选取因子数4来考察各算法对组分数的敏感性,表2~4分别列出了因子数为3和4时各算法解析柴油、汽油、煤油的预测质量浓度和回收率。由于噪声、共线性等影响,可以看出,ATLD与SWATLD解析结果的准确性低于PARAFAC与ACM,但PARAFAC在因子数为4时,出现对组分数敏感的问题,解析出不合理的结果(结果未列出)。

而ACM即使在选取因子数大于实际组分数时,也能给出准确的解析结果,表明ACM对组分数不敏感、解析结果稳定可靠。柴油、汽油、煤油的平均回收率分别为96.68%、97.83%、97.11%,质量浓度预测偏差较小,定量表明ACM分解算法对油类混合物具有良好的分辨率。

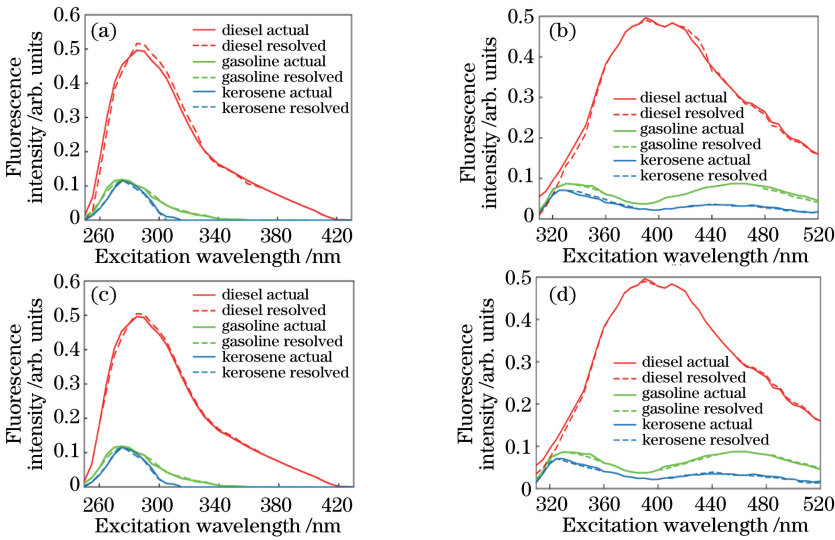


图 6 柴油、汽油和煤油的(a)(b) ATLD 解析光谱与(c)(d) ACM 解析光谱

Fig. 6 (a)(b) ATLD analytical spectra and (c)(d) ACM analytical spectra of diesel, gasoline and kerosene

表 2 不同组分数下各算法解析柴油所得结果

Table 2 Results obtained by each algorithm for analyzing diesel under different component numbers

Sample	Actual	N = 3				N = 4		
		ATLD	SWATLD	PARAFAC	ACM	ATLD	SWATLD	ACM
11	0.3	0.28	0.28	0.29	0.29	0.28	0.28	0.29
12	0.4	0.39	0.39	0.38	0.38	0.37	0.39	0.38
13	0.5	0.48	0.47	0.47	0.47	0.48	0.47	0.47
14	0.6	0.56	0.58	0.59	0.59	0.56	0.58	0.59
15	0.7	0.66	0.67	0.69	0.69	0.66	0.67	0.68
16	0.8	0.78	0.78	0.78	0.78	0.78	0.78	0.78
Average recovery / %		95.33	95.79	96.68	96.68	94.50	95.79	96.44

表 3 不同组分数下各算法解析汽油所得结果

Table 3 Results obtained by each algorithm for analyzing gasoline under different component numbers

Sample	Actual	N = 3				N = 4		
		ATLD	SWATLD	PARAFAC	ACM	ATLD	SWATLD	ACM
11	3.5	3.41	3.36	3.38	3.39	3.41	3.29	3.39
12	2.8	2.69	2.56	2.77	2.76	2.69	2.56	2.76
13	2.2	2.08	2.08	2.10	2.16	2.08	2.08	2.16
14	1.7	1.68	1.64	1.67	1.68	1.68	1.64	1.68
15	1.4	1.38	1.36	1.37	1.38	1.38	1.29	1.38
16	1.0	0.94	0.94	0.95	0.96	0.94	0.94	0.96
Average recovery / %		96.57	94.93	97.01	97.83	96.57	93.76	97.83

表 4 不同组分数下各算法解析煤油所得结果

Table 4 Results obtained by each algorithm for analyzing kerosene under different component numbers

Sample	Actual	N = 3				N = 4		
		ATLD	SWATLD	PARAFAC	ACM	ATLD	SWATLD	ACM
11	1.3	1.27	1.23	1.26	1.24	1.27	1.23	1.24
12	2.0	1.87	1.96	1.97	1.95	1.86	1.94	1.95
13	2.6	2.56	2.54	2.55	2.55	2.56	2.54	2.55
14	3.3	3.27	3.23	3.28	3.26	3.26	3.23	3.26
15	3.9	3.74	3.71	3.75	3.77	3.74	3.71	3.77
16	4.5	4.34	4.34	4.32	4.33	4.34	4.32	4.33
Average recovery / %		96.85	96.63	97.51	97.11	96.72	96.39	97.11

ACM 算法用于分离光谱重叠的油类混合物溶液体现了以下优势:1) 光谱解析结果精度高, 分析目标组分更加可靠;2) 消除较高噪声水平影响, 并保证共线性和组分数的不敏感性;3) 基于数据结构自行判断是否收敛于 ATLD、SWATLD 或 PARAFAC, 使解析过程更加高效;4) 集合 ATLD、SWATLD 和 PARAFAC 算法优势, 具有更高的普适性。

5 结 论

将 0# 柴油、95# 汽油、普通煤油视为环境石油类污染物, 以四氯化碳为溶剂配制不同质量浓度的混合溶液, 并对其进行光谱测量。由于每种石油产品的化学组分存在一定的相似性, 使其光谱有一定的重叠区域甚至极其相似, 给混合溶液光谱的区分带来了困难。结合 ATLD、SWATLD 和 PARAFAC 算法各自的优点, 提出了一种算法组合策略, 通过核一致诊断和残差平方和方法确定因子个数, 利用组合算法 ACM 结合三维荧光光谱实现油种识别和质量分数测量。结果表明, ACM 可以达到预期设计的目的, 并且可以在存在未知干扰的情况下实现对目标分析物的准确定性与定量分析。

参 考 文 献

- [1] Azcarate S M, Teglia C M, Karp F, *et al.* A novel fast quality control strategy for monitoring spoilage on mayonnaise based on modeling second-order front-face fluorescence spectroscopy data [J]. *Microchemical Journal*, 2017, 133: 182-187.
- [2] Wang J, Zhang F, Wang X P, *et al.* Three-dimensional fluorescence characteristics by parallel factor method coupled with self-organizing map and its relationship with water quality[J]. *Acta Optica Sinica*, 2017, 37(7): 0730003.
王娟, 张飞, 王小平, 等. 平行因子法结合自组织映射神经网络的三维荧光特征及其与水质的关系[J]. *光学学报*, 2017, 37(7): 0730003.
- [3] Wang Y T, Liu T T, Liu L F, *et al.* Determination of polycyclic aromatic hydrocarbons in water based on three dimensional fluorescence spectroscopy combined with wavelet compression and APTLD [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(4): 1171-1177.
王玉田, 刘婷婷, 刘凌妃, 等. 基于三维荧光光谱结合小波压缩与 APTLD 对水中多环芳烃测定[J]. *光谱学与光谱分析*, 2018, 38(4): 1171-1177.

- [4] Xu B Y, Ye Y, Ruan R Y, *et al.* Determination of illegal drugs in urine using excitation-emission matrix fluorescence coupled with second-order calibration [J]. *Chemical Journal of Chinese Universities*, 2015, 36(9): 1667-1673.
徐布一, 叶懿, 阮若云, 等. 三维荧光光谱结合二阶校正算法用于尿液中常见毒品的检测[J]. *高等学校化学学报*, 2015, 36(9): 1667-1673.
- [5] Wu H L, Li Y, Kang C, *et al.* Research progress of three-dimensional fluorescence coupled with chemical multiway calibration [J]. *Chinese Journal of Analytical Chemistry*, 2015, 43(11): 1629-1637.
吴海龙, 李勇, 康超, 等. 三维荧光化学多维校正方法研究新进展[J]. *分析化学*, 2015, 43(11): 1629-1637.
- [6] Tao C X, Ruan J, Shu S P, *et al.* Detection of fried oil in edible oil based on three-dimensional fluorescence spectroscopy [J]. *Chinese Journal of Lasers*, 2016, 43(1): 0115001.
陶春先, 阮俊, 舒顺朋, 等. 三维荧光光谱对食用植物油中煎炸油的检测[J]. *中国激光*, 2016, 43(1): 0115001.
- [7] Kong D M, Zhang C X, Cui Y Y, *et al.* Detection of oil species in mixed oil based on alternating penalty trilinear decomposition [J]. *Acta Optica Sinica*, 2018, 38(11): 1130005.
孔德明, 张春祥, 崔耀耀, 等. 基于交替惩罚三线性分解的混合油液油种成分的检测[J]. *光学学报*, 2018, 38(11): 1130005.
- [8] Bro R. PARAFAC. Tutorial and applications [J]. *Chemometrics and Intelligent Laboratory Systems*, 1997, 38(2): 149-171.
- [9] Wu H L, Shibukawa M, Oguma K. An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons [J]. *Journal of Chemometrics*, 1998, 12(1): 1-26.
- [10] Wang Y T, Liu L F, Wang S T, *et al.* Fluorescence spectra of aromatic hydrocarbons based on alternating trilinear decomposition [J]. *Laser & Optoelectronics Progress*, 2018, 55(9): 093001.
王玉田, 刘凌妃, 王书涛, 等. 基于交替三线性分解的芳烃类化合物荧光光谱[J]. *激光与光电子学进展*, 2018, 55(9): 093001.
- [11] Chen Z P, Wu H L, Yu R Q. On the self-weighted alternating trilinear decomposition algorithm—the property of being insensitive to excess factors used in

- calculation[J]. *Journal of Chemometrics*, 2001, 15(5): 439-453.
- [12] Yu Y J. Researches on theory and application of second-order calibration method and its development of application software [D]. Changsha: Hunan University, 2012.
于永杰. 化学计量学二阶校正理论及应用基础研究与实用软件开发[D]. 长沙: 湖南大学, 2012.
- [13] Bro R, Kiers H A L. A new efficient method for determining the number of components in PARAFAC models[J]. *Journal of Chemometrics*, 2003, 17(5): 274-286.
- [14] Maggio R M, de la Peña A M, Olivieri A C. Unfolded partial least-squares with residual quadrilinearization: a new multivariate algorithm for processing five-way data achieving the second-order advantage. Application to fourth-order excitation-emission-kinetic-pH fluorescence analytical data[J]. *Chemometrics and Intelligent Laboratory Systems*, 2011, 109(2): 178-185.
- [15] Du W. Missing data recovery and alternating residual trilinearization applied to multivariate calibration and calibration transfer [D]. Changsha: Hunan University, 2016.
杜文. 多元校正及模型转移中的缺损数据重构和交替残差多线性方法研究[D]. 长沙: 湖南大学, 2016.