

基于三维荧光光谱降维处理的矿物油识别研究

王玉田 徐 婧* 周艳岭

(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 以汽油、煤油和柴油为研究对象,并将其作为整体研究,不考虑其中具体组分,利用荧光光谱分析技术,提出一种通过将光谱数据矩阵首尾相连,求取其包络线,利用其表观统计量(选用标准偏差、峰度系数和偏度系数)将三维光谱降为二维光谱的方法,该方法对于光谱有一定程度重叠但光谱形状相似性较低的体系有较好的适用性。结合聚类方法,对油种进行识别,识别率达 98%,依据朗伯比尔定律,测定三种油类的荧光强度与浓度的拟合曲线,并对其进行定量分析,加标回收率可达 95% 以上。并与目前常用的平行因子法进行分辨效率的比较,在计算时间上提高了 43%。

关键词 光谱学;光谱降维;表观统计量;包络线;聚类分析

中图分类号 O436 文献标识码 A doi: 10.3788/CJL201441.1215002

Study on Mineral Oil Identification Based on a Dimension Reduction Method of Three Dimensional Fluorescence Spectra

Wang Yutian Xu Jing Zhou Yanling

(Measurement Technology and Instrumentation Key Laboratory of Hebei Province, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract Kerosene, diesel and gasoline are taken as research objects without considering each oil's components. By fluorescence spectroscopy analysis, a dimension reduction method that turns three-dimensional spectrum into two-dimensional spectrum is proposed by linking the spectral datum matrixes end to end, drawing its envelope and extracting the apparent statistic (the standard deviation, coefficient of kurtosis and coefficient of skewness are chosen). The method is applical to the systems of certain degree of spectral overlap but low spectral shape similarity. The recognition rate of the oils is over 98% with clustering analysis. According to the Lambert-Beer law, fitting curve of fluorescence intensity and concentration is used for quantitative analysis and standard addition recovery rate is over 95%. Distinguish efficiency is increased by 43%, compared with the parallel factor method used commonly.

Key words spectroscopy; spectral dimension reduction; apparent statistic; envelope; clustering analysis

OCIS codes 300.2530; 300.6170; 300.6280; 300.6390; 300.6540

1 引 言

荧光法灵敏快速的特性,常用于物质特性的研究^[1]。油类污染日益严重,危害人们的生活健康,海上溢油事件时有发生,对溢油种类的检测十分重要。由于矿物油中含有的芳香烃类物质使其可以受激发产生荧光光谱,同时由于不同种单一矿物油所含芳

香烃种类和含量有所不同,又使不同矿物油的荧光光谱有所不同,这是通过荧光法鉴别矿物油种类的理论基础。近年来许多学者利用荧光法对水中油类及其他污染物质进行了研究^[2-11]。本文通过绘制包络线并提取特征统计量,对油种进行识别,在运算时间上得到了改善。

收稿日期: 2014-05-23; 收到修改稿日期: 2014-07-02

基金项目: 国家自然科学基金(61201110,61071202)、河北省自然科学基金(F2010001312)

作者简介: 王玉田(1952—),男,博士,教授,主要从事光电检测、光纤传感和荧光分析法等方面的研究。

E-mail: y. t. wang@163. com

* 通信联系人。E-mail: ysuxujing@163. com

本文电子版彩色效果请详见中国光学期刊网 www. opticsjournal. net

2 主要研究内容

2.1 测量原理

三维荧光光谱(EEM)的常用表达形式为激发-发射-光强光谱矩阵形式^[12],其表达式为

$$[I] = aYX = [Y]_{I \times N} [X]_{N \times J}, \quad (1)$$

式中 $[I]$ 为荧光的光谱矩阵, a 是只与浓度有关的系数, X 为荧光的发射光谱, Y 为荧光的激发光谱, N 为组分数, I 为激发波长数, J 为发射波长数。

通过在发射谱方向上,将每个发射点对应的激发点列进行首尾相连,组成降维后的激发光谱,为了提高数据处理的速度,将数据进行离散化处理,并求取其包络线,提取表观统计量,用以对不同油种进行测定。包络线包含着荧光峰值的形状及走势,这正是对不同油种进行区分的理论依据,因此选用表征曲线所含信息量多少的标准偏差、表征数据分布对称性的峰度系数 k 和表征分布曲线陡缓程度的偏度系数 s 作为其特征统计量。

$$k = \frac{E(x - \mu)^3}{\sigma^3} \mu, \quad (2)$$

$$s = \frac{E(x - \mu)^4}{\sigma^4}, \quad (3)$$

式中 E 为数学期望, σ 为标准差, μ 为均值, x 为样本点取值。

2.2 实验部分

实验样品及器具主要有:取自秦皇岛某加油站的 0# 柴油、煤油和 97# 汽油,四氯化碳溶剂,四面透光石英比色皿 4 个,用于稀释缓冲溶液的 200 mL 广口杯 2 个,用于存放对应浓度溶液的玻璃锥形瓶 5 个(1 L),10 mL 注射器 3 只,5 mL 取液管 2 根,容量为 50、10 mL 的量筒各 1 个。

实验中配置油样与四氯化碳体积比为 1 : 1000 的标准母液,然后依次稀释得到不同体积分数的矿物油样本,依次稀释为体积分数为 85×10^{-6} , 65×10^{-6} , 45×10^{-6} , 25×10^{-6} , 5×10^{-6} , 2.5×10^{-6} , 1.25×10^{-6} , 0.625×10^{-6} 。

实验搭建的系统框图如图 1 所示,通过中央处理器(CPU)与控制电路来控制光源电源,调节电压及稳压控制,光源选取功率较大的脉冲氙灯,发出的复色光经激发单色器色散一部分进入参比电路,以排除光源波动带来的影响,另一部分进入样品池,对待测样品进行荧光激发,荧光通过发色单色器进入测量电路,进行信号检测,最终将信号输入上位机进行处理。具体仪器参数:激发光源功率为 450 W 的脉冲氙灯,激发单色器和发射单色器选用

典型 Czerny-Turner 对称结构设计的 M300 型光栅单色器,分辨率可达 0.05~0.1 nm,波长重复性可达 ± 0.1 nm,探测器选用 1P28 型光电倍增管。

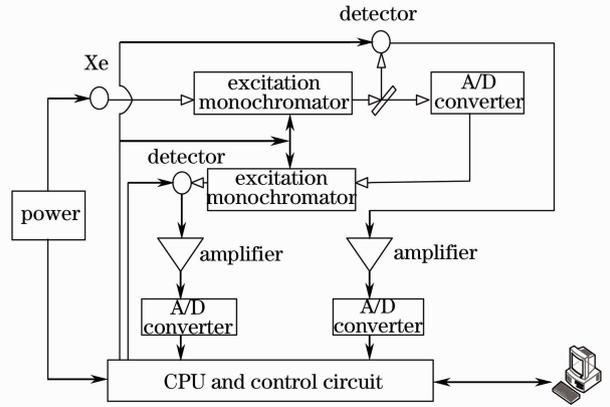


图 1 荧光光谱仪系统结构图

Fig. 1 Fluorescence spectrometer system structure

实验时为保证光源使用寿命,需等到冷却装置降温到 -20 °C 再开始实验,同时设定采样间隔为 0.1 nm,狭缝宽度为 0.25 mm,此时分辨率可达 1.0 nm,激发波长 (E_x, λ_{E_x}) 扫描范围为 250 ~ 500 nm,步长为 10 nm;考虑到杂散光的影响,选择发射波长 (E_m, λ_{E_m}) 扫描范围 $\lambda_{E_x} + 10$ nm ~ 500 nm,步长为 2 nm,可以较好的避免散射光的影响。

3 数据处理及结果分析

对数据进行标准化、离散化处理[下面均以柴油的标准化离散激发发射矩阵(EEM)图像为例进行说明],如图 2 所示。

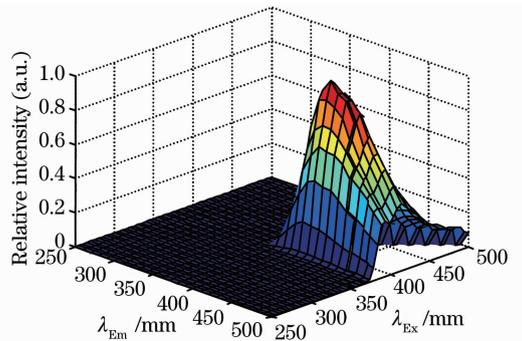


图 2 柴油的标准化离散 EEM 图像

Fig. 2 Diesel standardized discrete EEM image

对其进行降维处理,即将每个发射点对应的激发点列首尾相连,形成降维后的激发矩阵,如图 3 所示。对其求取包络线,如图 4 所示。

利用同样的方法求取汽油和煤油二维激发图像的包络线图像如图 5 和图 6 所示。

利用绘制包络线的方法,对 27 组待测油样进行

表观统计量的提取, 并进行标准化处理, 其结果如表 1 所示。

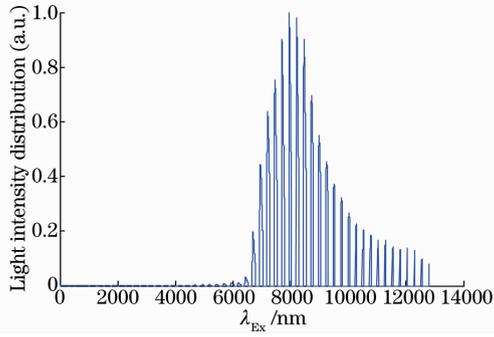


图 3 柴油的二维激光图像

Fig. 3 Two-dimensional excitation image of diesel

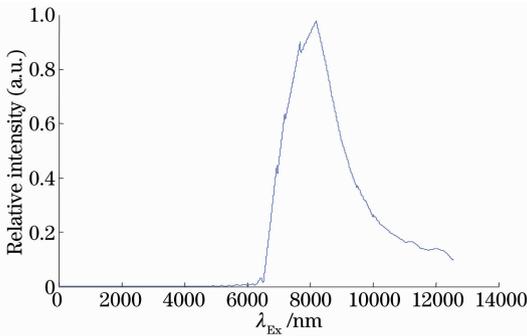


图 4 柴油的二维激发图像的包络线

Fig. 4 Envelope of diesel two-dimensional excitation image

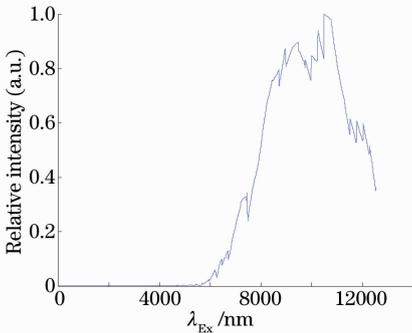


图 5 汽油的二维激发图像的包络线

Fig. 5 Envelope of gasoline two-dimensional

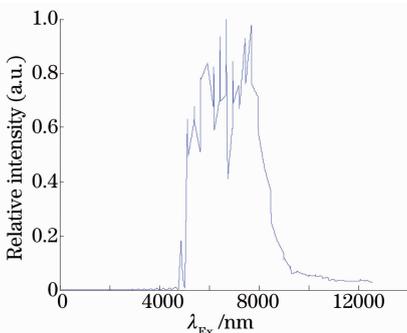


图 6 煤油的二维激发图像的包络线

Fig. 6 Envelope of kerosene two-dimensional excitation image

表 1 标准化的提取特征值列表

Table 1 Standardized characteristic value extraction table

Oil type	Number	Standard deviation	Kurtosis coefficient	Skewness coefficient
Gasoline	1	0.9143	-0.9924	-1.1768
	2	2.0693	-0.8972	-1.1424
	3	-0.2406	-1.0875	-1.2112
	4	1.3083	-0.9349	-0.9171
	5	0.5204	-1.0498	-1.4364
	6	2.2819	-0.9758	-0.9136
	7	-0.4418	-1.0433	-1.3936
	8	-0.4532	-1.0090	-1.4400
Kerosene	9	-0.1046	-0.3200	0.0507
	10	-0.3058	-0.3643	-0.1975
	11	-0.0550	-0.2309	0.3036
	12	-1.0668	-0.3996	-0.1509
	13	1.0178	-0.2193	0.2283
	14	-0.1556	-0.4093	-0.2024
	15	0.0953	-0.2758	0.2987
	16	-1.2283	-0.4208	-0.1271
Diesel	17	-0.5722	1.3372	1.3777
	18	-0.6374	1.3270	1.1547
	19	-1.5996	1.2474	0.9530
	20	-0.0819	1.3958	1.2010
	21	-1.1943	1.2581	1.1080
	22	0.3234	1.4065	1.3560
	23	-1.6393	1.3236	1.0796
	24	0.3631	1.3302	1.2295
Unknown	25	-0.7040	1.3167	0.9313
	26	0.8562	-0.2405	0.2521
	27	0.7301	-1.0718	-1.2153

将标准化后的数据利用欧氏距离 d_{ij} 作为标准进行最短距离聚类, 具体步骤为: 1) 计算样品间两两距离, 距离最接近的为为一类; 2) 计算类间距离, 类间距离最短两者合并, 至最终聚为一类。

$$d_{ij} = \left[\sum_{a=1}^3 (x_{ia} - x_{ja})^2 \right]^{1/2}, \quad (4)$$

式中 d_{ij} 为第 i 样品与第 j 样品间的距离, $a = 1, 2, 3$ 分别表示三个表观统计量, x_{ia}, x_{ja} 分别为样本 i , 样本 j 对应相应表观统计量的标准化值。通过选取的表观统计量进行聚类分析, 其结果如图 7 所示。

由图 7 可知, 聚类效果良好, 各油种均可被区分, 未知油种也被分到三组油中分别属于柴油煤油和汽油。

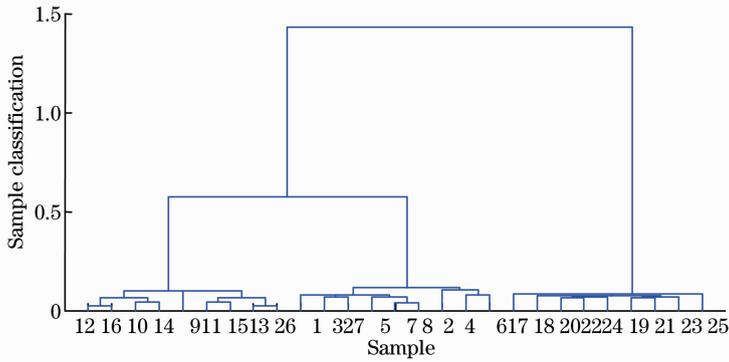


图 7 特征值的聚类分析

Fig. 7 Characteristic value clustering analysis

下面与常用的平行因子方法进行比较。利用平行因子法得到的残差平方和及核一致测试结果如图 8 所示,由图 8 可知,当选择因子数为 2 时,即可满足测量要求。平行因子模型激发波长分析及实测结果对比图如图 9 所示,图 10 为平行因子模型发射波长及实测结果对比图,可以看出模型效果良好。

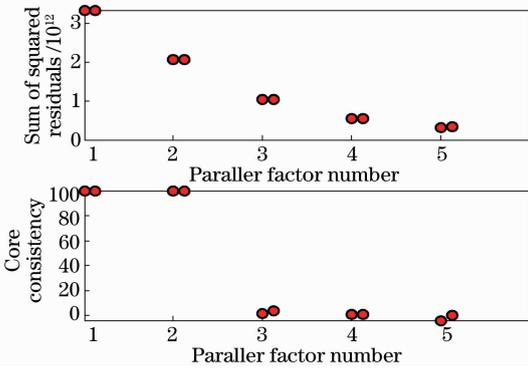


图 8 残差平方和及核一致诊断结果图

Fig. 8 Sum of squared residuals and core consistent diagnosis figure

经降维处理后的特征提取及聚类分析结果和常用的平行因子分析结果鉴别能力都可以达到 98% 以上。由表 2 可知,在分辨效率的比较中,实验中所采用方法有明显的优势,运算时间仅为平行因子方法的 58.7%,更适于对数据处理速度有更高要求的现场实时处理。

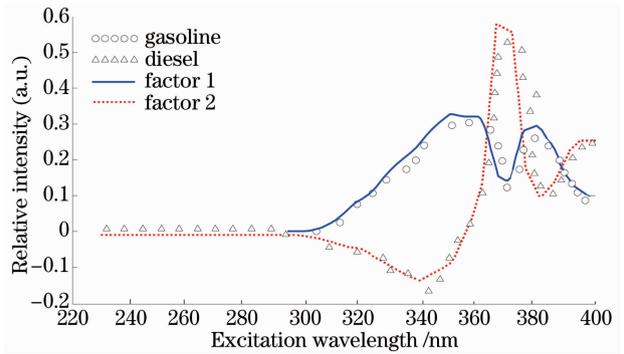


图 9 平行因子模型激发波长分析及实测结果对比图

Fig. 9 Contrast figure of parallel factor model excitation wavelength analysis and measured results

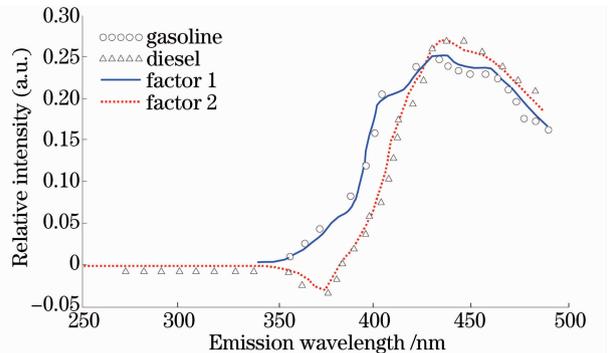


图 10 平行因子模型发射波长及实测结果对比图

Fig. 10 Contrast figure of parallel factor model emission wavelength and measured results

表 2 本文算法与平行因子分析耗时对比

Table 2 Time comparison of proposed method and parallel factor

Clustering analysis	Time /s	Parallel factor	Time /s
Three-dimensional image display	1.45	Three-dimensional image display	1.45
Discrete and standardization	0.45	Analysis of nuclear consistency and residual sum of squares	1.35
Dimension reduction	0.3	Parallel factor analysis excitation spectrum	1.24
Feature extraction	0.3	Parallel factor analysis emission spectrum	1.24
Clustering analysis	0.5	—	—
Total	3.0	Total	5.28

依据朗伯-比尔定律,在浓度较低可进行近似线性化的范围内,对汽油、煤油和柴油的荧光光强 y 与相应体积分数 C 进行线性拟合,其拟合曲线分别为

柴油: $y = 33.5783C + 1235.3728$,
 煤油: $y = 49.9785C + 1842.5341$,
 汽油: $y = 66.6737C + 595.6443$.

经验证其线性相关度均大于 95%,有良好的线性关系,三种矿物油的体积分数与荧光光强的拟合线如图 11 所示。

测定含有 5 mL 油样河水的加标回收率,其结果如表 3 所示。

由表 3 可知,加标回收率可达 95% 以上,证明本文所采用方法可以用于较准确地进行河水中矿物油的检测。式中 C_1 和 C_2 分别为试样浓度和加标试

样浓度。

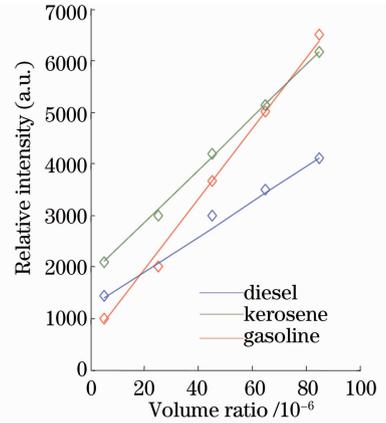


图 11 三种矿物油的浓度拟合线

Fig. 11 Concentration fitting lines of the three kinds of mineral oil

表 3 柴油、汽油和煤油的回收率

Table 3 Recovery rate of diesel, gasoline and kerosene

Sample name	Fortified volume fraction / 10^{-6}	$C_2 - C_1$ determination results / 10^{-6}	Recovery rate / %
Diesel	5.00	2.45, 2.40, 2.45, 2.35, 2.55, 2.45	97.67
Gasoline	5.00	2.44, 2.46, 2.45, 2.43, 2.47, 2.46	98.06
Kerosene	5.00	2.42, 2.45, 2.45, 2.41, 2.44, 2.46	97.53

4 结 论

对汽油、煤油和柴油进行定性和定量测量,提出一种通过将光谱数据矩阵首尾相连,将三维光谱降维二维光谱的方法,经过离散化和求取包络线处理后,对其进行特征提取,选取标准偏差,峰度系数,偏度系数作为表观特征统计量,结合聚类分析对样品的识别率达 98%,加标回收率可达 95% 以上,与广泛应用的平行因子方法相比,在保证达到同样的样品识别率的同时,运算时间得到了很大的提高,分辨效率提高了 43%。适于满足对运算时间有严格要求的场合使用。该方法对于汽油、煤油和柴油等光谱有一定的重叠,但光谱线形状有较大差异的体系适用性较好,但对于光谱重叠严重,且光谱线形状相似的体系将不再适用。对于光谱线形状相似但光谱重叠较小的体系,可在不同的光谱范围内应用此法,可达较好的效果。

参 考 文 献

1 Yan Suyuan, Zhang Weigang. Fluorescence polarization properties and detection of vitamin B2 solid powder[J]. Chinese J Lasers, 2012, 39(6): 0615001.
 严肃源, 张伟刚. 维生素 B2 固体粉末荧光偏振特性及其检测[J]. 中国激光, 2012, 39(6): 0615001.
 2 Li J, Fuller S, Cattle J. Matching fluorescence spectra of oil spills with spectra from suspect sources[J]. Analytica Chimica

Acta, 2004, 514(1): 51-56.
 3 Deepa S, Sarathi R, Mishra A K. Synchronous fluorescence and excitation emission characteristics of transformer oil ageing[J]. Talanta, 2006, 70(4): 811-817.
 4 Bugden J B C, Yeung C W, Kepkay P E, et al.. Application of ultraviolet fluorometry and excitation-emission matrix spectroscopy (EEMS) to fingerprint oil and chemically dispersed oil in seawater[J]. Marine Pollution Bulletin, 2008, 56(4): 677-685.
 5 Christensen J H, Tomasi G. Practical aspects of chemometrics for oil spill fingerprinting[J]. Journal of Chromatography A, 2007, 1169(12): 1-22.
 6 Smith C G, Sinski J F. Red-shift cascade: investigations into the concentration dependent wavelength shifts in 3-dimensional fluorescence spectra of petroleum samples[J]. Appl Spectrosc, 1999, 53(11): 1459-1469.
 7 Wang C, Li W, Luan X. Species identification and concentration quantification of crude oil samples in petroleum exploration using the concentration-synchronous-matrix-fluorescence spectroscopy [J]. Talanta, 2010, 81(1-2): 684-691.
 8 Divya O, Mishra A K. Understanding the concept of concentration-dependent red-shift in synchronous fluorescence spectra: Prediction of and optimization of λ_{SFS}^{max} and optimization of $\Delta\lambda$ for synchronous fluorescence scan[J]. Anal Chim Acta, 2008, 630(1): 47-56.
 9 Patra D, Mishra A K. Concentration dependent red shift: qualitative and quantitative investigation of motor oils by synchronous fluorescence scan[J]. Talanta, 2001, 53(4): 783-790.
 10 Yang Lili, Wang Yutian, Lu Xinqiong. Identification and measurement of petroleum pollutant by three-dimensional matrix fluorescence with second-order calibration methods[J]. Chinese J Lasers, 2013, 40(6): 0615002.
 杨丽丽, 王玉田, 鲁信琼. 三维荧光光谱结合二阶校正法用于石

- 油类污染物的识别和检测[J]. 中国激光, 2013, 40(6): 0615002.
- 11 Liu Jing, Liu Wenqing, Zhao Nanjing, *et al.*. Phytoplankton chlorophyll II fluorescence characteristics excited by various light qualities and intensities[J]. Acta Optica Sinica, 2013, 33(9): 0930001.
- 刘 晶, 刘文清, 赵南京, 等. 浮游植物在不同光质和光强激发下的叶绿素荧光特性[J]. 光学学报, 2013, 33(9): 0930001.
- 12 Wang Yanni, Fan Yangyu. The adaptive prediction algorithm of improved BP neural network[J]. Computer Engineering and Application, 2010, 46(17): 23-26.
- 王燕妮, 樊养余. 改进 BP 神经网络的自适应预测算法[J]. 计算机工程与应用, 2010, 46(17): 23-26.

栏目编辑: 史 敏