

【技术与方法】

近红外光谱技术结合主成分聚类分析 判别海面溢油种类

王 丽^{1,2}, 何 鹰^{1,2}, 王颜萍^{1,2}, 赵 英¹, 李 伟¹, 王小如², Frank Lee²

(1. 厦门大学 分析化学教育部重点实验室, 福建 厦门 361005; 2. 国家海洋局第一海洋研究所, 山东 青岛 266061)

摘要: 鉴于目前没有一种方法能独立解决溢油鉴别的所有问题, 本文提出了近红外光谱技术结合主成分聚类分析鉴别溢油种类的方法。通过有机溶剂萃取出自行配制的汽油、柴油和润滑油模拟样品中的溢油后记录其近红外光谱, 对 5 800 ~ 6 200 cm^{-1} 区段范围内的谱图经多元散射校正 (MSC)、Norris 一阶导数平滑预处理处理后求其主成分, 并在主成分的基础上引入 Ward 聚类分析法 (离差平方和法) 对样品分类。结果表明近红外光谱技术结合聚类分析能对体积分数在 0.4 ~ 0.8 mL/L 间的面溢油样品正确、快速分类, 近红外光谱技术结合主成分聚类可作为溢油鉴别的一种辅助方法。

关键词: 近红外; 主成分聚类分析; 溢油鉴别

中图分类号: X55 文献标识码: A 文章编号: 1007-6336(2004)02-0058-03

Oil Spill identification by near-infrared spectroscopy coupled with clustering analysis based on principle components

WANG Li^{1,2}, HE Ying^{1,2}, WANG Yan-ping^{1,2}, ZHAO Ying¹, LI Wei¹, WANG Xiao-ru², FRANK Lee²

(1. Department of Chemistry, the Key Laboratory of Analytical Science of MOE, Xiamen University, Xiamen 361005, China; 2. First Institute of Oceanography, State Oceanic Administration, Qingdao 266061, China)

Abstract: Owing to none of the existing method can do it all for oil identification, but a method by near-infrared spectroscopy (NIR) coupled with clustering based on principle components is proposed. Petroleum oil stimulating samples of gasoline, diesel fuel and lubricating oil were employed to develop the method. NIR spectra ranges of 5800 ~ 6200 cm^{-1} were recorded with the multiplicative signal correction and Norris first derivative filter pretreatments. The clustering analysis by Ward's method base on the different principle components was discussed. The results show that the method provides the rapid and accurate classification for spilled oil with concentration between 0.4 ~ 0.8 $\mu\text{L/L}$.

Key words: near-infrared spectroscopy; clustering analysis; principle components; spill oil identification

溢油一旦进入水体便迅速在水面扩散开来, 溢油的种类不同对水体的污染程度不同, 其相应的处理措施也不同。漂浮在水面的溢油不易肉眼分辨出种类。现有溢油鉴别方法包括中红外法、紫外光谱法、荧光光谱法、气相色谱法以及气质联用法等, 由于石油成分的复杂性, 还没有一种方法能独立快速解决溢油鉴别的问题^[1]。我国行业标准^[2] (海面溢油鉴别系统规范 HY 043-1997) 规定将气相色谱法、中红外法、荧光光谱法三种方法作为溢油鉴别的标准方法, 并规定了三种基本方法同时进行鉴别和三种基本方法的结果一致性鉴别原则, 即出现一种方法的结果与其他两种方法的结果不一致时,

不能论断, 而应采用辅助鉴别法。

鉴于上述原因, 发展一种快速鉴别溢油种类的方法是很必要的。近红外光谱技术是一种快速无损分析技术, 该技术结合聚类分析快速判断溢油种类作为溢油鉴别的辅助手段还报道甚少。本文作者自行配制了汽油、柴油和润滑油的海面溢油模拟样品, 有机溶剂萃取出样品中的溢油后记录其近红外光谱, 对 5 800 ~ 6 200 cm^{-1} 区段范围内的谱图经多元散射校正 (MSC)、Norris 一阶导数平滑预处理处理后求其主成分, 并在主成分的基础上引入了 Ward (离差平方和法) 聚类分析判别溢油的种类。结果令人满意。

收稿日期: 2003-07-08, 修改稿收到日期: 2003-11-19

基金项目: 国家高技术发展计划 (863) 资助项目 (2001AA635040)

作者简介: 王 丽 (1973-), 湖北武汉人, 在读硕士, 主要研究海洋环境监测技术研究。通读联系人: 王小如

1 材料与方 法

1.1 仪器与试剂

AVATAR 360N E. S. P 近红外光谱仪,透射测样附件(美国尼高力公司);奔 微机;0.8 cm 玻璃样品瓶; CCl_4 (分析纯)。

1.2 样品及制备

实验在 2003 年 2~4 份间每周从厦门大学加油站采集商品汽油、柴油、润滑油(不同种类的油一次只采一份样品,没有原油商品出售,因此本文没有采用原油作为研究对象),海水分三次采集于厦门白城海域。分别移取 200,300,400 μL 油到 500 mL 海水中,制得 42 个模拟样品。油的体积分数为 0.4、0.6、0.8 mL/L,充分摇匀使模拟油在海水中分散,静置 10 min 后油水分层,弃去下层海水约 495 mL 时止,用 2 mL CCl_4 萃取剩余油水混合物,取 CCl_4 层记录其近红外光谱图。

1.3 操作条件

近红外光谱图测定采用 OMNIC 软件(美国尼高力公司),测定范围 $4\ 000\sim 10\ 000\ \text{cm}^{-1}$,扫描次数 40 次,分辨率 $8\ \text{cm}^{-1}$,每个样品取五次测量的平均值。调用 TQ 软件^[3](美国尼高力公司)进行近红外谱图主成分分析,STATISTICA 软件进行聚类分析。

2 结果与讨论

2.1 聚类分析数据范围的选择

近红外光谱(图 1,a)记录了 $4\ 000\sim 10\ 000\ \text{cm}^{-1}$ 区域内的汽油、柴油和润滑油的吸收曲线,从图中可看出,三种油在近红外区域的响应模式是相同的,但在 $5\ 800\sim 6\ 200\ \text{cm}^{-1}$ 区域有明显不同,正是这些不同才使聚类分析成为可能,此区域以外的光谱对于表征样品的作用微小。为了减少计算量,仅选择该区域作为以下聚类分析的数据范围。

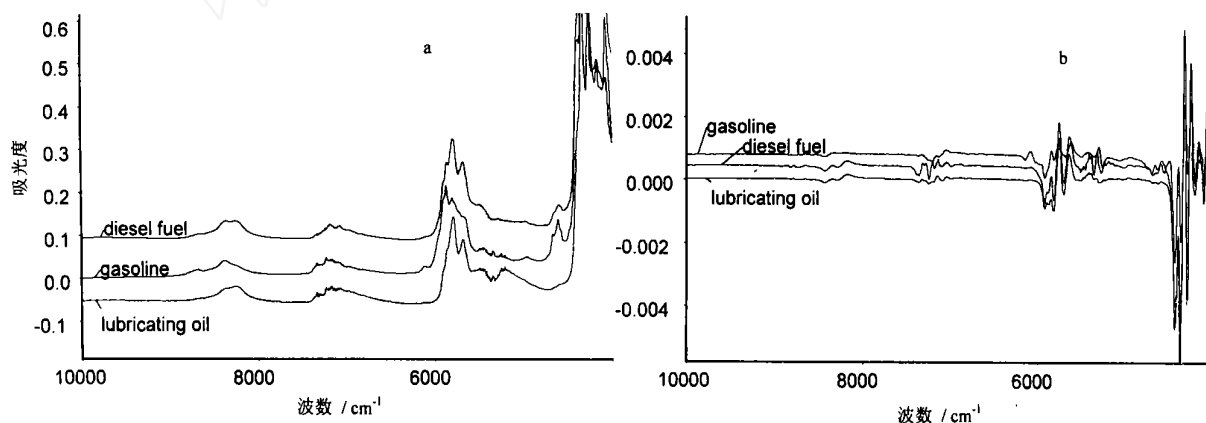


图 1 海水中汽油、柴油、润滑油溢油萃取液近红外图谱

(a: 原始光谱;b: 原始光谱经 MSC 和 Norris 一阶导数平滑)

Fig. 1 NIR spectra of gasoline, diesel fuel and lubricating oil extracted from sea water

(a: spectra without pretreatment ;b: spectra after pretreatment of MSC and Norris first derivate)

2.2 主成分聚类

主成分分析是一种数据降维方法,它将原变量进行转换,使少数几个新变量成为原变量的线性组合,这些新变量互相正交,尽可能多地表征了原变量的数据结构特征而消除了各变量间的相关性。因此,在主成分的基础上引入聚类分析可排除众多化学信息中相互重叠的部分而不丢失原始数据信息,使聚类分析的计算简化。

42 个样本的数据(其中 1~11 号样为柴油,12~26 号样为汽油,27~42 号样为润滑油,每种油的每个样品采集时间不同)。调用 TQ 软件将 $5\ 800\sim$

$6\ 200\ \text{cm}^{-1}$ 区域的近红外光谱,进行 MSC 处理后,算得主成分为 7 时,特征向量本征值累积达 99.9% 以上,将光谱在 MSC 处理的基础上,进一步进行 Norris 一阶导数平滑处理(见图 1,b),这时主成分为 4,特征向量本征值已达到 99.9% 以上。因此,对原始数据预处理的不同,应选取的主成分数也不相同。

采用 Ward 离差平方和法对 42 个样本经不同方法处理后的数据,即原始数据经 Norris 一阶导数平滑处理和未经 Norris 一阶导数平滑处理数据分别进行主成分聚类分析,聚类结果见图(2,3)。图 2 表明当欧氏距离在 10 000 以上的水平上,三

类油完全正确分开,聚类正确率 100%;而图 3 中, 中的干扰信息造成的。可见通过正确选择近红外当欧氏距离为 $1e^6$ 以上时,所有样本也分为三类, 光谱信息区域,经求导数平滑后,能消除干扰信息,但分类部分交叉,分类正确率不到 80%,这是数据 使分类结果准确性提高。

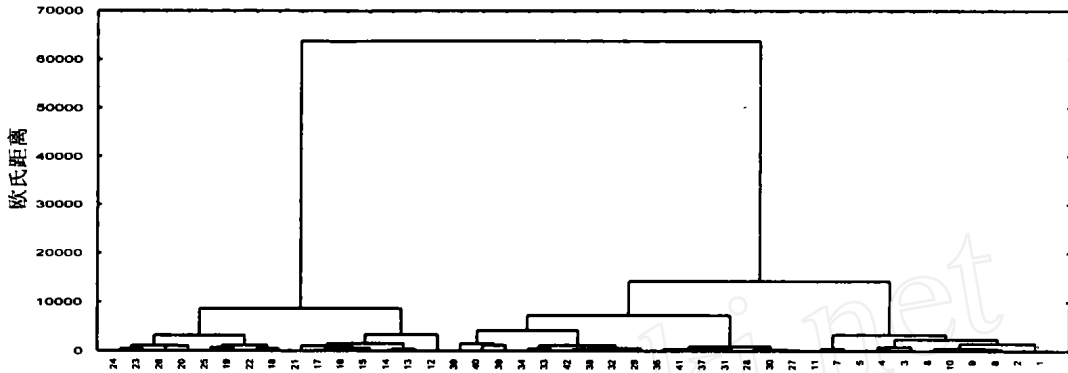


图 2 42 个样本经 Norris 一阶导数平滑处理的溢油样品的主成分聚类图(Ward's method)

Fig. 2 Clustering graph of 42 oil spill samples with pretreatment of Norris first derivate

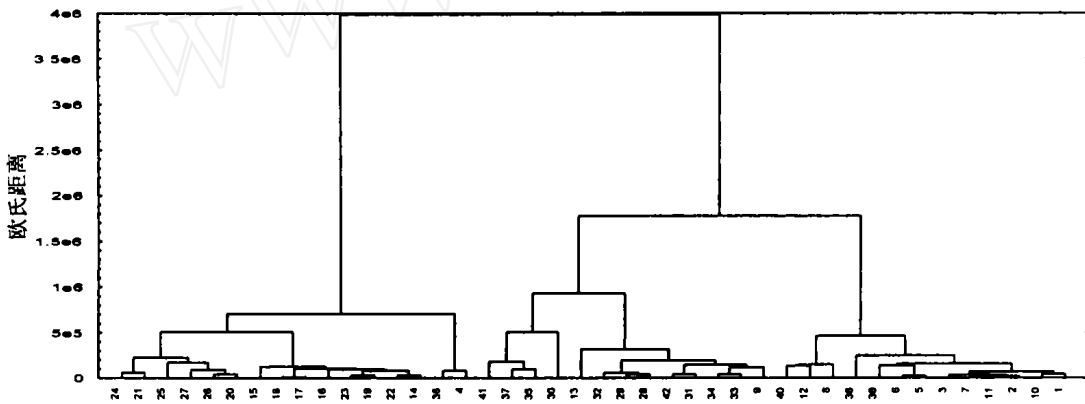


图 3 42 个样本未经 Norris 一阶导数平滑处理的溢油样品主成分聚类图(Ward's method)

Fig. 3 Clustering graph of 42 oil spill samples without pretreatment of Norris first derivate

3 结 语

通过以上工作,可得到以下结论:

- (1) 近红外光谱技术结合主成分聚类分析可用于体积分数在 0.4 ~ 0.8 μ L/mL 间的溢油的鉴别。近红外光谱技术结合主成分聚类可作为溢油鉴别的一种辅助方法;
- (2) 在主成分的基础上引入聚类分析,可以提取原始数据的非相关变量而不丢失信息,使聚类分析计算简化;
- (3) 通过选择适当的近红外光谱区域和光谱

预处理技术,如求导,平滑,可消除干扰信息,提高分类结果的准确性。

参考文献:

- [1] WANG Z D, FINGAS M. Developments in the analysis of petroleum hydrocarbons in oils, petroleum products and oil-spill-related environmental samples by gas chromatography [J]. Journal of chromatography: A, 1997, 774: 51-78.
- [2] HY 043-1997,海面溢油鉴别系统规范[S].
- [3] THERMO N. TQ analyst user's guide[M]. Madison: Thermo Nicolet Corporation, 1992. 27-69.