

Euclid 贴近度聚类分析鉴别海面溢油

徐恒振 尚龙生 周传光 刘 彤

(国家海洋局海洋环境保护研究所, 大连 116023)

提 要 于 1986 年对我国常见的 14 种油品和 1984 年 8 月大连湾海面溢油进行气相色谱分析, 用模糊数学方法建立 Euclid 贴近度聚类分析的海面溢油鉴别模式

$$r_{ij} = 1 - \sqrt{\frac{1}{m} \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]}$$

提出油种鉴别的 Euclid 贴近度置信水平 λ_i 的模糊区概念。油种及其海面溢油的 Euclid 贴近度聚类分析动态聚类图表明, 油种鉴别的置信水平 λ_i 主要受风化和实验误差的影响, 大庆原油、30# 重柴油、20# 重柴油、0# 轻柴油以及任丘原油各自分别同其它 13 种油可分类的置信水平 λ_i 的范围依次为: $[0.8726, 0.9953)$, $[0.8726, 0.9970)$, $[0.8726, 0.9857)$, $[0.8807, 0.9902)$ 和 $[0.8726, 0.9907)$ 。

关键词 Euclid 贴近度 模糊聚类分析 鉴别 溢油

在海面溢油鉴别的模糊数学法中, 徐恒振等(1987, 1990, 1992)曾应用模糊传递闭包法、模糊贴近度法和模糊最大矩阵元法较好地表征了海面溢油鉴别所具有的模糊性。由于传递闭包法和模糊贴近度法尚有遗失数据信息的缺陷, 而最大矩阵元法有聚类简便的特点, 对此, 本文用 Euclid 贴近度作为矩阵元, 用主对角线除外的最大矩阵元作为置信水平的聚类分析法, 以探讨海面溢油鉴别模式。

1 材料和方法

1.1 风化实验、油品制备及色谱条件 于 1986 年 8 月, 在 3 个盛有 5 000ml 洁净海水的烧杯中分别加入 30# 重柴油、0# 轻柴油(采自大连石油七厂质检科)和任丘原油(采自任丘油田任二联北大站), 形成约 3mm 厚的油膜, 置于通风向阳处风化一个月, 在风化 1, 5, 10, 15, 21 和 30d 时取样(下称风化油样)。期间多为晴天, 平均水温 33℃, 平均气温 28℃。

称取 3 种风化油样和 14 种未风化油样(表 1) 以及溢油样(采自 1984 年 8 月 23 日大连湾发生的涉及马鲁古海外轮和我国向阳号轮船的排油事件区域, 其中马鲁古海和向阳号两个轮船可疑油样分别为 1# 和 2#, 海面溢油样为 3#) 各 0.50g, 置于 10ml 离心管内, 分别加入 5ml 正己烷(A.R.) 振荡; 溶解后加 1g Na₂SO₄(A.R., 烘 3h), 振荡后离心分离上部溶液, 待色谱分析。用 GC-7A 气相色谱仪, 0.2mm × 45m 甲基乙烯基硅酮交联玻璃毛细管柱分析。氢焰检测器温度为 300℃, 初温 75℃(4min), 终温 280℃(16min), 升

* 自选课题。徐恒振, 男, 出生于 1957 年 2 月, 副研究员。

收稿日期: 1993 年 12 月 7 日, 接受日期: 1994 年 5 月 29 日。

温速率为 $10^{\circ}\text{C}/\text{min}$ 。载气和补偿气 (N_2 , 高纯) 分别为 $50\text{ml}/\text{min}$ 和 $60\text{ml}/\text{min}$, 分流比为 $1:25$ 。氢气和空气压力分别为 63.7kPa 和 127.5kPa , 进样量 $1-2\mu\text{l}$ 。

1.2 Euclid 贴近度聚类分析 将海面溢油的烃类含量比值数值归一化。取溢油信息点 m 个, 第 i 种油 x_i 的 m 个烃类含量比值为 $x'_{i1}, x'_{i2}, \dots, x'_{im}$; 第 i 种油第 k 个烃类含量比值归一化值为 $x_{ik} = x'_{ik} / \sum_{k=1}^m x'_{ik}$ 。

归一化后的 $x_{ik} (k = 1, 2, \dots, m)$ 满足 $0 \leq x_{ik} \leq 1$, 并可看作一个模糊向量, 故第 i 种油可用一个模糊向量表征为一个模糊子集 $A_i (A_i = 1, 2, \dots, n)$, 第 i 种油与第 j 种油的 Euclid 贴近度(陈贻源, 1984)可定义为: $\gamma_{ij} = 1 - \sqrt{\frac{1}{m} \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]}$ 。

设溢油的有限论域为 $\mathbf{U}, A_1, A_2, \dots, A_n$ 是 \mathbf{U} 上的 n 个模糊子集。由 γ_{ij} 计算的值组成一个 n 阶模糊相似矩阵 \mathbf{R} 。由于风化和分析误差的影响, 使得同一种油的重复样和风化样的贴近度不是 1 (徐恒振等, 1990), 因而, 溢油同溢油源和可疑油源间就有一个最大亲疏程度的问题。应用于连生(1982)模糊相似矩阵 \mathbf{R} 任一行(列)的主对角线除外的最大矩阵元与置信水平相等的原理, 由相似矩阵 \mathbf{R} 可直接得到溢油的模糊动态聚类图。根据最大隶属原则, 若 A_i 与 A_j 聚为一类时的置信水平 λ_i 为最大值, 则 A_i 属于 A_j , 即溢油 A_i 为溢油源 A_j 所排放。

2 结果与讨论

2.1 海面溢油鉴别信息点的选择 在海面溢油鉴别的研究中, 探讨各油种间的亲疏性和风化规律是必要的。通过对国内常见的 14 种油品(表 1)进行色谱分析可知, 14 种油品中 $n\text{C}_8-n\text{C}_{32}$ 之间的烃类有较好的分离, 其中 Pr (姥鲛烷) 与 $n\text{C}_{17}$ 、Ph (植烷) 与 $n\text{C}_{18}$ 有较高的分离度, 其色谱图可参见徐恒振等报道(1991); 9 种原油的烃类分布范围为 $n\text{C}_8-n\text{C}_{32}$, 高值区居 $n\text{C}_{11}-n\text{C}_{23}$ 之间, $n\text{C}_{24}$ 以后的烃类逐渐减少; 5 种燃料油的分布范围多为 $n\text{C}_{10}-n\text{C}_{24}$ 之间, 呈中间高两头低的态势分布; 30# 重柴油、0# 轻柴油和任丘原油风化一个月, $n\text{C}_{16}$ 以前的烃类基本上被风化掉, $n\text{C}_{16}$ 以后的烃类受风化影响较小, 尤其是 Pr 和 Ph, 油种间的差异显著, 且受风化影响更小, 具有较强的生物地球化学特征[30# 重柴油和任丘原油的风化见徐恒振等报道(1991)]。考虑到原油类和燃料油类烃类分布范围和受风化影响的不同, 正确表征油种的信息点应该是选择那些受风化影响小且油种间固有差别较大的物质(或参数值)。对此, 本模式油种表征量的数据处理点定为 Pr/ $n\text{C}_{17}$, Ph/ $n\text{C}_{18}$, Pr/Ph, $\frac{n\text{C}_{21} + n\text{C}_{22}}{n\text{C}_{19} + n\text{C}_{20}}$ 4 个比值, 前 3 个比值为油种生物标志物比值(马永安等 1988; 李从玲, 1990), 后一个比值为油种富存烃类比值, 用这 4 个比值处理数据, 既可消除进样量的影响, 又可减小风化的影响(徐恒振等, 1991)。表 1 中同一种油的这 4 个比值受实验误差的影响(用 4 个重复样的标准偏差表示)和受风化影响[用 6 个风化(1, 5, 10, 15, 21 和 30d)样品的标准偏差表示]较异种油(14 种油)间的固有差别为小。如表 1 中任丘原油的 4 个重复样和 6 个风化样的 Pr/ $n\text{C}_{17}$ 分别为 0.5457 ± 0.0133 和 0.5773 ± 0.0124 , 其中标准偏差 0.0133 和 0.0124 皆小于其均值 0.5457 和 0.5773 分别同该列中其余油种相应 Pr/ $n\text{C}_{17}$ 间的差值。同样, 表中其余标准偏差值亦有此种特性。可见, 用这 4 个比值表

表 1 14 个油种的分布特征和油种(包括重复样和风化样)的 4 个烃类含量比值

Tab. 1 Characteristics of distribution of 14 types of Oil and 4 ratios of hydrocarbon contents of 14 types of oil including repeatability and weathered samples

序号	油种	烃类分布范围	高值区	Pr/nC ₁₇	Pr/nC ₁₈	Pr/Ph	$\frac{nC_{21} + nC_{22}}{nC_{19} + nC_{20}}$ ¹⁾
1	渤海 7# 平台原油	nC ₈ -nC ₃₂	nC ₁₁ -nC ₂₃	0.4147	0.3303	1.3612	0.8435
2	大庆原油	nC ₈ -nC ₃₂	nC ₈ -nC ₂₃	0.1843	0.1611	1.2021	0.8763
2(1)-2(4)	大庆原油 4 个重复样	nC ₈ -nC ₃₂	nC ₈ -nC ₂₃	0.1918	0.1623	1.2285	0.8582
				±0.0051	±0.0034	±0.0411	±0.0173
3	渤海原油	nC ₈ -nC ₃₂	nC ₁₁ -nC ₂₃	0.4790	0.4233	1.3251	0.8332
4	辽河原油	nC ₈ -nC ₃₂	nC ₁₀ -nC ₂₃	0.6299	0.9666	0.7746	0.8576
5	胜利原油	nC ₈ -nC ₃₂	nC ₁₁ -nC ₂₃	0.7167	1.5558	0.5094	0.8287
6	南海单样原油	nC ₈ -nC ₃₂	nC ₁₀ -nC ₂₃	0.3930	0.2133	2.1790	0.8589
7	南海原油	nC ₈ -nC ₃₂	nC ₈ -nC ₂₃	0.4687	0.2467	2.2162	0.7898
8	大港原油	nC ₈ -nC ₃₂	nC ₁₂ -nC ₂₃	0.6759	0.5768	1.2119	0.8910
9	海军燃料油	nC ₉ -nC ₂₀	nC ₁₂ -nC ₁₈	0.1998	0.1488	2.4850	0.0000
10	30# 重柴油	nC ₉ -nC ₂₉	nC ₁₆ -nC ₂₆	0.2012	0.1914	0.8447	1.1082
10(1)-10(6)	30# 重柴油 6 个风化样	nC ₁₅ -nC ₂₉ (风化 30d)	nC ₁₇ -nC ₂₅ (风化 30d)	0.2016	0.1743	0.7994	1.0851
				±0.0047	±0.0034	±0.0186	±0.0102
11	1000 秒燃料油	nC ₁₀ -nC ₂₂	nC ₁₁ -nC ₁₉	0.1282	0.1515	1.1660	0.3164
12	20# 重柴油	nC ₁₀ -nC ₂₄	nC ₁₁ -nC ₂₂	0.1441	0.1373	1.2256	0.5099
12(1)-12(4)	20# 重柴油 4 个重复样	nC ₁₀ -nC ₂₄	nC ₁₁ -nC ₂₂	0.1469	0.1410	1.2020	0.4953
				±0.0019	±0.0056	±0.0684	±0.0035
13	0# 轻柴油	nC ₉ -nC ₂₃	nC ₁₂ -nC ₁₉	0.1724	0.1624	1.4060	0.3278
13(1)-13(4)	0# 轻柴油 4 个重复样	nC ₉ -nC ₂₃	nC ₁₂ -nC ₁₉	0.1794	0.1623	1.4519	0.3321
				±0.0049	±0.0017	±0.0187	±0.0061
13(5)-13(10)	0# 轻柴油 6 个风化样	nC ₁₅ -nC ₂₃ (风化 30d)	nC ₁₆ -nC ₃₁ (风化 30d)	0.1966	0.1627	1.4268	0.3455
				±0.0087	±0.0077	±0.0832	±0.0076
14	任丘原油	nC ₈ -nC ₃₂	nC ₁₃ -nC ₂₃	0.5688	1.4840	0.4109	0.9686
14(1)-14(4)	任丘原油 4 个重复样	nC ₈ -nC ₃₂	nC ₁₃ -nC ₂₃	0.5457	1.4389	0.4245	1.0447
				±0.0133	±0.0327	±0.0166	±0.0385
14(5)-14(10)	任丘原油 6 个风化样	nC ₁₅ -nC ₃₂ (风化 30d)	nC ₁₆ -nC ₂₆ (风化 30d)	0.5773	1.5115	0.3823	0.9661
				±0.0124	±0.0339	±0.0241	±0.0259

1) 表中凡有“±”的数值,前一个数为样本均值,后一个数为标准偏差。

征油种是比较客观和全面的。这里将其分别定义为溢油(或油种)的 4 个烃类含量比值 $x'_{ik}(k=1,2,3,4)$ 。

2.2 油种(溢油)鉴别的 Euclid 贴近度聚类分析结果及其风化和实验误差的影响 以表 1 中 4 个烃类含量比值为指标 ($k=4$), 以 14 种油为样本 ($n=14$), 进行 Euclid 贴近度聚类分析, 可得到 14 种油的聚类图(图 1)。从图 1 中可看出 14 种油间的亲疏程度, λ_i 可分类置信水平范围为 [0.8726, 0.9847], 其中最相似的是 1000 秒燃料油和 0# 轻柴油, 其次是南海单样原油和南海原油。同样将任丘原油及其 4 个重复样和 6 个风化样为样本 ($n=24$) 与其他 13 种油进行聚类分析, 可得到相应的聚类图(图 2)。仿此, 可得到大庆原油及其 4 个重复样 ($n=18$), 30# 重柴油及其 6 个风化样 ($n=20$), 20# 重柴油及其 4 个重复样 ($n=18$), 0# 轻柴油及其 4 个重复样和 6 个风化样 ($n=24$) 分别与除已外的 13 种油所形成的 4 个聚类图(略)。上述 5 种油及其重复样或风化样的相应

的 λ_i 值列于表 2 和表 3。从图 1 和表 2、表 3 中可看出,由于风化和实验误差的影响,使得同一种油间的 Euclid 贴近度不是 1,而是小于 1 但大于该油种同其它油种成类时的 λ_i 值。这里把风化或实验误差影响大庆原油、30# 重柴油、20# 重柴油、0# 轻柴油、任丘原油的各自 Euclid 贴近度的最小 λ_i 值 0.9953, 0.9970, 0.9857, 0.9902 和 0.9907 分别叫做这几种油的临界 Euclid 贴近度, 记为 λ_i^* 。将风化和实验误差所形成的油种鉴别的模糊区 $[\lambda_i^*, 1)$ 叫做油种鉴别的不可辨识区。将 $[0.8726, \lambda_i^*)$ (0# 轻柴油为 $[0.8807, \lambda_i^*)$, 这是由于风化影响改变了主对角线除外的最大矩阵元在模糊相似矩阵中的位置之故) 叫做油种鉴别的可辨识区。可见,在海面溢油鉴别过程中,关键是要找到溢油同所有可疑溢油源间的临界 Euclid 贴近度 λ_i^* 、油种不可辨识区和可辨识区。分析表 2、表 3 中这 5 种油的 λ_i 值,由于受风化和实验误差的影响,其 λ_i 的范围为 $[0.9857, 1)$, 表明了油种鉴别的 λ_i^* 具有随机性和模糊性双重不确定性。据此,在海面溢油鉴别时,根据最大隶属原则若溢油与某可疑溢油源间的 Euclid 贴近度大于 0.9857 时,则可大致认为两者为同源油,否则

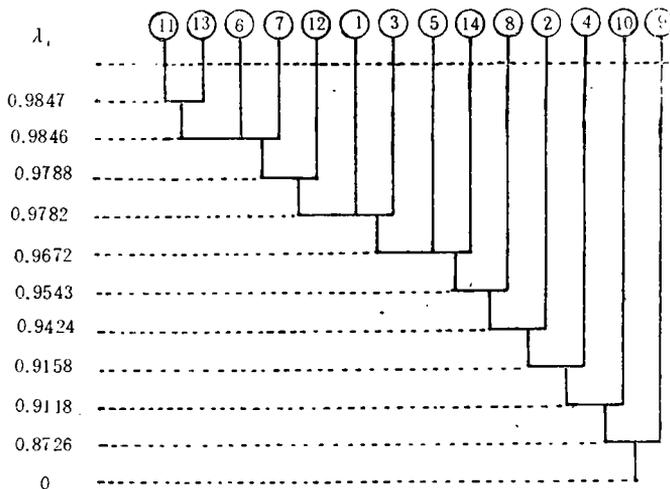


图 1 14 种油的聚类图

Fig. 1 Clustering charts of 14 types of oil

表 2 4 种油及其重复样的置信水平

Tab. 2 Confidence levels of 4 types of oil and their repeatability oil samples

大庆原油	序号	2	2(1)	2(2)	2(3)	2(4)
	λ_i	0.9982	0.9953	0.9953	0.9989	0.9989
20# 重柴油	序号	12	12(1)	12(2)	12(3)	12(4)
	λ_i	0.9977	0.9977	0.9857	0.9970	0.9970
0# 轻柴油	序号	13	13(1)	13(2)	13(3)	13(4)
	λ_i	0.9985	0.9985	0.9988	0.9984	0.9988
任丘原油	序号	14	14(1)	14(2)	14(3)	14(4)
	λ_i	0.9974	0.9963	0.9957	0.9968	0.9968

就不是。若从溢油鉴别方法学的角度出发,所提出的海面溢油鉴别 Euclid 贴近度临界值 λ_i^* 和 λ_i 模糊区等概念,其本质类同于溢油鉴别的最小角距离原则(戴云丛,1990),对

表 3 3 种油及其风化样的置信水平

Tab. 3 Confidence levels of 3 types of oil and their weathered oil samples

风化天数 (d)		0	1	5	10	15	21	30
30#重柴油	序号	10	10(1)	10(2)	10(3)	10(4)	10(5)	10(6)
	λ_i	0.9970	0.9971	0.9974	0.9974	0.9975	0.9975	0.9971
0#轻柴油	序号	13	13(5)	13(6)	13(7)	13(8)	13(9)	13(10)
	λ_i	0.9985	0.9966	0.9993	0.9933	0.9993	0.9984	0.9902
任丘原油	序号	14	14(5)	14(6)	14(7)	14(8)	14(9)	14(10)
	λ_i	0.9974	0.9974	0.9947	0.9973	0.9973	0.9931	0.9907

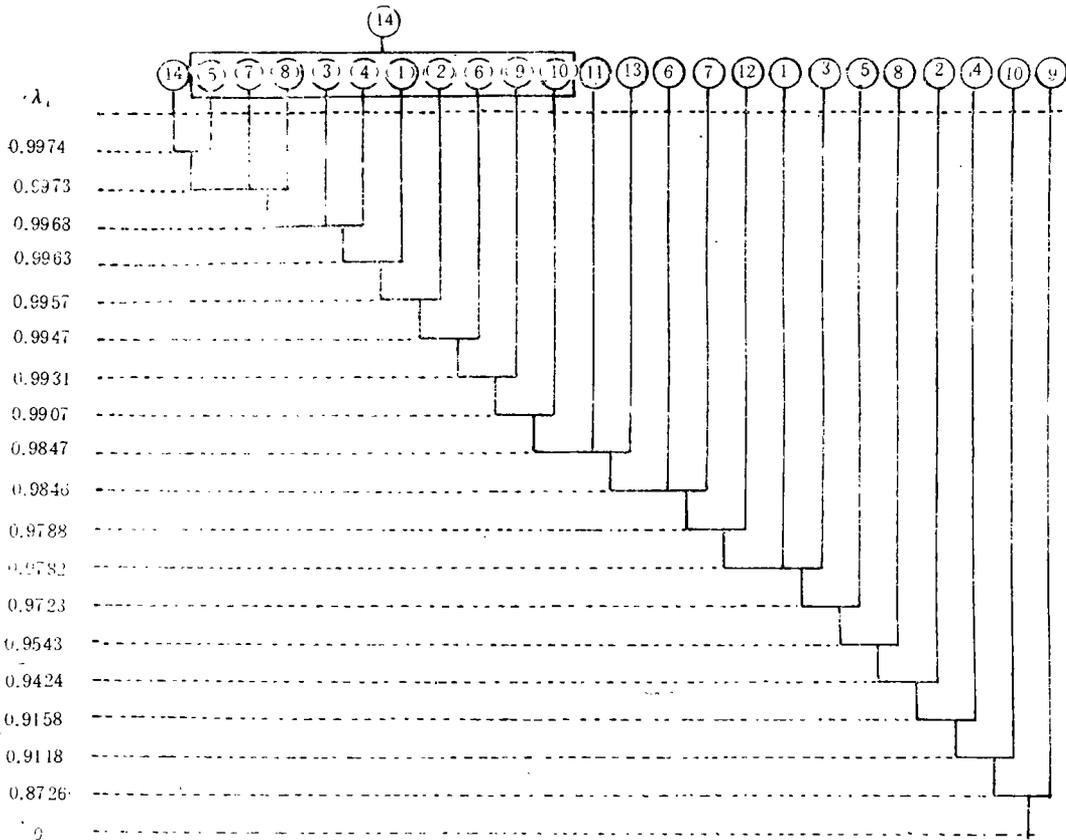


图 2 13 种油同任丘原油及其风化样、重复样的聚类图

Fig. 2 Clustering charts of 13 types of oil and Renqiu crude oil and its weathered and repeatability oil samples

数比值临界线原则(徐学仁,1990)和夹角余弦的模糊区原则(徐恒振等,1992)。

2.3 海面溢油鉴别案例 为了验证本方法对海面溢油鉴别的可行性,于1984年对大

连湾海域实际溢油案例进行鉴别,海面溢油的 Euclid 贴近度聚类分析见图 3。表明当 $\lambda_i^* = 0.9949$ 时, 1# 和 3# 为同源油,这一结果与马永安等(1988)、徐恒振等(1987,1990,1992)的报道一致。

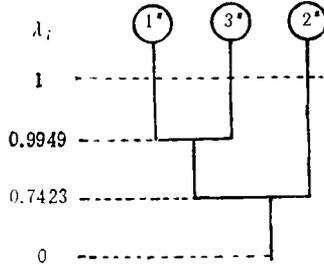


图 3 大连湾海面溢油聚类图

Fig. 3 Clustering charts of spilled oil at sea surface of Dalian Bay

3 结语

3.1 所选择的表征油种的 4 个烃类含量比值 Pr/nC_{17} , Ph/nC_{18} , Pr/Ph , $\frac{nC_{21} + nC_{22}}{nC_{19} + nC_{20}}$,

可以鉴别风化了一个月的油种和海面溢油。风化和实验误差的影响,使得海面溢油鉴别的 λ_i^* 值由 1 下降到 1 与 0.9857 之间的某一个值,从而形成溢油不可鉴别的 λ_i 模糊区域值,根据 λ_i^* 和油种可鉴别的 λ_i 区域值的大小,即可鉴别溢油,不受人为因素的影响。

3.2 大庆原油,30# 重柴油, 20# 轻柴油, 0# 轻柴油以及任丘原油分别用其它 13 种油可分类的置信水平 λ_i 的范围依次为: $[0.8726, 0.9953)$, $[0.8726, 0.9970)$, $[0.8726, 0.9857)$, $[0.8807, 0.9902)$ 和 $[0.8726, 0.9907)$ 。

参 考 文 献

马永安、李洪、徐恒振等,1988,毛细管色谱技术在鉴别海面溢油源中的应用,海洋环境科学,7(3): 82—89。
 于连生,1982,模糊聚类法的研究,吉林大学自然科学学报,4: 106—114。
 李从玲,1990,近代海洋沉积物(层)姥鲛烷/植烷比值及其地球化学意义,海洋地质与第四纪地质,10(4): 77—78。
 陈贻源,1984,模糊数学,华中工学院出版社(武昌),41—51。
 徐学仁,1990,气相色谱对数值法鉴别海面溢油源,海洋学报,12(6): 733—740。
 徐恒振、周传光、李洪,1987,GC-FTD 鉴别海上溢油及其模糊聚类分析,海洋环境科学, 6(4): 67—76。
 徐恒振、周传光、李洪,1990,模糊识别法在 GC-FID/FPD 鉴别海面溢油中的应用,海洋通报,9(5): 33—38。
 徐恒振、周传光,1991,灰色系统理论在建立海面石油烃风化模式中的应用,海洋学报,13(1): 51—59。
 徐恒振、李宗品,1992,模糊最大矩阵元鉴别海面溢油,环境科学,13(4): 67—69。
 戴云丛,1990,红外光谱计算分析鉴别海面溢油源的研究,分析化学,13(3),206—209。

CLUSTER ANALYSIS OF EUCLID APPROACH FOR IDENTIFICATION OF OIL SPILL AT SEA SURFACE

Xu Hengzhen, Shang Longsheng, Zhou Chuanguang, Liu Tong

(*Institute of Marine Environmental Protection, SOA, Dalian 116023*)

ABSTRACT

Fourteen common oil types of China and spilled oils at sea surface in Dalian Bay were analysed with gas chromatograms in 1984—1986. The distributions of hydrocarbons of nine crude oils and five fuel oils mostly ranged from nC_8 — nC_{32} and nC_{10} — nC_{24} . The impact of weathering on the oil chromatogram fingerprint is important. One month weathering experiment on 30# heavy diesel oil, 0# light diesel oil, and Renqiu crude oil on the seawater, showed the fluctuation of the weathered oil with carbon less than nC_{16} was large, than the fluctuation of the weathered oil with carbon more than nC_{16} , especially Pr and Ph, was less. Pr/ nC_{17} , Ph/ nC_{18} , Pr/Ph and $(nC_{21} + nC_{22})/(nC_{19} + nC_{20})$ oil type information points for data handling. A model is established for fuzzy mathematics clustering analysis of Euclid approach for identification of sea surface spilled oil. The Euclid approach between the types of oil is expressed as:

$$\gamma_{ij} = 1 - \sqrt{\frac{1}{m} \sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

where x_{ik} is the ratio of the hydrocarbon content k in the oil i , x_{jk} is the ratio of the hydrocarbon content k in the oil j . If the spilled oil and the suspected source oil come from the same source, their Euclid approach will be equal to or near 1. The concept of fuzzy area of confidence level λ_i of Euclid approach for identifying types of oil was introduced. The confidence level λ_i for the identification of the sea surface spilled oil is found to be greatly influenced not only by the types of the oil, but also by weathering and experimental error. In general, the threshold value of this effects is equal to or more than 0.9857. The results show that the confidence levels λ_i for differentiating Daqing crude oil, 30# heavy diesel oil, 20 heavy diesel oil, 0# light diesel oil, and Renqiu crude oil, from other thirteen types of oil, are [0.8726, 0.9953), [0.8726, 0.9970), [0.8726, 0.9857), [0.8807, 0.9902) and [0.8726, 0.9907), respectively.

Key words Euclid approach Fuzzy clustering analysis Identification Spilled oil