用相关系数法匹配色谱图形
鉴别海面溢油源

徐 学 仁

(国家海洋局海洋环境保护研究所，大连　116023)

提要　　用皮尔逊相关系数法匹配色谱图形，根据溢油与可疑油的相似性鉴别溢油源。通过风化实验，研究出六个受风化影响很小的油色谱峰作为信息点，用 K-最邻近分类法鉴别非常相似的油品。

关键词　　相关系数　色谱　鉴别　溢油

为了提高溢油鉴别工作的效率，增强人的识别能力，从 70 年代起，Brown 等人提出了许多溢油图形识别技术[2~4,7,11,12]。气相色谱法有较高的分辨率和丰富的指纹信息，是一种重要的溢油鉴别方法[6]。然而，石油受到风化后，指纹信息发生变化，用人的视觉往往难以辨认。为了提高人对风化油品的识别能力，必须寻找油品中固有不变的痕迹，根据这些痕迹提供的信息，用数学统计的方法识别谱图。有人曾用欧几里得空间距方法和对数比值法识别色谱图形[8]，这两种方法的缺点是必须对谱图标化，使工作量增加。

本文用相关系数法识别色谱图形，该法在仪器线性范围内不受浓度的影响，因此不须对谱图标化，对风化油品的识别能力有显著提高。相似油品的识别是很困难的，但是用 K-最近邻分类法可以有效的识别相似油品的图形。

一、实　验

1. 风化实验

1987 年 6 月至 7 月，在大连河口海湾制作 3 个聚乙烯膜围成的圆桶，长 3m，直径 0.5m，下无底，浮于海面。分别布置大庆原油，0# 轻柴油、20# 重柴油。油膜厚度约 0.5mm，水温 12.4—17.6℃，气温 17—24℃。

2. 谱图测量

所有的谱图都在日立 163 型色谱仪上测定，连接 C-R3A 型微处理机。色谱条件：柱长 0.31mm × 60m，OV-101SCOT 玻璃毛细管柱，柱程序升温 100℃—295℃，7.5℃/min，检测器与气化室温度是 310℃。

二、结 果 与 讨 论

1. 谱图特征的选择

特征峰选择可以看成是从原始的 n 维空间到一个 m 维空间（$m \ll n$）的映射。一张完整的油色谱图有几十个峰，如果把每一个峰都作为匹配测量的信息点，不仅需要很多的运算时间，而且匹配度也会因风化受到影响。因此，必须选择具有一定代表性、可分离性和耐风化性的少数峰作为匹配测量的信息点。各种油品的色谱峰分布特征有很大差别，原油的色谱图正构烷烃分布范围从 C_9—C_{36}，$0^\#$ 轻柴油从 C_9—C_{22}。为了使本方法既实用于原油又实用燃料油，必须在 C_9—C_{22} 之间选择特征峰。风化对油色谱指纹影响很大，我们用大庆原油海上模拟风化一个月，结果见图 1。图 1 用受风化影响很小的 C_{20} 标化每一个峰，从图可以看出，C_{16} 以前的峰受风化影响很大。峰高随风化时间呈线性下降，峰高与风化时间的关系为

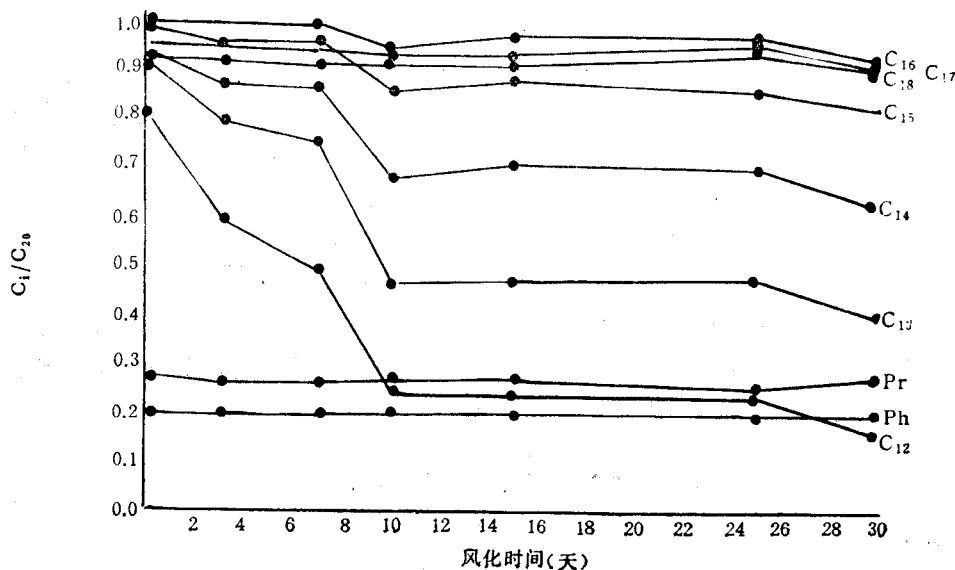$$Z = Z_0 e^{mt}$$

式中，Z 为峰高；Z_0 为截距；m 为斜率；t 为风化时间。

图 1 大庆原油海面上风化一个月的变化

Fig. 1 Variation of Daqing crude oil weathered for one month on sea surface

在 C_{16}—C_{22} 之间的老鲛烷（Pr）和植烷（Ph）是非常有特征的两个峰，有人曾用这两个峰作为油色谱图识别指标[1]。综合以上各因素，本文选择 C_{17}，Pr，C_{18}，Ph，C_{19}，C_{20} 六个峰作为匹配测量的信息点。这与资料报道基本一致[14]。

2. 图形识别原理

用皮尔逊相关系数测量两张谱图的相似性。如果溢油与可疑油是同一个污染源油品，相关系数 r 应等于 1 或接近于 1，相关系的定义如下：

$$r = \frac{\Sigma X_i Y_i - \frac{1}{n} \Sigma X_i \Sigma Y_i}{\left[\left(\Sigma X_i^2 - \frac{1}{n}(\Sigma X_i)^2\right)\left(\Sigma Y_i^2 - \frac{1}{n}(\Sigma Y_i)^2\right)\right]^{1/2}} \qquad (1)$$

式中，X_i 和 Y_i 分别代表溢油和可疑油色谱图第 i 峰的数值，在这里 $n=6$，等式(1)可用 i 物质的响应值表示

$$r = \frac{\Sigma S_{i1} S_{i2} - \frac{1}{n} \Sigma S_{i1} \Sigma S_{i2}}{\left[\left(\Sigma S_{i1}^2 - \frac{1}{n}(\Sigma S_{i1})^2\right)\left(\Sigma S_{i2}^2 - \frac{1}{n}(\Sigma S_{i2})^2\right)\right]^{1/2}} \qquad (2)$$

物质 i 的响应值定义是

$$S_i = \frac{A_i}{m_i} \qquad (3)$$

式中，S_i 为物质 i 的响应值；A_i 为物质 i 的峰面积；m_i 为物质 i 的进样量。

把(3)式代入(2)式得

$$r = \frac{\frac{1}{m_1 m_2}\Sigma A_{i1} A_{i2} - \frac{1}{n m_1 m_2}\Sigma A_{i1} \Sigma A_{i2}}{\left[\left(\frac{1}{m_1^2}\Sigma A_{i1}^2 - \frac{1}{n m_1^2}(\Sigma A_{i1})^2\right)\left(\frac{1}{m_2^2}\Sigma A_{i2}^2 - \frac{1}{n m_2^2}(\Sigma A_{i2})^2\right)\right]^{1/2}} \qquad (4)$$

(4)式化简得

$$r = \frac{\Sigma A_{i1} A_{i2} - \frac{1}{n}\Sigma A_{i1}\Sigma A_{i2}}{\left[\left(\Sigma A_{i1}^2 - \frac{1}{n}(\Sigma A_{i1})^2\right)\left(\Sigma A_{i2}^2 - \frac{1}{n}(\Sigma A_{i2})^2\right)\right]^{1/2}} \qquad (5)$$

从公式里消除了进样量。因此 r 值在仪器响应值的线性范围内不受浓度的影响。

3. 已知油品间的相互鉴别

收集了 4 种未风化原油，4 种未风化的燃料油用(5)式进行相关矩阵计算，结果见表1。

表 1　已知油品间的相关矩阵
Tab.1　Known oils correletion matrix

	大庆原油	胜利原油	辽河原油	30# 重柴油	渤海原油	1000 秒燃料油	0#轻柴油	20# 柴油
大庆原油	0.9996	0.5068	0.5561	0.8278	0.9858	0.7709	0.7920	0.4497
胜利原油		0.9919	0.3679	−0.2360	−0.5203	−0.4981	−0.4890	0.1596
辽河原油			0.9968	0.5772	0.5449	0.3810	0.4018	0.4029
30# 重柴油				0.9999	0.8421	0.2877	0.3222	−0.1738
渤海原油					0.9908	0.7438	0.7626	0.3871
1000 秒燃料油						0.9988	0.9778	0.9826
0#轻柴油							0.9984	0.9099
20# 柴油								0.9992

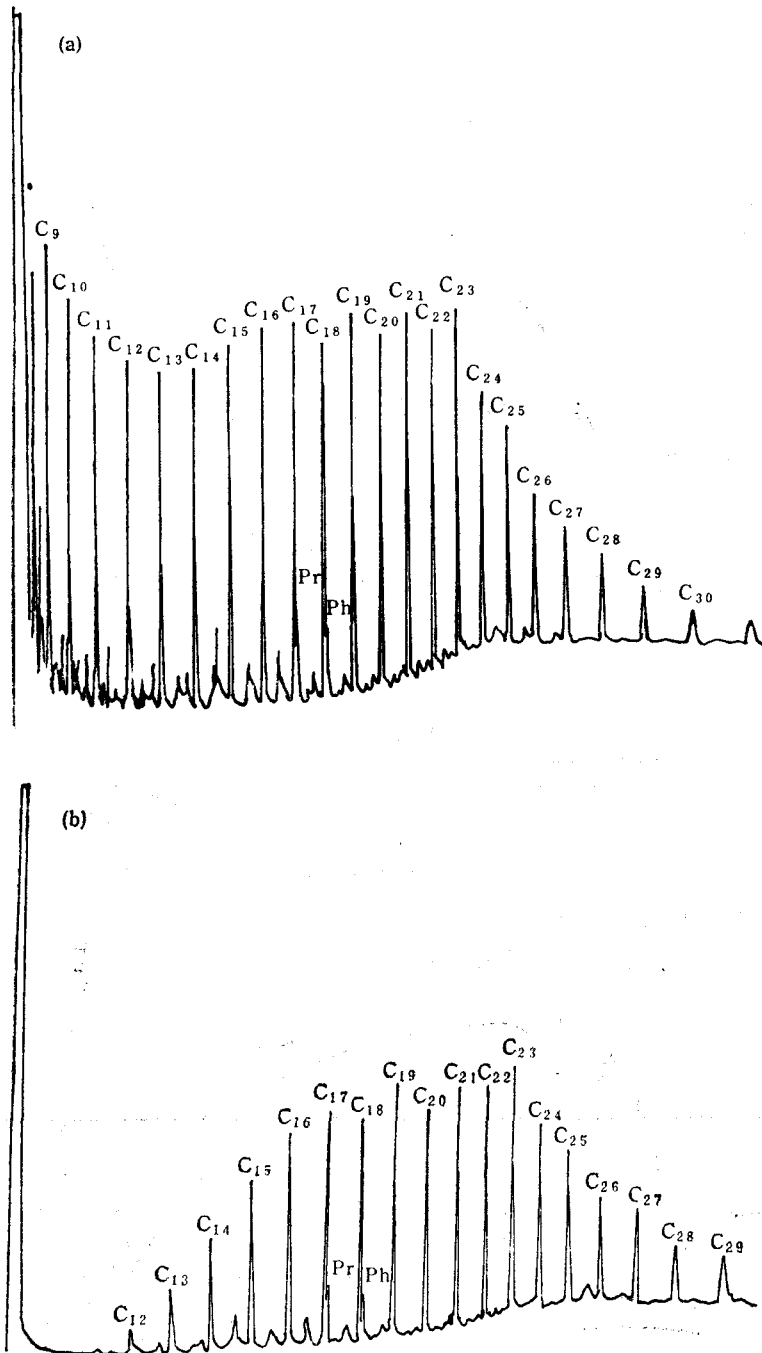从表 1 可以看出、各种油品的批间相关系数,远远小于油品的批内相关系数, 批内的相关系数都在 0.9900 以上。结果表明,所选择六个特征峰是具有一定的代表性和可分离性。

图 2 大庆原油未风化 (a) 和风化 28 天 (b) 的色谱图

Fig. 2 Gas chromatograms of original (a) and weathered 28 days(b) Daqing crude oil

风化对油品色谱指纹影响很大，图 2 是大庆原油未风化和风化 28 天的色谱图。可以明显看出，未风化与风化的色谱图指纹有较大的改变，然而在直观上却较难辨认。二者相关系数 $r = 0.9930$，仍有较大的匹配度。$O^\#$ 轻柴油风化 30 天后与未风化之间的相关系数 $r = 0.9931$，风化实验结果表明，本方法对风化油品的鉴别是非常有效的。

4. 用 K-最近邻法识别相似油品

本文所说的相似油品是指相关系数在 0.9000 以上，这类油品的直接识别是很困难的。例如两个不同批号的 $O^\#$ 轻柴油 $O_1^\#$ 和 $O_2^\#$，从色谱图上看峰的数目相同，Pr 和 Ph 的峰高基本相同，轮廓相似。相关运算的结果也非常相似 $O_1^\#$，的批内相关系数是 0.9984，$O_2^\#$ 的批内相关系数是 0.9991，$O_1^\#$ 和 $O_2^\#$ 的批间相关系数 0.9948。有些油品从总的图形分布看有非常明显的差别，但是在我们所选择六个特征峰范围内较相似，如 $O^\#$ 轻柴油与 1000 秒燃料油 $r = 0.9778$，大庆原油与渤海原油 $r = 0.9858$。这类油品尽管很相似，但是还是有微小的差别的，用 K-最近邻分类法可以有效的识别这类图形。所谓 K-最近邻法就是用未知的海面溢油与每一个已知的可疑油比较相似性，根据未知与每一个已知油品的相似程度给不同的权，最近邻的给最大的权，K 是最近邻的数目。该方法 1967 年提出[9]，后来被用于各种谱图识别[5,10,13]。加权的方法很多，本文的加权方法是 $\dfrac{1}{(1-r)^2}$，

r 是相关系数。通过最近邻计算，上述两个 $O^\#$ 轻柴油有明显的差别，$O_1^\#$ 的批内权值 39，$O_2^\#$ 的批内权值 123，$O_1^\#$ 和 $O_2^\#$ 的批间权值只有 2。由于权值很大，比较不方便，所以每个权值都缩小一万倍。1000 秒燃料油的批内权值 69，而与 $O_1^\#$ 的批间权值只有 0.2。

5. 实际溢油的鉴别

1987 年 8 月发生在某港区一起溢油事故，溢油量大约 200kg，现场调查发现，有两条船有较大嫌疑，这里把它们叫 A 船和 B 船，执法人员分别在 A 船机舱污水（A_1）、油柜（A_2）和 B 船机舱污水（B_1）采样分析。相关计算结果，海面溢油与 $A_1, r = 0.9967$；海面溢油与 $A_2, r = 0.9884$；海面溢油与 $B_1, r = 0.9758$；属于相似油品。K-最近邻计算结果 $A_1 = 92, A_2 = 7, B_1 = 2$，海面溢油与 A_1 有最大的相关性。

参 考 文 献

[1] 马永安等，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。
[2] Ahmaddjian, M., et al., 1976, Infrared spectra of petroleum weathered naturally and under simulated conditions, *Environ. Sci. Technol.*, 10: 777—781.
[3] Anderson, C. P., et al., 1987, Weighting factors for infrared data points in pattern recognition for oil identification. Pittsburgh Conference, P. 397.
[4] Anderson, C. P. and Killeen, T. J., 1980, Improved identification of spilled oil by infrared spectroscopy, *Environ. Sci. Technol.*, 14: 1230—1234.
[5] Baer, C. D. and Brown, C. W., 1977, Identifying the soure of weathered petroleum: matching infrared spectra with correlation coefficients. *Appl. Spectrosc.*, 31: 524—527.
[6] Bentz, A. P., 1976, Oil spill identification, *Anal. Chem.*, 48: 455A—470A.
[7] Brown, C. W. and Lynch, P. F., 1976, Identification of oil slicks by infrared spectroscopy. Final Report, U. S. Coast Guard Contract DOTCG-81-74-1099, NTIS Accession NO. ADA040975.
[8] Clark, H. A. and Jurs, P. C., 1975, Qualitative determination of petroleum sample type from gas chromatograms using pattern recognition techniques, *Anal. Chem.*, 47: 374—378.
[9] Cover, T. M. and Hart, P. E., 1967, Nearest neighbor pattern classification, IEEE Trans. information

Theory, IT-13, 21—27.

[10] Justice, J. B. and Isenhour, T. L., 1974, Information content of mass spectra as determined by pattern recognition methods, *Anal. Chem.*, **46**: 223—226.

[11] Killeen, T. J., 1981, Oil-matching by using a simple vector model for fluorescence spectra, *Talanta*, **28**: 1—6.

[12] Kawahara, F. K., 1969, Identification and differentiation of heavy residual oil and asphalt pollutants in surface waters by comparative ratios of infrared absorbances, *Environ. Sci. Technol.*, **3**: 150—153.

[13] Kowalski, B. R. and Bender, C. F., 1972, The K-nearest neighbor classification role (pattern recognition) applied to nuclear magnetic resonance spectral interpretation, *Anal. Chem.*, **44**: 1405—1410.

[14] Petrakis, L. and Weiss, F. T., 1980, Petroleum in the marrine environment, American Chemical Society Washingtion, D. C., 56—81.

IDENTIFYING THE SOURCE OF SPILLED OIL AT SEA BY MATCHING GAS CHROMATOGRAM PATTERN WITH CORRELATION COEFFICIENTS

Xu Xueren

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

ABSTRACT

Chromatograms peaks of six slightly weathered oil samples were selected as information points C_{17}, Pr, C_{18}, Ph, C_{19} and C_{20} in the field experiment. The degree of match between suspected spill oil chromatogram peaks (determined by the Peatson correlative coefficient) and the chromatogram area as a basic element was utilized to identify the spilled oil. If the spilled oil and the suspected source oil come from the same source, their chromatograms will inevitablly be the same and the correlation coefficient will be equal to or near 1. This method works within the instrument limitation, and therefore will not be influenced by the oil concentration. In this study, four crude oils and four light fuel oil samples were selected. Duplicated chromatograms were run for each sample under the instrumental conditions. After that, all the chromatograms were calculated according to the correlation matrix. The results showed that the correlation coefficients of samples within the batch were all > 0.9900, and ≤ 0.9000 between the batches. The impact of weathering to the oil chromatogram fingerprint is important. After one month weathering experiment on Daqing crude oil and $O^{\#}$ liyht fuel oil on the sea, the fluctuation of the weathered oil, especially $O^{\#}$ light fuel oil and oil with carbon less than C_{18}, chromatogram was larger, but it can not be recognized by the naked eye. By applying the six selected chromatograms as information point, the unmeathered and weathered oil chromatograms were calculated correlatively. The correlation coefficient was still over 0.9900. Identification of similar oil samples was regarded as difficult. Here, the so-called similar oil is that the correlation coefficients of two different oil samples are 0.9900 and slightly different at the 3rd and 4th after the point which can be effectively recognized by K-nearest neighbor classification. The same oil sample should have maximum similarity, therefore, maximum weighting value is given to the largest correlation coefficient. The method is $1/(1-\gamma)^2$ (γ is correlation coefficient).

Key words   Correlation coefficients, Gas chromatograms, Identifying, Spilled oil.