Matching fluorescence spectra of oil spills with spectra from suspect sources

Jianfeng Li [a], Steven Fuller [b], Julie Cattle [b], Christopher Pang Way [b], D. Brynn Hibbert [a,*]

[a] *School of Chemical Sciences, University of New South Wales, Sydney, NSW 2052, Australia*
[b] *Analytical Division, New South Wales Environment Protection Authority, Lidcombe, NSW 1825, Australia*

Abstract

Fluorescence spectroscopy is ideally suited to the analysis of oil spills as it allows chemical information of polycyclic aromatic hydrocarbons to be acquired quickly, sensitively and selectively. Unlike infrared spectra which have detailed peak information, many fluorescence spectra have only a few broad peaks. Nine different samples of crude and diesel oils were used for testing point-to-point matching across the spectral range. Five of them were discriminated by point-to-point matching algorithms and the other four very similar samples were not. Principal components analysis (PCA) did successfully discriminate among all similar samples. PCA could also distinguish the extent of weathering of different samples, an important factor in matching environmental spills.
© 2004 Published by Elsevier B.V.

1. Introduction

Petroleum oil pollution of the natural environment has been recognized as a major problem for many years. The rapid and reliable detection and characterization of petroleum oil contaminants is of crucial importance for environmental protection. Due to their complex constitution, petroleum oils display a wide range of physical and chemical characteristics. They can vary from colorless liquids to black, viscous, tar-like materials, some of which are solid at room temperature. Different techniques have been used for routine analysis of environmental samples. In recent years gas chromatography combined with mass spectrometry (GC/MS) [1–3] has become a method of choice. The method gives useful structural information of the oils in question and is usually successful in characterizing the sample. However, GC/MS is expensive to set up and run, has relatively poor reproducibility, the time for analysis can be long, especially when sample preparation is taken into

account, and data analysis is not straightforward (although modern software is improving this situation).

Fluorescence spectroscopy is a rapid, sensitive and selective method for identification of oil spills because most petroleum oils are fluorescent due to the presence of polycyclic aromatic hydrocarbons (PAHs) [4] and other species. Because of the limited ability as an exact diagnostic tool of conventional fluorescence, advanced and more complex fluorescence techniques such as synchronous excitation [5], fluorescence life time measurement [6], total synchronous fluorescence [7], and low temperature fluorescence [8] have been proposed.

The identification of spilled oil collected from the environment becomes increasingly difficult with the lapse of time due to weathering processes. When petroleum is weathered in the environment, its chemical composition is changed due to effects of evaporation, biodegradation, and, to lesser extent, dissolution and photo oxidation [9,10]. The extent and rate of weathering are different for each spill and are controlled by a number of spill conditions and natural processes such as the type of the spilled oil, environmental conditions and microbiological activity. The spectrum of a weathered oil sample may be quite different from that of a sample of

the same oil stored in a laboratory. It is traditionally thought that fluorescence analyses alone have little use where the petroleum is heavily weathered (e.g. in surface and ground-water) or when the materials are compositionally similar (e.g. diesel oil and mineral oil of similar boiling ranges and alkane composition).

Principal components analysis (PCA) [11] is an unsupervised data analysis technique. The main applications of PCA are: (1) to reduce the number of variables that are used to describe data and (2) to detect structure in the relationships between variables or objects, that is to classify variables or objects. Therefore, PCA is applied as a data reduction or structure detection method, and also it can be used as a display method for multidimensional data.

We shall show that with direct matching of spectra, or together with a multivariate technique such as principal components analysis, fluorescence can be useful as a rapid screening method for matching oil spills with candidate sources.

2. Experimental

2.1. Oil samples

Nine oil samples were provided by New South Wales Environment Protection Authority (NSW EPA). The oils, which are detailed in Table 1, covered several crude and refined oils and historical EPA samples of known origin. The weathering procedure for the oils was to form a 2–5 mm thick slick of the oil over water in an open beaker. The beakers were placed in an unprotected area on the roof of a building. The weathering periods were 2, 7 and 15 days during November–December 2002. In Sydney, at this time the maximum and minimum temperatures are 26 and 15 °C, respectively. The weathered samples were stored in sealed containers and preserved in a refrigerator.

2.2. Fluorescence spectroscopy

All fluorescence spectra were obtained on a Perkin-Elmer Model LS-50B luminescence spectrophotometer with a

Table 1
Oil samples used in the study

Code	Description
1	In-house EPA diesel fuel
5	Past EPA sample
7	In-house EPA fuel oil
10	Kuwait crude
11	Pennzoil HD SAE 30
12	Shell Diala S (transformer oil)
14	Past EPA sample
15	Past EPA sample
16	Past EPA sample

Table 2
Calculation of similarity measures used in matching FTIR spectra

Euclidean distance	$\sqrt{\dfrac{\sum_i (A_{1,i} \times A_{2,i})}{\sqrt{\sum_i A_{1,i}^2 \sum_i A_{2,i}^2}}}$
First derivative Euclidean distance	$\sqrt{\dfrac{\sum_i (\Delta A_{1,i} \times \Delta A_{2,i})}{\sqrt{\sum_i \Delta A_{1,i}^2 \sum_i \Delta A_{2,i}^2}}}$ where $\Delta A_{i,j} = A_{i+1,j} - A_{i,j}$
Correlation	$\dfrac{\text{Cov}(\boldsymbol{A}_1, \boldsymbol{A}_2)}{s_{A_1} s_{A_2}}$
First derivative correlation	$\dfrac{\text{Cov}(\Delta \boldsymbol{A}_1, \Delta \boldsymbol{A}_2)}{s_{\Delta A_1} s_{\Delta A_2}}$ where $\Delta A_{i,j} = A_{i+1,j} - A_{i,j}$

$\boldsymbol{A}_j$ is the vector of n absorbances of sample j. $A_{i,j}$ is the ith absorbance.

personal computer using FL-WinLab Version 2.01 (The Perkin-Elmer Corporation, 761 Main Ave., Norwalk, CT 06859-0012, USA). The oil solutions were prepared using spectroscopic pure cyclohexane as solvent. A known volume of sample was made up to the desired concentration (in v/v %) with cyclohexane. The excitation wavelength was 245 nm and spectra were recorded from 300 to 500 nm. The slit widths of excitation and emission beams were 10 and 2.5 nm, respectively. The scan speed was 120 nm s^{-1}. More detailed information can be found in reference [12]. Samples were analyzed independently 3 times for samples 7, 10, 11, 12 and 14, and 10–13 times for samples 1, 5, 15 and 16 (fresh samples and after weathering).

2.3. Data analysis

2.3.1. Measurement of the similarity of two spectra

Similarity measures between spectra were evaluated using repeat spectra of samples 7, 10, 11, 12 and 14. Each spectrum was normalized to its maximum intensity. Four similarity indices: first derivative Euclidean distance, absolute value, correlation, and first derivative correlation were calculated (Table 2).

2.3.2. Principal components analysis

PCA was carried out using STATISTICA (data analysis software system), Version 6 (StatSoft, Inc., 2003).

3. Results and discussion

3.1. Un-weathered samples

Fig. 1 shows the normalized fluorescence spectra of all un-weathered tested samples. The samples were grouped into two classes based on their spectra, ones with highly similar spectra (samples 1, 5, 15, 16, Fig. 1a) and samples that had very different spectra (samples 7, 10, 11, 12, 14, Fig. 1b).
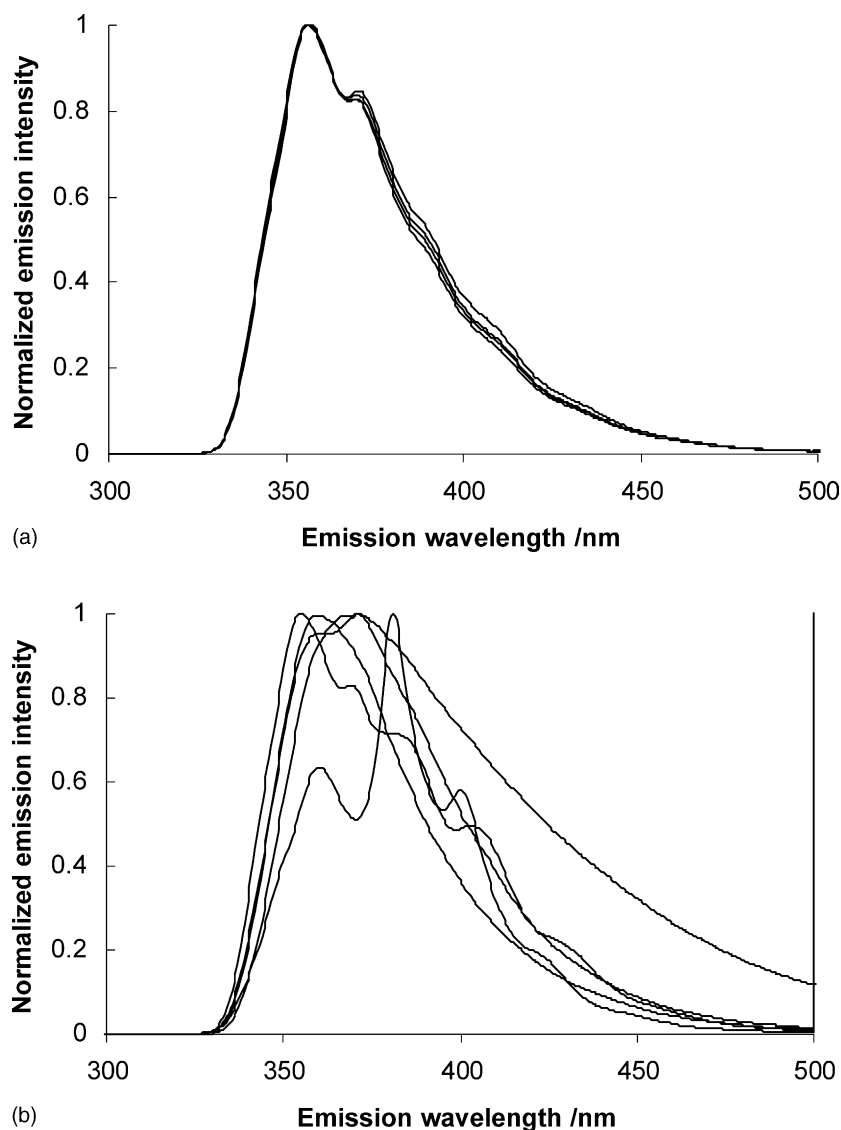
Fig. 1. Normalized fluorescence spectra of all un-weathered samples. Excitation wavelength 245 nm: (a) samples having similar spectra: 1, 5, 15, 16; (b) dissimilar samples: 7, 10, 11, 12, 14.

Each point-to-point similarity method could not distinguish between replicate spectra of the same sample and spectra of the highly similar group of Fig. 1a. They could, with different degrees of success, decide that the quite different spectra were indeed different from the 'similar' group and from each other. Fig. 2 shows this as histograms of the first derivative correlation index.

The results for each method are summarized in Fig. 3, which shows minimum, mean, and maximum values for each of the methods. First derivative correlation appears to give the best discrimination between similar and dissimilar spectra.

Point-to-point comparisons are possible when the spectra are very stable, both in terms of the wavelength and transmittance. Fluorescence spectra are sufficiently reproducible, partly because they have little fine structure (in contrast to,

for example, the fingerprint region of an infrared spectrum). However, the lack of fine structure means that it may be more difficult to pick up subtle differences between samples. In the case of a weathered sample point-to-point comparisons could not determine the extent of weathering. We therefore turned to a more information rich method, principal components analysis.

3.2. Principal components analysis

The extra dimensionality of PCA that arises from a treatment of a set of samples may have a greater ability to distinguish among samples compared with a single similarity index from a comparison of two spectra. It is an unsupervised pattern recognition method and thus requires interpretation. We will show that using PCA it is possible
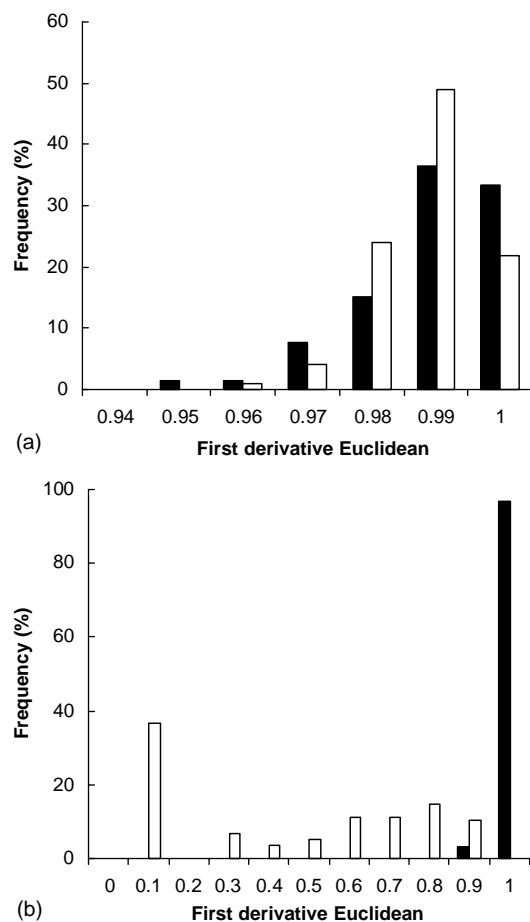
(a)

(b)

Fig. 2. Normalized histogram of first derivative Euclidean distance between fluorescence spectra of oil samples. Solid: between replicates of the same sample. No shading: between all combinations of different samples. (a) Among samples having similar spectra 1, 5, 15, 16; (b) among dissimilar samples 7, 10, 11, 12, 14.
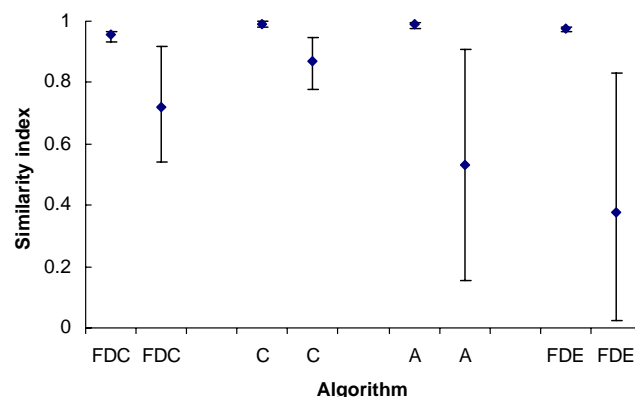

Fig. 3. Means of similarity indices applied to spectra of dissimilar samples 7, 10, 11, 12, 14. First point: between replicate spectra of the same sample; second point: between different samples. Algorithms: FDC = first derivative correlation, C = correlation, A = absolute value, FDE = first derivative Euclidean. Error bars show maximum and minimum.

to distinguish between different samples and degrees of weathering.

3.2.1. *PCA of similar un-weathered samples*

For mean-centered normalized spectra from samples 1, 5, 15 and 16, the first two principal components (PCs) explained 56 and 10% of the variance of the data. The scores of the first two PCs are plotted in Fig. 4.

Different samples occupy different areas and the boundaries are clear for each sample type. Two replicates of sample 15 appear to be separated from the area containing the bulk of the spectra of this sample. It is not known why this occurred but it does imply that the method may be sensitive to spectral outliers.
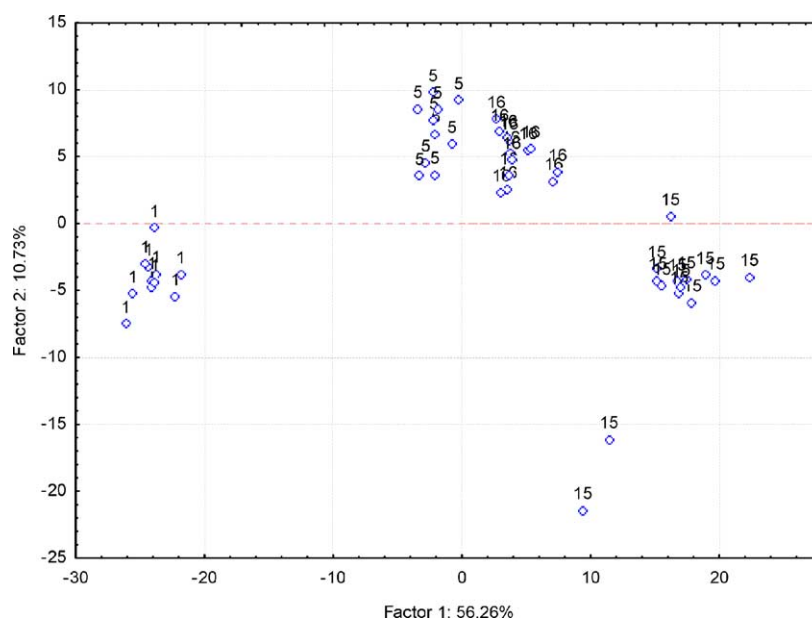

Fig. 4. Plot of the scores of the first two principal components for normalized spectral data from samples 1, 5, 15 and 16.

3.2.2. Feature selection

Four hundred and one fluorescence intensities were collected in one emission spectrum at 0.5 nm intervals between 300 and 500 nm. Some wavelengths have little information for classification, some of them are correlated with others and some of them are predominantly noise. The loadings of PCs can be used for selecting variables. The strategy used here was to select variables with projection distance greater than 0.8 from the origin of the variables on the PC1–PC2 plane. According to this criterion, 230 points remained for analysis by the PCA. The PCA was calculated again and similar results were obtained. Although PCA will usually relegate non-informative data to later PCs, feature selection both cuts down the number of data and does improve the discrimination and robustness of the PCA. This has proved useful in implementation of the method in a real-world application at the NSW EPA.

3.3. Weathering

The weathering process is complex and unpredictable. Each sample has a unique weathering pattern, which makes attempts to both characterize the oil and the exact extent of weathering impossible. However, weathering is known to follow a certain course, for example, with lighter components being lost before heavier ones, and so it may be possible to draw some conclusions as to oil type and weathering with suitable calibration and analysis. If the natural weathering processes can be experimentally mimicked, results from these may be compared with those from a spill (see Fig. 5).

The question therefore arises about the ability of any algorithm to pick a weathered sample from an un-weathered
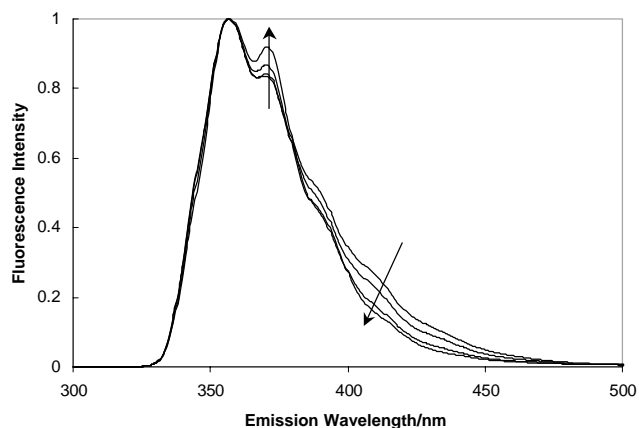

Fig. 5. Fluorescence spectra of sample 5 at different degrees of weathering. Arrows indicate direction of increasing weathering at 0, 2, 7 and 15 days.

or differently weathered, but different sample. The effect of weathering on the fluorescence spectrum clearly follows a progression from the un-weathered sample with time. A plot of the first principal component of the four similar samples weathered under the same conditions shows a correlation with the extent of weathering (Fig. 6). Inspection of other principal components showed no consistent trend with extent of weathering. Presumably the effect of weathering has the greatest influence on the variance of the samples and is therefore described in PC1. For these fresh oils with similar spectra, weathering has a similar effect on them. Given a known sample and prior information about the effects of weathering, it would be possible to determine the extent of weathering.
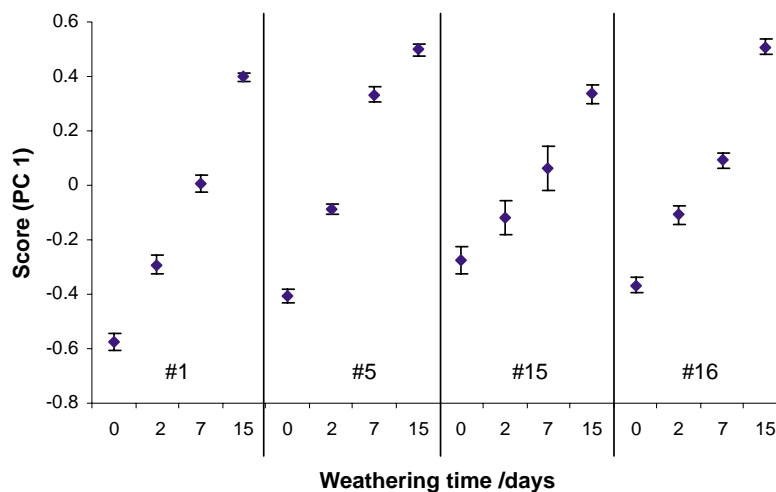

Fig. 6. The scores of the first principal components of fluorescence spectra of weathered samples of samples 1, 5, 15, 16 plotted against extent of weathering. The points are the means of a number (10–14) of test portions of the same oil and the bars are the range of values.

4. Conclusion

Using point-to-point spectral matching algorithms, fluorescence spectra of different samples can be distinguished quantitatively. Within all similarity indexes investigated in this work, first derivative Euclidean is the most robust for its ability to match extremely similar spectra, and distinguish between dissimilar spectra. However, the point-to-point spectral matching algorithms have limitations for distinguishing between the spectra of different oils that are of the same type. For these spectra, PCA was shown to discriminate individual samples of the same type of oil. To use PCA, however, a set of spectra of known origin is needed to provide the range in which the target spectra are embedded. Weathering is tracked excellently by PCA, and may be used with different samples of the same type of oil. Few exceptions have been observed when comparing different oils with the same degree of weathering. This demonstrates that multivariate data analysis methods like PCA can be used to characterize the same type oils and their weathered derivatives. It is possible to use artificial weathering to create data to classify oil spills against.

References

[1] D.S. Page, J.C. Foster, P.M. Fickett, E.S. Gilfillan, Mar. Pollut. Bull. 19 (1988) 107.
[2] Z. Wang, K. Li, M. Fingas, L. Sigouin, L. Menard, J. Chromatogr. A 971 (2002) 173.
[3] S.A. Stout, A.D. Uhler, K.J. McCarthy, Environ. Foren. 2 (2001) 87.
[4] U. Frank, Toxicol. Environ. Chem. Rev. 2 (1978) 163.
[5] J.B. Lloyd, J. Foren. Sci. Soc. 11 (1971) 83.
[6] A.G. Ryder, Appl. Spectr. 56 (2002) 107.
[7] D Patra, A.K. Mishra, Anal. Bioanal. Chem. 373 (2002) 304.
[8] D. Eastwood, S.H. Fortier, M.S. Hendrick, vol. 10, American Laboratory, Shelton, CT, United States 10 (1975) 45.
[9] G.S. Douglas, E.H. Owens, J. Hardenstine, R.C. Prince, Spill Sci. Technol. Bull. 7 (2002) 135.
[10] A.O. Barakat, Y. Qian, M. Kim, M.C. Kennicutt, Environ. Int. 27 (2001) 291.
[11] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, et al., Handbook of Chemometrics and Qualimetrics, Part A, 1997.
[12] ASTM, D 3650-93 Standard Test Method for Comparison of Waterborne Petroleum Oils by Fluorescence Analysis, American Society for Testing and Materials, Philadelphia, 1993.