

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/8496506>

Integrated Methodology for Forensic Oil Spill Identification

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · JUNE 2004

Impact Factor: 5.33 · DOI: 10.1021/es035261y · Source: PubMed

CITATIONS

67

READS

73

5 AUTHORS, INCLUDING:



Jan H Christensen

University of Copenhagen

61 PUBLICATIONS 1,108 CITATIONS

SEE PROFILE



Asger B Hansen

Haldor Topsøe

114 PUBLICATIONS 1,070 CITATIONS

SEE PROFILE



Giorgio Tomasi

University of Copenhagen

27 PUBLICATIONS 1,260 CITATIONS

SEE PROFILE



John Mortensen

Roskilde University

17 PUBLICATIONS 471 CITATIONS

SEE PROFILE

Integrated Methodology for Forensic Oil Spill Identification

JAN H. CHRISTENSEN,^{*,†,‡}
 ASGER B. HANSEN,[†] GIORGIO TOMASI,[§]
 JOHN MORTENSEN,[‡] AND
 OLE ANDERSEN[‡]

*Department of Environmental Chemistry and Microbiology,
 National Environmental Research Institute, Frederiksborgvej
 399, P.O. Box 358, 4000 Roskilde, Denmark, Department of
 Life Sciences and Chemistry, Roskilde University,
 Universitetsvej 1, P.O. Box 260, 4000 Roskilde, Denmark, and
 Department of Food Science, The Royal Veterinary &
 Agricultural University, Frederiksberg, Denmark*

A new integrated methodology for forensic oil spill identification is presented. It consists of GC–MS analysis, chromatographic data processing, variable-outlier detection, multivariate data analysis, estimation of uncertainties, and statistical evaluation. The methodology was tested on four groups of diagnostic ratios composed of petroleum biomarkers and ratios within homologous PAH categories. Principal component analysis (PCA) was employed and enabled the simultaneous analysis of many diagnostic ratios. Weathering was taken into account by considering the sampling uncertainties estimated from replicate spill samples. Statistical evaluation ensured an objective matching of oil spill samples with suspected source oils as well as classification into positive match, probable match, and nonmatch. The data analysis is further refined if two or more source oils are classified as probable match by using weighted least squares fitting of the principal components, local PCA models, and additional information relevant to the spill case. The methodology correctly identified the source of two spill samples (i.e., crude oils from Oseberg East and Oseberg Field Centre) and distinguished them from closely related source oils.

Introduction

Waterborne oil spills of unknown origin often occur in rivers, in open waters, and in coastal waterways. These spills range from the continuous leakage from land sources and illegal tank washings at sea to larger spill accidents. The liability associated with oil released into the environment warrants a comprehensive chemical characterization to defensibly determine the oil source(s), distinguish spilled oil from background hydrocarbons, and assess the impact on the ecosystem. Consequently, defensible differentiation and characterization of spilled oil is a critical part of many oil spill assessments.

Chemical fingerprinting is an important selection of techniques for solving such liability issues. Gas chromatography–flame ionization detection (GC–FID) is typically used for screening followed by a more comprehensive chemical

analysis by gas chromatography–mass spectrometry (GC–MS) (1, 2). GC–MS is the standard analytical technique for oil chemical fingerprinting, because it can resolve a broad range of petroleum biomarkers and polycyclic aromatic hydrocarbons (PAHs) including NSO compounds and because of the low cost of quadrupole instruments. Other instrumental techniques include metastable ion reaction monitoring GC–MS–MS (3), two-dimensional gas chromatography (GC–GC) (4), gas chromatography–isotope ratio mass spectrometry (GC–IRMS) (5), and excitation–emission fluorescence spectroscopy combined with multivariate data analysis (6).

Data analysis is a vital part of chemical fingerprinting, and a wide range of biomarkers are used in this respect. Particular attention has been given to tri-pentacyclic triterpanes and steranes, because they are indicators of source, maturation, and in-reservoir weathering and biodegradation (7), are found in high concentrations in most oils (7), and are recalcitrant when released into the environment (2). Petrogenic PAHs are dominated almost exclusively by the C₁–C₄ alkylated homologues of the parent compounds (in particular, naphthalene, phenanthrene, dibenzothiophene, fluorene, and chrysene) (8). The effect of short-term weathering processes (mainly evaporation and water washing) can be considerable on the distribution of C₀–C₄ homologues but is limited on the distribution of PAH isomers within a homologue category (2, 9). Consequently, it can be kept to a minimum by focusing the data analysis on ratios between concentrations, or peak areas, of biomarkers and of PAH isomers within a homologue category (diagnostic ratios). Conversely, the distribution of biomarkers and of PAHs within a homologue category can be altered by biodegradation because the individual isomers have varying susceptibility to microbial degradation (10).

Developments in oil fingerprinting procedures during the 1990s have concentrated on analytical techniques and on the application of new diagnostic ratios. Less attention has been paid to data analysis, which is often limited to visual comparison of ion chromatograms, bar charts of alkanes and PAH concentrations, and double plots of diagnostic ratios (8, 11). Multivariate methods have been used for data analysis in organic geochemistry since the 1980s (12, 13) and have only recently been adopted for oil spill identification (2, 14, 15). The advantages of such methods as compared to univariate ones are manifold; in particular, they allow for simultaneous analysis of a vast number of correlated variables by proper handling of redundant information.

In this paper, we present an integrated methodology for forensic oil spill identification based on GC–MS fingerprinting, multivariate data analysis, and statistical comparisons of the chemical composition of oil spill samples with that of source oils. The paper is organized in two sections: first, the methodology is outlined, focusing on multivariate data analysis and statistical evaluation; then it is tested on spill samples and source oils from a Nordtest Round-Robin oil spill exercise (16).

Methodology

Forensic oil spill identification deals with the process of defensibly linking spilled oil with its generic source. True identity in this context would imply all data to be identical, but as mineral oils consist of thousands of individual compounds and their composition are affected by weathering processes, this is an impossible task. Hence, the rationale of forensic oil spill identification is that identity prevails if no significant chemical differences can be demonstrated by

* Corresponding author telephone: +45-46301200; fax: +45-46301114; e-mail: jch@dmu.dk.

[†] National Environmental Research Institute.

[‡] Roskilde University.

[§] The Royal Veterinary & Agricultural University.

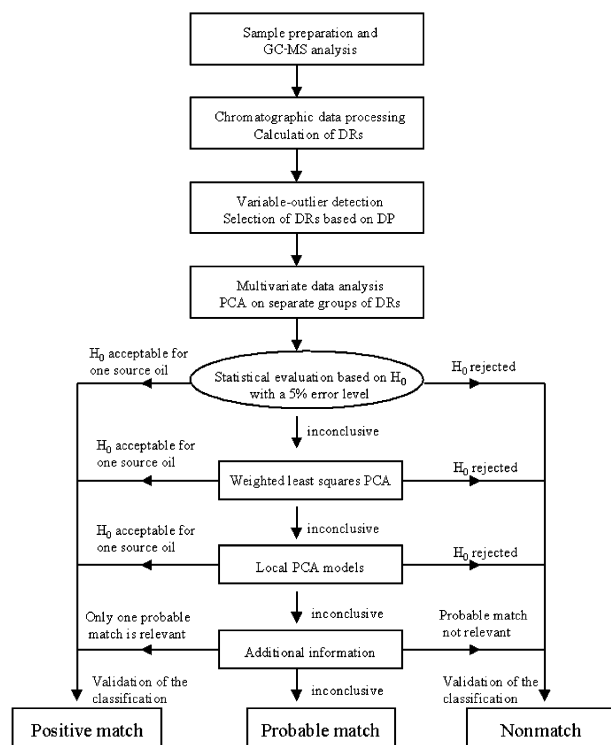


FIGURE 1. Flowchart for the forensic oil spill identification methodology.

focusing on recalcitrant and highly discriminating compounds. By pointing out significant differences between such compounds instead of proving an all-encompassing similarity, the evaluation of identity becomes viable and can be tested statistically. Our methodology complies with this overall concept; first, a vast number of discriminating diagnostic ratios are analyzed by principal component analysis (PCA). Subsequently, the spilled oil is compared statistically with source oils and identity prevails if no statistically significant differences can be observed, and this does not hold for other source oils. A flowchart of the methodology is shown in Figure 1.

Diagnostic Ratios. Throughout the data analysis, we use diagnostic ratios as defined in eq 1. Externally normalized ratios refer to recalcitrant compounds or to hydrocarbons within a homologue category with analogous physicochemical properties (e.g., boiling point and solubility). Ratios of the type $a_n/(a_n + a_{n^*})$ range between 0 (when compound n is absent) and 1 (when compound n^* is absent) and were preferred to a_n/a_{n^*} to circumvent problems with values rapidly approaching infinity.

$$DR^N = \frac{DR^S}{DR^S + DR^R} \quad (1)$$

where

$$DR^S = a_n^S / (a_n^S + a_{n^*}^S) \text{ (source or spill sample)}$$

$$DR^R = a_n^R / (a_n^R + a_{n^*}^R) \text{ (reference oil)}$$

and a_n^S is the area of peak n in an oil sample and a_n^R is the area of the corresponding peak in the reference. All diagnostic ratios refer to recalcitrant compounds or to hydrocarbons within a homologue category with analogous physicochemical properties (e.g., boiling point and solubility). Ratios of the type $a_n/(a_n + a_{n^*})$ range between 0 (when compound n is absent) and 1 (when compound n^* is absent) and were preferred to a_n/a_{n^*} to circumvent problems with values rapidly approaching infinity.

External normalization systematically reduces the influence of the measurement errors. For example, the relative analytical standard deviation (RSD_A) of the 137 diagnostic

ratios used in the case study ranged between 0.05% and 3.2% with normalization and between 0.09% and 5.1% without normalization. RSD_A is the random error arising from the chemical analysis, while the relative sampling standard deviation (RSD_S) is the combined random errors from the chemical analysis as well as the sample variation. The latter includes the effects of sample heterogeneity and weathering.

Chromatographic Data Processing. Multicomponent analysis is time-consuming and costly without automated chromatographic data processing. A semi-automated procedure for processing GC-MS data has been developed in our laboratory to provide fast and reproducible transformation of chromatographic data into diagnostic ratios (Christensen et al., unpublished). It consists of forming a database of chemical compounds, retention time alignment, estimation of the chromatographic background, iterative simplex estimation of chromatographic peak parameters assuming a Gaussian peak shape (peak width and peak height), and calculation of diagnostic ratios and their uncertainties. The retention time alignment, in particular, relies on frequent analyses of a laboratory reference oil.

Variable-Outlier Detection. Some ratios are heavily affected by measurement errors as well as weathering and biodegradation, thus a proper variable selection is important in order to keep the uncertainties to a minimum and to yield reliable results. However, when such selection is performed, there is a risk of omitting valuable fingerprinting information. Hence, the DRs to retain for the analysis were established based on the ability of the PCA to distinguish oil samples from different sources (the "resolution power"). Specifically, this was assessed by visual inspection of the scores of PCA models leaving out an increasing number of DRs starting with those with the lowest diagnostic power (DP). DP is defined as the relative standard deviation of a diagnostic ratio in oils with different origin (RSD_V) divided by RSD_S (eq 2):

$$DP = \frac{RSD_V}{RSD_S} \quad (2)$$

If RSD_S is unavailable, RSD_A can be used instead, although it disregards sample heterogeneity and weathering processes. Furthermore, the distribution within each DR^N for replicate analysis of the reference oil and replicate spill samples were evaluated from their skewness and kurtosis, and they were found to be approximately normally distributed, thus facilitating the use of parametric statistical methods.

Multivariate Data Analysis. PCA is a well-established method for data analysis (17). A data matrix (\mathbf{X}) of size $I \times J$ is decomposed into a product of a scores matrix, \mathbf{T} ($I \times K$), and a loadings matrix, \mathbf{P} ($J \times K$) plus a matrix of residuals \mathbf{E} ($I \times J$). Here, I , J , and K are the number of oil samples, diagnostic ratios, and principal components (PCs), respectively. To focus the analysis on between-samples variations, \mathbf{X} is typically column-wise mean-centered.

PCs are mutually orthogonal linear combinations of the original variables (diagnostic ratios) and are ordered according to the explained variation. Hence, the first principal component (PC1) describes the most prominent trend, and successive components (PC2, PC3, etc.) describe additional orthogonal trends in decreasing order of importance. In this manner, the redundant information is compressed into a small number of new variables explaining different phenomena/sources of variation. The optimal number of components in the oil spill case study was found by cross-validation (17) and verified by visual evaluation of scores and loadings (18).

Scores for new samples (\mathbf{T}_{new}) are computed by projecting the data (\mathbf{X}_{new}) on the loadings using

$$\mathbf{T}_{\text{new}} = \mathbf{X}_{\text{new}} \mathbf{P}(\mathbf{P}^T \mathbf{P})^{-1} = \mathbf{X}_{\text{new}} \mathbf{P} \quad (3)$$

PCA works well when the measurement errors are uncorrelated and their magnitudes are comparable. This is often not the case for GC-MS fingerprinting data, and modified PCA models adopting a weighted least squares (WLS) criterion can be more appropriate (19). In such models, variables with low weights are downscaled as compared to those with high ones. The MILES-PCA (maximum likelihood via iterative least-squares estimation-PCA) algorithm can be used for this purpose and also fits an offset that corresponds to mean centering in PCA. Scores for new oil samples in MILES-PCA are computed according to

$$\mathbf{T}_{\text{new}} = (\mathbf{X}_{\text{new}} - \mathbf{X}^{\text{off}}) \mathbf{Q} \mathbf{P}(\mathbf{P}^T \mathbf{Q} \mathbf{P})^{-1} \quad (4)$$

where \mathbf{X}_{new} is the data matrix consisting of new oil samples and \mathbf{X}^{off} is the offset. \mathbf{Q} (the weights) is a diagonal matrix of size $J \times J$, assuming uncorrelated measurement errors holding the inverses of RSDs. The data analysis was performed in MATLAB 6.5 (The MathWorks). The MILES-PCA algorithm can be downloaded from www.models.kvl.dk.

Estimation of Uncertainty. We estimated the sampling uncertainties in a two-step procedure assuming that the RSDs are consistent between oils. First, RSDs are calculated from replicate spill samples; then, they are used to estimate the standard deviation s_{ij} of the j th diagnostic ratio in the i th oil sample (eq 5):

$$s_{ij} \approx x_{ij} \text{RSD}_S^j \quad (5)$$

where x_{ij} is the value of the j th diagnostic ratio in the i th oil sample and RSD_S^j indicates the relative sampling standard deviation for the j th diagnostic ratio. The standard deviations of the scores (s_{ik}) can be calculated by eq 6, because least squares and weighted least squares PCA represent a linear transformation of the original data:

$$s_{ik} = \sqrt{\sum_{j=1}^J s_{ij}^2 p_{jk}^2} \quad (\text{PCA})$$

$$s_{ik} = \sqrt{\sum_{j=1}^J s_{ij}^2 \tilde{p}_{jk}^2} \quad \tilde{\mathbf{P}} = \mathbf{Q} \mathbf{P}(\mathbf{P}^T \mathbf{Q} \mathbf{P})^{-1} \quad (\text{MILES-PCA}) \quad (6)$$

where p_{jk} and \tilde{p}_{jk} are the loading coefficients of the j th diagnostic ratio in the k th principal component from PCA and MILES-PCA, respectively. Another procedure for estimating the sampling standard deviations on the scores relies on replicate samples of all considered oils. However, such an approach requires replicate source and spill samples rarely available in spill cases.

Statistical Evaluation. Each group of diagnostic ratios is analyzed by PCA, and the scores of source oils and spilled oils are compared for the PCs describing systematic variation. The statistical evaluation is based on the overall null hypothesis (H_0) that the spilled oil and the tested source oil are identical. The retained PCs in all models are tested independently using the inequality in eq 7a accepting a 5% error level ($\alpha = 0.05$). If the inequality is false in at least one of these tests, the overall H_0 is rejected and the tested source

oil is “beyond reasonable doubt” not the source of the spill:

$$\frac{|\bar{t}_k^{(\text{spill})} - \bar{t}_k^{(\text{source})}|}{s_k^{(\text{pooled})} \sqrt{(1/n_{\text{spill}}) + (1/n_{\text{source}})}} \leq q_{\alpha, \text{df}} \quad (7a)$$

where

$$s_k^{(\text{pooled})} = \sqrt{\frac{s_k^{(\text{spill})^2} + s_k^{(\text{source})^2}}{2}} \quad (7b)$$

and n_{spill} and n_{source} are the number of replicates used to calculate the mean scores along the k th principal component of the spilled oil ($\bar{t}_k^{(\text{spill})}$) and a source oil ($\bar{t}_k^{(\text{source})}$). $q_{\alpha, \text{df}}$ is the α -quantile from t -student's distribution where df is degrees of freedom and $s_k^{(\text{pooled})}$ is the pooled standard deviation for the k th PC. In this study, it was possible to calculate $s_k^{(\text{pooled})}$ from the estimated standard deviation of the scores for the spill ($s_k^{(\text{spill})}$) and for the tested source oil ($s_k^{(\text{source})}$) using eq 7b because the df were equal for the two variances, both estimated from RSDs.

Since multiple comparisons are performed, the risk for an overall type 1 error increases (i.e., that H_0 is rejected when it is true). One way to compensate for this is to apply the Bonferroni correction to the α value (20):

$$\alpha = \frac{\alpha_{\text{overall}}}{K_{\text{all}}} \quad K_{\text{all}} = \sum_{g=1}^G K_g \quad (8)$$

where K_g is the optimal number of PCs for the g th group of diagnostic ratios, G is the total number of groups, α is the quantile for the individual comparisons, and α_{overall} is the one for the entire set of K_{all} comparisons.

Depending on the outcome of these multiple tests, there are three possible classifications for the single source oils with respect to the spilled oil:

Positive Match. H_0 is acceptable (5% error level) for the tested source oil and the spill sample, and H_0 is rejected for all other source oils in the data set.

Probable Match. H_0 is acceptable for the tested source oil and spill sample, but the same holds for other source oils.

Nonmatch. H_0 is rejected.

Refinements of the Data Analysis. Prerequisites for defensibly linking the spilled oil to a ship are that the data set contains all possible source oils from the suspected ship(s), that the spill is composed of a single source and that the RSDs have been adequately estimated. When these are fulfilled, the data analysis is completed when source oils are classified as positive match or nonmatch. However, when source oils are classified as probable match, the data analysis needs to be refined to determine whether the spilled oil originates from one of them.

The probability of yielding an inconclusive answer (probable match) is expected to increase with the sampling uncertainty, the size of the data set, and its heterogeneity. The effect of high sampling uncertainties may be reduced by using WLS fitting of the PCs thereby downscaling the influence of the most uncertain DRs.

When closely related source oils are present in a large data set, it is likely that major trends, which are represented by the first PCs, mask the differences between these oils. The components that describe these minor differences may not be included in the optimal PCA model as they represent a minimal variation as compared to the total (17). Furthermore, if the differences between closely related source oils are described by few components compared to K_{all} , the Bonferroni adjustment may become too large, leading to an increase in the number of probable matches and in the risk of a type 2 error (i.e., H_0 is accepted when it is false). To

ensure that small but important variation is not disregarded as well as to decrease the risk of a type 2 error, PCA can be applied to a subset of source oils that lie close to the spilled oil (e.g., the probable matches) along the components retained in the original PCA model ("local" PCA). If results are still inconclusive, supplementary information of relevance to the spill case (e.g., backtracking) can be considered and/or additional diagnostic ratios added to the analysis.

Oil Spill Case Study

The oil spill identification methodology was tested on oil samples (i.e., crude oils, refined products, oil mixtures, and spill samples) from oil spill cases during the last 10 years stored at the forensic oil spill laboratory, National Environmental Research Institute, Denmark. A subgroup of these oils is part of a Nordtest Round-Robin oil spill exercise (16), and the largest Danish oil spill to date (the Baltic Carrier oil spill, Grønsund, Denmark, 2001) (21). The Round-Robin exercise was an artificial oil spill scenario consisting of two spill samples and five suspected sources. The two spill samples were oil/water emulsions with approximately 75% water. The five source oils were crude oils from Oseberg East (source A), Oseberg South East (source B), Oseberg Field Centre (source C), and two heavy fuel oils: IF 180-Shell refinery (source D) and IF 180-Esso refinery (source E). Source A, source C, spill I, and spill II were analyzed in duplicate.

Baltic Carrier oil spill samples were randomly sampled 0–14 days after the accident from the area affected by the spill. Oil on stones or vegetation was extracted with dichloromethane, and the extracts were cleaned up through a funnel with glass wool and anhydrous sodium sulfate and diluted to a total oil concentration of approximately 2000 mg/L. The extracts were stored at -20°C until analysis.

Chemical Analysis. Oil samples were analyzed on a HP-6890 gas chromatograph (Agilent Technologies) equipped with a 60-m HP-5MS capillary column (0.25 mm i.d. \times 0.25 μm film) and interfaced to a HP-5973 quadrupole mass spectrometer (Agilent Technologies) operating in EI mode. The 48 mass fragments were analyzed in six groups of 14–15 ions using selected ion monitoring. Data from the following ions (m/z values) were used in the subsequent data analysis: 191 (triterpanes); 217 and 218 (steranes); 231 (triaromatic steroids); 178, 192, 206, 220, and 234 (C_0 – C_4 -phenanthrenes); 184, 198, 212, 226, and 240 (C_0 – C_4 -dibenzothiophenes); 202 and 216 (C_0 – C_1 -pyrenes/fluoranthenes); 180 (C_1 -fluorenes), and 242 (C_1 -chrysenes). In the analytical sequence, both a blank and a laboratory reference oil were analyzed between every five oil samples. The reference oil was a 1:1 mixture of Brent crude oil (North Sea crude) and bunker oil from the Baltic Carrier. Data from 47 source oils, 4 Round-Robin spill samples, 21 replicate references, and 24 Baltic Carrier spill samples were used in the case study. Source oils were used to calculate the PCA models; the Round-Robin spill samples were projected on these models and subsequently matched to the source oils; references were used for external normalization of diagnostic ratios; and the Baltic Carrier oil spill samples were used for estimating the sampling uncertainties.

Variables and Outlier Detection. Four groups of diagnostic ratios were used in the case study: biomarker ratios, PAH ratios, C_2 -phenanthrene ratios, and C_3 - and C_4 -naphthalene ratios. The identities of the ratios are of minor importance here since our focus is on the principals of the methodology. The group of biomarker ratios comprised 17 of those most frequently used for chemical fingerprinting (1, 2, 7, 8). Their RSDs, RSD_v and DP are listed in the Supporting Information. Since none of the ratios were influenced by large measurement errors or weathering, they were all selected for analysis. The PAH group comprised 26 diagnostic ratios, mainly of single compounds within a homologue category of PAHs (data not shown). This was narrowed down

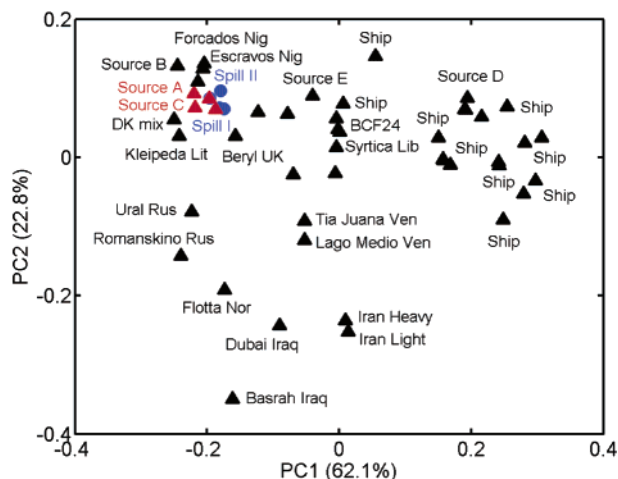


FIGURE 2. PCA with mean centering using 21 PAH ratios. PC1–PC4 describe 62.1%, 22.8%, 7.1%, and 3.3% of the variation in the ratios. The duplicates of source A and C (red triangles) and 43 additional source oils (black triangles) were used to calculate the PCA model, whereas duplicates of spills I and II (blue circles) were projected on the model using eq 3. Some legends have been removed to increase the readability.

to 21 by excluding the five diagnostic ratios with lowest DP (e.g., C_1 -chrysene ratios and ΣC_4 -phenanthrenes/ ΣC_4 -dibenzothiophenes influenced by large measurement errors). The distribution of C_2 -phenanthrenes (10 resolved peaks corresponding to 45 ratios), C_3 -naphthalenes (7 peaks and 21 ratios), and C_4 -naphthalenes (8 peaks and 28 ratios) are source and maturity indicators in oils (22, 23). We used all possible combinations of single compounds within each homologue category as potentially useful diagnostic ratios. The number of ratios was narrowed down from 45 and 49 to 30 and 20, respectively, again evaluated on the basis of the resolution power.

Principal Component Analysis. PCA with mean centering was applied separately to each group of diagnostic ratios using the 47 source oils. The K_g values were established via cross validation (two oils per validation segment) to 4 (biomarker ratios), 4 (PAH ratios), 3 (C_2 -phenanthrene ratios), and 3 (C_3 - and C_4 -naphthalene ratios), respectively. The root-mean-squared error in cross validation showed a sharp bend at the optimal number in all four cases. These choices were further verified by evaluating the loading and scores. For example, closely related oils from the Middle East (Basrah, Iraq, Iran Heavy, Iran Light, and Dubai Iraq), Venezuela (Lago Medio and Tia Juana), Russia (Romanskino and Ural), and Nigeria (Forcados and Escravos) form four separate clusters in a scores plot of PC1 versus PC2 calculated from the 21 PAH ratios (Figure 2). Furthermore, the Round-Robin spill samples are clustered with three of the suspected source oils (A–C). These oils can only be separated along PC4 (not shown) that describes a small percentage (3.3%) of the total variation in the data set.

Statistical Evaluation. Source oils were compared to spills I and II by testing the inequality in eq 7a. The tests comprised 14 individual comparisons per H_0 using 23 df for both $s_k^{(\text{spill})}$ and $s_k^{(\text{source})}$ ($\alpha_{\text{overall}} = 0.05$, $\alpha = 3.57 \times 10^{-3}$, $q_{\alpha,46} = 2.816$). The statistical analysis showed that H_0 was acceptable for source C and spill II and rejected for all other comparisons of source oils and spills. Consequently, source C is classified as positive match, and all other potential sources are classified as nonmatches. This classification was confirmed after the Round-Robin test since spill II is an artificially weathered crude oil from Oseberg Field Centre (16).

Only nonmatches were found for spill I when all the groups of diagnostic ratios were considered. However, leaving

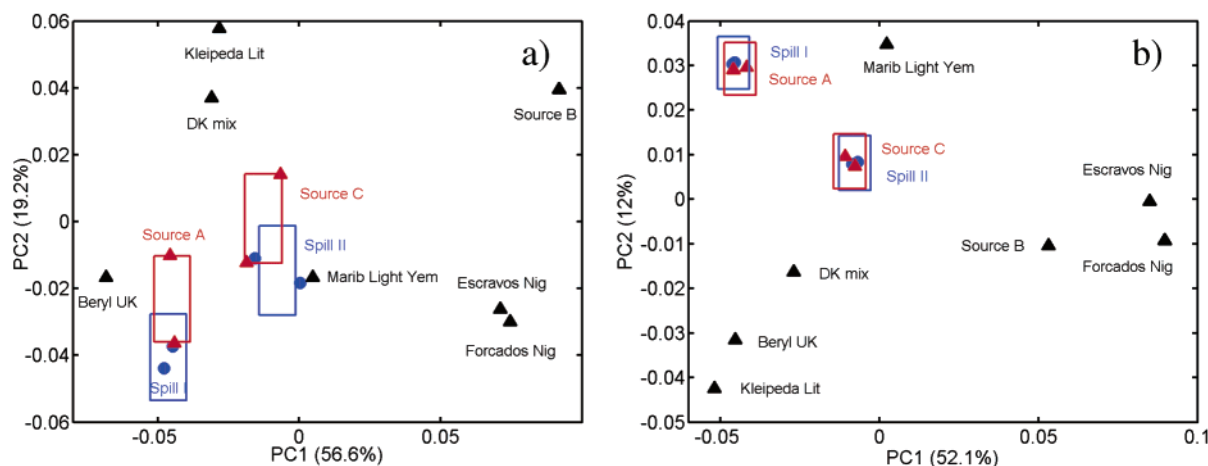


FIGURE 3. Comparison of the resolution power of (a) least squares and (b) weighted least squares PCA. The duplicates of sources A and C (red triangles), 7 additional source oils (black triangles) and 21 diagnostic PAH ratios were used in both analyses. The duplicate analyses of spills I and II (blue dots) were projected on the model. Rectangles illustrate H_0 for source A and spill I as well as source C and spill II, graphically allowing a 5% error level ($\alpha = 2.50 \times 10^{-2}$, $q_{\alpha,46} = 2.013$).

C_2 -phenanthrene ratios out of the statistical analysis yielded a positive match for spill I and source A. This was again confirmed by the Round-Robin test since spill I is a weathered crude oil from Oseberg East. Adequate estimations of the sampling uncertainties are a prerequisite for the statistical analysis. This is most likely not fulfilled for the C_2 -phenanthrene ratios, as it appears that the uncertainties have been clearly underestimated for the Round-Robin spill samples (RSD_s between 0.15% and 0.54%). Consequently, C_2 -phenanthrenes were not taken into account in the classification of source oils.

Refinements of the Data Analysis. The data set used in the case study was small, and positive matches were found by the initial PCA. However, as the data set becomes larger, the risk that oils are classified as probable matches increases. In such cases, the analysis can be refined by using the WLS approach and by focusing the PCA on a subset of source oils including all the probable matches.

Figure 3a,b shows the scores plots (PC1 vs PC2) from a local PCA using 21 PAH ratios (RSD_s varying between 0.08% and 3.3%) with least squares and WLS fitting of the PC model. The optimal number of PCs for this model was 2, and thus the statistical test of eq 7a can easily be illustrated graphically. In this study, $\bar{t}_k^{(spill)}$ and $\bar{t}_k^{(source)}$ were calculated from the two replicates of sources A and C and spills I and II. Consequently, for a test based solely on the PAH ratios (two comparisons), H_0 is acceptable if rectangles centered at $\bar{t}_k^{(spill)}$ and $\bar{t}_k^{(source)}$ and with side lengths of the confidence limit ($q_{\alpha,46} \times s_k^{(pooled)}$) overlap ($\alpha = 2.50 \times 10^{-2}$, $q_{\alpha,46} = 2.013$).

This is the case for both comparisons, but the resolution power of WLS (Figure 3b) is larger than for least squares (Figure 3a). Specifically, source C and Marib Light-Yem were both probable matches for spill II when using standard PCA but were positive match and nonmatch with WLS. Conversely, the two methods performed almost identically on ratios with RSD_s of the same magnitude (e.g., C_2 -phenanthrene ratios).

The refined procedure including WLS and local modeling was applied to all four groups of diagnostic ratios, although the C_2 -phenanthrene ratios were not considered in the classification. Figure 4 shows the results from these analyses, condensed into four plots (PC1 vs PC2) for (a) biomarker ratios, (b) PAH ratios, (c) C_2 -phenanthrene ratios, and (d) C_3 - and C_4 -naphthalene ratios. This was possible since the systematic information was retained within the first two PCs in all four cases. The statistical evaluation, which comprised six individual comparisons for each H_0 ($\alpha = 8.33 \times 10^{-3}$, $q_{\alpha,46}$

$= 2.485$), yielded the same positive matches as did ordinary PCA on the whole data set.

These refinements also improved the interpretability of the classification of oil samples by graphics (i.e., scores and loadings plots). For example, the suspected underestimation of the uncertainties for C_2 -phenanthrene ratios can be confirmed from visual inspection of the scores plot of PC1 vs PC2 (Figure 4c). Source A and spill I as well as source C and spill II are located close on the plot, but the small uncertainties still results in a rejection of the H_0 for spill I and source A.

Discussion

Numerous techniques have been suggested for assigning oil spilled into the environment to its source(s) (1, 2, 14, 15, 24), and it seems generally accepted that not one single method can exclusively be used to resolve this (24). Source assignments are therefore often time-consuming and require forensic chemistry at the highest skill and experience. Most methods for identification of spilled oil often utilize only a fraction of the available chemical information and lack a statistical evaluation. The methodology outlined in Figure 1 integrates analytical and statistical procedures and is fast and objective. It includes (i) sample preparation and GC-MS analysis, (ii) chromatographic data processing, (iii) variable-outlier detection based on DP, (iv) multivariate data analysis, (v) estimation of uncertainties, and (vi) statistical evaluation.

Diagnostic ratios are calculated using a semiquantitative procedure. Any combination of compounds can be calculated and selected based on their DP, which takes weathering into account. This selection need not be particularly refined, because the use of the WLS approach to some extent can account for the large uncertainties of some ratios.

This scheme is at variance with other fingerprinting methodologies where diagnostic ratios are selected based on a priori geochemical knowledge (e.g., source input, maturity and in-reservoir weathering, and biodegradation of specific markers) (1, 8). Our approach is less restrictive in this respect and can be used to identify ratios feasible for fingerprinting without prior assumptions. This is possible because ratios that contribute the most to a PC (i.e., those associated with large negative or positive loading coefficients) are also the main cause for the separation of oil samples in scores plots. Such ratios can be used to confirm the matches observed in the statistical analysis, thus finding a chemical explanation for the results. For example, two 25-norhopane

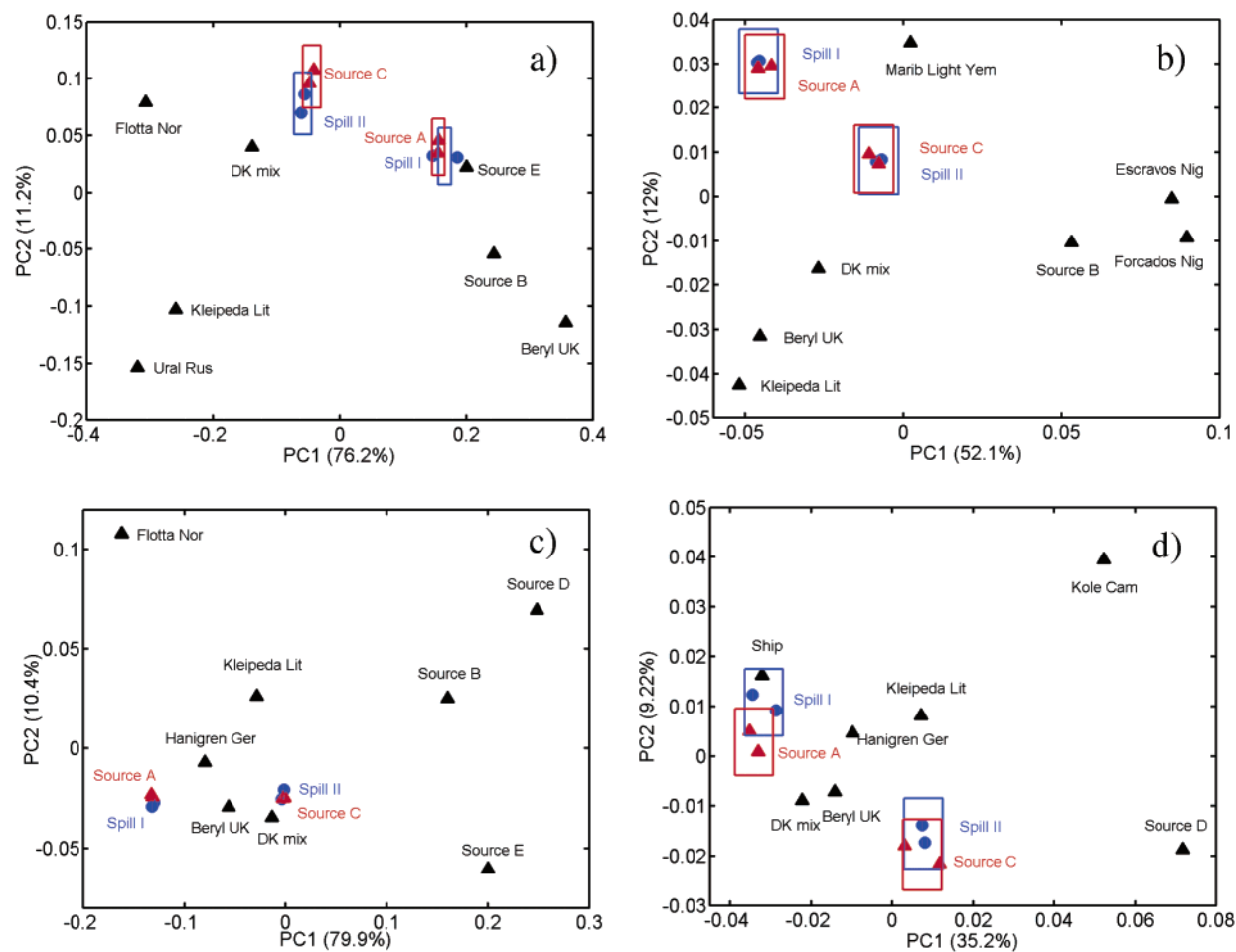


FIGURE 4. Refined data analysis using WLS and local modeling: (a) biomarker ratios using 17 ratios in the duplicates of sources A and C (red triangles), 7 additional source oils (black triangles), and the duplicates of spills I and II (blue dots); (b) PAH ratios (21 ratios and 11 source oils); (c) C_2 -phenanthrene ratios (30 ratios in 12 source oils); and (d) C_3 - and C_4 -naphthalene ratios (20 ratios in 11 source oils). Rectangles illustrate H_0 for source A and spill I as well as source C and spill II ($\alpha = 8.33 \times 10^{-3}$, $q_{\alpha,46} = 2.485$).

ratios ($17\alpha,21\beta$ -25-norhopane/($17\alpha,21\beta$ -25-norhopane + $17\alpha,21\beta$ -norhopane + $17\alpha,21\beta$ -hopane) and $17\alpha,21\beta$ -25-norhopane/($17\alpha,21\beta$ -25-norhopane + 18α -22,29,30-trisnorhopane + 17α -22,29,30-trisnorhopane)) have large coefficients in PC1 for biomarker ratios, which separates the three closely related crude oils from Oseberg East (source A), Oseberg South East (source B), and Oseberg Field Centre (source C). Faksness et al. (16) also found the relative amount of 25-norhopane to be particularly diagnostic with respect to the separation of these oils.

Furthermore, our methodology differs from standard methods for calculating diagnostic ratios since it employs a semiquantitative procedure. The use of semiquantitative analyses as opposed to quantitative ones has been a matter of discussion in the scientific community. For this data set, diagnostic ratios based on peak areas were normalized to a laboratory reference oil, which resulted in low analytical standard deviations (between 0.05% and 3.2%), comparable or smaller than those obtained by a quantitative method (2).

The statistical evaluation is a crucial part of the combined methodology that enables an objective and defensible classification of spill samples to suspected source oils. The criteria hereby proposed for positive match, probable match, and nonmatch is more strict compared to the criteria proposed by Daling et al. (1). A positive match is attained if the H_0 is acceptable only for one source oil, and thus it becomes increasingly difficult to obtain a positive match when the size of the database increases. This can be resolved by refining the data analysis, focusing on fewer closely related

source oils. Also, weighted least squares estimation of PC model can be useful since it decreases the effect of the sampling uncertainty and allows including potentially meaningful ratios that would otherwise be rejected in the variable-outlier detection if the choice were based solely on DP.

Weathering is considered implicitly in the statistical evaluation, dissimilarly from Daling et al. (1). In our methodology, a diagnostic ratio affected by weathering is either left out in the initial outlier detection or it is partially accounted for by the WLS approach. This is consistent with Stout et al. (2), who also implicitly considered weathering by variable selection prior to PCA. However, their approach does not include any statistical evaluation of the results.

The main assumptions applied in the combined methodology concern the estimation of the analytical and sampling uncertainties, where the latter are used for outlier detection and in the statistical evaluation. Whereas the analytical uncertainties of the individual ratios depend on, for example, the signal-to-noise ratio of peaks and the chromatographic quality, sampling uncertainties also depend on the sample heterogeneity and variations in the weathering state of the spill samples. Here, RSDs are calculated from Baltic Carrier oil spill samples exposed to weathering for 0–14 days and are used to estimate the absolute uncertainties in oil samples with a different chemical composition and degree of weathering. Consequently, there is a risk that these uncertainties are improperly estimated, as was indeed observed for the C_2 -phenanthrene ratios.

Underestimated uncertainties can lead to erroneous rejection of the H_0 such that identical oils are classified as nonmatch (e.g., C2-phenanthrenes). Conversely, overestimated ones could lead to erroneous acceptance of the H_0 , leading to a positive match or probable match for nonidentical oil samples. Hence, it is important to obtain good estimates of the uncertainties, and one should always try to obtain replicate analyses of source oils and spill samples to estimate the true analytical uncertainties. However, it may be difficult to obtain the true sampling uncertainties for individual oil samples since they can only be obtained from spill samples exposed to different degrees of weathering, which are rarely available in spill cases. Consequently, the sampling uncertainties can be substituted with the analytical ones if good estimates of the former cannot be obtained. However, it is then necessary to apply a priori knowledge and retain only ratios unaffected by weathering processes. Note that the latter approach does not account for sample heterogeneity, and it is likely that using the analytical uncertainties in this fashion would lead to a large number of false rejections of the null hypothesis.

Acknowledgments

The authors acknowledge Parmely Pritchard and Rasmus Bro for fruitful comments on the manuscript and Jørgen Avniskjold for technical assistance. This work was financed partly by the Department of Environmental Chemistry and Microbiology, National Environmental Research Institute, and partly by the Natural Sciences Research Council, Denmark.

Supporting Information Available

Two tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Literature Cited

- (1) Daling, P. S.; Faksness, L. G.; Hansen, A. B.; Stout, S. A. *Environ. Forensics* **2002**, *3*, 263–278.
- (2) Stout, S. A.; Uhler, A. D.; McCarthy, K. J. *Environ. Forensics* **2001**, *2*, 87–98.
- (3) Munoz, D.; Doumenq, P.; Guiliano, M.; Jacquot, F.; Scherrer, P.; Mille, G. *Talanta* **1997**, *45*, 1–12.
- (4) Frysinger, G. S.; Gaines, R. B.; Reddy, C. M. *Environ. Forensics* **2002**, *3*, 27–34.
- (5) Mansuy, L.; Philp, R. P.; Allen, J. *Environ. Sci. Technol.* **1997**, *31*, 3417–3425.
- (6) Beltran, J. L.; Ferrer, R.; Guiteras, J. *Anal. Chim. Acta* **1998**, *373*, 311–319.
- (7) Peters, K. E.; Moldowan, J. M. *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*; Prentice Hall: Englewood Cliffs, NJ, 1993.
- (8) Wang, Z. D.; Fingas, M.; Page, D. S. *J. Chromatogr. A* **1999**, *843*, 369–411.
- (9) Wang, Z. D.; Fingas, M. *Environ. Sci. Technol.* **1995**, *29*, 2842–2849.
- (10) Wang, Z. D.; Fingas, M.; Blenkinsopp, S.; Sergy, G.; Landriault, M.; Sigouin, L.; Foght, J.; Semple, K.; Westlake, D. W. S. *J. Chromatogr. A* **1998**, *809*, 89–107.
- (11) Bence, A. E.; Kvenvolden, K. A.; Kennicutt, M. C. *Org. Geochem.* **1996**, *24*, 7–42.
- (12) Telnaes, N.; Dahl, B. *Org. Geochem.* **1986**, *10*, 425–432.
- (13) Øygard, K.; Grahl-Nielsen, O.; Ulvøen, S. *Org. Geochem.* **1984**, *6*, 561–567.
- (14) Burns, W. A.; Mankiewicz, P. J.; Bence, A. E.; Page, D. S.; Parker, K. R. *Environ. Toxicol. Chem.* **1997**, *16*, 1119–1131.
- (15) Mudge, S. M. *Environ. Sci. Technol.* **2002**, *36*, 2354–2360.
- (16) Faksness, L. G.; Daling, P. S.; Hansen, A. B. *Environ. Forensics* **2002**, *3*, 279–291.
- (17) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer-Verlag: New York, 2002.
- (18) Martens, H.; Næs, T. *Multivariate Calibration*; John Wiley & Sons: Chichester, UK, 1996.
- (19) Bro, R.; Sidiropoulos, N. D.; Smilde, A. K. *J. Chemom.* **2002**, *16*, 387–400.
- (20) Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Some Important Hypothesis Tests. In *Handbook of Chemometrics and Quality Metrics: Part A*; Elsevier: Amsterdam, 1997; p 93.
- (21) Christensen, J. H. *Polycyclic Aromat. Compd.* **2002**, *22*, 703–714.
- (22) Budzinski, H.; Garrigues, P.; Connan, J.; Devillers, J.; Domine, D.; Radke, M.; Oudin, J. L. *Geochim. Cosmochim. Acta* **1995**, *59*, 2043–2056.
- (23) van Aarssen, B. G. K.; Bastow, T. P.; Alexander, R.; Kagi, R. I. *Org. Geochem.* **1999**, *30*, 1213–1227.
- (24) Wang, Z. D.; Fingas, M.; Sigouin, L. *Environ. Forensics* **2002**, *3*, 251–262.

Received for review November 12, 2003. Revised manuscript received February 20, 2004. Accepted March 3, 2004.

ES035261Y