7 Emerging CEN Methodology for Oil Spill Identification

Asger B. Hansen* (NERI, Denmark)

Per S. Daling, Liv-Guri Faksness, and Kristin R. Sörheim (SINTEF Chemistry, Norway)

Paul Kienhuis (RIZA, The Netherlands)

Rolf Duus (Standards Norway, Norway)

*Corresponding author: aha@dmu.dk

Since 1991 the existing Nordtest method on oil spill identification (Nordtest, 1991) has formed an important forensic "platform" in relation to oil spill identification, not only in the Scandinavian countries, but also in other European countries following its recommendation in and adoption to the Bonn Agreement[1] Counter Pollution Manual. Also countries outside Europe have adopted the concept and methodology for oil spill analysis and identification described in this Nordtest method. It basically incorporates a stepwise procedure including initial screening by GC/FID of all samples for characterization and to exclude obviously nonmatching candidate source samples. This step is followed by GC/MS fingerprinting of a few selected samples (spill and potentially matching candidate source samples) recording a suite of key or target petroleum compounds (biomarkers and polycyclic aromatic compounds, PACs). After comparing the GC/MS chromatograms in order to pinpoint possible differences, the conclusion according to the chemical analysis will either be identity (matching chromatograms) or nonidentity (nonmatching chromatograms). When evaluating the chromatograms (both GC/FID and GC/MS), weathering of the oil samples has to be taken into account as the criterion for identity excludes differences arising from weathering.

Experiences over the past 10–15 years with the existing Nordtest (1991) methodology for oil spill identification, however, has shown some need for improvements. For example, the evaluation of chromatograms was generally achieved by qualitative means (i.e., visual comparison) just as the influence of weathering was not always straightforward to interpret, facts that rendered the evaluation and final conclusions to be flawed by subjectivity and hence questionable. Besides, over this time period advances in both analytical and interpretive methods had opened the possibility to obtain more quantitative, objective, and defensible means for verification of the results. Thus in 2000, Nordtest[2] initiated Phase 1 of the

[1] The Bonn Agreement is a multilateral agreement by North Sea coastal states, which together with EU will offer: (1) mutual assistance and cooperation in combating pollution, and (2) surveillance as an aid to detecting and combating pollution and to prevent violations of antipollution regulations.

[2] Nordtest is an institution under the Nordic Council of Ministers acting as a joint Nordic body in the field of technical testing and standardization.

project called "Revision of the Nordtest Methodology for Oil Spill Identification" with participation of the national forensic oil spill laboratories in Denmark, Finland, Norway, Sweden, and the Battelle Memorial Institute (Duxbury), USA. Phase 1 (2000–2001) of that project included: (1) a review and assessment of recently published literature on oil spill identification and petroleum geochemistry, and (2) an update of the existing Nordtest (1991) method into a technically more robust and legally defensible oil spill identification methodology by introducing improved laboratory techniques for sample preparation and cleanup, chromatographic analysis, and data evaluation tools. The improved methodology was still based on tiered GC/FID screening and GC/MS fingerprinting procedures, but it also included more objective criteria for evaluation by introducing quantitative diagnostic ratios of key petroleum compounds that could be tested statistically and stricter analytical protocols including QA/QC criteria (Faksness et al., 2002a).

In Phase 2 of the project (2001–2002), the updated Nordtest methodology was evaluated through a Round Robin exercise arranged by SINTEF, Norway, and with the participation of 12 laboratories from 10 countries. Seven oil samples (two artificially weathered "spill" samples and five possible sources) were analyzed according to the recommended analytical protocols of the revised methodology. The Round Robin exercise was a "difficult case," because the two spill samples and three of the suspected sources were qualitatively similar and thereby quite highly correlated to one another. These samples were crude oils from the same oil field in the North Sea, but from different production wells. The Round Robin exercise, however, demonstrated the potential of the updated methodology as a strong technically defensible tool for oil spill identification due to its ability to distinguish between qualitatively similar oils from a spill and potential candidate source samples (Faksness et al., 2002b). Following the Round Robin exercise, the revised methodology has been implemented by several forensic laboratories in Europe and has been used in connection with recent major oil spills, e.g., the *Tricolor* (British Channel, December 14, 2002) and the *Prestige* (off the coast of Galicia, Spain, November 20, 2002).

In 2002, Nordtest proposed the revised methodology (Daling et al., 2002) as a new standard for oil spill identification to the European Committee for Standardization (CEN), which established a task force (CEN BT/TF 120) to evaluate the proposal and eventually prepare a new standard. The standardization process has been split into two work items: (1) sampling and (2) analytical methodology and interpretation of results, involving two working groups that accordingly should produce two guidelines (CEN Technical Reports) on oil spill identification.

7.1 Introduction

The objective of the emerging CEN methodology is to provide a forensic tool for the identification of waterborne oil by comparing samples from spills with those of suspected sources. This methodology, which has been based on a technical revision of the existing Nordtest method (Nordtest, 1991), should be capable of providing both administrative and legal support to the prosecution of an offender ("potential responsible party" — PRP) that has violated national or international regulations by illegally discharging mineral oil into the marine environment. The two working groups under the CEN task force BT/TF 120 have produced two draft guidelines (CEN Technical Reports) that describe the new methodology on oil spill identification — waterborne petroleum and petroleum products accordingly:

- Part 1 — Sampling (CEN, 2005)
- Part 2 — Analytical methodology and interpretation of results (CEN, 2006)

In 2004, RIZA organized a Round Robin (RR) exercise for oil spill identification. The first exercise (RR-2004) dealt with light fuel oil distillates (gas oils). Fifteen laboratories from nine countries participated in the study, and the results of this RR exercise (Kienhuis,

2004) have been taken into account for refining the CEN methodology. A second Round Robin exercise took place in 2005 (RR-2005) and involved bilge water samples. Again more than 10 oil spill laboratories from 9 countries participated (Dahlmann and Kienhuis, 2005), and the emerging CEN methodology now seems to have been implemented by many forensic oil spill laboratories in Europe.

In 2005, SINTEF (Norway) produced a standard oil mixture to facilitate the identification of target compounds and compound groups in oil samples. The mixture is a combination of three crude oils (from Russia, Sicily, and the North Sea) and a heavy bunker oil (IFO-180), and it contains all compounds mentioned in the CEN guidelines. The oil mixture can be obtained from SINTEF, Norway, together with all relevant and integrated chromatograms.

This chapter will describe Part 2 (CEN, 2006) of the emerging CEN methodology that covers the tiered analytical approach, the selected diagnostic target compounds, and the data treatment, whereas sampling techniques and handling of oil samples prior to their arrival at the forensic oil spill laboratory, as described in Part 1 (CEN, 2005), will not be covered here.

7.2 Scope of the CEN Methodology

The scope of the CEN methodology on oil spill identification is to provide a methodology to identify waterborne oils spilled in marine, estuarine, and aquatic environments based on detailed analytical and processing procedures for the comparison of samples from spills with those of suspected sources. When suspected sources are not available, the methodology may be used to characterize the spill as far as possible with respect to oil type and origin.

The methodology is restricted to petroleum and petroleum products containing a significant proportion of petroleum hydrocarbons with boiling points above 200°C. Examples are

- Crude oils
- Light refined products like diesel oils or gas oils
- Heavy refined products like heavy fuel oils, bunker oils, and vaccum residues
- Lubricating oils
- Mixtures of bilge and sludge samples

Still, while the general concepts described in the methodology have a limited applicability for some kerosenes and condensates, it may not be applicable for gasoline.

The described methodology is not intended for oil spilled to groundwater and soil. The chromatograms of oil extracted from soil and groundwater may contain reduced and/or additional compounds compared to the candidate source sample. Including such samples in the methodology would require additional extraction and cleanup methods together with an accounting of which compounds could possibly be reduced and/or which additional peaks could be expected, resulting in a diversion of the final conclusion. Such issues are beyond the scope of the CEN methodology. However, in such cases, including oil spilled to groundwater or soil where appropriate sample preparation and the potential for compound reductions and/or additions are appreciated, a positive match observed according to the CEN methodology would be valid.

7.3 Strategy for Identifying Oil Spills

When an oil spill has been observed, samples should be collected from the current spill and from potential responsible parties, such as suspected ships or other sources, in due course by appropiate authorized personnel. All samples should subsequently be sent either via an authorized "Sampling Coordinator" or directly to the analytical laboratory for forensic characterization and potentially for identification of the spill's source. In the context of the CEN methodology, the process of identifying the source of a spilled oil implies comparison of the chemical composition of the spilled oil with that of candidate source samples.

Conceptually, two results can be achieved in forensic oil spill investigation — "identity" and "nonidentity," depending on whether spill and candidate source samples are "identical" or

"nonidentical." "Identical" per se requires all measurable data to match exactly. This criterion is practically and technically impossible to fulfill, and therefore the definition of "identical" is rephrased in operational terms: two samples are identical beyond reasonable doubt if and only if (1) no significant differences in the recorded GC/FID and GC/MS data are observed (i.e., matching chromatograms) or (2) any observed differences are not genuine, but stem from changes introduced in the collected samples after the spill (e.g., due to weathering, contamination, mixing, or degradation). If these criteria are met, the condition of "identity" is achieved. Otherwise, samples are "nonidentical," and "nonidentity" is achieved. The task of looking for genuine significant differences in chemical composition by comparing chromatograms, instead of proving an all-encompassing match, is conceptually more logical and feasible to comply with. According to this definition, it is important to realize that only nonidentity based on genuinely significant differences between samples (i.e., nonmatching chromatograms) can be proved directly (rejection of the null hypothesis), whereas identity always must be based on the process of "ruling out" differences.

Identity must be tested by analyzing and comparing the detailed chemical composition of the selected samples by chemical fingerprinting of a suite of diagnostic or target petroleum compounds. If no or only insignificant differences (i.e., differences of chromatographic peaks being smaller than the analytical variance) are observed, identity could be concluded as being beyond reasonable doubt. On the other hand, if "true" differences (i.e., differences not related to changes in the chemical composition introduced after the spill, e.g., from weathering, etc.) are significant (i.e., larger than the analytical variance observed for the diagnostic compounds), nonidentity could be concluded. In practice, the process of comparing and evaluating chromatograms has been graduated in the CEN methodology by introducing four operational terms (i.e., positive match, probable match, inconclusive, and nonmatch) (cf. Section 7.4.6), to cover the comparison of samples. Only when a *positive match* has been obtained can *identity* be concluded as being beyond reasonable doubt.

In Europe, forensic oil spill identification is performed not only by laboratories that analyze oil samples on a regular basis, but also by laboratories that only compare samples a few times a year. Traditionally, the common practice has been to analyze oil samples qualitatively and compare the chromatograms and ion fragmentograms visually — as per the original Nordest (1991) protocol. The results of such comparisons depend heavily on the skill and experience of the analyst, and laboratories which rarely analyze oil samples may experience difficulties in reaching sound forensic conclusions. Therefore, the new CEN methodology has introduced the use of diagnostic ratios[3] as an additional and more objective and defensible tool for comparison. The criteria for selecting such ratios take into account the weathering and degradation behavior of the target petroleum compounds and their varying amount and composition in oils of different types and petrogenic origin (Faksness et al., 2002a). To reduce the analytical variance, ratios are preferably generated by using the peak area or height of target compounds, which preferably are recorded by the same *m/z* value and approximately within the same retention time window. The resulting ratios are successively compared using the repeatability limit ($\mathbf{r}_{95\%}$) as the test method. Optionally, diagnostic ratios can also be generated on the basis of chromatographic peaks that have been fully quantified by traditional quantitative chromatographic analysis. Strict QA/QC procedures should be observed when doing fully quantitative analyses.

The use of diagnostic ratios for comparison of oil samples is based on GC/MS data of 29 diagnostic ratios generated from a suite of alkylated polycyclic aromatic compounds

[3] Diagnostic ratios (DR) — ratios between the peak height or peak area of single compounds or groups of compounds selected for their diversity in the chemical composition in petroleum and petroleum products and their reported response to weathering and degradation processes.

(PACs) and petroleum biomarkers that are robust against weathering and that have been selected to cover genuine differences in oil samples and oil types. Whereas most of the ratios may be used when crude oils, bunker oils, and bilge samples are involved, only a limited number of ratios may be appropriate for lighter fuel oils (e.g., kerosene, paraffin, diesel, gas oil) because some of the higher-boiling compounds may not be present in such light-end refined products. In spill cases where the spilled oil has only been slightly exposed to weathering, three ratios involving acyclic isoprenoids (i.e., *n*-C17/pristane, *n*-C18/phytane, and pristane/phytane ratios) generated from GC/FID data and four ratios involving sesquiterpanes (in the *n*-C13 to *n*-C16 retention time window) generated from GC/MS data can also be included for comparison, provided a weathering check shows that these compounds have not been significantly influenced by weathering.

Before integrating (by area or height) the compounds required for generating the selected diagnostic ratios, a visual inspection of the relevant ion fragmentograms should be carried out in order to eliminate those target peaks not present in sufficient amount to fulfill the required signal-to-noise (S/N) criteria and hence not capable of generating robust diagnostic ratios. A visual comparison of the ion fragmentograms is also recommended to exclude obviously different samples and spot highly diagnostic compounds specific for the actual case (e.g., unusual biomarkers). After the comparison and evaluation of the diagnostic ratios using the repeatability limit as criteria, a second visual comparison of the ion fragmentograms of the relevant samples, one by one, should also be carried out to verify (ground truth) the conclusion.

7.4 Tiered Levels of Analysis and Data Treatment

7.4.1 Decision Chart for Identifying Oil Spills

The methodology for identification of oil spills is divided into three tiered levels of analytical procedures and data treatments according the decision or flowchart shown in Figure 7-1.

This operational flowchart guides the reader through the individual steps of the tiered procedures by linking each step of the procedure with the operation performed (analysis/evaluation) at the previous step and its resulting decision until a final conclusion regarding identity can be made. In the flowchart, squared gray boxes refer to operations to be performed, and rounded gray boxes refer to evaluations/conclusions to be made. The flowchart comprises the following steps:

- visual characterization of samples
- sample preparation and cleanup
- GC/FID screening of all samples — visual characterization/classification of oil types — weathering check — calculation of acyclic isoprenoid diagnostic ratios
- GC/MS fingerprinting of selected samples (spill(s) and candidate source(s)) — visual evaluation of chromatograms — calculation of diagnostic ratios — weathering check of PACs (optional)
- Comparison of diagnostic ratios (repeatability limit)
- Conclusion and reporting

7.4.2 Visual Characterization and Preparation/Cleanup of Oil Samples

When arriving at the analytical laboratory, all samples should be carefully examined and described with respect to type (e.g., spill or reference), matrix (e.g., water, sand, feather, etc.), amount, container (glass, plastic, etc.), and general condition. Eventually, all samples should be photographed to assist the visual description and document their condition at arrival at the laboratory. Especially, any sign of jeopardizing or "missing links" regarding the chain-of-custody should be reported.

The applied procedures for forensic identification of spilled oil must strictly observe that any manipulation performed during sample preparation and cleanup can alter its chemical composition and thus weaken the power of

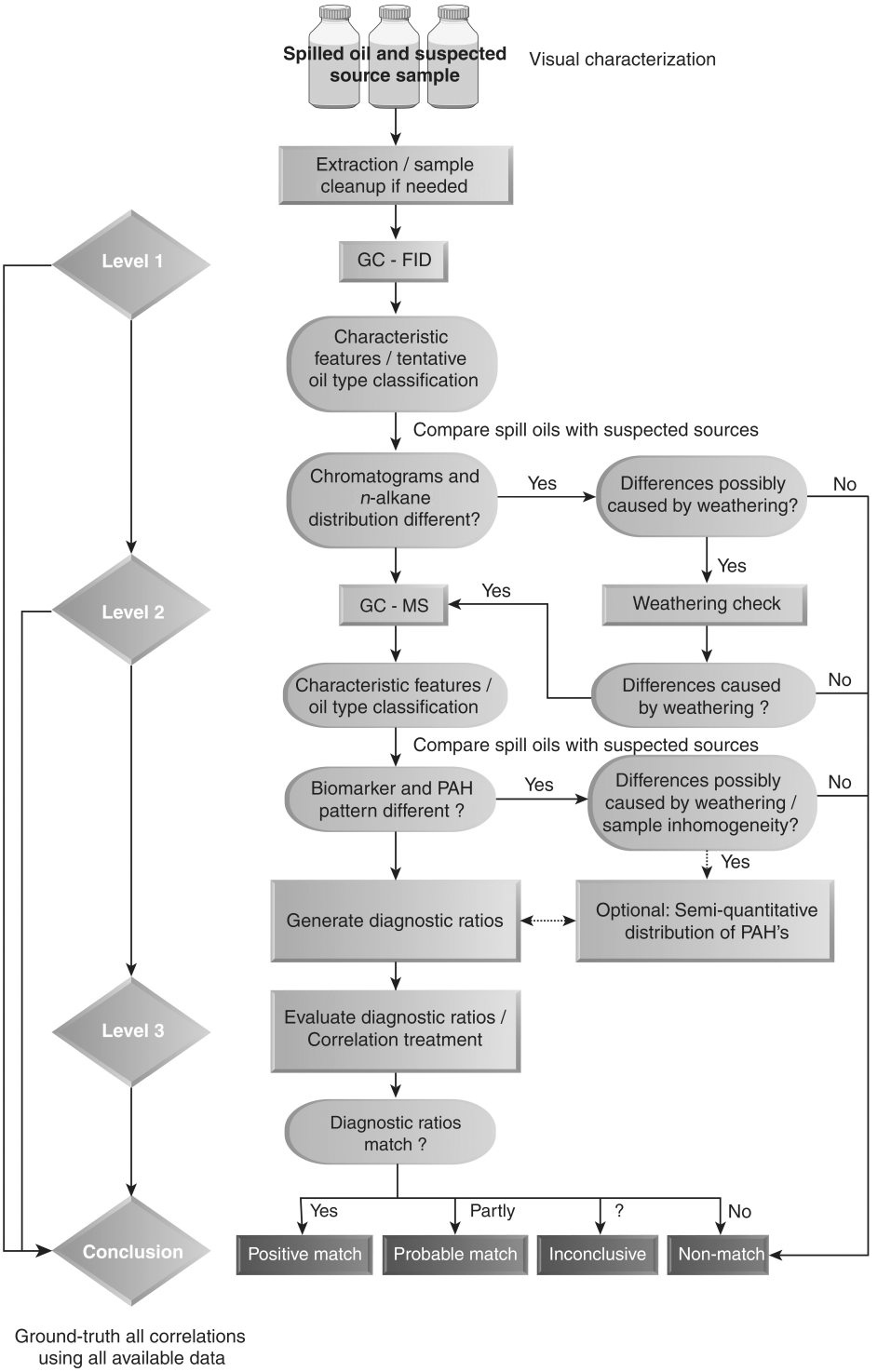

Figure 7-1 Decision/flowchart of the CEN oil spill identification methodology.

evidence. Therefore, sample preparation and cleanup should generally be restricted to a minimum, and in any case precisely stated in the final report. After removal of any pieces of debris (e.g., wood, fabric, feathers, etc.), preparation is generally performed by diluting the sample to appropriate concentrations for GC/FID and GC/MS analysis. For samples contaminated with polar compounds (e.g., bird feathers) or that contain a high amount of heavy components (e.g., heavy bunker oil) a simple cleanup on a silica/alumina column can be used. Descriptions of preparation and cleanup of oil samples are beyond the scope of this chapter, but more detailed procedures for the different sample types can be found in the CEN guidelines (CEN, 2006).

7.4.3 Level 1 — GC/FID Screening

After sample preparation, all samples (i.e., both samples from the oil spill and from suspected sources) are initially characterized by GC/FID screening. This will generally render a descriptive "picture" of the dominating petroleum hydrocarbons in the oil sample (e.g., the overall boiling range and prominence of individual resolved *n*-alkanes and major isoprenoids as illustrated in Figure 7-2). The GC/FID chromatograms also provide information on the weathering extent of the spilled oil and on any "characteristic features" or "contaminating" components present in the samples. Generally, the *n*-alkanes are the most characteristic and dominating peaks distributed regularly over the entire retention interval, except for extensively weathered oil samples, such as water samples collected from thin oil films (sheens), highly biodegraded crude oils (biodegradation in the reservoir), and certain refined petroleum products (like lubricating and hydraulic oils).

The GC/FID chromatograms also enable the calculation of diagnostic ratios derived from acyclic isoprenoids (diterpanes) like the *n*-C17/pristane, *n*-C18/phytane, and pristane/phytane ratios, unless the spill sample is too weathered, or the amount of these compounds is low relative to the UCM hump ("unresolved complex mixture"), as can often be observed in bunker oils. These ratios can also be indicative of biodegradation of the spilled oil, as they can monitor the effect of microbial degradation at the spill site by the preferential loss of *n*-alkane hydrocarbons compared to the isoprenoids. The acyclic isoprenoid ratios, however, are easily influenced by weathering and should therefore not be included in any evaluation without a thorough weathering check.

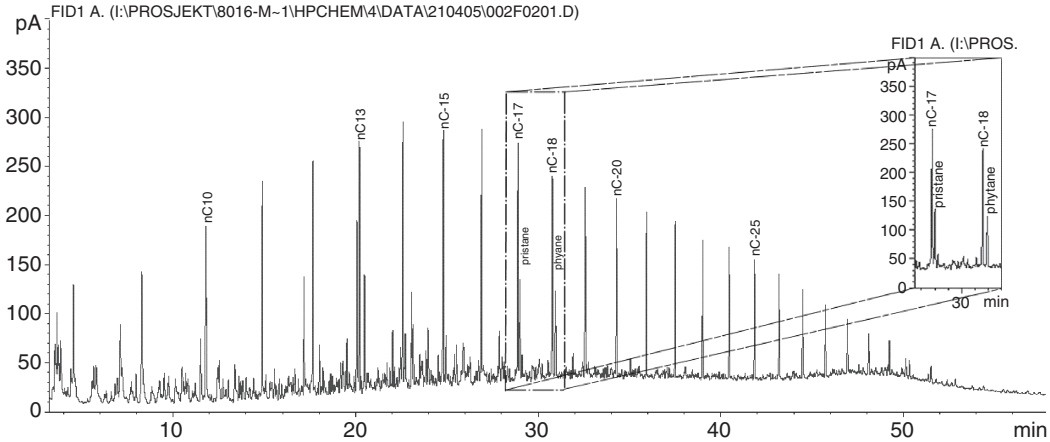

Figure 7-2 GC/FID chromatogram of a mildly weathered fuel oil displaying the dominating *n*-alkanes together with the acyclic isoprenoids, pristane and phytane (inserted).

7.4.3.1 Evaluation of Weathering

If the chromatograms of a spill and candidate source oil are different and the observed differences could possibly be caused by weathering, a "weathering check" is recommended. Figure 7-3A gives an example of a weathering check simply by overlaying chromatograms of a spill sample and a suspected source (in this case, gas oil). The normalization/manipulation of the chromatograms to a comparable attenuation can easily be done by expanding one of the chromatograms vertically until the peaks in the *n*-C20 to *n*-C24 range are of the same height. Alternatively, the weathering check can also be obtained by integrating (by peak height or area) a homologous series (e.g., the *n*-alkanes) in the GC/FID chromatograms, then normalizing the peaks to nonweathered compounds (e.g., the mean of *n*-C20 to *n*-C24), and eventually displaying the normalized peaks in a bar chart (e.g., prepared by Excel™) for comparison of the spill sample with a nonweathered source sample; see Figure 7-3B for an example of such a normalized *n*-alkane bar chart of two gas oil samples. The bar chart presentation of the *n*-alkane distribution in oil samples may also provide information on

Table 7-1 Diagnostic Ratios Derived from Selected *n*-Alkanes and Acyclic Isoprenoids (Diterpanes) Recorded by GC/FID

Diagnostic Ratio	*Definition*
DR-*n*C17/Pri	*n*-heptadecane/pristane
DR-*n*C18/Phy	*n*-octadecane/phytane
DR-Pri/Phy	pristane/phytane

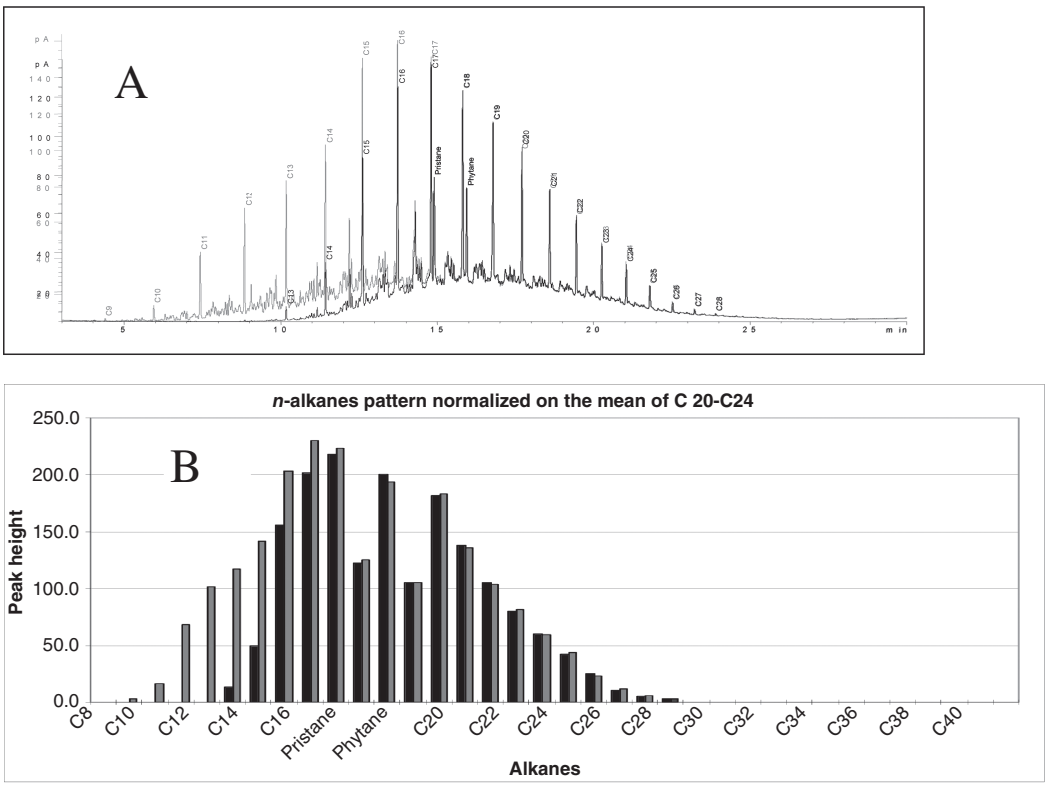

Figure 7-3 Examples of weathering checks of gas oils by GC/FID. A. Simple overlaying of chromatograms. B. Bar chart comparison of normalized *n*-alkane distributions.

possible wax/paraffin redistribution as a part of the weathering process (Strøm-Kristiansen et al., 1997).

If the comparison of the GC/FID chromatograms of the spill samples with the candidate source samples reveals differences in the hydrocarbon distribution, in the unresolved complex mixture distribution, and/or in the acyclic isoprenoid ratios that obviously are not caused by weathering and that are significantly higher than the analytical variance, then these source samples are nonmatching (i.e., nonidentity is achieved) and should be ruled out and eliminated from additional levels of analysis (cf. Figure 7-1). If, however, there are any doubts about the conclusions, the samples should still be considered as potential sources and analyzed in accordance with Level 2 of the flowchart (Figure 7-1).

7.4.4 Level 2 — GC/MS Fingerprinting

At this level, selected spill samples and those candidate source samples that have not been eliminated by the GC/FID screening (Level 1) are analyzed by tiered GC/MS fingerprinting, generally performed in selected ion-monitoring mode (GC/MS-SIM). The GC/MS analysis at this level is used for characterizing and assessing the content and distributions of a suite of diagnostic and target alkylated PAC and petroleum biomarker analytes, from which recommended diagnostic ratios can successively be generated.

7.4.4.1 Diagnostic Ratios from GC/MS Fingerprinting

Analytical data from PAC and biomarker compounds form the basis for generating a suite of diagnostic ratios. It is generally recommended that applied ratios are based on single compounds recorded at the same *m/z* value (e.g., *m/z* 192: DR-2-MPhe/1-MPhe; Table 7-2) to eliminate the mass spectrometer's varying response for different ions. In many cases, however, it may be of diagnostic value also to include ratios based on different *m/z* groups (e.g., *m/z* 212/206: DR-C2-Dbt/C2-Phe; Table 7-2) to assess different levels of various compound groups.

For the PACs, diagnostic ratios based on peak areas are generally recommended as most compounds are well-resolved peaks with smooth baselines that are easily integrated for area; besides, some diagnostic ratios are based on whole isomer groups (e.g., DR-C2-Dbt/C2-Phe; cf. Section 7.4.4.2) in which case the ratio has to be based on the total area. For the biomarkers, however, integration of peak heights are generally recommended; diagnostic ratios are often based on peaks that are not well-

Table 7-2 Recommended Diagnostic Ratios (DR) Derived from Alkylated PACs[1]

Ratio Name	*Definition*	m/z *Value*
DR-2-MPhe/1-MPhe[2]	2-Methylphenanthrene/1-Methylphenanthrene	192
DR-4-MDbt/1-MDbt	4-Methyldibenzothiophene/1-Methyldibenzothiophene	198
DR-C2-Dbt/C2-Phe	C2-Dibenzothiophenes/C2-Phenanthrenes	212/206
DR-C3-Dbt/C3-Phe	C3-Dibenzothiophenes/C3-Phenanthrenes	226/220
DR-C3-Dbt/C3-Chr	C3-Dibenzothiophenes/C3-Chrysenes	226/270
DR-Retene/C4-Phe	Retene (7-isopropyl-1-methylphenanthrene)/C4-Phenanthrenes	234
DR-BaF/4-MPy	Benzo(a)fluorene/4-Methylpyrene	216
DR-B(b+c)F/4-MPy	Benzo(b+c)fluorene/4-Methylpyrene	216
DR-2-MPy/4-MPy	2-Methylpyrene/4-Methylpyrene	216
DR-1-MPy/4-MPy	1-Methylpyrene/4-Methylpyrene	216

[1] C# here denotes the number of carbon atoms in the PAC alkyl substituent (i.e., C2 could either denote dimethyl or ethyl and C3 either trimethyl or ethyl-methyl).

[2] In cases where the double peaks of the methylphenanthrenes are not properly resolved, this diagnostic ratio could alternatively be generated from the area of the double peaks (i.e., DR-[(3-MPhe + 2-MPhe)/(9/4-MPhe + 1-MPhe)]).

resolved and with noisy baselines, in which case ratios based on peak heights are more robust. When integrating individual peaks, the use of coeluting peaks should be avoided, if possible. If peaks of, for example, C29Ts and C29αβ coelute (cf. Figure 7-6), use either the peak height of the combined peaks or the total area of the combined peaks in the combined ratio (C29Ts + C29αβ)/C30αβ.

Diagnostic ratios could be calculated either as A/B or A/(A + B), where the latter will always give values in the range of 0 to 1. The use of ratios compared to single components generally lowers the analytical variance, and ratios calculated as A/(A + B) result in lower relative standard deviations (RSDs) compared to ratios calculated as A/B. However, while ratios calculated as A/B give constant RSDs independent of the numerical values of A and B, this is not the case for ratios calculated as A/(A + B), where RSDs depend on the actual ratio value (Kienhuis, 2005; Sect. 3.3). Thus, ratios calculated as A/(A + B) are less suitable for comparison, since the confidence interval has to be adjusted to reflect that the RSD depends on the actual value. Generally, it is therefore recommended that diagnostic ratios (DR) between two distinct and well-resolved peaks (A and B) should be based on the ratio formula:

- DR = A/B, or
- DR = 100 × A/B (in %, depending on the preference of the user)

7.4.4.2 Diagnostic Ratios Derived from Alkylated Polycyclic Aromatic Compounds

Petroleum originating from different oil fields and petrogenic provinces generally has sufficient differences in the distribution of alkylated PAC isomers to be of diagnostic value. Several PAC diagnostic ratios have recently been described and applied for oil spill identification (Wang et al., 1999; Weiss et al., 2000; Daling et al., 2002; Stout et al., 2002; Wang and Fingas, 2003). The most common diagnostic groups of alkylated PAC isomers are the phenanthrenes (3-ring) and the heterocyclic dibenzothiophenes (3-ring) that can be found in all sample types mentioned in Section 7.2. However, other alkylated PAC isomer groups can also be of diagnostic value, like the fluorenes (3-ring) and the chrysenes (4-ring). In the CEN methodology, additional ratios of the methylfluoranthenes/methylpyrenes/benzofluorenes (4-ring) are also included, because compounds within this group may show significant variation between different oils. Thus, in total, 10 diagnostic ratios derived from alkylated PAC isomers are recommended as listed in Table 7-2.

To facilitate and ensure proper identification of the chromatographic peaks used for generating the recommended diagnostic ratios, the CEN methodology includes a series of relevant ion fragmentograms as shown in Figure 7-4A to 7-4I. Of particular importance are the methyl-phenanthrenes (*m/z* 192), as they may be used to differentiate between crude oils and heavy bunker fuel oils (cf. Figure 7-4A). For example, it has been reported that in crude oils the first pair of peaks (i.e., the 3-methyl- and 2-methylphenanthrenes) are generally less abundant then the second pair of peaks (i.e., the 9-/4- and 1-methylphenanthrenes), while in heavy fuel oil this is reversed (Dahlmann, 2003). The isomer compound cluster recorded by *m/z* 216 (i.e., the methylfluoranthenes, methylpyrenes, and benzofluorenes) may also provide valuable diagnostic characteristics that have shown to be relatively stable and suitable especially for comparing light fuel oil samples like gas oils. More detailed features of individual oil types encountered in most environmental oil spill situations have recently been characterized and described by Dahlmann (2003).

To provide additional information, a semi-quantitative histogram (e.g., Excel™ bar chart as in Figure 7-5) established from a suite of selected PAC homologues by their relative distribution [individual compounds or compound groups normalized to a nonweathered compound, e.g., C30-17α(H),21β(H)-hopane], may be used as a supplementary diagnostic fingerprint and as a weathering check of the PAC homologues.

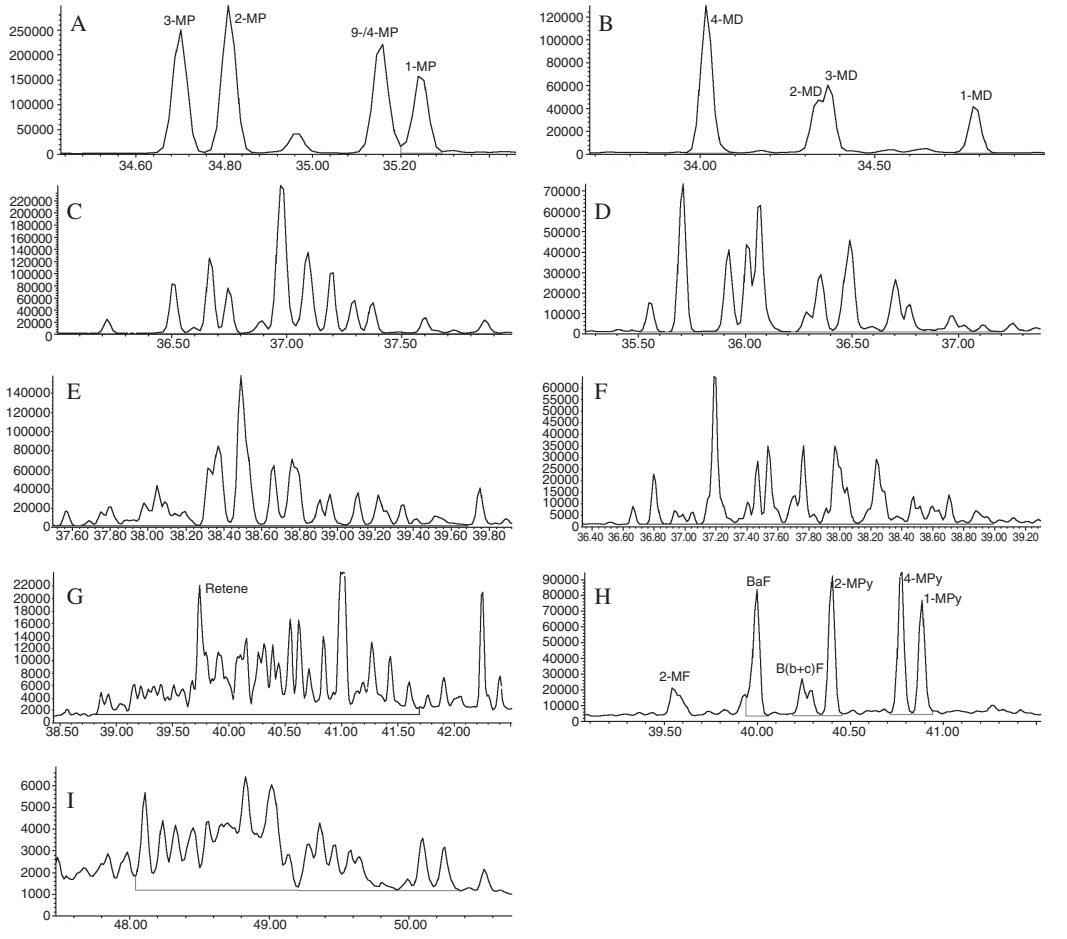

Figure 7-4 GC/MS ion fragmentograms of diagnostic alkylated PACs. 4A, methylphenanthrenes (*m/z* 192); 4B, methyldibenzothiophenes (*m/z* 198); 4C, C2-phenanthrenes (*m/z* 206); 4D, C2-dibenzothiophenes (*m/z* 212); 4E, C3-phenanthrenes (*m/z* 220); 4F, C3-dibenzothiophenes (*m/z* 226); 4G, C4-phenanthrenes including retene (*m/z* 234); 4H, methylfluoranthenes/methylpyrenes/benzofluorenes (*m/z 216*); 4I, C3-chrysenes (*m/z* 270).

7.4.4.3 Diagnostic Ratios Derived from Petroleum Biomarkers

Biomarkers are naturally occurring, ubiquitous, and stable hydrocarbons that are present in crude oils and most petroleum products. They are complex "molecular fossils" derived from once-living organisms. Biomarkers' specificity, diversity, complexity, and relative high resistance to weathering therefore make them extremely useful as diagnostic "markers" in the characterization and differentiation of spilled oils and candidate source oils (Stout et al., 2000; see also Chapter 3 herein). The most common biomarkers used by organic geochemists include sesquiterpanes (e.g., drimanes), triterpanes (e.g., hopanes), diasteranes/steranes (e.g., diacholestanes/cholestanes), and mono- and triaromatic steroids (Peters et al., 2005).

By exploiting the experience gained by petroleum exploration and production geochemistry, combined with the results of an extensive analysis of a large number of oils (Faksness et al., 2002a), a suite of 19 diagnostic biomarker ratios have been selected as

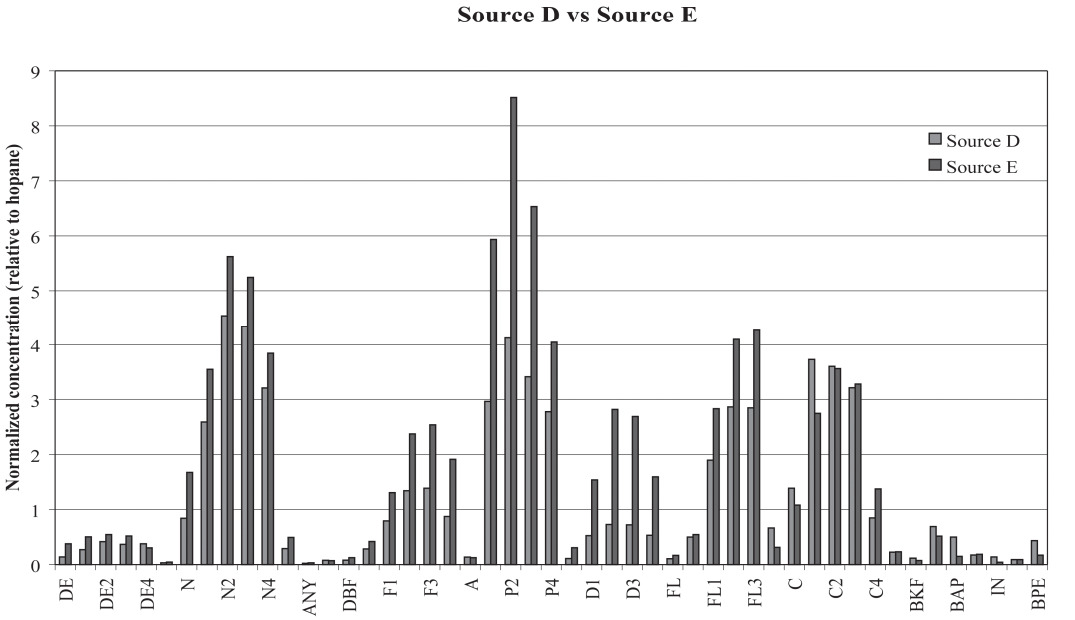

Figure 7-5 Relative distribution of PAC homologues in two different IFO-180 bunker oils (from different refineries).

technically defensible indices to differentiate among qualitative similar oils from spills and available sources.

Examples of ion fragmentograms displaying the target peaks of hopanes and other tri- and pentacyclic triterpanes, rearranged (diasteranes) and regular steranes, and triaromatic steroids are presented in Figures 7-6, 7-7, and 7-8, respectively. Accordingly, definitions of the recommended ratios within each group of biomarkers are listed in Tables 7-3, 7-4, and 7-5, respectively.

Aromatic steroid hydrocarbons (derived from steranes) may also be used as diagnostic compounds suitable for oil spill identification as has recently been described by Barakat et al. (2002) with respect to terrestrial spilled oil. In the CEN guideline, three ratios derived from triaromatic steroids are recommended as diagnostic tools. The monoaromatic steroids, typically recorded by *m/z* 253, are not recommended here as they may coalesce with higher paraffins (above C19) also recorded at this *m/z* value on GC/MS instruments with low mass resolution (e.g., quadropoles).

7.4.4.4 Optional Diagnostic Ratios Derived from Sesquiterpanes

Sesquiterpanes may be included as an optional group of biomarkers. Sesquiterpanes include a group of bicyclic (C14–C16 polymethyl-substituted decalins) biomarkers that comprise one of the largest of the terpenoid classes. Thus, sesquiterpanes, including drimane and eudesmane, are common components of crude oils and ancient sediments. In lighter to middle petroleum products like jet fuel and diesel, where refining processes have removed most of the higher-molecular-weight tetracyclic steranes and pentacyclic triterpanes, the lower-molecular-weight bicyclic sesquiterpanes are generally concentrated. In GC/MS chromatograms these compounds may be examined by their characteristic fragment ions (*m/z* 123, 179, 193, and 207), from where highly diagnostic ratios for correlation, differentiation, and source identification of lighter- to middle-range petroleum products may be acquired (Wang et al., 2005). Like any other low-boiling compounds, however, the sesquiterpanes are

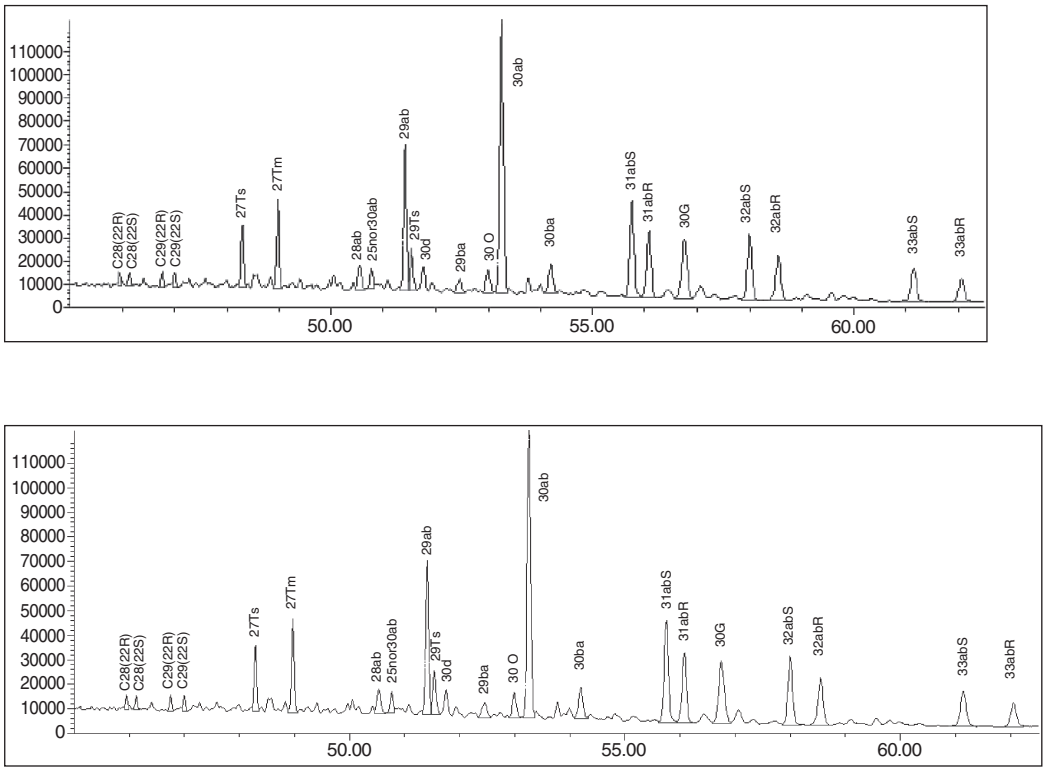

Figure 7-6 GC/MS ion fragmentogram of tri- and pentacyclic triterpanes ("hopanes") recorded at *m/z* 191 in the SINTEF standard oil mixture.

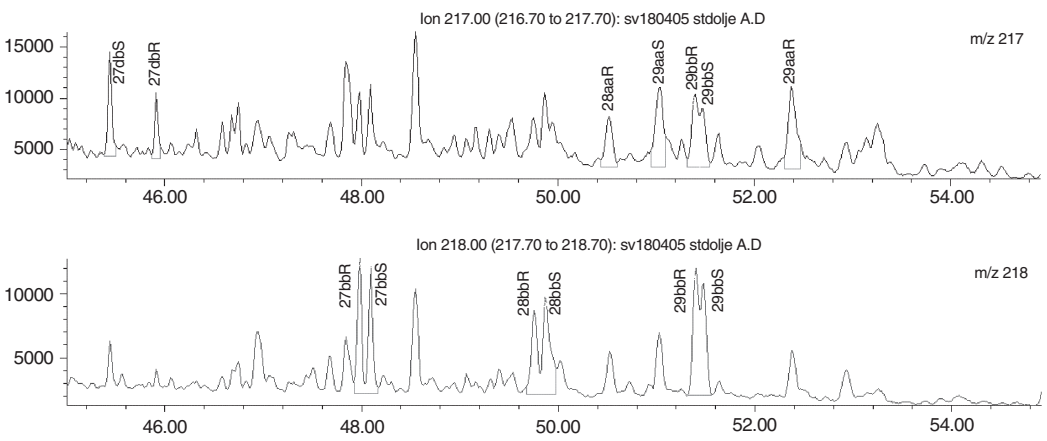

Figure 7-7 GC/MS ion fragmentogram of tetracyclic rearranged (diasteranes) and regular 17α(H)-steranes, and 17β(H)-steranes recorded at *m/z* 217 and *m/z* 218, respectively, in the SINTEF standard oil mixture.

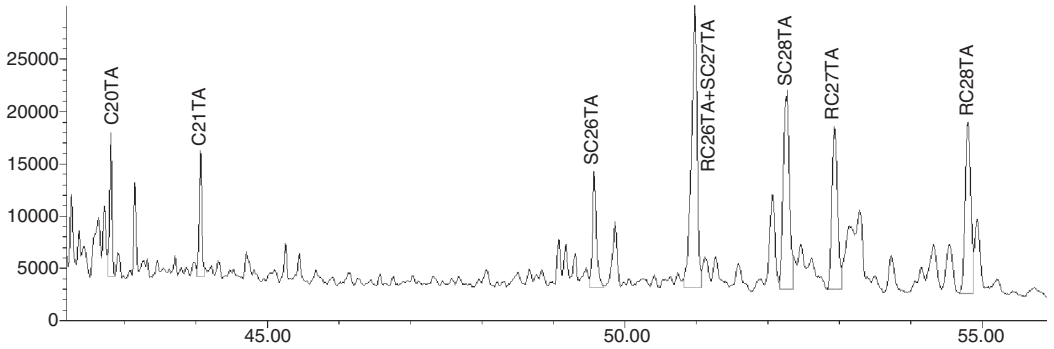

Figure 7-8 GC/MS ion fragmentogram of triaromatic steroids recorded at *m/z* 231 in the SINTEF standard oil mixture.

Table 7-3 Recommended Diagnostic Ratios (DR) Derived from Tri- and Pentacyclic Triterpanes

Abbreviation	*Compound Name*	m/z *Value*
C28 (22S)	C28 tricyclic triterpane (Cheilanthane)	191
C28 (22R)	C28 tricyclic triterpane (Cheilanthane)	191
C29 (22S)	C29 tricyclic triterpane (Cheilanthane)	191
C29 (22R)	C29 tricyclic triterpane (Cheilanthane)	191
C27Ts	C27 18α(H)-22,29,30-trisnorneohopane	191
C27Tm	C27 17α(H)-22,29,30-trisnorhopane	191
C28αβ	C28 17α(H),21β(H)-28,30-bisnorhopane	191
25norC29αβ	C29 17α(H),21β(H)-25-norhopane	191
C29αβ	C29 17α(H),21β(H)-30-norhopane	191
C29Ts	C29 18α(H)-30-norneohopane	191
C30d	C30 15α-methyl-17α(H)-27-norhopane (diahopane)	191
C29βα	C29 17β(H),21α(H)-30-norhopane (normoretane)	191
C30O	C30 18α(H)-oleanane	191
C30αβ	C30 17α(H),21β(H)-hopane	191
C30βα	C30 17β(H),21α-(H)-hopane (moretane)	191
C31αβS	C31 17α(H),21β(H),22S-homohopane	191
C31αβR	C31 17α(H),21β(H),22R-homohopane	191
C30G	C30 Gammacerane	191
C32αβS	C32 17α(H),21β(H),22S-bishomohopane	191
C32αβR	C32 17α(H),21β(H),22R-bishomohopane	191
C33αβS	C33 17α(H),21β(H),22S-trishomohopane	191
C33αβR	C33 17α(H),21β(H),22R-trishomohopane	191

Ratio Name	*Definition*	*Ratio Name*	*Definition*
DR-C28	C28(S+R)/C30αβ	DR-C29αβ	C29αβ/C30αβ
DR-C29	C29(S+R)/C30αβ	DR-C29Ts	C29Ts/C30αβ
DR-(C28+C29)	C28(S+R) + C29(S+R)/C30αβ	DR-C30d	C30d/C30αβ
DR-C27Ts	C27Ts/C27Tm	DR-C30O	C30O/C30αβ
DR-C28αβ	C28αβ/C30αβ	DR-C30G	C30G/C30αβ
DR-25norC30αβ	25norC29αβ/C30αβ		

Table 7-4 Recommended Diagnostic Ratios (DR) from Rearranged (Diasteranes) and Regular 14α(H)- and 14β(H)-Steranes

Abbreviation	*Compound Name*	m/z *Value*
C27dbS	C27 13β(H),17α(H),20S — diacholestane (diasterane)	217
C27dbR		217
C28ααR	C27 13β(H),17α(H),20R — diacholestane (diasterane)	217
C29ααS	C28 24-methyl-5α(H),14α(H),17α,20R — cholestane	217
C29ββR	C29 24-ethyl-5α(H),14α(H),17α,20S — cholestane	217
C29ββS	C29 24-ethyl-5α(H),14β(H),17β(H),20R — cholestane	217
C29ααR	C29 24-ethyl-5α(H),14β(H),17β(H),20S — cholestane	217
C27ββR	C29 24-ethyl-5α(H),14α(H),17α(H),20R — cholestane	218
C27ββS	C27 5α(H),14β(H),17β(H),20R — cholestane	218
C28ββR	C27 5α(H),14β(H),17β(H),20S — cholestane	218
C28ββS	C28 24-methyl-5α(H),14β(H),17β(H),20R — cholestane	218
C29ββR	C28 24-methyl-5α(H),14β(H),17β(H),20S — cholestane	218
C29ββS	C29 24-ethyl-5α(H),14β(H),17β(H),20R — cholestane	218
	C29 24-ethyl-5α(H),14β(H),17β(H),20S — cholestane	

Ratio Name	*Definition*
DR-C29ααS	C29ααS/C29ααR
DR-C29ββ	C29ββ(R+S)/C29αα(S+R)[1]
DR-C27ββSTER	C27ββ(R+S)/[C28ββ(R+S) + C29ββ(R+S)][1]
DR-C28ββSTER	C28ββ(R+S)/[C27ββ(R+S) + C29ββ(R+S)][1]
DR-C29ββSTER	C29ββ(R+S)/[C27ββ(R+S) + C28ββ(R+S)][1]

[1] Although it is generally recommended to use peak heights for generating the biomarker diagnostic ratios, the use of peak areas may in some cases be justified as with the steranes. Thus, as the ββR- and ββS-isomers are often not well-resolved, it is recommended to measure the C27ββ(R+S), C28ββ(R+S), and C29ββ(R+S) double peaks by integrating the whole area and hence use that area for generating the relevant diagnostic ratios as shown in Figure 7-7. For the DR-C29ααS ratio, either peak height or area can be used.

Table 7-5 Recommended Diagnostic Ratios (DR) Derived from Triaromatic Steroids

Abbreviation	*Compound Name*	m/z *Value*
C20TA	C20-triaromatic steroid (pregnane derivative)	231
C21TA	C21-triaromatic steroid (homopregnane derivative)	231
SC26TA	C26 20S-triaromatic steroid (cholestane derivative)	231
RC26TA+SC27TA	C26 20R- + C27 20S-triaromatic steroids	231
SC28TA	C28 20S-triaromatic steroid (ethylcholestane derivative)	231
RC27TA	C27 20R-triaromatic steroid (methylcholestane derivative)	231
RC28TA	C28 20R-triaromatic steroid (ethylcholestane derivative)	231

Ratio Name	*Definition*
DR-C21TA	C21TA/RC28TA
DR-SC26TA	SC26TA/SC28TA
DR-RC27TA	RC27TA/RC28TA

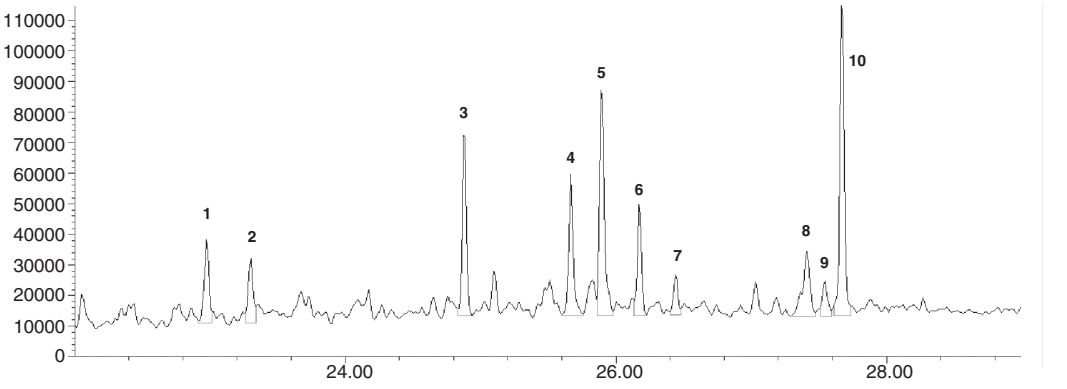

Figure 7-9 GC/MS ion fragmentogram of sesquiterpanes recorded at *m/z* 123 in the SINTEF standard oil mixture.

Table 7-6 Optional Diagnostic Ratios (DR) Derived from Sesquiterpanes

Abbreviation	*Compound Name (Peak No.)*	m/z *Value*
SES1	C14H26-sesquiterpane (1)	123
SES2	C14H26-sesquiterpane (2)	123
SES3	C15H28-sesquiterpane (3)	123
SES4	C15H28-sesquiterpane (4)	123
SES5	C15H28-8β(H)-drimane (5)	123
SES6	C15H28-sesquiterpane (6)	123
SES7	C15H28-sesquiterpane (7)	123
SES8	C16H30-sesquiterpane (8)	123
SES9	C16H30-sesquiterpane (9)	123
SES10	C16H30-8β(H)-homodrimane (10)	123
Ratio Name	*Definition (Incl. Peak No.)*	
DR-SES1/SES2	C14-sesquiterpane (1)/C14-sesquiterpane (2)	
DR-SES3/SES5	C15-sesquiterpane (3)/C15-8β(H)-drimane (5)	
DR-SES4/SES6	C15-sesquiterpane (4)/C15-sesquiterpane (6)	
DR-SES5/SES10	C15-8β(H)-drimane (5)/C16-8β(H)-homodrimane (10)	

subject to evaporative weathering, a fact that has to be taken into account before using them for diagnostic purposes. Figure 7-9 gives an example of a sesquiterpane GC/MS ion fragmentogram (*m/z* 123) of the SINTEF standard oil mixture, and Table 7-6 lists the identified peaks and recommended diagnostic ratios.

After analyzing the selected samples by GC/MS, the recorded chromatograms should be visually checked for characteristic features or obvious differences, which could possibly eliminate any suspected source sample from the candidate sources. If there is any doubt whether observed differences are due to genuine differences or the effect of weathering and degradation, an additional weathering check of the PAC distribution (cf. Section 7.4.4.2) should be performed. Together with the information obtained from the weathering check of the *n*-alkanes (cf. Section 7.4.3.1), it can now be concluded which compounds are most probably affected and hence not suitable for generation of diagnostic ratios. After the visual evaluation of the GC/MS chromatograms and the weathering check, the diagnostic ratios are generated using only

those compounds not or only slightly affected by weathering.

7.4.5 Level 3 — Treatment of Results

7.4.5.1 Comparison of Oil Samples Using Diagnostic Ratios

If two oil samples are identical, their chemical composition is by definition the same apart from those changes introduced after the spill as the result of weathering, contamination, mixing, and degradation (cf. Section 7.3). Accordingly, measured ratios between any pair of compounds should also match — up to a certain analytical related statistical confidence level — in identical samples. Various ratios between individual compounds or group of compounds have been described extensively in the literature and used to compare oil samples. In many cases such ratios may be referred to as "diagnostic" if they possess the quality to discriminate genuinely different oils. This is often the case with ratios derived from biomarkers. Other ratios, like some based on PACs, may or may not possess similar diagnostic significance; especially regarding refined products, refinery processes (e.g., cracking and reforming) may have affected the genuine petrogenic PAC distribution. However, genuinely diagnostic or not, any ratio not influenced by weathering or degradation should still display the same value for identical samples and thus be of "diagnostic" value in a specific spill case and contribute in discriminating and ruling out "nonsource" oils.

It is important to realize, however, that the suite of diagnostic PAC (Table 7-2) and biomarker ratios (Table 7-3 to 7-6) are neither all-inclusive nor appropriate for all oil spill identification cases. In some instances it may be prudent to include a certain characteristic feature of the spilled oil that is recognized as particularly diagnostic. In other situations, the abundance of some compounds necessary for determining the recommended diagnostic ratios are below the recommended S/N ratio. Thus, maintaining flexibility in the selection of diagnostic ratios to be used in a specific spill case is very important.

7.4.5.2 Criteria for Selecting, Eliminating, and Evaluating Diagnostic Ratios

In the CEN methodology, the criteria for selecting diagnostic ratios have generally been based on (1) specificity and diversity, (2) resistance to weathering, and (3) analytical precision/complexity (e.g., ratios calculated using different *m/z* values are generally not recommended).

Basically, only those ratios that can be measured with low variation should be evaluated for comparing candidate sources to spilled oil (Stout et al., 2000). Peaks with low S/N ratios have an increased variance and should only be used for a visual comparison and not for the comparison of diagnostic ratios. To accommodate both for the limitation in analytical precision and impact of sample heterogeneity, Stout et al. (2000) have suggested a protocol by which candidate diagnostic ratios are evaluated in order to identify those that are most useful for further correlation analysis. The evaluation of the diagnostic ratios was conducted by a simple statistical test (relative standard deviation of triplicate analyses) to identify those ratios that were unaffected by, for example, sample heterogeneity or low analytical precision. However, the use of triplicate analyses of one or some samples in order to decide whether a ratio should be used or not may not be particularly robust and result in highly varying RSDs among the diagnostic ratios.

Another approach was suggested by Faksness et al. (2002a) who applied a Student's t-test on triplicate analyses of one sample and the 95% and 98% confidence levels as critical limits to decide whether two ratios matched or not. Again this approach is not very robust, as the triplicate analyses sometimes may result in very low variances for some ratios, and in the end this may result in very small critical differences (cf. Equation 7-2) and hence non-matching ratios (i.e., false negative or the type I error).

For the CEN methodology, a comparable yet different approach has been applied. It recommends a fixed RSD of 5% for all diagnostic ratios to overcome the variation in critical differences. Before applying this approach,

however, the laboratory should validate its analytical method by analyzing a sample at least 7 times to test that it complies with the 5% RSD limit. This limit is applied as a quality criterion, because methods producing higher RSD values should not be used to analyze and compare oil samples in a forensic context. A weakness with this approach, however, could be that small peaks may have relatively high RSDs. Therefore, it is generally recommended to analyze some of the samples from a spill case in duplicate and to compare the ratios of the duplicates. When small peaks result in differences larger than the critical difference, these peaks should only be used for a visual comparison.

To decide whether or not a particular ratio is sufficiently precise and robust to be used for comparison, the CEN methodology describes two consecutive tests to evaluate diagnostic ratios as described by the protocol/decision chart shown in Figure 7-10. The two consecutive tests used for selection and elimination of the diagnostic ratios comprise:

- elimination by means of a signal-to-noise (S/N) tests (Section 7.4.5.4)
- elimination by means of the comparison of the duplicate analyses (Section 7.4.5.5)

7.4.5.3 Repeatability Limit and Critical Difference

To estimate an acceptable difference between two analytical results, the standard deviation of the analysis method is used (ISO, 1994a). Repeatability (**r**) is applied as a test method to compare individual ratios, assuming that the two samples to be compared originate from the same source. Repeatability conditions are met when the samples are analyzed in one series. If the repeatability limit[4] is exceeded, it is beyond reasonable doubt that this assumption is not valid and that the samples originate from different sources.

The repeatability limit of validation is based on standard normal distribution. Determining the repeatability standard deviation s_r of an analytical method depends on the quality assurance (QA) system of the laboratory. In general, s_r is calculated by analyzing samples relevant for a method at least seven times in one series when a new or revised method is being implemented and if the method has to be validated. Calculation of the standard deviation or relative standard deviation (RSD) generates **r**.

The repeatability limit $\mathbf{r}_{95\%}$ (i.e., 95% confidence level) is calculated by multiplying the fixed RSD (s_r) with a factor of 2.8 by using Eq. (7-1):

$$\mathbf{r}_{95\%} = 2.8 * 5\% = 14\% \qquad (7\text{-}1)$$

This implies that when samples are analyzed under repeatability conditions, any ratio with an s_r of 5% to be used for the evaluation must not differ more than 14% relatively.

The corresponding critical difference (CD) is calculated by using Eq. (7-2):

$$CD = (\text{mean} * \mathbf{r}_{95\%})/100 \qquad (7\text{-}2)$$

When the absolute difference between a pair of corresponding ratios of two samples to be compared is lower than the critical difference (CD), based on the repeatability limit $\mathbf{r}_{95\%}$, then the comparison gives a positive match; if it is higher, they do not match. This test has to be performed for every diagnostic ratio applied in the evaluation. The repeatability limit $\mathbf{r}_{95\%}$ is a test at the 95% confidence level, which implies that 5%, or 1 out of 20 of the pair of ratios slightly above the CD, is acceptable without jeopardizing the conclusion. Table 7-7 shows an example of a ratio comparison.

7.4.5.4 Elimination of Diagnostic Ratios Using Signal-to-Noise (S/N) Test

Only peaks with S/N > 3 to 5 should be used for comparing diagnostic ratios. The S/N criterion is based on the method recommended by IUPAC (Ettre, 1993). N is the peak-to-peak

[4] The repeatability limit (**r**) is the difference between two test results; the associated standard deviation is $\sigma\sqrt{2}$. In normal statistical practice, for examining the differences between these two values, the critical difference (CD) used is f times this standard deviation, i.e., $f * \sigma\sqrt{2}$. For a normal distribution at 95% probability level, f is 1.96 and $f * \sqrt{2}$ then is 2.77. As the purpose of this guideline is to give some simple "rules of thumb" to be applied by nonstatisticians when examining the test results, a "rounded" value of 2.8 has been suggested instead of $f * \sqrt{2}$ (ISO, 1994b). Therefore, the repeatability limit $\mathbf{r}_{95\%}$ is calculated by multiplying s_r with 2.8.

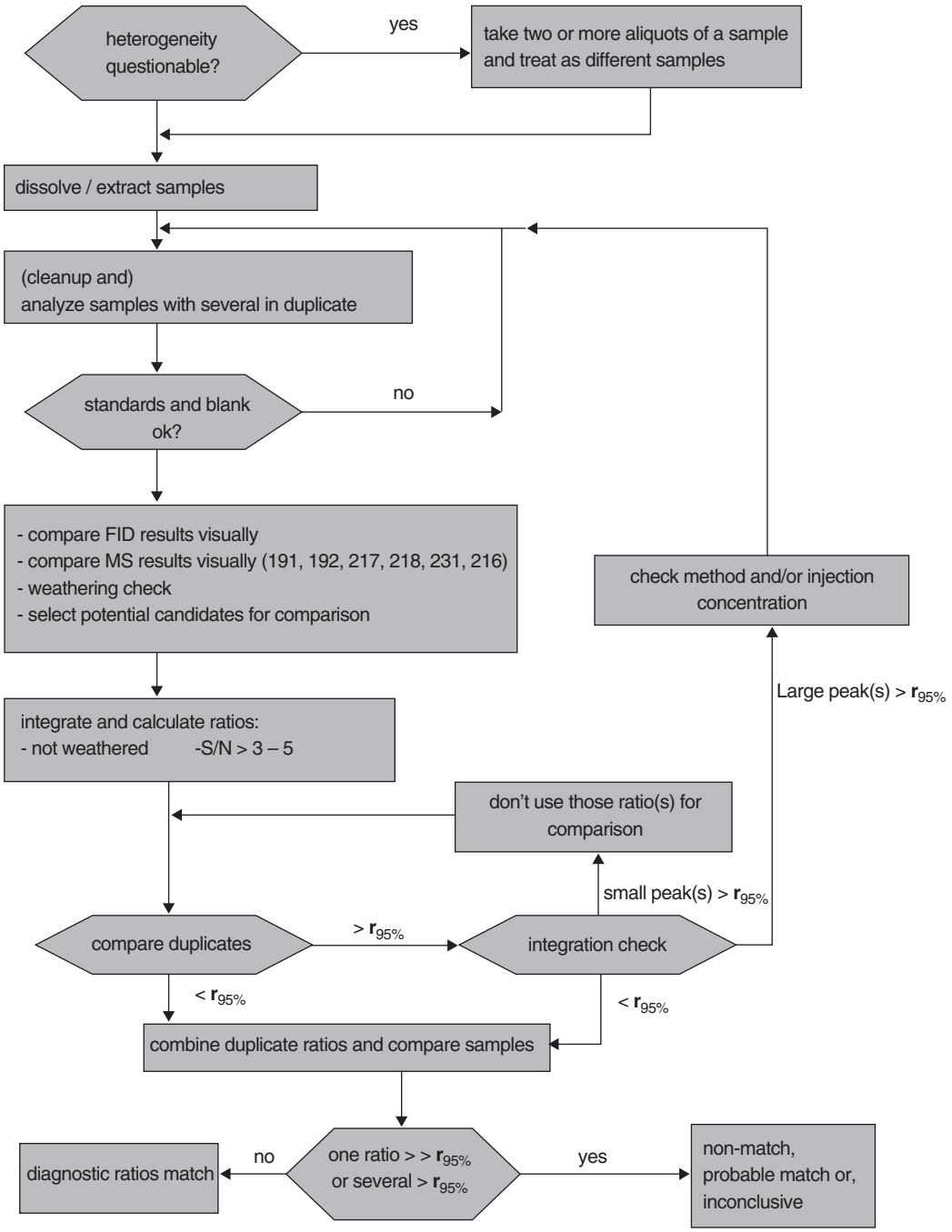

Figure 7-10 Protocol/decision chart for selection/elimination of diagnostic ratios.

Table 7-7 Example of Comparison of Diagnostic Ratios Based on a Repeatability Limit of 14%

Diagnostic Ratios (in %)		*Mean*	*Absolute Difference*	*Critical Difference*	*Conclusion*
Sample A	*Sample B*	*(%)*	*(%)*	*(CD, %)*[1]	
48	52	50	52 − 48 = 4	50 ∗ 14/100 = 7	Match
35	45	40	45 − 35 = 10	40 ∗ 14/100 = 5.6	Nonmatch

[1] The critical difference is based on the repeatability limit ($\mathbf{r}_{95\%}$) of 14% (at 5% RSD); cf. Eq. 7-2.

value of a part of the noise around a peak and S the peak height. A decision range of 3 to 5 is given, because it may often be difficult to estimate the noise precisely due to the many small peaks present in the chromatographic fine structure of oil samples.

As an example, a part of the *m/z* 191 ion fragmentogram (from C28 22R-tricyclic triterpane to C27 17α(H)-22,29,30-trisnorhopane) of a bunker oil sample is shown in Figure 7-11. Normally, only a part of the baseline around the peak would be used to calculate the noise. A comparison of the noise from 37.6 min–40 min with the noise of another section of the chromatogram [e.g., from 51.5 min to 55 min (see insert)] shows that the noise may not only be caused by the instrument, but that a significant part arises from the very small peaks present in the fine structure of the chromatogram, as mentioned above. In this example, S = 1950 (for the C28 22R-tricyclic triterpane) and N is 450 resulting in S/N = 4.5; therefore, the C28 22R-tricyclic triterpane peak could be either excluded or included in the DR comparison test.

Even more difficult, however, is the integration of groups of isomer compound. For example, in a propeller shaft lubricating oil, many of the methylated PACs mentioned in Table 7-2 are present at a low concentration showing only the highest peaks of an isomer group. In such cases, the isomer pattern should only be used for visual comparison, and in general the S/N > 3 to 5 criterion should be used as a "rule-of-thumb" to facilitate the elimination of small peaks and groups of peaks from the ratio comparison.

7.4.5.5 Elimination of Diagnostic Ratios Using Duplicate Analyses

Duplicate analyses of two aliquots taken from a homogenous sample or the duplicate injections of the same extract will provide information about the actual performance of the analytical system for the samples involved. To improve the value of this test, it is strongly advised to analyze all samples in duplicate in cases with two or three samples. In cases with many samples, at least some of the samples should be analyzed in duplicate.

Ideally, duplicate analyses of an extract should result in identical diagnostic ratios, and observed differences can only be caused by analytical variation. As a result of this test, several options are possible according to the repeatability test (cf. Section 7.4.5.3):

(1) Ratios of large peaks (i.e., with S/N ≫ 5) differ more than 14% (cf. Section 7.4.5.3, Eq. 7-1), then the integration of the peaks involved (retention time and baseline drawing), instrumental conditions, and/or injection concentration should be rechecked and the sample (e.g., after cleanup) preferably re-analyzed.
(2) Ratios of small peaks (i.e., S/N ≈ 5) differ more than 14%, then again the integration of the peaks involved should be rechecked (retention time and baseline drawing); if one or two ratios are still above the critical difference (cf. Section 7.4.5.3, Eq. 7-2), the peaks involved should only be used for a visual comparison, while if several ratios are still above the critical difference, proceed as with large peaks.

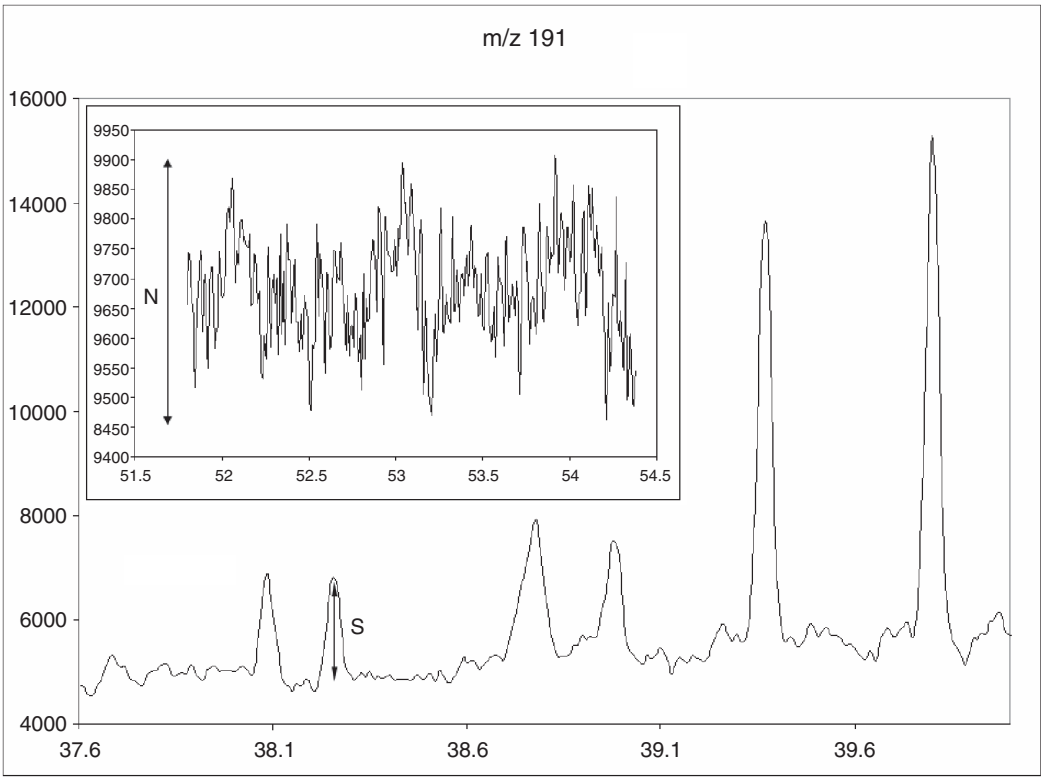

Figure 7-11 Signal-to-noise calculation. The distance between the highest and lowest peaks of a part of the noise is used as N (see insert) and the peak height as S. In this example, S/N is 4.5.

(3) All ratios differ less than 14%, then proceed to the next step of the decision chart (Figure 7-10) and compare the samples involved. Use the mean value of the diagnostic ratios of the samples analyzed in duplicate; the mean value has a lower variance and comes closer to the actual real ratio of the sample.

Generally, duplicate or triplicate analyses make the comparison of diagnostic ratios statistically more robust and reduce the relative standard deviation (RSD) by $\sqrt{n}$. Hence, the RSD of the mean value (m) of duplicate analyses becomes RSD/$\sqrt{2}$. For simplicity, however, it has been decided in this guideline not to reduce the repeatability limit of 14% and not to make a distinction between samples analyzed just once or in duplicate.

7.4.5.6 Optional Comparison of Diagnostic Ratios Using Multivariate Statistics

Multivariate statistical methods like principal component analysis (PCA) can be used as a tool to compare diagnostic ratios between multiple samples (e.g., Stout et al., 2002; Christensen et al., 2004) instead of the purely univariate approach described in previous sections. One advantage of multivariate compared to univariate statistical methods is the ease by which relationships between multiple samples and variables (e.g., diagnostic ratios) can be resolved and visualized by so-called factor scores and loading plots. Additional advantages include noise reduction, obtained by multiple measurements of the same phenomenon (i.e., interrelated variables), and their ability to detect outliers.

The fundamentals of PCA as a data treatment tool in chemical fingerprinting are that it summarizes the information in many correlated diagnostic ratios into a few so-called principal components that are weighted sums of these ratios. Hence, a model with few principal components (e.g., PC1 and PC2) describes the most prominent trends in data. By plotting principal components [e.g., the first principal component (PC1) vs. the second (PC2)], chemical information retained in numerous diagnostic ratios can be compared simultaneously. Oil samples with similar chemical composition will plot closely in score plots, whereas the opposite is the case for dissimilar oils.

The comparison of oil samples by PCA can be further refined by taking into account the uncertainty of the individual diagnostic ratios. This ensures an objective matching of oil spill samples and suspected sources based on all the available information (see Chapter 9, this volume). An example of the application of PCA to an actual oil spill case is shown by the score plot in Figure 7-12. Here the spill sample plots closely to the source sample (Statfjord crude), while nonsource samples (other North Sea crudes) plot more distantly.

7.4.6 Final Evaluation and Conclusions

The final conclusion should be based on a total evaluation using all available data. The comparison of the diagnostic ratios is an important part of the evaluation of the data, however, not all conclusive. It is important to visually inspect all the chromatograms and identify possible characteristic features, and not only to evaluate the measured ratios, before final conclusions are made. In accordance to the flowchart in Figure 7-1, the identification of an oil spill using this new CEN methodology should be concluded with respect to one of the four terms: positive match, probable match, inconclusive, or nonmatch.

A *positive match* states that source and spill samples are identical beyond reasonable doubt. A visual inspection of the chromatograms (GC/FID) and ion fragmentograms (GC/MS) shows only differences which can be

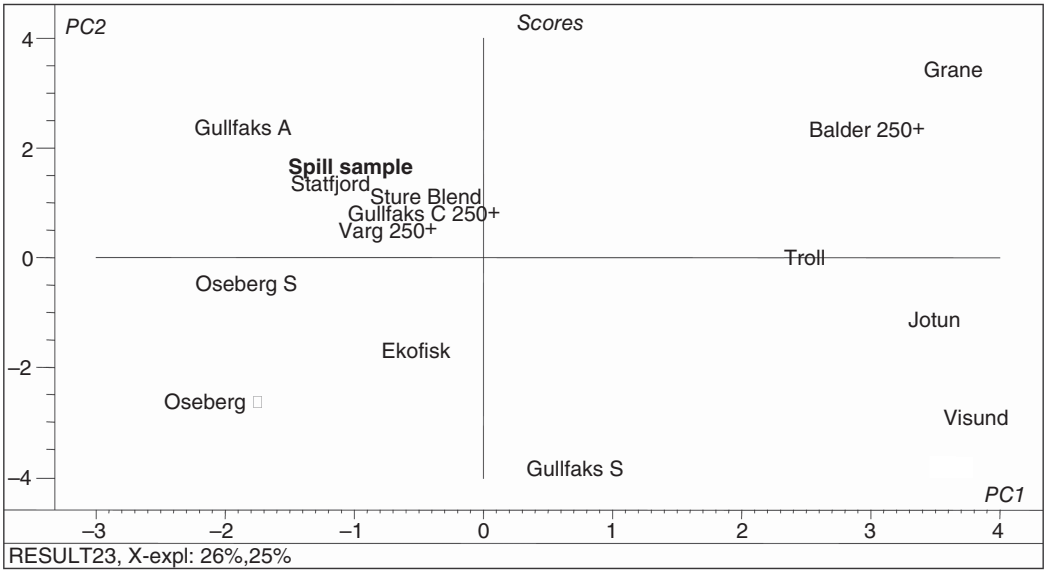

Figure 7-12 PCA score plot of data (diagnostic ratios) from a recent mystery oil spill in the North Sea. Several possible source samples (production samples from nearby offshore fields) together with other production samples were included in the PCA.

explained by weathering/degradation, and all observed differences between diagnostic ratios are below the repeatability limit.

A *nonmatch* applies when differences between chromatograms and diagnostic ratios cannot be explained by weathering/degradation, and when several pairs of ratios are outside the repeatability limit. If only very small differences (close to the repeatability limit) are observed, if contamination is expected, or if just one pair of ratios is clearly outside the repeatability limit, a *probable match* is the obvious alternative conclusion.

If the total amount of oil in a sample is very low, and consequently there is higher analytical variance of the diagnostic peaks, it might result in differences between diagnostic ratios (based on repeated analyses of the same sample) that are higher than the recommended repeatability limit. This would eventually imply the elimination of so many of the diagnostic ratios from further comparisons that it would render the test *inconclusive*.

7.5 The CEN Methodology in Practice: A Case Study

7.5.1 The Spill Case

In an actual spill case from 2004, a waterborne oil slick was discovered, and samples were collected from the spill. According to observations, a specific vessel was suspected of the illegal discharge of oil, and samples were collected from the vessel's bilge tank as candidate source samples. All samples were sent to the analytical laboratory for analysis and eventual identification of the spill.

7.5.2 GC/FID Screening

A spill sample and a bilge sample from the suspected vessel were extracted with dichloromethane, and the extracts dried over sodium sulphate and diluted to appropriate concentrations before being analyzed by GC/FID. Resulting chromatograms of the spill and candidate source samples are shown in Figure 7-13A and B, respectively. Both samples were analyzed in duplicate.

Both samples were characterized as bilge oils according to their GC/FID chromatograms, and no differences were observed apart from those related to the weathering of the spill sample. By overlaying the two chromatograms (see Figure 7-14), the degree of weathering could be evaluated. The spill sample appeared to be affected by evaporation up to about *n*-C18.

As the spill sample did not seem heavily affected by weathering above *n*-C18, the acyclic isoprenoid ratios (see Table 7-1) could be suitable for comparison. As is listed in Table 7-8, the comparison of these diagnostic ratios based on a repeatability limit of 14% shows that the absolute differences between the pairs of ratios for the spill and candidate source samples are all well below the repeatability limit; thus, these ratios give a positive match.

7.5.3 GC/MS Fingerprinting

As the GC/FID screening did not reveal significant differences between the spill and candidate source samples, the evaluation was continued to level 2, GC/MS fingerprinting, according to the flowchart in Figure 7-1. The GC/MS analysis included all the *m*/*z* values listed in Tables 7-1 to 7-5, while the optional sesquiterpanes (cf. Table 7-6) were not included as these compounds could be affected by weathering. After analyzing the samples and before generating any diagnostic ratios, the ion fragmentograms were compared visually to check for obvious differences between the two samples. As no obvious differences were observed, all peaks relevant for the generation of diagnostic ratios were integrated and the diagnostic ratios calculated.

7.5.4 Evaluation and Comparison of Diagnostic Ratios

According to the flowchart in Figure 7-10, diagnostic ratios were first eliminated on the S/N criteria. Hence, peaks with S/N < 3–5 were not used for generating diagnostic ratios; this included the following triterpane ratios: DR-C28αβ, DR-25norC29αβ, DR-C30O (cf.

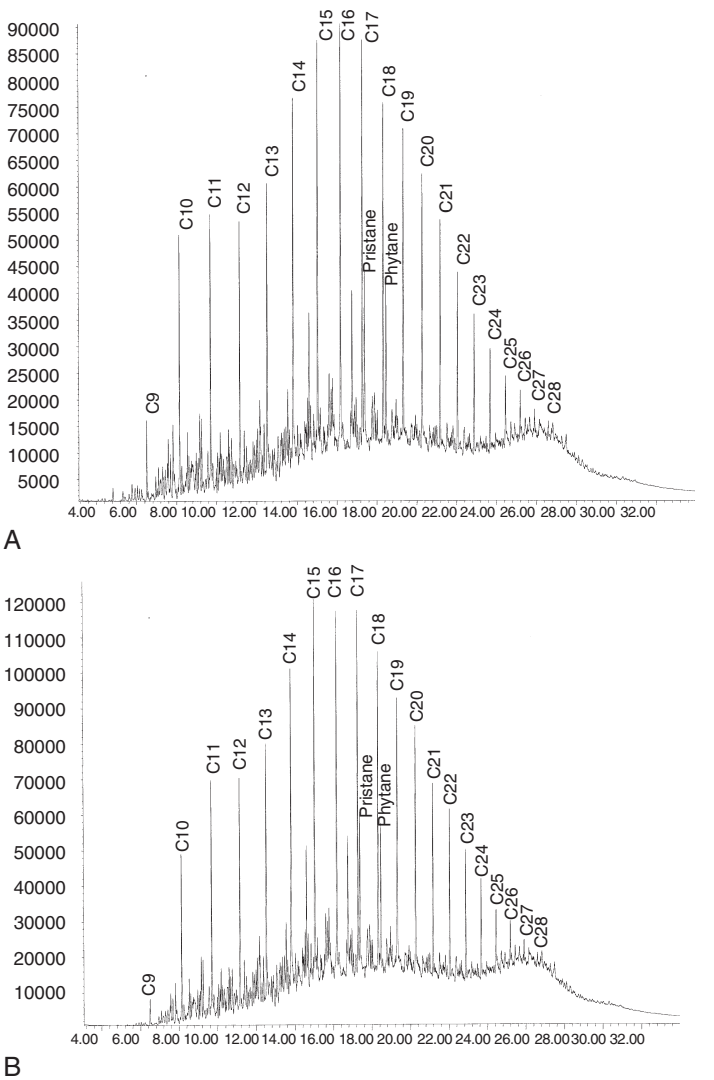

Figure 7-13 GC/FID chromatogram of extracts of (A) candidate source sample (bilge), and (B) spill sample.

Table 7-4), and the three triaromatic steroid ratios: DR-C21TA, DR-SC26TA and DR-RC27TA (cf. Table 7-5). In the next step, ratios were evaluated and eventually eliminated based on duplicate analyses using the repeatability limit. Table 7-9 gives an example on how this selection was achieved for some of the biomarker ratios.

Only one ratio, the DR-C30d, exceeded the repeatability limit of 14%. A closer look at the chromatogram showed that the C30d peak was small, and although it was not eliminated on the S/N criteria, the peak size and the noisy baseline made it difficult to obtain robust values for this ratio, which was therefore eliminated. According to the flowchart in Figure 7-1, Level 3, the remaining ratios were eventually compared on the basis of the repeatability limit, as shown in Table 7-10.

The comparison of 22 diagnostic ratios based on the repeatability limit showed that none of these ratios had absolute differences that exceeded the critical difference and that according to that criteria they all revealed a positive match as was also observed with the three diagnostic ratios derived from acyclic

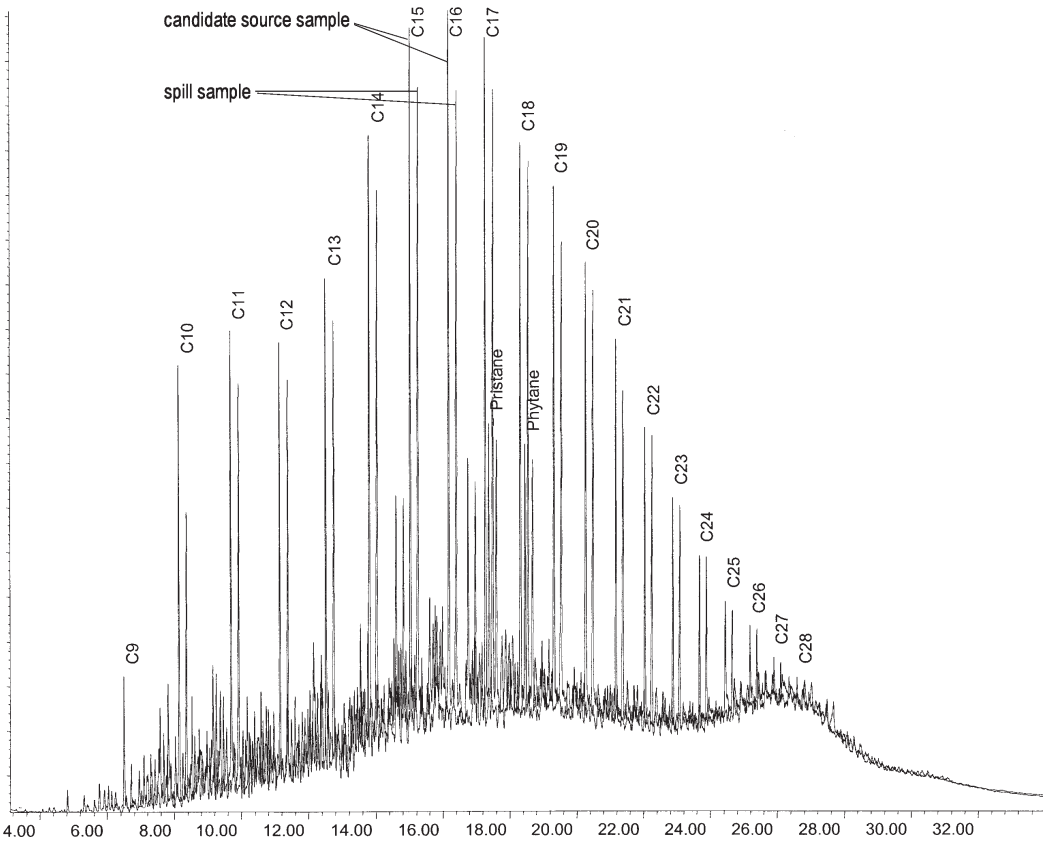

Figure 7-14 Weathering check. Overlay of GC/FID chromatograms of spill sample and candidate source sample.

Table 7-8 Comparison of Diagnostic Ratios of Acyclic Isoprenoids Using the Repeatability Limit

Diagnostic Ratio	*Spill Sample*[1]	*Candidate Source Sample*[1]	*Mean Value*	*Absolute Difference*	*Critical Difference*[2]
DR-*n*C17/Pri	1.94	2.06	2.00	0.12	0.28
DR-*n*C18/Phy	2.12	2.15	2.13	0.03	0.30
DR-Pri/Phy	1.23	1.22	1.22	0.01	0.17

[1] DR values are means of duplicate analyses.
[2] The critical difference is based on the repeatability limit ($\mathbf{r}_{95\%}$) of 14% (at 5% RSD) (cf. Eq. 7-2).

Table 7-9 Comparison of Diagnostic Ratios (in %) of Some Triterpanes Using the Repeatability Limit

Diagnostic Ratio	*Spill Sample Duplicate 1*	*Spill Sample Duplicate 2*	*Mean Value (%)*	*Absolute Difference (%)*	*Critical Difference (%)*[1]
DR-C29	115.7	112.9	114.3	2.8	16.0
DR-C27Ts	78.2	77.6	77.9	0.6	10.9
DR-C29Ts	23.2	24.3	23.7	1.1	3.3
DR-C30d	5.0	4.1	4.6	***0.9****	***0.6****
DR-C30G	8.4	8.6	8.5	0.2	1.2

[1] The critical difference is based on the repeatability limit ($\mathbf{r}_{95\%}$) of 14% (at 5% RSD) (cf. Eq. 7-2).
* Exceeded the repeatability limit, see text for description.

Table 7-10 Comparison of Diagnostic Ratios (in %) of Spill and Candidate Source Samples Using a Repeatability Limit ($\mathbf{r}_{95\%}$) of 14%

	Spill Sample	*Candidate Source Sample*	*Mean Value*	*Absolute Difference*	*Critical Difference (%)*[1]
DR-2-MPhe/1-MPhe	112, 81	109, 23	111, 02	3, 58	15, 54
DR-4-MDbt/1-MDbt	364, 39	385, 03	374, 71	20, 64	52, 46
DR-C2-Dbt/C2-Phe	65, 39	61, 91	63, 65	3, 48	8, 91
DR-BaFl/4-MPy	94, 47	98, 43	96, 45	3, 96	13, 50
DR-B(b+c)Fl/4-MPy	32, 81	35, 16	33, 99	2, 35	4, 76
DR-2-MPy/4-MPy	61, 27	66, 98	64, 12	5, 7	8, 98
DR-1-MPy/4-MPy	52, 87	58, 16	55, 52	5, 29	7, 77
DR-C3-Dbt/C3-Phe	86, 15	84, 43	85, 28	1, 72	11, 94
DR-Retene/C4-Phe	4, 93	5, 28	5, 11	0, 35	0, 72
DR-C3-Dbt/C3-Chr	3756, 7	3535, 6	3646, 1	221, 1	510, 5
DR-(C28+C29)	95, 48	92, 20	93, 84	3, 28	13, 14
DR-C28	40, 16	38, 98	39, 57	1, 18	5, 54
DR-C29	55, 32	53, 22	54, 27	2, 09	7, 60
DR-C27Ts	77, 90	78, 66	78, 28	0, 75	10, 96
DR-C29αβ	114, 3	106, 6	110, 5	7, 69	15, 47
DR-C29Ts	23, 73	22, 84	23, 29	0, 89	3, 26
DR-C30G	8, 52	9, 23	8, 87	0, 71	1, 24
DR-C29ααS	114, 1	115, 8	115, 0	1, 69	16, 09
DR-C29ββ	131, 8	134, 8	133, 3	2, 96	18, 66
DR-C27ββSTER	41, 51	41, 79	41, 65	0, 28	5, 83
DR-C28ββSTER	27, 07	25, 94	26, 51	1, 13	3, 71
DR-C29ββSTER	31, 42	32, 27	31, 84	0, 85	4, 46

[1] The critical difference is based on the repeatability limit ($\mathbf{r}_{95\%}$) of 14% (at 5% RSD) (cf. Eq. 7-2).

isoprenoids. Thus, all 25 diagnostic ratios evaluated and compared in this study revealed a positive match. As the final visual comparison (ground truth) of all chromatograms neither revealed any significant differences apart from weathering effects, it could be concluded that the spill and candidate source samples matched positively and that the spill sample was identical to the bilge sample from the suspected vessel beyond reasonable doubt.

7.6 Summary

The emerging CEN methodology for oil spill identification presented here is based on a three-level tiered approach, including (1) GC/FID screening of all involved samples (Level 1); (2) GC/MS fingerprinting of selected spill and candidate source samples (Level 2), from which up to 29 diagnostic ratios of selected PAHs and biomarkers are derived, diagnostic ratios are selected and evaluated on the basis of their analytical variability and changes due to weathering; and (3) correlation of spill and candidate oil samples based on those diagnostic ratios that can be precisely measured and are resistant to weathering effects (Level 3). By statistical treatment of the ratios and an overall assessment of results from all analytical levels, the oil spill identification using this methodology can be concluded with respect to one of four operational and technically defensible terms: positive match, probable match, inconclusive, or nonmatch.

The revised methodology has been implemented by many European forensic oil spill laboratories and as such has been used recently in connection to several oil spill identification cases in, for example, Norway (e.g., Almås and Daling, 2001), Denmark (Hansen et al., 2002), and the Netherlands (e.g., the *Tricolor* incident in the British Channel, and the *Prestige* incident in Spain and France) (Guyomarch, 2002).

The methodology has been demonstrated to be a strong, technically defensible tool capable of differentiating among qualitatively similar oils from a spill and available candidate sources.

Acknowledgment

Nordtest, Standards Norway (SN), and the Norwegian Pollution Control Authorities (SFT) are gratefully acknowledged for co-financing this project.

References

Almås, I.K. and P.S. Daling, *Identification of Oil Spill Observed by "KV Lafford,"* October 2001, SINTEF report STF66 F01178, 2001 (in Norwegian).

Barakat, A.O., Y. Qian, M. Kim, and M.C. Kennicutt II, Compositional changes of aromatic steroid hydrocarbons in naturally weathered oil residues in the Egyptian western desert, *Environmental Forensics*, 2002, **3**, 219–225.

Christensen, J.H., A.B. Hansen, G. Tomasi, J. Mortensen, and O. Andersen, Integrated methodology for forensic oil spill identification, *Environmental Science & Technology*, 2004, **38**, 2912–2918.

Christensen, J.H. and G. Tomasi, Multivariate statistical methods for petroleum hydrocarbon fingerprinting and spill source identification, Chapter 9, this volume.

Dahlmann, G., Characteristic Features of Different Oil Types in Oil Spill Identification, Berichte des BSH, 2003, **31**, (Available at http://www.bsh.de/de/Produkte/Buecher/Berichte/Bericht31/index.jsp).

Dahlmann, G. and P. Kienhuis, Oil Spill Identification — Round Robin 2005, The Comparison of Four Bilge Samples, RIZA Report No. 2006.043X, 2006.

Daling, P.S., L.-G. Faksness, A.B. Hansen, and S. Stout, Improved and standardized methodology for oil spill fingerprinting, *Environmental Forensics*, 2002, **3**, 263–278.

Ettre, L.S., Nomenclature for chromatography, *IUPAC Recommendations*, 1993, Section 4.3, *Pure & Applied Chemistry*, 1993, **65**(4), 819–872.

CEN, European Committee for Standardization, Oil Spill Identification — Waterborne Petroleum and Petroleum Products — Part 1: Sampling, TC/BT TF 120 WI CSS27004, 2005.

CEN, European Committee for Standardization, Oil Spill Identification — Waterborne Petroleum and Petroleum Products — Part 2: Analytical Methodology and Interpretation of Results, TC/BT TF 120 WI CSS27003, 2006.

Faksness, L.G., H. Weiss, and P.S. Daling, Revision of the Nordtest methodology for oil spill identification — technical report, SINTEF report STF66 A01028, 2002a.

Faksness, L.G., P.S. Daling, and A.B. Hansen, CEN/BT/TF 120 Oil Spill Identification, Summary Report: Round Robin Test Series B; SINTEF report STF66 A02038; 2002b.

Faksness, L.-G., P.S. Daling, A.B. Hansen, P. Kienhuis, and R. Duus, New guidelines for oil spill identification of waterborne petroleum and petroleum products, *Proc 28th Arctic and Marine Oilspill Program* (*AMOP*) *Technical Seminar*, Calgary, Canada, June 7–9, 2005, Environment Canada, Ottawa, **1**, 183–202.

Guyomarch, J., Identification du fuel du Prestige-2002, CEDRE Report no. GC.02-15; 2002 (in French).

Hansen, A.B., J. Avnskjold, and C.A. Rasmussen, Application of PAH and biomarker diagnostic ratios in forensic oil spill identification by the revised Nordtest methodology, *Oil and Hydrocarbon Spills III*, *Modeling, Analysis*, *and Control*, WIT Press, Southampton, UK, 2002, 59–66.

ISO 5725-2, Accuracy (trueness and precision) of measurement methods and results, Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method, International Organization for Standardization, ISO TC69/SC6, 1994a.

ISO 5725-6, Accuracy (trueness and precision) of measurement methods and results, Part 6: Use in practice of accuracy values, International Organization for Standardization, ISO TC69/SC6, 1994b.

Kienhuis, P.G.M., Oil spill identification — Round robin 2004, the comparison of three gas oil samples, RIZA Report No. 2005.042X, 2005.

Nordtest, *Oil Spill Identification*, NT Chem 001, 2nd ed., 1991.

Peters, K.E., C.C. Walters, and J.M. Moldowan, *The Biomarker Guide*, 2nd ed., Cambridge University Press, Cambridge, UK, Vol. 2, *Biomarkers and Isotopes in the Petroleum Exploration and Earth History*, 2005, 475–1155.

Stout, S., W.P. Naples, A.D. Uhler, K.J. McCarthy, L.G. Roberts, and R.M. Uhler, Use of quantitative biomarker analysis in the differentiation and

characterization of spilled oil, Paper prepared for the SPE International Conference on Health, Safety, and the Environment, Stavanger, Norway, 26–28 June, 2000.

Stout, S.A., A.D. Uhler, and K.J. McCarthy, A strategy and methodology for defensibly correlating spilled oil to source candidates, *Environmental Forensics*, 2001, **2**, 87–98.

Stout, S.A., A.D. Uhler, K.J. McCarthy, and S. Emsbo-Mattingly, Chemical fingerprinting of hydrocarbons, *Introduction to Environmental Forensics*, B. Murphy and R. D. Morrison (eds.), Academic Press, San Diego, CA, 2002, 137–260.

Strøm-Kristiansen, T., A. Lewis, P.S. Daling, J.N. Hokstad, and I. Singsaas, Weathering and dispersion of naphthenic, asphaltenic and waxy crude oils, *Proc. 1997 Inte. Oil Spill Conf,* 1997, 631–636.

Wang, Z., M. Fingas, and D.S. Page, Oil spill identification, *J. Chromatography A,* 1999, **843**, 369–411.

Wang, Z. and M. F. Fingas, Development of oil hydrocarbon fingerprinting and identification techniques, *Marine Pollution Bull*., 2003, **47**(9–12), 423–452.

Wang, Z., C. Yang, M. Fingas, B. Hollebone, X. Peng, A.B. Hansen, and J.H. Christensen, Characterization, weathering, and application of sesquiterpanes to source identification of spilled lighter petroleum products, *Environmental Science & Technology*, 2005, **39**, 8700–8707.

Weiss, H.M., A. Wilhelms, N. Mills, J. Scotchmer, P.B. Hall, K. Lind, and T. Brekke, *NIGOGA — The Norwegian Industry Guide to Organic Geochemical Analyses*, Edition 4.0, Published by Norsk Hydro, Statoil, Geolab Nor, SINTEF Petroleum Research, and the Norwegian Petroleum Directorate, 2000, (Available at http://www.npd.no/engelsk/nigoga/nigoga.pdf.)