



BSI Standards Publication

Oil spill identification — Waterborne petroleum and petroleum products

Part 2: Analytical methodology and
interpretation of results based on GC-FID
and GC-MS low resolution analyses

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

raising standards worldwide™

National foreword

This Published Document is the UK implementation of CEN/TR 15522-2:2012. It supersedes PD CEN/TR 15522-2:2006 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee PTI/12, Petroleum Measurement and Sampling.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2012. Published by BSI Standards Limited 2012

ISBN 978 0 580 78751 5

ICS 75.080

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 October 2012.

Amendments issued since publication

Date	Text affected
------	---------------

TECHNICAL REPORT
RAPPORT TECHNIQUE
TECHNISCHER BERICHT

CEN/TR 15522-2

October 2012

ICS 75.080

Supersedes CEN/TR 15522-2:2006

English Version

Oil spill identification - Waterborne petroleum and petroleum products - Part 2: Analytical methodology and interpretation of results based on GC-FID and GC-MS low resolution analyses

This Technical Report was approved by CEN on 13 August 2012.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Former Yugoslav Republic of Macedonia, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and United Kingdom.



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Management Centre: Avenue Marnix 17, B-1000 Brussels

Contents

Page

Foreword.....	6
Introduction	7
1 Scope	9
2 Normative references	9
3 Terms, definitions and abbreviated terms	9
3.1 General.....	9
3.2 Sample comparison.....	10
3.3 Conclusions	11
3.4 Abbreviated terms	11
4 Strategy for the identification of oil spills sources	12
4.1 Introduction	12
4.2 Basis for reliable conclusions – Numerical comparisons.....	12
4.3 Overview of the procedure	13
4.3.1 Sampling and sample preparation	13
4.3.2 GC-FID and GC-MS analysis.....	14
4.3.3 Conclusions and reporting	14
5 Sample preparation	16
5.1 General.....	16
5.2 Visual examination and description of samples	16
5.3 Preparation	16
5.3.1 General.....	16
5.3.2 Water samples.....	16
5.3.3 Oil samples from an Ethylene tetrafluoroethylene (ETFE) net	17
5.3.4 Thick oil and emulsified oil samples	17
5.3.5 Tar balls and emulsified lumps	17
5.3.6 Samples from oiled birds, fish and other animals and vegetation.....	17
5.4 Sample clean-up	18
5.4.1 General.....	18
5.4.2 Biogenic materials	18
5.4.3 Black oil/HFO (removing of asphaltenes and/or soot particles).....	18
5.5 Recommended injection concentration	18
6 Characterisation and evaluation of analytical data.....	19
6.1 General.....	19
6.2 Characterisation by GC-FID – Level 1	20
6.2.1 General.....	20
6.2.2 Evaluation of the influence of weathering on sample comparison	21
6.2.3 Acyclic isoprenoids ratios	22
6.2.4 Level 1 Conclusions	22
6.3 Characterisation by GC-MS – Level 2.....	22
6.3.1 General.....	22
6.3.2 Visual inspection and overall characterisation - Level 2.1.....	23
6.3.3 Treatment of the GC-MS results – Level 2.2	23
6.4 Treatment of the results using the MS-PW-plot– Level 2.2	23
6.4.1 General.....	23
6.4.2 PW-plot calculations	24
6.4.3 Evaluation of the variability of the analysis and peak integration	24
6.4.4 Evaluation of weathering	26
6.4.5 Evaluation of mixing.....	29

6.5	Treatment of the results using ratios – Level 2.2.....	31
6.5.1	General	31
6.5.2	Diagnostic ratios calculation	32
6.5.3	Normative diagnostic ratios	32
6.5.4	Analytical error	35
6.5.5	Match-criterion for ratios	35
6.5.6	Criteria for selecting, elimination and evaluating diagnostic ratios	36
6.5.7	Optional: Evaluation of diagnostic ratios using conventional or multivariate statistics	39
6.6	Conclusions	40
Annex A	(normative) GC-FID analysis	43
A.1	General	43
A.2	Analytical standards for GC-FID analyses	43
A.2.1	N-alkanes	43
A.2.2	Injection concentration of the standard GC-FID	43
A.2.3	Storage of standard solutions.....	44
A.3	Suggested instrumental conditions	44
A.4	Measures to improve and verify the accuracy of the method – GC-FID	44
A.4.1	Mass discrimination	44
A.4.2	Column resolution	45
A.4.3	Calibration range	46
A.4.4	Mid-level concentration	46
A.4.5	Variance.....	47
A.5	Sample analysis with GC-FID	47
Annex B	(normative) GC-MS analysis	48
B.1	General	48
B.2	Analytical standards for GC-MS analyses	48
B.2.1	General	48
B.2.2	SINTEF oil mixture.....	49
B.2.3	Analytical standards for PAH homologues	49
B.2.4	Storage of standard solutions.....	49
B.3	Suggested instrumental conditions	49
B.3.1	GC conditions for the exchange of analytical results.	49
B.3.2	MS conditions for full-scan analysis.....	52
B.3.3	MS preparation for selected ion monitoring (SIM) analysis.....	52
B.4	Measures to improve and verify the accuracy of the GC-MS method	53
B.4.1	Relative retention time	53
B.4.2	Mass discrimination	53
B.4.3	Peak symmetry and column resolution	53
B.4.4	Patterns	54
B.4.5	Calibration range	54
B.4.6	Mid-level concentration	54
B.4.7	Variance.....	54
B.5	Sample analysis with GC-MS	54
Annex C	(informative) List of PAHs and biomarkers analysed by GC-MS-SIM	55
Annex D	(informative) Alkyl homologue patterns of PAHs	57
Annex E	(informative) Diagnostic ratios	65
E.1	Diagnostic ratios of PAHs	65
E.2	Diagnostic ratios of biomarkers	69
Annex F	(informative) General composition of oils – chemical groups.....	76
F.1	Introduction.....	76
F.2	Hydrocarbons	76
F.3	Paraffins	76
F.4	Naphthenes	77
F.5	Aromatics	77
F.6	Heteroatomic organic compounds	77
F.7	Resins	77

F.8	Asphaltenes	77
Annex G (informative)	Weathering of oils spilled on water	79
G.1	Introduction	79
G.2	Evaporation	80
G.3	Dissolution	82
G.4	Re-distribution of chemical composition	83
G.5	Biodegradation	86
G.6	Photooxidation	86
G.7	Contamination	88
Annex H (informative)	Characteristic Features of Different Oil Types in Oil Spill Identification	89
H.1	Introduction	89
H.2	Light fuel oil (gas oil, diesel, fuel No 2)	89
H.2.1	General	89
H.2.2	Analysis, GC screening	90
H.2.3	MS analysis (alternative parameters)	92
H.2.4	Addition of biodiesel	94
H.3	Lubricating oil	95
H.3.1	General	95
H.3.2	Analysis	95
H.4	Heavy fuel oil (HFO, Bunker C, Fuel No 6)	99
H.4.1	General	99
H.4.2	Analysis	99
H.5	Waste oil (bilge oil, sludge, slops)	107
H.5.1	General	107
H.5.2	Analysis	108
H.6	Crude oil	113
H.6.1	General	113
H.6.2	Analysis	113
H.7	Conclusion	118
Annex I (informative)	Example of internal documentation – technical report of an oil spill case	120
I.1	General	120
I.2	Sample information	120
I.2.1	General	120
I.2.2	Contact information	120
I.2.3	Request	120
I.2.4	Photo(s) of the samples	121
I.3	Sample preparation and analyses	121
I.4	Quality assurance	124
I.5	GC-FID results	125
I.6	GC-MS results	128
I.6.1	General	128
I.6.2	Comparison of the surface water samples	129
I.6.3	Comparison of the spill samples with bilge Sample 6	130
I.7	Conclusions	131
I.7.1	Surface water Sample 1 with bilge Sample 6	131
I.7.2	Surface water Sample 2 with bilge Sample 6	132
I.7.3	Final conclusion	132
Annex J (informative)	Example of external documentation – identification report of an oil spill identification case	133
J.1	Introduction	133
J.2	Sample information	133
J.3	Analytical procedure	133
J.3.1	Method	133
J.3.2	Dilution/extraction	133
J.3.3	Analyses	133
J.4	Results	133
J.5	Interpretation	134

J.5.1	General	134
J.5.2	Positive match	134
J.5.3	Probable match.....	134
J.5.4	Inconclusive	134
J.5.5	Non-match	134
J.6	Conclusions	134
Bibliography		135

Foreword

This document (CEN/TR 15522-2:2012) supersedes CEN/TR 15522-2:2006, which was prepared by CEN/BT/TF 120 "Oil Spill Identification" (now disbanded).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN [and/or CENELEC] shall not be held responsible for identifying any or all such patent rights.

This document supersedes CEN/TR 15522-2:2006.

CEN/TR 15522 is composed of the following parts:

- *Part 1: Sampling;*
- *Part 2: Analytical methodology and interpretation of results based on GC-FID and GC-MS low resolution analyses* (the present document).

Introduction

This Technical Report describes and recommends a forensic methodology for characterising and identifying the source of waterborne oils resulting from accidental spills or intentional discharges. The methodology may be used in support of the legal process as evidence for prosecuting offenders ("potential responsible party" – PRP). This methodology is a technical revision of CEN/TR 15522-2 Version 1 published in December 2006.

This methodology is composed of two parts that are described by the following CEN documents:

- *Part 1 – Sampling*: describes sampling techniques and the handling of oil samples prior to their arrival at the forensic laboratory;
- *Part 2 – Methodology*: covers the general concepts and laboratory procedures of oil spill identification methodology, analytical techniques, data processing, data treatment, and interpretation/evaluation and reporting of results.

Oil spill source identification is a complex methodology due to the large variation in samples and oil spill situations that can be encountered. Part 1 is a compilation of instructions and experiences from experts all over the world which will guide the user in sampling, storing and delivering oil samples for laboratory analysis. Part 2 will guide the reader through the analytical process. It prescribes how to prepare and analyse oil samples using GC-FID and GC-low-resolution mass spectrometry (MS). Any chemical differences found between samples are only relevant if a difference is larger than the variability of the method itself. Good analytical performance and strict quality assurance are therefore essential. In the Annexes of Part 2, relevant information concerning different types of oil and oil comparison techniques is presented.

The main purpose of the methodology described in this Technical Report (TR) is to defensibly identify the source of oil spills in marine, estuarine and other aquatic environments by comparing the chemical compositions of samples from spills with those of suspected sources. The underlying basis for this method is the widely variable nature of oils with respect to their specific chemical compositions, which allows oils from different sources to be readily distinguished using the appropriate analytical methods. The method relies upon detailed chemical characterisation and statistical comparison between samples' (i.e., a spilled oil and a suspected source) diagnostic features in order to determine whether they "match". To minimise the danger of "false positive matches", good laboratory practices are necessarily maintained. Even so, a "positive match" between a spilled oil and a suspected source may not be used alone to identify the "potential responsible party" (PRP), but this result is often a critical piece of evidence in proving a case within the legal process.

However, in some oil spill identification cases, both the oil spill and also suspected source(s) may not necessarily be unique or homogeneous in nature, e.g., due to the changing/variable nature of oil in the bilge tanks or due to mixing of oils spilled from several sources in a case of a larger incident. The risk therefore exists that the chemical composition of the available source samples may not match to that of the available spill samples. In such cases, oil spill identification methodologies in general will have limitations and may not necessarily lead to unequivocal conclusions. In other words, the success of this methodology in defensibly identifying a spilled oil's source depends upon the samples available for chemical study. To minimise the danger for "false positive" or "false non-matches", good sampling practice, and particularly the need to obtain appropriate reference/suspect source samples, is crucial (as described in *Part 1: Sampling*).

When oil from suspected sources is not available, this methodology may still be used to characterise the spilled oil in order to determine the spilled oil type and any specific characteristics. The characterisation of a spilled oil sample can still be useful for several reasons:

- If the source of an oil pollution event is unknown, the investigating authorities should be advised on the type of oil in order to aid in the identification of a possible source. For example, in the case of a "mystery" spill, the mere differentiation between pure, unused refined petroleum products (e.g. diesel fuel versus heavy fuel oil) or versus crude oil or waste oil (e.g., bilge residues, sludge, slops) can provide potentially

valuable information as the possible source(s) for the spill. In such instances, the type of oil spilled should be identified rapidly because the chances of identifying and collecting candidate source oils generally decrease with time.

- In some court trials, the differentiation between pure refined products and waste oil may be very important because it allows conclusions to be drawn regarding the cause of an oil discharge, e.g. technical failure, accidental discharge, intentional discharge.
- In some countries, photos (e.g. taken from an airplane) from a plume behind a ship, combined with the evidence that the plume contains mineral oil, is enough for a condemnation.
- Finally, characterisation of the spilled oil provides a baseline against which future impacts to the affected area/environment might be compared.

This Technical Report is the result of advancements in the field of oil spill identification [e.g., 13, 21, 44, 46 and 50] that have been made since the Nordtest Method [35, 36] was first introduced in 1991. These have included:

- advancements in analytical methodologies;
- improved understanding of the specific chemical compositions and diagnostic features of oils;
- improved understanding of how an oil's composition may change in the environment (e.g., due to weathering);
- improvements in the statistical and numerical analysis of chemical data.

These advancements have been made by researchers around the world and documented in a wide range of peer-reviewed literature. In addition, numerous round robin tests have been conducted to evaluate and improve upon the methodology. Since 2004, in the framework of Bonn-OSINET (Bonn-greement Oil Spill Identification Network), annual round robin tests are organised jointly by RWS-WD (Rijkswaterstaat - Center for Water Management in the Netherlands) and BSH (Bundesamt für Seeschifffahrt und Hydrographie in Germany) in which laboratories from around the world participate. The round robin tests have covered oil spill cases dealing with light fuel oil distillates (diesel oils), bilge water samples (a mixture of gas oils and lube oil), crude oils and heavy fuel oils. Findings from these RR-tests have been discussed at annual meetings by the participating scientists and have been taken into account for refining the suggested methodology described herein. The final reports of the RR-tests can be downloaded for free from the Bonn-OSINET part of the Bonn-agreement website [7].

1 Scope

This Technical Report (TR) describes a methodology to firstly identify the specific nature of oils spilled in marine, estuarine and aquatic environments and secondly compare the chemical composition of spilled oil or oily samples with that of suspected sources. Specifically, the TR describes the detailed analytical methods and data processing specifications for identifying the specific nature of waterborne oil spills and establishing their correlation to suspected sources. Even when samples or data from suspected sources are not available for comparison, establishing the specific nature (e.g., refined petroleum, crude oil, waste oil, etc.) of the spilled oil may still help constrain the possible source(s) of the spilled oil.

This methodology is restricted to petroleum and petroleum products containing a significant proportion of hydrocarbon-components with a boiling point above 200°C. Examples are: crude oils, higher boiling condensates, diesel oils, residual bunker or heavy fuel oils, lubricants, and mixtures of bilge and sludge samples. While the specific analytical methods may not be appropriate for lower boiling oils (e.g. kerosenes, jet fuels, or gasoline), the general concepts described in this methodology, i.e., statistical comparison of weathering-resistant diagnostic ratios, may have applicability in spills involving lower boiling oils.

This method is not directly intended for oil spills impacting groundwater, vegetation, wildlife/tissues, soils, or sediments, and although its application in these matrices is not precluded, it requires caution. The reason for caution is that the extractable compounds in these matrices may alter and/or contribute additional compounds compared to the source sample, which if left unrecognised, can lead to “false non-matches”. Including these “non-oil” matrices in this oil spill identification method may require additional sample preparation (e.g. clean-up) in the laboratory prior to analysis and consideration of the extent to which the matrix may affect the correlation achieved. Evaluating the possible effects in these matrices is beyond the scope of this guideline. Whether the method can be used for this kind of “non-oil” matrices may depend on the oil concentration compared to the “matrix concentration” of the samples. In “non-oil” matrices containing a relative high concentration of oil, a positive match can still be concluded. In “non-oil” matrices containing a relative low concentration of spilled oil, a non-match or an inconclusive match could be achieved due to matrix effects.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

CEN/TR 15522-1, *Oil spill identification – Waterborne petroleum and petroleum products – Part 1: Sampling*

3 Terms, definitions and abbreviated terms

For the purposes of this document, the following terms and definitions apply.

3.1 General

3.1.1

chain of custody

line of recorded actions taken for samples collected from spill and suspected sources at court for safe surveillance and storing; to ensure that the samples have not been tampered with or altered accidentally

3.1.2

mixing

mixing can include chemical changes brought about by (a) the mixing of multiple oils, (b) mixing with pre-existing background compounds from biogenic or anthropogenic sources, or (c) heterogeneity within the sample(s) (e.g., within a vessel, tank, or oil slick)

3.1.3

sample heterogeneity

non-representative or non-homogenous character of samples caused for example by variable degrees of mixing within a vessel, tank or oil slick

3.1.4

contamination

changes in oil composition which take place during/after the spillage, by mixing with additional compounds, including naturally-occurring chemicals or other products

3.1.5

waterborne oil

petroleum and petroleum products borne by water or available in the water column from marine, estuarial and aquatic environments

Note 1 to entry: These environments include lakes and rivers but exclude groundwater.

3.1.6

weathering

changes in oil composition which take place after the spillage, including evaporation, dissolution, emulsification, oxidation and biological decomposition

Note 1 to entry: See also Annex G.

3.1.7

bilge water

mixture including water and oil collected in the bilge of the machinery space in a ship as a result of leakage, drainage, etc.

3.1.8

slop

mixture of water and oil residues from cargo tanks in oil tankers that may contain oil/water emulsions, wax, sediments and other tank residues

3.1.9

sludge

deposits, generally from the purification of fuel and lubrication oils, consisting of mixtures including oil, wax, sand and water

3.1.10

tank washings

tank washing water containing cargo tank residues including oil, wax, sediment and other foreign matter such as tank cleaning chemicals

3.2 Sample comparison

3.2.1

PW-plot

graph based on GC-FID or GC-MS data of two samples normalised to a non-weathered compound or group of compounds and sorted on boiling point or retention time

Note 1 to entry: The name "PW-plot" is originally a reference to Per Wrang, who introduced the plot in the Nordtest method [36]. In this TR, the name PW-plot will be used as an abbreviation of a "Percentage Weathering" plot.

3.2.2

diagnostic ratios (DR)

ratios between the peak height or peak area of single compounds or compound groups selected by their diversity in chemical composition in petroleum and petroleum products and on their known behaviour in weathering processes

3.2.3

critical difference (CD)

14 % of the mean value of a ratio for two different samples

Note 1 to entry: The fixed value of 14 % is based on the maximum allowable relative standard deviation of 5 % for the ratios [23, 24, 25].

3.3 Conclusions

3.3.1

positive match

differences in the chromatographic patterns and diagnostic ratios of samples submitted for comparison are lower than the variability of the method or can be explained unequivocally, for example by weathering

Note 1 to entry: The samples are considered to match to a high degree of scientific certainty.

3.3.2

probable match

differences in chromatographic patterns and diagnostic ratios do not permit an unequivocal positive match, but they can be explained reasonably by external factors, for example weathering in combination with mixing or by non-representative or heterogeneous properties of the available samples

Note 1 to entry: The samples are considered to match to a reasonable degree of scientific certainty.

Note 2 to entry: Unavailability of a representative source sample. For example, if a vessel that has discharged all of its bilge water, a representative source sample may not be available. Therefore, when comparing an individual source lubricating oil with the lubricating oil component of a bilge sample, certain differences have to be expected due to other bilge sample components besides the source lubricating oil.

3.3.3

inconclusive

differences in the chromatographic patterns and diagnostic ratios of the samples submitted for comparison, do not permit a probable or non-match conclusion; for example in case the concentration of the contaminant in a sample is too low

3.3.4

non-match

differences in the chromatographic patterns and diagnostic ratios of the samples submitted for comparison are pronounced and are larger than the variability of the method, and such differences cannot be explained by any external factors such as weathering, contamination or heterogeneity

Note 1 to entry: The samples are concluded to not match to a high degree of scientific certainty.

3.4 Abbreviated terms

CD	Critical difference
DR	Diagnostic ratio
FID	Flame ionisation detection
GC	Gas chromatography
HFO	Heavy Fuel Oil
IUPAC	International Union of Pure and Applied Chemistry
LCO	Light cycle oil

LFO	Light fuel oil
MS	Mass spectrometry
NR	Normative ratio
PCA	Principal component analysis
RSD	Relative standard deviation
TR	Technical report

4 Strategy for the identification of oil spills sources

4.1 Introduction

Identification of spilled oils in the context of this Technical Report (TR) implies the comparison of the total chemical composition of the spilled oil with that of candidate source samples.

The likeness of a source and spill sample should be tested by analyzing the samples with GC-FID and/or low resolution GC-MS and by comparing their detailed chemical compositions using a suite of generic and diagnostic petroleum components. If no or only insignificant differences (i.e. differences being smaller than the analytical variability) are observed, a “positive match” should be concluded. 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, mixing or heterogeneity), that are larger than the variance of the analysis are observed within these diagnostic compounds, it should be concluded that the samples are a “non-match”. Some investigations may result in conclusions intermediate to “positive match” and “non-match”, such as “probable match” or “inconclusive” (see definitions in 3.3).

NOTE It is practically and technically impossible to measure and compare every chemical in spilled oil and its prospective source in order to conclude a positive match exists. Therefore, in practice, two samples are considered to be a positive match if no statistically significant differences in diagnostic metrics determined by GC-FID and GC-MS analysis are present, that cannot be explained by weathering, mixing or heterogeneity. This approach, i.e., looking for differences in diagnostic features instead of similarity among every possible feature, is conceptually more logical and more practically and technically achievable. As such, only distinct differences between samples can be proved. Therefore, when no statistically significant differences between samples are observed, a “positive match to a high degree of scientific certainty” should be concluded.

4.2 Basis for reliable conclusions – Numerical comparisons

The usual practice is to analyse samples qualitatively and then compare the chromatograms and ion chromatograms visually, such as described in the original Nordtest Method (1991) [36] and the ASTM methods [2, 3]. The outcome of such qualitative comparisons is subjective and depends on the experience or bias of the analyst. Because of the high complexity of oils, and the many details that can be compared in the often very complex chromatograms, qualitative comparisons should always form an integral part of oil sample comparisons. However, in order to make conclusions more objective, reproducible, and therefore, more reliable, this Technical Report also requires the use and comparison of quantitative metrics; i.e., specific peaks or groups of peaks have to be measured and peak ratios have to be calculated and compared.

These ratios are produced in two different ways:

- a) Measurements of single compounds normalised to hopane – or, if hopane is not sufficiently present, to tetra-methylphenanthrene (T-M-phe; see Figure E.2).

Comparing the normalised abundances of compounds spanning the boiling range of oil in the form of PW-plots, reveals:

- 1) the variance of the analysis and integration,
- 2) concentration differences between compounds (e.g. biomarkers and PAH's) and
- 3) the effects of weathering on individual compounds and compound groups, and in turn, on prospective ratios based upon those compounds.

When the effects of weathering on a spilled oil are understood, any affected ratio can be eliminated or cautiously considered when comparing a weathered spill sample to an unweathered source sample.

- b) Diagnostic Ratios (DRs): for keeping the analytical error low, measurements of compounds (peaks) that are primarily recorded by the same m/z value via GC-MS within a limited retention time range.

NOTE 1 The term "diagnostic ratio" originates from geochemistry, and is related to the diversity in chemical composition of crude oils from different geologic provinces, oil fields, individual wells or reservoirs. DRs are characteristics, which collectively may lead to a unique classification and identification of oils ("diagnosis": "sorting into categories"). The term "Diagnostic Ratio" is used here in order to pronounce the general differentiation possibilities of these ratios and thus their importance, (see 3.2.2). In addition, this term is widely used in oil spill identification literature [44, 46].

Whereas most of the DRs described in this TR may be used when spills involving crude oil, bunker oil and bilge samples are investigated, only a limited number of DRs may be useful for spills involving lighter fuel oils (e.g. kerosene, diesel, gas oil). The reason for this is because the DRs based upon higher boiling biomarkers may not be present in lighter refined products. Similarly, the higher boiling lubricating oils, normally rich in biomarkers, may not or not sufficiently contain aromatics and n-alkanes, rendering DRs based on these compounds useless.

Because of the need for some flexibility when investigating spills of different oil types, a specific list of normative diagnostic ratios (NRs) generated from potentially diagnostic and more weathering resistant alkanes, PAH compounds and biomarkers has been developed for crude oil and five other types of petroleum (see 6.4.4). This list should be treated as a minimum list, which can be expanded for specific cases and/or oil types. Examples of additional "informative" ratios are given in Annex E. A much more comprehensive review of potentially diagnostic ratios, which may be used/are useful in oil spill identifications, can be found in [46]. The resulting ratios are compared using a critical difference of 14 % (see 3.2 and 6.4.6).

The combined information derived from DRs, PW-plot, and the visual comparison of chromatograms (see Figure 11 and I.7) gives a deep insight in the differences and similarities between two oil samples.

NOTE 2 The present methodology is based on semi-quantitative analyses for establishing diagnostic ratios of whole samples (no fractionation in aliphatic and aromatic compounds [1]). For those laboratories who work daily on a "quantitative analytical" level (i.e. internal standards are added and concentrations of individual compounds or compound groups are calculated), and that have good laboratory QA/QC protocols for doing quantitative analysis of the recommended diagnostic compound peaks, the ratios recommended in these guidelines may be established based on quantitative analysis in which the absolute concentrations are used to calculate DRs [15, 16, 23].

4.3 Overview of the procedure

4.3.1 Sampling and sample preparation

When an oil spill has been discovered, samples should be taken from the current spill and from any potential responsible parties such as suspected ships or other sources. The sampling shall be carried out in accordance with CEN/TR 15522-1. When suspected sources are not available, the methodology can still be used to characterise the spilled oil.

All samples should be sent under chain-of-custody either via an authorised "Sampling Coordinator" or directly to a forensic laboratory for oil spill identification.

At the laboratory the samples are visually characterised, described, see 5.2, and prepared for analysis (conservation, extraction and clean-up if needed), see 5.3

4.3.2 GC-FID and GC-MS analysis

Oil sample comparison by GC-FID (see Annex A) and GC-MS (see Annex B) is conducted in both cases by visual (qualitative) comparisons and by numerical (quantitative) comparisons (Figure 1).

GC-FID is used for the screening of all samples. Based upon the visual comparison of the gas-chromatograms, source samples that are clearly a different type of petroleum than that which comprises the spill sample(s) may be ruled out as “non-matching”. For example, a diesel fuel potential source sample may be eliminated from further consideration in a crude oil spill. Differences due to weathering of the spill sample are recognised and visualised by overlays of the gas-chromatograms (Level 1.1). For source samples which are not clearly a different type of oil than the spill, n-alkanes and isoprenoids, normally the most abundant components of oils, are measured and compound ratios are calculated and compared. A proper visualisation of the results is achieved by GC-PW-plots, where alkane-ratios of the spill sample(s) are plotted against those of the comparison sample(s) (Level 1.2).

NOTE Level 1.2 is optional. Its utility lies in the potential to eliminate some clearly distinct source sample for analysis by GC-MS (Level 2) if justified. However, if it has already been decided to conduct Level 2 GC-MS analysis, Level 1.2 is not essential, except for spills involving light petroleum products like diesel. In this type of spill the very low concentration of biomarkers in light products and the high concentration of alkanes, the alkane pattern, the isoprenoid ratios and the GC-PW-plot can provide essential input to the final decision.

If there are no obvious or only smaller significant differences evident following Level 1, i.e. differences, which are below or only slightly above the analytical error, Level 2 (Figure 1) is conducted:

GC-MS is used for detailed fingerprinting of the spill and candidate source samples. The steps are the same as given above. In Level 2 however, a large number of compounds representing different compound groups is investigated and compared between the samples. Based upon the visual comparison of the mass-chromatograms (may be supported by overlays), at this point candidate source samples that are clearly different from the spill sample(s) due to reasons other than those that might be attributable to weathering, mixing, or sample heterogeneity, may be ruled out as “non-matching” (Level 2.1). For the remaining samples, results are visualised by plotting ratios of the spill sample(s) against those of the comparison sample(s) in form of PW-plots (Level 2.2). Also at this point, a list of “representative” target compounds is measured and DRs are calculated and compared (Level 2.2).

If there are no or only smaller significant differences, the final conclusion is reached, whereupon the results from all four steps (Levels 1.1 through 2.2; Figure 1) are collectively evaluated.

Level 1 may also be conducted by means of a GC-MS full-scan run, if GC-FID is not available.

4.3.3 Conclusions and reporting

Following completion of the evaluation an internal documentation, a Technical Report is prepared by combining and evaluating all information and all results (see 7.2 and Annex I). A final conclusion(s) that includes the reasons on which the conclusion(s) is based is clearly stated in the internal documentation.

An external documentation - identification report - is prepared for the client/authorities, if requested (see 7.3 and Annex K).

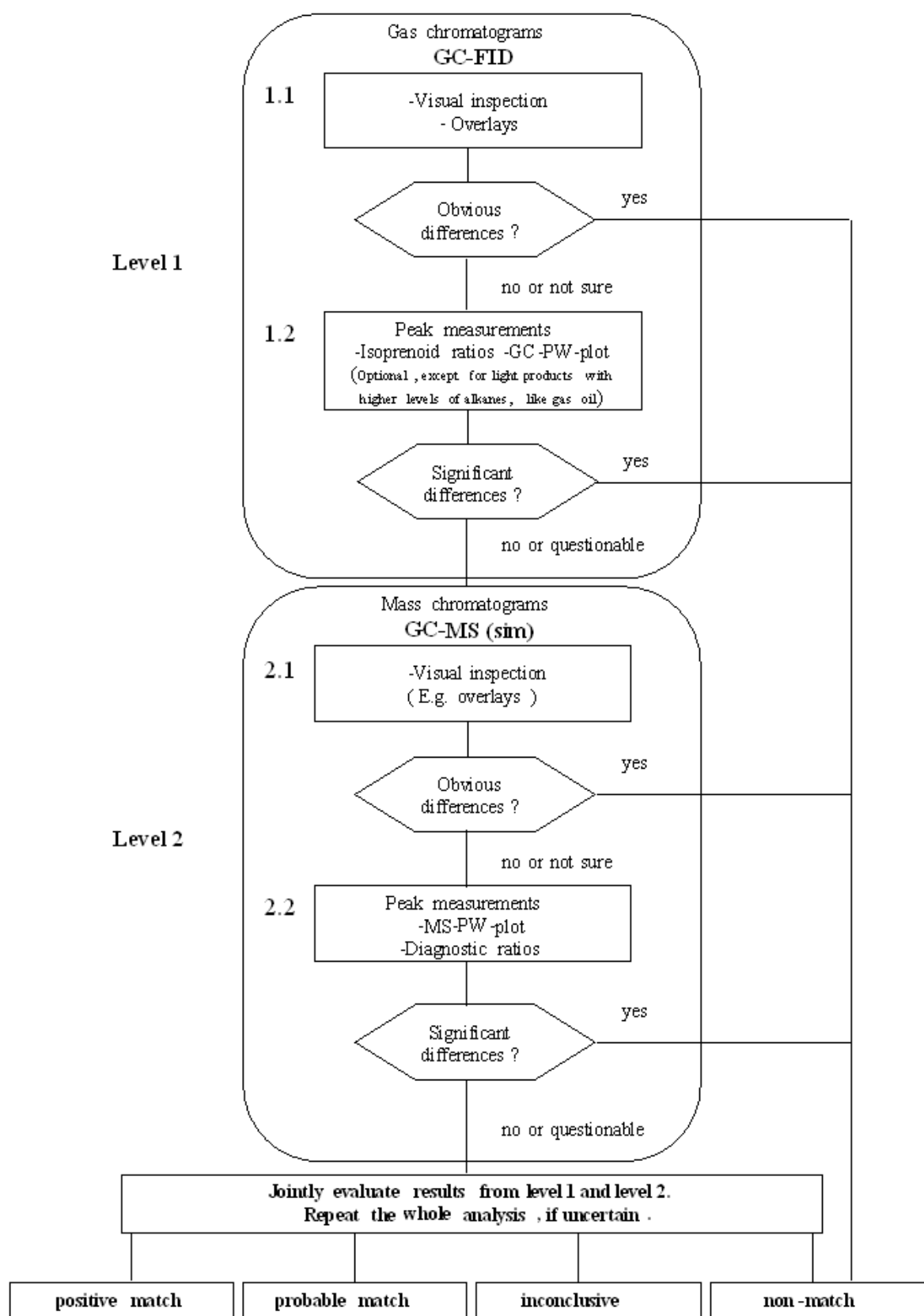


Figure 1 — Protocol/decision chart for oil spill identification

5 Sample preparation

5.1 General

The sample preparation is based on the analysis of an aliquot prepared from the 'whole oil', which benefits from minimising the handling of the sample prior to analysis. The same sample preparation procedure should be used to prepare a single extract for analysis by both GC-FID and GC-MS, i.e., the same extract should be analysed by both instruments. Minimising the handling of the samples prior to analysis minimises any variability potentially introduced during the sample preparation. Maintaining low variability/high precision is essential for the numerical comparison of samples so that small but authentic differences between samples can be recognised. As such, the fractionation of the sample into separate aliphatic and aromatic hydrocarbon fractions prior to analysis is not recommended or necessary [1].

5.2 Visual examination and description of samples

Sampling method (e.g. ETFE net, see 5.3.3) colour, odour, viscosity [4, 5, 31] and content of any free water and debris should be visually noted and recorded. It is also recommended that each sample is photographed in order to document its "as-received" condition. Any pieces of wood, fabric or other debris should be removed from the sample with a pair of tweezers and noted in the report.

5.3 Preparation

5.3.1 General

If a sample contains a significant amount of oil/emulsion, and an aliquot of the sample cannot be immediately prepared after arriving at the laboratory, the sample bottle should be stored in a refrigerator at $(5 \pm 3) ^\circ\text{C}$ or less. Undisturbed storage of the sample may allow oil to separate from water facilitating the oil's collection for preparation. In any case, sample preparation should be carried out as soon as possible, preferably within a week, even in samples comprised of oil/water emulsions.

If a sample contains primarily water and it cannot be immediately extracted, dichloromethane (DCM)¹ should be immediately added after arrival, (e.g. 10 ml to a 1 L water sample). After shaking, the sample should be stored in a refrigerator at $(5 \pm 3) ^\circ\text{C}$. Further sample preparation should be carried out as soon as possible, preferably within a week.

5.3.2 Water samples

If samples primarily consist of water (no visual oil on the surface), the water is transferred to a separator funnel and extracted using DCM. Depending on the amount of oil, the sample should be extracted one by one or serially (e.g. 60 ml, 30 ml and 30 ml of solvent if the sample volume is 1 L, less volume if the water sample is smaller). The inside of the sample bottle is rinsed with DCM before pouring it into the water sample for extraction. The combined extract is dried with sodium sulphate and diluted or concentrated to a suitable injection concentration. If a concentration step is needed, use a gentle evaporation technique (e.g. Kudena Danish, Turbo Vap, N-Evap).

1) If DCM is not allowed, use another appropriate solvent e.g. hexane.

5.3.3 Oil samples from an Ethylene tetrafluoroethylene (ETFE) net

ETFE nets²⁾ are recommended for sampling thin oil films/sheens [49]. The oil should be thoroughly rinsed off the ETFE net using DCM or the ETFE net can be extracted with DCM. After drying with sodium sulphate, the extract should be diluted or concentrated to a suitable injection concentration.

A direct DCM extraction of water with an amount of oil can easily result in emulsions:

- Depending on the layer thickness of the oil film on water in the sample bottle, a small piece of ETFE-net can be useful to take apart (a part of) the oil from the sample in an easy way, followed by a simple extraction of the net with DCM.
- For samples containing water and an ETFE-net in the same bottle: decant the water in a clean vial and start with extracting the oil from the ETFE-net first. In general, enough oil is present for analysis; if not the water can be extracted additionally according 5.3.2.

Depending on batch and/or manufacturer, it might be necessary to pre-clean the nets by rinsing with DCM before use.

5.3.4 Thick oil and emulsified oil samples

Samples consisting primarily of thick oil or emulsified oil should be at room temperature before sample preparation. Any free water should be decanted into a separating funnel. The oil/emulsion fraction is gently homogenised before a 20 mg to 50 mg aliquot is weighed in a pre-weighted volumetric flask (5 ml) and subsequently diluted with DCM to the final volume. The extract should be dried with sodium sulphate. Samples consisting of solid oil or sludge, e.g. high elastic/viscous HFO, should be adjusted to room temperature. When the homogeneity of the sample is questionable, take two (or more) aliquots, 20 mg to 50 mg each, at different locations within the sample container and treat the aliquots as different samples.

5.3.5 Tar balls and emulsified lumps

Oils weathered for a long time at sea and/or on shore may form very viscous oil lumps ("tar balls") often centimetres in diameter. Their degree of weathering may not be homogenous within the lumps, thus in such cases, it is recommended to split the lump and obtain two aliquot samples of oil (20 mg to 50 mg each), one from the middle (inside) of the lump and one from the more exposed surface of the lump. Samples should be diluted in DCM as described in 5.3.4 before analysis.

5.3.6 Samples from oiled birds, fish and other animals and vegetation

Samples from oiled birds, fish and other animals or vegetation should be handled with the necessary care and precaution (e.g. to prevent carryover of contagious diseases). Visible surface oil should be wiped off using a piece of ETFE net. The ETFE net should be rinsed with DCM and, if necessary, the extract should be filtered through a glass fibre filter to remove feathers and other particles. A clean-up of the extract (as described below) may also be necessary if only a small quantity of oil is recovered. If the oil cannot be wiped off, the feathers or other surfaces may be rinsed with hexane and followed by an extract clean-up procedure as described in 5.4.2. When homogeneity is questionable, take multiple aliquots from different areas and treat the aliquots as different samples.

2) ETFE is most well-known by the DuPont brand name Tefzel® and is an example of a suitable commercially available product. This information is given for the convenience of users of this CEN Technical Report and does not constitute an endorsement by CEN of this product. An example is a Tefzel® net manufactured by SEFAR – Internet: www.sefar.com (Sefar Fluortex Product ref.09-150/36 or 9-250/39).

5.4 Sample clean-up

5.4.1 General

Matrix interferences may be caused by contaminants that are co-extracted from the sample. Since biogenic materials and other naturally occurring materials may cause interferences, alumina and/or silica clean-up is recommended in order to eliminate or minimise matrix interference prior to GC-MS and GC-FID analysis.

High boiling compounds not eluting from the GC column strongly influence the performance of the GC injector, column and detector. Therefore “black oils”, such as heavy fuel oils, should also be cleaned using a silica or florisil clean-up in order to remove the high amount of asphaltenes and/or soot particles, that typically are detrimental to instrumentation.

All matrices (oils, sediments/soils, vegetation, wildlife, emulsions) within a given investigation shall be cleaned in the same manner in order to maintain the comparability of the samples.

5.4.2 Biogenic materials

If the potential for contamination of an extract with, for example lipids or fatty acids is very high, it may be necessary to clean the extracts using an alumina column before the silica column [22].

The alumina column should be prepared just prior to use: Pack the tip of a 19 mm ID column with a small piece of glass wool. Add circa 25 ml DCM to the column. Weigh 20,0 g of alumina into the beaker and add 30 ml of DCM. Swirl to remove any bubbles. Slowly pour the alumina into the column while rinsing the beaker with DCM. Allow to settle. Add approximately 2 g of dried sodium sulphate. Drain the DCM from the column until the head of the liquid in the column is just above the sodium sulphate. Close the stopcock to stop solvent flow. Add 50 ml of hexane and drain to the top of the packing material. Transfer the 1 ml sample extract (in DCM) onto the column. Rinse out the extract vial with 1 ml DCM and immediately add it to the column. Allow the sample to drain to the top of the packing material. Add 100 ml of DCM and slowly drain the column (approximately 2 ml/min). Collect the eluate in a 250 ml EM-flask. Alternative clean-up procedures are described in [34].

5.4.3 Black oil/HFO (removing of asphaltenes and/or soot particles)

Use self-packed or pre-packed silica or florisil columns (500 mg of silica or florisil, e.g. Varian, Supelco or Isolute). Apply a thin layer of dried sodium sulphate to the pre-packed columns before use.

Self-packed columns can be prepared from glass Pasteur pipets (typical volume 2,5 ml). Fill each pipet with a small piece of glass wool, (500 ± 50) mg silica or florisil (60 mesh to 100 mesh) and a thin layer of dried sodium sulphate. A pipet pump can be used to increase the flow rate when needed and to remove eventually the DCM from the column.

Clean the column with three bed volumes DCM. Drain the DCM from the column until the head of the liquid in the column is just above the sodium sulphate. Add 2 ml to 5 ml of the sample extract dissolved in DCM at the injection concentration to the column. Extract the column with three bed volumes of DCM and adjust the final extract volume to the applied sample extract volume.

Test each batch of pre-packed columns or each batch of silica or florisil for the self-packed columns, when used for the first time. It is advised to run a reference oil before and after each series (see B.4.6). This reference oil solution can also be used for testing the clean-up column. Treat the reference oil solution as a sample according to the method described above. A match should be found when comparing the resulting GC-MS chromatogram after clean-up with chromatograms of the reference oil without clean-up. Note the result according to the QA system of the laboratory.

5.5 Recommended injection concentration

Oil samples contain a range of compounds in a range of concentrations. For oil spill identification investigations, both high and low concentration levels of target compounds (i.e., those used in calculating

diagnostic ratios) have to be evaluated. Therefore, it is recommended that spill and source samples have about the same pre-injection concentration so that gross concentration differences do not potentially influence the final comparison. It is recommended to use a pre-injection concentration approximating the mid-level injection concentration that is determined during the validation of the instrument/method. The mid-level injection concentration for GC-FID may be determined by analyzing the calibration range of a normal alkane standard mixture (see A.2). For GC-MS, a reference crude oil (see B.2) may be used to find the best peak shape for both high and low level compounds, and a sufficient S/N ratio (see 6.4.7.2) for the low level biomarkers in one injection.

Table 1 shows some recommended pre-injection concentrations of common oil types. When the GC-FID and GC-MS systems are provided with the columns suggested in the Annexes (A.3 and B.3.1), it is strongly advised not to increase the injection concentration to prevent overloading of the GC column(s).

As newer instruments tend to be more sensitive, this table should be adjusted to the laboratory's instrument(s) being used. A laboratory can determine the mid-level concentration as described above for one or two different oil types and reduce the recommended pre-injection concentrations in Table 1 proportionately.

Table 1 — Recommended pre-injection concentrations for different oil types

Oil Type	Injection concentrations mg/ml	Oil type	Injection concentrations mg/ml
Gas oil	1,5 (e.g. 20 µl -> 10 ml DCM)	HFO (Very viscous to solid)	5 to 10
Lubricating oil	2	Bilge	2
Light bunker oil	1,5	Paraffin (white)	0.3
Bunker oil	2	Paraffin (yellow)	2

6 Characterisation and evaluation of analytical data

6.1 General

A duplicate analysis of two aliquots taken from a homogenous sample or the duplicate injection of an extract gives information about the actual performance of the analytical system for the samples involved. Duplicate preparation and analysis of samples provides a measure of the handling and analytical portions of the analytical system, while replicate analyses of a single extract provides a measure of the analytical portion of the system.

For spill investigations involving only oils, replicate analyses of an extract are sufficient, since the preparation or handling of pure oils prior to instrument analysis is minimal. However, for spill investigations involving matrices other than pure oils (e.g., vegetation, wildlife, soils, or sediments) or oils suspected to exhibit heterogeneity, then duplicate samples should be prepared and analysed in order to monitor the precision of both the handling and analytical portions of the analytical system.

Therefore, in order to improve the robustness of the method and to check the variance of the diagnostic ratios used in an investigation, a minimum number of samples should be analysed in duplicate as follows:

- in investigations with two samples: the extracts from both samples;
- in investigations where there are many samples, and duplicate analyses of all samples is impractical: at least 10 %, with a minimum of two samples representing the apparent variety of the samples;

- in investigations of very viscous and solid samples (e.g., tar balls): duplicate analyses of some of the multiple aliquots of potentially heterogeneous samples should be prepared and analysed (see also 5.3.4 to 5.3.6).

All samples' extracts should be analysed under the same analytical conditions as were used in the analysis of analytical standards. Use and preparations of appropriate analytical standards are described in Annex A for GC-FID and Annex B for GC-MS.

Analysis of an external standard is required at the beginning of an analytical sequence, after approximately every tenth injection, and at the end of a sequence in order to check instrument performance and stability during the sequence. Special attention should be given to maintaining the intensity of higher boiling compounds (i.e., monitoring any mass discrimination) and the resolution of close-eluting peaks. It is recommended to log the intensity of some peaks at the beginning, middle and end of the chromatogram and some of the normative ratio's in, for example, a control chart to track system performance during and between sequences.

For labs analyzing oil spill samples in a regular basis, it may be sufficient to run an external standard daily and to log the results in a control chart.

An example of the evaluation on an oil case with six samples is illustrated in Annex I.

6.2 Characterisation by GC-FID – Level 1

6.2.1 General

Following the sample preparation steps described above, the chemical fingerprinting analysis in the laboratory starts with the GC-FID screening analysis of all samples (Figure 1). Annex A describes the minimum and recommended specifications for standards and instrumental conditions for GC-FID equipment used in the screening analysis. If a GC-FID is not available, see B.3.2.

The data from the GC-FID screening should be used for:

- characterising the spill and source sample(s) by determining the overall boiling (carbon) range of the oils, i.e., the total distribution of hydrocarbons including n-alkanes from C10 to C40 if present (for an overview of boiling points of n-alkanes, see Annex F);
- visual inspections of the chromatograms for possible characteristic features and, to the degree possible, a classification to a type(s) of oil comprising the spilled and source oils (see Annex H);
- eliminating source samples clearly unrelated to the spilled oil from the GC-MS analysis (level 2) if they are obvious "non-matches";
- determining the degree of weathering;
- establishing selected normative ratios (C17/pristane; C18/phytane and pristane/phytane);
- adjusting the pre-injection concentration of the samples to the appropriate mid-level concentration (see 5.5) for the GC-MS analysis.

Annex H presents and discusses examples of characteristic GC-FID-chromatograms of different oil spill sample types, including light fuel oils, heavy fuel oils, lubricating oils, ships' bilge, sludge samples, different crude oils and a mixture of gas oil with biodiesel. Examples in Annex H may prove helpful to provide a tentative classification of the type of spilled oil.

6.2.2 Evaluation of the influence of weathering on sample comparison

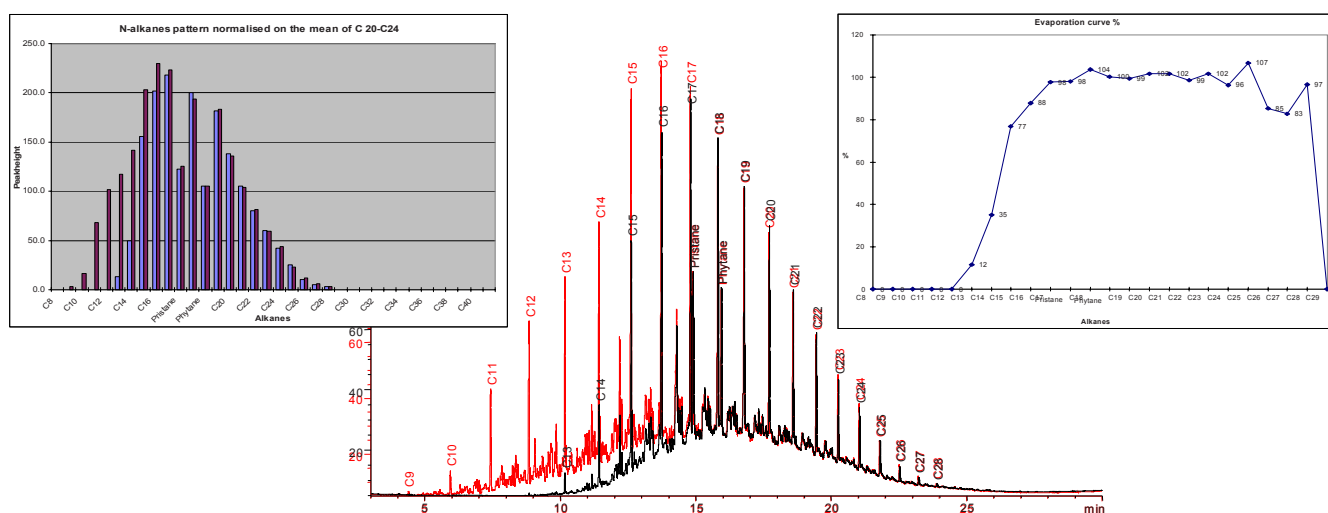
6.2.2.1 General

In oil spill identification investigations, weathering is defined as all the chemical, physical, or biological processes that affect the composition of the spilled oil after the oil enters the environment. When petroleum or petroleum products are spilled onto water, weathering can rapidly alter the spilled oil's "fingerprint" from that of the original oil. Differences caused by weathering should not preclude that a match exist between an unweathered source oil and a weathered spill oil. Therefore, it is important to have a good understanding of how the various weathering processes, such as evaporation, dissolution, photo-chemical oxidation, microbial degradation, wax-enrichment/-depletion, can influence the analytical results. See Annex G and [50] for additional information. Section 6.4 gives examples how weathering can be evaluated in an oil spill identification case.

Several tools can be used to estimate weathering from the data of the GC-FID analysis:

6.2.2.2 Chromatogram overlay

Figure 2 gives an example by a simple overlay of chromatograms of a spilled gas oil sample (in black) and a suspected gas oil source sample (red). The overlay was achieved by normalisation/manipulation these samples' GC-FID chromatograms to a comparable attenuation within the non-weathered portions of each, i.e., C19 to C28 range using, e.g., Power Point™³.



In black: the chromatogram of the spill sample showing a gradual loss of lighter compounds.

Figure 2 — Bar chart, overlays of chromatograms, and GC-PW-plot from an investigation dealing with gas oil.

6.2.2.3 Bar chart (histogram)

Figure 2, left side, shows an example of a bar chart of alkane abundances for the samples shown in Figure 2 that was constructed using Excel³⁾. Compounds abundances were determined by integrating peaks in the GC-FID chromatograms and normalising their areas or heights in order to compensate for concentration

3) Power Point™ and Excel are the trade names of products supplied by Microsoft. This information is given for the convenience of users of this Standard and does not constitute an endorsement by CEN of the product named. Equivalent products may be used if they can be shown to lead to the same results.

differences. The n-alkanes of the two gas oil samples have been normalised to the mean areas of n-C20 to n-C24, which in this case have not been influenced by weathering. The reduced abundances of lower boiling alkanes in the spilled oil (blue) are consistent with losses attributable to evaporation affecting the spilled oil.

6.2.2.4 GC-PW-plot

The right side of Figure 2 shows a “percentage weathering” plot (GC-PW-plot) based on the same data as the bar chart. After normalisation to the mean abundances of C20 to C24, the percentage remaining of each alkane in the spill sample can be calculated when compared to the source sample (see 6.4.2). For example, in the spill sample of Figure 2 only about 1/3 of the concentration of C14 appears to remain, which is actually calculated to be 35 %. On the other hand, C20 in the spill sample is still at the same peak height and is shown as 100 % (or 0 % loss). The higher boiling alkanes in the GC-PW-plot show more variation but this is because they are represented by small peaks with a higher variability in analysis and peak integration.

Except for biodegraded oil samples, the changes in the n-alkanes distributions are caused mainly by evaporation, which progressively reduces the abundances of the lower boiling n-alkanes (i.e., those with higher vapor pressures) with decreasing carbon number. The GC-PW-plot shows this progressive loss of the lower boiling alkanes below C18 and therefore, it can be concluded that the differences between these two gas oil samples is caused by evaporative weathering. As such, diagnostic ratios relying upon compounds boiling below C18 are likely altered by weathering and therefore, should not be used to compare the unweathered source sample to the weathered spill sample.

6.2.3 Acyclic isoprenoids ratios

The acyclic isoprenoids ratios of n-C17/pristane, n-C18/phytane and pristane/phytane are potentially important diagnostic ratios that can be determined from the Level 1 GC-FID chromatograms. However, if the spill sample is too weathered, or if the concentration of these compounds in a sample is low relative to the unresolved complex mixture (UCM), e.g., as in many lubricating oils or HFO's, these ratios are less or unreliable. This is e.g. relevant for the ratios n-C17/pristane and pristane/phytane in the comparison of the samples shown in Figure 2. In such cases, these ratios should be handled with care or excluded from the numerical analysis of the data at level 1.2.

C17 and pristane and also C18 and phytane are progressively affected by evaporation based upon their respective boiling points/vapor pressures. Biodegradation however has a stronger effect on C17 and C18 compared to pristane and phytane (see Figure H.4) and therefore, this weathering process should also be considered when reviewing level 1 data. The comparison of samples by means of these and other ratios is shown and discussed in 6.5.

6.2.4 Level 1 Conclusions

If the GC-FID chromatograms of spill sample(s) and suspected sources show differences in the hydrocarbon distribution, the unresolved complex mixture distribution, and/or the acyclic isoprenoid ratios that obviously are not caused by weathering, mixing, or potential heterogeneities in the spill samples, the suspected source samples can be concluded to be “non-matches” to the spill and eliminated from Level 2 analysis (see Figure 1). If there is any doubt following the Level 1 analysis, prospective source samples should not be eliminated based upon the Level 1 results, and should be analysed using GC-MS in accordance with Level 2.

6.3 Characterisation by GC-MS – Level 2

6.3.1 General

At Level 2, the spill and candidate source samples should be analysed using gas chromatography combined with mass spectrometry (GC-MS). Annex B describes the recommended specifications for standards and instrumental conditions for GC-MS equipment used in the Level 2 analysis and Annex C lists the target compounds.

This GC-MS analysis is necessary for determining the content and distributions of a suite of diagnostic petroleum biomarkers and PAH target analytes. Data from this analysis should be used for:

- visual inspection of the extracted ion (mass) chromatograms;
- possible determination of characteristic features of the spilled oil that may aid in describing the general nature or type of spilled oil;
- classification of the oil type of the spilled oil;
- generation of MS-PW-plots of PAH and biomarker compounds to assess the analytical variability of the method, potential effects of weathering, heterogeneity and mixing;
- calculation of a suite of diagnostic ratios based on an evaluation of the abundances of the diagnostic PAH and biomarkers.

6.3.2 Visual inspection and overall characterisation - Level 2.1

The extracted ion chromatograms of the biomarkers and the PAH in each candidate source oil should be visually inspected and compared to the spilled oil in order to identify any obvious non-matching source samples. Inspection of the spilled oils' extracted ion profiles will also reveal any particularly diagnostic features that could be useful in its comparison to candidate sources and, if appropriate, development of new diagnostic ratios to include in Level 2.2 (see informative Annexes D, E and H).

For any of the candidate source samples that cannot be obviously ruled out as possible sources following the qualitative visual inspection of extracted ion chromatograms, the biomarkers and PAH peaks necessary to establish the MS-PW-plots and to calculate the appropriate diagnostic ratios should be integrated. Integrations should be performed in consistent manner among all samples analysed. Annex D and Annex E give information about retention times and integration ranges for the suite of target compounds.

6.3.3 Treatment of the GC-MS results – Level 2.2

The results obtained from Level 2.1 should be used for:

- assessing the analytical variability (as based upon replicate analyses), integration of compounds and the impact of weathering on the samples by means of MS-PW-plots;
- determining which, if any, samples should be re-analysed because of apparent heterogeneity;
- deciding which diagnostic ratios are appropriate for numerical comparison based on the oil type, any particular features of the spilled oil, the analytical variability and the degree of weathering of spill and sometimes source samples;
- conducting a numerical comparison of appropriate diagnostic ratios in the spill samples and each source sample;
- a final visual comparison of extracted ion chromatograms to check the results of the numerical comparison by means of PW-plots and diagnostic ratios (ground-truth all data).

6.4 Treatment of the results using the MS-PW-plot– Level 2.2

6.4.1 General

The MS-PW-plot is used to evaluate results and to visualise the analytical data. In 6.2.2.2, the GC-PW-plot (based on alkanes and isoprenoids) has been introduced as a tool to evaluate weathering effects on spill samples. The PW-plot, however, is not only proof that differences in the lower boiling part have, or have not, been caused by weathering. The plot also gives information about the analytical error and the quality of peak

integration, and it shows the similarity of oil samples [28, 29]. The MS-PW-plot may even be used to confirm that observed differences have been caused by mixing.

For the MS-PW-plots shown and discussed here, more compounds and also groups of compounds have been integrated than are needed to calculate the normative ratios of Table 3. This has been done in order to have data points over the whole retention time range. Four different compound groups are used and indicated with different markers.

6.4.2 PW-plot calculations

A data point ($\% C_{Nspill}$) in the PW-plot represents the concentration of a compound or compound group in the spill sample (C_{Nspill}) relative to the concentration of the same compound or compound group in the source sample ($C_{Nsource}$). The data points are calculated from the peak height or peak area of a compound. To compensate for the total amount of oil injected, the result is normalised to a compound present in both samples. The calculation is expressed in Formula 1:

$$\% C_{Nspill} = \frac{C_{Nspill} / C_{HOPANE\ spill} \times C_{HOPANE\ source} \times 100\%}{C_{Nsource}} = \frac{C_{Nspill} / C_{HOPANE\ spill}}{C_{Nsource} / C_{HOPANE\ source}} \times 100\% \quad (1)$$

When comparing two samples by means of a PW-plot any compound may be used as a normalisation compound, but in case of an oil spill it is preferable to use a compound that is robust to weathering. For this guideline, the normative ratio compound hopane (C_{HOPANE}) has been chosen as normalisation compound and the normative ratio compound tetramethylphenantrene (T-M-phe, Figure E.2) is used as an alternative for hopane in case of a low concentration or absence of hopane.

6.4.3 Evaluation of the variability of the analysis and peak integration

In case of a duplicate analysis of an oil extract, all concentrations are the same and theoretically all $\% C_{Nspill}$ of a PW-plot should be 100 %. In practice, the instrumentation and integration of the compounds causes some variation resulting in a range of data points around 100 % as shown in Figure 3. This MS-PW-plot is the result of a duplicate analysis of sample 2 of the case discussed in Annex I.

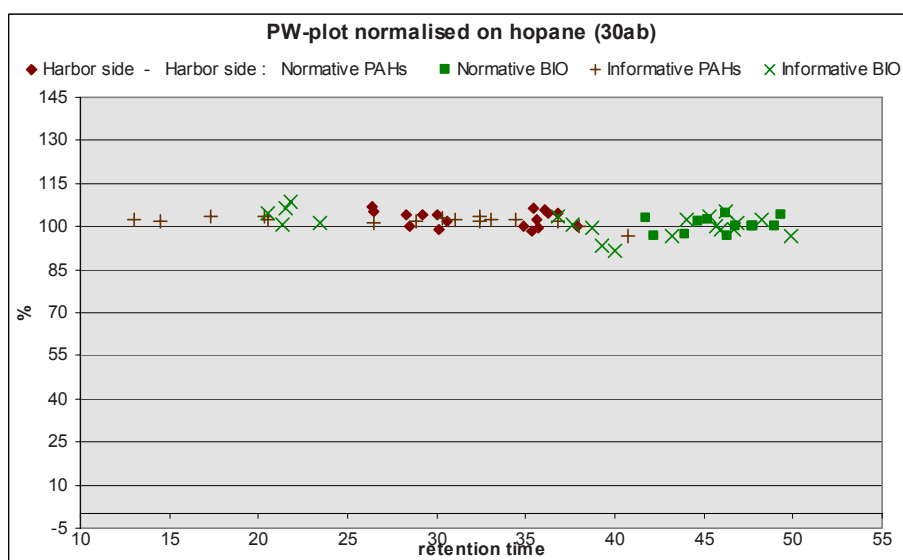


Figure 3 — MS-PW-plot of duplicate injections of an extract of a surface water sample

To assess the error range of the MS-PW-plot, two samples containing the same extract of a bilge sample were tested in the intercalibration round of 2009 [29]. The standard deviation was calculated from the data points of the resulting PW-plots. To eliminate outliers, the data points were sorted by participants starting with the results showing the lowest variance (see Figure 4).

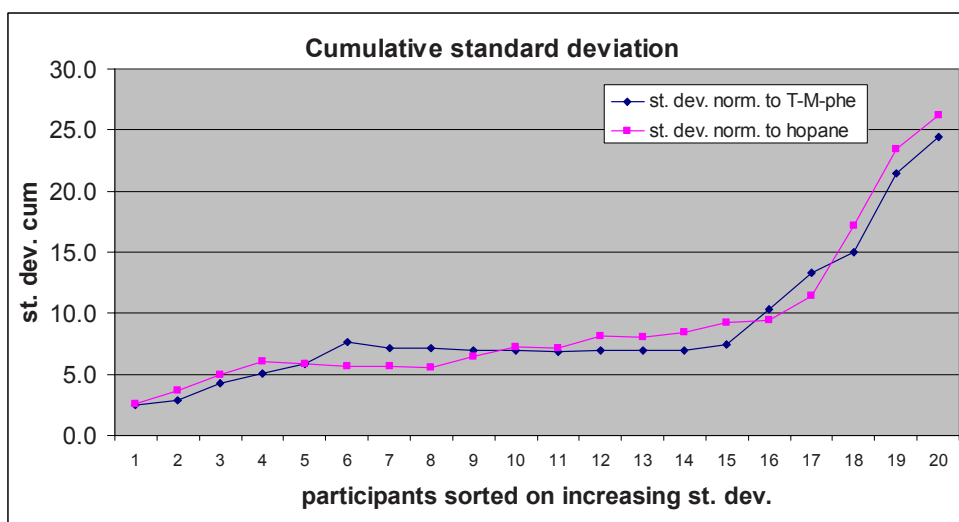


Figure 4 — Cumulative standard deviation of the MS-PW-plots of the intercalibration round of 2009 normalised to T-M-phe and hopane, respectively

The standard deviation shown for “participant 1” is the standard deviation of the data points of the PW-plot of participant 1.

The standard deviation shown for “participant 2” is calculated from the combined results of participant 1 and participant 2.

The standard deviation shown for “participant 3” is calculated from the combined results of participants 1, 2 and 3.

The standard deviation shown for participant “20” is calculated from the combined results of all participants.

Use of the flat part of the graph as an acceptable standard deviation of the MS-PW-plot results in a value of 7 to 8. So 95 % of the data points should be in the range of $2 * \text{standard deviation} = 15$ and the 95 % error range will be $100 \pm 2 * 7,5 = 100 \% \pm 15 \%$.

From a mathematical point of view however, the error range should be 85 % to 118 %.

Choosing one of the samples as 100 % and calculating the results of the other sample relative to it has an effect on the difference to the mean of 100 %. The mathematical effect is shown in Table 2.

Table 2 — MS-PW-plot data point calculation. Influence on the error range by choosing one of the samples as reference for the other

Sample 1 peak height	100	100	100	100	100	100	100
Sample 2 peak height	100	95	90	85	80	75	50
Sample 2/Sample 1 in %	100	95	90	85	80	75	50
Sample 1/Sample 2 in %	100	105	111	118	125	133	200

The assessment to estimate the variance of the MS-PW-plot results in a range of 85 % to 118 %. This range is assigned as a maximum acceptable range when comparing duplicate analyses of sample extracts. If the comparison leads to outliers, the integration and maybe even the method should be checked and improved.

In translating this range into a range to be used in real cases, it should be emphasised that the:

- range is calculated/estimated on the basis of duplicate analyses in which variations are caused exclusively by the method used. In reality, differences in sample treatment, sample composition and injection concentration will also increase the variation;
- calculations are based on the results of participants most of whom are not familiar with MS-PW-plots. The PW-plot itself informs about outliers based on mistakes in the integration. Therefore, it may be expected that variance will decrease as users are gaining experience with the plot.

In general, the 85 %-118 % range should be used as guidance for an acceptable range when comparing oil spill identification samples.

6.4.4 Evaluation of weathering

Lighter compounds of an oil spill evaporate faster than compounds having higher boiling points. As GC separation is based mainly on boiling points of the components, a PW-plot with the retention time on the x-axis provides information about the evaporation of compounds after the spillage.

Figure 5 shows a MS-PW-plot of an investigation comparing two bilge samples. A range of PAH and biomarker compounds have been integrated in both the source and spill sample and each compounds' abundance has been normalised to that of hopane (30ab).

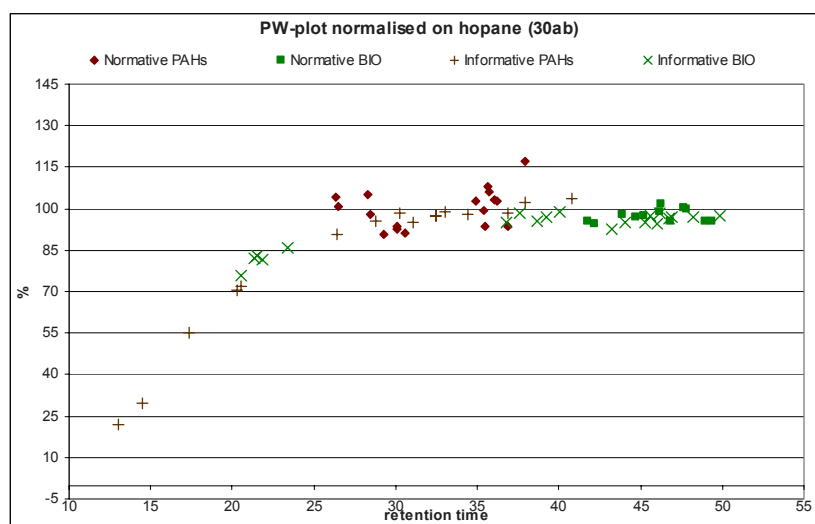


Figure 5 — MS-PW-plot showing the effect of evaporation

Figure 5 reveals that weathering has reduced the abundance of the earlier eluting PAH and biomarker compounds in the spill sample. This is a common and expected pattern for a PW-plot in which evaporation is the most important weathering effect. All other compounds and compounds groups in the spill oil are unweathered and remain present in abundances around 100 %, ranging from 91 % to 117 % relative to the source oil (i.e., there is no significant difference between the abundances of the analytes in the spill and source oils). Due to the variability of the analysis, a range of 85 % to 118 % is used as acceptable range. The data point of 117 % is related to benzo-naphto-thiophene (BNT) and the higher percentage might be caused by the very low sulphur content of the samples involved.

The MS-PW-plot in Figure 6 is more complex. It results from the comparison of a spill sample with a bilge sample. Although sampled from a bilge, the aromatic sterane pattern (m/z 231) indicates that the samples consist exclusively of gas oil. The ion chromatogram of m/z 231 shows the normal aromatic steranes pattern

(see Figure H.11). The hopane peak is close to noise level. Therefore, the peak height of T-M-phenanthrene has been used for normalisation.

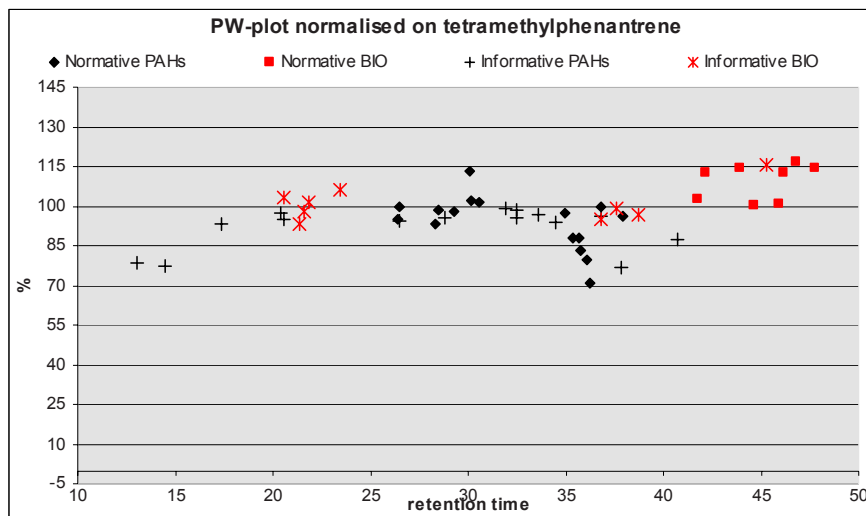


Figure 6 — MS-PW-plot of the comparison of a bilge sample with a surface water sample showing the effect of photo-oxidation

The PW-plot shows almost no reduction of compounds. Only the peaks before a retention time of 15 min are slightly reduced.

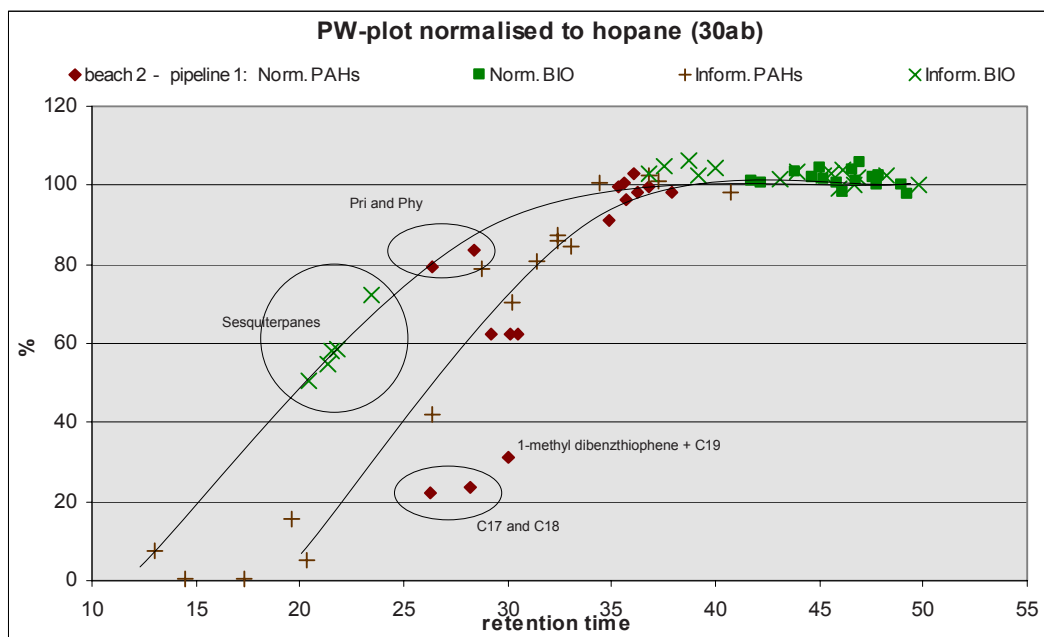
The normative biomarkers, eluting after a retention time of 40 min, are just above or at noise level; therefore only some of the higher peaks have been integrated. The PW-plot shows that most of the peaks are around 115 % even after a careful check of the integration. Table 3 indicates that the ratios of m/z 191, 217 and 218 are normative for a bilge sample, but that only a visual inspection is necessary for gas oil. As the analytical results had shown that the bilge sample consisted almost completely of gas oil, it was decided to compare the normative biomarkers only visually.

The informative biomarkers, shown in Figure 6 (sesquiterpanes at a retention time between 20 min and 25 min and tricyclic triterpanes C23Tr, C24Tr and C25Tr), are part of the distillation range of gas oil. Their peak heights are far above noise level and the data points are close to 100 %.

The normative PAHs vary between (100 ± 15) % except for the three methylpyrenes eluting at about 36 min 1-MPy is reduced more strongly than 4-MPy, which in turn is reduced more strongly than 2-MPy. This behaviour of the methylpyrenes is most likely caused by photo-oxidation (see section G.6).

The informative PAHs shown in Figure 6 after a retention time of 15 min are above 85 % except for the compound group C2-fluoranthenes pyrenes (m/z 230) eluting at a retention time of 38 min. This is also most likely related to photo-oxidation.

Figure 7 shows the effect of weathering on a layer of crude oil on water in laboratory test device for a month [53]. The resulting PW-plot shows the effect of a combination of evaporation, dissolution and biodegradation.



compounds	Retention time	% in spill	compounds	Retention time	% in spill
C1-dekalin	13.03	8	NR-2-methyl phenanthrene	30.08	62
Naphthalene	14.52	0	Methylantracene	30.24	71
C1-Naphthalenes	17.31	0	NR-1-methyl phenanthrene	30.52	62
C2 benzothiophenes	19.57	16	C2-dibenzothiophenes	31.43	81
C2-Naphthalenes	20.36	5	C2-highest peak phe. a...	32.42	87
SES1	20.46	50	C2 phenanthrenes anthr...	32.42	86
SES2	21.3	55	C3 dibenzothiophenes	33.03	85
SES3	21.51	58	C3 phenanthrenes anthr...	34.42	101
SES4	21.81	59	NR-2-Methylfluoranthene	34.88	91
SES8	23.4	72	NR-benzo(a)-fluorene	35.34	100
NR-C17	26.24	22	NR-benzo(b+c)fluorene	35.6	100
NR-Pristane	26.35	79	NR-2-methylpyrene	35.71	97
C1-Fluorenes	26.4	42	NR-4-methylpyrene	36.1	103
NR-C18	28.16	24	NR-1-methylpyrene	36.21	98
NR-phytane	28.35	83	C23 Tr	36.76	103
C2-Fluorenes	28.77	79	NR-tetramethyl-phenant...	36.79	100
NR-4-Methyl Dibenzothi...	29.23	63	C4 phenanthrenes anthr...	36.79	103
NR-1-Methyl Dibenzothi...	30.01	31			

The plot shows the combined weathering effects of evaporation, dissolution and biodegradation. Compounds with a retention time below 37 min are shown in the table of Figure 7. Compound names starting with NR- are normative compounds (see Table 3). Biomarkers are indicated in red

Figure 7 — MS-PW-plot normalised to hopane of crude oil on water for a month, compared with the source oil.

Two lines are drawn in Figure 7. The upper line combines compounds that have a higher resistance against dissolution and biodegradation, i.e. C1-decalins, sesquiterpanes, pristane and phytane. Their concentration reduction is primarily caused by evaporation.

The lower line combines compounds with a higher solubility in water (naphthalene and the C1- and C2-naphthalenes eluting between 18 min and 22 min have the highest solubility of the aromatic compounds integrated). Their concentration reduction is caused by a combination of evaporation and dissolution.

N-Alkanes are easily biodegraded and can be found below the lower line (N-alkanes C17 and C18). Strange is the low value for 1-methyldibenzthiophene (1-MDBT) eluting at a retention time of 30,01 min. It is far more reduced than 4-methyldibenzthiophene eluting at a retention time of 29,23 min. The reason is that 1-MDBT co-elutes with C19 (C19 generates m/z 198 at 0,5 % relative to m/z 85) at the settings used (see B.3.1).

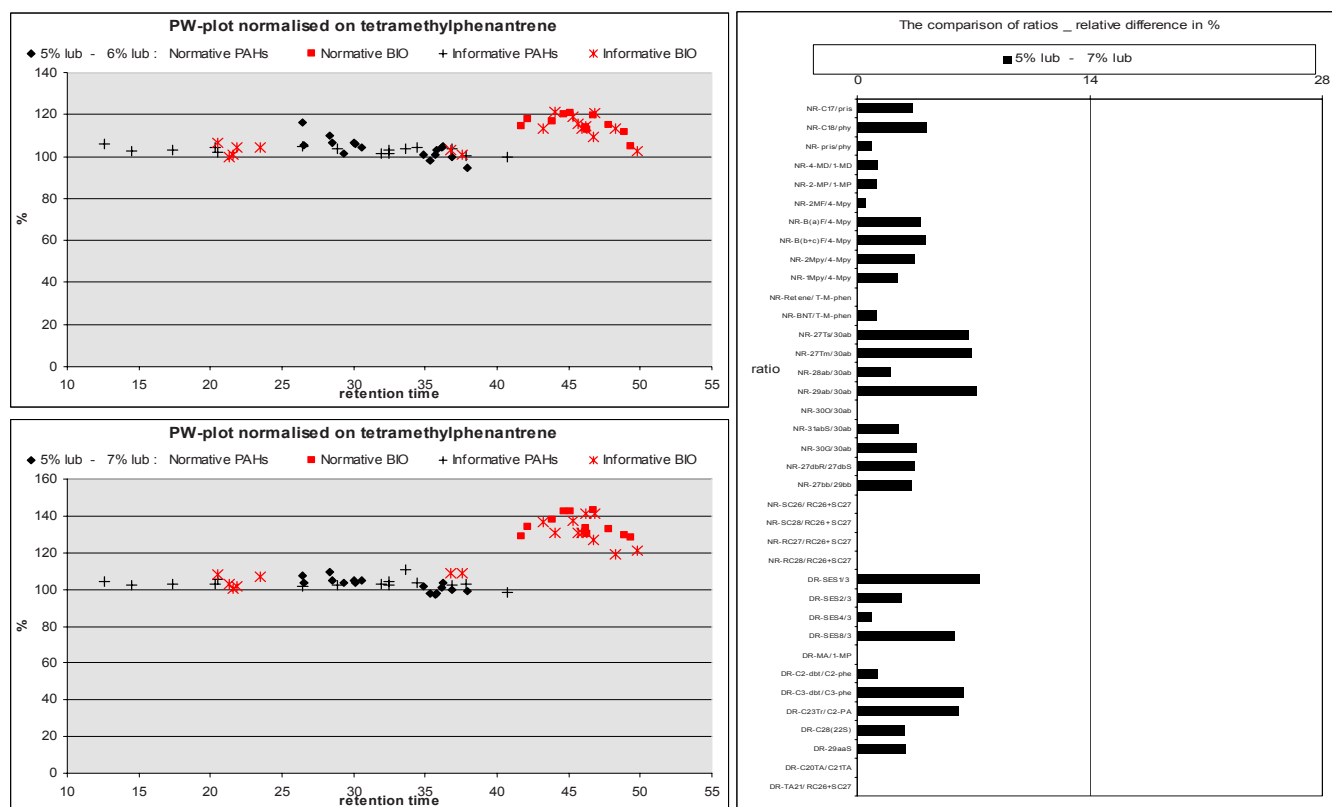
NOTE The co-elution of compounds is difficult to prevent in oil analyses. For a sufficient GC-separation on a 0,25 mm ID column, each compound needs a retention time window of about 3 s. Therefore, twenty compounds can be separated in one minute and 1 200 in one hour. The MS gives extra separation power, but most compounds generate many ion fragments in EI ionisation, so a factor of 10 is reasonable. As a result, about 12 000 compounds can be separated in an ideal situation. Crude oil is however estimated to contain about 250 000 compounds [48], so some co-elution can be expected in oil-spill identification.

6.4.5 Evaluation of mixing

Many oil spills involve mixtures of different oil types. Some examples are:

- fuel oils and lubricating oils that are spilled on board and collected in bilges, sludge and slop tanks;
- heavy fuel oil, which is a mixture of large molecules and particles dissolved/suspended in a solvent like light cycle oil (LCO, also called cutter-stock). Concentrations of the different components may differ even between tanks filled at the same bunker station.

The PW-plot helps to identify concentration differences of oil components. Figure 8 shows two PW-plots as a result of the mixing of 100 g gas oil with 5, 6 and 7 g lubricating oil [29].



Upper left plot: 5 % versus 6 %; lower left plot: 5 % versus 7 %. To the right: the ratio comparison graph between gas oil with 5 % and 7 % lubricating oil added.

Figure 8 — MS-PW-plots of the comparison of three mixtures of gas oil with 5 %, 6 % and 7 % lubricating oil, respectively.

Gas oil contains small amounts of biomarkers eluting after a retention time of 40 min. Lubricating oil contains a higher amount of biomarkers and a relatively small amount of PAHs.

In the upper left graph of Figure 8, the 5 % mixture is compared with the 6 % mixture. The plot is normalised to the PAH compound tetramethylphenantrene and is based on the 5 % mixture (the peak heights/areas of the compounds in the 5 % mixture are used as reference = 100 %). Because the gas oil concentration of the mixture is almost the same (95 % and 94 % respectively) the left PW-plot shows the compounds eluting close to 100 % within 40 min. The compounds eluting after 40 min show the difference in lubricating oil concentration. As the 5 % mixture is used as 100 % reference, the biomarkers of the 6 % mixture should be at about 120 %. This is very visible in the upper left plot.

For the lower left plot, in which the 5 % mixture is compared with the 7 % mixture, the biomarkers of the 7 % mixture should be at about 140 %.

NOTE 120 % and 140 % are estimates. The exact values depend on the actual concentrations of tetramethylphenantrene and of the other compounds used to generate the plot, in the gas oil and lubricating oil analysed. Exact values can be calculated when the samples are analysed with a full quantitative method, but this is beyond the scope of the method.

The difference in addition of lubricating oil is not indicated by the ratio comparison method (see 6.5) shown in Figure 8. The highest difference for the normative ratios of the 5 % to 7 % comparison is far below the critical difference of 14 %. It is important to note that the normative ratios have been selected to show differences within compound groups like the hopanes, steranes etc., not between compound groups like biomarkers and PAH's.

According to

- the match definition (see 3.3.1),
- the MS-PW-plot shape and variance,
- the likeliness of small concentration differences between gas oil and lubricating oil in bilge samples,
- the results of the diagnostic ratio comparison,
- the visual inspection of the ion chromatograms,

the 5 % to 7 % comparison results in a match.

It is still important to be aware of concentration differences of, e.g., PAHs versus biomarkers and to mention it in the final report. In an oil spill case, it is important to be aware of the fact that samples can originate from an inhomogeneous source.

6.5 Treatment of the results using ratios – Level 2.2

6.5.1 General

If two oil samples are identical, their chemical composition is by definition the same, apart from those changes experienced by the spilled oil as a result of weathering or contamination (mixing). Thus, measured ratios between any pair of compounds unaffected by weathering or contamination should ideally be the same (up to a certain statistical analysis related confidence level) in identical samples. Various ratios between individual compounds or groups of compounds have been described extensively and used to compare oil samples [46]. In many cases, such ratios may be referred to as "diagnostic" if they possess the quality to genuinely discriminate between different oils. This is often the case with ratios between petroleum biomarkers, which typically provide a high degree of specificity among different crude oils [37]. Other ratios such as those based on PAH can often be diagnostic in crude oils, but may or may not have a similar diagnostic significance in refined products in which their distributions can be affected by refinery processes (e.g. cracking, reforming, and hydro-treatment), see H.4. However, genuinely diagnostic or not, any ratio not influenced by weathering or contamination should still show the same value in any identical samples and thus be of "diagnostic" value in a specific spill case and contribute to discriminating and ruling out "non-source" oils.

Annex C includes a list of target compounds or compound groups and Annex E includes a suite of potentially diagnostic ratios suitable for most oil spill investigations. It is important to realise, however, that this list of PAH and biomarker analytes and suite of diagnostic ratios are neither all inclusive nor appropriate for all oil spill identification cases. As discussed in Annex H, the compounds and ratios of most use depend on the type of oil spilled. In some spill cases, it may be prudent to include a certain characteristic feature of the spilled oil that is recognised as being particularly diagnostic. In other situations, the abundance of some compounds necessary to determine the suggested ratios are below analytical measurable limits or have been influenced by weathering or contamination. Thus, maintaining flexibility in the selection of the target compounds and ratios to be applied in any given oil spill identification investigation is very important. Such flexibility can be maintained by acquiring raw data for a broad range of target compounds representing various petroleum hydrocarbon groups, such as those listed in Annex C and used in Table B.2. From this suite of target compounds, potentially diagnostic ratios unaffected by weathering or contamination should be calculated and compared as described below.

The diagnostic ratios and the compounds, which are used for producing the MS-PW-plots, depend on each other: in both cases the same target analytes are used. MS-PW-plots turned out to be powerful tools for the recognition of the different kinds of weathering effects. When the effects of weathering on a spilled oil are understood, any affected DRs can be eliminated or cautiously considered in when comparing a weathered spill sample to an unweathered source sample.

6.5.2 Diagnostic ratios calculation

Diagnostic ratios (DRs, see definition in 3.2.) should preferably be based on peak heights for individual peaks and on peak areas for compound groups such as certain alkylated PAH. Peak areas of individual peaks can also be used as long as appropriate resolution and integrations are carefully and consistently performed. For example, when integrating individual peaks, co-eluting peaks should be avoided or carefully and consistently integrated. For example, if peaks of 29Ts and 29ab co-elute, it may be preferable to use either the peak height of the combined peaks or the total area of the combined peaks in calculating a combined ratio of (29Ts+29ab)/30ab.

The diagnostic ratio (DR) between two distinctly peaks (A and B), should be based on the ratio formula:

$$DR = A/B; \quad (2) \text{ or}$$

$$DR = 100 \times A/B \text{ (in \% , depending on the preference of the user).} \quad (3)$$

The result of the ratio formula A/B has a fixed RSD (relative standard deviation).

NOTE The ratio formula $DR = \frac{A}{(A+B)}$ is also often used to calculate diagnostic ratios as it has the benefit of normalising ratios between 0 and 1. However, this ratio format yields a ratio-dependent RSD [21]. Because of this, DRs based upon $A/(A+B)$ cannot be utilised for the statistical comparison method described in the TR, which relies upon a fixed RSD (section 6.4.5). The revised Nordtest protocol [12] relies upon an $A/(A+B)$ formula format and employs a different statistical metric for comparing DRs among oil samples.

6.5.3 Normative diagnostic ratios

To identify any potential source samples as “non-matches” with the spill sample(s), a distinct chromatographic pattern or just one distinct DR (unaffected by weathering or mixing) can be sufficient. However, in order to prove a “positive match” all chromatographic patterns (unaffected by weathering or mixing) should be the same and all DRs (unaffected by weathering or mixing) should statistically match. Some minimum set of DRs is needed to represent the different “aspects” of the particular type of oil spilled to make the comparison test meaningful. Toward this end, normative sets of diagnostic ratios for different oil types have been selected from the suite of ratios given in Annex E.

The selection of the normative diagnostic ratios for a particular oil type is based on the following criteria for the target compounds involved:

- commonly present at a higher concentration in the oil type involved;
- well separated single peaks or known co-eluting peaks (e.g. RC26TA+SC27TA);
- based on comparable boiling compounds within the same m/z extracted ion chromatograms;
- robust against weathering or with known weathering characteristics.

When a normative DR is excluded from the statistical ratio comparison the reason (e.g. DR-30O/30ab: in many crude oils 30-O (oleanane) is not present) should be noted in the internal documentation (7.2).

The normative selection of diagnostic ratios in Table 3 should be treated as a minimum list, which can be extended with the appropriate ratios described in Annex E or from the peer-reviewed literature. In addition, if the spill oil exhibits any particular features not represented by existing ratios it is possible to introduce new ratio(s) that reflect those features.

The ratio DR-T-M-phe/ BNT provides information about the sulfur content of the spill [32]. Many labs routinely determine the concentrations of dibenzthiophenes for this purpose and calculate the ratios DR-C3-dbt/C3-phe and DR-C2-dbt/C2-phe [46]. Despite their widespread use, these compound groups have not been included in the normative ratios in Table 3, because they are not based on single peak integrations, but rather on area

integrations including many isomers. However, when the laboratory can demonstrate validation of the method for these compound groups and the results of duplicate measurements show that these ratios can be calculated with a standard deviation < 5 %, DR-T-M-phe/ BNT can be replaced by the ratios DR-C3-dbt/C3-phe and DR-C2-dbt/C2-phe (see Annex E).

Table 3 — Normative diagnostic ratios

Ratio	Light Fuel	Lub. oil	Bilge	HFO	Sludge	Crude oil
GC						
DR-GC-C17/Pr	X		X		X	X
DR-GC-C18/Ph	X		X		X	X
DR-GC-Pr/Ph	X		X		X	X
MS						
m/z 85 DR-C17/Pr	X		X	X	X	X
DR-C18/Ph	X		X	X	X	X
DR- Pr/Ph	X		X	X	X	X
m/z 192 DR-2-MP/1-MP	X		X	X	X	X
m/z 198 DR-4-MDBT/1-MDBT	X		X	X	X	X
m/z 216 DR-2MFL/4-MPy	X		X	X	X	X
DR-B(a)F/4-MPy	X		X	X	X	X
DR-B(b+c)F/4-MPy	X		X	X	X	X
DR-2MPy/4-MPy	X		X	X	X	X
DR-1MPy/4-MPy	X		X	X	X	X
m/z 234 DR-Retene/ T-M-phe	X		X		X	X
DR-BNT/ T-M-phe	X		X	X	X	X
m/z 191 DR-27Ts/30ab		X	X	X	X	X
DR-27Tm/30ab		X	X	X	X	X
DR-28ab/30ab		X	X	X	X	X
DR-29ab/30ab		X	X	X	X	X
DR-30O/30ab		X	X	X	X	X
DR-31abS/30ab		X	X	X	X	X
DR-30G/30ab		X	X	X	X	X
m/z 217 DR-27dbR/27dbS		X	X	X	X	X
m/z 218 DR-27bb/29bb		X	X	X	X	X
m/z 231 DR-SC26/ RC26+SC27				X	X	X
DR-SC28/RC26 + SC27				X	X	X
DR-RC27/RC26+ SC27				X	X	X
DR-RC28/RC26+SC27				X	X	X

Light Fuel represents distillation fractions of crude oils of which all hydrocarbons elute < C30 to C31. It includes distilled intermediates and fuels, such as gas oil and marine diesel. Typically, the higher boiling biomarkers are not present in light fuel. However, it is quite common for refineries to add light cycle oil (LCO,

also called cutter-stock) to the distilled light oil fraction. LCO is produced in a cat cracker and contains mainly aromatic compounds and a low concentration of biomarkers. As a result of this mixing, the chromatographic pattern and concentration of the C1-fluoranthrenes/pyrenes can be very different, while the alkane pattern is very similar. Because the concentration level of the higher boiling biomarkers is still low in such mixtures, it is advised not to calculate ratios based upon these compounds, but to visually inspect and compare this retention time area. The lower boiling biomarker compound C23 Tricyclic triterpane (C23Tr) elutes just behind n-alkane C23 and is often present in light fuel oil at a higher level. The ratio DR-C23Tr/C2-Phe-hp (see E.1) can be used to provide information about the biomarker abundance in light fuel oil.

Lubricating oil is used to grease engines and should have a high flame point. Therefore, the paraffinic fraction is removed and the aromatic compounds are hydrogenated. As a result the alkanes, PAH and triaromatic steranes (m/z 231) are at a low abundance or are absent in the chromatograms of unused lubricating oil. These features can be different for used lubricating oil, which can be modified during use or mixed with other petroleum products (e.g., in bilge). Therefore, for spills involving lubricating oil it is advised to check the chromatograms for these features and to extend the number of DRs as may be appropriate.

The mineral oil part of the composition of hydraulic oil is very similar to that of lubrication oil and cannot really be distinguished from lubricating oil in spill samples. Both lubricating and hydraulic oil can be a mixture of mineral and synthetic oil. CEN/TR 15522-2 compares the mineral oil parts.

Bilge, in the context of this TR, is related to ships with a large open space of several cubic metres below older diesel engines which use large amounts of lubricating oil. The mixture of mainly light fuel oil and lubricating oil is collected in the bilge below the engine(s). The concentration of the biomarkers depends on the composition of the mixture. Typically, the concentration of the triaromatic steranes is low.

NOTE The word "bilge" is a nautical term which is used to refer to a several concepts. It dates from 1513, when "bilge" was first used to refer to the lowest compartment inside the hull of a ship, where the two sides meet at the bottom. However, the term is also used to discuss the matter which collects in the bilge: water drains from the decks of the ship into the bilge, dragging detritus from topside down below. The water that collects in the bilge can be brackish and also have a foul odor, so bilge has also come to be a term which refers to anything foul or unpleasant.

Heavy fuel oil is a residual product of a refinery. It may have gone through a cat cracker and as a result the aromatic patterns can have been changed (sometimes becoming more or less standardised). As a result, the methylphenanthrene pattern has changed, the concentration of methylanthracene has increased significantly and the concentration of retene reduced strongly. The normative ratio DR-2-MP/1-MP is useful in comparing the methylphenanthrene patterns, but the DR-Retene/ T-M-phe is omitted.

Sludge is in general an HFO or a mixture of HFO and lubricating oil. Depending on the relative proportions of HFO and lubricating oil, the character of sludge can vary widely. Retene will probably not be present, but it is possible that other petroleum products like marine diesel are present, so the ratio DR-Retene/ T-M-phe should be looked at.

Crude oil can vary very widely and mostly contains all normative ratios, although the biomarker compounds 28,30-bisnorhopane, oleanane and/or gammacerane can be absent.

Bilge versus Sludge

Smaller ships do not use HFO, thus do not produce sludge, which consists mainly of residues from HFO-precipitates produced in fuel tanks on board. Thus, with regard to the illegal discharges, the most commonly encountered oily waste is bilge discharged by smaller ships. On the other hand, it does not make much sense for bigger ships, which use HFO, to discharge oil, which is collected in the ship's bilge because here the amount is much smaller than the amount of waste oil, which results from HFO-separation. Those ships have a special sludge tank (must have by law, with special dimensions) and the relatively small amount of waste oil in the ships bilge can easily be transferred into the sludge tank and thus added to the sludge. Consequently, it can be confidently concluded that most spills consisting of diesel/lubricating oil mixtures originate from a bilge.

For more information about petroleum and the different petroleum products relevant for the guideline, see Annexes F and H.

6.5.4 Analytical error

In order to test whether a difference in two measured values is significant or not, the analytical error of the method used to measure the values is needed [24, 25, 26]:

- Because of the complex and variable composition of oil samples, this error is not only different from DR to DR, but can change from case to case. Performing a validation test with a reference oil sample, a good system and careful integration can easily lead to RSDs of 3 % or lower for some of the diagnostic ratios. However, as mentioned above, the composition of spill and source samples in real cases can be very different from the composition of the reference oil. For example, the concentration of single compounds may be much smaller or target peaks may be more or less overlapped by others, leading to the fact that diagnostic ratios in authentic spill samples can only be determined with less precision than in a reference oil sample.
- To determine the standard deviation of the individual DRs in an oil case, at least eight measurements for some of the samples are needed to be able to calculate statistically sound results. As a result, it is very time consuming to determine the precision of the method for all ratios in every single case.
- Working with actual error values will result in different match criteria for each oil case. Even reanalyzing an oil case by another analyst in the same laboratory can lead to different match criteria for the same oil case.
- Laboratories that like to speed up the analytical part for commercial reasons, might also accept higher analytical errors, what increases the risk of false match conclusions.

As a result, a default value of 5 % relative standard deviation (RSD) for diagnostic ratios is adopted in this TR, representing the upper limit of the acceptable performance of the method (see also A.4.5 and B.4.7). The value of 5 % is based on the experience of several laboratories, on criteria used in literature [44] and in the meanwhile used in several inter-calibration tests with good results [28, 29, 53]

Laboratories should therefore ensure that an RSD of 5 % for all relevant normative ratios of the duplicate analyses within an oil case are not exceeded. If diagnostic ratio values exceeding the analytical error are observed here, any oil identification analyses should be stopped, and the reasons for those differences (e.g. integration, a leakage in the system or a dirty ion source) should be found and corrected before resuming any oil spill sample analyses⁴⁾.

6.5.5 Match-criterion for ratios

Crucial for oil sample comparison is the consistent analytical performance of the GC-FID and GC-MS instruments during the analysis of one batch of related samples. For samples analysed in one batch, any differences in their compositions should reflect real differences between the samples and should not be caused by changes due to instrument performance e.g. the sensitivity of the detector by contamination, leaking and a reducing sensitivity for higher masses.

When control charts are produced by means of recurrent analysis of a standard reference oil, the difference between any two ratios of two consecutive samples should only in one of twenty cases exceed the repeatability limit $r_{95\%}$. [25]:

$$r_{95\%} = 2,8 \times \text{RSD (repeatability)} \quad (4)$$

Using an RSD of 5 % as given above leads to a repeatability limit of 14 %.

The critical difference (CD) (see 3.2.3) is calculated by using the following formula:

4) The use of e.g. a fast GC program to decrease the analysis time in combination with a high number of masses in one SIM section, reduces the number of scans over a peak and significantly increases the variance of the method (refers to B.3).

$$CD = ((\text{mean}_x \times r_{95\%}) / 100) \quad (5)$$

In close analogy to this principle, ratios measured in samples from real cases originating from the same source and not influenced by weathering and/or mixing should show the same repeatability, when they are analysed within a single batch run.

Consequently, any measured difference in a diagnostic ratio between two samples (i.e. the absolute difference in percent based on the mean of the two ratios) higher than 14 % should be regarded as significant, meaning that these two ratios are statistically different. This same comparison has to be evaluated for each diagnostic ratio being considered in any particular oil identification investigation. An example of two ratio comparisons – one showing no statistical difference and one showing a statistical difference - is given in Table 4.

Table 4 — Example of a ratio comparison based on the repeatability limit

Ratios to be compared in %		Mean	Absolute difference	Relative difference	Conclusion
Ratio sample 1	Ratio sample 2				
48	52	50	$52 - 48 = 4$	$4/50 \times 100 = 8\%$	$8\% < 14\%$ => No difference
35	45	40	$45 - 35 = 10$	$10/40 \times 100 = 25\%$	$25\% > 14\%$ => Different

For a presentation of the relative differences in bar graph format for comparing a large number of DRs in two samples see Figure 8. In this figure, each ratio difference falls to the left of the 14 % line indicating that there is no statistically significant difference between the two samples being compared.

NOTE The standard deviation of the combined variance of the DR's used for a comparison of two samples can be calculated from the differences between the individual ratios, by calculating the standard deviation from paired data [9, section 3.7.1]. For example, for the ratios of the ratio comparison graph of Figure 8 the standard deviation of the paired ratios is 2.9. It gives a good indication of the standard deviation of the ratios for the samples involved.

6.5.6 Criteria for selecting, elimination and evaluating diagnostic ratios

6.5.6.1 General

Generally, only those diagnostic ratios that can be measured with a high precision should be evaluated by comparing candidate sources to the spilled oil [43]. Small peaks at or near the baseline level have an increased variance and should only be used for a visual, qualitative comparison and not for the quantitative DR comparison. In order to evaluate the precision for an actual analytical series for the samples involved and the precision of the individual diagnostic ratios, the preparation and analysis of duplicate samples is essential (see 6.1).

Figure 9 shows a protocol/decision chart for DR evaluation.

For DRs determined to be unaffected by weathering, the chart shows two methods by which the DRs should be evaluated in order to assure their reliability for use in the DR comparison. Two bases exist by which DRs might be eliminated from the comparison, viz.:

- elimination by means of a signal to noise (S/N) test (see 6.5.6.2);
- elimination by means of the comparison of the duplicate analyses of extracts (see 6.5.6.3).

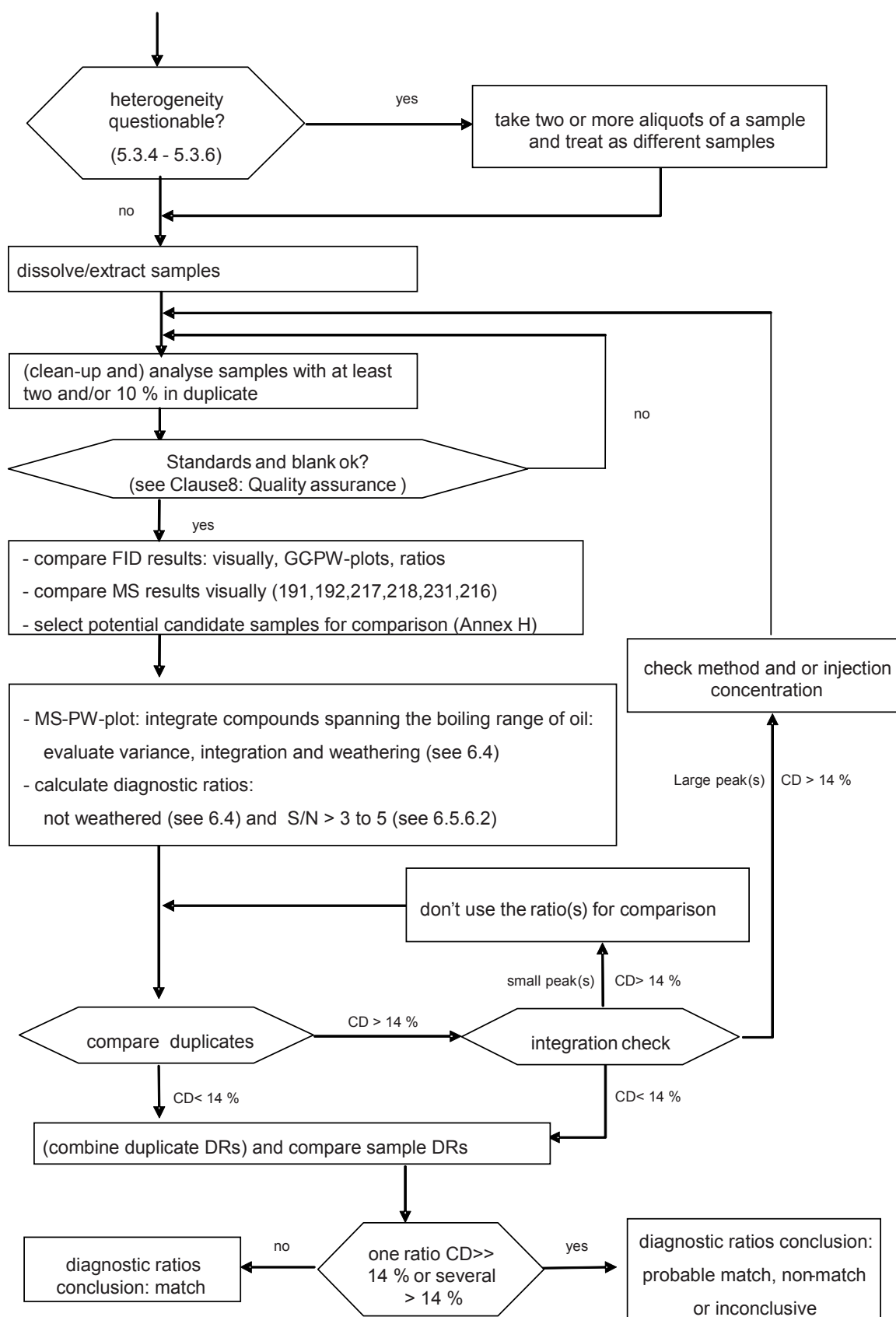


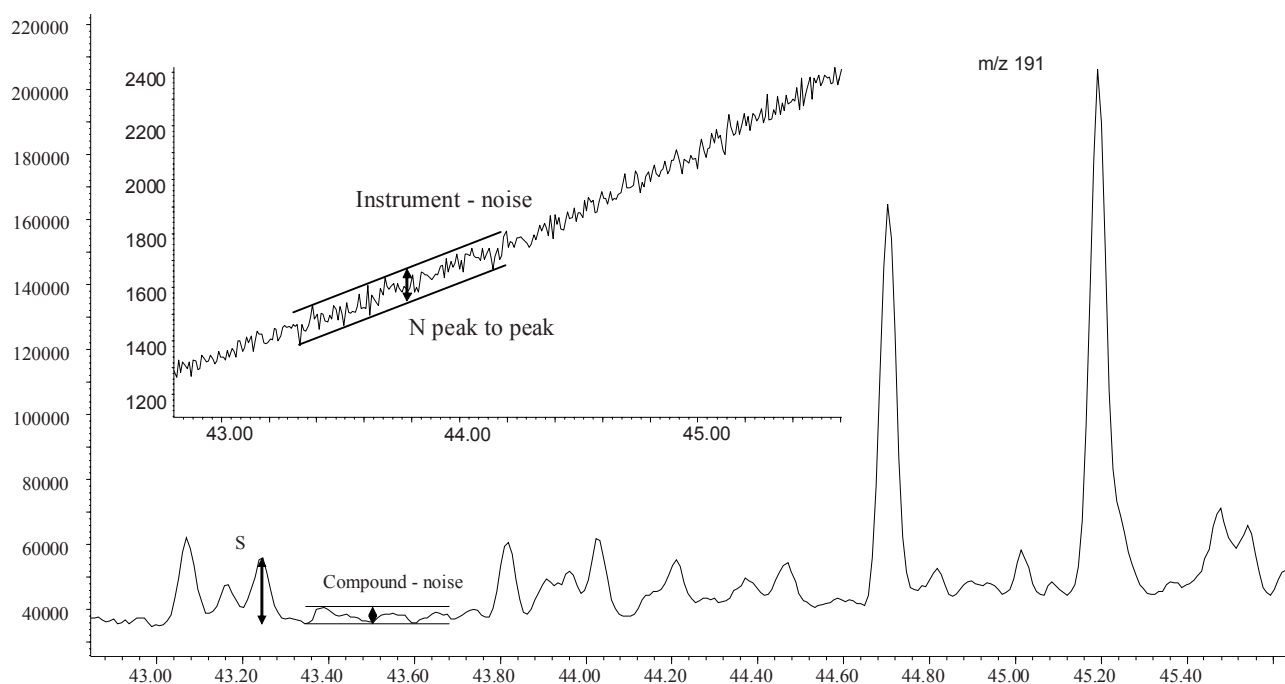
Figure 9 — Protocol/decision chart for diagnostic ratio evaluation

6.5.6.2 Elimination by means of a signal to noise (S/N) test

Only peaks with an $S/N > 3$ to 5 should be used for the DR comparison test. The S/N criterion is based on the method described by IUPAC [18, 27] and shown in Figure 10. N is the peak-to-peak value of a part of the noise around a target peak and S the target peak height. A decision range of 3 to 5 is given, because it is often difficult to estimate the noise, which is caused by the many peaks present in the chromatograms of oil samples.

As an example, a part of the m/z 191 ion-chromatogram (C28 (22R) to 27 Tm) of the SINTEF oil mixture is shown in Figure 10. Inset within the figure is a part of the ion chromatogram of an injection of clean DCM (Blank injection) over the same retention time section, demonstrating the noise calculation. Normally a part around the peak should have been taken to calculate the noise. Comparing, however, the mainly electronic noise from the blank injection with the "chemical noise" of the oil sample, shows that the baseline in the oil sample is an ion signal of compounds.

In this example, the noise of the blank $N_{bl} = 112$. The "compound noise" in the oil sample is $N_{sample} = 4975$ and the signal of C28 (22S) $S_{sample} = 20\,319$ resulting in $S/N_{sample} = 4,1$. Compared to the decision range of 3 to 5, both decisions are valid. Smaller peaks should not be integrated for use in diagnostic ratios, but can still be useful for qualitative visual inspection. Larger peaks should be included in the DR comparison test.



The distance between the highest and lowest peak of a part of the noise is used as N and the peak height as S . S/N is 4,1 for this example

Figure 10 — Signal to Noise (N) calculation

Even more difficult is the decision to integrate areas representing groups of compounds, rather than discrete peaks. For example, in propeller shaft lubricating oil many of the alkylated PAHs mentioned in Table C.1 are present at a low relative concentration and are represented by only the highest peaks of a group. In such cases, the pattern should only be used for a visual comparison and can be eliminated from any DRs reliant upon alkylated PAHs. An exception to this guidance has to be made for retene. Retene elutes within the C4-phenanthrene group of compounds and should be integrated when it is clearly higher than the peaks in the retention time range of 35 to 36 min (see Annex E, Figure E.2 where retene should be integrated). Besides m/z 234, retene also generates an ion at m/z 219, which can be useful to recognise the presence of retene.

In general, the $S/N > 3$ to 5 criterion should be used as a rule of thumb to assess and potentially eliminate DRs based upon integrations of small single peaks from the DR statistical comparison.

6.5.6.3 Elimination by means of the comparison of the duplicate analyses of extracts

A duplicate analysis of an extract should result in statistically identical DR values. Differences are usually caused by the method, the instrumentation and the data processing (peak integration). Because of this test, several options are possible:

- ratios of large peaks differing more than 14 %;

Check the integration of the peaks involved (retention time and baseline drawing), instrumental conditions and/or injection concentration. For example, a high injection concentration of compounds, not analysed by the SIM method, can still influence the properties of the peaks to be analysed. Eventually, perform a clean-up and reanalyse the samples.

- ratios of small peaks differing more than 14 %;

Check the integration of the peaks involved (retention time and baseline drawing). If one or two ratios are still above the CD, the peaks involved should only be used for a visual comparison and not for calculating ratios. If many ratios are still above the CD, act in the same way as for large peaks differing more than 14 %.

- all ratios differing less than 14 %.

Go to the next step of the chart (Figure 9) and compare the samples involved. Use the mean value of the diagnostic ratios of the samples analysed in duplicate for comparison to the other samples. The mean value has a lower variance and comes closer to the ratio of the sample.

NOTE Comparing the mean of a duplicate analysis makes the comparison more robust and reduces the value of the relative standard deviation (RSD) by $\sqrt{n}$. Therefore, the RSD of the mean value of a duplicate analysis becomes $RSD/\sqrt{2}$. For this TR however, it has been decided not to reduce the repeatability limit of 14 %, because the RSD of 5 % is a chosen value applied as criterion for the quality of the sequence and yields no difference between samples analysed in singular or in duplicate.

6.5.7 Optional: Evaluation of diagnostic ratios using conventional or multivariate statistics

On occasion, it is necessary to compare diagnostic ratios from a large number of samples, for example, in the case of a mystery spill compared to a large database of oils [19]. In this situation, a univariate comparison of each oil sample in the database to the mystery spill can be facilitated through the use of conventional or multivariate statistics. Of course, there is the need for some caution when comparing DRs obtained over extended periods of time, such as may exist in the case of long-standing database. Under such circumstances it may be prudent to compare DRs of laboratory reference oils analysed over time (e.g. early in the database's development versus recently in the database's development) in order to assure that the laboratory's performance over time has remained comparable, thereby allowing comparison of the "new" mystery spill data to the database.

Conventional statistical metrics such as various similarity indices or correlation tools (e.g. correlation coefficients) can be used to compare a spill's diagnostic ratios versus ratios in a large database as a means of identifying the database's oil that are most similar to the spill oil. These most similar oils can then be the focus of purely univariate comparison to the mystery spill according to the protocol of Figure 9.

In addition, multivariate statistical methods like the principal component analysis (PCA) can be used as a tool to compare diagnostic ratios between multiple samples [8, 10, 44] instead of the purely univariate approach described in section 6.4. One advantage of multivariate statistical methods compared to univariate statistical methods is the ease by which relationships between multiple samples and variables (diagnostic ratios) can be resolved and visualised by scores and loading plots. Additional advantages include noise reduction, obtained by multiple measurements of the same phenomenon (i.e. inter-related variables), and their ability to detect outliers.

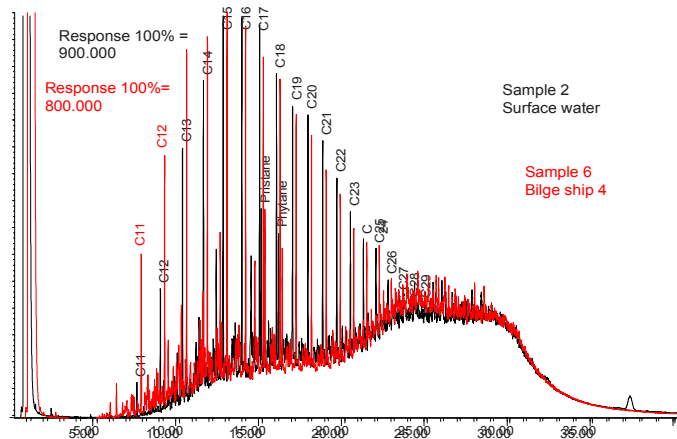
The fundamentals of PCA as a data treatment tool in chemical fingerprinting, is that it summarises the information in many correlated diagnostic ratios into a few so-called principal components that are weighted sums of these ratios. Hence, a model with few components (e.g. one or two) describes the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

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.

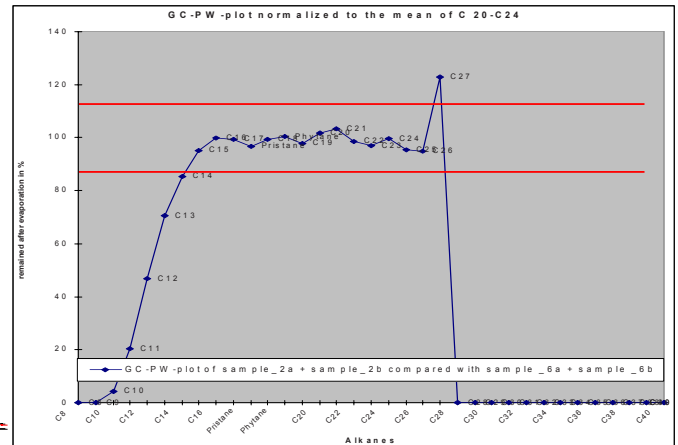
Conventional and multivariate statistical method can be used to rule out potential sources but it cannot replace the ratio comparison method and the visual inspection of the chromatograms to conclude a positive match. Small differences and compounds not used in the statistical treatment could not be recognised by the multivariate method.

6.6 Conclusions

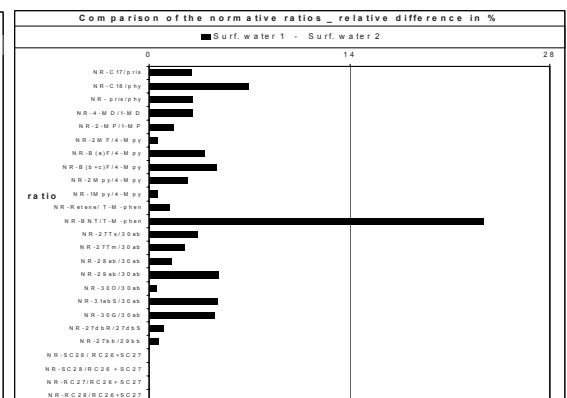
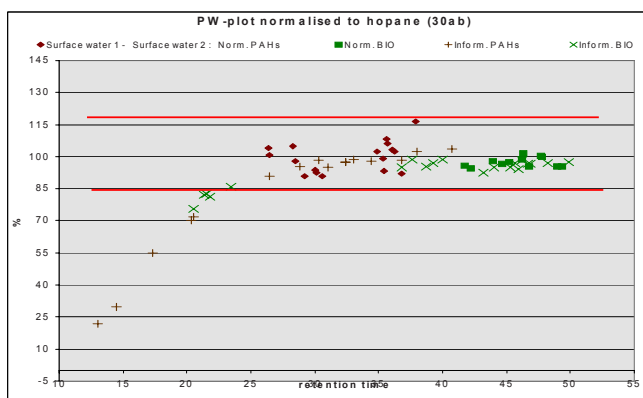
The final conclusion of an oil spill identification investigation should be based on a total evaluation of the analytical results. A summary of the means given in this TR for comparing two oil samples is shown in Figure 11.



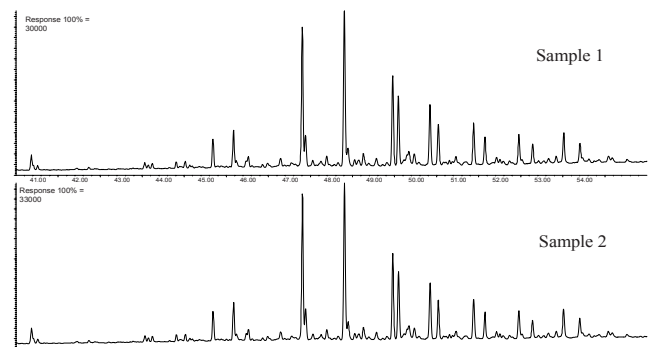
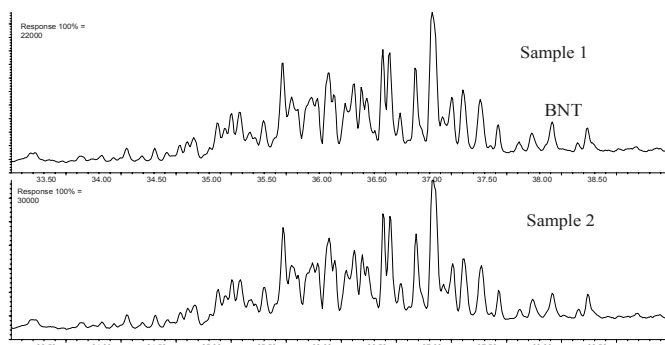
Comparison of the n-alkane distribution and the unresolved complex mixture by overlays



Measuring of n-alkanes and isoprenoids, calculation of n-alkane ratios and cross-plotting (GC-PW-plots)



Measuring of target compounds in relation to hopane and crossplotting (MS-PW-plot) and the comparison of diagnostic ratios.



Detailed comparison of complex complex clusters

In this example, much of the necessary information is jointly at hand, permitting a final conclusion to be defensibly drawn.

Figure 11 — Example of combined analytical results comparing a spill and source oil

In accordance with the decision charts in Figures 1 and 9, the results of a comparison between a spilled oil and a prospective source sample using this methodology should be concluded using one of the four terms: positive match, probable match, inconclusive or non-match (see 3.3).