Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1090 (2005) 133–145
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Multivariate statistical methods for evaluating biodegradation of mineral oil

Jan H. Christensen [a,*], Asger B. Hansen [b], Ulrich Karlson [b], John Mortensen [c], Ole Andersen [c]

[a] *Department of Natural Sciences, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark*
[b] *Department of Environmental Chemistry and Microbiology, National Environmental Research Institute, Frederiksborgvej 399, 4000 Roskilde, Denmark*
[c] *Department of Life Sciences and Chemistry, Roskilde University, Universitesvej 1, 4000 Roskilde, Denmark*

Received 28 February 2005; received in revised form 24 June 2005; accepted 6 July 2005
Available online 1 August 2005

Abstract

Two methods were developed for evaluating natural attenuation and bioremediation of mineral oil after environmental spills and during in vitro experiments. Gas chromatography–mass spectrometry (GC–MS) in selected ion monitoring (SIM) mode was used to obtain compound-specific data. The chromatographic data were then preprocessed either by calculating the first derivative, retention time alignment and normalization or by peak identification, quantification and calculation of diagnostic ratios within homologue series of polycyclic aromatic compounds (PACs). Finally, principal component analysis (PCA) was applied to the preprocessed chromatograms or diagnostic ratios to study the fate of the oil. The methods were applied to data from an in vitro biodegradation experiment with a North Sea crude oil exposed to three mixtures of bacterial strains: R (alkane degraders and surfactant producers), U (PAC degraders) and M (mixture of R- and U-strains) over a 1-year-period with five sampling times. Assessment of variation in degradability within isomer groups of methylfluorenes (*m*/*z* 180), methylphenanthrenes (*m*/*z* 192) and methyldibenzothiophenes (*m*/*z* 198) was used to evaluate the effects of microbial degradation on the composition of the oil. The two evaluation methods gave comparable results. In the objective pattern matching approach, principal component 1 (PC1) described the general changes in the isomer abundances, whereas M samples were separated from U and R samples along PC2. Furthermore, in the diagnostic ratio approach, a third component (PC3) could be extracted; although minor, it separated R samples from U and M samples. These results demonstrated that the two methods were able to differentiate between the effects due to the different bacterial activities, and that bacterial strain mixtures affected the PAC isomer patterns in different ways in accordance with their different metabolic capabilities.

Keywords: Biodegradation; Oil spills; PCA; COW; GC–MS; Chemical fingerprinting

1. Introduction

Crude oil and refined petroleum products released into the environment are subject to a number of weathering processes that change the oil composition, including physical (e.g. evaporation, emulsification, natural dispersion, dissolution and sorption), chemical (photodegradation) and biological (mainly microbial degradation) processes [1]. The physical processes result in redistribution of oil components in compartments of the environment, while photodegradation and microbial degradation lead to transformation of the original compounds.

Numerous authors have investigated microbial degradation of mixtures of petroleum hydrocarbons, in situ [2–5] and under laboratory conditions [6–9]. Evaluation of the role of biodegradation in environmental oil samples is, however, often difficult due to the inherent complexity of the chemical hydrocarbon mixtures, and due to the variety of weathering processes affecting oil composition [10]. Consequently, most bioremediation studies have been limited to few 'model' compounds, and to a selection of bacterial strains capable of utilizing petroleum hydrocarbons for growth [6,11–14]. Although, hundreds of such investigations have been per-

* Corresponding author. Tel.: +45 35282456; fax: +45 35282398.
E-mail address: jch@kvl.dk (J.H. Christensen).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.07.025

formed since the 1980s, the development of evaluation techniques has been limited. Gas chromatography-flame ionization detection (GC-FID) and Gas chromatography–mass spectrometry (GC–MS) are the standard methods for the analysis of petroleum hydrocarbons [7,9,15].

Bulk properties such as total petroleum hydrocarbon (TPH) concentration, measured by GC-FID, and gravimetric analysis of the aliphatic, aromatic and polar fractions of weathered oil samples have been used frequently to evaluate the effects of oil weathering [2,16,17]. GC–MS analysis can, on the other hand, resolve a broad range of petroleum hydrocarbons including individual PACs within homologue series. Thus, evaluation of compound-specific data provided by GC–MS can lead to more comprehensive assessments of weathering and biodegradation. The distribution of alkylated PAC homologues of naphthalene, dibenzothiophene, phenanthrene, fluorene and chrysene are commonly employed in this respect [9,18,19]. In general, increased molecular complexity leads to a decrease in the susceptibility to microbial attack. Specifically, the rate of PAC degradation in the environment decreases with increasing compound size (number of rings) and degree of alkylation [9,20,21]. Thus, the rate of degradation of alkylated PACs is $C_0 > C_1 > C_2 > C_3 > C_4$, where C_x denotes the total number x of side chain carbon atoms. Similar effects occur for physical weathering such as evaporation and dissolution, since the physicochemical properties of PACs (i.e. boiling point and solubility) also depend on number of rings and degree of alkylation [9].

Normalization to a conservative internal marker compound such as 17α,21β-hopane [22–25], 17α,21β-norhopane [17] or vanadium [26] has been used frequently in weathering studies to correct for heterogeneity encountered in field samples and to calculate the percent loss of oil or individual analytes [21]. Percent losses based on preserved internal markers describe the combined effects of physical, chemical and biological weathering processes. Only during controlled laboratory experiments can the isolated effects of microbial degradation be described by subtracting the loss in sterile controls. Conversely, in samples collected in the field after oil spills, the low molecular weight hydrocarbons such as heptadecane (nC$_{17}$), octadecane (nC$_{18}$), pristane (Pr), phytane (Ph) and 2–3 ring PACs are often heavily affected by physical weathering processes.

Several authors have observed that microbial degradation is isomer specific [5,7,9,15,27,28]. Changes in nC$_{17}$/Pr and nC$_{18}$/Ph have long been used as indicators for biodegradation [4,29]. Likewise, preferential degradation of specific isomers within homologue PAC series have been described since the 1980s [7,28,30]. Recently, Wang et al. [5,9,31] observed differential susceptibility to degradation within series of alkylated PAC homologues of C$_1$–C$_3$-naphthalenes, methylfluorenes (MF), methylphenanthrenes (MP) and methyldibenzothiophenes (MD) by subjective pattern matching and univariate comparison of diagnostic ratios. In particular, Wang et al. observed isomer specific microbial degradation as a decrease in the ratios (3MP + 2MP)/(4 + 9MP + 1MP), (2 + 3MD)/1MD and $\sum$MF/1MF [5,9].

Changes in isomer patterns within alkylated PAC series and in the corresponding diagnostic ratios are highly specific for biodegradation, as no such changes occur during physical weathering [9]. In contrast, a few studies have shown that chemical weathering can lead to changes in the isomer patterns, although the sequences of alteration are different from those observed during biodegradation [32]. Jacquot et al. observed increasing photodegradability of methylphenanthrenes for the sequence 2MP < 1MP < 3MP < 4 + 9MP [32] where xMP denotes the position x of the methyl substituent in the phenanthrene skeleton. These differences may be used to distinguish between chemical and biological weathering effects.

Methods for evaluating the effects of weathering, and specifically biodegradation, on the composition of spilled oil have mainly been limited to subjective pattern matching and univariate plots. Multivariate statistical methods such as principal component analysis (PCA) provide useful tools for more extensive and objective chemical characterization based on chromatographic data. Multivariate methods have been used regularly for data analysis in organic geochemistry [33,34], but they have only recently been adopted for oil spill identification [35–39] and for studying the fate of petroleum hydrocarbons in the environment [27,40]. The advantages of multivariate over univariate methods are manifold. In particular, they allow for simultaneous analysis of a large number of correlated variables, and thus facilitate a more comprehensive evaluation of the effects of biodegradation.

Here, we present two novel methods for evaluating in situ and in vitro biodegradation of mineral oil. The aim has been to reduce time and cost, to increase the amount of data considered and to increase the objectivity. The methods are based on chemometric data analysis of aligned and normalized regions of GC–MS chromatograms and a selection of diagnostic ratios of PAC isomers within homologue series. These methods were applied to data from an in vitro biodegradation experiment where a North Sea crude oil was exposed to three mixed bacterial inocula over a 1-year-period with five sampling times. Chromatographic data from tiered GC–MS analysis of methylfluorene (m/z 180), methylphenanthrene (m/z 192) and methyldibenzothiophene (m/z 198) groups were used for evaluating effects of microbial degradation.

2. Methods

The methodology employed consists of four steps: sample preparation and GC–MS analysis, chromatographic preprocessing, multivariate statistical data analysis by PCA or weighted-least-squares PCA (WLS-PCA) and data evaluation. Fig. 1 presents a flowchart of this general methodology.

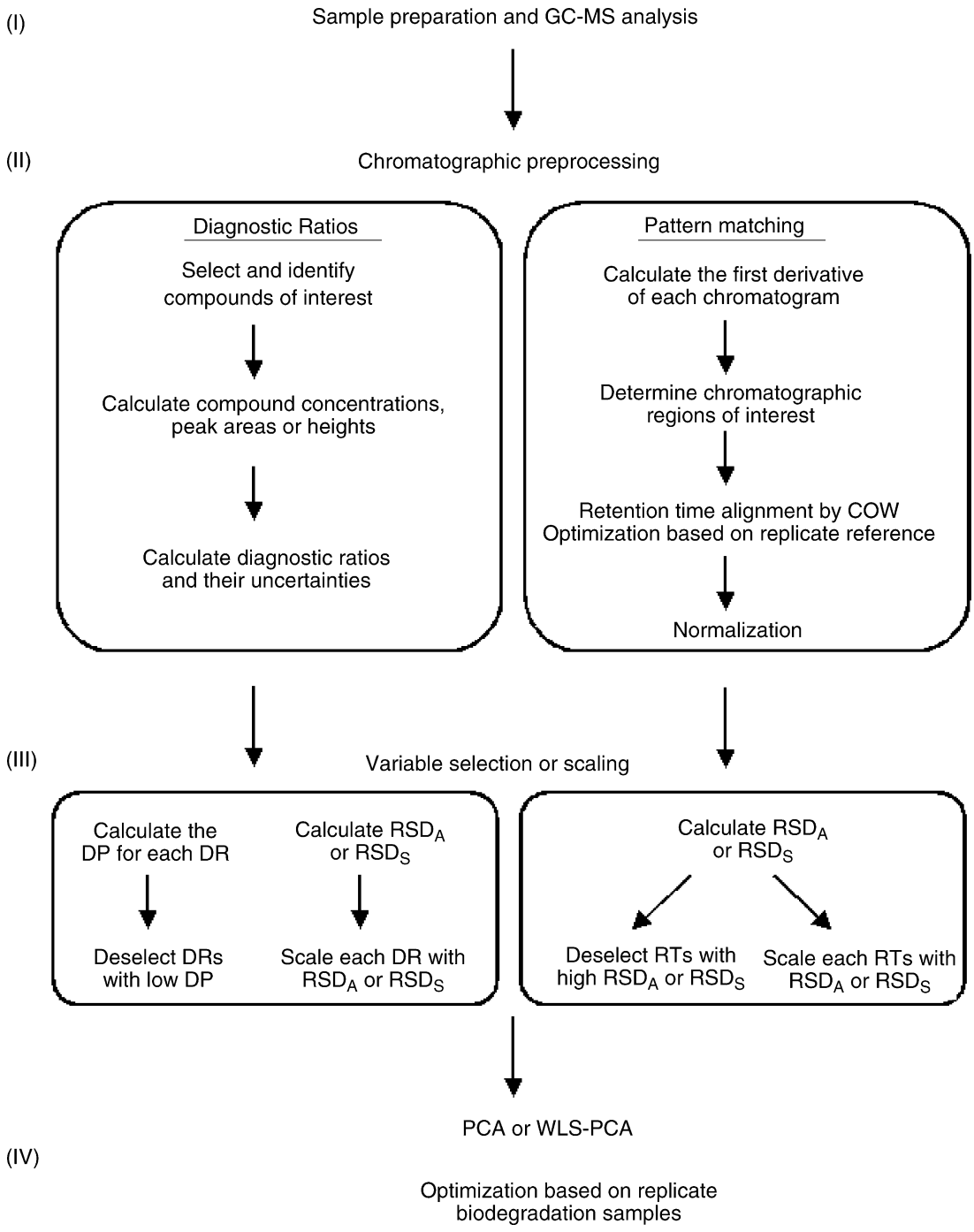

Fig. 1. Flowchart of the general method based on multivariate statistical data analysis for evaluating biodegradation of mineral oil.

The techniques used in the method are automated and objective extensions of standard evaluation techniques often used in the literature, based on previously published methods for chemical fingerprinting using chemometrics [35,38]. Until now, assessment of environmental oil degradation has to a large degree been based on subjective evaluation of changes in GC–MS chromatograms and univariate plots of concentrations and diagnostic ratios.

2.1. In vitro biodegradation experiment

2.1.1. Bacterial strain mixtures

Three mixtures of bacterial strains were used, R, U and M where M was a mixture of R- and U-strains:

R was a mixture of bacterial long-chain alkane degraders and biosurfactant producers, including *Pseudomonas aeruginosa* RR1, *Weeksella* sp. RR7, *Acinetobacter calcoaceticus*

RR8, *Burkholderia cepacia* RR10, *Rhodococcus* spp. RR12 and RR14 and *Xylella fastidiosa* RR15, which have been characterized elsewhere [41].

U was a mixture of bacterial PAC degraders including unclassified Gram-positive strains VM445, VM450 and VM504 and *Sphingomonas* sp. VM506 [42], *Mycobacterium frederiksbergense* Fan9 [43] and a mixture of strains previously described in detail [44]: *Sphingobium* sp. EPA505, *Pseudomonas aeruginosa* CRE11, *Mycobacterium vanbaalenii* PYR-1, *Mycobacterium gilvum* GJ-3p, *Sphingobium chungbukensis* LB126, *Mycobacterium* sp. VF1, *Mycobacterium frederiksbergense* LB501T and *Nocardia asteroides* VM451. Previously, each strain in the U-mixture was found to degrade one or several of the following unsubstituted PACs: phenanthrene, anthracene, fluorene, fluoranthene, pyrene or biphenyl by incubation with ^{14}C-labelled pure compounds (data not shown). In addition, their ability to grow in seawater-modified M9 medium was ensured in a pre-experiment using a complex medium (data not shown).

2.1.2. Incubation

A sample of Brent Crude oil was autoclaved (121 °C, 30 min) and incubated at 10 °C in trace element-supplemented [45] M9 medium [46] (modified by replacing 9/10 of the M9 with autoclaved Baltic Sea water) on a rotary shaker (200 rpm) in the dark. The experimental units were 100 mL Erlenmeyer flasks each containing 25.5 mL medium and 100 mg oil. The experimental treatments included a series of sterile control flasks (SC), and three series of flasks exposed to either R mixture, initial density 3.1×10^6 cells per mL, U mixture, 4.0×10^6 cells per mL or M mixture, 1.1×10^6 cells per mL.

At 20, 54, 132, 224 and 364 days after the start of the experiment, two complete experimental units (replicates *a* and *b*) per treatment plus one sterile control were sacrificed. Hence, the complete data set comprised four samples taken at day 0, seven samples per incubation time plus four analytical replicates of one extract of a R-treatment sacrificed at day 132 (sample 3*a*-R), leading to a total of 43 samples. The extraction of two samples (1*a*-R and 5-SC) failed, leading to a total of 41 samples in the entire data set. Sample labels are explained in Fig. 6. The aim of the in vitro experiment was to assess the biodegradation of oil exposed to any of the three mixtures of bacterial strains, while the sterile controls were used to isolate the effects of physical weathering.

2.2. Sample preparation and GC–MS analysis

The experimental procedure used here is based on a semi-quantitative approach with limited sample preparation and cleanup. The sample units (oil and water) were extracted three times with ~10 mL dichloromethane by shaking the flasks. The water and dichloromethane phases were then separated in a 100 mL separatory funnel, and the dichloromethane phase cleaned through a funnel with glasswool and sodium sulphate to remove particles and water. Each sample was adjusted to a precise volume of 50 mL with dichloromethane resulting in a nominal oil concentration of 2000 mg/L at the beginning of the experiment. A selection of surrogate standards was spiked to samples prior to extraction to enable compound quantification using the internal standard method, but the standards were not used for quantification in this study.

The GC–MS analyses were performed in one GC-run covering a large variety of petroleum hydrocarbon groups including, alkanes, alkylbenzenes, petroleum biomarkers and the C_0-C_4-alkylated PACs (e.g. naphthalenes, dibenzothiophenes, fluorenes, phenanthrenes, biphenyls and chrysenes). Oil extracts were analyzed on a Finnigan TRACE DSQ™ Single Quadrupole GC–MS (Thermo Electron Corporation) operating in EI mode and equipped with a 60 m HP-5MS capillary column (0.25 mm I.D. × 0.25 μm film). One microlitre of aliquots were injected in PTV splitless mode starting at 35 °C and increasing with 14.5 °C/s to 315 °C and hold for 1 min during transfer. Column temperature program: 35 °C (2 min), 60 °C/min to 100 °C, 5 °C/min to 315 (20 min) and transfer line and ion source temperatures: 300, and 250 °C, respectively. Fortyfour mass fragments were acquired in eight groups of 12 ions using selected ion monitoring (SIM). GC–MS/SIM chromatograms of methylfluorenes (*m*/*z* 180), methylphenanthrenes (*m*/*z* 192) and methyldibenzothiophenes (*m*/*z* 198) were used in this study. The chemical structures of methylfluorene, methylphenanthrene and methyldibenzothiophene are shown in Fig. 2.

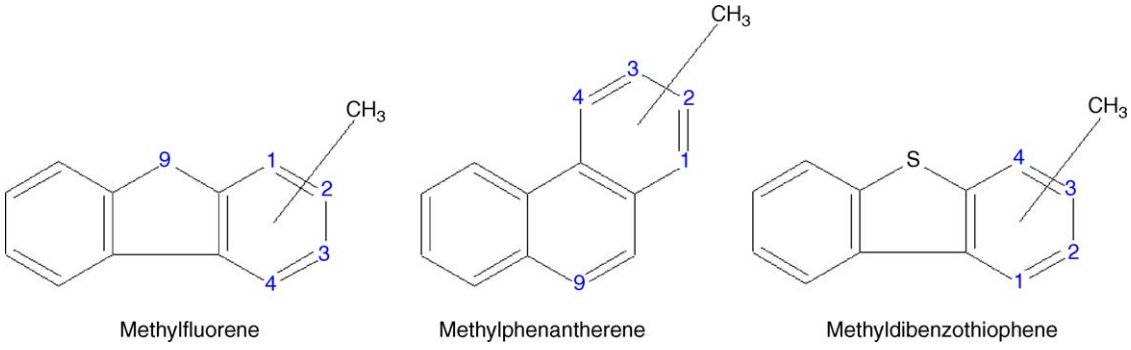

Fig. 2. Chemical structures of methylfluorenes, methylphenanthrenes and methyldibenzothiophenes.

A blank and a laboratory reference oil solution were analyzed between every eight sample extracts (samples were analyzed as part of a larger sequence of samples, blanks, references and quantification standards). The reference oil solution was a 1:1 mixture of Brent crude oil and a heavy fuel oil with a nominal oil concentration of 2500 mg/L. The references were used for quality control, optimizing the data preprocessing and calculating the relative analytical standard deviation (RSD_A).

Although the above sample preparation scheme and GC–MS analytical protocol are recommended as part of a hyphenated method, standard sample preparation and fractionation procedures, as those described by Wang et al. [47,48], can be used alternatively. This would increase the analytical time and costs, but may further enhance the analytical selectivity and robustness.

2.3. Data preprocessing

Multivariate data analyses of sections of GC–MS/SIM chromatograms and diagnostic ratios respectively, require different preprocessing to reduce variations in oil composition unrelated to biodegradation. A selection of preprocessing procedures, including baseline removal, normalization and chromatographic alignment was performed using a modification of the method to preprocess fingerprinting data recently described by Christensen et al. [38]. Conversely, the diagnostic ratio approach required selection and identification of compounds of interest as well as peak quantification.

A proper variable selection or weighting of variables in the fitting of the PCA model is important to exclude or to scale down noisy variables, leading to an increase in the resolution power of the PCA model. Resolution power is defined as the ability of the PCA model to describe the effects of biodegradation, which is the ratio between the variations within replicate biodegradation samples to the total variation described by the model (c.f. Eqs. (5a)–(5c)).

2.3.1. Preprocessing of sections of chromatograms

Chromatographic baselines (see Fig. 3a–c) can negatively affect both warping [49,50] and normalization [38] and are thus, removed by calculating the first derivative. The first derivative is calculated numerically as the difference between consecutive points, which makes integration straightforward by cumulative summation.

The most severe impediment to multivariate statistical analysis of sections of chromatograms is retention time shifts caused mainly by deterioration of the capillary column [51]. Hence, PCA will model both such time shifts and effects of biodegradation as genuine variations in sample composition. This hinders any sensible interpretation of the PCA and thereby creates a need for alignment. Here, the GC–MS/SIM chromatograms were aligned by correlation optimized warping (COW). In COW, a ‘target chromatogram’ is divided into segments, and the optimal boundary positions for corre-

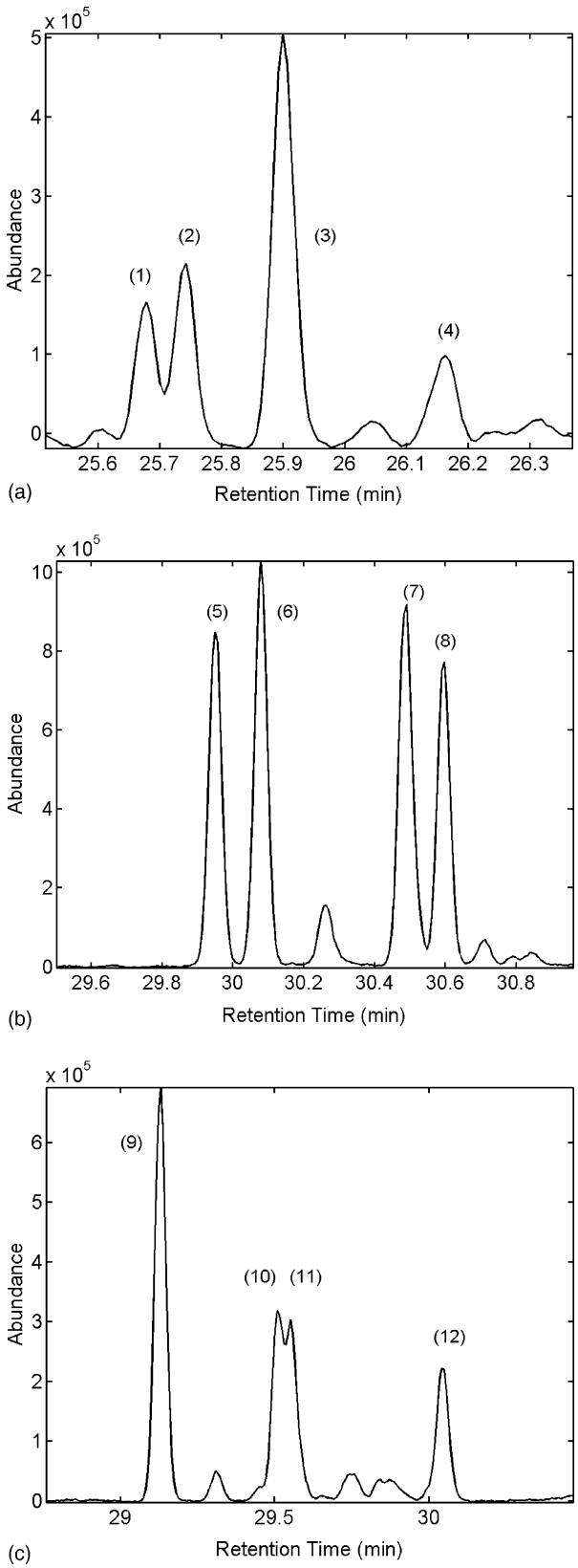

Fig. 3. GC–MS/SIM chromatograms of (a) *m/z* 180, (b) *m/z* 192 and (c) *m/z* 198 in unweathered Brent crude oil.

sponding segments are determined for each of the remaining chromatograms ('sample chromatograms'). Segment length and slack are used as inputs in the COW algorithm, where the slack parameter is the number of data points each boundary is allowed to move. Furthermore, a new parameter, padsize, defined as the number of segments comprised of noise that are added to each end of the chromatograms is introduced. It allows for adequate corrections in the parts retained for the PCA. The noise regions (i.e. pads), added to each chromatogram prior to warping, are removed immediately after warping.

The method described by Christensen et al. [38] to optimize the warping parameters was used. Hence, the optimal choice of warping parameters is the one that maximizes the first singular value in the PCA model without mean centering based on the replicate reference oils. The use of COW for aligning chromatograms has been described in detail elsewhere [38,49,50].

Normalization using Eq. (1) as described by Christensen et al. [38] was applied to compensate for concentration effects and sensitivity changes,

$$x_{nj}^{N} = \frac{x_{nj}}{\sqrt{\sum_{j=1}^{J} x_{nj}^{2}}} \quad (1)$$

where x_{nj} is the first derivative of the n-th chromatogram at the j-th retention time and J is the total number of retention times. In this study, a more complex normalization scheme was also adopted by using only a limited set of retention times. The selection criterion was based on the RSD_A's calculated from aligned and normalized replicate references. Hence, only chromatographic abundances with RSD_A less than a predefined threshold were used for normalization. Under such circumstances, the denominator in Eq. (1) is modified to the sum of the squared first derivatives of selected abundances. In addition, the relative importance of the individual m/z chromatograms (m/z 180, 192 and 198) in the data set, can be adjusted by multiplying each data point with the number of retention times in the respective chromatogram, termed relative scaling of chromatograms.

Preprocessing was performed in MATLAB 6.5.1 (The MathWorks). The COW algorithm was downloaded from http://www.models.kvl.dk.

2.3.2. Data preprocessing of diagnostic ratios

Diagnostic ratios can be calculated either from quantitative (i.e. compound concentrations) or from semi-quantitative data (i.e. peak areas or heights). Ratios of compound concentrations or double-normalized diagnostic ratios, as those proposed by Christensen et al. [35], describe the relative chemical composition. In this study, diagnostic ratios were calculated by using Eq. (2),

$$\mathrm{DR} = \frac{a_n^S}{(a_n^S + a_{n*}^S)} \quad (2)$$

where a_n^S is the peak area, peak height or concentration of compound n or the sum of several compounds in an oil sample. The commercial GC–MS software Xcalibur v. 1.3 was used for peak identification and integration. The diagnostic ratios and the RSD_A's were calculated in Microsoft Excel from peak areas and exported to MATLAB 6.5.1 for chemometric data analysis.

2.4. Variable selection

Some diagnostic ratios or chromatographic abundances are more informative than others, while still others may lead to incorrect conclusions, if kept in the data set. Hence, a proper variable selection is paramount in order to keep the uncertainties to a minimum and yield reliable results.

The approach used here to select diagnostic ratios and retention times for PCA corresponds to the method suggested by Christensen et al. which are based on the diagnostic power (DP) (Eq. (3)) and RSD_A, respectively [35,38].

$$\mathrm{DP} = \frac{\mathrm{RSD_V}}{\mathrm{RSD_A}} \quad (3)$$

where RSD_A is calculated from the replicate reference oils and RSD_V is the relative standard deviation of the 41 laboratory biodegradation samples. The denominator in Eq. (3) can be substituted by RSD_S, which in biodegradation studies would be the relative standard deviation calculated from samples affected by physical or chemical weathering processes. Subsequently, either an increasing number of variables can be excluded from the data analysis starting with the one with the smallest DP and removing one at a time until exhaustion, or, alternatively, variables with RSD_A below a predefined threshold are excluded from the PCA.

2.5. Chemometric data analysis

The preprocessed data are collected in a two-way data matrix ($\mathbf{X}$) of size I (oil samples) $\times$ J (diagnostic ratios or chromatographic abundances). Subsequently, $\mathbf{X}$ is bilinearly decomposed by PCA into products of scores, $\mathbf{t}$ ($I \times 1$), and loading vectors, $\mathbf{p}^T$ ($1 \times J$) (i.e. T in superscript means the transposed matrix of $\mathbf{p}$), plus residuals, $\mathbf{E}$ ($I \times J$). The bilinear decomposition with K principal components is defined in Eq. (4).

$$\mathbf{X} = \left(\sum_{k=1}^{K} \mathbf{t}_k \times \mathbf{p}_k^T\right) + \mathbf{E} \quad (4)$$

In addition, PCA was fitted according to a weighted-least-squares criterion (WLS-PCA). We applied the PCAW algorithm [52] for this purpose using the inverse of RSD_A as weights. All chemometric data analyses were performed in MATLAB 6.5.1 The PCAW algorithm was downloaded from http://www-its.chem.uva.nl/research/pac.

2.6. Optimal preprocessing and data analysis

The optimal combination of preprocessing and data analysis is determined by maximizing the resolution power. Specifically, the optimization was done by minimizing the variance of replicate samples s with respect to their average scores $\bar{\mathbf{t}}_S$ (Eq. (5a)) compared to the variance explained by the model (Eq. (5b)) using a predefined number of principal components K.

$$d_{\mathrm{Rep}} = \sum_{s} (n_s - 1)^{-1} \sum_{i \in S_s} (\mathbf{t}_i - \bar{\mathbf{t}}_s)(\mathbf{t}_i - \bar{\mathbf{t}}_s)^T \quad (5a)$$

$$d_{\mathrm{All}} = \sum_{i=1}^{I} \mathbf{t}_i \mathbf{t}_i^T \quad (5b)$$

$$r = \frac{d_{\mathrm{Rep}}}{d_{\mathrm{All}}} \quad (5c)$$

where S_s are the row indexes for the n_s replicates of oil sample s, $\mathbf{t}_i$ is the i-th row of T, and r is dimensionless. Thus, the PCA with highest resolution power minimizes r (Eq. (5c)), where d_{All} is the normalization factor that accounts for the increase in d_{Rep} due to a larger number of principal components.

3. Results and discussion

The multivariate statistical approaches for evaluating biodegradation of mineral oil were tested on oil samples from the in vitro biodegradation experiment. The 'complete data set' comprised 74 samples including 33 replicate references ('the reference set') and the 41 biodegradation samples ('the sample set'). For each sample, GC–MS data from methylfluorene (m/z 180, Fig. 3a), methylphenanthrene (m/z 192, Fig. 3b) and methyldibenzothiophene (m/z 198, Fig. 3c) isomer groups were used for evaluating effects of microbial degradation, based on relative changes in isomer abundances. The identity of individual isomers and abbreviations used in the remaining sections are listed in Table 1. The former was determined by comparing peak retention times and mass-spectra with that of standards purchased from Chiron AS, Norway except for 3-methylfluorene, which was tentatively identified. The 9-methylfluorene isomer elutes 2 min prior to the peak cluster of methylfluorenes used in this study.

Table 1
Description and abbreviation of individual isomers of methylfluorenes, methylphenanthrenes and methyldibenzothiophenes

No.	Compound	Abbreviation
1	3-Methylfluorene (tentatively identified)	3MF
2	2-Methylfluorene	2MF
3	1-Methylfluorene	1MF
4	4-Methylfluorene	4MF
5	3-Methylphenanthrene	3MP
6	2-Methylphenanthrene	2MP
7	4- and 9-Methylphenanthrene (coeluting peaks)	49MP
8	1-Methylphenanthrene	1MP
9	4-Methyldibenzothiophene	4MD
10	2-Methyldibenzothiophene	2MD
11	3-Methyldibenzothiophene	3MD
12	1-Methyldibenzothiophene	1MD

3.1. Evaluation of biodegradation based on time warping and PCA

3.1.1. Retention time alignment

The capillary column was trimmed during the analytical sequence, thus chromatograms obtained after the column trimming were manually shifted (33 data points) prior to warping to avoid the need for inexpedient high flexibility (i.e. high slack) that could result in insufficient alignment. After manual shifting, calculation of the first derivatives and exclusion of noisy data points from the beginning and end of the chromatograms (five points at each side), the three m/z chromatograms comprised 700 (m/z 180), 700 (m/z 192) and 790 (m/z 198) data points.

Initially, using combinations of segment lengths of 100, 150, 200 and 250, slacks of 2, 3 and 4 and padsizes of 2, 4 and 6 times the applied segment length were tested to find the optimal warping parameters for the reference set. A reference sample analyzed halfway through the analytical sequence was selected as the target chromatogram in COW to reduce the need for correction. After warping, the three GC–MS/SIM chromatograms were reduced to 90, 175 and 225 data points, respectively, by omitting parts before and after the relevant peak clusters. The explained variance of a one-component PCA model, which is proportional to the first singular value, was calculated for the normalized data sets in the reference set (33×90, 33×175 and 33×225, respectively). It increased from 43.6, 49.1 and 44.7% (without warping) to 97.6, 98.5 and 98.4% for the optimal warping attained for segments of 200, 150 and 150, slacks of 4, 4 and 4 and padsizes of 4, 6 and 4 times the optimal segment lengths.

The complete data sets (74×90, 74×175 and 74×225 data points) were aligned using the optimal warping parameters determined from the reference set. Chromatograms in each data set were then normalized using Eq. (1) and combined for each sample in the reference set (33×490 data points) and in the sample set (41×490 data points).

The effects of the optimal time warping procedure on 10 randomly selected normalized and combined reference chromatograms are illustrated in Fig. 4. Note that the plot regions are extended compared to the size of the sample set. The addition of sections of noise prior to warping was important and allowed for adequate corrections in the part retained for the PCA. The improvement is evident and, as also observed by Christensen et al. [38], the residual shift after warping was at most one point when considering the complete data set. Compounds present in individual GC–MS/SIM chromatograms (tri- and tetra-cyclic steranes and methylated PAC homologues) have similar physicochemical properties. Hence, they are affected similarly by changes in the column properties,

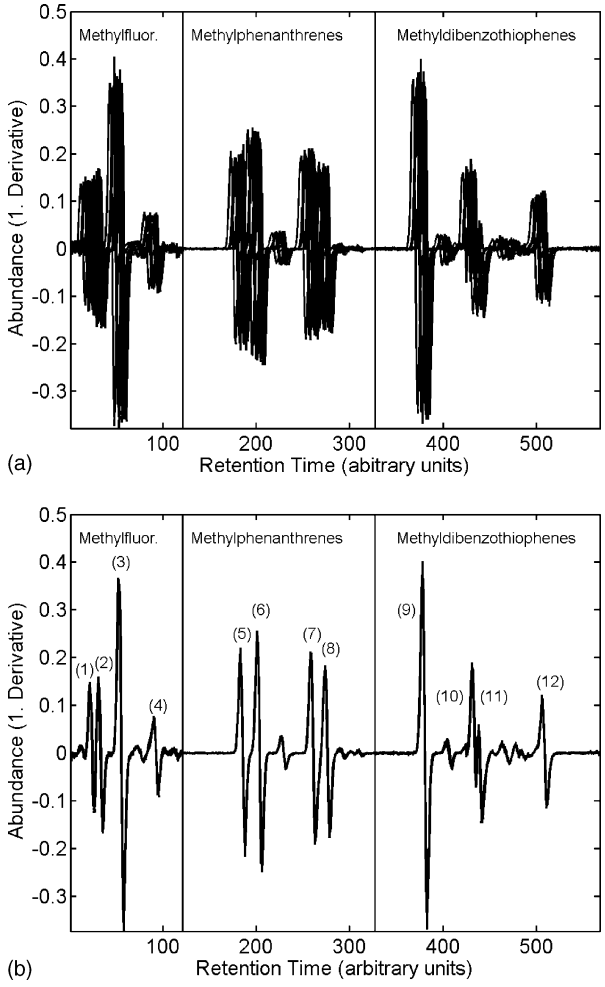

Fig. 4. First derivative of the normalized and combined chromatograms of *m/z* 180, 192 and 198 for 10 references: (a) without time alignment and (b) with time alignment. The optimal COW parameters consisted of segment lengths of 200, 150 and 150, slack of four for all three *m/z* values and padsize of 4, 6 and 4 times the optimal segment lengths.

which is the most likely reason for the consistently high warping quality of COW on GC–MS/SIM data.

Conversely, in chromatograms based on less selective detectors such as GC-FID, retention time shifts may appear more irregular since the variation in physicochemical properties of the compounds present generally are greater than in a GC–MS/SIM chromatogram with more closely related compounds. Thus, closely related isomers (as detected by GC–MS/SIM) are not expected to change their order of elution even if significant changes in the column stationary phase take place, while compounds with different physicochemical properties (as detected by GC-FID) often do, and COW cannot adequately correct for such changes. This may explain why some authors [49,50] have experienced that optimal chromatographic alignment is achieved if the segment length is of the same order of magnitude as the peaks, since a high flexibility is necessary to correct for irregular retention time shifts.

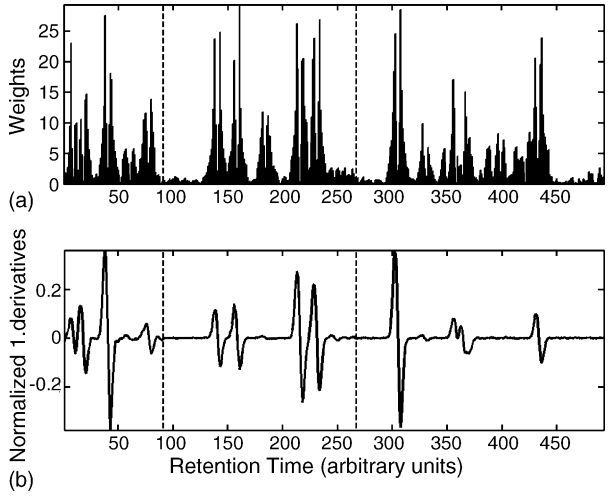

Fig. 5. (a) Weights (RSD_A^{-1}) used for WLS-PCA and (b) the aligned and normalized, mean combined chromatogram of unweathered Brent crude oil. The mean was calculated from four replicate samples.

3.1.2. Variable selection

The RSD_A's were calculated from the aligned and normalized reference set using the optimal warping parameters and the complete chromatographic sections (90, 175 and 225 data points, respectively) for normalization. The RSD_A's were used for variable selection by excluding retention times with RSD_A below a predefined threshold thereby retaining the peak regions. Fig. 5a shows the weights (RSD_A^{-1}) used in WLS-PCA to scale the importance of retention times with respect to their analytical uncertainty, and Fig. 5b shows the first derivative of the aligned and normalized mean combined chromatogram of the unweathered Brent crude oil.

The peak and noise regions have, respectively, high and low weights. Thus, the importance of peak regions is high, compared to those of noise regions, during fitting of the PCA model (Fig. 5).

3.1.3. Chemometric data analysis

PCA was performed on the mean-centered sample set using combinations of normalization, relative scaling of the three GC–MS/SIM chromatograms, PCA methods and optimal number of principal components (PCs). The chromatographic sections (90, 175 and 225 data points, respectively) were applied for normalization using Eq. (1). Furthermore, thresholds from 0.1 to 1.0 with increments of 0.1 were imposed, thereby retaining between 13 and 78 data points in *m/z* 180, 23–116 in *m/z* 192 and 17–133 in *m/z* 198. Subsequently, data were either left unchanged or multiplied by the sizes of the corresponding data sets. This increased the relative importance of *m/z* 198 (225 data points) compared to *m/z* 192 (175 data points) and *m/z* 180 (90 data points). Finally, data were analyzed by PCA with variable selection or WLS-PCA using two or three PCs. The optimal preprocessing and data analysis was obtained by minimizing *r* in

Eq. (5) (maximizing resolution power) using the duplicate biodegradation samples to calculate d_{Rep}.

The resolution power of PCA models with three PCs was significantly lower ($r = 0.0346–0.231$) than for PCA models with two PCs ($r = 0.0188–0.0477$). Furthermore, from visual inspection of loadings it was clear that PC3 and higher order components described the residual shifts rather than systematic changes in chemical composition. Residual shifts show up in the cumulative sum of loadings as first derivative peaks. As an increasing number of PCs is extracted from data, the ratio between the systematic information and the variations caused by insufficient alignment decreases until the latter becomes the most pronounced. Christensen et al. [38] extracted four reliable components from the calibration set of 61 oil samples × 1231 data points. In this study only two components could be extracted, since the number of independent variations was smaller.

Normalization and relative scaling of chromatograms influenced the resolution power. Normalization consistently affected the *r*-value for WLS-PCA from 0.206 (with a threshold at 0.1) to 0.230 (using all data for normalization). Relative scaling diminished the resolution power of the PCA model, because the *m*/*z* 198 chromatogram became more influential and thereby introducing more noise in data (e.g. the *r*-value of PCA with variable selection using a threshold of 0.2 increased from 0.0188 without scaling to 0.0263 with scaling).

The highest resolution power ($r = 0.0188$ to 0.0198) was obtained by PCA with variable selection using a threshold of 0.2–0.3. However, WLS-PCA with normalization to data points with $RSD_A < 0.1$ gave comparable resolution power ($r = 0.0206$), and it is a more attractive method since no data are excluded from the analysis. Furthermore, the resolution power may not always be the best criterion for choosing the optimal combination of preprocessing and data analysis following time alignment. Exclusion of small but important variations caused by decreasing the threshold may by chance lead to an increased resolution power.

The score plot of PC1 versus PC2 using WLS-PCA with mean-centering, normalization to data points with $RSD_A < 0.1$ and no scaling is shown in Fig. 6. The two PCs describe 82.9% of the total variation in the sample set.

Sterile control samples, R samples incubated from 20 to 132 days (sampling time 1–3) and M and U samples incubated for 20 days cluster together with oil samples from day 0, at low PC1, and ∼0 in PC2 (Fig. 6). Thus, the changes in relative isomer abundances of methylfluorenes, methylphenanthrenes and methyldibenzothiophenes appear to be insignificant in these samples. The low variation within the cluster of unaffected oil samples, including the sterile controls, shows that most of the variations unrelated to the chemical composition have been effectively removed by the chromatographic preprocessing procedure. Likewise, the within-class variations of replicate references (not shown) are low compared to the class-to-class variations (from one sample to the next sample) observed in the score plot.

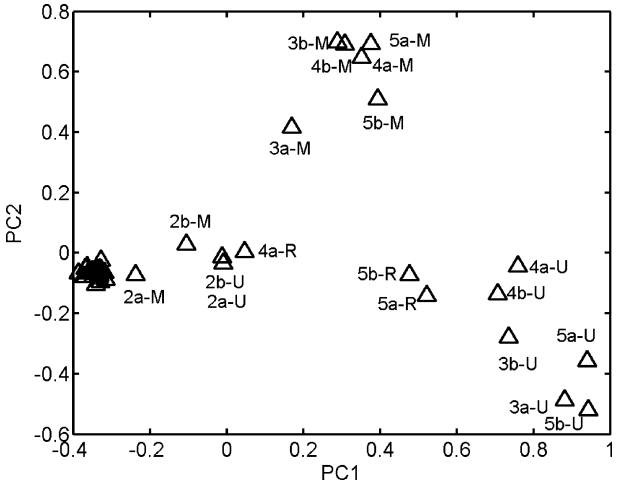

Fig. 6. Score plot PC1 vs. PC2 based on the sample set (41 × 490 data points), using WLS-PCA and normalization to data points with $RSD_A < 0.1$. The sample labels signify; 0–5, sampling time 0, 20, 54, 132, 224 and 364 days; *a* and *b*, duplicates; R, U and M, the bacterial strain mixture; SC, sterile control. The labels for the cluster of unaffected samples located at low PC1 and around 0 for PC2 have been left out.

M samples incubated for 54–364 days are located at increasing PC1 and PC2 depending on incubation time. For instance, samples incubated for 54 days (2*a*-M and 2*b*-M) are located at lower PC1 and PC2 than samples incubated for 132 days (3*a*-M and 3*b*-M) and so forth (e.g. 5*a*-M and 5*b*-M). Likewise, U and R samples are located at increasing PC1, but decreasing PC2 as incubation time increases. The results indicate that PC1 describes the general biodegradation pattern of the Brent crude oil and signifies an increasing change in the PAC isomer patterns due to preferential degradation of isomers. The U- and M- mixtures degrade the PAC methyl homologues faster than the R mixture.

The biological duplicates 4*a*-R/4*b*-R and 3*a*-M/3*b*-M have a greater variability in the score plot than the remaining duplicates. The reason is most likely that these samples are collected during periods where the rate of microbial degradation (and hence the degree of preferential degradation) of C1-PAC isomers is high, since the biological variability becomes more pronounced during such periods.

The cumulative sums of the PC1 and PC2 loadings are compared in Fig. 7a and b with the normalized mean combined chromatogram of unweathered Brent crude oil.

The general degradation order along PC1 within each individual homologue series can be deduced by comparing the normalized loadings of individual peaks with the normalized mean chromatogram of the unweathered Brent crude oil (Fig. 7b). The PC1 loadings for methylphenanthrene isomers compared to the mean chromatogram are for instance lowest (most negative) for 2MP and 3MP followed by 1MP and 49MP, where the latter has the most positive loading compared to the mean chromatogram. Thus, the relative concentrations of 2MP and 3MP are lowest in oil samples located at high PC1 and highest for samples located at low

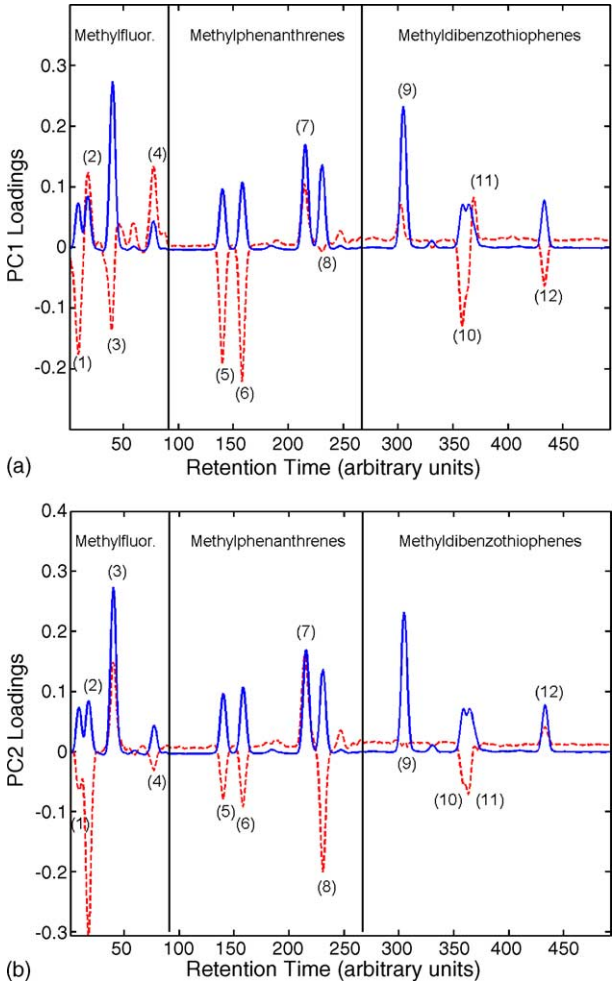

Fig. 7. Loading plots based on the sample set (41 × 490), using WLS-PCA and normalization to data points with $RSD_A < 0.1$. (a) The cumulative sum of the PC1 loadings (red and broken line) compared with the cumulative sum of the aligned and normalized mean chromatogram of unweathered Brent crude oil (blue line) and (b) The cumulative sum of the PC2 loadings (red and broken line) compared with the unweathered Brent crude oil (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

PC1 (the unaffected samples), whereas the opposite is the case for 49MP. The general degradation orders observed are 2MP = 3MP > 1MP > 49MP for the methylphenanthrene homologue series, 3MF > 1MF > 2MF > 4MF for methylfluorenes and 2MD > 1MD > 4MD > 3MD for methyldibenzothiophenes. It is of special interest that the co-eluting peaks 2- and 3-methyldibenzothiophenes (23MD), which usually are considered as one common peak cluster, show preferential degradation. Hence, 2MD is the isomer most susceptible to microbial degradation by R, U and M mixtures, whereas 3MD is the least degradable. In fact, due to the significant peak overlap, the retention time of 3MD changes due to loss of 2MD, a trend which can be observed in the loading plot as a shift to higher retention time of the 3MD isomer peak.

Another interesting observation in the score plot is that the M versus the U and R samples are separated along PC2.

Table 2
Description of diagnostic ratios, RSD_A, RSD_V and DP for each diagnostic ratio calculated from the reference set

Diagnostic ratios	RSD_V (%)	RSD_A (%)	DP
2MP/(1MP + 2MP)	72.4	0.7	101.3
3MP/(3MP + 1MP)	66.2	0.7	100.8
3MP/(3MP + 49MP)	67.8	0.8	88.8
2MP/(2MP + 49MP)	74.0	0.9	79.1
1MF/(1MF + 4MF)	11.2	0.3	35.2
3MF/(3MF + 4MF)	72.1	2.2	33.1
2MD/(2MD + 1MD)	73.8	2.3	31.8
49MP/(49MP + 1MP)	15.0	0.5	28.7
2MF/(2MF + 4MF)	33.0	1.4	24.2
2MF/(2MF + 1MF)	50.8	2.2	22.6
3MF/(3MF + 1MF)	74.0	3.4	21.7
3MF/(3MF + 2MF)	66.3	3.2	20.5
3MP/(3MP + 2MP)	10.1	0.5	18.7
2MD/(2MD + 3MD)	74.7	4.8	15.6
4MD/(4MD + 1MD)	7.5	0.7	10.9
4MD/(4MD + 2MD)	12.2	1.6	7.7
3MD/(3MD + 1MD)	20.9	2.9	7.3
4MD/(4MD + 3MD)	5.1	2.0	2.6

Thus M, the combined mixture of bacterial strains in U and R, affected the PAC isomer pattern differently than did both the U and the R mixtures. The PC2 loadings are compared with the aligned and normalized mean chromatogram of unweathered Brent crude oil in Fig. 7b. The comparison indicates a less preferential degradation of 1MF (most positive loading) over 2MF (most negative) and 1MD (most positive) over ∑23MD (most negative) in M samples than in U and R samples. Furthermore, a more preferential degradation of 1MP (most negative) over 49MP (most positive) is observed.

3.2. Evaluation of biodegradation based on diagnostic ratios and PCA

3.2.1. Data and variable selection

All possible combinations of diagnostic ratios of single compounds within each of the homologue series of C_1-PACs were calculated. Hence, 18 diagnostic ratios (Table 2) with potential sensitivity to biodegradation due to preferential degradation of specific isomers were included in this study. The reference set (33 samples × 18 diagnostic ratios) was used to calculate the RSD_A's, whereas the sample set (41 samples × 18 diagnostic ratios) was used to calculate RSD_V and the PCA models. The RSD_A, RSD_V and DP for each diagnostic ratio are listed in Table 2.

3.2.2. Chemometric data analysis

PCA was performed on the mean-centered sample set using PCA with variable selection (including 1–18 diagnostic ratios) and WLS-PCA, with 2 and 3 PCs leading to 38 individual PCAs. The optimal data analysis was determined by optimizing the resolution power based on the duplicate biodegradation samples.

The number of PCs was determined by cross validation to three, and further verified by evaluating the loadings and

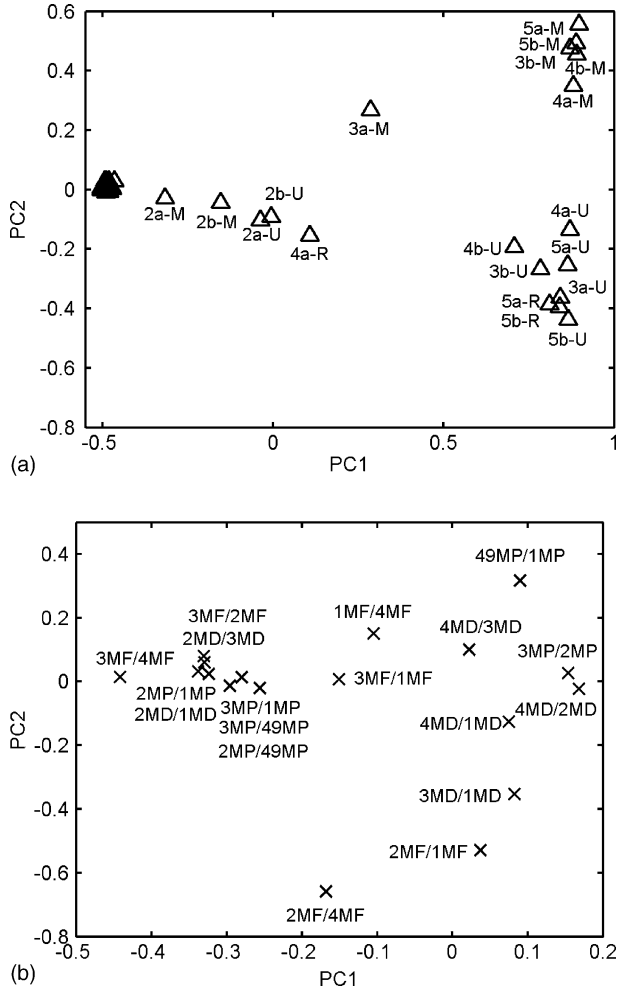

Fig. 8. WLS-PCA on mean-centered data using 3 PCs. (a) Score plot of PC1 vs. PC2 based on the sample set (41 × 18) and (b) loading plot of PC1 vs. PC2. The labels of diagnostic ratios are simplified.

scores. The optimal number of PCs varied depending on the number of diagnostic ratios retained for the analysis. The highest resolution power ($r = 0.0207$–0.0251) was obtained by PCA with variable selection using between 5 and 13 diagnostic ratios with highest DP. However, the resolution power was comparable for the WLS-PCA model ($r = 0.0252$).

The score and loading plots of PC1 versus PC2 for the three-component WLS-PCA model with mean-centering are shown in Fig. 8a and b, respectively. The three PCs describe 98.9% of the total variation in the sample set.

PC1 describes the general biodegradation pattern of the Brent crude oil. The effects of microbial degradation on the relative isomer abundances are shown as an increase (ratios located at high PC1 in the loading plot) or decrease (ratios located at low PC1) in the diagnostic ratios. The general biodegradation trends observed along PC1 causes among others a decrease in 3MP/(3MP + 1MP), 2MD/(2MD + 3MD) and 3MF/(3MF + 4MF) and an increase in 4MD/(4MD + 1MD). The M versus U and R samples are separated along PC2. The cause of this separation can be explained by a stronger increase in 49MP/(49MP + 1MP) and a weaker increase in 2MF/(2MF + 1MF) and 3MD/(3MD + 1MD). Hence, 1MD is less degraded compared to ∑MD and 1MP more degraded than 49MP in M samples than in U and R samples. The within-class variations (e.g. the unaffected oil samples) are, as for the evaluation based on time-warping and PCA, much less than the class-to-class variations. The conclusions made from Fig. 8a and b correspond to those made from Figs. 6 and 7 using time warping and PCA. In addition, although a minor component, data for U and R samples are separated by PC3 due to less preferential degradation of e.g. 1MF over 2MF and 3MP over 2MP in R samples compared to U samples.

Co-elution of compounds at low concentrations that interfere with target compounds, such as C_5-naphthalenes with C_1-dibenzothiophenes in *m/z* 198, may affect the results negatively. The risk of introducing undesirable variability from such minor components is higher when using preprocessed chromatographic sections compared to the diagnostic ratio approach based on peak identification and quantification. Furthermore, the biodegraded samples are more likely to be affected since the relative abundances of these minor components may change due to selective weathering processes. However, since the concentrations of the C_1-PAC isomers are high in Brent crude oil the results are, in the present study, most likely unaffected by co-elution of minor components.

Time warping combined with PCA represents a highly objective alternative to standard evaluation methods. The complete data treatment, including data preprocessing and analysis, can be performed with only limited human intervention. However, whether or not the method provides an appropriate tool for assessing effects of weathering in a specific case study depends on the ratio between the inherent variability in the data set and the variability due to insufficient alignment. Two components were extracted, which was one less than by the diagnostic ratio approach. Although the third component (PC3) explained only a minor percentage of the total variance it was important for separating R samples from U (and M) samples.

Furthermore, variations in peak shape (e.g. from symmetrical to tailing) during column deterioration will affect the multivariate data analysis negatively due to changes in intensity distribution of adjacent retention times within peak regions. Peak quantification is less affected by these factors since peak areas and heights are relatively independent of peak shape.

4. Conclusions and perspectives

Assessments of biodegradation in the field after oil spills and during in vitro experiments are difficult due to the inherent chemical complexity of mineral oil and to the multitude of processes changing its composition. The two novel methods presented in this paper enabled a comprehensive analysis of the effects of biodegradation with limited time consump-

tion and cost and high objectivity compared with standard assessment methods. In the first method based on analyses of sections of chromatograms, the large data set (41 samples × 490 retention times) was evaluated in three plots (i.e. one score plot and two loading plots). In the diagnostic ratio approach, a third component was extracted from the data set (41 samples × 18 diagnostic ratios) distinguishing R from U samples. The results showed that the PAC degradation pattern was different for each of the three bacterial strain mixtures. Thus, preferential degradation patterns in the natural environment cannot be established from data obtained during in vitro experiments using simple mixtures of bacterial strains.

A wide range of potential applications exists for the two evaluation methods described in this paper such as assessment of other complex mixtures of contaminants (e.g. persistent organic pollutants) and other weathering processes. The relative changes in the homologue series of *n*-alkanes (e.g. *m*/*z* 85) can for example be used to deduce the effects of evaporation, whereas ratios of *n*-alkanes and isoprenoids can describe the initial phase of biodegradation.

Furthermore, by normalizing to one or several surrogate standards (which corresponds to the internal quantification method) or by using the normalization factor of the closest reference for normalization (which corresponds to the external quantification method), concentration effects can be retained in data. Such modifications would allow the monitoring of levels of complex mixtures of pollutants in environmental samples with limited use of resources and high objectivity compared to standard tiered analytical approaches.

Acknowledgements

The authors acknowledge Lotte Frederiksen for technical assistance, D. Springael for providing several of the PAC-degrading bacterial strains, and the European Commission (contracts “BIOSTIMUL”, QLRT-1999-00326, and “ALARM”, GOCE-CT-2003-506675), the Danish Natural Sciences Research Council and the National Environmental Research Institute, Denmark, for funding.

References

[1] Z.D. Wang, M. Fingas, J. Chromatogr. A 712 (1995) 321.
[2] G. Mille, D. Munoz, F. Jacquot, L. Rivet, J.C. Bertrand, Estuar. Coast. Shelf. S 47 (1998) 547.
[3] R.C. Prince, E.H. Owens, G.A. Sergy, Mar. Pollut. Bull. 44 (2002) 1236.
[4] T.C. Sauer, J.S. Brown, P.D. Boehm, D.V. Aurand, J. Michel, M.O. Hayes, Mar. Pollut. Bull. 27 (1993) 117.
[5] Z. Wang, M. Fingas, H. Blenkinsopp, G. Sergy, M. Landriault, L. Sigouin, P. Lambert, Environmen. Sci. Technol. 32 (1998) 2222.
[6] J.D. Leblond, T.W. Schultz, G.S. Sayler, Chemosphere 42 (2001) 333.
[7] S.J. Rowland, R. Alexander, R.I. Kagi, D.M. Jones, A.G. Douglas, Org. Geochem. 9 (1986) 153.
[8] G. Thouand, P. Bauda, J. Oudot, G. Kirsch, C. Sutton, J.F. Vidalie, Can. J. Microbiol. 45 (1999) 106.
[9] Z.D. Wang, M. Fingas, S. Blenkinsopp, G. Sergy, M. Landriault, L. Sigouin, J. Foght, K. Semple, D.W.S. Westlake, J. Chromatogr. A 809 (1998) 89.
[10] R.C. Prince, R.M. Garrett, R.E. Bare, M.J. Grossman, T. Townsend, J.M. Suflita, K. Lee, E.H. Owens, G.A. Sergy, J.F. Braddock, J.E. Lindstrom, R.R. Lessard, Spill. Sci. Technol. B 8 (2003) 145.
[11] B.V. Chang, S.H. Wei, S.Y. Yuan, J. Environ. Sci. Heal. B 36 (2001) 177.
[12] W.R. Cullen, X.F. Li, K.J. Reimer, Sci. Total. Environ. 156 (1994) 27.
[13] A.T. Law, K.S. Teo, J. Mar. Biotechnol. 5 (1997) 162.
[14] J.C. Roper, F.K. Pfaender, Environ. Toxicol. Chem. 20 (2001) 223.
[15] H. Budzinski, N. Raymond, T. Nadalig, M. Gilewicz, P. Garrigues, J.C. Bertrand, P. Caumette, Org. Geochem. 28 (1998) 337.
[16] T.K. Dutta, S. Harayama, Environ. Sci. Technol. 34 (2000) 1500.
[17] J. Oudot, F.X. Merlin, P. Pinvidic, Mar. Environ. Res. 45 (1998) 113.
[18] P.D. Boehm, G.S. Douglas, W.A. Burns, P.J. Mankiewicz, D.S. Page, A.E. Bence, Mar. Pollut. Bull. 34 (1997) 599.
[19] D. Munoz, P. Doumenq, M. Guiliano, F. Jacquot, P. Scherrer, G. Mille, Talanta 45 (1997) 1.
[20] C.E. Cerniglia, Biodegradation 3 (1992) 351.
[21] G.S. Douglas, A.E. Bence, R.C. Prince, S.J. McMillen, E.L. Butler, Environ. Sci. Technol. 30 (1996) 2332.
[22] J.R. Bragg, R.C. Prince, E.J. Harner, R.M. Atlas, Nature 368 (1994) 413.
[23] E.L. Butler, G.S. Douglas, W.G. Steinhauer, On-Site Bioreclamation, 2001, p. 515.
[24] R.C. Prince, D.L. Elmendorf, J.R. Lute, C.S. Hsu, C.E. Halth, J.D. Senius, G.J. Dechert, G.S. Douglas, E.L. Butler, Environ. Sci. Technol. 28 (1994) 142.
[25] D. Venosa, M.T. Suidan, D. King, B.A. Wrenn, J. Ind. Microbiol. Biot. 18 (1997) 131.
[26] T. Sasaki, H. Maki, M. Ishihara, S. Harayama, Environ. Sci. Technol. 32 (1998) 3618.
[27] J.H. Christensen, Polycycl. Aromat. Comp. 22 (2002) 703.
[28] M.C. Kennicutt, Oil. Chem. Pollut. 4 (1988) 89.
[29] L.B. Christensen, T.H. Larsen, Ground. Water. Monit. R 13 (1993) 142.
[30] J.K. Volkman, Org. Geochem. 6 (1984) 619.
[31] Z. Wang, M. Fingas, Environ. Sci. Technol. 29 (1995) 2842.
[32] F. Jacquot, M. Guiliano, P. Doumenq, D. Munoz, G. Mille, Chemosphere 33 (1996) 671.
[33] N. Telnaes, B. Dahl, Org. Geochem. 10 (1986) 425.
[34] K. Øygard, O. Grahl-Nielsen, S. Ulvøen, Org. Geochem. 6 (1984) 561.
[35] J.H. Christensen, A.B. Hansen, G. Tomasi, J. Mortensen, O. Andersen, Environ. Sci. Technol. 38 (2004) 2912.
[36] W.A. Burns, P.J. Mankiewicz, A.E. Bence, D.S. Page, K.R. Parker, Environ. Toxicol. Chem 16 (1997) 1119.
[37] S.A. Stout, A.D. Uhler, K.J. McCarthy, Environ. Forens. 2 (2001) 87.
[38] J.H. Christensen, G. Tomasi, A.B. Hansen, Environ. Sci. Technol. 39 (2005) 255.
[39] J.H. Christensen, A.B. Hansen, J. Mortensen, O. Andersen, Anal. Chem. 77 (2005) 2210.
[40] C.D. Simpson, C.F. Harrington, W.R. Cullen, Environ. Sci. Technol. 32 (1998) 3266.
[41] L. Yuste, M.E. Corbella, M.J. Turiégano, U. Karlson, A. Puyes, F. Rojo, FEMS Microbiol. Ecol. 32 (2000) 69.
[42] D. Springael, Personal Communication, 2005.
[43] P.A. Willumsen, U. Karlson, E. Stackebrandt, R.M. Kroppenstedt, Int. J. Syst. Evol. Micr. 51 (2001) 1715.
[44] A.R. Johnsen, U. Karlson, Appl. Microbiol. Biot 63 (2004) 452.
[45] T. Bauchop, S.R. Elsden, J. Gen. Microbiol 23 (1960) 457.

[46] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Spring Harbor, New York, 1989.
[47] Z.D. Wang, M. Fingas, K. Li, J. Chromatogr. Sci. 32 (1994) 367.
[48] Z.D. Wang, M. Fingas, K. Li, J. Chromatogr. Sci. 32 (1994) 361.
[49] N.P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[50] G. Tomasi, F. van den Berg, C. Andersson, J. Chemometr. 18 (2004) 231.
[51] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 72 (2000) 4154.
[52] H.A.L. Kiers, Psychometrika 62 (1997) 251.