Development and application of a hyphenated analytical and chemometric method in petroleum characterisation, identification and degradation studies

J. H. Christensen\textsuperscript{1,2}, A. Hansen\textsuperscript{2}, O. Andersen\textsuperscript{1} & J. Mortensen\textsuperscript{1}

\textsuperscript{1}Roskilde University – Department of Life Science and Chemistry, Denmark.
\textsuperscript{2}National Environmental Research Institute – Department of Environmental Chemistry, Denmark.

Abstract

In this ongoing study a hyphenated chemical and statistical method for classification, identification and biodegradation of environmental petroleum spills has been developed. A robust chemometric model (PCA) has been developed using 23 crude oils and applied for identification of an artificial oil spill sample. The utility of the method for fast and improved integration of GC-MS peaks in highly complex mixtures is documented and its potential in identification, characterisation and degradation studies is discussed and illustrated by examples.

Introduction

Petroleum spilled in the environment either at sea or on land constitutes an increasing problem in the world today. One can only imagine how much oil is spilled world-wide as many spills are not reported and smaller spills are not accounted for. In 1998 nearly 108000 tonnes of oil was spilled world-wide into the marine and inland environments from all sources as estimated from 215 reported incidents of spills greater than 34 tonnes [1]. In 1999 more than 425 oil samples from spills along the Danish coastal line were analysed at the National Environmental Research Institute, Department of Environmental Chemistry, Denmark. The variability of spills is enormous ranging from the continuous
leaking of gasoline from gasstations to larger spills at sea as the Exxon Valdez spill amounting to more than 11 million gallons of crude oil [2]. Therefore, and especially in countries with large coastal lines like Denmark, the development of methods applicable in oilspill identification and degradation studies is an important issue.

For identification purposes GC-MS followed by visual comparison of GC-MS fingerprints is the analytical method most frequently applied, and together with GC/FID, it is included in the Nordtest spill identification system [3]. The feasibility of modern computerised GC-MS system provides the analyst with a massive amount of data, as petroleum consists of thousands of individual compounds. So far however multivariate decomposition methods have been applied only to a small extent as a tool in identification studies [4][5].

In biodegradation studies complex petroleum samples are rarely used; often degradation of model compounds like Pyrene [6], Flouranthene [7] and Benzo(a)pyrene [8] are examined and results extrapolated to degradation of complex mixtures. In this ongoing study a combined analytical, computerised and statistical method has been developed. A short review of the general principles and some examples of its applicability in petroleum analysis are given below.

**Theory**

The combined method is illustrated in Figure 1 and consists of three main parts; chemical analysis (extraction and detection), preliminary computerised treatment of data and statistical data evaluation. Samples are extracted with organic solvents, using methods depending on the sample matrix. Then extracts are analysed by GC-MS.

![Diagram](image)

**Figure 1:** Diagram of the developed hyphenated analytical and chemometric method. Samples are analysed using GC-MS (SIM), between 100 and 500 peaks are integrated applying a fast and improved integration method for complex mixtures [9]. Data are then analysed using multivariate techniques both during oilspill identification and biodegradation assessment.
The mass spectrometer applied in these studies is a double focusing high-resolution VG 70-250S (VG Analytical Ltd., Manchester, UK) sector instrument with on-column injection. Any high resolution mass spectrometer could have been applied.

GC-MS peaks are usually integrated using data acquisition software associated with the mass spectrophotometer. When analysing a vast number of compounds, peak integration can be a time consuming procedure encumbered with major uncertainties. The computerised peak integration procedure developed in this study lowers the analytical time considerably and simultaneously results in a more precise peak integration especially in the case of poorly separated peaks [9]. The computerised method assumes that GC-MS peaks can be described by gaussian distributions. Ion-fragmentograms are numerically differentiated and parameters, needed in a complete description of the gaussian distributions, e.g. $\chi_0$, $\delta$ and $A$, are approximated and used to develop the gaussian peak approximations. Only the uppermost part of a GC-MS peak is utilised, which is the main reason why the description of poorly separated peaks is improved by applying the method. Figure 2 illustrates the high quality approximations obtained by applying the computerised method.

It should be noticed that the only difference separating the approximations from the real peaks is that the spectral background has been removed during the preliminary computerised treatment.

Figure 2: A section of ion-fragment 183.2107; linear alkanes $C_{18} - C_{22}$ and the isoprenoids, phytane and IC21. The punctured line indicates the real GC-MS peaks, while the approximated gaussian distributions of the chosen peaks are drawn in solid.
In this example the peaks are well separated and it is obvious that the gaussian distribution result in an excellent description of GC-MS peaks. At higher retention times increasing peak tailing is often observed, but even at the largest retention times the tailing is of only minor importance, in relation to the uncertainties attached to GC-MS analysis in general [9].

The approximated gaussian peaks are then numerically integrated and the integrals are transferred to a multivariate software, Unscrambler 7.01. Principal component analysis is performed using group compound ratios, diagnostic ratios or normalised compound integrals. The specific input variables employed and the complete multivariate procedure, in the case of identification and biodegradation will be described in future reports.

Results and Discussion

Oilspill identification

When developing a multivariate model for use in either oilspill identification or biodegradation assessment it is very important that the model is robust. To create a robust model two conditions are of principal importance: (1) all variables must be normally distributed and (2) variations between samples must be uniformly distributed, meaning that no sample must be located in an outlier position in the system of co-ordinates spanned by the principal components. A multivariate model consisting of 23 crude oils has been developed, with the specific purpose of identifying a North Sea Crude to its source even though it has been artificial weathered for 14 days in a laboratory sediment/water experiment. The variables applied in the model consist of 19 normalised integrals of tricyclic and pentacyclic triterpanes (hopanes). These compounds belong to the group of petroleum biomarkers and have often been applied in oilspill identification studies because of their high resistancy towards weathering and biodegradation. Models applying compound group ratios and specific diagnostics ratios as variables have also been developed in this study, and the following PCA model is just an example of the applicability of multivariate analysis as a statistic tool in oilspill identification.

The developed model consists of an optimum of five principal components where PC1, PC2 and PC3 together describe 86 % of the variation in the x-matrix. A plot of PC1 vs PC3 (Scores) are shown in figure 3. The punctured circles describe the main groupings in the crude oils used in the model. Five distinct groupings (expected) related to the geographical locations are observed; the North Sea, Nigeria, Venezuela, Middle East and the North African region. Samples lying close in the score plot are similar while samples located with increasing distance contain larger dissimilarities. Oils encircled into the same area therefore have a great similarity in its biomarker distribution.

Identification of another group, encircled by a solid line, located in the area describing the North Sea crudes is easy. Here three samples NS1, NS2 and NS3 are identified to its source North Sea Crude even though they have been weathered for 14 days in a sediment/water laboratory experiment [9]. The
differences observed in the score plot are in all three cases ascribed to experimental uncertainties (GC-MS detection). Fully applicable collections of multivariate models applied in oilspill identification are summarised in [9].

Knowledge concerning which original variables are accountable for similarities and dissimilarities between samples can be obtained from a loading plot. Loadings describe the cohesion between the x-variables and the principal components, and are defined as cosine to the angles between the original descriptors (variables) and the respective principal components [10]. Figure 4 shows loadings describing the model developed in this section.

Figure 3: Score plot of a multivariate oilspill model (PC1 vs PC3) consisting of 23 crude oils described by 19 normalised integrals of individual tricyclic and pentacyclic triterpanes.

By examination of the loading plot it is obvious that the main descriptive power along PC1, describing 58% of the variation in x, is effected by the relative amount of tricyclic vs pentacyclic triterpanes. A large part of the observed separation along the third principal axis is caused by \( T_s/T_m, C_{29}\) and \( C_{30}\) pentacyclic triterpane. This is consistent with the fact that the ratios \( T_s/T_m \) and \( C_{30}/C_{29} \) have been applied in the literature as useful diagnostic ratios describing maturity. When only 19 variables are applied for identification purposes as in this example relatively large changes usually occur when one compound is
added to or removed from the variable matrix. The uncertainties diminish when more ratios or isomeric patterns are introduced to the variable matrix, resulting in a model with higher predictive powers [9].

![Diagram](image)

**Figure 4:** Loadings describing which original has the largest influence on PC1 and PC3. Tricyclic and Pentacyclic triterpanes with high loadings on the first principal axis are encircled.

**Biodegradation**

Numerous studies have investigated the fate of hydrocarbons in marine sediments, especially effects of biodegradation. Most studies have concerned the microbial degradation of model compounds under laboratory conditions by well-suited genotypes [6][7][8]. Wang [11] quantified more than 100 individual target aliphatic, aromatic and biomarker compounds. The study revealed distinct changes in relative compound concentrations due to microbial degradation by six microbial strains (three aliphatic and three aromatic degraders. The degradation by natural bacterial communities of several hundred compounds in a complex mixture was examined in the present study, but so far only few results have been obtained, and multivariate analysis has therefore not yet been applied. Figure 5 shows one of the most promising results so far. A small sediment/water experiment using the North Sea Crude also applied in the Oilspill model resulted
in isomeric degradation of alkyltoluenes already after 3 weeks of incubation of oxic sediment. The ion-fragment 106.0777 (alkyltoluenes) indicate that the bacterial community clearly prefers the meta-isomer of toluene’s relative to the ortho-isomer.

Figure 5: Changes of the isomeric patterns during three weeks of microbial biodegradation. Solid lines indicate partly degraded compounds (m-alkyl-toluenes), while the punctured arrows indicates more persistent compounds.

Even though multivariate analyses have not yet been applied in biodegradation studies information about its applicability is still sought. Figure 6 shows a theoretical multivariate model describing three of the most significant weathering/degradation processes known; evaporation, photooxidation and biodegradation. The normalised isomeric patterns of linear alkanes, alkyltoluenes and some PAHs are applied as variables in the model. It must be
emphasised that the model is purely hypothetical but it illustrates some general principles, when applying multivariate models in weathering and degradation studies of complex compound mixtures. In the Score plot the non-weathered/degraded North Sea Crude is shown by a filled circle.

Figure 6: A Score plot of the hypothetical multivariate model. The filled circle denote non-weathered/degraded North Sea Crude while open circles, crosses and triangles, respectively, denote samples from the three hypothetical weathering/degradation processes. The letters A-E denote the time series, e.g. increasing evaporation. The Loading plot shows which compounds change during the different processes described.
The expected effects of application of one of the three hypothetical treatments are calculated and shown on the score plot. The loading plot indicates those compounds having the greatest effect on the time dependent changes. Light sensitive PAHs have the greatest effect on the direction of changes during photooxidation. Light hydrocarbons have effect on evaporation while m-alkyl-toluenes have the largest effect in the biodegradation scenario.

Conclusion

Application of the developed hyphenated method for petroleum spilled in the environment results in a considerable lowering of analytical time, which enables more comprehensive investigations to be performed both in identification and biodegradation studies. The use of a fast and improved integration method gives more precise integration of each peak especially in the case of poorly separated peaks. When applied to a real oil-spill case the multivariate analysis can result in distinct differentiation of the analysed oils and offers a statistical evaluation of the identity of the spill, even though numerous and highly complex data are analysed. Furthermore these techniques can in weathering/degradation studies provide a tool for evaluation of effects on complex matrixes, e.g. petroleum samples. Determination of the importance of different variables, e.g. isomeric patterns, compound group- or diagnostic ratios during oilspill identifications or weathering/degradation studies are of high interest especially in the latter.

References


