Fuel Processing Technology 106 (2013) 181–185

Contents lists available at SciVerse ScienceDirect

Fuel Processing Technology

journal homepage: www.elsevier.com/locate/fuproc

Classification of crude oil samples through statistical analysis of APPI FTICR mass spectra

Stefano Chiaberge, Tiziana Fiorani, Alberto Savoini, Anna Bionda, Stefano Ramello, Monica Pastori, Pietro Cesti *

Centro Ricerche per le energie non convenzionali, Istituto Eni Donegani, Via Fauser, 4, 28100 Novara, Italy

A R T I C L E I N F O

Article history:
Received 23 April 2012
Received in revised form 10 July 2012
Accepted 20 July 2012
Available online 11 August 2012

Keywords:
Crude oil
Photoionization
FTICR mass spectrometry
Principal component analysis
Hierarchical clustering analysis
Geochemistry

A B S T R A C T

Characterization and classification of crude oil is done following different approaches. Geochemistry investigations on one hand are focused on finding and quantifying markers characteristic of the oil, to provide detailed information about the geological history, the conditions of sedimentation of the oil and of the source rock. On the other hand Fourier transform ion cyclotron mass spectrometry (FTICR MS) has been proved to be a powerful tool in the analysis of crude oil at the molecular level. In this study, a group of fourteen samples of crude oil coming from different fields around the world has been analyzed through Atmospheric Pressure Photoionization (APPI) coupled to FTICR MS. Since the comparison of the many complex APPI FTICR mass spectra is complicated, we have applied statistical methods such as Principal Component Analysis (PCA) and hierarchical clustering (HCA) considering the peaks of the mass spectra. The statistical analysis has found to be able to group the oils according to the well where they have been extracted and from their geographical origin.

1. Introduction

Crude oil geochemical characterization is mainly addressed to obtain information about the source rock from which the oil was generated. Petroleum geochemical knowledge, in fact, provides a wide variety of useful information such as maturity of the sediment, hydrocarbon accumulations, structure and connectivity of the reservoir [1,2]. Analytical investigations are commonly aimed at finding and quantifying oil specific markers to provide detailed information about the oil and its source rock, such as their geological history, conditions of sedimentation (diagenesis), and age [3]. The relative quantities of these markers have been correlated to specific properties of the reservoir and have been successfully used to characterize different types of oil [4].

The advent of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) offers a new tool for the characterization of petroleum samples at the molecular level [5–7].

The correlation of this analytical approach together with others techniques such as Infrared Spectroscopy (FTIR) and nuclear magnetic resonance (NMR) and the properties of oil and its derivatives is the main subject of the field known as Petroleomics [8]. Different ion sources in direct infusion mode like Electrospray Ionization (ESI) [9] and more recently [10] Atmospheric Pressure Photoionization (APPI) [11] coupled with FTICR MS have been applied to make a molecular characterization of crude oil and its fractions such as deasphalted oils and Asphaltenes [12]. Positive ion mode APPI in particular has been successfully applied for the determination of polycyclic aromatic sulfur heterocyclic compounds (PASHs) in the oil [13,14] and has found to be very efficient (thousands of peaks per mass spectrum) especially when dealing with heavy crude oils rich in sulfur.

The ultrahigh mass resolution together with the mass accuracy allows to determine the molecular formula of thousands of components per mass spectrum [15]. Given the high complexity of FTICR MS spectra of petroleum, data elaboration is a key aspect in the upgrading of the technique for different applications in oil research. A single APPI-FTICR MS mass spectrum of crude oil in positive ion mode for example, routinely contains about 5000–15,000 peaks with a relative abundance higher than the instrumental noise threshold. It has been reported [16] that a single FTICR MS mass spectrum contained as many as 50000 peaks. The common criteria of data elaboration is to group the molecular formula according to the heteroatom content [9] and Kendrick mass [17] has been applied to identify class of homologues. In order to compare oil mass spectra Kendrick mass Defect (KMD) and Van Krevelen plots have been used [18]. In KMD plots the molecular formulas identified in the mass spectra are grouped according to the value of KMD and the nominal mass, and compounds whose compositions differs by specific masses related to structural units such as CH2, CH2O, COOH, can be identified on the two dimension as straight or diagonal lines [19]. In the Van Krevelen plots the ratio between the elements in the molecular formulas and the relative abundances is displayed. In this way the difference of composition between samples of oil can be visualized. This approach has been

---

* Corresponding author.
E-mail address: Pietro.Cesti@eni.it (P. Cesti).

0378-3820/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fuproc.2012.07.023

proved to be very useful in crude oil characterization, however the extraction of information out of such complex mass spectra, especially when dealing with the comparison of a high number of samples, remains complicated.

In order to compare different samples of oil, statistical analysis has recently been applied to high-resolution APPI FTICR MS mass spectra [20,21].

The present work describes the use of Principal Component Analysis (PCA) and hierarchical clustering (HCA) to group 14 samples of crude oil coming from different wells considering the APPI-FTICR MS mass spectra in positive ion mode. Each peak of the mass spectra has been considered for PCA and HCA. The statistical treatment of the FTICR mass spectra have been found to be able to group the samples of crude oil according to the geographical origin and to the specific well from which the oils have been extracted.

Statistical results were found to be similar to those obtained considering a geochemical marker on the same group of crude oils [22].

2. Materials and methods

2.1. APPI-FTICR MS

Fourteen samples of crude oils have been considered in the present study. The samples have been diluted to 0.5 mg/mL with a 50:50 v/v solution of toluene/methanol. Toluene acts both as a solvent and as dopant in the ionization process [11]. Toluene and methanol HPLC grade were purchased from Carlo Erba. The solutions were subsequently directly injected with a flow rate of 50 ul/min on a 7 T FTICR Mass spectrometer (LTQ-FT Ultra Thermo Scientific). The ion source is equipped with a vacuum ultraviolet gas (Kripton) discharge lamp. Nitrogen was used as nebulizing gas. Mass spectra have been obtained in positive mode with a mass range of 100–1000 m/z and with an average resolving power of 400,000 at m/z 400. Spectra have been recorded in profile mode, microscan 1, max injection time 1000, and automatic gain control (AGC) on the ICR cell was set to 10 [6]. A minimum of 100 scan were collected and averaged for each analysis to improve the signal to noise ratio. An average of 90% of the peaks with magnitude above a threshold of 3σ (standard deviation) of baseline noise have been assigned a molecular formula and considered in the data evaluation.

Data have been processed by the software Xcalibur (Thermo Fisher Scientific Company) after the selection of some restrictions of the range of elements: 10–60 C [12], 0–2 C [13], 10–100 H, 0–6 N, 0-6 ^{32}S, 0-1 ^{34}S, 0-6 O and the error range was set to 2.5 ppm. These restrictions are required because of the great number of possible different combinations of elements that can be generated from a single accurate mass. First step of data evaluation has been done below 400 Da. At low masses the molecular formulas assignment is more reliable due to the lower number of possible combination for a single mass. The lists of masses and molecular formulas were subsequently elaborated and distinguished with custom built software (ISOMASS) [12]. Afterward masses with relative intensities lower than 0.5% has been cut out. Each sample has been analyzed in triplicate for statistical significance, and the consequent list of masses, with relative molecular formulas and their relative abundances, has been first of all averaged and then unified in order to have a single matrix of data, which has been then used in the PCA and HC statistical analysis.

2.2. Statistical data treatment

To classify the oil samples after APPI FTICR MS analysis, hierarchical clustering (HCA) and principal data analysis (PCA) were performed on the experimental results. HCA applies a number of different methods for assigning a group(clusters) to a set of objects of similar kind. Detailed information about Cluster Analysis can be found in the literature [23]. In our study we used the Euclidean distance and the unweighted pair group method with arithmetic mean, also known as average linkage clustering. The Euclidean distance is the geometric distance in the multidimensional space:

$$\left(\sum_{i=1}^{N}(x_i-y_i)^2\right)^{1/2}$$ where x_i (or y_i) is the coordinate of x (or y) in dimension i.

while unweighted pair group method with arithmetic mean is a simple agglomerative method where the distance between two clusters (A and B) is calculated as the average distance between all pairs of objects in the two different clusters:

$$\frac{1}{|\mathcal{A}|\cdot|\mathcal{B}|}\sum_{x\in\mathcal{A}}\sum_{y\in\mathcal{B}}d(x,y)$$

The description of PCA method can be found in the literature [24]. Briefly, PCA reduces the number of variables through a linear combination of the original variables.

The relative abundances of the attributed peaks of each of the 14 samples analyzed were selected as the variables. HCA has been carried out considering all the peaks and their relative abundances and in a further elaboration considering only the molecular formulas common to all the oil samples and then considering those containing sulfur (1-3 sulfur atoms). PCA has been carried out only on the common peaks and on the corresponding sulfur containing molecular formulas.

Exploratory data analysis was applied to the data matrix constituted of the 14 samples (as rows) and this group of variables (as columns). Statistica version 9 (StatSoft, www.statsoft.com, USA) was used to perform both HCA and PCA.

3. Results and discussion

The fourteen petroleum samples considered in the present paper have different origin. The first 12 samples come from Italian oil fields. The Italian oils A1,A2,A3,A4 and C1,C2,C3,C4 come from the same two wells, respectively from South East Italy and North Italy but they have been extracted in different years. Sample B1 comes from the same geographical area of the A series but from different well. The samples of the South West (SW) Italy series are heavy oils rich in sulfur, at about 6% in weight, and come from the same geographical area, but from different wells (D,E,F). The Asian oil samples (H1,K1) have been extracted from different wells and different geographical area. The origin, the API gravity, the sulfur content, the total number of assigned molecular formula for each APPI FTICR mass spectrum for the fourteen samples of crude oil is reported in Table 1. Noticeably the number of assigned molecular formulas is relative to the protonated molecular ions formed during the ionization process without considering the isotopic contributes of 13 C, 34 S, 18O and 15 N. That is the reason why the number is much lower

Table 1
Crude oil properties.

Origin of crude oil	Sample name	Well	API gravity	Sulphur %	no. of molecular formulas APPI-FTMS
SE Italy	A1	A	36.1	2.1	7031
SE Italy	A2	A	34.0	2.1	7050
SE Italy	A3	A	35.2	2.1	7045
SE Italy	A4	A	35.8	2.2	7441
SE Italy	B1	B	13.1	7.1	7616
North Italy	C1	C	40.5	0.2	5058
North Italy	C2	C	39.8	0.2	4731
North Italy	C3	C	41.2	0.2	5321
North Italy	C4	C	39.6	0.2	5787
SW Italy	D1	D	32.1	6.6	6434
SW Italy	E1	E	33.4	6.1	6604
SW Italy	F1	F	32.7	6.3	6566
Asia	K1	K	36.4	1.2	6488
Asia	L1	L	24.2	3.8	6566