PLS regression on spectroscopic data for the prediction of crude oil quality: API gravity and aliphatic/aromatic ratio

Ouissam Abbas [a], Catherine Rebufa [b,*], Nathalie Dupuy [b], Albert Permanyer [c], Jacky Kister [b]

[a] Quality of Agricultural Products Department, Walloon Agricultural Research Centre, 24 Chaussée de Namur, 5030 Gembloux, Belgium
[b] LISA-E.A. 4672, Equipe METICA, Aix Marseille Université - Pôle de l'étoile, Case 451, 13397 Marseille Cedex 20, France
[c] Department de Geoquímica, Petrologia i Prospecció Geològica, Universitat de Barcelona, Martí i Franquès s/n, 08028 Barcelona, Spain

A B S T R A C T

This work describes a chemometric approach for predicting quality parameters of crude oils by using the information present in spectroscopic data as Fourier Transform Infrared–Attenuated Total Reflectance (FTIR–ATR) absorption and Synchrounous Ultra Violet Fluorescence (SUVF). Using multivariate analysis such as Partial Least-Square (PLS) analysis, the predictive ability of spectroscopic techniques has been explored to estimate the American Petroleum Industry (API) gravity usually determined using standard physical methods and infrared structural/functional indices characterizing the repartition of aliphatic and aromatic structures present in crude oils. Giving global information on chemical compounds present in oil, FTIR–ATR also appears to be a rapid analytical method for quantifying changes in abundances of aliphatic and aromatic structures with the help of the infrared indices calculated from area ratio of specific bands. Then, a PLS model based on MIR data allows to predict this aliphatic/aromatic ratio for various crude oils and avoid time-consuming step of infrared peaks integration.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Crude oils are complex mixtures of saturated and aromatic hydrocarbons and NSO (Nitrogen, Sulfur, Oxygen) compounds as also named resins and asphaltenes. There is considerable variation in the color, gravity, odour, sulfur content and viscosity of oils from different areas. Oils produced in different regions of the world have compositional characteristics depending essentially on source rock type, maturity and post generation phenomena (e.g. biodegradation). Physical properties such as specific gravity and viscosity have an important impact on field development decisions and the economic value of petroleum. In recent years, there have been successful applications of spectroscopy concentrated on the determination of the quality of petroleum products. To extract rapidly the spectral information, multivariate analysis were often used for relating the spectra to properties of interest. It was demonstrated that the quality of crude petroleum and its derivatives, as assessed from the API gravity (usually obtained from standard procedures established by the American Petroleum Institute [1,2]), could be predicted directly from the molecular composition of crude oils. Pasquini and Bueno [3] used Near Infrared (NIR) absorption spectroscopy to predict API gravity (extended from 13.2° to 49.6° with a standard deviation of 1,63° determined only for some mixtures

of crude oils) in order to characterize 122 petroleum samples mainly produced in various regions in Brazil. A specific NIR spectral zone (5000–3900 cm^{-1}, interval referring to the combination bans of C–H bonds) was chosen to improve the performance of the developed Partial Least Square (PLS) model. For API prediction of 22 petroleum samples, their results showed that NIR technique presented repeatability similar to that obtained by reference method (procedure described in ASTM D 1298) with seven latent variables, a correlation coefficient (R^2) equal to 0.923 and RMSEP (Root Mean Square Error of Prediction) equal to 0.24 for the external validation of PLS model. Hidajat and Chong [4] have shown that density of crude oils (105 samples originating from more than 12 countries) can be characterized by NIR spectroscopy (that extend from 780 to 2500 nm) using PLS regression with a cross validation. Their developed NIR PLS model (10 latent variables, r^2 = 0.994, RMSEP = 0.0022 for a calibration range of density from 0.7980 to 0.9038) has shown a reproducibility limit equivalent to the ASTM procedure. Ryder [5] discussed how suitable Total Synchronous Fluorescence Scan (TSFS) 3-D spectra normalized at the point of maximum fluorescence emission could be used to evaluate the maturity of 18 crude oils. TSFS contour trend recorded from 250 to 700 nm over a wavelength interval of 40–200 nm using a front surface sampling accessory has allowed discrimination of low maturity oils (API > 40°) from other types. Heganzi and Hamda [6] estimated the grade of six crude oils from Saudi Arabia with Time-Resolved Fluorescence (TRF) spectra which showed the presence of a shoulder near 380 nm whose

---

intensity systematically decreased according to the grade of the oil (from 50.24° to 24.00°). Contour diagrams of TRF spectra have been elucidated by associating ratios of different areas under the curves in TRF spectra (355–390 nm, 432–480 nm and 480–545 nm) with API gravity in order to obtain conversion curves at different time gates (within the leading and trailing edges). Stasiuk and Snowdon [7] demonstrated a graphic correlation between fluorescence microscopy parameters (L_{max}, wavelength of maximum emission intensity and Q, ratio of intensities at 650 nm and 500 nm) and the density of crude oil entrapped as hydrocarbon fluid inclusions (API values covering a range between 19.0° and 44.3°). Only Q has been correlated positively with API gravity. Quantitative fluorescence investigations were reported by Hagemann and Hollerbach [8] who showed that there is a characteristic shift from green to blue from low to high maturity oils. Degraded oils have a maximum intensity in the yellow region.

Another way for assessing petroleum quality is to obtain a characterization of molecular average structures of crude oils. The classical approach is to measure the aromatic, aliphatic and polar contents determined from SARA fractionation [9], separating crude oil into Saturates, Aromatics, Resins and Asphaltenes (SARAs). Aske et al. [10] reported that Infrared (IR) spectroscopy seems to provide slightly better performance than Near Infrared (NIR) in predicting the saturate and aromatic fractions of crude oils already obtained using High-Performance Liquid Chromatography (HPLC). They concluded that standard error of PLS prediction model was comparable to the experimental error in the HPLC fractionation procedure: 2.5% and 2.2% for saturate and aromatic fractions respectively from IR data. Many studies have examined IR spectroscopic features of different types of fossil organic matter (coal, kerogen, macerals, oils…) in order to determine NIR or MIR derived ratios of the specific bands to predict oil quality before its production. Mullins et al. [11] showed that the ratio of Near Infrared (NIR) peak areas can be used to obtain a reliable estimate of the Gas Oil Ratio (GOR) used to classify crude oils in terms of black oils, volatile or condensates. About oil maturity, a large bibliography was produced by Lis et al. [12] in order to introduce their own FTIR aromatic/aliphatic absorption ratio which has a strong correlation with the vitrinite reflectance of kerogens (Type-II). They encouraged the use of FTIR-based indices as proxies for maturity. The advantage of the approach is that vitrinite is not available in sufficient quantity or quality in marine sediments.

In previous papers, we promoted these spectroscopic techniques because they contain data rich in information embedded in multiple absorption bands in spectra. Hence, spectral data treated using chemometric tools have allowed discrimination of oils from different origins, determination of reservoir compartments, assessment of oil biodegradation on the basis of the relative abundance of compound classes such as aliphatic, aromatic and polar species [13–16]. Structural and functional indices calculated from ratio of IR absorption bands areas were developed to characterize the chemical composition of crude oils. Theses relative indices were allowed to obtain an indirect measure reflecting the relative abundance of a particular functional group describing specific compounds families like aromatic, aliphatic or polar ones. It was well known that the relative abundance of saturated and aromatic hydrocarbons could be considered as maturity indicators influencing directly oils quality. Here, two of these indices were used to estimate oil quality from the knowledge of the repartition of aliphatic and aromatic structures by calculating the ratio between the aliphatic index and the aromatic one. The aliphatic index involved the absorption bands at 1460 cm^{-1} (asymmetric C–H deformation vibrations of CH$_3$ group and scissor vibration of –CH$_2$ group) and the one at 1370 cm^{-1} (assigned to symmetric C–H deformation vibrations of CH$_3$ group). Among the three spectral zones usually describing aromatic absorptions (3000–3100 cm^{-1}

for the aromatic C–H stretching, 1600 cm^{-1} for aromatic ring stretching and 700–900 cm^{-1} for aromatic C–H out-of-plane deformation bands), only one spectral zone was chosen around 1600 cm^{-1} because this region yielded the strongest correlation between aromaticity. The 700–900 cm^{-1} region was not considered because it was used to define substitution indices of aromatic rings [13]. The low absorption intensity in spectral zone at 3000–3100 cm^{-1} resulted in large errors. To avoid the integration step, using Fourier deconvolution in some cases because of aromatic absorption at 1600 cm^{-1} overlapped with carbonyl absorption at 1700 cm^{-1}, indices determination was possible by the use of multivariate technique like PLS to establish the relationship between FTIR data and functional indices.

This study discusses the predictive ability of PLS models based on FTIR–ATR and SUVF data to estimate API gravity, a key-factor in petroleum exploration and production. It also demonstrates the utility of FTIR ratios of absorption bands areas for rapid estimation of the abundance of aliphatic and aromatic compounds which greatly influences oil maturity. The aim of using spectroscopic techniques coupled with chemometric tools is to obtain rapidly, without preparation of sample, an estimated value of these quality factors with the sole knowledge of geographic origin of oils.

2. Methods and materials

2.1. Sample

Fifty-two crude oils from seven geographic locations (Algeria, Brazil, Venezuela, Egypt, the Equator, the Mediterranean Sea and the Pyrenees) were used. These crude oils exhibit wide variations in composition and properties because they were taken from different fields and from different production depths in the same well.

2.1.1. Algerian oils
Eleven samples, provided by the Cepsa-Sonatrach companies, came from two fields located at the east of the Berkine Basin. Oils (601 to 604 and 626–631) came from the field named Rhourde El Krouf (RKF). Only one sample (795) came from the Ourhoud field (ORD). Samples 626, 627, 628 and 631 are respectively the same oils that 601, 602, 603 and 604 but they were taken 1 year after values of API gravity between 44° and 45° indicate a high maturity for these oils.

2.1.2. Brazilian oils
Six oils, provided by Rio of Janeiro Federal University, were from two fields located on Brazilian marginal basins, whose names remained confidential. The API gravity is unknown. Three samples (619, 620 and 624) are considered as non biodegraded, on the basis of gas chromatography–mass spectrometry (GC–MS) data [17] by contrast to the biodegraded ones noted 621, 622, 623 and 625.

2.1.3. Venezuelan oils
Eleven oils (533–539) from the El Furrial field (Maturin Basin, East of Venezuela) have API gravity between 21° and 28°, suggesting a medium maturity.

2.1.4. Egyptian oils
Six oils (26, 27, 30, 146, 148 and 149) provided by the REPSOL company, are from Khalda permit in the Western Desert. The API gravity ranges between 31° and 41°.

2.1.5. Equator oils
Eleven samples (701–711) are from the Oriente Basin with API gravity between 22° and 32°. Samples were provided by Occidental Oil Company.