Introducing a Novel Approach for Oil-Oil Correlation based on Asphaltene Structure: X-ray Diffraction

Zahra SADEGHTABAGHI1, Ahmad Reza RABBANI1,*, and Abdolhossein HEMMATI-SARAPARDEH2,3,4

1 Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
2 Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
3 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
4 Faculty of Environmental and Chemical Engineering, Duy Tan University, Da Nang 550000, Vietnam

Abstract: Asphaltenes have always been an attractive subject for researchers. However, the application of this fraction of the geochemical field has only been studied in a limited way. In other words, despite many studies on asphaltene structure, the application of asphaltene structures in organic geochemistry has not so far been assessed. Oil-oil correlation is a well-known concept in geochemical studies and plays a vital role in basin modeling and the reconstruction of the burial history of basin sediments, as well as accurate characterization of the relevant petroleum system. This study aims to propose the X-ray diffraction (XRD) technique as a novel method for oil-oil correlation and investigate its reliability and accuracy for different crude oils. To this end, 13 crude oil samples from the Iranian sector of the Persian Gulf region, which had previously been correlated by traditional geochemical tools such as biomarker ratios and isotope values, in four distinct genetic groups, were selected and their asphaltene fractions analyzed by two prevalent methods of XRD and Fourier-transform infrared spectroscopy (FTIR). For oil-oil correlation assessment, various cross-plots, as well as principal component analysis (PCA), were conducted, based on the structural parameters of the studied asphaltenes. The results indicate that asphaltene structural parameters can also be used for oil-oil correlation purposes, their results being completely in accord with the previous classifications. The average values of distance between saturated portions (d₁) and the distance between two aromatic layers (d₄) of asphaltene molecules belonging to the studied oil samples are 4.69Å and 3.54Å, respectively. Furthermore, the average diameter of the aromatic sheets (L₃), the height of the clusters (L₄), the number of carbons per aromatic unit (Cₐ), the number of aromatic rings per layer (Rₐ), the number of sheets in the cluster (M₄) and aromaticity (fₐ) values of these asphaltene samples are 10.09Å, 34.04Å, 17.42Å, 3.78Å, 10.61Å and 0.26Å, respectively. The results of XRD parameters indicate that plots of d₁ vs. d₄, d₄ vs. M₄, d₄ vs. fₐ, d₄ vs. L₄, L₄ vs. L₅, and fₐ vs. L₅ perform appropriately for distinguishing genetic groups. A comparison between XRD and FTIR results indicated that the XRD method is more accurate for this purpose. In addition, decision tree classification, one of the most efficacious approaches of machine learning, was employed for the geochemical groups of this study for the first time. This tree, which was constructed using XRD data, can distinguish genetic groups accurately and can also determine the characteristics of each geochemical group. In conclusion, the obtaining of structural parameters for asphaltene by the XRD technique is a novel, precise and inexpensive method, which can be deployed as a new approach for oil-oil correlation goals. The findings of this study can help in the prompt determination of genetic groups as a screening method and can also be useful for assessing oil samples affected by secondary processes.

Key words: oil-oil correlation, petroleum characterization, X-ray diffraction, Fourier-transform infrared spectroscopy, decision tree classification

1 Introduction

Oil-oil correlation is a widely-known concept in geochemical studies, defined as finding geochemical relevance between oil samples (Dembicki, 2016). There is a striking principle which indicates that every oil sample can preserve its chemical characteristic throughout the passage of time. This property, referred to as a fingerprint, can be used for distinguishing differences between one oil sample and others, or detecting similarities between oil samples of a genetic family (Christensen and Tomasi, 2007; Speight, 2014; Hashemi-Nasab and Parastar, 2019). One of the most significant applications of oil-oil correlation is determining the number of active source rocks in a given study area. This can be achieved by identifying existing genetically-distinct oil families in the basin (Dembicki, 2016). Oil-oil correlation also plays a vital role in basin modeling and reconstruction of the burial history of basin sediments (Soltani et al., 2019), as well as precise characterization of the petroleum system.
Determining the geochemical relationships among several oils, in other words oil-oil correlation, can specify the migratory paths from the individual source rock to more than two reservoirs (Dembicki, 2016). Identification of distinct compartments in reservoirs, as well as the reduction of exploration risk, can also be considered as benefits of oil-oil correlation (Peters, 2005; Wang et al., 2018).

The tools which are applied generally for this goal are biomarker ratios and isotopic composition (Galimov, 1973; Peters and Moldowan, 1993; Galimov and Rabbani, 2001). Most researchers have published their achievements based on these two aforementioned geochemical tools, which are measured in crude oil fractions or whole oil samples. Biomarker ratios and stable carbon isotope results are used extensively in oil-oil correlation studies (Telnaes and Dahl, 1986; Galimov and Rabbani, 2001; Rabbani, 2007, 2008; Kamali et al., 2013; Rabbani et al., 2014; Sourour et al., 2015; Rabbani and Baniasad, 2017; Alizadeh et al., 2018; El Diasty et al., 2019; Kobræi et al., 2019), whereas several authors have used trace elements (Hitchon and Filby, 1984; Ellrich et al., 1985; Filby, 1994; López et al., 1998; Akinlua et al., 2008, 2010, 2015; Abdula et al., 2018) and diamondoids (Wang et al., 2006; Baniasad et al., 2014) to achieve this objective.

It must be emphasized that some researchers have focused on the role of asphaltene in the broader scope of geochemical studies. For instance, pyrolysis of the asphaltene fraction has been used for maturity assessment (Cassani and Eginton, 1986; Nali et al., 2000). Several authors have utilized the surrounded hydrocarbon inside the asphaltene structure for geochemical objectives (Pan et al., 2002; Russell et al., 2004; Liao et al., 2006). Oil-oil correlation in such investigations has been done by using molecular weight (Escobar et al., 2007), kinetic parameters (Soltani et al., 2019) and FTIR spectra of asphaltenes (Asemani and Rabbani, 2016; Taherinezhad et al., 2019). The FTIR spectra of asphaltene samples have also been applied for similarity assessments of oil fingerprints (Asemani et al., 2020).

As mentioned, biomarker fingerprints play a significant role in oil-oil correlation studies, but their precision has sometimes been a matter of debate. It should be noted that alteration processes can lead to eliminating all or a part of the indicator biomarkers that can influence geochemical interpretations (Dembicki, 2016; Soltani et al., 2019). For instance, severe thermal alteration can partly or completely change the geochemical characteristics of oil samples, different degrees of biodegradation also having the ability to affect biomarker distributions. The most problematic consideration lies in situations where alterations have occurred, but without being obvious, thus rendering all results spurious (Dembicki, 2016). In some cases, it has been known for geochemists that biomarkers are associated with a series of inconsistencies and even mistakes. Using another distinctive method seems necessary in order to make accurate decisions, the most appropriate approach in such circumstances as these being the application of further methods in order to reduce the level of possible errors.

Asphaltene has always been considered a problematic issue in the oil industry and several efforts have been made to assess precipitation behavior and treatment methods (Duran et al., 2019; Fakher and Imqam, 2019; Mohammadi et al., 2019; Enayat et al., 2020; Tazikeh et al., 2020). Although asphaltenes are a problem in the realm of oil production, they can be considered as the proper choice to apply for correlation purposes, because they are structurally and chemically similar to kerogen (Behar et al., 1984; Hall et al., 2014; Sultana and Bhasin, 2014). This similarity can help genetic studies (Behar et al., 1984) and, since this fraction of crude oil is very resistant, it can be preserved against probable variations. Thus, this special property can render asphaltenes a suitable geochemical indicator (Rubinstein et al., 1979). Inspite of the fact that biomarkers are affected by the imposed processes, asphaltene, as the heaviest fraction of crude oil, is usually unchanged. Thus, studying asphaltene can be very worthwhile in altered oils (Reyes et al., 2015) and as such they can be extremely useful and efficacious in oil-oil correlation (Reyes et al., 2015; Snowdon et al., 2016).

Several efforts have been made to obtain a comprehensive assessment of the asphaltene structure using the X-ray diffraction (XRD) method. XRD is a non-destructive method for the investigation of carbonaceous structures, which can be applied for the determination of asphaltene crystallite parameters and molecule stacks (Tuinstra and Koenig, 1970; Iwashita and Inagaki, 1993; Christopher et al., 1996). The first attempt belonged to Yen et al. (1961) who studied eight asphaltene samples structurally. Thereafter, many scientific manuscripts were published on asphaltene structure studies by XRD, or a combination of this technique and some others such as Nuclear Magnetic Resonance (NMR), Raman spectroscopy and Fourier-transform infrared spectroscopy (FTIR) (Schwager et al., 1983; Zhang et al., 1989; Christopher et al., 1996; Shirokoff et al., 1997; Siddiqui et al., 2002; Tanaka et al., 2004; Andersen et al., 2005; Bouhadda et al., 2007; Daou et al., 2009; Molina et al., 2016; Hemmati-Sarapardeh et al., 2018; Jiang et al., 2019; Nguyen et al., 2019; Ahooei et al., 2020). None of the aforementioned studies dealt with geochemical aspects.

The main goal of this current study is a novel application of asphaltenes as a geochemical tool for oil-oil correlation through application of the method of X-Ray Diffraction. The XRD method is a non-destructive, simple, quick and inexpensive approach, so can be considered as a useful and applicable method for genetic assessment, when applied to the asphaltene fraction as the fraction of crude oil most resistant to secondary processes and migration, as well as the most similar fraction of crude oil to kerogen. As already mentioned, previous studies on asphaltene structure obtained by the XRD method were all related to topics other than geochemistry and this study aims to show the application of asphaltene structure in this field of study for the first time. Conversely, previous geochemical research on asphaltenes have been limited to the FTIR and synchronous ultraviolet fluorescence (SUVF) methods (Permancy et al., 2002, 2005, 2007; Asemani and Rabbani, 2016), but this study assesses the
effectiveness of both XRD and FTIR methods, comparing their effectiveness. The use of a data mining method like a decision tree for genetic classification is also unprecedented.

In this study, 13 asphaltene samples from the Iranian section of the Persian Gulf region were selected and then analyzed by two popular techniques of XRD and FTIR for the determination of genetic groups. In addition to using cross-plots for this objective, a decision tree approach was applied for studying the genetic groups. This tree was modeled using the obtained XRD parameters of asphaltene. All results were verified by biomarker ratios and isotopic values. In the following section, materials and methods are presented which describe the oil samples and the two applied methods of XRD and FTIR. The results of the structural analysis are then discussed in two major parts of the study area in the results and discussion section, in addition to the statistical investigations. Finally, the achievements are noted in the summary and conclusions section. The general steps of this study are illustrated in Fig. 1. The nomenclature terms used in this paper could be found in Appendix A.

2 Materials and Methods

2.1 Oil samples
Thirteen oil samples from different reservoirs in the Kharg, Dorood, Foroozan, Salman, Balal, Bahregansar, Hendijan, Nowrooz, Resalat, Reshadat, SirriD and Nosrat fields, belonging to the Iranian sector of the Persian Gulf, were selected for this study. These oil samples had already been investigated geochemically, the complete results of their biomarker ratios, carbon isotopes and genetic classification being published in 2014 (Rabbani et al., 2014). The oil samples were identified in four main genetic groups. The Kharg, Dorood and Foroozan fields were classified in group I. These crude oils possessed high values of sulfur, as well as ratios of hopane (C29/C30) and tricyclic terpane (C22/C21). The second group consisted of Salman and Balal crude oil samples, commonly having high API and low sulfur content, as well as the highest values of C19/C23 tricyclic terpanes. The members of group III (Bahregansar, Hendijan and Nowrooz) possessed high values of sulfur and asphaltene contents. The lightest isotope carbon of crude oil samples also belonged to this group. Finally, group IV was the fourth genetic group in the Persian Gulf. Unlike group III, this group had the heaviest isotope carbon of the whole oil samples. Moreover, the lowest values of tricyclic terpanes (C19/C23) belonged to the members of this group (Resalat, Reshadat, SirriD and Nosrat) (Rabbani et al., 2014).

Fig. 2 shows the location of the selected fields, a simple stratigraphic column being shown in Fig. 3. Due to the geographical extent and geological complexity of the Persian Gulf area, the stratigraphic column of this aforementioned area possesses more detail. It must be noted that since this paper does not pursue geological objectives, the presented column is merely included for the purpose of introducing the salient formations. For more information regarding geological settings and basin conditions, the reader is referred to Rabbani et al. (2014). In this study, the asphaltene fractions of these aforementioned samples were analyzed using XRD and FTIR methods.

2.2 Asphaltene precipitation
Asphaltene fractions of oil samples {Kharg (Manifa reservoir), Dorood (Manifa and Yamamma reservoirs), Foroozan (Burghan reservoir), Salman (Buwaib reservoir), Balal (Khatiah reservoir), Bahregansar (Ghar reservoir), Hendijan (Sarvak reservoir), Nowrooz (Nahreumr reservoir), Resalat (Mishrif reservoir), Reshadat (Mishrif reservoir), SirriD (Mishrif reservoir) and Nosrat (Mishrif reservoir)} were precipitated by using n-hexane solvent through the IP/143 procedure (ASTM International, 2012). Two samples of Dorood (Manifa reservoir) and Reshadat (Mishrif reservoir) were twice precipitated and analyzed for a repeatability assessment. The results of the asphaltene percentages are summarized in Table 1.

2.3 X-ray diffraction
XRD analysis provides inherent high quality information about the macrostructure of asphaltenes. The dimensions of asphaltene units or crystallite parameters can be determined through this technique, whereas other methods such as elemental analysis only provide the chemical composition of the asphaltenes and do not present any insight into cell dimensions (AlHumaidan et al., 2015; Hemmati-Sarapardeh et al., 2018). The XRD
technique determines distances between saturated chains, as well as between aromatic sheets. Also, the size of the asphaltene aggregates can be defined using this technique. Other valuable parameters, such as the diameter of the aromatic sheets and their numbers in aggregate, as well as the number of sheets in the asphaltene clusters, are examples of other results of asphaltene XRD analysis. Moreover, the value of aromaticity can be characterized by XRD analysis of asphaltene (Okolo et al., 2015; Jiang et al., 2019).

The aforementioned parameters are deduced from the XRD pattern by their Bragg and Scherrer equations (Okolo et al., 2015). A hypothetical asphaltene pattern, as well as an asphaltene structure, is shown in Fig. 4. As can be seen, the XRD spectrum of asphaltenes consists of three main detected peaks, named the g-band (approximately $2\theta = 19^\circ$), the 002-band (approximately $2\theta = 25^\circ$) and the 10-band (approximately $2\theta = 45^\circ$) (Fig. 4a).

The g-band, 002-band and 10-band are related to saturated portion distances, aromatic sheet distances and aromatic unit sizes, respectively (Yen et al., 1961). The position, intensity and peak area are the three principal factors in structural parameter calculations. The distance between aromatic sheets ($d_m$) and the distance between saturated rings or chains ($d_i$) are calculated by Bragg’s law as follows (Fig. 4b):

$$d_m = \frac{\lambda}{2\sin \theta_{002}}$$  \hspace{1cm} (1)

$$d_i = \frac{\lambda}{2\sin \theta}$$  \hspace{1cm} (2)

where $\lambda$ is the wavelength of applied radiation and $\theta$ is the Bragg angle. The average diameter of the aromatic sheets in the asphaltene stack ($L_a$) and the average height of an
Table 1 Asphalten e fraction percentages of the studied crude oil samples

<table>
<thead>
<tr>
<th>Field</th>
<th>Reservoir</th>
<th>Asphaltene (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kharg</td>
<td>Manifa</td>
<td>3.38</td>
</tr>
<tr>
<td>Dorood</td>
<td>Yamam ma</td>
<td>5.15</td>
</tr>
<tr>
<td>Dorood*</td>
<td>Yamam ma*</td>
<td>5.17</td>
</tr>
<tr>
<td>Foroozan</td>
<td>Burgh an</td>
<td>2.81</td>
</tr>
<tr>
<td>Dorood</td>
<td>Manifa</td>
<td>6.04</td>
</tr>
<tr>
<td>Salman</td>
<td>Buwaib</td>
<td>0.83</td>
</tr>
<tr>
<td>Balal</td>
<td>Khatia</td>
<td>6.18</td>
</tr>
<tr>
<td>Bahregansar</td>
<td>Ghar</td>
<td>1.98</td>
</tr>
<tr>
<td>Hendjian</td>
<td>Sarvak</td>
<td>8.6</td>
</tr>
<tr>
<td>Nowrooz</td>
<td>Nahramr</td>
<td>10.31</td>
</tr>
<tr>
<td>Resalat</td>
<td>Mishrif</td>
<td>3.81</td>
</tr>
<tr>
<td>Reshadat</td>
<td>Mishrif</td>
<td>4.46</td>
</tr>
<tr>
<td>Reshadat*</td>
<td>Mishrif*</td>
<td>4.30</td>
</tr>
<tr>
<td>Sirri D</td>
<td>Mishrif</td>
<td>3.51</td>
</tr>
<tr>
<td>Nosrat</td>
<td>Mishrif</td>
<td>2.35</td>
</tr>
</tbody>
</table>

Note that samples marked (*) are those selected for repeatability assessment.

Fig. 4. The positions of the three main peaks in the XRD pattern of the asphalten e samples (a), as well as a hypothetical asphalten e structure (b).

asphalten e cluster, which is made up of several sheets lying perpendicular to their plane ($L_c$) can be determined through the following equations (Fig. 4b):

$$L_s = \frac{1.84\lambda}{\omega \cos \theta_{002}}$$ (3)

$$L_c = \frac{0.9\lambda}{\omega \cos \theta_{002}}$$ (4)

where $\omega$ is the full width at half the maximum of the related band. Also, the average number of carbon atoms per aromatic structure ($C_{au}$), the average number of aromatic rings per sheet ($R_a$) and the average number of aromatic sheets in an asphalten e cluster ($M_e$) can be obtained via equations (5) to (7):

$$C_{as} = \frac{L_s + 1.23}{0.65}$$ (5)

$$R_a = \frac{L_c}{2.66\lambda}$$ (6)

$$M_e = \frac{L_c}{d_m} + 1$$ (7)

Finally, the aromaticity value, which is defined as the proportion of aromatic units to aliphatic portions, can be calculated with the following equation:

$$f_a = \frac{A_{002}}{A_{002} + A_{\gamma}}$$ (8)

where $A$ is the area under the relevant peak (Yen et al., 1961; Trejo et al., 2007; Liu and Li, 2015).

In this study, the Inel diffractometer (EQUINOX 3000) with Cu Kα radiation was applied for XRD analysis at room temperature on powdered samples with a resolution of 2$\theta = 0.03^\circ$.

2.4 Fourier-transform infrared spectroscopy

FTIR is another appropriate tool for asphalten e investigation. Although this technique is unable to assist in structural parameter determination, it can provide significant information about the functional groups present (Asemani and Rabbani, 2020), the efficiency of FTIR in characterizing geochemical genetic groups already having been proven (Asemani and Rabbani, 2016). This technique can determine the values of the contributions of aromatic and aliphatic compounds, in the form of aromaticity and aliphaticity indices. Moreover, branched carbon chains, long carbon chains (carbon chains with more than four carbon atoms), carbonyl (C=O bonds), sulfoxide (S=O bonds), substitution 1 (one shared hydrogen atom present in the aromatic structure) and substitution 2 (two or three shared hydrogen atoms present in the aromatic structure) indices, are the other factors that can be calculated through analysis of asphalten e spectra using FTIR. These values are determined by peak area, the relevant equations being presented in Appendix B.

In this study, Thermo FTIR was used for obtaining asphalten e spectra in the range of 400 cm$^{-1}$ to 4000 cm$^{-1}$ with a 2 cm$^{-1}$ resolution. In order for this to be achieved, asphalten e samples were mingled with KBr powder to convert them into transparent tablets.

3 Results and Discussion

According to geochemical and statistical studies, Rabbani et al. (2014) demonstrated that there are four principal genetic groups in the Iranian sector of the Persian Gulf’s range. They used 14 geochemical parameters, which were both dependent on the source rock and unaffected by secondary processes, for statistical assessment by hierarchical cluster analysis (HCA), which
classified 33 crude oil samples into four distinct genetic groups. Fig. 5 shows selected samples for this study from their classification, based on biomarkers and isotopic ratios. Several oil samples from the northwestern Persian Gulf (Kharg field Manifa reservoir, Dorood field Yamamma reservoir, Foroozan field Burghan reservoir and Dorood field Manifa reservoir) belong to group I. This group of oils has a carbonate source rock, deposited in anoxic conditions, with an estimated age of Jurassic or older. The other group in this area is group III. The oil samples of this family (Bahregansar field Ghar reservoir, Hendijan field Sarvak reservoir and Nowrooz field Nahreumr reservoir) have originated from a calcareous shale source rock in the middle Cretaceous. Also, two other genetic groups are active in the southeast. Group II includes oil samples from a carbonate source rock in the Jurassic or older (Salman field Buwaib reservoir and Balal field Khatiah reservoir), the environmental condition being from a Cretaceous source rock (Rabbani et al., 2013; Molina et al., 2016; Poveda et al., 2009; Yasar et al., 2009; Yi et al., 2009; Hosseinpour et al., 2007; Soleimani Nazar and Bayandori, 2008; Daaou et al., 2005; Michael et al., 2005; Trejo et al., 2000, 2007; Bansal et al., 2004; Tanaka et al., 2004; Andersen et al., 2005; Michael et al., 2005; Trejo et al., 2007; Soleimani Nazar and Bayandori, 2008; Daou et al., 2009; Yasar et al., 2009; Yi et al., 2009; Hosseinpour et al., 2013; Molina et al., 2016; Poveda-Jaramillo et al., 2014).

In this study, several oil samples from each family were selected, in order to introduce a novel approach for oil-oil correlation. The selected samples are shown in Fig. 5. This study attempts to assess whether the results of this new method (XRD) are in accordance with previous results or not. For this purpose, Kharg (Manifa), Dorood (Yamamma), Foroozan (Burghan) and Dorood (Manifa) from group I, Salman (Buwaib) and Balal (Khatiah) from group II, Bahregansar (Ghar), Hendijan (Sarvak) and Nowrooz (Nahreumr) from group III, and ultimately Resalat (Mishrif), Reshadat (Mishrif), Sirrid D (Mishrif) and Nosrat (Mishrif) from group IV were chosen (Fig. 2).

### 3.1 XRD and FTIR results

The asphaltene fractions from these samples were precipitated and then analyzed by XRD, the resulting patterns being shown in Fig. 6. Afterwards, the structural parameters were calculated, based on equations 1-8, the results of the structural parameters being presented in Table 2. In this table, the values of the average distance between aromatic sheets (d_{as}), the average distance between saturated portions (d_{s}), the average diameter of aromatic sheets in the asphaltene stack (L_{a}), the average height of an asphaltene cluster perpendicular to the plane of the sheets (L_{c}), the average number of carbon atoms per aromatic structure (C_{au}), the average number of aromatic rings per sheet (R_{a}), the average number of aromatic sheets in an asphaltene cluster (M_{a}) and aromaticity (f_{a}) are listed for all 13 asphaltene samples. In addition, both repeated analyses of Dorood (Yamamma) and Reshadat (Mishrif) are presented in this table. The ranges of d_{s}, d_{as}, L_{c}, L_{a}, C_{au}, R_{a}, M_{a} and f_{a} are equal to 4.69Å, 3.54Å, 10.09Å, 34.04Å, 17.42Å, 3.78Å, 10.61Å and 0.26Å, respectively. All these values are in the ranges of other studied asphaltene fractions (Yen et al., 1961; Asaoka et al., 1983; Sadeghi et al., 1986; Shirokoff et al., 1997; Bouhadda et al., 2000, 2007; Bansal et al., 2004; Tanaka et al., 2004; Andersen et al., 2005; Michael et al., 2005; Trejo et al., 2007; Soleimani Nazar and Bayandori, 2008; Daou et al., 2009; Yasar et al., 2009; Yi et al., 2009; Hosseinpour et al., 2013; Molina et al., 2016; Poveda-Jaramillo et al., 2014).

![Fig. 5. Hierarchical cluster analysis of four studied groups in the Persian Gulf (modified from Rabbani et al., 2014).](image)

The names refer to the 13 samples selected for this study.

### Table 2 The calculated structural parameters of the studied samples

<table>
<thead>
<tr>
<th>Field</th>
<th>Reservoir</th>
<th>d_{s} (Å)</th>
<th>d_{as} (Å)</th>
<th>L_{c} (Å)</th>
<th>L_{a} (Å)</th>
<th>C_{au}</th>
<th>R_{a}</th>
<th>M_{a}</th>
<th>f_{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kharg</td>
<td>Manifa</td>
<td>5.468</td>
<td>3.544</td>
<td>10.194</td>
<td>32.130</td>
<td>17.576</td>
<td>3.822</td>
<td>10.065</td>
<td>0.267</td>
</tr>
<tr>
<td>Dorood</td>
<td>Yamamma</td>
<td>4.670</td>
<td>3.528</td>
<td>9.999</td>
<td>35.803</td>
<td>17.275</td>
<td>3.749</td>
<td>11.147</td>
<td>0.265</td>
</tr>
<tr>
<td>Dorood*</td>
<td>Yamamma*</td>
<td>4.703</td>
<td>3.507</td>
<td>10.495</td>
<td>32.704</td>
<td>18.039</td>
<td>3.935</td>
<td>10.322</td>
<td>0.282</td>
</tr>
<tr>
<td>Foroozan</td>
<td>Burghan</td>
<td>4.670</td>
<td>3.528</td>
<td>9.999</td>
<td>35.803</td>
<td>17.275</td>
<td>3.749</td>
<td>11.147</td>
<td>0.265</td>
</tr>
<tr>
<td>Dorood</td>
<td>Manifa</td>
<td>4.703</td>
<td>3.507</td>
<td>10.495</td>
<td>32.704</td>
<td>18.039</td>
<td>3.935</td>
<td>10.322</td>
<td>0.282</td>
</tr>
<tr>
<td>Salman</td>
<td>Buwaib</td>
<td>4.614</td>
<td>3.539</td>
<td>10.450</td>
<td>35.209</td>
<td>19.992</td>
<td>3.923</td>
<td>10.976</td>
<td>0.276</td>
</tr>
<tr>
<td>Balal</td>
<td>Khatiah</td>
<td>4.637</td>
<td>3.570</td>
<td>10.065</td>
<td>32.172</td>
<td>17.378</td>
<td>3.774</td>
<td>10.011</td>
<td>0.267</td>
</tr>
<tr>
<td>Bahregansar</td>
<td>Ghar</td>
<td>4.761</td>
<td>3.578</td>
<td>9.009</td>
<td>37.474</td>
<td>15.755</td>
<td>3.378</td>
<td>11.472</td>
<td>0.197</td>
</tr>
<tr>
<td>Hendijan</td>
<td>Sarvak</td>
<td>4.761</td>
<td>3.578</td>
<td>9.009</td>
<td>37.474</td>
<td>15.755</td>
<td>3.378</td>
<td>11.472</td>
<td>0.197</td>
</tr>
<tr>
<td>Nowrooz</td>
<td>Nahreumr</td>
<td>4.619</td>
<td>3.541</td>
<td>9.819</td>
<td>38.312</td>
<td>16.999</td>
<td>3.681</td>
<td>11.818</td>
<td>0.233</td>
</tr>
<tr>
<td>Resalat</td>
<td>Mishrif</td>
<td>4.695</td>
<td>3.539</td>
<td>10.122</td>
<td>36.803</td>
<td>17.464</td>
<td>3.795</td>
<td>11.397</td>
<td>0.203</td>
</tr>
<tr>
<td>Reshadat</td>
<td>Mishrif</td>
<td>4.761</td>
<td>3.522</td>
<td>9.927</td>
<td>34.315</td>
<td>17.164</td>
<td>3.722</td>
<td>10.740</td>
<td>0.221</td>
</tr>
<tr>
<td>Reshadat*</td>
<td>Mishrif*</td>
<td>4.604</td>
<td>3.542</td>
<td>9.420</td>
<td>36.859</td>
<td>16.385</td>
<td>3.532</td>
<td>11.404</td>
<td>0.229</td>
</tr>
<tr>
<td>Nosrat</td>
<td>Mishrif</td>
<td>4.712</td>
<td>3.531</td>
<td>10.776</td>
<td>34.926</td>
<td>18.471</td>
<td>4.041</td>
<td>10.889</td>
<td>0.278</td>
</tr>
</tbody>
</table>

Note that samples marked (*) are those selected for repeatability assessment.
Fig. 6. The XRD patterns of the studied asphaltene samples. Samples marked (*) are those selected for repeatability assessment.
Following XRD analysis, the asphaltene samples were analyzed using FTIR. The spectra from FTIR analysis were smoothed and normalized, as shown in Fig. 7. The aforementioned indices were then calculated by equations A1 to A8 (Appendix B). Detailed results for FTIR are presented in Table 3: the values of aromatic, aliphatic, branched carbon chains, long carbon chains, carbonyl, sulfoxide, substitution 1 and substitution 2 indices for all 13 asphaltene samples as well as the repeat samples are presented in this table. As can be seen, aromatic and aliphatic indices vary within the small ranges of 0.04 to 0.07 and 0.20 to 0.23, respectively. By comparing the results presented in Tables 2 and 3, it can be determined that the XRD technique is more precise in the determination of aromaticity values than FTIR. Indeed, the calculated values of aromaticity obtained by the XRD patterns are in good agreement with other techniques such as NMR (Zhang et al., 1989; Shirokoff et al., 1997; Okolo et al., 2015; Baysal et al., 2016; Poveda-Jaramillo et al., 2016), the calculated aromaticity values based on FTIR spectra being smaller than the actual values.

In the following section, in order to better understand the role of the XRD results in oil-oil correlation, the study area is discussed in two distinct geographical components as northwest and southeast.

### 3.1.1 Southeast of the Persian Gulf

Six oil samples from two major oil families (groups II and IV) have been assessed in this section. Salman (Buwaib) and Balal (Khatiah) are from group II. The star diagram and the Galimov curves of oil samples from Kharg (Manifa), Sarvak, Ghar, Bahregamsar, Sarvak, Nahrumr, Resalat, Mishrif, Reshadat, Mishrif, Reshadat, and Sirri D are discussed in two distinct geographical components as northwest and southeast.

<table>
<thead>
<tr>
<th>Field</th>
<th>Reservoir</th>
<th>Aromaticity index</th>
<th>Aliphatic index</th>
<th>Branched index</th>
<th>Long chains index</th>
<th>Carbenyl index</th>
<th>Sulfoxide index</th>
<th>Substitution 1 index</th>
<th>Substitution 2 index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kharg</td>
<td>Manifa</td>
<td>0.060</td>
<td>0.219</td>
<td>0.261</td>
<td>0.064</td>
<td>0.014</td>
<td>0.028</td>
<td>0.248</td>
<td>0.444</td>
</tr>
<tr>
<td>Dorood</td>
<td>Yamamma</td>
<td>0.062</td>
<td>0.226</td>
<td>0.278</td>
<td>0.064</td>
<td>0.009</td>
<td>0.024</td>
<td>0.234</td>
<td>0.407</td>
</tr>
<tr>
<td>Dorood*</td>
<td>Yamamma*</td>
<td>0.048</td>
<td>0.210</td>
<td>0.334</td>
<td>0.061</td>
<td>0.007</td>
<td>0.020</td>
<td>0.153</td>
<td>0.453</td>
</tr>
<tr>
<td>Foroosan</td>
<td>Burghah</td>
<td>0.068</td>
<td>0.209</td>
<td>0.264</td>
<td>0.031</td>
<td>0.017</td>
<td>0.036</td>
<td>0.252</td>
<td>0.368</td>
</tr>
<tr>
<td>Dorood</td>
<td>Manifa</td>
<td>0.059</td>
<td>0.217</td>
<td>0.275</td>
<td>0.026</td>
<td>0.009</td>
<td>0.022</td>
<td>0.244</td>
<td>0.394</td>
</tr>
<tr>
<td>Salman</td>
<td>Buwaib</td>
<td>0.072</td>
<td>0.224</td>
<td>0.276</td>
<td>0.048</td>
<td>0.034</td>
<td>0.034</td>
<td>0.302</td>
<td>0.354</td>
</tr>
<tr>
<td>Balal</td>
<td>Khatiah</td>
<td>0.061</td>
<td>0.218</td>
<td>0.255</td>
<td>0.050</td>
<td>0.010</td>
<td>0.017</td>
<td>0.292</td>
<td>0.362</td>
</tr>
<tr>
<td>Bahregamsar</td>
<td>Ghar</td>
<td>0.058</td>
<td>0.221</td>
<td>0.264</td>
<td>0.035</td>
<td>0.020</td>
<td>0.043</td>
<td>0.192</td>
<td>0.375</td>
</tr>
<tr>
<td>Hendijan</td>
<td>Sarvak</td>
<td>0.077</td>
<td>0.223</td>
<td>0.261</td>
<td>0.034</td>
<td>0.023</td>
<td>0.033</td>
<td>0.278</td>
<td>0.394</td>
</tr>
<tr>
<td>Nowrouz</td>
<td>Nahrumr</td>
<td>0.069</td>
<td>0.237</td>
<td>0.260</td>
<td>0.020</td>
<td>0.014</td>
<td>0.013</td>
<td>0.230</td>
<td>0.419</td>
</tr>
<tr>
<td>Resalat</td>
<td>Mishrif</td>
<td>0.077</td>
<td>0.232</td>
<td>0.263</td>
<td>0.043</td>
<td>0.007</td>
<td>0.045</td>
<td>0.273</td>
<td>0.643</td>
</tr>
<tr>
<td>Reshadat</td>
<td>Mishrif</td>
<td>0.070</td>
<td>0.226</td>
<td>0.261</td>
<td>0.029</td>
<td>0.036</td>
<td>0.033</td>
<td>0.264</td>
<td>0.332</td>
</tr>
<tr>
<td>Reshadat*</td>
<td>Mishrif*</td>
<td>0.068</td>
<td>0.229</td>
<td>0.278</td>
<td>0.014</td>
<td>0.031</td>
<td>0.032</td>
<td>0.262</td>
<td>0.340</td>
</tr>
<tr>
<td>Sirri D</td>
<td>Mishrif</td>
<td>0.075</td>
<td>0.219</td>
<td>0.259</td>
<td>0.037</td>
<td>0.009</td>
<td>0.016</td>
<td>0.285</td>
<td>0.406</td>
</tr>
<tr>
<td>Nosrat</td>
<td>Mishrif</td>
<td>0.077</td>
<td>0.215</td>
<td>0.254</td>
<td>0.058</td>
<td>0.017</td>
<td>0.022</td>
<td>0.283</td>
<td>0.362</td>
</tr>
</tbody>
</table>

Note that samples marked (*) are those selected for repeatability assessment.
Fig. 7. FTIR spectra of the studied asphaltenes samples. Samples marked (*) are those selected for repeatability assessment.
Dorood (Yamamma), Foroozan (Burghan) and Dorood (Manifa) (group I) and samples from Bahregansar (Ghar), Hendijan (Sarvak) and Nowrooz (Nahreumr) (group III) are depicted in Figs. 12 and 13, respectively. As can be seen, the members of group III exhibit some discrepancies and, despite the other groups, there is no consistent trend in its members. In particular, the sample from Hendijan (Sarvak) differs from two other members of this group.

Rabbani et al. (2014) identified several factors, such as migration from other adjacent basins and the existence of several source rock ‘kitchens’ in the study area, that could be considered as reasons for these break downs. To dismiss this issue, they considered some subgroups for assessing group III (Rabbani et al., 2014; Baniasad et al., 2017). As such, the Hendijan (Sarvak) sample is in a subgroup that is separated from the two other members of this group. Rabbani et al. (2014) identified several factors, such as migration from other adjacent basins and the existence of several source rock ‘kitchens’ in the study area, that could be considered as reasons for these break downs. To dismiss this issue, they considered some subgroups for assessing group III (Rabbani et al., 2014; Baniasad et al., 2017). As such, the Hendijan (Sarvak) sample is in a subgroup that is separated from the two other members of this group (Rabbani et al., 2014). In this study, two samples of Bahregansar (Ghar) and Nowrooz (Nahreumr) are considered as subgroup A, with Hendijan (Sarvak) in subgroup B. The distinction between these two subgroups can be clearly seen in Fig. 13.

To evaluating the XRD method in this area, cross-plots of $d_r$ vs. $d_m$ (Fig. 14a), $d_r$ vs. $M_e$ (Fig. 14b), $d_r$ vs. $f_a$ (Fig. 14c), $d_m$ vs. $L_c$ (Fig. 14d), $L_c$ vs. $L_a$ (Fig. 14e) and $f_a$ vs. $L_c$ (Fig. 14f) were drawn. As illustrated, the separation of the two subgroups and even the two groups is suitable and acceptable. Fig. 14 shows that the XRD results can be applied as a proper tool for oil-oil correlation in this area too. As can be seen, the discrepancy between the two subgroups of group III, which was demonstrated by extensive ranges of biomarker and isotope data (Rabbani et al., 2014), can be detected in the XRD cross-plots. In some cases, some blurring in the distinction of these two groups can be seen, whereas the two previous groups (groups II and IV in the southeastern section) were completely distinguishable. It must be emphasized that the geochemical characteristics of groups I and III indicate that the members of these groups possess some similarities (Rabbani et al., 2014). In other words, even the biomarkers or isotopic ratios show the similarity between
groups I and III. Although there are geological and geochemical complications in this area that make correlation difficult, even with broad sets of geochemical data, the results of XRD plots are as effective as this geochemical data. Fig. 14 also shows that, in spite of the aforementioned complexities of these two genetic groups,
plots of $d_t$ vs. $M_e$ (Fig. 14b), $d_m$ vs. $L_c$ (Fig. 14d), $L_a$ vs. $L_c$ (Fig. 14e), and $f_a$ vs. $L_c$ (Fig. 14f) can illustrate the separation between groups I and III, as well as even two subgroups of group III. In the next step, FTIR cross-plots were drawn and are shown in Fig. 15. As illustrated, plots of substitution 1 vs. aromaticity (Fig. 15a) and sulfoxide + carbonyl vs. aromaticity + aliphaticity (Fig. 15d) perform better than aliphaticity vs. aromaticity (Fig. 15b) and sulfoxide + carbonyl vs. aromaticity (Fig. 15c) in discerning groups and subgroups.

Finally, it can be stated that the results of assessing asphaltene cross-plots indicate that the asphaltene structural parameters can provide the same results for oil-oil correlation as biomarkers, while consuming less time and money.

### 3.2 Statistical classification

#### 3.2.1 Principal component analysis

Using chemometrics is a prevalent way for oil-oil or oil-source correlation objectives. PCA or principal component analysis is one of the most used approaches in this field. The fundamental basis of this technique is dimension reduction. In geochemical studies, wide ranges of geochemical characteristics, such as biomarker ratios, are considered as inputs. This approach has already been extensively used (Oygard et al., 1984; Telnaes and Dahl, 1986; Raju and Mathur, 1995; Napitupulu et al., 2000; Eneogwe and Ekundayo, 2003; Gürgey, 2003; Zumberge et al., 2005; Justwan et al., 2006; He et al., 2012; Rabbani et al., 2014; Wang et al., 2014; Mashhadi and Rabbani, 2015; Alizadeh et al., 2017).

For the samples in this paper, the PCA method has been used for classification in a previous publication (Rabbani et al., 2014). This study used 14 biomarkers and isotopic ratios that were related to source rock for PCA classification. In this study, the PCA was conducted based on the XRD and FTIR parameters of the studied asphaltenes. In other words, 12 asphaltene parameters (indices of aromaticity, aliphaticity, branched, long chains, carbonyl, sulfoxide, substitution 1 and substitution 2 from FTIR and $d_m$, $L_a$, $L_c$ and $f_a$ from XRD results) were used for this purpose. For a better comparison, Fig. 16 illustrates the PCA plot based on biomarker and isotopic ratios (Fig. 16a) as well as the PCA plot based on
calculated XRD and FTIR parameters of asphaltene samples (Fig. 16b). As seen in this figure, PCA based on asphaltene parameters can differentiate between four groups appropriately and performs as excellently as in the previous study.

The most significant point about PCA based on asphaltene results is its simplicity and the necessity for fewer data. For instance, Rabbani et al. (2014) used 14
source-related parameters. On the other hand, several publications used 52 (Telnaes and Dahl, 1986), 15 (Raju and Mathur, 1995), 17 (Napitupulu et al., 2000), 14 (Eneogwe and Ekundayo, 2003), 23 (Wang et al., 2014), 16 (Mashhadi and Rabbani, 2015) and 18 (Zumberge et al., 2005) source rock geochemical parameters for PCA classification, whereas the structural parameters of asphaltene obtained by the XRD and FTIR methods can provide as satisfactory a classification as the traditional methods.

3.2.2 Decision tree classification

Data mining is a process of exploring patterns and trends in extremely large quantities of data that works by extracting efficient and impressive data from an extensive range of data. The decision tree (DT) is one of the most significant machine learning techniques and is considered to be one of the useful approaches of data mining subjects (Jadhav and Channe, 2016; Sharma and Kumar, 2016). In this study, a classification tree was developed for the four studied genetic groups, which is illustrated in Fig. 17. The input data for this tree are structural parameters obtained by XRD. This tree is composed of nodes and branches. The root node (the first one in the first level of the tree) and the leaf nodes (nodes that indicate the oil groups) are connected through the branches. As shown in this figure, this classification tree can represent the bounds of the asphaltene structural parameters in each group.

Fig. 17 indicates the applied classification tree using \( d_{\text{mr}} \), \( d_{\text{r}} \), \( M_e \) and \( L_c \) parameters for genetic classification. As
illustrated, values of $d_i$ lower than 4.68Å, as well as $M_i$ lower than 10.02, are characteristic of group II. When $d_i$ is lower than 4.68Å, $M_i$ is more than 10.02 and $L_i$ is lower than 37.05Å, the samples belonging to group I. Also, group IV is specified with $d_i$ values more than 4.68Å and $d_m$ values between 3.51Å and 3.55Å. This feature of the decision tree approach, that can distinguish various genetic groups with the most impressive parameters and specification of boundaries in each group, seems excellent and unparalleled in geochemical interpretations. However, inherent discrepancies and differences between members of group III resulted in three distinct paths in the classification tree, despite the three other groups having merely one unique route. Ultimately, it has to be emphasized that the classification tree of Fig. 17 can also be employed for genetic group prediction for possible subsequent samples. The probability of membership in each group is also determined and shown in the boxes. For example, if the value of $d_i$ is more than 4.68Å, the probabilities of groups I and II are zero, whereas the probabilities of groups III and IV are approximately 33% and 66%, respectively.

### 3.3 The intrinsic structure of asphaltene

The discussion in this paper has indicated that the structure of asphaltenes (obtained by XRD) can be a means of separating geochemical groups. This shows that the structure of asphaltenes is a genetic concept that is related to the origin of the oil samples. In other words, the oil samples of a genetic family are the same in their biomarker and isotopic ratios, as well as in their structural foundations. It can thus be deduced that a particular source rock produces oils with the same asphaltene structures. It can therefore be concluded that similar structures of different asphaltene samples can represent a unique source rock.

The average values of the structural parameters of each group are presented in Table 4. For greater clarity, the average values of $d_i$, $d_m$, $L_i$, and $L_c$ that determine the sizes of the asphaltene clusters in each group are depicted in Fig. 18. As shown in this figure, the values of $d_i$ are ascending from group I to group IV (Fig. 18a). Group II possesses the highest value for the distance between aromatic portions (Fig. 18b). The average diameter of aromatic sheets in group IV is more than the other groups, with group II having the lowest diameter (Fig. 18c). The largest asphaltene clusters belong to group III and then to group IV (Fig. 18d). These values indicate that each genetic family of oils possesses an exclusive structure of asphaltene. For an easier explanation of this subject, hypothetical structures of all studied groups are illustrated in Fig. 19. These structures are drawn based on the average values in each group. Fig. 19 can provide a visual representation of the asphaltene structure in each group.

### 3.4 Advantages, disadvantages, drawbacks and recommendations

As a final point, it should be noted that using the XRD method for geochemical assessment of asphaltenes has several significant advantages. The technique costs much less than other analyses, such as biomarkers and carbon isotope investigation, while using the fraction of crude oil (asphaltene) most similar to kerogen. The sample preparation for XRD analysis is simple and quick. The required amount of the sample is small. Moreover, using the structure of asphaltene for oil-oil correlation reduces the effects of secondary processes affecting crude oil to an acceptable level. Since this approach is based on asphaltene structural assessment and identification of the genetic characteristics of crude oil are the objectives of oil-oil correlation investigation, the sampling procedure (surface or downhole) is not significant and a few amounts of extracted asphaltene are enough for this approach. Although this technique can provide valuable information for geochemists, it has some shortcomings, like any other analysis. For instance, the determination of a baseline and curve fitting for the detected peaks are somewhat dependent on the operator. The diffractograms of asphaltene possess only three detected peaks, thus limiting the information obtained. Also, despite other secondary processes, maturity can affect asphaltene, to some extent. It is predicted that under high maturity conditions, due to the change in asphaltene structure, the efficiency of this method will decrease. It is also to be acknowledged that this test study also faces limitations. The crude oil samples are only sourced from within the Persian Gulf and it is therefore unknown as to how well this method might perform in other samples or regions, where it might require the application of other auxiliary methods.

Other geochemical tools like biomarkers and isotopic ratios do not individually provide complete information. This novel approach should also be supported by other techniques, such as FTIR. However, this new technique, by providing an overall insight, can be regarded as a screening method in the oil industry.

Another point that must be emphasized is the possible errors for such studies. The procedure for asphaltene precipitation can cause some experimental errors. Since the aim of this study is identifying the genetic characteristics of asphaltene samples through comparison, the circumstances and conditions of asphaltene precipitation must be identical to reduce the adverse effects of experimental errors. It should be noted that there are several methods for asphaltene precipitation and asphaltenes are insoluble in aliphatic solvents. The type of solvent used affects the extracted asphaltene samples, the asphaltene sample which is derived by n-Hexane differing from the one that is extracted by n-Heptane (Liu et al., 2018). For oil-oil correlation objectives, similar conditions and a similar solvent are mandatory, since the samples

**Table 4 The average values for the parameters obtained by the XRD method for each genetic group**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i$ (Å)</td>
<td>4.619</td>
<td>4.626</td>
<td>4.694</td>
<td>4.793</td>
</tr>
<tr>
<td>$d_m$ (Å)</td>
<td>3.542</td>
<td>3.555</td>
<td>3.542</td>
<td>3.530</td>
</tr>
<tr>
<td>$L_i$ (Å)</td>
<td>9.986</td>
<td>10.257</td>
<td>9.774</td>
<td>10.366</td>
</tr>
<tr>
<td>$L_c$ (Å)</td>
<td>33.851</td>
<td>30.794</td>
<td>36.164</td>
<td>34.285</td>
</tr>
<tr>
<td>$C_m$</td>
<td>17.256</td>
<td>17.673</td>
<td>16.930</td>
<td>17.840</td>
</tr>
<tr>
<td>$R_m$</td>
<td>3.7444</td>
<td>3.846</td>
<td>3.665</td>
<td>3.886</td>
</tr>
<tr>
<td>$M_i$</td>
<td>10.558</td>
<td>9.660</td>
<td>11.204</td>
<td>10.710</td>
</tr>
<tr>
<td>$f_i$</td>
<td>0.251</td>
<td>0.338</td>
<td>0.237</td>
<td>0.251</td>
</tr>
</tbody>
</table>
should be comparable. Therefore, the unique n-Hexane solvent and identical conditions have been utilized for this study. Also, the curve-deconvolution of recorded XRD patterns of asphaltene samples should be performed by one operator, with the same procedure being utilized to diminish the errors of peak recognition and parameter calculation.

4 Conclusions

In this study, the obtaining of asphaltene's structure through use of the XRD technique was introduced as a geochemical tool for oil-oil correlation. For this purpose, 13 asphaltene samples from the Iranian section of the Persian Gulf region were selected and then analyzed using two popular techniques of XRD and FTIR for the determination of genetic groups. Cross-plots for genetic group determination and decision tree classification were applied, based for the first time on XRD data. The obtained results were validated by biomarkers and isotope ratios.

The main achievements of this study are as follows:

1. The structural characteristics obtained for studied asphaltenes in the Persian Gulf indicate that the distance between saturated chains varies between 4.58Å to 5.00Å, whereas the distance between two aromatic sheets of asphaltene stacks changes across the limited range of 3.50Å to 3.57Å. The average height of the asphaltene clusters is between 29.41Å to 38.31Å.

2. The average structural values of the four studied groups indicate that group I has the lowest $d_r$ value (4.61Å) and group II is characterized by the highest $d_m$ (3.55Å) and lowest $L_c$ (30.79Å). The smallest diameters of aromatic sheets belong to group III (9.77Å). Also, group IV is recognized through the highest values of $d_r$ (4.79Å) and $L_a$ (10.36Å), as well as the shortest distance between aromatic sheets (3.53Å).

3. The results of the structural parameters obtained by the XRD technique are in good agreement with other traditional methods, such as biomarker and isotopic ratios. Therefore, this novel, precise and inexpensive method can be applied as a new approach for oil-oil correlation objectives.

4. It was also found that the precision of the XRD method for dissociation of genetic groups is higher than FTIR.

5. Based on FTIR results, plots of substitution 1 vs. aromaticity and sulfoxide + carbonyl vs. aromaticity +
aliphaticity perform better than aliphaticity vs. aromaticity and sulfoxide + carbonyl vs. aromaticity, in detecting groups and subgroups.

(6) Principal component analysis based on XRD and FTIR methods can provide results that are as effective as the results of PCA based on traditional techniques, with a smaller number of data.

(7) The decision tree used \( d_m, d_s, M_0 \) and \( L_c \) parameters of XRD for genetic classification.

(8) Decision tree classification can be regarded as a suitable tool for determining the characteristics of genetic groups and even for predicting the genetic group of possible asphaltene samples from the study area.

(9) The structures of asphaltenes are dependent on their source rock and as every geochemical group possesses unique geochemical features, so it also possesses unique asphaltene structure.

Appendix A. Nomenclature terms

- \( \lambda \): Radiation wavelength (Å)
- \( \theta \): Diffraction angle (°)
- \( \omega \): The full width at half maximum (°)
- \( d_r \): Distance between saturated chains (Å)
- \( d_m \): Distance between two aromatic sheets (Å)
- \( L_s \): The diameter of aromatic sheets (Å)
- \( L_t \): The average height of the cluster perpendicular to the sheets plain (Å)
- \( c_m \): Number of carbons in the aromatic unit
- \( R_s \): Number of aromatic rings in aromatic sheets
- \( M_0 \): Number of aromatic sheets in a cluster
- \( f_a \): Aromaticity index

Appendix B. Equations of the FTIR method (Permanyer et al., 2005, 2007; Asemani and Rabbani, 2016)

Aromatic index: \( \Sigma A = A_{1470} + A_{1700} + A_{1600} + A_{1460} + A_{1376} + A_{1030} + A_{864} + A_{814} + A_{743} + A_{724} + A_{(2953, 2923, 2862)} \)

Aliphatic index: \( A_A = (A_{2923} + A_{2953}) / \Sigma A \)

Long chain index: \( A_L = A_{2864} / (A_{2864} + A_{2923} + A_{2953}) \)

Branched index: \( A_B = (A_{2864} + A_{2923} + A_{2953}) / A_{2864} \)

Carbonyl index: \( A_C = A_{1600} / A_{1470} \)

Sulfoxide index: \( A_S = A_{1030} / A_{1470} \)

Substitution index 1: \( A_{1376} / (A_{2864} + A_{2923} + A_{2953}) \)

Substitution index 2: \( A_{1376} / (A_{2864} + A_{2923} + A_{2953}) \)

\( \Sigma A = A_{1200} + A_{1000} + A_{1150} + A_{1170} + A_{1010} + A_{864} + A_{814} \)

\( A_{1376} / A_{1200} + A_{1000} + A_{1150} + A_{1170} + A_{1010} + A_{864} + A_{814} \)

References


Rabbani, A.R., Kotarba, M.J., Baniasad, A.R., Hosseiny, E., and
Rabbani, A.R., and Baniasad, A.R., 2017. Application of
Rabbani, A.R., 2008. Geochemistry of crude oil samples from
Permanyer, A., Douifi, L., Dupuy, N., Lahcini, A., and Kister, J.,
Pan, C.C., Geng, A.S., Liao, Z.W., Xiong, Y.Q., Fu, J.M., and
Øygard, K., Grahl
Okolo, G.N., Neomagus, Hein.W.J.P., Everson, R.C., Roberts,
Nguyen, N.T., Kang, K.H., Lee, C.W., Kim, G.T., Park, S., and
Nali, M., Caccialanza, G., Ghiselli, C., and Chiaramonte, M.A.,
118
Geochimica et Cosmochimica Acta, 43: 1
of asphaltenes: A source of geochemical information.
Cadernos de Geociências, 12: 84
petroleum: Asphaltenes as a source of geochemistry
Maciel, I.C., and de Oliveira, O.M.C., 2015. Geochemistry of
Geochemistry, 23: 55
part of Upper Assam Basin, India: A brief overview. Organic
Chemistry, 31: 1325–1332
Chemical characterization of free
561–567.
Pan, C.C., Geng, A.S., Liao, Z.W., Xiong, Y.Q., Fu, J.M., and
Permanyer, A., Douifi, L., Dupuy, N., Lahcini, A., and Kister, J.,
Permanyer, A., Douifi, L., Lahcini, A., Lamontagne, J., and

About the first author
Zahra SADEGHTABAGHI, female, born in 1991 in Iran, graduated from Amirkabir University of Technology, and received Ph.D. of petroleum engineering in 2021. She is now interested in the petroleum geochemistry and basin modeling. E-mail: zahrasadeghtabaghi@aut.ac.ir.

About the corresponding author
Ahmad Reza RABBANI, currently works as Professor at the Department of Petroleum Engineering, Amirkabir University of Technology. Rabbani does research in reservoir geochemistry, and reservoir continuity assessment by integrated geochemical approach, asphaltene geochemistry, environmental geochemistry, H₂S source study and mitigation approach in the oil and gas reservoir. E-mail: rabbani@aut.ac.ir.