Iranian Journal of Oil & Gas Science and Technology, Vol. 8 (2019), No. 4, pp. 01-17
http://ijogst.put.ac.ir

Chemometrics-enhanced Classification of Source Rock Samples Using their Bulk Geochemical Data: Southern Persian Gulf Basin

Majid Alipour[1*], Bahram Alizadeh[1,2], Scott Ramos[3], Behzad Khani[4], and Shohreh Mirzaie[5]

[1] Assistant Professor, Department of Geology, Faculty of Earth Sciences, Shahid Chamran University of Ahvaz, Iran
[2] Professor, Petroleum Geology and Geochemistry Research Center (PGGRC), SCU, Ahvaz, Iran
[3] Professor, Infometrix, Inc. 11807 North Creek Parkway South, Suite B-111, Bothell, WA 98011
[4] Ph.D., Research Institute of Petroleum Industry (RIPI), Tehran, Iran
[5] Ph.D., Pars Petro Zagros Engineering & Services Co. (PPZ), Tehran, Iran

Received: October 20, 2018; *revised*: November 03, 2018; *accepted*: January 19, 2019

Abstract

Chemometric methods can enhance geochemical interpretations, especially when working with large datasets. With this aim, exploratory hierarchical cluster analysis (HCA) and principal component analysis (PCA) methods are used herein to study the bulk pyrolysis parameters of 534 samples from the Persian Gulf basin. These methods are powerful techniques for identifying the patterns of variations in multivariate datasets and reducing their dimensionality. By adopting a “divide-and-conquer” approach, the existing dataset could be separated into sample groupings at family and subfamily levels. The geochemical characteristics of each category were defined based on loadings and scores plots. This procedure greatly assisted the identification of key source rock levels in the stratigraphic column of the study area and highlighted the future research needs for source rock analysis in the Persian Gulf basin.

Keywords: Chemometric Classification, Source Rock Geochemistry, Rock-Eval Pyrolysis Data, HCA, PCA.

1. Introduction

Geochemical screening is one of the main concerns for petroleum system studies. The procedure often includes identifying the source intervals which might exist within the stratigraphic column as well as defining their areal extent and organic attributes. For this purpose, the analytical data from large numbers of samples (analyzed at regular intervals within oil wells) are treated using Microsoft Excel by considering certain thresholds for various parameters. Multivariate methods, on the other hand, can provide a much clearer picture of the structure of the dataset by considering all the parameters simultaneously.

Factor-based chemometric methods are used to reduce the dimensionality of large sets of data (Ramos et al., 1986). Principal component analysis (PCA) is an exploratory method which is a cornerstone of many multivariate studies and is based on the characterization of variance in a data set. HCA is a complementary method that describes differences among samples based on a multivariate distance. After doing an exploratory analysis using one or both of these methods, subsequent modeling can be

* Corresponding author:
Email: alipour@scu.ac.ir

performed for either classification (Kowalski et al., 1972) or some regression modeling (Peters et al., 2013).

During the past few decades, statistical methods have been successfully applied to various purposes, including oil geochemistry, oil classification, correlation, and detailed biomarker studies.

In their landmark study of the greater Ekofisk field in the North Sea, Hughes and coworkers used PCA for disentangling the effects of source and maturity on bulk and molecular geochemical data (Hughes et al., 1985). Other investigators have used decision-tree chemometrics to identify Circum-Arctic petroleum systems (Peters et al., 2007). These authors have used PCA in combination with K-nearest neighbor and soft independent modeling of class analogy (SIMCA) models to classify a large number of oil samples into genetically distinct families using source-related biomarker parameters. Also, a similar work was previously completed on oils from Italian Po basin (Oberrauch et al., 1987). Recently, Peters and coworkers have used a similar method for the chemometric differentiation of oil families in the San Joaquin Basin (Peters et al., 2013).

Other applications include studies by Telnaes and Dahl (1985), who tried to correlate oil samples from the Norwegian sector of the North Sea using multivariate analysis (PCA). These authors could define a relationship between the three main principal components and important geochemical processes (Telnaes and Dahl, 1986). Based on a similar approach, Kvalheim used the distributions of C_6 and C_7 saturates for oil-source correlation (Kvalheim, 1987). In the same context, Telnaes and Cooper (1991), correlated Norwegian oil samples with their corresponding source layers based on multivariate statistical results (Telnæs and Cooper, 1991).

Detailed biomarker research has also benefited greatly from statistical methods to better identify the relationship between the groups of compounds (Telnaes and Dahl, 1986), to study the distribution of specific compounds (Fabiańska, 2004), and to elucidate the source of polluting hydrocarbons in coastal systems (Grimalt et al., 1993). Farrimond and coworkers interpreted detailed biomarker data from black shale samples representing the Toarcian anoxic event in Northern Italy to demonstrate fluctuations in the organic production and preservation during shale deposition (Farrimond et al., 1994). Kruge (2000) used PCA to determine the thermal maturity and type of the organic matter by analyzing the distribution of polycyclic aromatic compounds in the solvent extracts of different shales, coals, and kerogen samples (Kruge, 2000).

This study is the first report on the successful application of chemometric methods to the Persian Gulf basin, which addresses the bulk geochemical parameters of source rock samples. We have used two exploratory algorithms (HCA and PCA) for grouping the samples into main families via the patterns of distinctive organic properties. These groupings represent bodies of the samples with globally uniform geochemical parameters which may be associated with certain stratigraphic levels or geographic regions in the area. In this way, we have been able to identify several key source rock categories and describe their organic geochemical properties in the southern Persian Gulf basin. These, in turn, have provided invaluable inputs to petroleum system modeling studies in the same area (Alipour et al., 2016a; Alipour, 2017; Alipour et al., 2017) through a procedure called “descriptive source rock quality mapping” (Alipour et al., 2016b).

2. Geological background

The area of the current study is located in the Iranian part of the Persian Gulf basin (Figure 1). The area has recently been the focus of extensive exploration activities due to the very rich hydrocarbon deposits which are, in some cases, ranked as the world biggest oil/gas deposits. Following the recent geophysical survey, interest in locating hitherto undiscovered reserves has been aroused (Alipour, 2017; Alipour et al., 2017; Alizadeh et al., 2017). In this context, basin and petroleum system studies are of special importance for studying the essential elements and processes responsible for the occurrence of hydrocarbons in multiple dimensions (Peters et al., 2012).

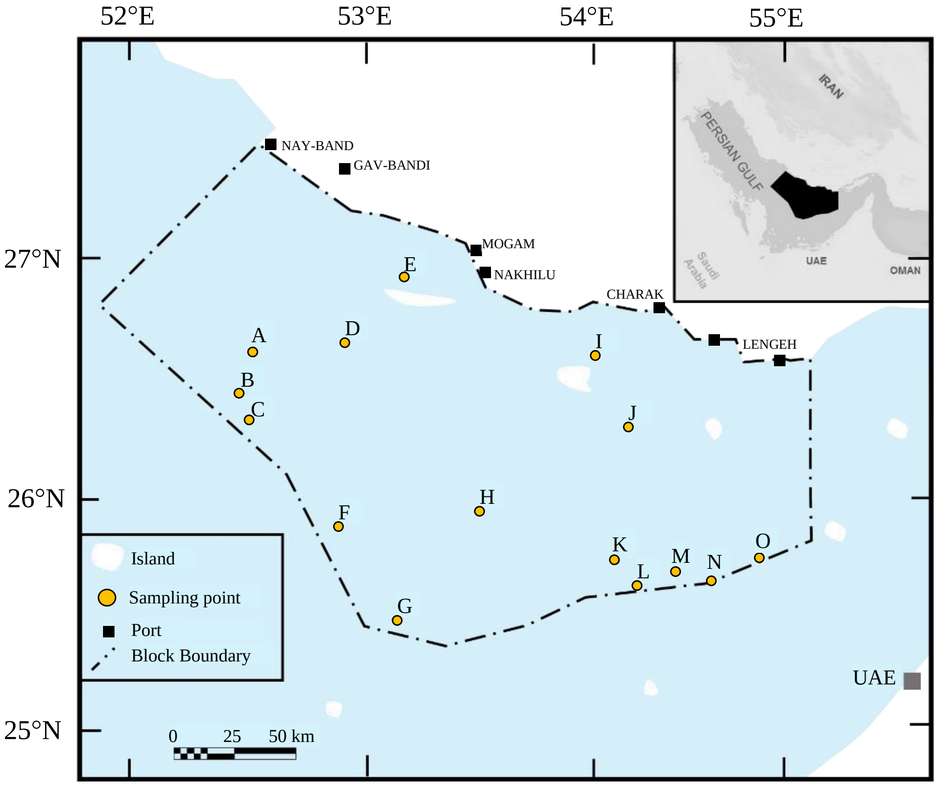

Figure 1

The generalized map of the study area indicating the location of oil wells (across 15 fields) from which the samples of this study were collected.

The Persian Gulf basin encompasses a thick sedimentary succession with alternating clastics, carbonates, and evaporate sediments, which makes the area particularly prolific for hosting large hydrocarbon deposits (Ghazban, 2007) (Figure 2). The oldest sediments in the area are believed to be the evaporates, shales, and dolomites of the Late-Precambrian Hormuz Series (Kent, 1979; Nasir et al., 2008). Generally speaking, there is little data available about the sedimentary history of the Lower Paleozoic in the Persian Gulf region, and the sedimentary record comprises mostly shales and sandstone with rare carbonates in Devonian and Early Carboniferous (Murris, 1980).

During the Permian, carbonate shelf deposits of the Dalan formation were laid under warm, shallow-water conditions (Alsharhan, 1989). More arid conditions during Mid-Late Triassic formed the evaporate deposits of the Dashtak formation, which marks the end of the carbonate cycles (Alsharhan and Kendall, 1986). A sea-level drop in the early Jurassic caused clastic deposition (the Neyriz formation), with a subsequent unconformity over an extensive area (Sharland et al., 2001). The Middle Jurassic sediments mainly consist of normal marine shelf carbonates (the Surmeh formation) which are capped with extensive evaporates deposited under very shallow conditions (Hith formation) during Tithonian (Murris, 1980; Lasemi and Jalilian, 2010).

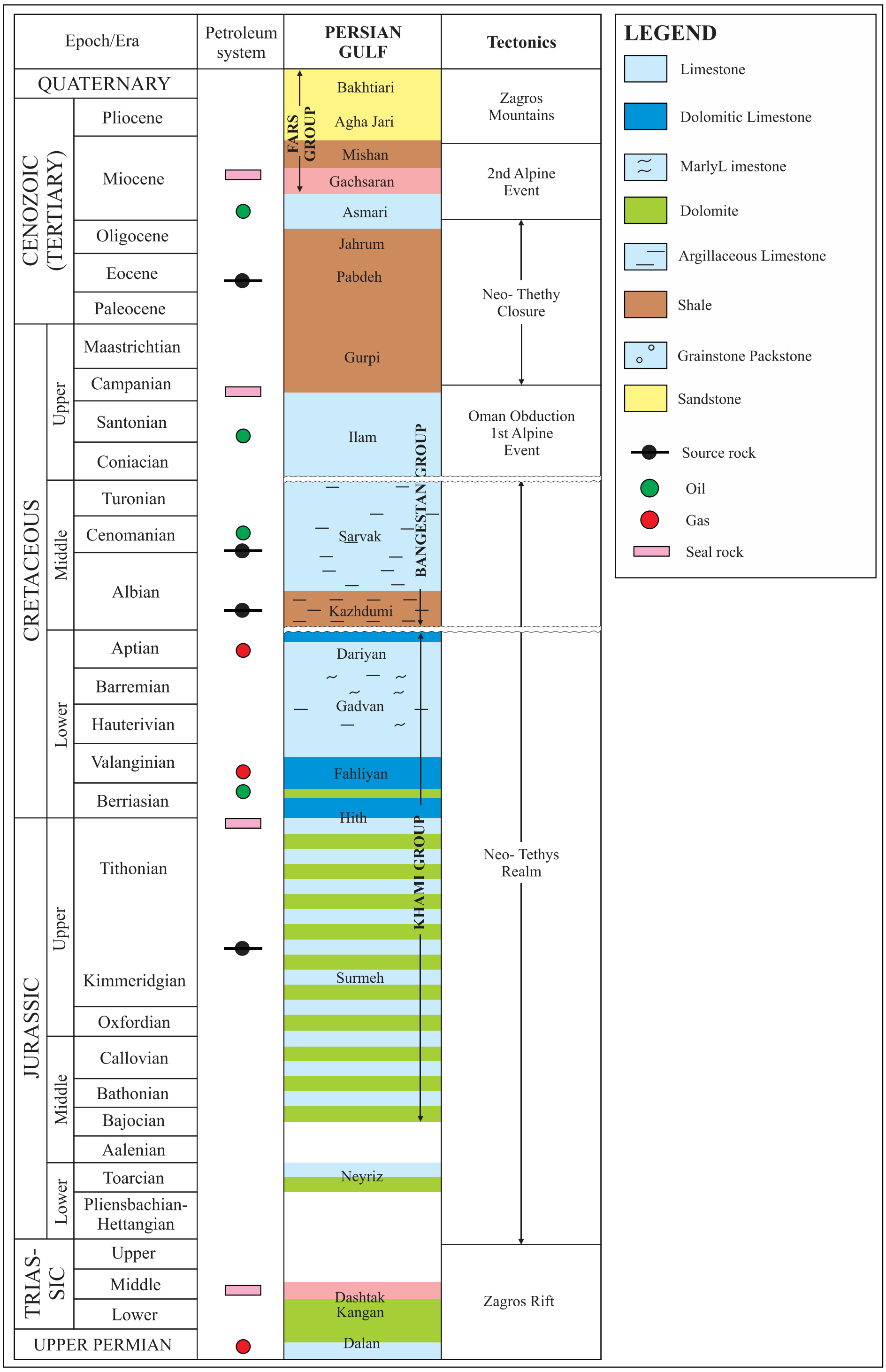

Figure 2

The generalized stratigraphic column of the Persian Gulf region (modified after Sharland et al., 2001).

During Cretaceous, three main stratigraphic sequences (Figure 2) are recorded in the Persian Gulf area: the Lower Cretaceous deposits of Fahliyan, Gadvan, and Dariyan formations; the Middle Cretaceous sediments constituting Kazhdumi and Sarvak formations; and the Upper Cretaceous deposits of Ilam, Laffan, and Gurpi formations (Harris et al., 1984; Jordan et al., 1985). A regional unconformity marks the end of the Cretaceous and the boundary between the Late Cretaceous and Early Tertiary sediments (between Pabdeh and Gurpi formations) (Sharland et al., 2001). The Lower Tertiary in Figure 2 consists of Pabdeh and Jahrum formations, followed by bioclastic carbonate sediments of the Oligocene Asmari formation. These sediments are followed by evaporate deposition during Middle Miocene (Gachsaran formation). The orogenic folding of the adjacent Zagros during the Late Tertiary resulted in rapid uplift and extensive eroded deposits (Alsharhan and Nairn, 1995), which formed the thick clastic deposits of the Zagros growth (Figure 2).

In this study, the drill cutting samples of 20 different stratigraphic levels (Table 1) drilled in 15 different fields (Figure 1) were statistically analyzed in terms of their bulk organic parameters.

Table 1

List of formations, their sampling location, and the number of samples used in this study (see Figure 1 for the location of the oilfields).

Age (Ma)	Formation	Field (number of samples)	Total number of samples
245	Kangan	G(1)	1
172	Neyriz	H(2)	2
20	Asmari	K(2)	2
15	Gachsaran	K(1), N(3)	4
135	Fahliyan Middle	C(2), D(1), A(1)	4
175-230	Dashtak	H(5)	5
145	Hith	A(2), J(5)	7
80	Laffan	F(2), I(1), N(5)	8
122	Gadvan	A(4), B(5)	9
125	Dalan	G(10)	10
35	Jahrum	C(5), N(2), D(6)	13
80	Ilam	I(3), K(6), O(4), N(1)	14
140	Fahliyan Lower	C(1), B(7), J(6)	14
130	Fahliyan Upper	G(7), D(3), A(3), B(6)	19
100	Kazhdumi	C(5), F(6), I(2), D(4), B(2), L(3), M(5), J(2)	29
115	Dariyan	C(1), F(8), G(1), D(6), A(6), B(2), H(1), L(4), M(15)	44
155	Surmeh	G(9), D(30), H(4), J(22), E(12)	77
90	Sarvak	F(15), L(21), G(1), O(6), N(4), A(3), H(1), J(3), M(3)	57
60	Gurpi	I(4), K(25), O(16), N(38), O(2), L(5), M(4)	94
40	Pabdeh	I(9), K(36), O(14), N(52), D(1), L(6), M(3)	121

3. Material and methods

3.1. Samples and analytical methods

A total of 534 drill cutting samples, collected from various stratigraphic units penetrated in 15 producing fields, were analyzed by a Vinci Rock-Eval 6 apparatus. The instrument was set to operate under analytical conditions published elsewhere (Peters et al., 2005). Briefly, aliquots of pulverize rock samples are heated at 300 °C for 3 minutes to release the free hydrocarbons present in the rock (S_1 parameter). Subsequently, the oven temperature is ramped to 650 °C at a rate of 25 °C/min to release the pyrolyzable fraction of the organic matter (S_1 parameter). The temperature at the maximum yield of pyrolyzable hydrocarbons is recorded as T_{max}. In addition, several geochemical parameters are calculated based on the obtained parameters (e.g., hydrogen index, oxygen index, production index, total organic carbon, residual carbon, etc.). Nine important parameters including *TOC*, *HI*, *OI*, S_1, S_2, S_3, T_{max}, Production Index (*PI*), and Residual Carbon (*RC*) were used for chemometric analyses in the current study (see Figure 4c). Standard guidelines for the application of these data in geochemical interpretations are listed in Table 2.

Table 2

Guidelines used for source rock interpretation (Peters et al., 2005; Peters and Cassa, 1994).

Potential	*TOC* (wt.%)	<1.0	1.0-2.0	2.0-3.0	3.0-4.0	>4.0
	S_2 (mg HC/g rock)	<2.0	2.0-5.0	5.0-10.0	10.0-20.0	>20.0
		Poor	Moderate	Good	Very Good	Excellent
Generation	*HI* (mg HC/g TOC)	<50	50-200	200-400	400-600	>600
		Inert	Gas-prone	Mix	Oil-prone	Very oil-prone
Maturation	T_{max} (°C)	<435	435-440	440-450	450-165	>465
		Immature	Early-Mature	Peak oil	Wet gas	Dry gas

3.2. Software and statistical analysis

Pirouette 4.5 software (Infometrix, Bothell, USA) has been used for chemometric classifications by means of the exploratory HCA and PCA algorithms. A preprocessing step is required before exploratory analyses since multivariate algorithms rely on variance patterns within the independent variables. In this study, the “auto scale” preprocessing option was used, which is simply mean-centering followed by variance scaling.

The HCA method operates by calculating and comparing the distance between pairs of samples; in fact, when distances between samples are small, this implies that the samples are similar. The most common system to compute distance in multivariate analyses is the “Euclidean distance”, which is also applied to this study. The primary purpose of HCA is to define natural groupings in the dataset and present them in the form of a dendrogram. After the distances between all the pairs of samples are computed, the two most similar samples are linked as a single cluster. Individual clusters are linked to each other by seeking the smallest inter-cluster distance. There are different linkage methods for doing the HCA analysis. The “incremental linkage” method works better than other methods where groups of samples differ only slightly. Therefore, considering the nature of the geochemical parameters used in this study, this method is employed for cluster definition.

The PCA method is a powerful exploratory analysis to reduce the effective dimensionality of the data. The algorithm is based on finding the linear combination of variables that account for maximal amounts of variation, i.e. the principal components or factors shown in loading plots. The loadings represent the relative contribution of each geochemical parameter to a principal component axis. PCA results are used to ascribe geochemical meanings to the data patterns emerging in the dataset. For additional information about various functionalities of the Pirouette software and analytical procedures, the reader is referred to the software’s user guide.

Following processing by both algorithms, three main objects were displayed inside the Pirouette work space: a main object from the HCA (the dendrogram), and two objects from the PCA (the preprocessed data and the scores). The number of clusters was determined by the evaluation of the groupings in both the dendrogram and the scores. For the initial analysis, performed on the full data set, a similarity of 0.4 is chosen, and 3 separate families are tentatively identified. These families are separately analyzed by performing HCA and PCA analyses, and additional subgroupings are identified.

4. Results and discussion

The first step in doing a chemometric analysis is to identify and eliminate problematic samples. From a total of 534 rock samples used for the purpose of this study, only one sample (S-613) is considered to be an outlier based on a line-plot view of the full data in Pirouette (Figure 3). This sample with a TOC of 0.05 wt.% and an OI value of 1481 (mg CO_2/g TOC) can be problematic and is therefore removed from the dataset before starting the analyses.

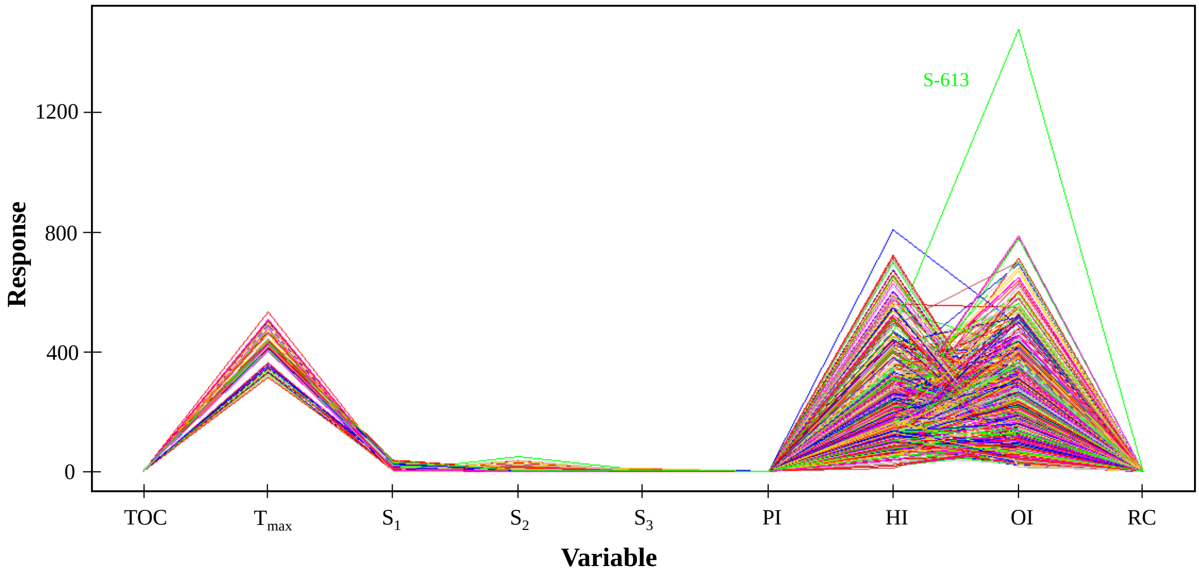

Figure 3

The line-plot view of the full data to identify the problematic samples before beginning the exploratory analyses.

4.1. Main families

Using the HCA results and the two PCA objects mentioned above, a similarity degree of 0.4 was chosen for the initial grouping of the samples. In general, a similarity level between 0.4 and 0.6 is preferable for interpreting dendrograms. However, under certain circumstances, where a clear differentiation of clusters is needed, a similarity level of 0.2 may be justified (see below). There appear to be three main groupings in the collection of 534 samples in our dataset (Figure 4a).

Experience has shown that the most convenient means of exploring sample groupings in complex datasets is through the “divide-and-conquer” approach. In this technique, we begin by breaking the dataset into a small number of main families (in this case 3) and subsequently run separate exploratory analyses on each family to reveal its inherent subgroupings. In this manner, various subfamilies and sub-subfamilies are defined under each main family (Figure 5).

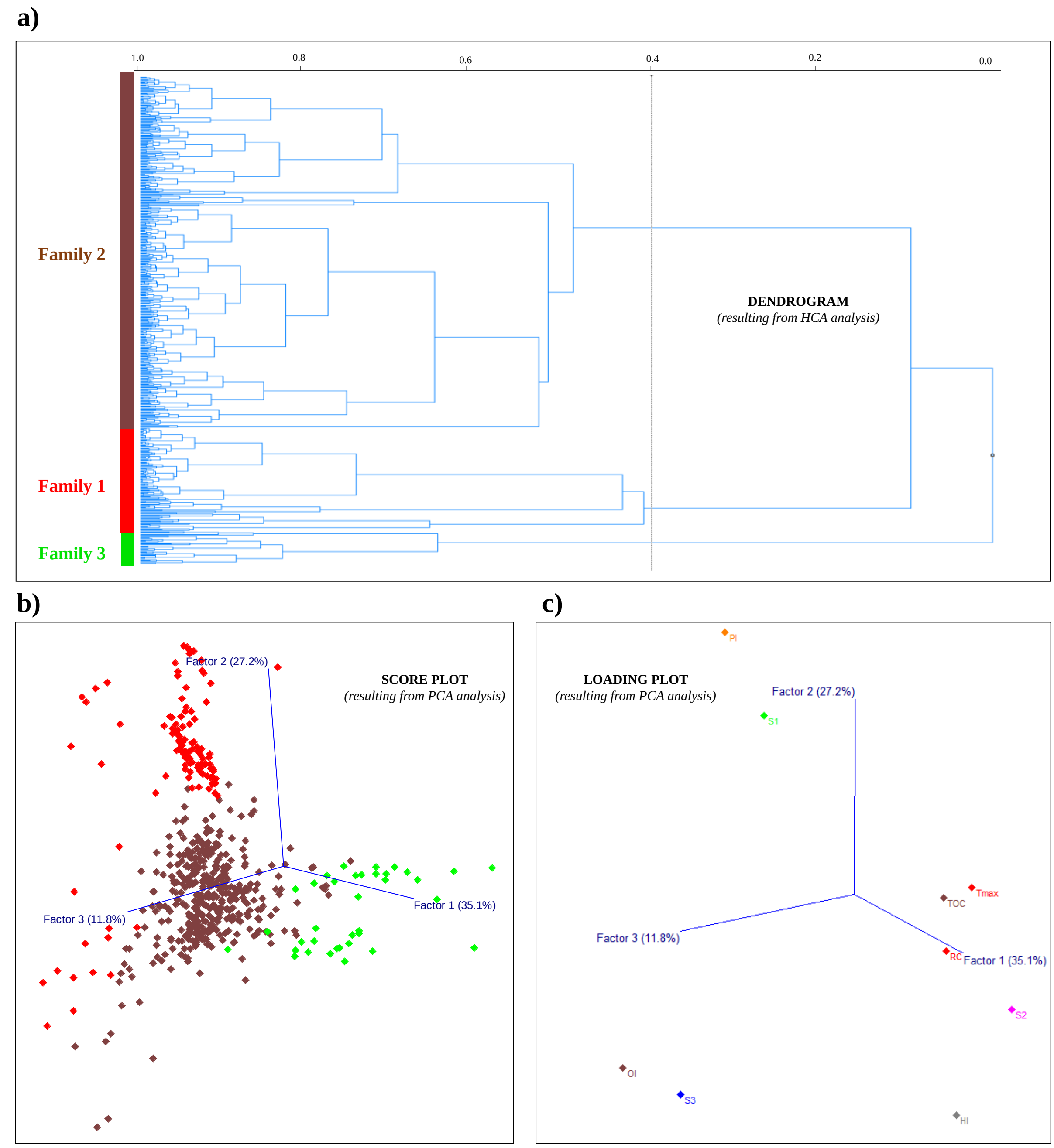

Figure 4

HCA dendrogram showing a) the 3 main families identified within the full dataset, b) the corresponding scores plot, and c) the loadings.

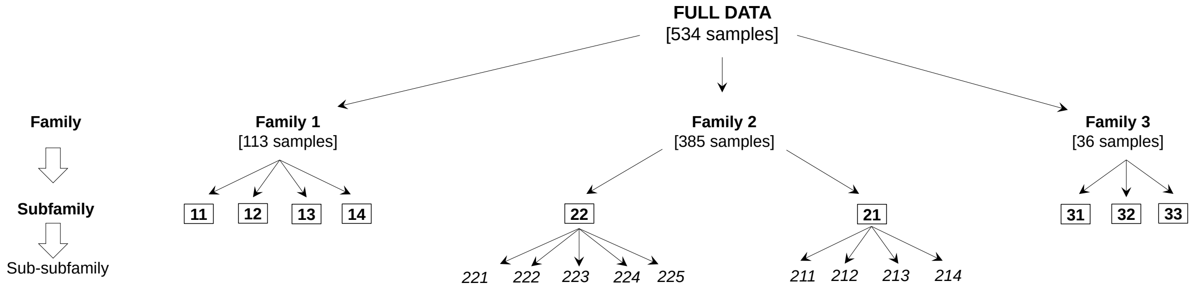

Figure 5

Chemometric analysis tree diagram, indicating the main families, subfamilies, and sub-subfamilies obtained from the study of the full dataset.

The discussions presented in the remainder of this paper are narrowed down to the main families in order to better elucidate their underlying geochemical characteristics.

a. Family 1

Family 1 includes the red-colored samples which are generally described by Factor 2 (Figure 4b). The implication could be that Family 1 samples have a relatively high percentage of free hydrocarbons in their matrix (Figure 4c). However, some Family 1 samples are cast onto Factor 3, which might indicate the presence of oxidized organic matters (Figures 4b and 4c). Therefore, further exploratory analysis of this family could be useful for accurate characterization. To this end, Family 1 samples are independently examined by exploratory HCA and PCA analyses, which reveals the presence of three individual subfamilies as shown by the dendrogram presented in Figure 6a.

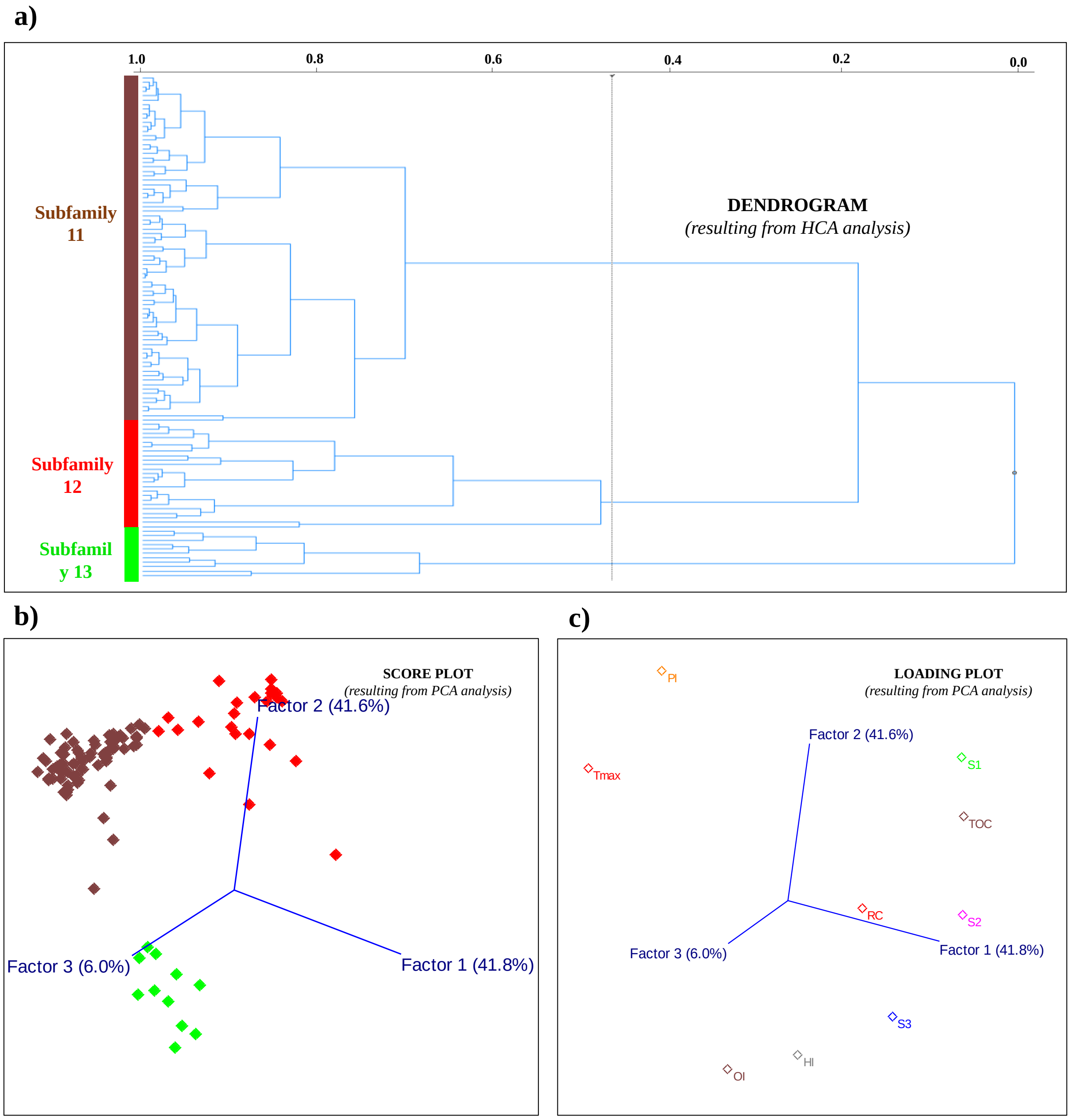

Figure 6

a) Family 1 HCA dendrogram and PCA results including b) the scores plot and c) the loadings.

As mentioned above, Family 1 samples are typically rich in free hydrocarbons in their matrix, so most related subfamilies share the common feature of having high free hydrocarbons (red and brown samples in Figure 6b). However, it is known that large concentrations of free hydrocarbons are likely to be a result of ongoing hydrocarbon generation within source rocks. Interestingly, Family 1 samples indicating appropriate thermal maturities (those grouped under subfamily 11) could be differentiated from their immature counterparts (subfamily 12 samples) (Figures 6b and 6c). Therefore, the red-colored samples representing subfamily 12 are very likely contaminated by organic-based drilling fluids; in other words, care should be taken when using their analytical data in geochemical interpretations. However, we suggest that additional information about the burial/temperature histories of these units, in conjunction with appropriate kinetic data, be needed for better geochemical interpretations. Subfamily 13 represents samples containing oxidized organic matters in their matrix and may have the negligible potential for hydrocarbon in the studied area (Figures 6b and 6c).

b. Family 2

Family 2 constitutes the main part of the dataset and includes brown-colored samples in Figure 4a. These samples occupy the central position on the scores plot (Figure 4b). However, some information about the structure of Family 2 samples can be found in higher factors (such as Factor 3 and 5) (see Figure 4b). Therefore, additional exploratory analyses are critical to the better separation of existing subfamilies, as will be discussed in the following.

The independent analysis of Family 2 samples by exploratory (HCA and PCA) methods has indicated the presence of two main subfamilies (Figure 7). However, these subfamilies require additional characterization due to their complex structures (see below).

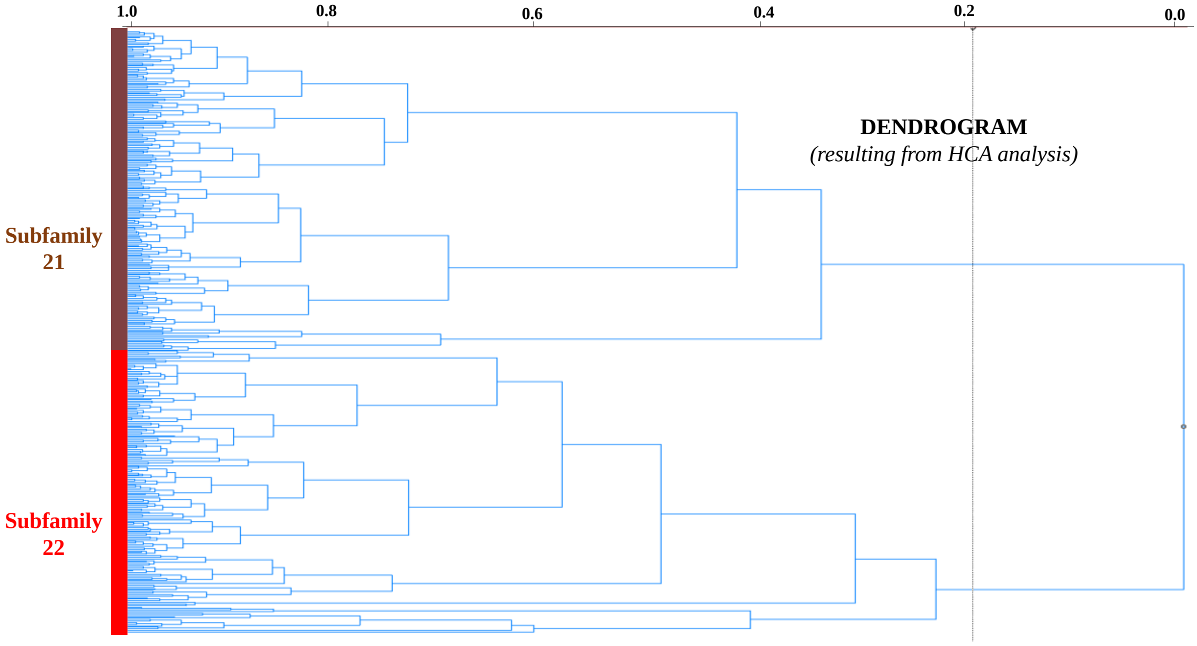

Figure 7

Family 2 classification and the resulting subfamilies; it should be noted that, due to the complex structure, these subfamilies require independent analysis.

Subfamily 21 generally comprises samples from younger geological formations with moderate source rock potential using the guidelines listed in Table 2. The exploratory analysis of this subfamily has identified four sub-subfamilies (Figure 8a). According to the scores plot (Figure 8b), we can make some predictions about the geochemistry of separated sub-subfamilies. Magenta samples can represent source rock layers with the noticeable potential for petroleum since they show strong loadings on HI, TOC, and S_2 parameters (Figure 8c). However, the contained organic matter is likely deposited under more oxic conditions since these samples are similarly loaded into S_3 parameter. On the other hand, green-colored samples can represent rock intervals with high residual carbon contents. The brown-colored samples can generally represent rock units containing elevated concentrations of free hydrocarbons. Finally, the red-colored samples are supposed to contain oxidized organic matters due to the strong loading on OI parameter (Figure 8b).

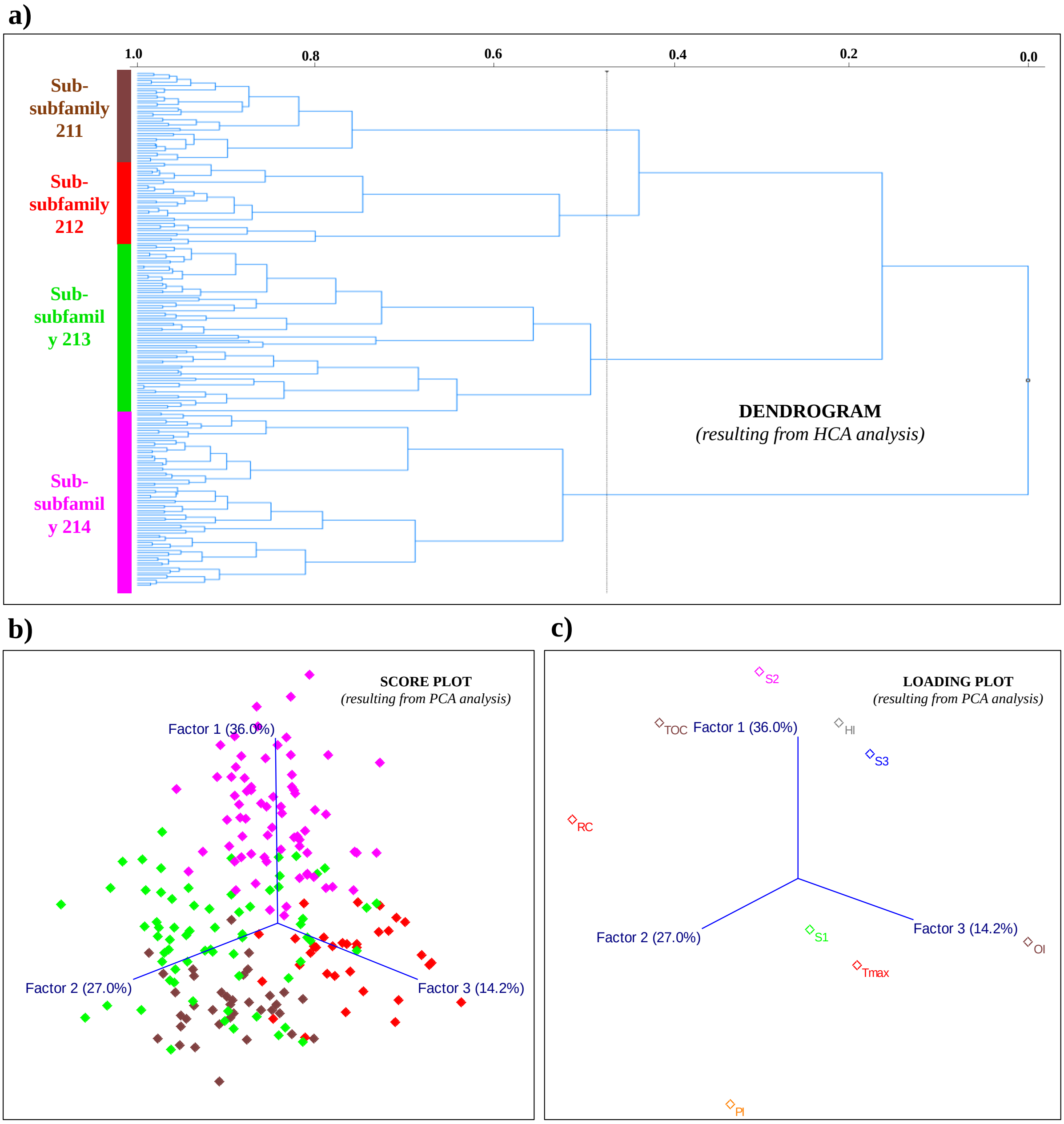

Figure 8

Subfamily 21 classification indicating a) the HCA dendrogram, b) the scores plot, and c) the loadings.

The samples of subfamily 22 are separately studied by exploratory HCA and PCA analyses, and 5 sub-subfamilies are identified (Figure 9a). The blue samples represent rock intervals with relatively good potential for source rock (Figure 9b) since they are loaded into HI, S_2, and TOC parameters (Figure 9c). In contrast, the magenta and brown-colored samples are likely to contain oxidized organic matters, i.e. high loadings into OI and S_3 parameters in Figure 9c. The red and green samples occupy the central parts of the scores plot and may represent the rock intervals of moderate source rock potential. However, the green-colored samples have relatively higher PI readings compared to the red samples.

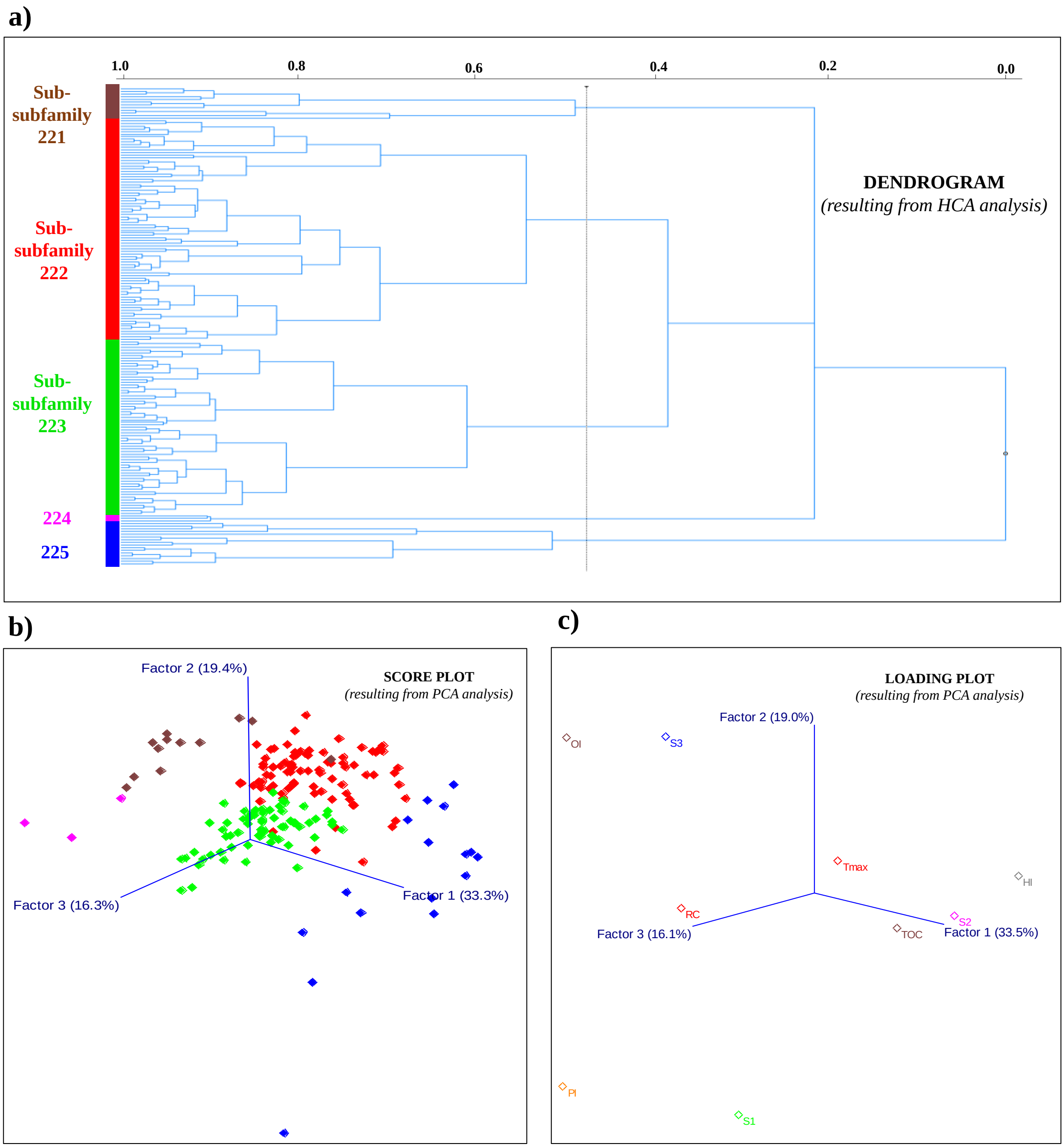

Figure 9

Detailed analysis of subfamily 22 indicating a) the HCA dendrogram, b) the scores plot, and c) the loadings.

c. Family 3

Family 3 includes the green samples described by Factor 1 on the scores plot of Figure 4b. Considering the reference to the loading plot on one side (Figure 4c) and the general source rock interpretation guide on the other side (Table 2), the samples in this family might represent good-quality source rock intervals in the area of study. The further characterization of Family 3 samples by the independent exploratory (HCA and PCA) analyses has better revealed its geochemical characteristics (Figure 10).

It is noteworthy that most Family 3 samples generally belong to a single stratigraphic unit in the area of study. However, three main subfamilies (31, 32, and 33) could be detected in this family as shown by the dendrogram in Figure 10a. A closer inspection of the scores plot and the loadings in Figure 10 indicates that the red-colored samples have a higher loading into T_{max}, S_1, and PI parameters (Figure 10c). From a geochemical viewpoint, these conditions may imply mature source rocks with considerable hydrocarbon saturation. On the other hand, the spatial arrangement of the brown-colored samples could convey rock units containing high-quality organic matters with relatively lower maturity. Finally, the green-colored samples might represent rock intervals with relatively more oxic facies (Figures 10b and 10c).

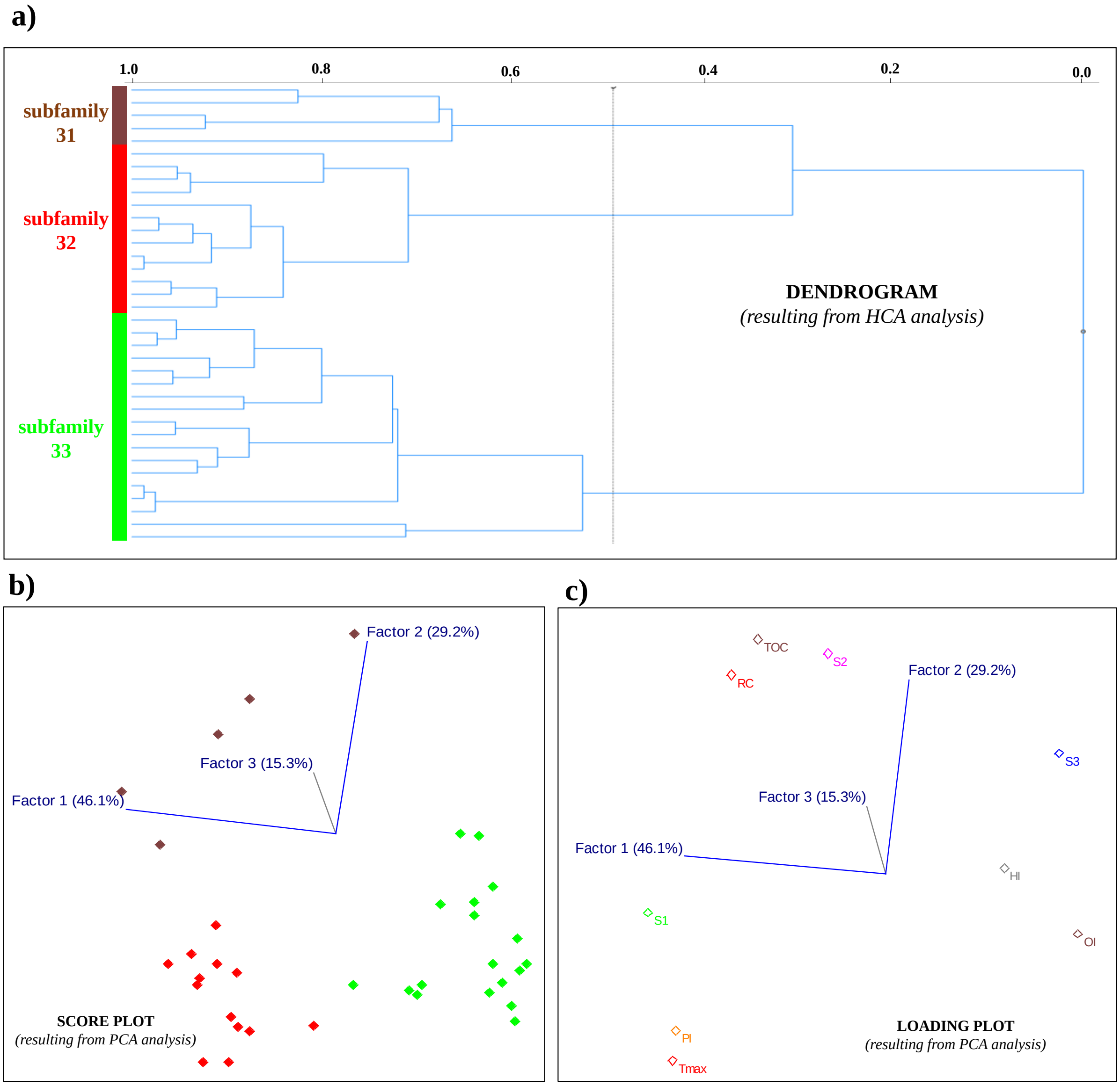

Figure 10

Family 3 analysis showing a) the dendrogram, b) the scores plot, and c) the loadings.

The results of this study show that chemometric methods can enhance the classification and interpretation of bulk geochemical data. It should be emphasized that manual techniques may provide misleading results because they usually rely on applying certain cut-off thresholds to individual parameters such as filtering samples with higher S_1 readings. The application of chemometric methods to a collection of 534 samples in this study greatly helped to discern the structure of the dataset as a whole. By the careful examination of the relationships between the samples and the variables, we could point out geochemical reasons which give rise to the structure of the dataset. This, in turn, greatly facilitated the sample screening procedures and provided strong feedback for identifying key source rock levels in the studied area.

Table 3 summarizes the results of the exploratory analyses performed in this study with the geochemical significance of each sample subgrouping highlighted. The sample groupings tagged with high concentrations of oxidized organic matters (e.g., 13, 212, 214, 221, and 224) or residual carbon (e.g., 213) in their matrix can be safely discarded. Similarly, the samples contaminated by drilling additives (e.g., 11, 12, 211, and 223) should be excluded from further considerations. On the other hand, the sample groupings interpreted as moderate/good source rocks are critical candidates for petroleum system modeling studies; in other words, their stratigraphic/geographic distributions should be mapped.

Accordingly, it has been suggested that chemometric methods should be superior to conventional screening procedures because their results can a) provide a preview of the structure of geochemical datasets, b) guide us into focusing our attention on potential source rock intervals in the study area, c) identify information gaps, d) take the necessary precautions of data interpretations, and e) provide help for selecting samples for complementary analyses.

Table 3

Results from the exploratory analyses of the full dataset with the geochemical interpretations on sample groupings.

Family	Subfamily		Geochemical significance	Interpretation
Family 1	11		Contamination (relatively high T_{max})	Further investigation
	12		Contamination	Reject
	13		Oxidized organic matters	Reject
Family 2	21	211	Contamination	Reject
		212	Oxidized organic matters	Further investigation
		213	Non-source rock interval (relatively high RC)	Reject
		214	Moderate source rock (more oxic facies)	Further investigation
	22	221	Oxidized organic matters	Reject
		222	Moderate source rock	Accept
		223	Moderate source rock (relatively high PI)	Further investigation
		224	Oxidized organic matters	Reject
		225	Good source rock	Accept
Family 3	31		Good source rock (relatively low T_{max})	Accept
	32		Good source rock (relatively high T_{max})	Accept
	33		Good source rock (more oxic facies)	Accept

5. Conclusions

The chemometric-enhanced classification of source rock samples using their bulk geochemical parameters is performed for the first time on a total of 534 samples from the Persian Gulf basin. After a detailed exploratory study by HCA and PCA methods, three main families and a number of smaller subdivisions were identified in the sample set. According to our results, most of Family 1 samples are characterized by excess amounts of free hydrocarbons with a high probability of being contaminated. In addition, Family 2 samples are generally characterized by moderate/good potential for source rocks and, in some cases, belong to strongly oxic depositional settings. However, Family 3 samples represent good source rocks which are subjected to varying levels of thermal maturation due to differing depths of burial. The results of this novel classification approach proved useful for a) clarifying the groupings of the samples, b) obtaining a big-picture view of the existing geochemical variations, c) identifying potential source rock layers, and d) gaining information about the quality of key source rock units before starting basin and petroleum system models. More importantly, the analysis of large datasets by multivariate techniques sidesteps the issues related to the subjectivity of manual data filtration. Therefore, an initial chemometric analysis prior to the ordinary graphical representation of bulk Rock-Eval parameters can tell us about the main sample groupings inherent in the database.

Acknowledgements

National Iranian Offshore Oil Company (IOOC) is acknowledged for providing the data. Petroleum Geology and Geochemistry Research Center (PGGRC) of Shahid Chamran University of Ahvaz is also acknowledged for scientific cooperation. The authors are sincerely thankful to the Infometrix licensing group, in particular Paul Bailey, for providing a static license of Pirouette® 4.5 for the purpose of this study. The authors also acknowledge the useful comments of anonymous reviewers on an earlier version of this manuscript, which greatly improved the quality of the work.

Nomenclature

2D modeling	Two dimensional modeling
HCA	Hierarchical cluster analysis
HI	Hydrogen index
OI	Oxygen index
PCA	Principal component analysis
PI	Production index
RC	Residual carbon
S_1	Free hydrocarbons released up to 300 °C within a pyrolysis oven
S_2	Generative (i.e. remaining) potential
SIMCA	Soft independent modeling of class analogy
T_{max}	Pyrolysis oven temperature at maximum hydrocarbon yield
TOC	Total organic carbon

References

Alipour, M., Organic Geochemistry of Source Rocks and Unconventional Resources; Habitat and Alteration of Hydrocarbons in Block a of the Persian Gulf, Shahid Chamran University of Ahvaz, p. 245, 2017.

Alipour, M., Alizadeh, B., and Chehrazi, A., A Thermal Maturity Analysis of the Effective Cretaceous Petroleum System in Southern Persian Gulf Basin, Iranian Journal of Oil and Gas Science and Technology, Vol. 6, p. 1-17, 2017.

Alipour, M., Alizadeh, B., Chehrazi, A., Mirshahani, M., and Khani, B., Sequence Stratigraphic Control on Active Petroleum System in the Eastern Block A, Persian Gulf, the 1st International Conference on Science and Basic Research, Kharazmi Higher Institute of Science and Technology, Iran, p. 151-153, 2016a.

Alipour, M., Alizadeh, B., Chehrazi, A., Mirzaie, S., Shakib, S., Khani, B., and Ramos, L.S., Descriptive Source Rock Quality Mapping Based on Chemometric Classification of Bulk Geochemical Data, Persian Gulf Basin, 78th EAGE Conference and Exhibition, Vienna, Austria, 2016b.

Alizadeh, B., Alipour, M., Chehrazi, A., and Mirzaie, S., Chemometric Classification and Geochemistry of Oils in the Iranian Sector of the Southern Persian Gulf Basin, Organic Geochemistry, Vol. 111, p. 67-81, 2017.

Alsharhan, A., Petroleum Geology of the United Arab Emirates, Journal of Petroleum Geology, Vol. 12, p. 253-288, 1989.

Alsharhan, A. and Nairn, A., Tertiary of the Arabian Gulf: Sedimentology and Hydrocarbon Potential, Paleogeography, Paleoclimatology, Paleoecology, Vol. 114, p. 369-384, 1995.

Alsharhan, A.S. and Kendall, C.G.S.C., Precambrian to Jurassic Rocks of Arabian Gulf and Adjacent Areas: their Facies, Depositional Setting, and Hydrocarbon Habitat, American Association of Petroleum Geologists Bulletin, Vol. 70, p. 977-1002, 1986.

Fabiańska, J.M., GC-MS Investigation of Distribution of Fatty Acids in Selected Polish Brown Coals, Chemometrics and Intelligent Laboratory Systems, Vol. 72, p. 241-244, 2004.

Farrimond, P., Stoddart, D., and Jenkyns, H., An Organic Geochemical Profile of the Toarcian Anoxic Event in Northern Italy, Chemical Geology, Vol. 111, p. 17-33, 1994.

Ghazban, F., Petroleum Geology of the Persian Gulf, Tehran University, Tehran, Iran, 2007.

Grimalt, J.O., Canton, L., and Olive, J., Source Input Elucidation in Polluted Coastal Systems by Factor Analysis of Sedimentary Hydrocarbon Data, Chemometrics and Intelligent Laboratory Systems, Vol. 18, p. 93-109, 1993.

Harris, P., Frost, S., Seiglie, G., and Schneidermann, N., Regional Unconformities and Depositional Cycles, Cretaceous of the Arabian Peninsula, in: Schlee, J.S. (Ed.), Interregional Unconformities and Hydrocarbon Accumulation (Memoir 36), American Association of Petroleum Geologists, p. 67-80, 1984.

Hughes, W., Holba, A., Miller, D., and Richardson, J., Geochemistry of Greater Ekofisk Crude Oils, in: Thomas, B.M. (Ed.), Petroleum Geochemistry in Exploration of the Norwegian Shelf. Springer, p. 75-92, 1985.

Jordan, C.F., Connally, T.C., and Vest, H.A., Middle Cretaceous Carbonates of the Mishrif Formation, Fateh Field, Offshore Dubai, UAE, in: Roehl, O.P., Choquette, P.W. (Eds.), Carbonate Petroleum Reservoirs. Springer, p. 425-442, 1985.

Kent, P., The Emergent Hormuz Salt Plugs of Southern Iran, Journal of Petroleum Geology, Vol. 2, p. 117-144, 1979.

Kowalski, B., Schatzki, T., and Stross, F., Classification of Archaeological Artifacts by Applying Pattern Recognition to Trace Element Data, Analytical Chemistry, Vol. 44, p. 2176-2180, 1972.

Kruge, M.A., Determination of Thermal Maturity and Organic Matter Type by Principal Components Analysis of the Distributions of Polycyclic Aromatic Compounds, International Journal of Coal Geology, Vol. 43, p. 27-51, 2000.

Kvalheim, O.M., Oil-Source Correlation by the Combined Use of Principal Component Modelling, Analysis of Variance and a Coefficient of Congruence, Chemometrics and Intelligent Laboratory Systems, Vol. 2, p. 127-136, 1987.

Lasemi, Y. and Jalilian, A., The Middle Jurassic Basinal Deposits of the Surmeh Formation in the Central Zagros Mountains, Southwest Iran: Facies, Sequence Stratigraphy, and Controls, Carbonates and Evaporites, Vol. 25, p. 283-295, 2010.

Murris, R.J., Middle East Stratigraphic Evolution and Oil Habitat, American Association of Petroleum Geologists, Special Volumes, p. 353-372, 1980.

Nasir, S., Al-Saad, H., Alsayigh, A., and Weidlich, O., Geology and Petrology of the Hormuz Dolomite, Infra-Cambrian: Implications for the Formation of the Salt-cored Halul and Shraouh Islands, Offshore, State of Qatar, Journal of Asian Earth Sciences, Vol. 33, p. 353-365, 2008.

Oberrauch, E., Salvatori, T., Novelli, L., and Clementi, S., Oils of the Po Basin: A Geochemical-chemometric Study, Chemometrics and Intelligent Laboratory Systems, Vol. 2, p. 137-147, 1987.

Peters, K.E., Coutrot, D., Nouvelle, X., Ramos, L.S., Rohrback, B.G., Magoon, L.B., and Zumberge, J.E., Chemometric Differentiation of Crude Oil Families in the San Joaquin Basin, California, American Association of Petroleum Geologists Bulletin, Vol. 97, p. 103-143, 2013.

Peters, K.E., Curry, D.J., and Kacewicz, M. Basin Modeling: New Horizons in Research and Applications, American Association of Petroleum Geologists, Houston, Texas, , 2012.

Peters, K.E., Ramos, L.S., Zumberge, J.E., Valin, Z.C., Scotese, C.R., and Gautier, D.L., Circum-Arctic Petroleum Systems Identified Using Decision-tree Chemometrics, American Association of Petroleum Geologists Bulletin, Vol. 91, p. 877-913, 2007.

Peters, K.E., Walters, C.C., and Moldowan, J.M., The Biomarker Guide: Biomarkers and Isotopes in the Environment and Human History, Cambridge University Press, New York, 2005.

Ramos, L.S., Beebe, K.R., Carey, W.P., Sanchez M, E., Erickson, B.C., Wilson, B.E., Wangen, L.E., and Kowalski, B.R., Chemometrics, Analytical Chemistry, Vol. 58, 294R-315R, 1986.

Sharland, P.R., Archer, R., Casey, D.M., Davies, R., Hall, S.H., Heward, A.P., Horbury, A.D., and Simmons, M., Arabian Plate Sequence Stratigraphy, Gulf Petrolink, Manama Bahrain, 2001.

Telnæs, N. and Cooper, B.S., Oil-source Rock Correlation Using Biological Markers, Norwegian Continental Shelf, Marine and Petroleum Geology, Vol. 8, p. 302-310, 1991.

Telnaes, N. and Dahl, B., Oil-oil Correlation Using Multivariate Techniques, Organic Geochemistry, Vol. 10, p. 425-432, 1986.