

APPLICATION OF MULTIVARIATE STATISTICAL ANALYSIS TO BIOMARKERS IN SE-TURKEY CRUDE OILS

K. Gürgey ^{*1}, S. Canbolat ¹

¹ Department of Petroleum and Natural Gas Engineering Near East University, Near East Avenue, 99138, Nicosia, Mersin 10, Turkey - (kadir.gurgey, serhat.canbolat)@neu.edu.tr

KEY WORDS: Geostatistics, Biomarker, Terpane, Crude oil, SE-Turkey

ABSTRACT:

Twenty-four crude oil samples were collected from the 24 oil fields distributed in different districts of SE-Turkey. API and Sulphur content (%), Stable Carbon Isotope, Gas Chromatography (GC), and Gas Chromatography-Mass Spectrometry (GC-MS) data were used to construct a geochemical data matrix. The aim of this study is to examine the genetic grouping or correlations in the crude oil samples, hence the number of source rocks present in the SE-Turkey. To achieve these aims, two of the multivariate statistical analysis techniques (Principle Component Analysis [PCA] and Cluster Analysis) were applied to data matrix of 24 samples and 8 source specific biomarker variables/parameters. The results showed that there are 3 genetically different oil groups: Batman-Nusaybin Oils, Adiyaman-Kozluk Oils and Diyarbakir Oils, in addition to a one mixed group. These groupings imply that at least, three different source rocks are present in South-Eastern (SE) Turkey. Grouping of the crude oil samples appears to be consistent with the geographic locations of the oils fields, subsurface stratigraphy as well as geology of the area.

INTRODUCTION

SE-Turkey embodies the northwest end of the Persian Gulf sedimentary basin and covers an area of 90,000 km² (Fig. 1). It is the extension of Zagros oil province and the most significant oil prone basin of Turkey. Types of oils in SE-Turkey vary from the heavy oils (12-25 API gravity) to medium-light oils (25-36 API gravity), (Gürgey, 1991). Several oil fields have been discovered since the discovery of the Raman oil field (e.g., Located in Batman area; Fig. 1) in 1954. Currently, Turkey as an import-dependent country for energy, imports over 90 percent of its crude oil needs.

Considering this, it is important for Turkey to develop more scientific models on the discovered oil fields which should help eventually to discover new oil fields. In this sense, “petroleum system” model studies of the basins were proved to serve the petroleum geologists and geophysicists to make help new oil discoveries. On the other hand, a reliable genetic grouping and correlation of the discovered oils is one of the prerequisites of petroleum system studies (Zumberge, 1987). Hence, the aim of this study is to develop a genetic correlations model of the SE-Turkey crude oils and to recognize their number of source rocks. To achieve these aims, multivariate statistical techniques such as principle component analysis (PCA) and cluster analysis were used.

1. METHODOLOGY

2.1 Sampling

In this study, a total of 24 crude oil samples were collected from the 24 different oil fields located at the different districts of SE-Turkey (Fig.1). The samples were collected at the wellhead at atmospheric conditions. In order to prevent possible reactions between the container and the fresh crude oil samples, a glass jar with teflon cover was used. The samples were carried to TPAO (Turkish Petroleum Corporation Research Center) within an icy box.

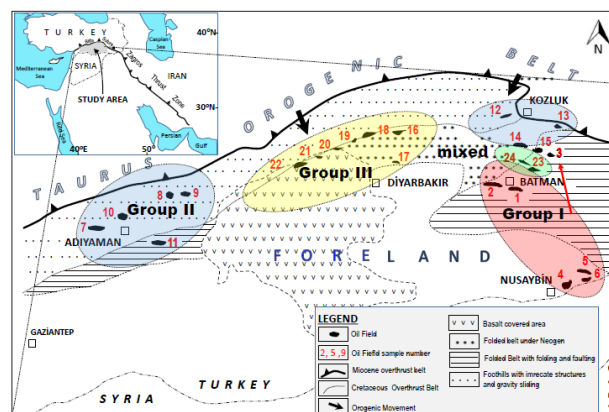


Figure 1. Location map SE-Turkey and collected crude oil samples from 24 different oil fields. Genetic grouping of the oils on the basis of this study is also shown.

2.2 Experimental work

API gravity and sulphur content (%) measurements were performed by using Carl Zeiss-89578 model Abbe refractometer and atomic BaSO₄ method, respectively. Furthermore, gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS) analyses were conducted by Varian 3700 Model FID capillary gas chromatograph and Hewlett Packard 5988A Model mass spectrometer, respectively. All the analyses were performed at the TPAO Research Centre laboratory, Ankara whereas available carbon isotope values of the saturate and aromatic fractions were obtained from the foreign oil companies.

2.3. Methods

In this study, PCA and cluster analyses methodology of the 2009 WinSTAT statistic software were applied to the geochemical data matrix. The data consists of 24 crude oil samples and 8 source

specific variables (24X8) of which 6 variables are molecular-biomarker ratios. These are calculated by measuring peak heights on the gas (Pr/Ph; pristane/phytane ratio) and mass chromatograms (Fig.2; C_{24}^*/C_{26} , $T_m+T_s/C_{28}+C_{29}$, $C_{29}NH/C_{30}H$, T_m/T_s and C_{23}/C_{24}^* ratios). In addition to these biomarker ratios, carbon isotope values of saturate ($\delta^{13}C_{sat}$) and aromatic fractions ($\delta^{13}C_{aro}$) were used to complete 24 X 8 data matrix. In this study, the variables selected are sensitive to source rock paleoenvironmental conditions. A dendrogram from cluster analysis as well as loading and score plots from PCA are then interpreted to reach conclusion about the number of genetic groups among the 24 SE-Turkey crude oil samples.

2. RESULTS AND DISCUSSION

All the oil samples were initially analysed for their API gravity (12-37°) and sulphur content (0.22-7.18%) for screening purposes. Then, selected oils were analyzed using GC and GC-MS. Stable carbon isotope ratios of saturates and aromatics range from $\delta^{13}C_{max} = -27.1$ to $\delta^{13}C_{min} = -28.5$ ‰ and $\delta^{13}C_{max} = -27.2$ to $\delta^{13}C_{min} = -28.5$ ‰, respectively. A representative m/z 191 terpane chromatogram of the Turkish crude oils is given in Figure 2. Each peak on this chromatogram is called a biomarker molecule ratios calculated by using these peaks are then used to construct the data matrix.

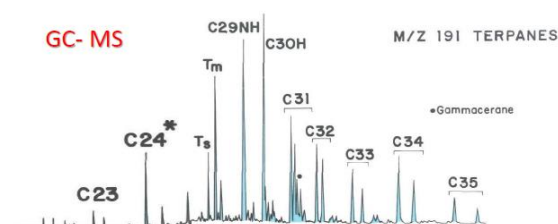


Figure 2. A representative m/z 191 chromatogram of a Turkish crude oil. Each peak on this chromatogram is a biomarker molecule ratios of which is the part of the geochemical data matrix of this study.

Variables/parameters used for grouping within the data matrix are specific to conditions of source rock depositional environment and are presumably not affected by the geochemical processes, such as migration, maturation, biodegradation and water washing (Moldowan et al., 1985). Cluster analysis can be used to investigate the relationship among geochemical samples. The method is used to find similarities among samples and to produce a graphic display (e.g., dendrogram) of how the samples are clustered. Its advantage over PCA is that clusters are assembled using all the variance in the data matrix whereas PCA carry information 60-90% typically represented by the first few PCs (Xue et al., 2011). Using this concept, three oil clusters e.g., groups) (Fig. 3). It is interesting to note that oil groups correspond to geographical locations of crude oil samples: Group I, II, and III oils are distributed mainly in the Batman-Nusaybin area, Adiyaman-Kozluk area and the Diyarbakir Area, respectively. Mixed group oils are take place between areas of group I and group II oils (Fig.1).

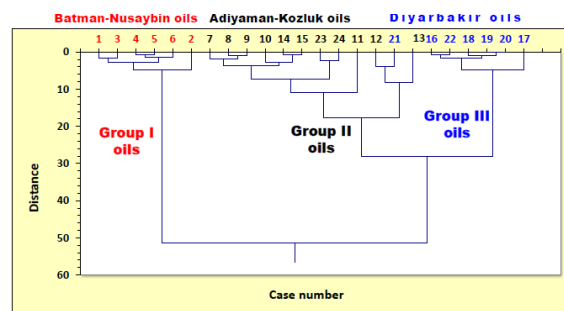


Figure 3. A dendrogram showing the genetic grouping of the SE-Turkey crude oils. The dendrogram is based on the 24 crude oils and 8 source specific variables. Oil numbers from 1 to 24 correspond the oil sample numbers in Fig.1

The PCA is a chemometric or multivariate statistical technique used to extract maximum amount of information from the data matrix and grouping of the samples on the PCs score plots (Zumberge, 1987). PCA is also useful method examining correlations among variables in the original data matrix on a newly selected axes. The projection of each variable on a new axis is called its “loading” that indicates the relative importance of each variable on that axis whereas the projection of each sample in the new axis is called its “score”. Hence, the classification may be done on the basis of scores and characteristic of each PC are established by the interpretation of loadings. Sample score plot describe the relative position of each sample in the PC space.

In this study, the PCA was applied to the Turkish crude oil data matrix (24X8) as used for the cluster analysis. As a result of that application, two PCs (e.g., principles components) were found: PC1 and PC2 which carry 59.79 % and 18.82 % of the total variance of the whole data matrix, respectively. Constructed PC1 versus PC2 score plot given in Fig.4 shows grouping of the SE-Turkey crude oils. As seen, this grouping of PCA is consistent with that of cluster analysis shown in Fig. 3. Importance of this grouping for petroleum exploration will be the subject of another study, mostly because the limited space.

In brief, application of the multivariate statistical analysis to the geochemical datasets is a powerful tool for the understanding of interrelations among crude oil samples.

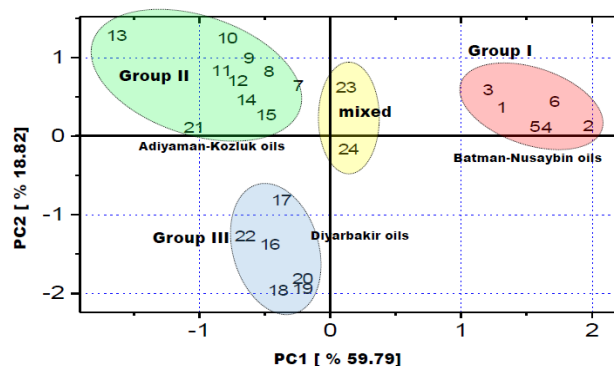


Figure 4. A score plot showing the genetic groupings between the SE-Turkey oils. Notice that PC1 and PC2 together carry the 78.61 % variance of the total data matrix. Numbers from 1 to 24 correspond the oil field/sample numbers in Fig.1.

3. CONCLUSIONS

Application multivariate statistical analyses to the geochemical data matrix of 24 SE-Turkey crude oil samples (e.g., collected from 24 different oil fields) and 8 variables let us to recognize three oil groups in addition to one mixed group. The group I and II oils are geographically distributed in the Batman-Nusaybin area and the Adıyaman-Kozluk area whereas the group III oils are distributed in the Diyarbakır area. Mixed group oils are geographically located between the group II Kozluk and group I Batman oils. Grouping appears to be consistent with the geographic locations of the oils fields, subsurface stratigraphy as well as geology of the area.

ACKNOWLEDGEMENTS

We are grateful to Turkish Petroleum Corporation (TPAO) for providing crude oil samples and related data and for the permission to published this data.

REFERENCES

- Gürgey, K., 1991. Genetic classification of the SE-Turkey crude oils and oils and delineation of source rock type with the use of biological markers. Middle East Technical University, Ankara, Turkey. *PhD Thesis*, 300p.
- Moldowan, J. M., Seifert, W. K., Gallegos, E. J., 1985. Relationship between petroleum composition and depositional environment of petroleum source rocks: *AAPG Bulletin* 69, pp. 1255-1268.
- Zumberge, J. E., 1987. Prediction of source rock characteristics based on terpane biomarkers in crude oils: A multivariate statistical approach: *Geochimica et Cosmochimica Acta* 51, pp.1625-1637.
- Xue, J., Lee, C., Wakeham, G. S., Armstrong, A. R., 2011. Using principle components analysis (PCA) with cluster analysis to study the organic geochemistry of sinking particles in the ocean: *Organic Geochemistry* 42, pp. 356-367.