https://doi.org/10.3176/oil.1994.1.02

UDC 533.983 : 553.06

K. UROV, A. EBBER

ONCE MORE ON CLASSIFICATION OF OIL SHALES USING CLUSTER ANALYSIS

Abstract

On the basis of the results of the multivariate statistical analysis of 56 oil shales and shale-like rocks using their 6 essential characteristics these caustobioliths have been subdivided into three classes: lipokerogenous shales originating mainly from the lipid fraction of initial organic material and characterized by a high yield of semicoking oil, kerogen basis, heterokerogenous shales containing organic matter thoroughly transformed by multi-stage metabolic processes and yielding little oil on low temperature pyrolysis and intermediate shales.

Introduction

In [1] it has been established that by cluster analysis of oil shales from 44 deposits, according to their 13 characteristic indices, oil shales can be divided into two principal classes differing among themselves primarily by the semicoking oil yield from their organic matter (kerogen). Since the original data (characteristics of oil shales) were used without any pre-processing, this could to some extent influence the results obtained.

In this work, in order to enlarge the examined shales list the number of indices used was reduced to 6, but, as a result, the number of shales increased up to 56. Before calculations the shales characteristics were centered and normalized.

In both cases the data published in [2] were made use of; in the same work additional information on the shales viewed is available.

Methods

The characteristics of oil shales used in calculations are presented in Table. This matrix of multivariate data can be represented as a swarm of 56 points in a 6-dimensional space.

Our objective being to seek for a presumed grouping of similar shales into clusters on the base of their known properties, two methods have been used:

1) searching for the shales groups in original multidimensional space according to interpoint distances;

2) projection of data into a lowerdimensional space; more comprehensible 2- and 3-dimensional spaces were used.

Prior to calculations, the available date were mean-centered (i.e. from each column the mean value of data was subtracted) and scaled so that each column had standard deviation 1.0.

Characteristics of oil shales and shale-like rocks

No	Oil shale Deposit, outcrop (country, region)	Geological age	Characteristics, %						Euclidean distance from No 1
			Mineral carbon dioxide $(CO_2)^d_M$	Organic matter $(OM)^d$	Total sulfur S^d_t	Yield of semi-coking oil, kerogen basis, T^o_{sK}	Yield of the semi-coke organic part, kerogen basis, $(sK)^o$	Content of nitrogen in kerogen N^o	
1	Dictyonema (Estonia, Maardu)	Lower Ordovician	0.2	20.6	2.5	16.7	54.3	2.5	0.0000
2	Omolon (Russia, Magadan)	Lower Triassic	0.2	21.5	0.7	13.0	54.9	2.9	0.1416
3	Omolon (Russia, Magadan)	Middle Triassic	0.2	17.6	1.2	13.6	54.0	3.5	0.2060
4	Olenek (Russia, Yakutia)	Cambrium	0.8	30.2	0.9	22.8	59.3	1.6	0.2238
5	Tremembé-Taubaté massive shale (Brazil)	Tertiary	0.3	17.4	0.3	23.0	39.1	2.8	0.2251
6	Alyuisk (Russia, Irkutsk)	Jurassic	0.3	38.9	1.0	23.5	53.2	1.7	0.2372
7	Voronye-Voloskovsk (Russia, Vyatka)	Jurassic	4.6	20.1	1.7	27.6	37.8	2.1	0.2415
8	Dictyonema (Estonia, Toolse)	Lower Ordovician	0.1	19.4	5.9	9.3	58.8	2.1	0.2447
9	Menilitic (Ukraine, Carpathians)	Tertiary	0.5	19.9	2.1	10.5	58.0	1.1	0.2691
10	Novodmitrovsk (Ukraine)	Tertiary	4.8	21.1	4.6	24.2	35.1	1.7	0.3089

Characteristics of oil shales and shale-like rocks (continued)

No	Oil shale Deposit, outcrop (country, region)	Geological age	Characteristics, %						Euclidean distance from No 1
			Mineral carbon dioxide $(CO_2)^d_M$	Organic matter $(OM)^d$	Total sulfur S^d_t	Yield of semi-coking oil, kerogen basis, T^o_{sK}	Yield of the semi-coke organic part, kerogen basis, $(sK)^o$	Content of nitrogen in kerogen N^o	
11	Maomin (China)	Tertiary	1.4	25.2	1.1	33.0	34.8	1.6	0.3171
12	Kenderlyk, the Kalyn-Kara seam (Kazakhstan)	Upper Carboniferous	0.0	48.4	0.9	21.6	38.0	1.8	0.3238
13	Sysol (Russia, Komi)	Jurassic	5.9	21.4	3.2	23.2	31.6	1.6	0.3258
14	Borov Dol (Bulgaria)	Tertiary	3.3	19.7	1.9	41.6	39.1	1.7	0.3295
15	Ust-Kamenogorsk (Kazakhstan)	Carboniferous-Permian	3.2	22.8	1.1	35.6	28.9	2.5	0.3344
16	Manturovo (Russia, Nizny Novgorod)	Jurassic	3.5	39.2	4.6	30.2	34.4	1.9	0.3397
17	Simbirsk (Russia)	Jurassic	6.1	31.9	3.8	28.8	41.1	1.3	0.3407
18	Autun (France)	Permian	6.5	19.8	0.8	41.4	35.4	2.4	0.3573
19	Kapali (Uzbekistan)	Tertiary	10.2	29.7	6.1	25.6	45.5	2.6	0.3664
20	Irati (Brazil)	Carboniferous-Permian	1.8	34.0	2.0	40.9	34.1	1.6	0.3685
21	Kvarntorp (Sweden)	Cambrian-Silurian	1.0	20.0	5.3	24.2	50.9	0.8	0.3705
22	Fushun (China)	Tertiary	3.4	21.2	0.5	36.8	27.8	3.0	0.3731
23	St. Hilaire (France)	Permian	3.4	25.3	2.3	37.5	44.3	1.3	0.3749
24	Garibak (Tadjikistan)	Tertiary	0.2	48.6	2.4	34.0	55.6	1.1	0.3782

Characteristics of oil shales and shale-like rocks (continued)

No	Oil shale Deposit, outcrop (country, region)	Geological age	Characteristics, %						Euclidean distance from No 1
			Mineral carbon dioxide $(CO_2)^d_M$	Organic matter $(OM)^d$	Total sulfur S_t^d	Yield of semi-coking oil, kerogen basis, T^o_{sK}	Yield of the semi-coke organic part, kerogen basis, $(sK)^o$	Content of nitrogen in kerogen N^o	
25	Westwood (Great Britain, Scotland)	Lower Carboniferous	3.2	19.0	0.7	43.2	29.5	2.6	0.3811
26	Boltysh (Ukraine)	Tertiary	3.6	34.9	1.1	41.5	35.8	1.6	0.3817
27	Ermelo (South Africa)	Lower Carboniferous	1.8	54.2	2.0	34.0	52.9	1.3	0.3866
28	Pahinsk (Russia, Perm)	Carboniferous-Permian	0.1	26.5	0.2	22.6	48.7	0.6	0.3914
29	Baisun (Uzbekistan)	Tertiary	6.8	38.0	5.1	35.5	29.7	2.2	0.4119
30	Koprinka (Bulgaria)	Tertiary	1.1	15.6	0.9	38.5	25.0	1.6	0.4130
31	Eastern Urtabulak (Uzbekistan)	Tertiary	6.3	39.8	6.6	26.6	30.4	3.1	0.4403
32	Pirin (Bulgaria)	Tertiary	4.6	34.5	2.2	40.3	21.2	2.0	0.4412
33	Gurkovo (Bulgaria)	Tertiary	5.9	10.8	0.6	38.9	23.1	1.9	0.4433
34	Budagovo, humic saprope-lite (Russia, Irkutsk)	Jurassic	2.9	68.6	0.5	7.6	59.2	1.5	0.4571
35	Kenderlyk, the Karaungur seam (Kazakhstan)	Lower Permian	1.9	21.7	0.7	45.9	20.7	2.2	0.4584
36	Bogoslovsk (Russia, Yekaterinburg)	Triassic-Jurassic	3.0	60.9	4.2	36.0	38.8	1.4	0.4664

Characteristics of oil shales and shale-like rocks (continued)

No	Oil shale Deposit, outcrop (country, region)	Geological age	Characteristics, %						Euclidean distance from No 1
			Mineral carbon dioxide $(CO_2)^d_M$	Organic matter $(OM)^d$	Total sulfur S_t^d	Yield of semi-coking oil, kerogen basis, T^o_{sK}	Yield of the semi-coke organic part, kerogen basis, $(sK)^o$	Content of nitrogen in kerogen N^o	
37	Kultak-Zevardy (Uzbekistan)	Jurassic	14.6	22.8	3.2	27.2	25.4	2.8	0.4757
38	Cerro Largo (Uruguay)	Permian	0.0	21.6	3.1	19.4	14.8	1.0	0.4880
39	Kashpir (Russia, Samara-Simbirsk)	Jurassic	11.3	30.5	5.0	39.3	33.8	1.2	0.5029
40	Ashinsk (Russia, Bashkiria)	Middle Devonian	15.5	14.8	2.2	43.2	32.8	2.1	0.5102
41	Tremembé-Taubaté, paper shale (Brazil)	Tertiary	0.2	39.5	0.2	53.4	28.4	1.6	0.5112
42	Eastern-Chandyr (Uzbekistan)	Tertiary	15.1	18.8	3.0	26.6	20.2	2.5	0.5132
43	Kimmeridge (Great Britain, England)	Jurassic	2.5	59.8	6.6	42.6	33.4	2.0	0.5307
44	Obshchy Syrt (Russia, Samara)	Jurassic	10.0	33.8	4.5	34.3	22.2	0.8	0.5631
45	Turovo (Byelorussia)	Upper Devonian	12.7	17.2	2.6	47.3	29.1	0.8	0.5858
46	Krasava (Bulgaria)	Tertiary	13.6	10.9	0.8	48.6	19.3	1.7	0.6042
47	Tshernozatonsk (Kazakhstan)	Upper Devonian	9.8	54.0	5.1	42.2	21.1	0.8	0.6503

Characteristics of oil shales and shale-like rocks (end)

No	Oil shale Deposit, outcrop (country, region)	Geological age	Characteristics, %						Euclidean distance from No 1
			Mineral carbon dioxide $(CO_2)^d_M$	Organic matter $(OM)^d$	Total sulfur S_t^d	Yield of semi-coking oil, kerogen basis, T^o_{sK}	Yield of the semi-coke organic part, kerogen basis, $(sK)^o$	Content of nitrogen in kerogen N^o	
48	Yarenga (Russia, Komi)	Jurassic	1.6	76.0	7.9	37.5	34.6	1.5	0.6521
49	Glen Davis (Australia)	Carboniferous-Permian	0.1	26.5	0.2	22.6	48.7	0.6	0.6595
50	Tshagansk (Russia, Orenburg)	Jurassic	7.6	56.7	8.3	43.9	22.4	1.1	0.6998
51	Mandra (Bulgaria)	Tertiary	13.6	27.7	2.6	65.0	17.0	1.0	0.7354
52	Lyubansk (Byelorussia)	Upper Devonian	17.2	11.7	1.2	52.5	18.3	0.6	0.7469
53	Green River (USA, Utah)	Tertiary	19.0	19.4	1.1	59.3	11.8	2.1	0.7587
54	Mae Sot (Thailand)	Tertiary	11.0	21.0	0.9	65.6	15.3	0.4	0.7759
55	Budagovo, the true sapropelite (Russia, Irkutsk)	Jurassic	1.0	89.2	0.6	51.1	16.1	1.4	0.7787
56	Kukersite (Estonia)	Middle Ordovician	18.0	35.5	1.7	65.6	18.3	0.4	0.8405

For typifying the shales according to the first method, the Euclidean distances between the points in original 6-dimensional space have been determined and considered as a measure of objects similarity:

$$E_{ij} = \sqrt{\sum_{k=1}^{K} (x_{ik} - x_{jk})^2}$$

where E_{ij} is the Euclidean distance between objects i and j, x_{ik} and x_{jk} are the values of k-th variable correspondingly for i-th and j-th objects.

For cluster analysis results of which can be visualized the principal component analysis [3] has been used. The procedure for finding principal components (PC), or factors, has been implemented according to NIPALS algorithm [4]. Calculations with 1...3 factors showed that the first PC explained 38.8 % of variance in the initial data, introduction of the second factor enabled to explain 62.7 % of the data and that of the third factor 79.8 %. Since involving the third PC did not improve the feasibility of classification only PC1 and PC2 and, respectively, a projection of data onto a two-dimensional plane have been used.

Results and Discussion

A histogram of 56 oil shales scanned in this work according to semicoking oil yield, kerogen basis, is presented in Fig. 1. It supports the assumption made in [1] that oil shales are not distributed in the nature evenly as to their composition and properties, but form certain groups. This, in its turn, may be regarded as a pre-requisition for their motivated genetic classification.

In Table the shales are arranged in consecutive order accordingly to their Euclidean distance from Dictyonema shale (Estonia) that, after some preliminary experimentation and using our *a priori* knowledge, had been found to be a suitable starting point due to its in many respects exceptional nature. This shale is characterized by a very low semicoking oil yield, organic matter basis, brought about by a low hydrogen content in its kerogen, relatively high nitrogen content, near absence of mineral carbon dioxide, considerable content of heavy metals and other properties typical of so-called "black" shales. The latter are on borderline of true oil shales and caustobioliths of humin origin. Therefore, if a regularity of some kind does exist in the case of oil shales, Dictyonema shale may be expected to be at the end (or beginning) of the succession.

Distribution of the shales according to their interpoint distances from Dictyonema shale is shown in Fig. 2. The curve obtained may be roughly divided into three parts: a section with a sharp rise, a more or less horizontal part and again a steep-rising region. Let us divide all shales tentatively into three classes according to this index.

Results of cluster analysis carried out in the above-described way are demonstrated in Fig. 3. It is evident that the classes proposed on the basis of interpoint distances (Fig. 2) are also clearly differentiated here. In comparison with the results of the work [1] the groups I and VI differentiated there are close to classes I and III here. Thus, results of both studies are similar though normalization of the initial data caused a little different distribution of shales.

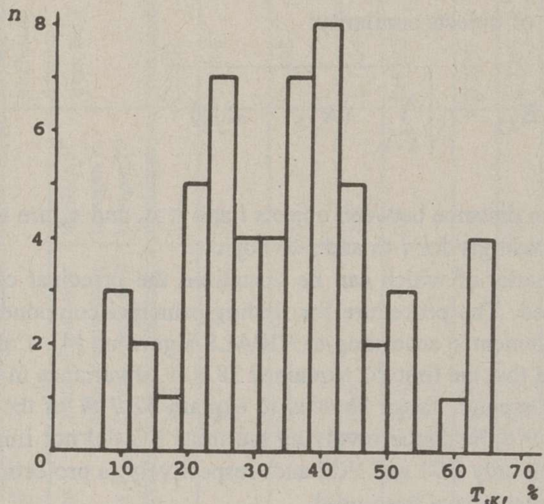

Fig. 1. Histogram of oil shales according to their semicoking oil yield, kerogen basis ($T°_{sK}$);
n - number of shales with respective oil yield

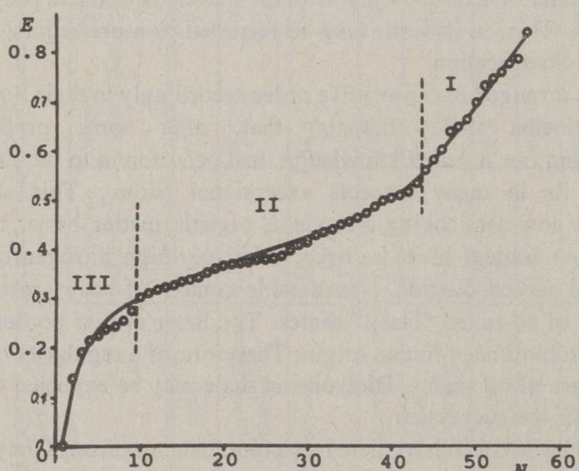

Fig. 2. Distribution of oil shales according to interpoint distances: *E* - Euclidean distance from Dictyonema shale; *N* - successive number of the shale according to Table; I, II and III - classes of oil shales

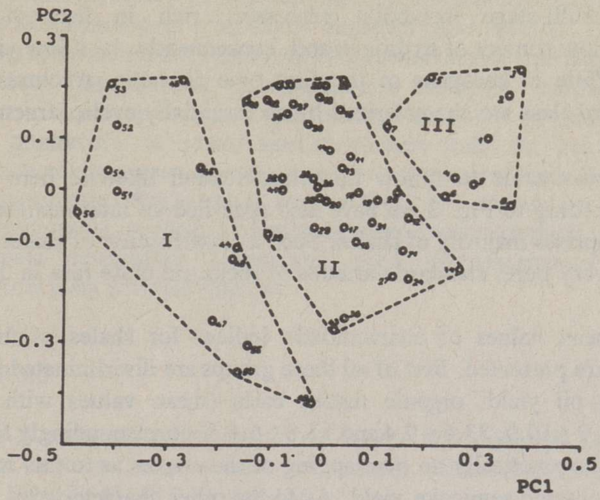

Fig. 3. Projection of the mean-centered and normalized data on oil shales onto 2-dimensional plane defined by first two principal components: PC1 - the first principal component; PC2 - the second principal component; I, II and III - classes of oil shales

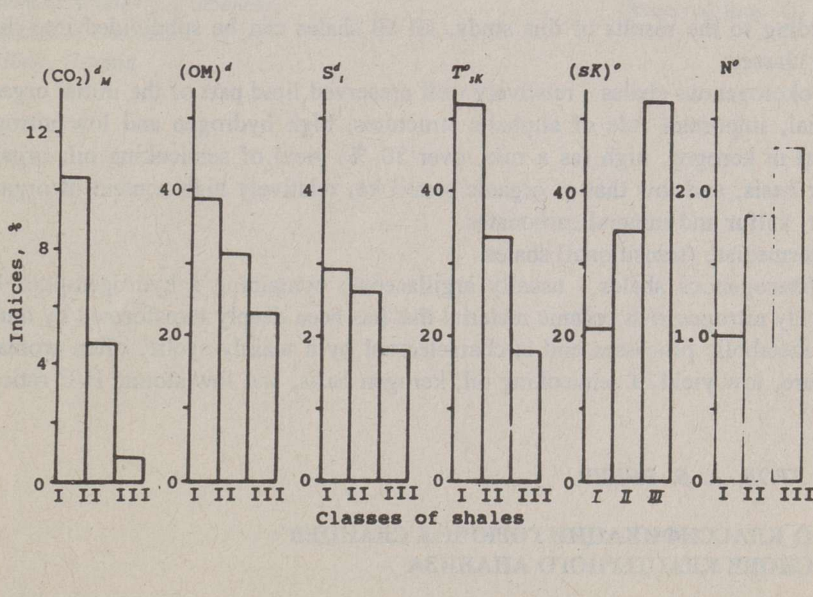

Fig. 4. Mean values of characteristic indices for various shale classes

In the previous investigation [1], the shales of the group I were specified as lipokerogenous, with transformation products of the initial organic material lipid fraction prevailing in their composition, and those of the group VI (here class III) as heterokerogenous shales with a leading role of organic material, profoundly transformed by multi-stage metabolic processes, rich in heteroatoms and characterized by a low content of hydrogen and, consequently, by a low yield of oil on semicoking. While in kerogens of the first type aliphatic structures prevail, kerogens of the third class are characterized by an essentially cyclic structure, often of aromatic nature.

It seems to us reasonable to follow this classification likewise here with the difference that according to Fig. 3 we have also specified an intermediate class II, and this group comprises majority of shales. Such a classification of shales seems to be justified as the very pure, standard varieties of rocks are quite rare in the earth's crust.

In Fig. 4 the mean values of characteristic indices for shales of the above-mentioned classes are presented. Best of all these groups are discriminated in respect to the semicoking oil yield, organic matter basis (mean values with standard deviations being 51.9 ± 10.9; 33.6 ± 9.4 and 17.8 ± 6.6 % correspondingly for classes I, II and III), there is practically no overlapping of the classes as to this index. The same goes for the organic semicoke yield. As to the other characteristics used, the shale classes are not so well differentiated though certain tendencies can be defined: from class I to class III organic matter, mineral carbon dioxide and sulfur content tend to diminish, but nitrogen content in kerogen is inclined to grow.

Conclusions

According to the results of this study, all oil shales can be subdivided into three major classes:

I. Lipokerogenous shales - relatively well preserved lipid part of the initial organic material, important role of aliphatic structures, high hydrogen and low nitrogen content in kerogen, high (as a rule, over 30 %) yield of semicoking oil, organic matter basis, and low that of organic semicoke, relatively high content of organic matter, sulfur and mineral carbonates.

II. Intermediate (transitional) shales.

III. Heterogenous shales - usually argillaceous, containing a hydrogen-poor but relatively nitrogen-rich organic material that has been deeply transformed by multi-stage metabolic processes and is characterized by a mainly cyclic, often aromatic structure, low yield of semicoking oil, kerogen basis, and low atomic H/C ratio.

К. Э. УРОВ, А. В. ЭББЕР

ЕЩЕ О КЛАССИФИКАЦИИ ГОРЮЧИХ СЛАНЦЕВ НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА

Резюме

В продолжение работы [1] проведен многомерный статистический анализ данных для 56 различных горючих сланцев с учетом 6 показателей их состава и свойств

(таблица); для расчетов исходные данные были центрированы и нормализованы к единичному стандартному отклонению.

Судя по гистограмме распределения рассматривавшихся сланцев по выходу смолы полукоксования в расчете на кероген (рис. 1), они действительно распределяются в природе не равномерно, а образуют определенные группы, что может служить основой для естественной генетической классификации этого вида каустобиолитов.

Основываясь на кривой распределения сланцев согласно эвклидовым расстояниям, характеризующим удаленность их точек от диктионемого сланца Эстонии, принятого за точку начала отсчета (рис. 2), все горючие сланцы предлагается подразделить на три класса: I - липокерогеновые, II - переходные, III - гетерокерогеновые. Кластерный анализ (рис. 3) показал, что эти классы ясно дифференцированы, особенно по выходу низкотемпературной смолы на кероген.

Определены тенденции изменения всех учитывавшихся показателей в генетическом ряду сланцев (рис. 4).

REFERENCES

1. *Уров К., Кадарпик В., Сумберг А.* О выборе показателей для классификации горючих сланцев и граничных значениях параметров // Изв. АН Эстонии. Химия. 1991. Т. 40, № 3. С. 149-156.
2. *Уров К. Э., Сумберг А. И.* Характеристика горючих сланцев и сланцеподобных пород известных месторождений и проявлений. - Таллинн, "Валгус", 1992.
3. *Wold S., Esbensen K., Geladi P.* Principal Component Analysis // Chemom. Intell. Lab. Syst. 1987. V. 2. P. 37—52
4. *Martens H., Naes T.* Multivariate Calibration. - Chichester, New York, Brisbane, Toronto, Singapore, "John Wiley & Sons", 1991.

Estonian Academy of Sciences,
Institute of Chemistry
Tallinn, Estonia

Received June 02, 1993