

## Дифференциация образцов нефти по изомерному составу парафинов с использованием методов кластерного анализа

\*Ю.П. Туров, М.Ю. Гузниева

Сургутский государственный университет, Российская Федерация, 628412, Тюменская область, Ханты-Мансийский автономный округ – Югра, г. Сургут, пр. Ленина, д. 1

\*Адрес для переписки: Туров Юрий Прокопьевич, E-mail: yuri\_tom@rambler.ru

Поступила в редакцию 8 ноября 2021 г., после доработки – 24 января 2022 г.

Предложен метод контроля источников образцов нефти по компонентному составу *n*-парафинов для их дифференциации, идентификации и установления происхождения (конкретных скважин эксплуатируемого месторождения) с использованием агломеративной иерархической кластеризации данных с применением алгоритма Д. Уорда (J. Ward). Исходной информацией служат данные по составу и относительным содержаниям *n*-парафинов C<sub>17</sub>H<sub>36</sub> ÷ C<sub>34</sub>H<sub>70</sub>, изопреноидов пристана C<sub>19</sub>H<sub>40</sub> и фитана C<sub>20</sub>H<sub>42</sub> в образцах нефти, полученные методом хроматомасс-спектрометрии в варианте мониторинга выбранных ионов. Возможности метода иллюстрируются результатами обработки матрицы данных размерностью 20×27 по относительным содержаниям парафинов в 27 шифрованных анонимных образцах. При этом выделены 3 кластера, в состав каждого из которых были объединены образцы, отобранные из конкретной скважины. Таким образом, дифференцированы источники образцов, из которых они были отобраны. Решить эту задачу традиционным методом сопоставления усредненных для каждой скважины содержаний индивидуальных парафинов оказалось невозможным. С учетом неизбежных погрешностей в результатах анализов и возможных изменений состава извлекаемой нефти в процессе эксплуатации скважин, значения относительных содержаний парафинов в образцах нефти из всех трех скважин перекрываются и источники остаются неразличимыми. Предлагаемый аналитический метод, основанный на получении информации о компонентном составе парафинов методами хроматомасс-спектрометрии в варианте мониторинга выделенных ионов и обработке ее методами кластерного анализа, расширяет возможности аналитиков при контроле источников образцов нефти в органической геохимии и химии окружающей среды, которые осложнены широкой вариабельностью состава образцов нефти, изменчивостью ее состава под воздействием природных и техногенных факторов и отсутствием возможности оценок реальных неопределенностей данных.

**Ключевые слова:** нефтяная система, кластерный анализ, состав парафинов, хроматомасс-спектрометрия, принятие решения в условиях неопределенности.

For citation: *Analitika i kontrol'* [Analytics and Control], 2022, vol. 26, no. 1, pp. 31-40

DOI: 10.15826/analitika.2022.26.1.006

## Differentiation of oil samples by isomeric composition of paraffins using cluster analysis methods

\*Iu.P. Turov, M.Iu. Guzniaeva

Surgut State University (SurGU), 1, Lenin Ave., Surgut, 628412, Russian Federation

\*Corresponding author: Iu.P. Turov, E-mail: yuri\_tom@rambler.ru

Submitted 08 November 2021, received in revised form 24 January 2022

A new method is proposed for controlling the sources of oil samples by chemical composition for their differentiation, identification and establishment of their origin (specific wells of the operating field) using the agglomerative hierarchical data clustering and applying the J. Ward algorithm. The initial information is the data on the composition and relative contents of *n*-paraffins C<sub>17</sub>H<sub>36</sub> ÷ C<sub>34</sub>H<sub>70</sub>, isoprenoids of pristane C<sub>19</sub>H<sub>40</sub> and phytane C<sub>20</sub>H<sub>42</sub> in oil samples obtained by selected-ion monitoring gas chromatography/mass spectrometry.

The solution to the problem was carried out in the absence of the information about the magnitude of the total uncertainty in the data, without using the standard samples for their assessment and without the preliminary assumption of the number of expected clusters. The proposed technique's possibilities were illustrated by the results of  $20 \times 27$  data matrix processing based on the relative paraffin content in 27 encrypted anonymous samples. As an outcome, three clusters were identified, each of which included objects sampled from the specific well. Differentiating the sources of the samples using the traditional methods by comparing the amounts of *n*-paraffins and isoprenoids proved to be impossible. Due to the inevitable errors in the analysis results and the possible changes in the composition of the recovered oil, the values of the relative contents of paraffins in the oil samples from all three wells overlap and the sources remain indistinguishable. The newly proposed method made it possible to solve the differentiation problem of oil samples in organic geochemistry and environment control which are complicated by the wide variability of the composition of oil samples, the lability of petroleum system under the influence of natural and technogenic factors and the absence of the estimate of the actual errors in the data.

**Keywords:** oil system, data mining, cluster analysis, paraffin composition, gas chromatography/mass spectrometry, decision making under uncertainty.

## Введение

Задачи идентификация образцов нефти по результатам анализа материального состава основной массы вещества и доминирующих в них компонентов возникают в различных областях деятельности химиков-аналитиков: в природоохранном контроле при установлении источников нефтяного загрязнения, в геохимических методах поиска нефтяных месторождений, при исследовании состава рассеянного органического вещества пород, при оценках запасов углеводородов разведанных месторождений, при геохимическом сопровождении процессов нефтеизвлечения на отдельных добывающих скважинах и эксплуатируемых месторождений в целом. С позиций требований аналитической химии все эти задачи объединяет вероятностный характер конечного результата – вследствие чрезвычайной многокомпонентности и вариабельности состава образцов и широкой изменчивости его под воздействием трудно контролируемых внешних факторов обеспечить строгую метрологическую оценку неопределенности в результатах анализа традиционными методами невозможно. Под традиционными методами в данном случае рассматривается обеспечение прослеживаемости аналитических результатов и оценка неопределенности в них использованием для этого стандартных образцов. При этом следует учесть, что в масштабе мировой аналитической практики методики использования стандартных образцов качественных свойств для оценки и выражения неопределенности в качественном анализе находятся пока в состоянии разработки [1]. Тем более, что вследствие многостадийности процедур нефтяного анализа (пробоотбор, транспортировка и хранение образцов, пробоподготовка перед проведением непосредственно измерений аналитического сигнала) неизбежные погрешности каждого этапа накапливаются, и решить проблему оценки суммарной неопределенности результата даже при гипотетической реализации создания таких стандартных образцов не удастся.

В настоящее время задачи классификации и идентификации образцов нефти и нефтепродуктов

решают путем сопоставления компонентного и изомерного состава различных классов нефтяных соединений. Вследствие своей широкой распространенности наиболее популярным классом веществ при этом являются парафины. Парафины присутствуют в составе нефти практически всех месторождений, и в большинстве образцов нефти месторождений на территории Ханты-Мансийского автономного округа являются одним из доминирующих классов соединений. Кроме того, относительные содержания различных изомеров парафинов (в том числе так называемых биомаркеров) и рассчитываемые по ним геохимические параметры или характеристические соотношения содержаний *n*-парафинов и изопреноидов [2] обладают высоким идентификационным потенциалом и дифференцирующими (характеристическими) свойствами. Их широко используют при классификации образцов нефти по их генетическим признакам в органической геохимии. Обоснованность такого подхода при сопоставлении состава образцов нефти различных месторождений подтверждена многолетней популярностью применения геохимических индексов в органической геохимии, начиная со второй половины XX века [3, 4] и до наших дней [5]. Детальное метрологическое обоснование применимости характеристических отношений содержаний изопреноидных и нормальных парафинов для дифференциации образцов нефти разных месторождений по результатам анализа методом газо-жидкостной хроматографии было дано в работе [6].

Но отсутствие стандартных образцов изомерного состава нефтяных компонентов и принципиальная невозможность создания таковых из-за чрезвычайно широкой вариабельности состава образцов нефти и их изменчивости под действием природных и техногенных факторов вынуждает аналитиков в большинстве случаев ограничиваться оценками вариабельности характеристик состава нефти различных месторождений [6].

Однако существует ряд задач, связанных с необходимостью получения более строгих оценок неопределенности в результатах сопоставления состава различных нефтяных образцов в отсут-

ствии априорной информации (или вне зависимости от нее) об их происхождении или природе. Такие задачи возникают при идентификации источников нефтяного загрязнения компонентов окружающей среды [7], оптимизации режимов нефтеизвлечения с учетом миграции пластовых флюидов, балансовых расчетах притоков из различных источников в многопластовых месторождениях [8]. Источником дополнительных трудностей в решении всех этих задач является хорошо известный нефтяникам феномен изменения состава нефти в процессе эксплуатации как отдельных нефтяных скважин, так и месторождений в целом [9, 10].

Широко используемые в органической геохимии критерии генетической классификации образцов по углеводородному составу [2, 6] не имеют четких границ, и не могут служить надежной основой для принятия решений. Хорошо отработанные в аналитической химии приемы идентификации химических веществ [11] методом сопоставления характеристик объекта идентификации с эталонами или библиотечными данными здесь бессильны, так как ни того, ни другого при решении перечисленных задач нет. Тем не менее, оценка надежности и состоятельности получаемых результатов анализа при сопоставлении состава образцов необходима во всех этих случаях – для того, чтобы подтвердить обоснованность принимаемых на их основе технологических и организационных решений.

Большое число изомеров, обнаруживаемых в составе нефтяных парафинов, и многомерность информации, получаемой при анализе состава образцов нефти хроматографическими методами, позволяют использовать приемы и методы «data mining», с помощью которых решают ключевые вопросы многомерного анализа сложных систем:

- можно ли считать идентичными анализируемые объекты и за счет каких признаков можно объяснить их возможные отличия?
- как можно объединить отобранные объекты в группы?
- существует ли пространственная или временная изменчивость описанных объектов и каковы ее структурные особенности?

- а также позволяющие оценивать уровни неопределенностей в больших наборах данных без каких-либо предположений об их природе и о структуре информации в массиве данных [12].

Методы «data mining» широко используют для обобщения, структурирования и интерпретации данных в науках о Земле – географии, геологии, геохимии и геофизике, при этом наиболее популярным инструментом является кластерный анализ [13-18]. Популярность многомерных математических методов для обработки химико-аналитических данных также непрерывно увеличивается одновременно с совершенствованием аналитических приборов [19-21].

Общий методологический подход, заложенный в основу «data mining» – систематизация и описание

структуры данных без предварительного постулирования априорной физической и/или химической модели. Цель этих методов – объяснение дисперсии (изменчивости) данных с использованием минимальных или полного отсутствия предположений о природе обрабатываемых данных.

При решении задач классификации и идентификации источников образцов нефти по их составу потенциальные возможности «data mining» представляют особый интерес, так как из-за отсутствия стандартных образцов состава и динамичности нефтяной системы заранее оценить величину неопределенностей в данных не представляется возможным. Тем не менее, чтобы принимать решение о значимости или незначимости наблюдаемых различий в составе разных образцов нефти, необходимо получить оценку возможности их дифференцирования при отсутствии оценок суммарной неопределенности в результатах анализов.

Основная задача, решению которой посвящена данная работа заключается в том, чтобы оценить возможно ли дифференцировать образцы по их источникам на основе вариаций относительных содержаний парафинов в разных образцах с учетом суммарных (но неизвестных) погрешностей определения относительных содержаний, а также того, что с течением времени состав извлекаемой из каждой скважины нефти может изменяться. Правда, насколько значимо – тоже неизвестно.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Объекты исследования.** Изучены 27 образцов нефти, отобранных на устье трех добывающих скважин одного из месторождений на территории Ханты-Мансийского автономного округа тремя партиями в течение полугода с интервалом ~2 месяца. Во время каждого пробоотбора из каждой скважины отбирали по три пробы нефти. На скважинах в полугодовой период времени, предшествующий пробоотбору, и в процессе его не проводили никаких мероприятий, связанных с физико-химическими воздействиями на продуктивные пласты и инженерное оборудование скважин. По имеющимся сведениям, скважины № 1 и № 3 извлекают нефть из одного продуктивного пласта, а № 2 – из другого. Контрольная проба нефти для проверки возможности идентификации источника образца была отобрана из скважины № 1 через полтора месяца после завершения периодических пробоотборов.

**Методы анализа и аппаратура.** Пробы сырой нефти разделяли на фракции насыщенных, ароматических веществ, смол и асфальтенов в соответствии с методикой SARA-анализа [22]. Состав парафинов во фракциях насыщенных веществ образцов был исследован методами газовой хроматографии (ГХ) с пламенно-ионизационным детектором и хроматомасс-спектрометрии (ГХ/МС). В процессе исследований было установлено, что наилучшая повторяемость и воспроизводимость результатов



анализа при сопоставлении данных, полученных разными методами – ГХ, ГХ/МС с регистрацией хроматограмм по полному ионному току и методом мониторинга выбранных ионов (при использовании идентичных колонок и режимов хроматографического разделения образцов), достигается при использовании селективной ионной регистрации хроматограмм по суммарной интенсивности характеристических для масс-спектров класса парафинов фрагментных ионов с массами 57, 71, 85 и 99 [23, 24]. Далее в работе использовали данные, полученные этим методом.

ГХ/МС анализ проведен с использованием хроматомасс-спектрометра GCMS-TQ8040 (Shimadzu, Japan). Хроматографическое разделение проводили с использованием капиллярной колонки размерами 30 м × 0.25 мм × 0.25 мкм с неподвижной фазой Gs BP-5MS. Режим работы хроматографа: газ-носитель – гелий, расход газа 1.0 мл/мин; начальная температура термостата 50 °С, изотерма 2 мин, затем подъем температуры со скоростью 5 °С/мин до 290 °С с выдержкой при этой температуре в течение 45 мин; температура испарителя и интерфейса 280 °С. Режим работы масс-спектрометра: ионизация электронами с энергией 70 эВ; температура ионизационной камеры 280 °С; сканирование спектра в диапазоне масс от 33 до 460 Да со скоростью 4 полных масс-спектра в секунду. Сбор и первичную обработку результатов эксперимента осуществляли с использованием программного комплекса Shimadzu GCMS solution, для идентификации веществ и поиска ближайших аналогов использовали библиотеку масс-спектров NIST14.

**Обработка результатов.** Идентификация *n*-парафинов, пристана и фитана проведена с использованием времен удерживания и масс-спектров соответствующих хроматографических пиков. Определение относительных содержаний парафинов осуществляли по площадям пиков на хроматограммах, реконструированных в режиме мониторинга выбранных ионов по суммарной интенсивности ионов с массами 57, 71, 85 и 99. Так как основной задачей работы было сопоставление относительных содержаний веществ в разных образцах, коэффициенты чувствительности масс-спектрального детектора ко всем веществам принимали равными единице. Обработку реконструированных хроматограмм проводили методом внутренней нормализации: нормированное относительное содержание вещества в процентах получали путем деления площади пика соответствующего вещества на сумму площадей пиков пристана, фитана и всех *n*-парафинов от  $C_{17}H_{36}$  до  $C_{34}H_{70}$ .

Выбор группы алканов  $C_{17}H_{36}$  ÷  $C_{34}H_{70}$  обусловлен тем, что для геохимической классификации образцов нефти из разных источников и рассеянного органического вещества пород чаще всего используют информацию о составе парафинов именно этого диапазона [2–5, 8, 10]. Кроме того, трудно контролируемая история проб нефти до их попадания в

аналитическую лабораторию (различия в сроках и условиях хранения и транспортировки образцов) может приводить к потерям наиболее легколетучих компонентов и существенному изменению состава легких парафинов с числом атомов углерода в молекуле менее 17.

Набор данных по относительным содержаниям пристана, фитана и всех *n*-парафинов от  $C_{17}H_{36}$  до  $C_{34}H_{70}$  в 27 образцах нефти (три пробоотбора по три образца из каждой скважины) представляет собой матрицу размерностью 20×27. Кластерный анализ этих данных проведен с использованием пакета прикладных программ MATLAB (пакет *Statistics Toolbox*, функции *pdistc* использованием евклидовой метрики между объектами, *linkage* «ward» для расчета иерархического дерева кластеров по методу Уорда и *dendrogram* для графического представления результатов – построения дендрограммы). Расчеты доверительных интервалов и вариаций содержания веществ проводили с использованием встроенных функций табличного процессора Excel (Windows 2010).

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В табл. 1 приведены результаты анализа относительных содержаний парафинов в образцах нефти. Порядковый номер образцов и номер партии отражают последовательность их поступления в лабораторию.

Попытка дифференцировать источники образцов традиционными методами представлена данными в табл. 2. В ней приведены усредненные значения относительных содержаний парафинов в пробах нефти из каждой скважины (табл. 1) с учетом доверительных интервалов, рассчитанных в соответствии со следующими предположениями: изменения состава извлекаемой нефти за время наблюдений пренебрежимо малы; наблюдаемые расхождения в составе образцов носят случайный характер; стандартные отклонения состава проб из каждой скважины по трем пробоотборам можно использовать для оценки доверительных интервалов.

Как следует из табл. 2, различия в значениях усредненных содержаний парафинов в пробах не позволяют надежно дифференцировать их источники, т.е. скважину, из которой они были отобраны – с учетом доверительных интервалов относительные содержания индивидуальных парафинов в образцах нефти из трех исследованных скважин перекрываются. Поэтому мы попытались дифференцировать образцы с использованием множественных характеристик состава парафинов, сопоставляя весь набор их относительных концентраций.

Эффективность использования методов кластеризации при обработке многомерной информации и ее интерпретации в науках о Земле [13, 15] обусловили нашу попытку использовать методы кластерного анализа для решения поставленных задач при обработке данных об относительных

Таблица 1

Относительные содержания *n*-парафинов и изопреноидов в пробах нефти, % отн.

Table 1

Relative amounts of *n*-paraffins and isoprenoids in oil samples, % rel.

Номер пар- тии образ- цов	<i>n</i> -парафины и изопрено- иды	Порядковые номера образцов и скважин								
		1	2	3	4	5	6	7	8	9
		скважина № 1			скважина № 2			скважина № 3		
I	C <sub>17</sub>	6.21	5.92	5.84	6.32	6.61	6.74	6.98	6.81	6.90
	Pr	5.13	5.33	5.13	5.64	5.42	5.65	4.73	4.71	5.11
	C <sub>18</sub>	6.44	6.34	5.97	6.69	7.10	6.88	6.61	6.99	6.70
	Phy	6.83	6.29	6.72	6.93	7.59	7.01	7.26	7.38	7.41
	C <sub>19</sub>	6.23	6.59	6.06	7.06	7.01	6.90	6.85	6.80	6.87
	C <sub>20</sub>	6.97	7.41	6.75	7.62	7.80	7.23	7.92	7.93	7.47
	C <sub>21</sub>	6.42	6.63	6.64	7.18	6.76	7.13	6.92	6.53	7.09
	C <sub>22</sub>	6.84	6.32	6.86	6.91	7.52	6.84	7.17	6.79	6.82
	C <sub>23</sub>	6.06	6.22	5.86	6.24	6.52	6.00	5.97	5.84	6.33
	C <sub>24</sub>	6.42	6.49	6.42	6.41	6.29	6.46	6.93	6.83	6.43
	C <sub>25</sub>	6.04	5.83	5.98	5.58	5.69	5.73	6.00	5.66	5.75
	C <sub>26</sub>	6.30	6.89	6.41	6.05	5.48	6.00	6.51	6.43	6.79
	C <sub>27</sub>	4.33	4.26	4.59	4.02	4.02	3.84	3.82	3.83	3.92
	C <sub>28</sub>	3.98	4.11	4.16	3.54	3.37	3.44	3.37	3.61	3.36
	C <sub>29</sub>	3.57	3.37	3.64	3.07	3.12	3.12	3.01	2.82	2.97
	C <sub>30</sub>	3.09	3.26	3.10	2.80	2.75	2.70	2.71	2.65	2.74
	C <sub>31</sub>	2.82	2.86	2.93	2.56	2.37	2.62	2.49	2.39	2.62
C <sub>32</sub>	2.60	2.61	2.63	2.15	2.09	2.18	2.08	2.04	2.02	
C <sub>33</sub>	2.17	2.10	1.99	1.65	1.72	1.77	1.58	1.72	1.61	
C <sub>34</sub>	1.62	1.66	1.73	1.41	1.35	1.35	1.45	1.55	1.42	

Таблица 1 (продолжение)

Table 1 (continued)

Номер пар- тии образ- цов	<i>n</i> -парафины и изопрено- иды	Порядковые номера образцов и скважин								
		10	11	12	13	14	15	16	17	18
		скважина № 1			скважина № 2			скважина № 3		
II	C <sub>17</sub>	6.16	6.14	6.12	6.60	6.86	6.29	6.78	6.47	6.64
	Pr	5.34	5.35	5.06	6.04	5.92	6.02	4.69	4.85	4.48
	C <sub>18</sub>	6.27	6.40	6.07	6.54	6.55	6.78	6.54	6.40	6.50
	Phy	6.70	6.79	6.88	7.45	7.26	7.71	6.99	7.11	7.20
	C <sub>19</sub>	6.41	6.41	6.59	6.52	6.76	6.38	6.46	6.62	6.39
	C <sub>20</sub>	7.36	7.33	7.60	7.28	7.32	7.28	7.06	7.18	7.07
	C <sub>21</sub>	6.76	6.58	6.57	6.73	6.60	6.58	6.76	7.01	6.55
	C <sub>22</sub>	6.85	6.85	6.75	6.58	6.77	6.59	7.21	6.99	7.05
	C <sub>23</sub>	6.22	6.01	6.19	5.81	5.52	5.99	6.12	6.38	6.09
	C <sub>24</sub>	6.59	6.55	6.61	6.36	6.26	6.53	6.92	7.09	6.57
	C <sub>25</sub>	6.06	5.97	6.18	5.31	5.15	5.52	5.83	5.81	5.95
	C <sub>26</sub>	6.55	6.44	6.52	5.08	5.24	5.17	6.22	6.19	6.33
	C <sub>27</sub>	4.48	4.55	4.61	4.17	4.01	4.13	3.81	3.80	4.05
	C <sub>28</sub>	3.91	3.89	3.92	3.61	3.58	3.79	3.79	3.81	3.54
	C <sub>29</sub>	3.44	3.48	3.57	3.51	3.51	3.58	3.27	3.34	3.38
	C <sub>30</sub>	2.98	2.94	3.00	2.88	3.04	2.80	2.92	2.93	2.98
	C <sub>31</sub>	2.77	2.75	2.80	2.91	2.75	2.91	2.77	2.85	2.80
C <sub>32</sub>	2.22	2.26	2.09	2.64	2.61	2.72	2.27	2.23	2.43	
C <sub>33</sub>	1.80	1.79	1.83	2.17	2.04	2.17	1.84	1.87	1.84	
C <sub>34</sub>	1.39	1.44	1.43	1.68	1.70	1.73	1.70	1.62	1.67	

Таблица 1 (окончание)

Table 1 (end)

Номер партии образцов	Н-парафины и изопреноиды	Порядковые номера образцов и скважин									
		19	20	21	22	23	24	25	26	27	28
		скважина № 1			скважина № 2			скважина № 3			***
III	C <sub>17</sub>	5.84	6.19	5.97	6.41	6.39	6.52	6.60	6.51	6.65	6.06
	Pr	5.29	5.64	5.41	5.79	5.79	6.06	4.92	4.99	4.77	5.62
	C <sub>18</sub>	6.08	6.09	6.15	6.75	6.74	6.40	6.48	6.70	6.44	5.86
	Phy	6.91	6.82	6.90	7.24	7.17	7.16	7.01	6.83	7.24	6.78
	C <sub>19</sub>	6.55	6.25	6.44	6.78	6.56	6.73	6.60	6.45	6.81	6.24
	C <sub>20</sub>	7.50	7.65	7.43	7.04	7.47	7.03	7.50	7.40	7.62	7.81
	C <sub>21</sub>	6.79	6.57	6.57	6.73	6.50	7.11	6.73	6.60	6.65	6.73
	C <sub>22</sub>	6.89	6.94	6.97	6.77	6.85	7.02	6.85	7.12	6.73	7.15
	C <sub>23</sub>	6.34	6.27	6.17	6.04	5.64	6.16	6.08	6.09	5.97	6.08
	C <sub>24</sub>	6.92	6.40	6.70	6.39	6.53	6.30	6.66	7.00	6.38	6.71
	C <sub>25</sub>	5.96	5.91	6.01	5.57	5.56	5.78	5.95	5.89	6.06	6.09
	C <sub>26</sub>	6.18	6.04	6.23	5.69	5.57	5.84	6.47	6.85	6.30	6.17
	C <sub>27</sub>	4.12	4.27	4.20	3.98	3.92	4.00	4.00	4.07	3.78	4.17
	C <sub>28</sub>	4.03	3.84	3.94	3.58	3.72	3.38	3.57	3.60	3.65	3.90
	C <sub>29</sub>	3.52	3.43	3.51	3.29	3.20	3.44	3.17	3.16	3.22	3.53
	C <sub>30</sub>	3.14	3.13	3.15	2.91	2.80	2.93	2.88	2.84	2.98	3.08
C <sub>31</sub>	2.75	2.83	2.71	2.66	2.58	2.80	2.80	2.81	2.67	2.80	
C <sub>32</sub>	2.17	2.17	2.24	2.52	2.59	2.57	2.26	2.29	2.26	2.26	
C <sub>33</sub>	1.75	1.75	1.78	2.02	2.19	1.94	1.84	1.78	1.82	1.82	
C <sub>34</sub>	1.39	1.37	1.43	1.61	1.75	1.55	1.62	1.68	1.60	1.45	

**Примечание:** \*\*\* контрольная проба из скважины № 1, отобранная через 1.5 месяца после III партии образцов.

содержаниях парафинов в образцах нефти. Как известно, методы кластерного анализа позволяют структурировать информацию на основе разбиения ее на фрагменты таким образом, что различия характеристик представителей одного фрагмента (кластера) минимизируются при одновременной максимизации различий между разными фрагментами (кластерами), и служат одним из наиболее популярных инструментов «data mining» в геохимии [16-18]. Кластерный анализ полученных данных об относительных содержаниях парафинов в наших образцах был проведен с использованием евклидовой метрики расстояний между объектами и алгоритма Уорда для расчета иерархического дерева кластеров. Агломеративная иерархическая кластеризация по алгоритму Уорда [25] на первом шаге рассматривает каждый образец как однокомпонентный «кластер», а весь набор данных – как массив из числа «кластеров», равных количеству образцов. Затем, начиная с произвольного первого «кластера», среди всех остальных образцов алгоритм находит наиболее близкий по характеристикам к первому, используя в качестве меры сходства суммарную дисперсию признаков (среднеквадратичное отклонение), и присоединяет такой наиболее похожий образец к первому, формируя уже двухкомпонентный кластер, и уменьшая размерность исходного массива на единицу. Процедура повторяется, и к первому двух-

компонентному кластеру присоединяется следующий образец с минимальной дисперсией признаков по отношению к уже усредненным характеристикам (центроиду) первого кластера и т.д. Таким образом формируется первый кластер, в который объединяются образцы с минимальной дисперсией признаков. По этому же принципу формируется второй, третий и т.д. кластеры из оставшихся образцов, затем начинается объединение кластеров первого уровня в более крупные на основании минимизации евклидовых расстояний между центрами тяжести (центроидами) первичных кластеров. Процесс формирования последовательности вложенных кластеров, при котором число кластеров уменьшается, продолжается до тех пор, пока не будет сформирован один кластер, объединяющий все образцы.

Наиболее привлекательной особенностью кластеризации по Уорду является отсутствие априорных предположений об ожидаемой кластерной структуре массива данных (числе кластеров, степени их различия и т.п.) – она автоматически формируется в соответствии с близостью характеристик и уровнями неопределенности характеристик объектов, и их объединение в кластеры оптимизируется в смысле метода наименьших квадратов.

Графическое представление результатов кластеризации принято изображать в виде дендрограммы, иллюстрирующей состав кластеров разных

**Таблица 2**

Средние значения относительных содержаний парафинов в пробах нефти с учетом оценок доверительных интервалов, % отн.

**Table 2**

Average values of the relative paraffin amounts in oil samples considering the estimates confidence interval, % rel.

н-алканы и изопреноиды	Скважина № 1	Скважина № 2	Скважина № 3
C <sub>17</sub>	6.0 ± 0.2	6.5 ± 0.2	6.7 ± 0.4
Pг	5.3 ± 0.3	5.8 ± 0.6	4.8 ± 0.3
C <sub>18</sub>	6.2 ± 0.2	6.7 ± 0.4	6.6 ± 0.4
Phy	6.8 ± 0.3	7.3 ± 0.4	7.2 ± 0.4
C <sub>19</sub>	6.4 ± 0.2	6.7 ± 0.6	6.6 ± 0.4
C <sub>20</sub>	7.3 ± 0.6	7.3 ± 0.5	7.5 ± 0.8
C <sub>21</sub>	6.6 ± 0.1	6.8 ± 0.5	6.8 ± 0.2
C <sub>22</sub>	6.8 ± 0.3	6.9 ± 0.6	7.0 ± 0.2
C <sub>23</sub>	6.2 ± 0.3	6.0 ± 0.6	6.1 ± 0.2
C <sub>24</sub>	6.6 ± 0.3	6.4 ± 0.1	6.8 ± 0.2
C <sub>25</sub>	6.0 ± 0.2	5.5 ± 0.5	5.9 ± 0.2
C <sub>26</sub>	6.4 ± 0.5	5.6 ± 0.9	6.4 ± 0.4
C <sub>27</sub>	4.4 ± 0.4	4.0 ± 0.2	3.9 ± 0.1
C <sub>28</sub>	4.0 ± 0.2	3.6 ± 0.2	3.6 ± 0.3
C <sub>29</sub>	3.5 ± 0.1	3.3 ± 0.5	3.2 ± 0.5
C <sub>30</sub>	3.1 ± 0.2	2.9 ± 0.2	2.9 ± 0.3
C <sub>31</sub>	2.8 ± 0.2	2.7 ± 0.4	2.7 ± 0.4
C <sub>32</sub>	2.3 ± 0.6	2.4 ± 0.7	2.2 ± 0.3
C <sub>33</sub>	1.9 ± 0.4	2.0 ± 0.6	1.8 ± 0.3
C <sub>34</sub>	1.5 ± 0.4	1.6 ± 0.4	1.6 ± 0.2

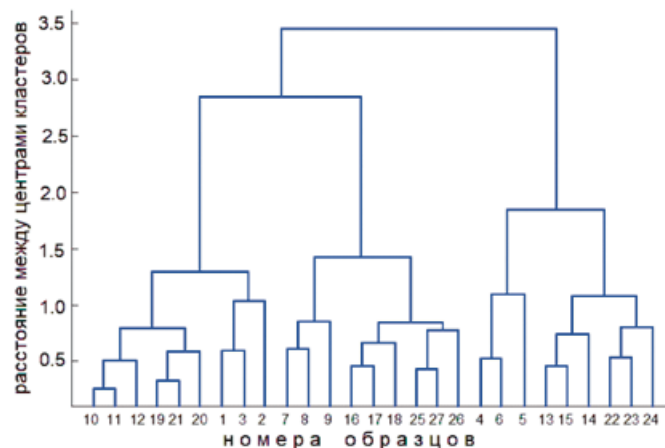
уровней и их взаимосвязи, что облегчает смысловую интерпретацию результатов. Кластеризация Уорда является одной из наиболее популярных для обработки многомерных данных в науках о Земле, том числе для решения задач нефтяной геологии [17, 18]. Так как нашей целью является исследование возможности дифференцирования образцов нефти по компонентному составу парафинов в

условиях неизвестной неопределенности данных, кластеризация на основе минимизации дисперсий в составе объединяемых в кластер образцов без априорного задания числа формируемых кластеров представляется наиболее рациональной, и поэтому был выбран метод Уорда.

На рис. 1 представлены в форме дендрограммы результаты кластерного анализа данных табл. 1, в которых происхождение каждого образца оставалось анонимным и было зашифровано указанием его порядкового номера.

Как известно, эффективность применения кластерного анализа при решении прикладных задач определяется прежде всего возможностью смысловой интерпретации полученных результатов [12-15]. Из приведенной на рис. 1 дендрограммы следует, что по относительным содержаниям парафинов все 27 образцов нефти на первом уровне кластеризации объединяются для каждой скважины в группы по 3 образца в соответствии со временами пробоотбора. На уровне дистанций между центрами кластеров ~1.8 все образцы объединяются в три кластера, что соответствует числу скважин, из которых были отобраны образцы нефти. При этом в каждый кластер объединены образцы, полученные из одной и той же скважины, а на уровне расстояния между центрами ~2.8 первые два кластера, соответствующие образцам из скважин № 1 и № 3, объединяются в один. Таким образом, полученная дендрограмма отражает геологическое происхождение образцов, т.е. их принадлежность к различным добывающим скважинам. Объединение двух кластеров в один является отражением близости состава соответствующих образцов, обусловленной происхождением из одного продуктивного нефтеносного горизонта, что соответствует имеющейся геологической информации о принадлежности двух скважин из трех к одному горизонту.

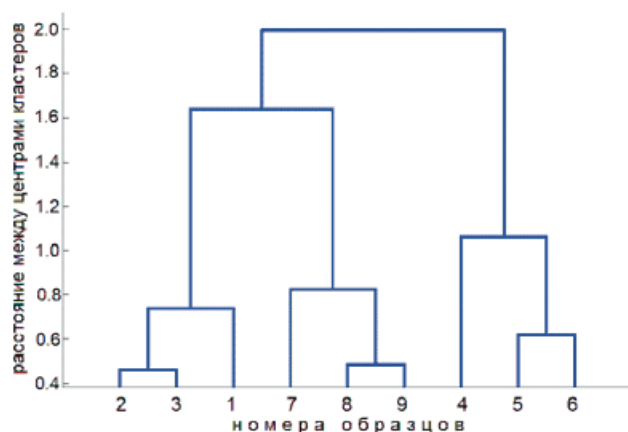
Следует отметить, что кластерная структура вариаций состава парафинов сохраняется при замене составов индивидуальных образцов на средние их



**Рис. 1.** Результаты кластерного анализа при обработке данных табл. 1, представленные в виде дендрограммы.

**Fig. 1.** Results of cluster analysis by processing Table 1 data presented as a dendrogram.





**Рис. 2.** Результаты кластерного анализа при обработке усредненных по каждому пробоотбору относительных содержаний парафинов в 9 образцах нефти из трех скважин.

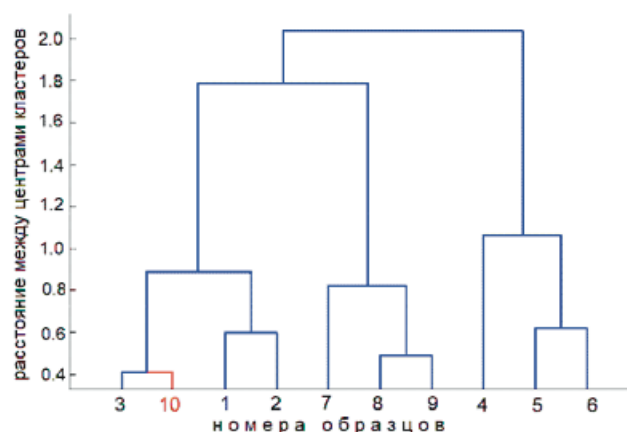
**Fig. 2.** Results of cluster analysis when processing averages over each sampling relative paraffin amounts in 9 oil samples from three wells.

значения по каждому пробоотбору. В этом случае соответствующий массив данных, извлеченных из табл. 1, имеет размерность 20×9, а результаты кластерного анализа в виде дендрограммы представлены на рис. 2. Данная дендрограмма более наглядно демонстрирует структуру вариаций состава парафинов в образцах. В этом случае образцам из скважины № 1 соответствуют порядковые номера 1, 2 и 3; из скважины № 2 – 4, 5 и 6; из скважины № 3 – 7, 8 и 9.

Иллюстрация возможности использовать полученную кластерную структуру, характеризующую составы парафинов и их вариации в образцах нефти из трех скважин, для идентификации источника образца нефти приведена на рис. 3, где приведена дендрограмма результатов кластерного анализа расширенного массива данных за счет включения в него состава парафинов контрольной пробы нефти с № 10 (№ 28 из табл. 1), отобранной из скважины № 1 через полтора месяца после окончания регулярных периодических пробоотборов. Из рис. 3 следует, что контрольный образец помещен в кластер, образованный пробами нефти из первой скважины, тем самым продемонстрирована возможность идентификации источника образца нефти по компонентному составу парафинов с применением процедур кластерного анализа.

Полученные результаты позволяют сделать следующие выводы:

– результаты ГХ/МС анализов состава парафинов методом мониторинга выбранных ионов позволили решить задачу дифференцирования источников и определения происхождения образцов путем разделения массива данных по относительным концентрациям парафинов и их вариациям на кла-



**Рис. 3.** Пример идентификации источника образца нефти (№ 10) с использованием кластерной структуры данных по относительным содержаниям парафинов в составе нефти, извлекаемой из трех скважин одного месторождения.

**Fig. 3.** Example of identifying the source of the oil sample (No. 10) using the clustered data structure for the relative paraffin amounts of oil extracted from three wells of the same field.

стеры, соответствующие геологическим источникам образцов (различным скважинам и продуктивным горизонтам);

- различия в составе парафинов в образцах нефти из трех скважин одного месторождения превышают суммарные погрешности в результатах анализов и позволяют провести дифференциацию образцов по источникам проб;
- за время наблюдений (~ 6 месяцев) возможные изменения в составе образцов нефти, добываемой из изученных скважин, оказались незначимыми и не помешали объединению образцов в кластеры, связанные с их геологическими источниками;
- объединение двух кластеров первого уровня в один означает близость состава парафинов в нефти, добываемой из первой и третьей скважин, и подтверждает ее извлечение из одного продуктивного горизонта, что согласуется с имеющейся геологической информацией о распределении скважин по нефтеносным горизонтам.

Эти результаты можно рассматривать как демонстрацию возможностей геохимических методов сопровождения процессов нефтеизвлечения для повышения эффективности и регулирования режимов эксплуатации добывающих скважин, в том числе для обнаружения межпластовых перетоков нефтяного флюида, при балансовых расчетах притоков из разных продуктивных горизонтов в процессе эксплуатации многопластовых месторождений, и как источник дополнительной геохимической информации о строении эксплуатируемых месторождений в промысловой нефтяной геологии.



## ЗАКЛЮЧЕНИЕ

В настоящей работе предложен метод контроля источников образцов нефти для их дифференциации и идентификации источника происхождения (конкретных скважин эксплуатируемого месторождения) по компонентному составу парафинов. Для решения этих задач рекомендовано использовать результаты анализа относительных содержаний *n*-парафинов  $C_{17}H_{36} \div C_{34}H_{70}$  и изопреноидных алканов пристана  $C_{19}H_{40}$  (*Pr*) и фитана  $C_{20}H_{42}$  (*Phy*) методами мониторинга выбранных ионов (ГХ/МС) в пробах нефти из разных скважин одного месторождения с применением методов «data mining» в варианте кластерного анализа для обработки данных. Такие задачи неразрешимы при прямом сопоставлении относительных содержаний индивидуальных парафинов в образцах.

Обработка результатов анализа по относительным содержаниям парафинов в 27 пробах нефти с помощью алгоритмов кластерного анализа позволила дифференцировать источники образцов по составу парафинов в масштабах одного месторождения с идентификацией конкретной скважины, из которой получен образец нефти, в условиях отсутствия оценок неопределенности в данных, т.е. погрешностей анализов.

Установлено, что для изученных трех эксплуатируемых скважин возможные изменения состава парафинов в добываемой нефти за полугодовой период наблюдений оказались незначимыми и не помешали установлению источников образцов.

Предложенный метод дифференцирования источников и идентификации образцов нефти по составу парафинов открывает новые возможности для решения практически важных задач, связанных с повышением эффективности процессов нефтедобычи путем замены дорогостоящих индикаторных методов обнаружения и мониторинга межпластовых перетоков анализами компонентного состава парафинов в извлекаемой нефти.

## ЛИТЕРАТУРА

1. Барановская В.Б., Медведевских М.Ю., Карпов Ю.А. Актуальные проблемы качества химического анализа // Аналитика и контроль. 2021. Т. 25, № 4. С. 273-279.
2. Peters K.E., Walters C.C., Moldowan J.M. The Biomarker Guide. Second ed. Cambridge University Press, Cambridge, United Kingdom, 2005. 1155 p.
3. Powell T.G., McKirdy D.M. Relationship between Ratio of Pristane to Phytane, Crude Oil Composition and Geological Environment in Australia // Nat. Phys. Sci. 1973. V. 243, № 5. P. 37-39.
4. McKirdy D.M., Aldridge A.K., Yrma P.J.M. A Geochemical Comparison of some Crude Oils from pre-Ordovician Carbonate Rocks // Advances in Organic Geochemistry. Wiley, Chichester, 1981. P. 99-107.
5. Jiang L., George S.C. Geochemical Comparison of three Oil Families from the Gippsland Basin, SE Australia // Marine and Petroleum Geology. 2020. V. 121. P. 1-23.

6. Жаманбаева А.Б., Куртова О.Ю., Алимжанова М.Б., Зенкевич И.Г. Особенности определения характеристических отношений изопреноидных алканов  $C_{15}-C_{20}$  в нефтях Казахстана // Аналитика и контроль. 2019. Т. 23, № 2. С. 237-246.
7. Темердашев З.А., Павленко Л.Ф., Корпакова И.Г., Ермакова Я.С., Экилик В.С. О некоторых методических аспектах оценки нефтяного загрязнения водных объектов с учетом деградации нефтепродуктов во времени // Аналитика и контроль. 2016. Т. 20, № 3. С. 225-235.
8. Wenlong Jiang, Ablimit Imin, Xueyong Wang, Tao Wang, Wenjian Guo. Geochemical Characterization and Quantitative Identification of Mixed-Source Oils from the Baikouquan and Lower Wuerhe Formations in the Eastern Slope of the Mahu Sag, Junggar Basin, NW China // Journal of Petroleum Science and Engineering. 2020. V. 191, 107175.
9. Титов В.И., Жданов С.А. Изменение состава пластовых нефтей при разработке месторождений (обзор) // Нефтяное хозяйство. 1988, № 8. С. 26-28.
10. Jun Niu, Haiping Huang, Wenlong Jiang. Geochemical Characteristics and Correlation of Continuous Charge Mixing and Biodegradation of Heavy Oil in Southeastern Dongying Sag, Bohai Bay basin, China // Journal of Petroleum Science and Engineering. 2018. V. 166. P. 1-12.
11. Мильман Б.Л. Введение в химическую идентификацию. С.-Пб.: ВВМ, 2008. 180 с.
12. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 2017. [Электронный ресурс]: <https://github.com/ranalytics/data-mining> (дата обращения 25.12.2021).
13. Parks J.M. Cluster Analysis Applied to Multivariate Geologic Problems // Journal of Geology. 1966. V. 74. № 5. Part 2. P. 703-715.
14. Дулин С. К., Розенберг И. Н., Уманский В. И. Методы кластеризации в исследовании массивов геоданных // Системы и средства информ. 2009. Доп. выпуск. С. 86-113.
15. Fouedjio F. A Hierarchical Clustering Method for Multivariate Geostatistical Data // Spatial Statistics. 2016. V. 18. Part B. P. 333-351.
16. Elrod L.W. Geochemistry and Petroleum Asset Life // Intertek Westport Technology Center. [Электронный ресурс]: <https://www.intertek.com/oil-and-gas/geochem-asset> (дата обращения 25.12.2021).
17. Edwards J., Lallier F., Caumon G., Carpentier C. Uncertainty Management in Stratigraphic Well Correlation and Stratigraphic Architectures: A Training-based Method // Comput. Geosci. 2017. P. 1-17.
18. Geochemical Characteristics and Genetic Relationship of the Grude Oils from the Western Part of West Siberian Basin, Russia / V. Kuznetsova [et al.] // 29th Intern. Meeting on Organic Geochemistry (Gothenburg, Sweden). Conf. Proc. 2019. Houten: EAGE, V. 2019. P. 1-2.
19. de Juan A., Casassas E., Tauler R. Soft Modeling of Analytical Data. In Encyclopedia of Analytical Chemistry / R.A. Meyers (Ed.). John Wiley & Sons Ltd, Chichester, 2000. P. 9800 - 9837.
20. Мильман Б.Л., Журкович И.К. Большие данные в современном химическом анализе // Ж. аналит. химии. 2020. Т. 74, № 4. С. 316-326.
21. Мильман Б.Л., Журкович И.К. Обобщенные критерии идентификации химических соединений методами хроматографии – масс-спектрометрии // Аналитика и контроль. 2020. Т. 24, № 3. С. 164-173.
22. Kharrat A.M., Zacharia J., Cherian V.J., Anyatonwu A. Issues with Comparing SARA Methodologies // Energy & Fuels. 2007. V. 21. № 6. P. 3618-3621.

23. Harvey D.J. Gas Chromatography | Gas Chromatography/Mass Spectrometry // Encyclopedia of Analytical Science. Third Edition. Elsevier Ltd. 2019. V. 4. P. 169-179.
24. Gosetti F., Marengo E. Mass Spectrometry | Selected Ion Monitoring // Encyclopedia of Analytical Science. Third Edition. Elsevier Ltd. 2019. V. 6. P. 500-510.
25. Ward J.H. Hierarchical Grouping to Optimize an Objective Function // J. Amer. Statist. Assoc. 1963. V. 58. Iss. 301. P. 236-244.

## REFERENCES

1. Baranovskaya V.B., Medvedevskikh M.Yu., Karpov Yu.A. [Current quality issues in chemical analysis]. *Analitika i kontrol'* [Analytics and Control], 2021, vol. 25, no. 4, pp. 273-279. doi: 10.15826/analitika.2021.25.4.005 (in Russian).
2. Peters K.E., Walters C.C., Moldowan J.M. *The Biomarker Guide*. Second ed. Cambridge University Press, Cambridge, United Kingdom, 2005. 1155 p. doi: 10.1017/CBO9781107326040
3. Powell T.G., McKirdy D.M. Relationship between Ratio of Pristane to Phytane, Crude Oil Composition and Geological Environment in Australia. *Nature Physical Science*, 1973, vol. 243, no. 5, pp. 37-39. doi: 10.1038/physci243037a0
4. McKirdy D.M., Aldridge A.K., Ypma P.J.M. A Geochemical Comparison of some Crude Oils from Pre-Ordovician Carbonate Rocks. *Advances in Organic Geochemistry*. Wiley, Chichester, 1981, pp. 99-107.
5. Jiang L., George S.C. Geochemical Comparison of three Oil Families from the Gippsland Basin, SE Australia. *Marine and Petroleum Geology*, 2020, vol. 121, pp. 1-23. doi: 10.1016/j.marpetgeo.2020.104575
6. Zhamanbayeva A.B., Kurtova O.Yu., Alimzhanova M.B., Zenkevich I.G. [Determination of C<sub>15</sub>-C<sub>20</sub> isoprenoid alkanes characteristic ratios in the oils from Kazakhstan fields]. *Analitika i kontrol'* [Analytics and Control], 2019, vol. 23, no. 2, pp. 237-246. doi: 10.15826/analitika.2019.23.2.011 (in Russian).
7. Temerdashev Z.A., Pavlenko L.Ph., Karpokova I.G., Ermakova Ya.S., Ekilik V.S. [Some methodological aspects of oil pollution evaluation of water bodies based on the degradation of petroleum products over time]. *Analitika i kontrol'* [Analytics and Control], 2016, vol. 20, no. 3, pp. 225-235. doi:10.15826/analitika.2016.20.3.006 (in Russian).
8. Wenlong Jiang, Ablimit Imin, Xueyong Wang, Tao Wang, Wenjian Guo. Geochemical Characterization and Quantitative Identification of Mixed-Source Oils from the Baikouquan and Lower Wuerhe Formations in the Eastern Slope of the Mahu Sag, Junggar Basin, NW China. *Journal of Petroleum Science and Engineering*, 2020, vol. 191, 107175. doi: 10.1016/j.petrol.2020.107175
9. Titov V.I., Zhdanov S.A. [Changing the Composition of Reservoir Oils in the Mining (Overview)]. *Neftyanoe khozyaistvo [Oil Industry]*, 1988, no. 8, pp. 26-28 (in Russian).
10. Jun Niu, Haiping Huang, Wenlong Jiang. Geochemical Characteristics and Correlation of Continuous Charge Mixing and Biodegradation of Heavy Oil in Southeastern Dongying Sag, Bohai Bay basin, China. *Journal of Petroleum Science and Engineering*, 2018, vol. 166, pp. 1-12. doi: 10.1016/j.petrol.2018.03.035
11. Milman B.L. *Vvedenie v himicheskuyu identifikatsiyu [Introduction to chemical identification]*. Saint Petersburg, VVM, 2008. 180 p. (in Russian).
12. Shitikov V.K., Mastitsky S.E. [Classification, Regression, and Other Data Mining Algorithms Using R]. 2017. Available at: <https://github.com/ranalytics/data-mining> (accessed 5 July 2021) (in Russian).
13. Parks J.M. Cluster Analysis Applied to Multivariate Geologic Problems. *Journal of Geology*, 1966, vol. 74. No. 5, part 2, pp. 703-715.
14. Dulin S.K., Rosenberg I.N., Umansky V.I. [Clustering techniques for exploring geodatabases]. *Sistemy i sredstva informatiki [Systems and Means of Informatics]*. 2009, pp. 86-113 (in Russian).
15. Fouedjio F. A Hierarchical Clustering Method for Multivariate Geostatistical Data. *Spatial Statistics*, 2016, vol. 18, part B, pp. 333-351. doi: 10.1016/j.spasta.2016.07.003.
16. Elrod L.W. Geochemistry and Petroleum Asset Life. Intertek Westport Technology Center, 2021. Available at: <https://www.intertek.com/oil-and-gas/geochem-asset> (accessed 25 December 2021).
17. Edwards J., Lallier F., Caumon G., Carpentier C. Uncertainty Management in Stratigraphic Well Correlation and Stratigraphic Architectures: A Training-based Method. *Computer & Geosciences*, 2017, pp. 1-17. doi: 10.1016/j.cageo.2017.10.008.
18. Kuznetsova V., Ablya E., Manko I., Baniasad A., Littke R. Geochemical Characteristics and Genetic Relationship of the Grude Oils from the Western Part of West Siberian Basin, Russia. *29th Intern. Meeting on Organic Geochemistry. Conference Proceedings*. Houten: EAGE, 2019, vol. 2019, pp. 1-2. doi: 10.3997/2214-4609.201902738
19. de Juan A., Casassas E., Tauler R. Soft Modeling of Analytical Data. In *Encyclopedia of Analytical Chemistry* / R.A. Meyers (Ed.). John Wiley & Sons Ltd, Chichester, 2000, pp. 9800-9837. doi: 10.1002/9780470027318.a5208
20. Milman B.L., Zhurkovich I.K. Big Data in Modern Chemical Analysis. *Journal of Analytical Chemistry*, 2020, vol. 75, pp. 443-452. doi: 10.1134/S1061934820020124
21. Milman B.L., Zhurkovich I.K. [Summarized criteria of chemical compounds identification by chromatography-mass spectrometry]. *Analitika i kontrol'* [Analytics and Control], 2020, vol. 24, no. 3, pp. 164-173. doi: 10.15826/analitika.2020.24.3.003 (in Russian).
22. Kharrat A.M., Zacharia J., Cherian V.J., Anyatonwu A. Issues with Comparing SARA Methodologies. *Energy & Fuels*, 2007, vol. 21, no. 6, pp. 3618-3621. doi: 10.1021/ef700393a.
23. Harvey D.J. Gas Chromatography/Mass Spectrometry. *Encyclopedia of Analytical Science*. Third Edition. Elsevier Ltd. 2019, vol. 4, pp. 169-179.
24. Gosetti F., Marengo E. Selected Ion Monitoring. *Encyclopedia of Analytical Science*. Third Edition. Elsevier Ltd. 2019, vol. 6, pp. 500-510. doi: 10.1016/B978-0-12-409547-2.14418-X
25. Ward J.H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 1963, vol. 58, iss. 301, pp. 236-244.