

## Hard- и Soft-PLS-DA методы в идентификации нефтепродуктов на основе данных газовой хроматографии

**\*А.С. Мухаметганеева, А.Р. Байгузина**

*Экспертно-криминалистическая служба – региональный филиал  
Центрального экспертно-криминалистического таможенного управления  
г. Екатеринбург  
Российская Федерация, 450112, г. Уфа, ул. Калинина, 29*

*\*Адрес для переписки: Мухаметганеева Альфия Салиховна, E-mail: alfiya.mukhametganeeva@mail.ru*

Поступила в редакцию 15 сентября 2023 г., после доработки 29 ноября 2023 г.

Таможенная экспертиза для идентификации общей или различной групповой принадлежности нефтепродуктов в целях Товарной номенклатуры внешнеэкономической деятельности Евразийского экономического союза (ТН ВЭД ЕАЭС) является важной задачей и использование хемометрического подхода при изучении химического состава дает дополнительные возможности интерпретации получаемых хроматографических данных. В основу предложенного варианта идентификации входят предварительное разделение основных компонентов методом газожидкостной хроматографии и обработка полученных данных с помощью хемометрического метода анализа – дискриминантного анализа с помощью регрессии на латентные структуры (PLS-DA). В статье показана целесообразность и эффективность использования хемометрических методов в решении задач идентификации на уровне установления общей групповой принадлежности нефтепродуктов. Метод PLS-DA является наглядным при визуальной оценке и группировке объектов, имеющих близкие показатели компонентного состава. Рассмотрено применение метода Soft-PLS-DA при проведении сравнительного исследования 31 анализа (пробы нефтепродуктов: автомобильные бензины разных марок, дистилляты газового конденсата, фракции ароматических углеводородов) по данным газовой хроматографии с пламенно-ионизационным детектором (ГХ-ПИД). Полученная модель позволяет идентифицировать образцы нефтепродуктов при наличии соответствующих данных по выборкам (коллекциям). Проведена проверка полученной модели путем включения в набор двух дополнительных образцов (выбросов), отличающихся по своему составу (характеристикам) от других объектов, которая показала достоверность и устойчивость модели. Результаты проведенного анализа свидетельствуют о перспективности применения хемометрического подхода для первичной идентификации и группировки проб нефтепродуктов по результатам ГХ-ПИД с применением метода Soft-PLS-DA в экспертных приложениях.

**Ключевые слова:** хемометрика, газовая хроматография, ГХ-ПИД, идентификация нефтепродуктов, метод проекций на латентные структуры – дискриминантный анализ (PLS-DA)

For citation: Analitika i kontrol' [Analytics and Control], 2023, vol. 27, no. 4, pp. 276-284

DOI:10.15826/analitika.2023.27.4.009

## Hard- and Soft-PLS-DA methods in the identification of petroleum products on the basis of gas chromatography data

**\*A.S. Mukhametganeeva , A.R. Bayguzina**

*Yekaterinburg Regional Branch of Central Expert-Criminalistic Customs Administration  
29 Kalinin Str., Ufa, 450142, Russian Federation*

*\*Correspondence: Alfiya S. Mukhametganeeva, E-mail: alfiya.mukhametganeeva@mail.ru*

Submitted 15 September 2023, received in revised form 29 November 2023

Customs examination for the identification of general or various group affiliation of petroleum products for the purposes of the Commodity Nomenclature of Foreign Economic Activity of the Eurasian Economic

Union (THE CUSTOMS CODE OF THE EAEU) is an important task, and the use of a chemometric approach in the study of chemical composition provides additional opportunities for interpretation of the chromatographic data obtained. Proposed identification technique is based on preliminary separation of main components by gas-liquid chromatography and processing the obtained data using a chemometric analysis method – discriminant analysis using regression on latent structures (PLS-DA). The article demonstrates the appropriateness and effectiveness of applying chemometric methods for solving identification problems at the level of establishing the general group affiliation of petroleum products. The PLS-DA method is clear for visual assessment and grouping of objects with similar component composition indicators. Application of the Soft-PLS-DA method in performing a comparative study of 31 analytes (petroleum product samples: automotive gasoline of different brands, gas condensate distillates, aromatic hydrocarbon fractions) using gas chromatography with a flame-ionization detector (GC-FID) is considered. The obtained model allows identifying samples of petroleum products if the data on the corresponding sets of samples (collections) are available. The obtained model was tested by including two additional samples (outliers) in the set, which differ in composition (characteristics) from other objects, and the reliability and stability of the model was demonstrated. The results of the analysis confirmed the prospects for using chemometric approach for primary identification and grouping of petroleum product samples on the basis of GC-FID results using the Soft-PLS-DA method in expert applications.

**Keywords:** chemometrics, gas chromatography, GC-FID, identification of petroleum products, partial least squares–discriminant analysis (PLS-DA)

## ВВЕДЕНИЕ

Хемометрика – это синтетическая дисциплина, находящаяся на стыке химии и математики. Согласно определению Д. Массарта (D. Massart) хемометрика – это химическая дисциплина, применяющая математические, статистические и иные методы, основанные на формальной логике, для построения (или отбора) оптимальных методов измерения (и планирования эксперимента), а также для извлечения наиболее важной информации из экспериментальных данных [1]. Анализ химических данных – это важнейшая часть хемометрики. В последнее время это направление очень быстро и плодотворно развивается, предлагая аналитикам не только новые методы обработки данных, но и новые подходы к постановке экспериментов [1 – 3].

Хемометрические подходы находят применение в различных областях. Математическому аппарату хемометрики посвящено большое количество статей, обзоров, монографий и учебных пособий [1 – 11].

Как известно, для получения полезной информации в хемометрике используют методы сжатия данных (в отличие от классического варианта, когда из массива данных выделяют отдельные, особо значимые результаты измерения). Сущность этих методов заключается в том, что исходные данные представляют таким образом, чтобы выявить новые скрытые переменные. Методы сжатия данных позволяют представить полезную информацию в более компактном виде, удобном для визуализации и интерпретации. Еще одной важной особенностью этих методов является рассмотрение многофакторных данных в совокупности, без выделения предварительно одного или нескольких значимых данных или переменных [2 – 4].

Простейший случай – это одномерные данные, то есть просто одно число, например значение оптической плотности, которое может быть получено на монохроматическом фотометре. Более сложный случай – это многомерные, одномерные данные,

то есть набор из нескольких измерений, относящихся к одному образцу. Примерами таких данных являются спектр или хроматограмма.

В результате построенные многомерные эмпирические модели позволяют обнаружить новые связи или явления, так как учитывают скрытые объективно существующие связи внутри объекта [2 – 4].

При производстве таможенных экспертиз для идентификации в целях ТН ВЭД ЕАЭС [12] экспертам приходится решать вопрос общей или различной групповой принадлежности нефтепродуктов. Большинство исследований нефтепродуктов основывается на использовании хроматографических методов анализа, которые позволяют достоверно определить качественный и количественный состав нефтепродуктов [13]. Газовая хроматография с пламенно-ионизационным детектированием (ГХ-ПИД) один из наиболее популярных аналитических методов, так как ГХ-ПИД является стандартным подходом для первоначального скрининга проб нефти и нефтепродуктов [14, 15].

Как известно, многокомпонентные смеси нефтепродуктов можно анализировать как с разделением компонентов с помощью высокоэффективной жидкостной хроматографии (ВЭЖХ), ГХ, капиллярного электрофореза, так и без разделения методами спектрофотометрии в УФ, видимой или ИК-области и спектрофлуориметрии. В зависимости от типа решаемой задачи применяют один из этих подходов. Для определения единичного вещества (аналита) или группы родственных аналитов используют как спектральные, так и хроматографические методы. Преимущества хроматографических методов реализуются в полном объеме при решении задач другого типа – при одновременном и отдельном определении ряда аналитов, а также при проведении структурно-группового анализа. Следовательно, основным способом структурно-группового анализа бензинов стала газовая хроматография (ГХ-ПИД).

Хроматографическое разделение компонентов перед их детектированием снимает наиболее сложную проблему присущую молекулярно-спектроскопическим методам. Так, при использовании спектрофотометрических методов анализа зачастую наблюдаются спектральные наложения сигналов разных компонентов, что осложняет интерпретацию результатов.

Использование хемометрических методов в сочетании с новыми разработками в области газовой хроматографии и масс-спектрометрии (ГХ-МС, ГХ-ПВД, ГХ-ГХ и др.) позволили повысить чувствительность, селективность при идентификации нефти и нефтепродуктов [14 – 20].

Для решения задач идентификации на уровне установления общей групповой принадлежности были применены методы главных компонент (РСА) и кластерного анализа при проведении сравнительного исследования образцов газового конденсата по данным газожидкостной хроматографии [21]. Метод РСА также был использован для классификации и идентификации бензинов на основе данных, полученных методом ГХ-МС [22, 23].

Двумерную комплексную газовую хроматографию (ГХ-ГХ) в сочетании с методом РСА применяли для решения задачи распознавания реактивных топлив [24].

Несмотря на то, что газовая хроматография является одной из наиболее востребованных при анализе нефтепродуктов при проведении судебных экспертиз [25], количество работ по идентификации нефтепродуктов по данным ГХ с хемометрической обработкой методами частичных наименьших квадратов (PLS) или дискриминационного анализа на основе частичных наименьших квадратов (PLS-DA) ограничено, значимой среди них является работа [26], в которой разработан метод выявления разливов в почве различных видов топлив (ракетного керосина, авиационного и дизельного топлива). Предлагаемый вариант идентификации основан на предварительном разделении основных компонентов методом газовой хроматографии и их масс-спектрометрическом детектировании с последующей обработкой данных, полученных хемометрическими методами анализа (анализ главных компонент и прогнозирование на скрытые структуры с дискриминантным анализом). Основным преимуществом метода PLS-DA является возможность проведения одновременного многоклассового анализа [4].

Цель настоящей работы заключалась в применении хемометрического подхода для первичной идентификации и группировки проб нефтепродуктов, наиболее часто встречающихся в экспертной практике таможенных экспертов (автомобильные бензины разных марок, дистилляты газового конденсата, фракции ароматических углеводородов) по результатам ГХ-ПВД с применением метода PLS-DA.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

### Реагенты и оборудование

Для установления количественного состава пробы нефтепродуктов в нативном виде исследовали методом газовой хроматографии с использованием хроматографа «Хроматэк-Кристалл 5000.2» (ЗАО СКБ «Хроматэк», г. Йошкар-Ола, Россия) при следующих технических условиях: программа для управления и сбора данных – Хроматэк-Аналитик 2.6 (детектор пламенно-ионизационный ПВД, кварцевая капиллярная колонка CR-1 PONA 100 м\*0.25 мм\*0.5 мкм), в следующих условиях: температура инжектора – 250 °С, интерфейса детектора – 275 °С; линейная скорость газа носителя (гелия) – 30 см/с; объем вводимой пробы – 1,0 мкл; режим инжектора – деление потока 1:201.4; анализ проведен в режиме программирования температуры колонки: 1) начальная температура – 35 °С (время выдержки 13 мин); 2) подъем температуры от 35 °С до 45 °С со скоростью 10 °С/мин (время выдержки 15 мин), 3) подъем температуры от 45 °С до 60 °С со скоростью 1 °С/мин (время выдержки 15 мин), 4) подъем температуры от 60 °С до 200 °С со скоростью 2 °С/мин (время выдержки 40 мин). Пробу вводили в автоматическом режиме (autosampler).

### Материалы

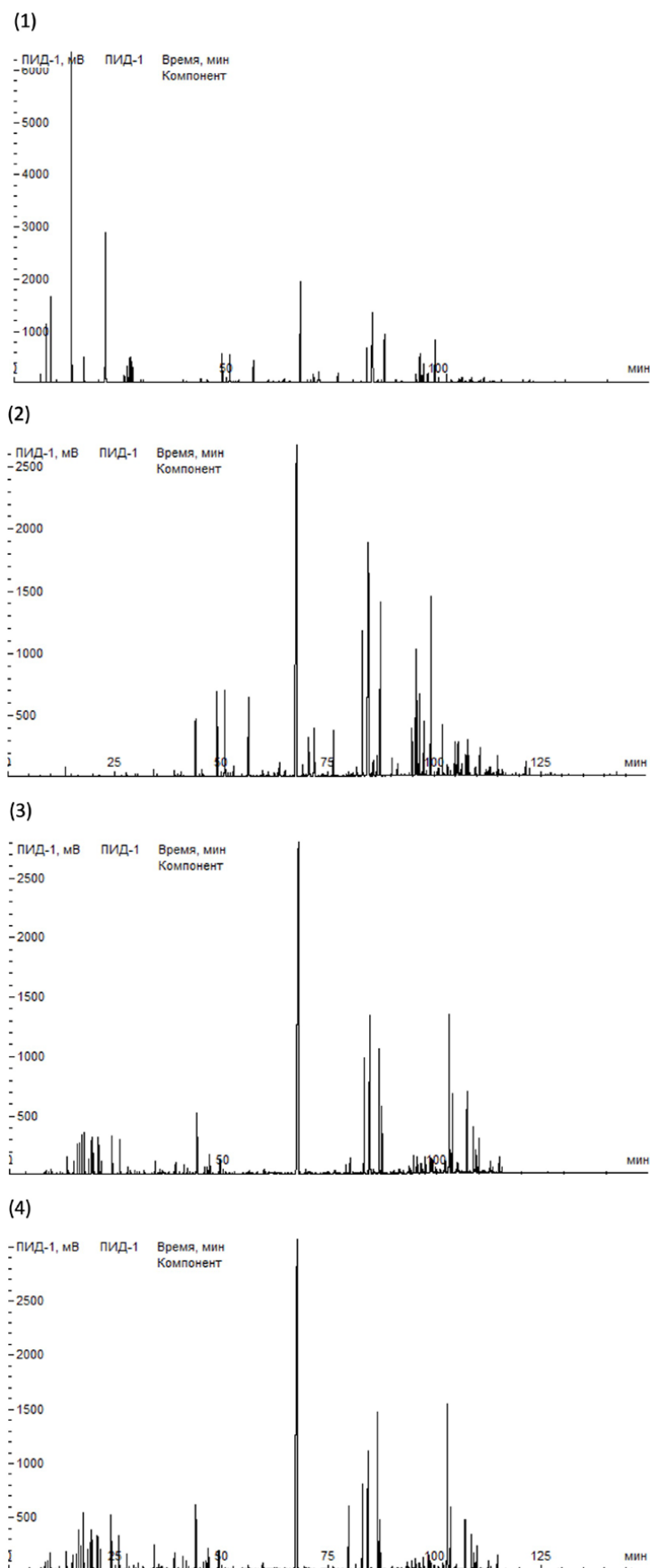
Данными для исследования является содержание (массовая доля, %) различных углеводородов (парафины и изопарафины; олефины; ароматические углеводороды), идентифицированных методом ГХ-ПВД.

Для разработки первичного скрининга нефтепродуктов методом газовой хроматографии с целью идентификации были использованы 33 пробы нефтепродуктов: 16 проб автомобильного неэтилированного бензина марок АИ-92 и АИ-95 производства ООО «Газпром нефтехим Салават», ПАО «Орскнефтеоргсинтез» и АО «Газпром-нефть-ОНПЗ»; 8 проб дистиллята газового конденсата марки Б производства ООО «Газпром нефтехим Салават»; 7 проб фракций ароматических углеводородов производства АО «Уралоргсинтез», 2 пробы пиро-конденсата производства ООО «ЗапСибнефтехим».

Для всех проб получены ГХ-хроматограммы (ГХ-ПВД), типичные представлены на рис. 1. Полные результаты газохроматографического исследования проб приведены в табл. 1. Полученные характеристики предложены в качестве параметров для моделирования.

### Методы

Для экспресс-идентификации проб нефтепродуктов в условиях небольшого размера выборки был выбран метод ГХ-ПВД в сочетании с методом PLS-DA.



**Рис. 1.** Типичные хроматограммы ГХ-ПИД нефтепродуктов: автомобильный бензин (1), дистиллят газового конденсата (2), фракция ароматических углеводородов (3), пироконденсат (4)

**Fig. 1.** Typical GC-FID chromatograms of petroleum products: motor gasoline (1), gas condensate distillate (2), aromatic hydrocarbons fraction (3), pyrocondensate (4)

Таблица

Результаты газохроматографического исследования нефтепродуктов

Table

Results of gas chromatographic analysis of petroleum products

№ пробы	Массовая доля, %																		
	н-парафины							изо-парафины							олефины		ароматические углеводороды		
	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	C <sub>5</sub>	C <sub>6</sub>	бензол	толуол	этилбензол
1	0.00	0.06	0.06	0.02	0.01	1.26	0.05	0.01	0.32	0.11	4.92	2.06	1.16	1.50	0.01	3.05	0.33	26.92	22.38
2	0.00	0.06	0.06	0.02	0.01	1.30	0.46	0.01	0.30	0.12	4.88	2.12	1.18	1.84	0.01	0.00	0.25	27.33	22.66
3	0.00	0.01	0.09	0.02	0.01	1.30	0.42	0.02	0.41	0.17	4.42	2.11	1.11	1.02	0.01	0.04	0.26	26.21	23.16
4	0.00	0.00	0.03	0.02	0.01	1.41	0.46	0.00	0.09	0.06	4.70	2.31	1.18	1.08	0.00	0.00	0.32	25.28	23.47
5	0.00	0.00	0.02	0.02	0.01	1.18	0.36	0.00	0.05	0.05	5.54	1.90	0.93	0.29	0.00	0.00	0.80	27.73	22.07
6	0.00	0.00	0.34	0.08	0.02	1.08	0.32	0.00	0.18	0.07	7.54	1.76	0.82	1.71	0.00	0.00	1.95	26.39	21.92
7	0.00	0.00	0.03	0.21	0.01	1.33	0.45	0.00	0.17	0.13	7.27	2.07	1.08	0.33	0.01	0.01	1.68	22.50	22.03
8	0.00	0.03	0.05	0.05	2.35	1.91	0.56	0.03	0.69	0.13	1.26	8.55	2.19	0.32	0.08	0.24	0.00	30.05	26.36
9	0.12	2.37	1.55	1.40	0.01	0.68	0.13	0.92	20.63	12.11	6.57	1.15	0.22	0.03	0.36	1.41	0.99	12.83	13.96
10	0.14	2.11	1.24	1.30	0.01	0.69	0.13	0.99	18.56	12.62	6.93	2.05	0.12	0.02	0.26	1.87	0.82	12.51	11.96
11	0.15	3.05	1.34	1.44	0.01	0.73	0.14	1.51	17.94	12.22	6.97	2.52	0.45	0.03	0.04	1.73	0.92	12.70	14.94
12	0.06	1.89	1.24	1.18	0.01	0.77	0.15	0.78	19.33	13.19	6.65	2.68	0.30	0.03	0.04	1.48	0.99	12.56	14.20
13	0.00	0.63	3.68	0.52	0.07	0.81	0.32	0.528	19.73	3.50	0.33	2.38	0.74	0.35	2.44	4.03	0.20	12.74	9.97
14	0.00	0.15	2.34	0.94	0.09	0.94	0.38	0.11	11.07	2.38	0.40	4.32	1.57	0.58	2.11	4.21	0.22	12.72	13.27
15	0.00	0.28	2.52	0.78	0.08	0.84	0.30	0.24	14.60	4.70	0.38	4.54	1.59	0.63	2.81	4.81	0.18	10.07	10.20
16	0.12	3.22	1.62	1.17	0.03	0.79	0.22	0.89	17.33	8.08	0.68	4.08	0.93	0.30	1.54	3.18	0.108	11.17	9.59
17	0.00	0.65	3.14	0.93	0.07	0.80	0.27	0.67	19.80	4.27	0.36	3.97	1.34	0.51	2.80	4.08	0.20	11.85	10.30
18	0.00	1.82	0.99	0.02	0.01	0.75	0.19	0.33	24.96	11.38	3.75	0.94	0.17	0.04	0.30	3.58	0.04	12.98	13.46
19	0.01	0.54	3.42	1.06	0.07	0.79	0.27	0.51	16.67	3.91	0.44	3.47	0.65	0.1940	2.09	4.61	0.16	11.21	8.61
20	0.01	0.52	3.23	1.02	0.07	0.85	0.29	0.42	17.60	4.26	0.44	3.65	0.57	0.54	2.07	3.52	0.16	10.81	11.66
21	0.00	1.20	7.54	3.32	0.02	1.20	0.24	0.15	8.68	2.65	0.36	1.66	0.67	0.05	0.38	7.20	0.08	14.07	15.27
22	0.00	0.77	4.08	1.98	0.05	0.90	0.30	0.53	9.05	3.25	0.78	3.15	1.30	0.53	2.32	6.26	0.13	10.52	9.56
23	0.00	0.60	3.29	1.38	0.27	0.72	0.27	0.26	6.00	1.53	0.83	2.30	0.99	0.80	5.64	8.65	0.62	8.77	8.74
24	0.00	2.51	0.75	0.80	0.07	0.61	0.36	0.60	11.26	14.41	10.24	1.72	2.37	0.82	4.15	4.04	0.49	10.24	4.32
25	0.00	0.12	1.70	1.28	0.01	1.70	0.16	0.01	0.83	0.71	0.97	0.40	6.89	1.31	5.65	1.51	1.29	25.59	10.45
26	0.00	0.11	1.69	1.16	0.01	0.14	0.09	0.01	1.53	0.75	1.08	0.36	5.01	1.32	6.52	1.36	1.41	27.65	10.72
27	0.00	0.05	0.91	0.55	0.01	0.13	0.08	0.00	0.47	0.53	1.06	0.38	4.36	1.26	6.63	1.51	1.49	29.34	12.57
28	0.01	0.26	0.74	1.64	0.45	0.40	0.19	0.04	0.95	0.73	0.61	0.24	6.63	1.45	6.06	1.82	5.84	26.28	9.63
29	0.00	0.27	0.80	1.75	0.94	0.40	0.18	0.04	1.02	0.79	1.03	0.32	6.52	1.38	6.56	1.84	6.36	26.21	10.22
30	0.00	0.12	1.17	0.84	0.01	0.14	0.82	0.01	0.78	0.94	1.29	0.33	6.30	1.00	8.32	1.47	1.09	26.71	6.48
31	0.00	0.11	1.96	1.16	0.00	1.14	0.09	0.01	1.53	0.75	1.08	0.36	5.01	1.32	6.52	1.50	1.41	27.65	12.22
32	0.00	0.19	3.44	0.03	0.03	0.01	0.00	0.00	0.54	0.10	2.09	0.023	0.19	0.34	7.86	3.71	43.12	9.60	5.92
33	0.00	0.15	3.82	0.36	0.02	0.09	0.00	0.00	0.60	0.08	2.07	0.023	0.19	0.36	6.91	3.71	43.49	9.76	6.31

Перед применением PLS-DA в работе использовали метод простой нормировки данных. Он основан на предположении о том, что все вещества, взятые в равном количестве, дают одну и ту же площадь пика, независимо от их строения. Это выполняется в том случае, если вещества химически сходны, а в качестве газа-носителя применяют газ с высокой теплопроводностью (в частности, гелий). Для использования этого метода необходимо, чтобы на хроматограмме были зарегистрированы все компоненты, входящие в состав анализируемой смеси. Сумму площадей всех пиков принимают за 100 %.

Статистическую обработку данных газохроматографического анализа проводили с использованием шаблонов Excel [27, 28].

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

При обработке и интерпретации данных, необходимых для сравнительного анализа, возникают сложности, связанные с большим количеством компонентов в составе нефтепродуктов. На правильность конечных результатов влияет не только методология математической обработки, но и процесс непосредственного получения хроматограмм, а также единообразие их автоматической разметки. Так, в работе [26] предложен метод для идентификации разливов ракетных и авиационных керосинов в почвах, авторами установлено, что наиболее подходящими методами обработки данных газовой хроматографии с масс-селективным детек-



тированием (ГХ-МС) являются PCA и PLS-DA. Оба хемометрических метода (PCA и PLS-DA) успешно классифицировали все исследуемые виды топлив, так как зоны их распространения не пересекались и были достаточно удалены, но в случае PLS-DA границы разброса переменных были более узкими и четкими.

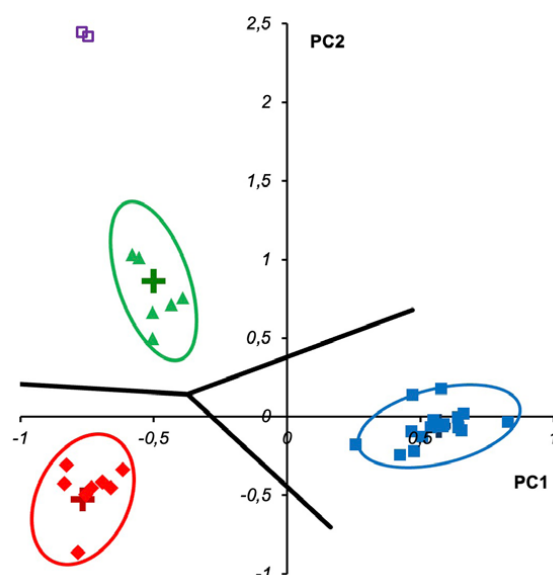
Метод проекций на латентные структуры – дискриминантный анализ (PLS-DA) является наглядным при визуальной оценке и группировке объектов, имеющих близкие показатели компонентного состава. PLS-DA является инструментом жесткой классификации, главный недостаток которого заключается в его неспособности правильно классифицировать образцы, не принадлежащие ни к одному из заранее определенных классов. Причина – отсутствие мягких решений типа «образец вообще не принадлежит ни одному классу». Поэтому можно рассмотреть 2 подхода к классификации: «строгий» или «жесткий» (Hard) и «нестрогий» или «мягкий» (Soft) способ распределения выборки по целевым классам. Первый метод предполагает, что каждая выборка в обязательном порядке относится только к одному классу. Второй позволяет разделить выборку более чем на один класс или даже оставить ее неклассифицированной. Основываясь на этой концепции, Померанцев А.Л. и Родионова О.Е. предложили два метода PLS-DA: традиционный Hard-PLS-DA и новый подход Soft-PLS-DA [4].

Для построения PLS-DA модели использовали обучающие наборы образцов, которые предварительно были разделены на базовые группы: фракции ароматических углеводородов, автомобильные бензины, дистилляты газового конденсата. Обучающая группа содержит 21 образец: 12 образцов бензинов, 5 образцов дистиллята газового конденсата, 4 образца фракции ароматических углеводородов. В данной работе кодировка экспериментальных данных не проводилась, так как данные изначально имеют числовое значение. Нормировка и выравнивание пиков не проводились, так как использовался автоматический ввод пробы.

В данной работе для идентификации нефтепродуктов на основе данных газовой хроматографии мы применили Hard- и Soft- версии метода PLS-DA. Визуализация примененных версий метода PLS-DA показана на рис. 2.

На рис. 2 видно, что базовые группы не перекрываются между собой и расположены в разных областях двухмерной плоскости, что подтверждает различия между ними, как при использовании Hard-PLS-DA метода, так и Soft- PLS-DA метода. Цвет и форма значков соответствуют классу: красный ромб – дистиллятов газового конденсата, синий квадрат – автомобильных бензинов, зеленый треугольник – фракций ароматических углеводородов, фиолетовый квадрат – пироконденсатов.

В случае применения версии Hard-PLS-DA двухмерная плоскость «строго» разделена прямыми



**Рис. 2.** Классификация нефтепродуктов с использованием методов Hard- и Soft-PLS-DA

- - бензины,
- ▲ - фракции ароматических углеводородов,
- ◆ - дистиллят газового дистиллята,
- - пироконденсаты

**Fig. 2.** Classification of petroleum products using Hard- and Soft-PLS-DA methods

- - motor gasolines,
- ▲ - aromatic hydrocarbon fraction,
- ◆ - gas condensate distillate,
- - pyrocondensate

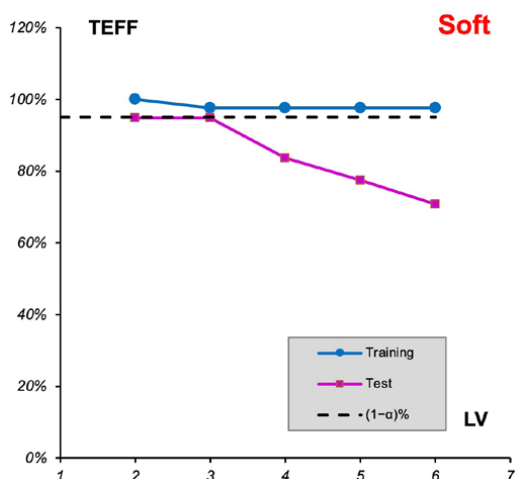
на 3 сектора, в одном из них сгруппированы бензины, во втором – фракции ароматических углеводородов, а в третьем секторе – дистилляты газового конденсата, при этом классы нефтепродуктов не перекрываются между собой и располагаются в разных областях, что подтверждает различия между ними (рис. 2).

При применении новой версии PLS-DA, называемой Soft-PLS-DA [4], дискриминация нефтепродуктов проводится «нестрого», с помощью доверительных эллипсов, что является главным отличием от традиционного Hard-PLS-DA.

Для выбора оптимального числа LV (где, LV – число латентных переменных) был использован проверочный набор из трех базовых групп состоящий из 10 образцов: 4 образца бензина, 3 образца дистиллята газового конденсата, 3 образца фракции ароматических углеводородов. Из графика (рис. 3), отражающего полную эффективность (TEFF) для метода Soft-PLS-DA с уровнем достоверности  $\alpha = 0.05$ , видно, что оптимальное LV равно 3. После этой точки эффективность для проверочного набора начинает падать.

Затем экспериментальные данные из проверочного набора добавили в обучающий набор, на основе которого построили PLS-DA модель с LV=3.

Качество полученной модели было проверено на тестовом наборе, полученная чувствительность



**Рис. 3.** График полной эффективности для метода Soft-PLS-DA классификации по результатам ГХ-ПИД

**Fig. 3.** Total efficiency graph for Soft-PLS-DA method of classification by GC-FID results

SNS для метода Hard-PLS-DA составила 100 %, специфичность SPC – 100 %, чувствительность SNS для метода Soft-PLS-DA – 90 %, специфичность SPC – 100 % (рис. 4, 5). При этом чувствительность для метода Soft-PLS-DA соответствует выбранной нами достоверности  $\alpha = 0.05$ , что подтверждает правильность этого метода.

Таким образом, исследования, проведенные на определенной ограниченной выборке, позволяют отметить высокую чувствительность газовой хроматографии относительно компонентов нефтепродуктов (автомобильные бензины разных

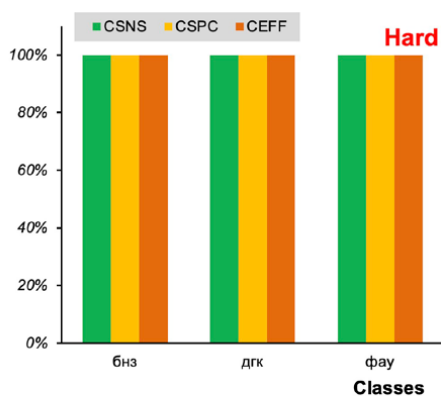
марок, дистилляты газового конденсата, фракции ароматических углеводородов) как к колебаниям их компонентного состава, так и к изменениям концентраций отдельных компонентов.

Как видно из рис. 2 применение обеих версий метода PLS-DA при ГХ-анализе индивидуальных нефтепродуктов (автомобильные бензины разных марок, дистилляты газового конденсата, фракции ароматических углеводородов), позволяет осуществить их сравнение и объединить в группы, обладающие общими количественными и качественными признаками.

Полученная модель позволяет сравнивать неизвестные образцы с контрольными, существующими в базе данных, и на этой основе классифицировать новые образцы, поступающие для проведения экспертных исследований. При помощи нее легко выявляются выбросы – объекты, которые существенно отличаются по своему составу (характеристикам) от других объектов, предоставленных на исследование.

Для подтверждения последнего положения исследованы две пробы пироконденсата (Таблица, № проб 32, 33). Как видно из рис. 2, пироконденсаты не попадают ни в одну группу при классификации методом Soft-PLS-DA, при применении метода Hard-PLS-DA пироконденсаты оказываются в одной группе с фракцией ароматических углеводородов. Метод Soft-PLS-DA проявил себя как более селективный метод дискриминации аналитов.

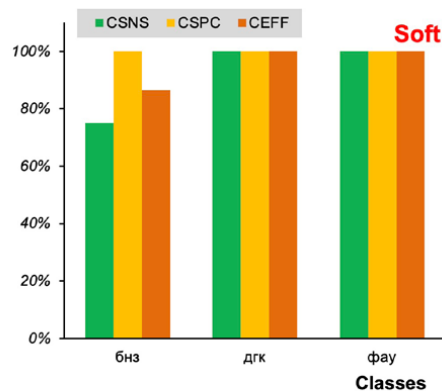
Следует отметить, что на данном этапе развития общедоступных основ хемометрики подобный подход может применяться лишь на начальном



Общая чувствительность	TSNS	100%
Общая специфичность	TSPC	100%
Общая эффективность	TEFF	100%

**Рис. 4.** Диаграмма общей чувствительности, специфичности, эффективности для метода Hard-PLS-DA CSNS – классовая чувствительность, CSPC – классовая специфичность, CEFF – классовая эффективность

**Fig. 4.** Diagram of the total sensitivity, specificity and efficiency of the Hard-PLS-DA method CSNS – class sensitivity, CSPC – class specificity, CEFF – class efficiency



Общая чувствительность	TSNS	90%
Общая специфичность	TSPC	100%
Общая эффективность	TEFF	95%

**Рис. 5.** Диаграмма общей чувствительности, специфичности, эффективности для метода Soft-PLS-DA CSNS – классовая чувствительность, CSPC – классовая специфичность, CEFF – классовая эффективность.

**Fig. 5.** Diagram of the total sensitivity, specificity and efficiency of the Soft-PLSDA method CSNS – class sensitivity, CSPC – class specificity, CEFF – class efficiency

этапе сравнения – при выборе объектов, которые могут относиться к одной или разным группам, чтобы существенно сократить время, необходимое для проведения идентификации. Окончательное решение должно приниматься непосредственно экспертом.

## ЗАКЛЮЧЕНИЕ

В настоящей работе показано применение метода проекций на латентные структуры PLS-DA при проведении сравнительного исследования 31 анализа (автомобильный бензины разных марок, дистиллят газового конденсата, фракции ароматических углеводородов) по данным только газовой хроматографии (ГХ-ПИД).

Нами разработан подход для первичной идентификации проб нефтепродуктов на основе данных газовой хроматографии (ГХ-ПИД) на примере нефтепродуктов, таких как автомобильные бензины разных марок, дистилляты газового конденсата, фракции ароматических углеводородов с помощью хемометрического метода PLS-DA (PLS-Discriminant Analysis). Представленная модель обеспечивает хорошую идентификацию и группировку нефтепродуктов по общим признакам, что свидетельствует о ее достоверности.

## БЛАГОДАРНОСТИ

Авторы выражают благодарность за помощь в подготовке настоящей работы доктору физико-математических наук, главному научному сотруднику Института химической физики им. Н.Н. Семенова Российской Академии наук Алексею Леонидовичу Померанцеву и доктору физико-математических наук, главному научному сотруднику Института химической физики им. Н.Н. Семенова Российской Академии наук Оксане Евгеньевне Родионовой.

## ACKNOWLEDGEMENTS

The authors express their gratitude for the help in preparing this work to Doctor of Physical and Mathematical Sciences, Principal Researcher of the Institute of Chemical Physics named after N.N. Semenov of the Russian Academy of Sciences Aleksey L. Pomerantsev and Doctor of Physical and Mathematical Sciences, Principal Researcher of the Institute of Chemical Physics named after N.N. Semenov of the Russian Academy of Sciences Oksana E. Rodionova.

## ЛИТЕРАТУРА

1. Chemometrics / D.L. Massart [et al.]. New York, Elsevier. 1988. 488 p.
2. Родионова О.Е. Хемометрический подход к исследованию больших массивов химических данных // Рос. хим. ж. 2006. Т. 50, № 2. С. 128-144.

3. Родионова О.Е., Померанцев, А.Л. Хемометрика: достижения и перспективы // Успехи химии. 2006. Т. 75, № 4. С. 302-321.
4. Pomerantsev A.L., Rodionova O.Ye. Multiclass partial least squares discriminant analysis: Taking the right way – A critical tutorial // J. Chemometrics. 2018. V. 32, № 8. Article 3030.
5. Lavine B., Workman J. Chemometrics // Anal. Chem. 2008. V. 80, № 12. P. 4519-4531.
6. Факторный, дискриминантный и кластерный анализ: пер. с англ./ Дж.-О. Ким [и др.]. М.: Финансы и статистика. 1989. 215 с.
7. Gan G., Ma C., Wu J. Data Clustering: Theory, Algorithms, and Applications. Philadelphia: SIAM. 2007. 184 p.
8. Дребуцка Т.Н. Введение в хемометрику: учеб. пособие. Новосибирск: Новосиб. гос. ун-т, 2013. 89 с.
9. Brereton R.G. Chemometrics. Data analysis for the laboratory and chemical plant. Chichester: Wiley, 2003. 489 p.
10. Шараф М.А., Иллман Д.Л., Ковальски Б.Р. Хемометрика: пер. с англ. Л.: Химия, 1989. 272 с.
11. Hughes I.G., Hase T.P.A. Measurements and their uncertainties. Oxford: Oxford Univ. Press Inc., 2010. 136 p.
12. [Электронный ресурс] <http://www.eurasiancommission.org/ru/act/trade/catr/ett/Pages/default.aspx> (дата обращения: 14.09.2023).
13. Sandercock P.M.L. Fire investigation and ignitable liquid residue analysis – a review: 2001–2007 // Forensic Sci. Int. 2008. V. 176, № 2-3. P. 93-110.
14. Chemometric analysis of gas chromatography with flame ionisation detection chromatograms: A novel method for classification of petroleum products / N.J. Nielsen [et al.]. // J. Chromatogr. A. 2012. V. 1238. P. 121-127.
15. Applications of chemometrics in oil spill studies / L.G. Oliveira [et al.] // Microchem. J. 2005. V. 166. P. 106216-106228.
16. Detection of discoloration in diesel fuel based on gas chromatographic fingerprints / B. Krakowska [et al.] // Anal. Bioanal. Chem. 2015. V. 407. P. 1159-1170.
17. Okop I.J., Bassej C.E., Okorie F.S. Chemometric determination of chemical and physical assortment of total hydrocarbons within crude oil spilled soil // Inter. Res. J. Nat. Sci. 2013. V. 1, № 2. P. 23-33.
18. Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline / D.L. Flumignan [et al.] // Chemom. Intell. Lab. Syst. 2008. V. 92, № 1. P. 53-60.
19. Improving oil classification quality from oil spill fingerprint beyond six sigma approach / H. Juahir [et al.] // Mar. Pollut. Bull. 2017. V. 120, № 1-2. P. 322-332.
20. Bayona J.M., Dominguez C., Albaiges J. Analytical developments for oil spill fingerprinting // Trends Environ. Anal. Chem. 2015. Т. 5, P. 26-34.
21. Роль хемометрических методов в криминалистической экспертизе нефтепродуктов и горюче-смазочных материалов / В.А. Руднев [и др.] // Теорія та практика судової експертизи і криміналістики, 2016. Т. 16. С. 249-258.
22. Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks / P. Doble [et al.] // Forensic Sci. Int. 2003. V. 132. P. 26-39.
23. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration / V.L. Skrobot [et al.] // Energ. Fuel. 2007. V. 21. P. 3394-3400.
24. Johnson K.J., Synovec R.E. Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis // Chemom. Intell. Lab. Syst. 2002. V. 60, P. 225-237.



25. Hendrikse J. ENFSI collaborative testing programme for ignitable liquid analysis: A review // *Forensic Sci. Int.* 2007. V. 167, № 2-3. P. 213-219.
26. Use of chemometric methods of data analysis for the identification and typification of petroleum and petroleum products/ T.A. Bolotnik [et al.] // *J. Anal. Chem.* 201., V. 74. P. 1336-1340.
27. Померанцев А. Хемометрика в Excel. Томск: Изд-во ТПУ, 2014. 435 с.
28. PLS-DA templates. [Электронный ресурс] <http://chemometrics.chph.ras.ru/PLS-DA/> (дата обращения: 14.09.2023).

## REFERENCES

1. Deming S.N., Michotte Y., Massart D.L., Kaufman L., Vandeginste B.G.M. *Chemometrics*. New York, Elsevier, 1988. 488 p.
2. Rodionova O.Ye. [Chemometric approach to the study of large arrays of chemical data]. *Rossiyskiy Khimicheskij Zhurnal [Russ. Chem. J.]*, 2006, vol. 50, no. 2, pp. 128-144 (in Russian).
3. Rodionova O.Ye., Pomerantsev A.L. Chemometrics: achievements and prospects. *Russ. Chem. Rev.*, 2006, vol. 75, no. 4, pp. 271-287. doi: 10.1070/RC2006v075n04ABEH003599 (in Russian).
4. Pomerantsev A.L., Rodionova O.Ye. Multiclass partial least squares discriminant analysis: Taking the right way – A critical tutorial // *J. Chemometrics*. 2018. V. 32, № 8. Article 3030.
5. Lavine B., Workman J. Chemometrics. *Anal. Chem.*, 2008, vol. 80, no. 12, pp. 4519-4531. doi: 10.1021/ac800728t.
6. Kim J.O., Mueller C.W., Klekka W.R., Aldenderfer M.S., Blashfield R.K. *Factorial, Discriminant and Cluster Analysis*. Translated to Russian. Moscow, Finance and Statistics, 1989. 215 p. (Russ. ed.: Kim Dzh.-O., M'juller Ch.U., Klekka U.R., Oldenderfer M.S., Bljeshfild R.K. *Fakturnyj, diskriminirnyj i klasternyj analiz*. Per. s angl. Moskva, Finansy i statistika, 1989. 215 p.).
7. Gan G., Ma C., Wu J. *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, SIAM, 2007. 184 p. doi:10.1137/1.9780898718348.
8. Drebushhak T.N. *Vvedenie v hemometriku: ucheb. Posobie [Introduction to chemometrics: a tutorial]*. Novosibirsk, Novosib. State Un-t, 2013. 89 p. (in Russian).
9. Brereton R.G. *Chemometrics. Data analysis for the laboratory and chemical plant*. Chichester, Wiley, 2003. 489 p.
10. Sharaf M.A., Illman D.L., Kowalski B.R. *Chemometrics*. New York, Wiley, 1986. 352 p. (Russ. ed.: Sharaf M.A., Illman D.L., Koval'ski B.R. *Hemometrika*: per. s angl. L.: Himija, 1989. 272 p.).
11. Hughes I.G., Hase T.P.A. *Measurements and their uncertainties*. Oxford, Oxford Univ. Press Inc., 2010. 136 p.
12. THE CUSTOMS CODE OF THE EAEU (2023). Available at: <http://www.eurasiancommission.org/ru/act/trade/catr/lett/Pages/default.aspx> (access: 14.09.2023).
13. Sandercock P.M.L. Fire investigation and ignitable liquid residue analysis – a review: 2001–2007. *Forensic Sci. Int.*, 2008, vol. 176, no. 2-3, pp. 93-110. doi: 10.1016/j.forsciint.2007.09.004.
14. Nielsen N.J., Ballabio D., Tomasi G., Todeschini R., Christensen J.H. Chemometric analysis of gas chromatography with flame ionisation detection chromatograms: A novel method for classification of petroleum products. *J. Chromatogr. A.*, 2012, vol. 1238, pp. 121-127. doi: 10.1016/j.chroma.2012.03.062.
15. Oliveira L.G., Araujo K.C., Barreto M.C., Bastos M.E.P., Lemos S.G., Fragoso W.D. Applications of chemometrics in oil spill studies. *Microchem. J.*, 2021, vol. 166, pp. 106216-106228. doi: 10.1016/j.microc.2021.106216.
16. Krakowska B., Stanimirova I., Orzel J., Daszykowski M., Grabowski I., Zaleszczyk G., Sznajder M. Detection of discoloration in diesel fuel based on gas chromatographic fingerprints. *Anal. Bioanal. Chem.*, 2015, vol. 407, pp. 1159-1170. doi: 10.1007/s00216-014-8332-4.
17. Okop I.J., Bassey C.E., Okorie F.S. Chemometric determination of chemical and physical assortment of total hydrocarbons within crude oil spilled soil. *Inter. Res. J. Nat. Sci.*, 2013, vol. 1, no. 2, pp. 23-33. doi:10.37745/irjns.13.
18. Flumignan D.L., de Oliveira Ferreira F., Tininis A.G., de Oliveira J.E. Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline. *Chemom. Intell. Lab. Syst.*, 2008, vol. 92, no. 1, pp. 53-60. doi:10.1016/j.chemolab.2007.12.003.
19. Juahir H., Ismail A., Mohamed S.B., Toriman M.E., Kassim A.M., Zain S.M., Ahmad W.K.W., Wah W.K., Zali M.A., Retnam A., Taib M.Z.M., Mokhtar M. Improving oil classification quality from oil spill fingerprint beyond six sigma approach. *Mar. Pollut. Bull.*, 2017, vol. 120, no. 1-2, pp. 322-332. doi:10.1016/j.marpolbul.2017.04.032.
20. Bayona J.M., Dominiguez C., Albaiges J. Analytical developments for oil spill fingerprinting. *Trends Environ. Anal. Chem.*, 2015, vol. 5, pp. 26-34. doi:10.1016/j.teac.2015.01.004.
21. Rudnev V. A., Klimchuk A. F., Dan'shina, K. M., Karnozhickij P. V. [The role of chemometric methods in the forensic examination of petroleum products and fuels and lubricants]. *Teoriya ta praktika sudovoi ekspertizi i kriminalistiki [Theory and Practice of Forensic Science and Criminalistics]*, 2016, vol. 16, pp. 249-258. (in Russian).
22. Doble P., Sandercock M., Du Pasquier E., Petocz P., Roux C., Dawson M. Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic Sci. Int.*, 2003, vol. 132, pp. 26-39. doi:10.1016/S0379-0738(03)00002-1.
23. Skrobot V.L., Castro R., Pereira C., Pasa V. M., Fortes I. C. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. *Energ. Fuel.*, 2007, vol. 21, pp. 3394-3400. doi:10.1021/ef0701337.
24. Johnson K.J., Synovec R.E. Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis. *Chemom. Intell. Lab. Syst.*, 2002, vol. 60, p. 225-237. doi:10.1016/S0169-7439(01)00198-8.
25. Hendrikse J. ENFSI collaborative testing programme for ignitable liquid analysis: A review. *Forensic Sci. Int.*, 2007, vol. 167, no. 2-3, pp. 213-219. doi:10.1016/j.forsciint.2006.06.058.
26. Bolotnik T.A., Timchenko Y.V., Plyushchenko I.V., Levkina V.V., Pirogov A.V., Smolenkov A.D., Popik M.V., Shpigun O.A. Use of chemometric methods of data analysis for the identification and typification of petroleum and petroleum products. *J. Anal. Chem.*, 2019, vol. 74, pp. 1336-1340. doi: 10.1134/S1061934819130045.
27. Pomerantsev A.L. *Khemometrika v Excel [Chemometrics in Excel]*. Tomsk, Publishing house TPU, 2014. 435 p. (in Russian).
28. PLS-DA templates. Available at: <http://chemometrics.chph.ras.ru/PLS-DA/> (access: 14.09.2023).