

ФИЗИКО-МАТЕМАТИЧЕСКИЕ АСПЕКТЫ ЛИКВИДАЦИИ ПОСЛЕДСТВИЙ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ИСПОЛЬЗОВАНИЯ ОПТИЧЕСКОЙ СПЕКТРОСКОПИИ С МНОГОМЕРНЫМИ ИЗМЕРЕНИЯМИ ДЛЯ ФОРМИРОВАНИЯ АНАЛИТИЧЕСКИХ ОБРАЗОВ ЖИДКИХ МИНЕРАЛЬНЫХ ТЕХНИЧЕСКИХ МАСЕЛ

Н.И. Егорова, кандидат физико-математических наук.

Санкт-Петербургский университет ГПС МЧС России.

И.О. Конюшенко, кандидат физико-математических наук;

В.М. Немец, доктор технических наук, профессор.

Санкт-Петербургский государственный университет

Показана возможность применения статистического метода распознавания образов в варианте метода главных компонент для решения задач идентификации нефтепродуктов.

Ключевые слова: метод распознавания образов, оптическая абсорбционная спектроскопия, нефтепродукты

INVESTIGATION OF POSSIBILITIES OF USING OPTICAL SPECTROSCOPY WITH MULTIDIMENSIONAL MEASUREMENTS FOR CREATION OF ANALITICAL IMAGES OF LIQUID MINERAL TECHNICAL OILS

N.I. Egorova. Saint-Petersburg university of the State fire service of EMERCOM of Russia.

I.O. Konyushenko; V.M. Nemetz. Saint-Petersburg state university

The possibility of using the statistical method of pattern recognition in a variant of the method of principal components to solve the problem of petroleum products identification is shown.

Keywords: pattern recognition, absorption spectroscopy, identification of oil-products

В настоящее время довольно широко развиваются методы распознавания образов (МРО) применительно к тестированию (классификация и идентификация жидких объектов сложного молекулярного состава [1, 2]). Одним из направлений такого подхода являются разработки с использованием многомерных измерений в оптической абсорбционной спектроскопии.

На этапе разработки МРО осуществляются исследования эффективности информационных признаков исследуемых объектов.

Результаты подобного рода исследований возможностей абсорбционной спектроскопии применительно к таким нефтепродуктам, как технические масла, являются предметом обсуждения в предлагаемой работе.

1. Сущность метода и основные представления.

Основой группы МРО во всех приложениях является понятие образа объекта исследования.

Понятие «образ объекта» имеет достаточно широкий смысл, что связано с широким спектром направлений исследований, в которых применяются МРО. В общем случае понятие «образ» характеризует совокупность информационных признаков, настолько полную, что имеется возможность надежно (с установленной вероятностью ошибки) отличить между собой однородные объекты, составляющие класс. В аналитике под словом «класс» подразумевается группа объектов, как правило, однородных по составу (по крайней мере, по основным компонентам) или иным признакам, включающая в себя все объекты, идентичность контролируемого вещества одному из которых может быть установлена. Подчеркнем, что МРО применяются для идентификации исследуемого объекта, то есть установления его идентичности одному из известных веществ, составляющих класс по тому или иному параметру.

При решении задачи идентификации того или иного объекта, то есть установления его идентичности одному из объектов данного класса, построенного по некоторым их характерным признакам, в качестве последних могут фигурировать различные параметры. Например, класс могут образовать технические масла одного производителя, но разных марок, или технические масла одной марки, но разных производителей, конденсаты, и др.

В случаях, когда в качестве образа используется спектроскопическая информация, его информационными признаками могут быть значения интенсивности излучения на большом числе определенных длин волн, а также значения коэффициентов, характеризующих оптические свойства веществ (например, коэффициент поглощения и др.). Следовательно, оптическим образом в данном случае может служить тот или иной спектр оптических характеристик объекта, зарегистрированный в достаточно широком диапазоне длин волн.

Безусловно, далеко не всякая оптико-спектральная характеристика объекта может являться его надежным образом. Качество образа определяется, с одной стороны, его пригодностью к идентификации объекта внутри класса объектов, принятых за стандартные, а с другой стороны, минимальностью информационных признаков, составляющих образ. Так, например, при использовании в качестве образа нефтепродукта его спектральных характеристик требование минимальности числа признаков и надежности идентификации приводят к необходимости выбора для формирования образа таких участков спектральной характеристики, различия в которых при переходе от образа к образу внутри класса объектов (нефтепродуктов) максимальны.

Конкретный вид образа зависит не только от состава объекта, но и от условий формирования образа. Поэтому оптико-спектральный образ, то есть сложный спектр как единое целое представляет собой образ объекта, взятого в конкретных условиях, сформированный в соответствии с конкретным методом, на конкретном приборном обеспечении. В процессе регистрации спектра должны контролироваться внешние и внутренние параметры объекта (температура, давление, влажность и др.). Кроме того, необходимым требованием к образу с аналитической точки зрения должна служить воспроизводимость образа в фиксированных условиях, а также его независимость (или известная зависимость) от времени.

Таким образом, первым этапом аналитического процесса для идентификации исследуемого объекта с использованием распознавания образов следует считать формирование оптико-спектрального образа одним из аналитических спектроскопических методов (эмиссия, поглощение, рефракция, флуоресценция и др.).

Говоря о методах математической обработки оптико-спектральных образов, следует заметить, что при решении поставленной задачи формируемые образы объектов – вещества, составляющих класс, считаются реализациями многомерных случайных величин. В рамках используемой модели вид распределения для каждого из объектов класса считается нормальным. Параметры нормального распределения (математическое ожидание и ковариационная матрица) могут быть оценены путем многократного формирования образа данного вещества, то есть из нескольких реализаций одной случайной величины. На рис. 1 приведена иллюстрация для случая двух веществ и двумерного статистического пространства, то есть образ вещества представляет собой двумерный вектор вида (x_1, x_2) . По вертикальной оси отложена плотность распределения случайной величины. Сформированный образ объекта в рамках метода является реализацией одной из случайных величин класса.

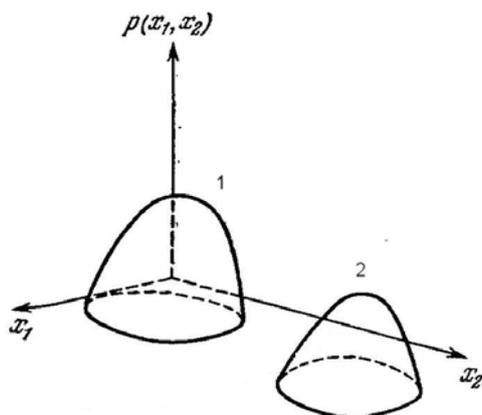


Рис. 1. Случай двух веществ и двумерного пространства

В целом задача процедуры математической обработки состоит в определении, к какой области из случайных величин данная реализация относится с наибольшей вероятностью. Указанная процедура строится на понятии условной вероятности и теореме Байеса [3].

Необходимым элементом МРО является оптимизация размерности образа. С одной стороны, высокая размерность образа повышает его информативность и селективность, а, следовательно, вероятность его надежной идентификации. С другой стороны, чрезвычайно высокая размерность образа затрудняет построение систем математической обработки информации, увеличивает время работы обрабатывающих систем. Применительно к оптико-спектральному подходу, образы объектов могут иметь чрезвычайно высокую размерность, так как можно взять сколько угодно большое число точек на шкале длин волн (это число ограничивается разрешающей способностью и спектральным диапазоном). В этом случае для формирования образа априорно берется избыточное число точек, а затем решается задача сокращения размерности образа до оптимальной. Оптимальной считается размерность, сведенная к минимуму, при условии, что при этом не произошло существенной потери информативности образа.

В данной работе для сокращения размерности оптико-спектральных образов применен известный метод главных компонент (МГК) [4], представляющий собой способ выявления наибольших различий между объектами внутри класса. Смысл МГК может быть пояснен как поворот первоначального многомерного статистического пространства определенным образом. Первый орт нового базиса определяет направление, вдоль которого различия между образами максимальны из всех возможных. Второй орт определяет направление, вдоль которого различия максимальны из оставшихся, после определения первого орта и т.д. Иными словами, первый орт новой системы наиболее информативен, второй – наиболее информативен после первого и т.д. Выбирая для дальнейшего

рассмотрения определенное число ортов с последовательными номерами, можно сформировать пространство любой размерности, более низкой, чем первоначальная, и сохранить при этом наибольшее количество информации. Реально выбирается такое число ортов, при котором обеспечивается достаточно высокая вероятность взаимного соответствия образов различных образцов, принадлежащих одному и тому же подклассу. И если такое достигнуто, то созданная аналитическая система и обеспечивающий ее работу аппарат многофакторной математической статистики готовы для выполнения процесса идентификации объектов рассматриваемого класса.

В целом разработка аналитического подхода с использованием МРО с целью идентификации веществ состоит в следующем.

1) С учетом особенностей решаемой конкретной задачи выбор способа формирования образов и варианта метода их распознавания.

2) Создание автоматизированной аналитической системы, способной формировать, хранить и сопоставлять образы различных веществ – объектов, принадлежащих к одному определенному классу, каждое из которых являет собой тот или иной подкласс.

3) Для выбранного класса объектов на основе установленной совокупности информационных признаков и их статистического разброса формируются образы каждого из объектов класса. Каждый из таких объектов условно определяется как стандартный, и его образ хранится в памяти аналитической системы.

4) В тех же условиях, на той же самой аналитической системе и по тому же методу формируется образ контролируемого объекта (неизвестного, но принадлежащего к данному классу в реальной аналитической задаче идентификации).

5) Образ контролируемого объекта сопоставляется с образами известных объектов – членов класса, хранящимися в памяти системы.

6) Формируется заключение в виде ответов на вопросы:

– какому объекту класса соответствует тестируемый;

– какова вероятная ошибка установленного соответствия.

Ниже иллюстрируется практическая применимость такого подхода для целей идентификации нефтепродуктов на примере такого класса веществ, как моторные масла.

2. Создание методики идентификации моторных масел МРО на основе абсорбционных оптико-спектральных измерений.

Применение изложенного выше подхода для идентификации образцов моторных масел состоит, прежде всего, в исследовании оптико-спектральных особенностей отобранных условно стандартных объектов этого класса (класс – моторные масла). Затем следует разработка методики формирования абсорбционно-спектральных образов объектов этого класса, отработка процедуры оптимизации размерности образов, построение математического аппарата идентификации, создание действующего макета идентификационной аналитической системы и ее испытание.

Реализация такого алгоритма включает:

1) Подборку максимально большого числа различных образцов моторных масел, являющихся условно стандартными объектами.

2) Абсорбционно-спектральные исследования отобранных образцов моторных масел с целью:

– определения областей спектра, обеспечивающих наиболее эффективные измерения;

– разработки методики формирования образов масел;

– оценки воспроизводимости формируемых образов.

3) Разработка на основе методов многомерной математической статистики (метод главных компонент – МГК) методики сокращения размерности образов с целью ее оптимизации, обеспечивающей допустимое время работы систем, обрабатывающих результаты измерений.

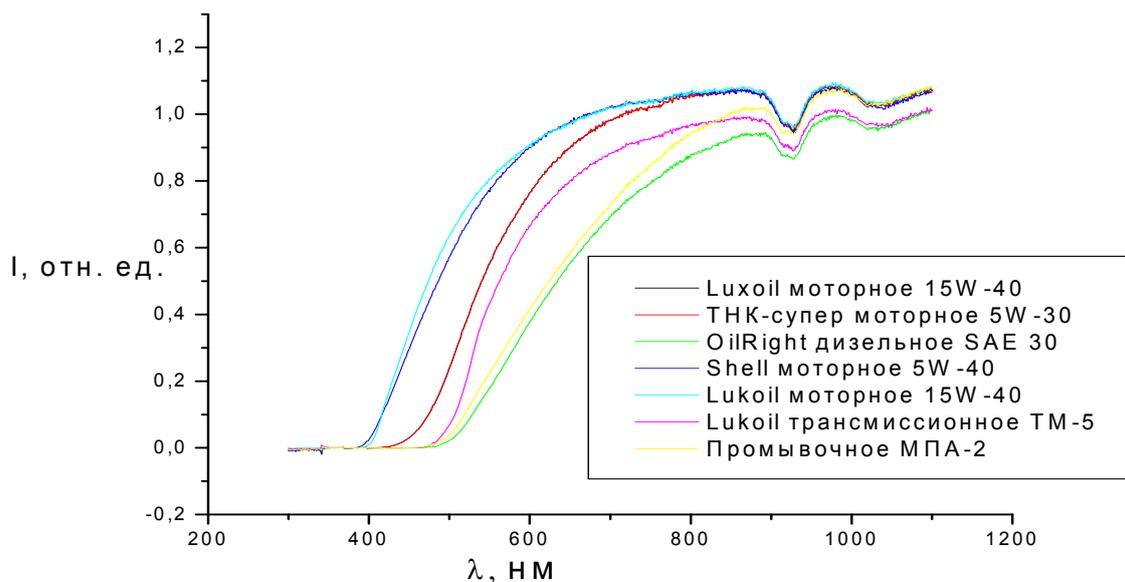
4) Разработка математического аппарата идентификации образов с числом членов класса, равным числу отобранных «стандартных» образцов масел.

5) Создание аналитической автоматизированной идентификационной системы и ее проверка на условно стандартных объектах.

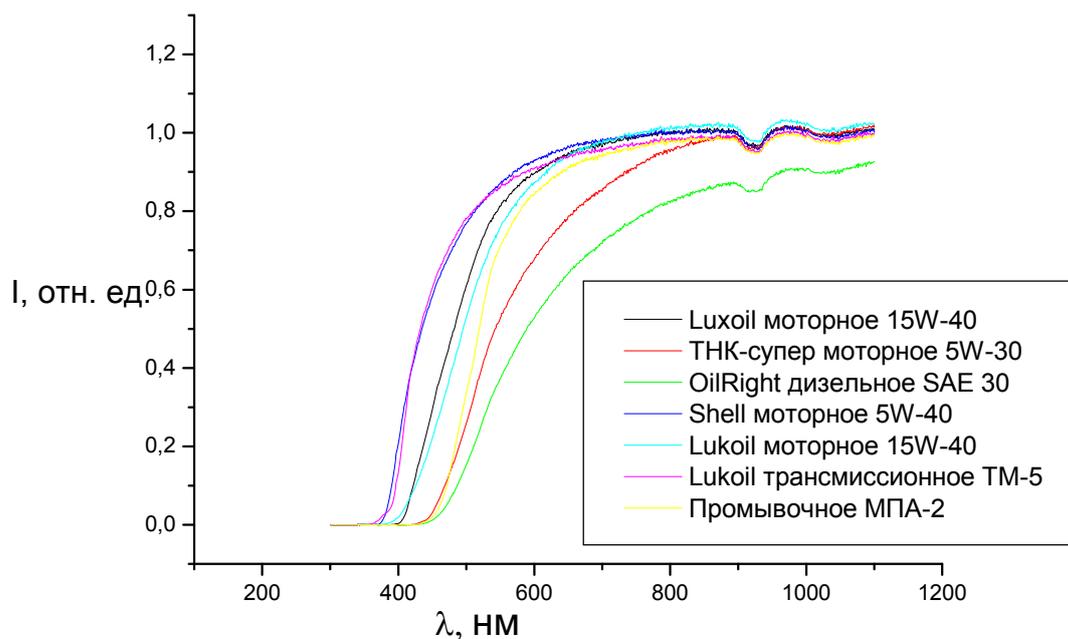
б) Применение методики для идентификации моторных масел.

В предлагаемой работе исследованы возможности реализации изложенного выше подхода в идентификации моторных масел с использованием оптической абсорбционной спектроскопии. Спектры получали на серийном УВИ-спектрофотометре СФ-56 производства ЗАО «ОКБ СПЕКТР» (Санкт-Петербург). Спектры масел регистрировали с шагом 1 нм. Спектральная ширина щели спектрофотометра также 1 нм, динамический диапазон 100.

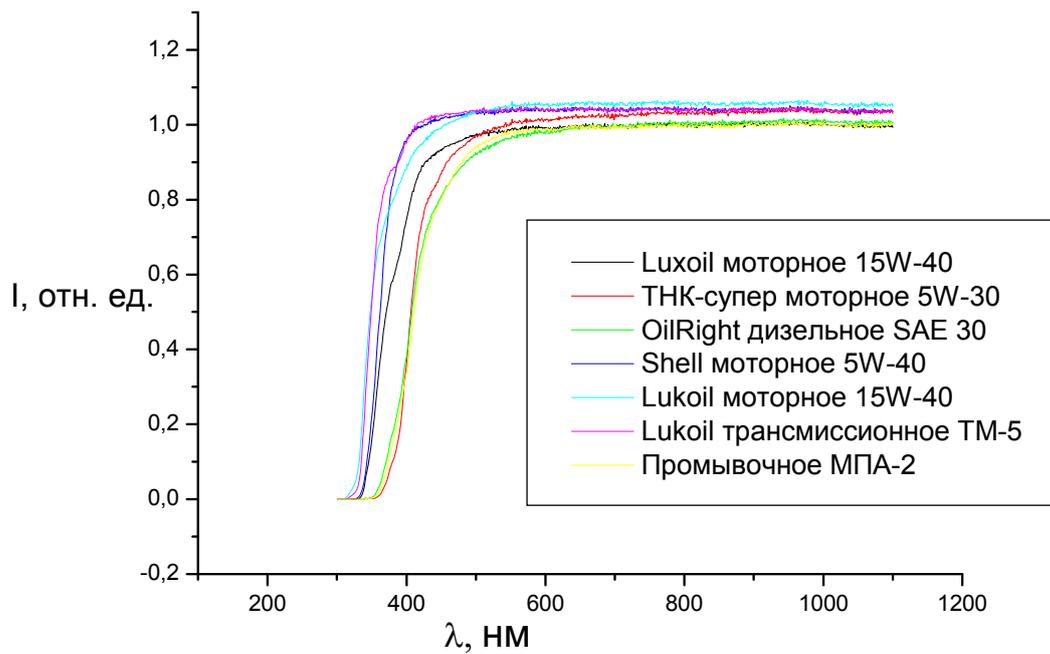
На рис. 2 а), б), в), г) приведены полученные спектры пропускания моторных масел в диапазоне длин волн от 300 до 1100 нм при использовании кювет различной толщины. Для сравнения был получен спектр подсолнечного масла (рис. 3).



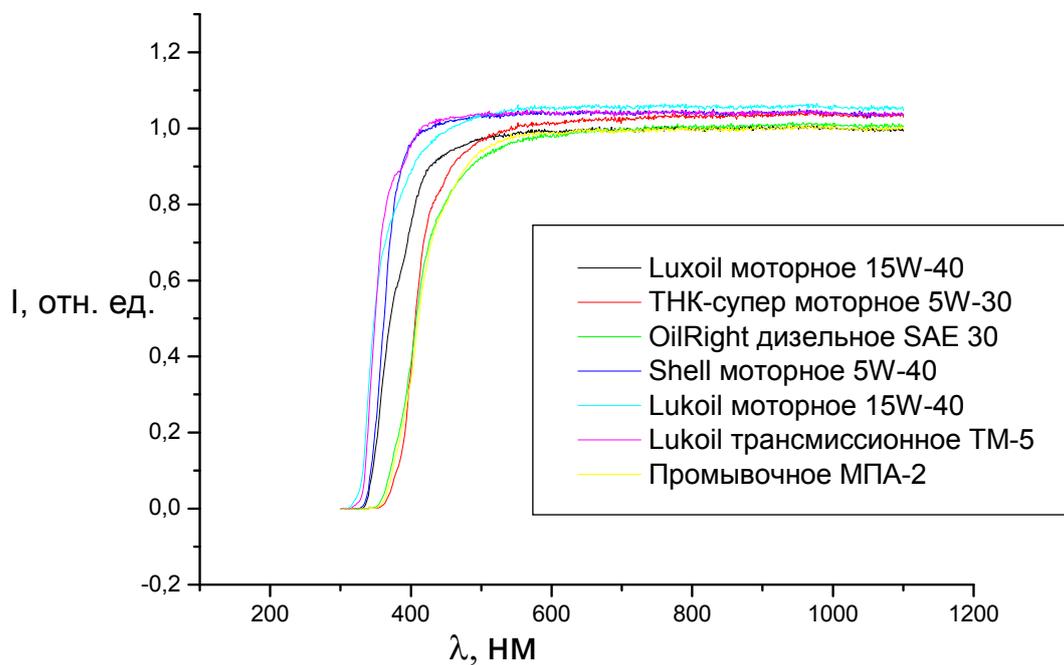
а)



б)



в)



г)

Рис. 2. Спектр коэффициента пропускания в кювете:
 а) 10 мм; б) 4 мм; в) 2 мм; г) 0,2 мм.
 I – нормированная интенсивность

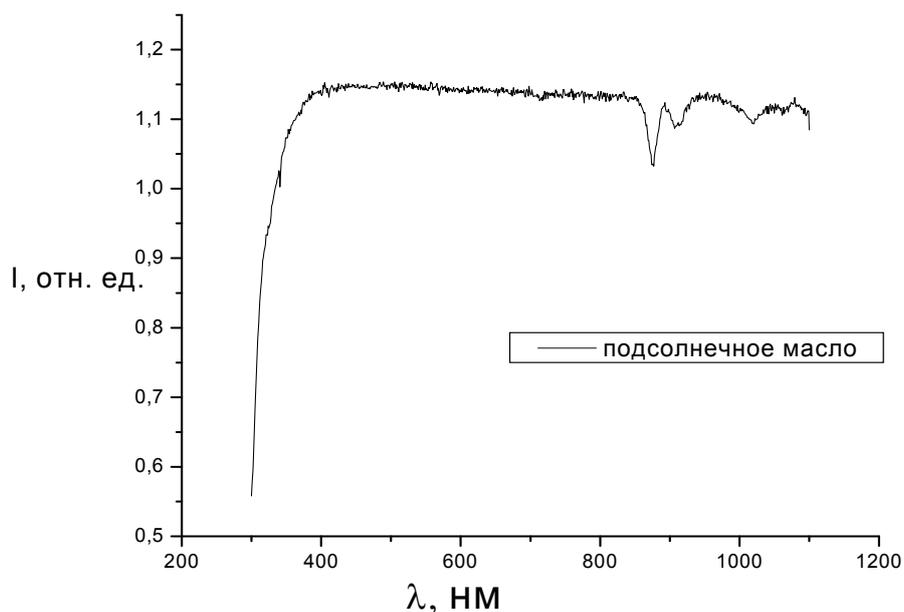


Рис. 3. Спектр пропускания подсолнечного масла в кювете 10 мм.
I – интенсивность

Видно, что в ближней ультрафиолетовой (УФ) области наблюдается сильное поглощение. При увеличении длины волны отмечается монотонный рост прозрачности. Незначительные полосы поглощения наблюдаются в области от 850 до 1100 нм. С целью обеспечения максимальной информативности результатов исследования спектров, они были сняты с использованием кювет различной толщины. Этот эксперимент показал, что поведение кривой пропускания у разных масел различный. Это, возможно, является отражением того, что они произведены по различным технологиям. Если сопоставить спектры пропускания моторных масел и спектр пропускания подсолнечного масла, можно отметить, что край полосы поглощения подсолнечного масла существенно смещен в УФ-область, а полосы поглощения, расположенные в инфракрасной области, находятся на разных длинах волн. Из этого можно сделать вывод о том, что спектральные полосы поглощения, расположенные в области 850–950 нм, характерны для нефтепродуктов.

Для оценки случайного разброса результатов эксперимента и последующей статистической обработки спектральных характеристик были измерены спектральные распределения коэффициента пропускания 12 образцов моторных масел. Для каждого образца процедура измерения была повторена 10 раз, что в результате позволило получить 120 различных спектральных образов. Их общая картина представлена на рис. 4.

В представленном массиве из 120 распределений существуют два типа различий. Один из них обусловлен различиями в составе масел, и именно этот тип различия необходимо выделить при статистическом исследовании. Другой тип различий обусловлен экспериментальным разбросом результатов измерений. Наличие этого типа различия затрудняет статистический анализ распределения результатов в целом. Поэтому традиционное применение МГК ко всему массиву из 120 спектральных распределений приводит к смешению двух типов различий, так, что отделить одно от другого невозможно.

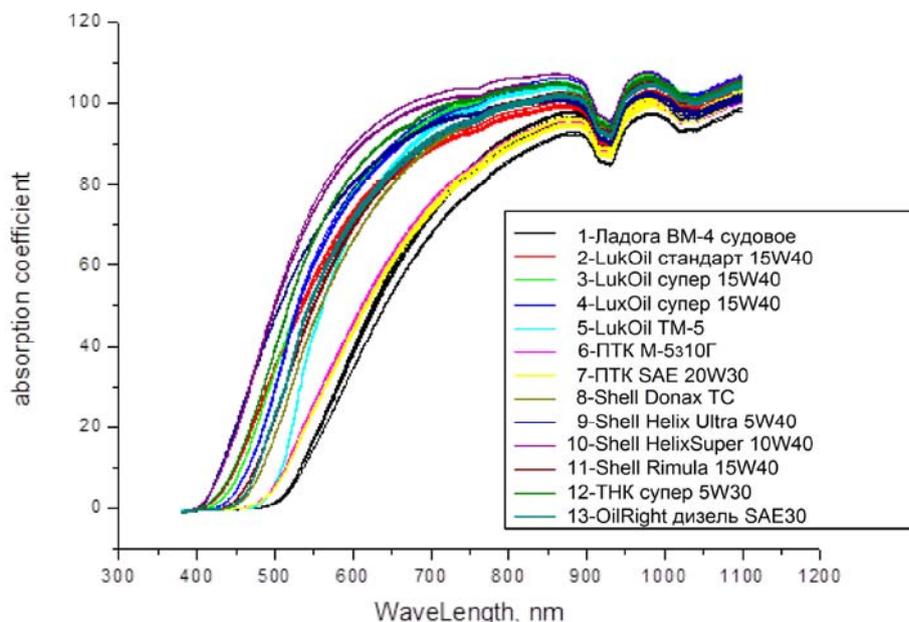


Рис. 4. Распределение коэффициента пропускания жидких минеральных технических масел

Примененная в данном случае модификация МГК предполагает первоначальное усреднение спектральных распределений, относящихся к одному и тому же образцу. Система главных компонент выстраивается на усредненных по каждому образцу данных (то есть на 13 усредненных образах), после чего проводится их отображение в результирующее пространство.

Для иллюстрации работы описанной процедуры на рис. 5 приведен результат отображения в трехмерное пространство образов трех типов масел с учетом экспериментального разброса. Приведенная картина обладает наглядностью и позволяет сделать вывод о возможности разделения различных типов масел.

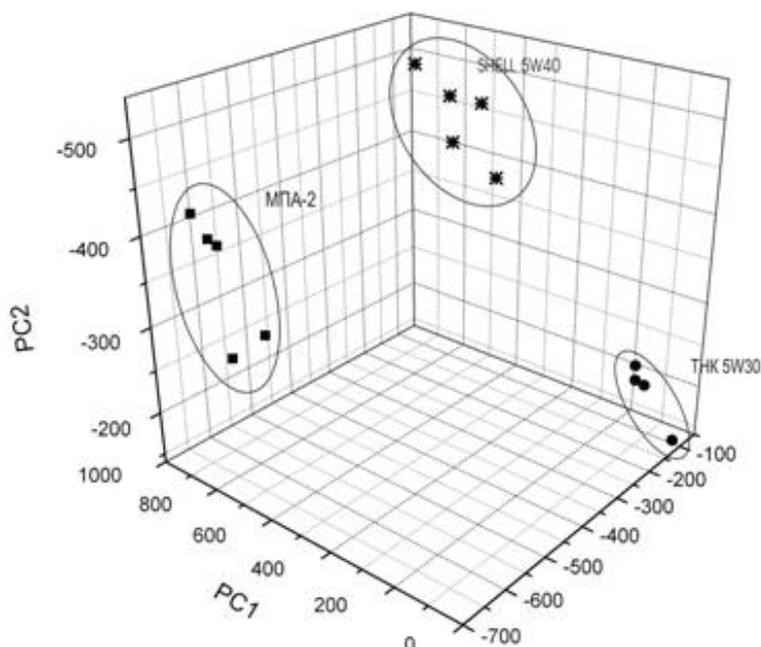


Рис. 5. Взаимное расположение образов трех типов масел в пространстве первых трех ГК

Результаты процедуры отображения образов всех 13 типов масел с учетом экспериментального разброса на плоскость первых двух главных компонент представлены на рис. 6.

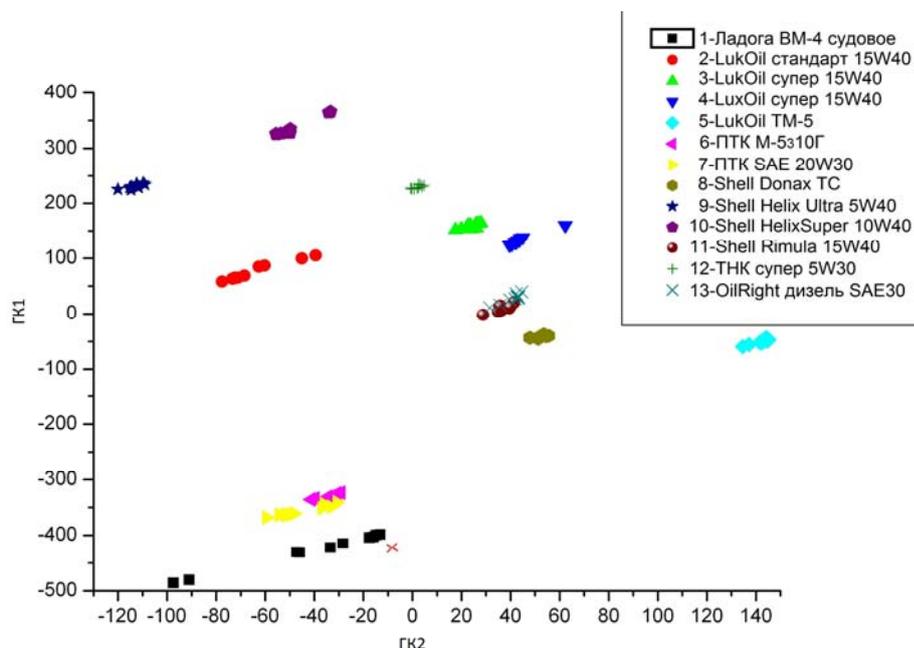


Рис. 6. Расположение образов известных масел и образа «неизвестного» масла на плоскости первых двух ГК

Для тестирования описанного метода идентификации неизвестного образа были использованы образцы тех же масел, что и используемые для построения классификатора. Эти образцы были получены независимо и не входили в совокупность образцов, использовавшихся для построения системы МГК. На рис. 6 диагональным крестом отмечен образ условно неизвестного масла. В таблице приведены значения условной вероятности при соотнесении образа условно неизвестного образца с образцами известных «стандартных» образцов. Максимум значения условной вероятности наблюдается для образца № 1, что соответствует действительности. Ошибка идентификации в данном, достаточно простом случае составляет менее 1 %.

Таблица. Оценки условной вероятности, полученные при соотнесении образа условно неизвестного образца с образцами известных образцов

Номер образца	Значение условной вероятности, $\times 10^{-7}$
1	1063,5
2	5,1
3	19,7
4	9,1
5	53,1
6	103,7
7	56,7
8	42,0
9	14,9
10	7,0
11	27,6
12	31,5

Полученные в целом результаты дают основание судить о возможности применения статистического метода распознавания образов в варианте МГК для решения задач идентификации нефтепродуктов.

Описанный подход может быть применен для идентификации и других нефтепродуктов. В полной мере это относится и к такому нефтепродукту, как газовые конденсаты, для которых необходим быстрый, надежный и автоматизированный входной контроль соответствия техническим условиям.

Литература

1. Доленко С.А., Гердова И.В. Лазерная флуориметрия смесей сложных соединений с использованием искусственных нейронных сетей // Квантовая электроника. 2001. Т. 31. № 9. С. 834–838.

2. Simon M. Scott, David James, Zulfegur Ali, William T. O'Hare, Fred. J. Rowel. Total luminescence spectroscopy with pattern recognition for classification of edible oils // Analyst. 2003. V. 128. P. 966–973.

3. Смирнов Н.В., Дудин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969. 512 с.

4. Дронов С.В. Многомерный статистический анализ. Барнаул: Изд-во Алт. гос. ун-та, 2003. 246 с.