

**Федеральное государственное автономное образовательное учреждение
высшего образования «Московский физико-технический институт
(национальный исследовательский университет)»**

На правах рукописи

Саакян Арам Ваганович

**РАЗВИТИЕ МЕТОДОВ АБСОРБЦИОННОЙ, ФЛУОРЕСЦЕНТНОЙ И
РАМАНОВСКОЙ СПЕКТРОСКОПИИ С ХЕМОМЕТРИЧЕСКОЙ
ОБРАБОТКОЙ ДАННЫХ ДЛЯ АНАЛИЗА ЖИДКИХ СРЕД И АЭРОЗОЛЕЙ**

2.2.4 - Приборы и методы измерения (по видам измерений)

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель
д. т. н., А.Д. Левин

Москва 2025 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
ГЛАВА 1. АНАЛИТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ	11
1.1 Методы флуоресцентной спектроскопии	11
1.2 Методы спектроскопии КР (рамановской).....	17
1.3 Методы хеометрики и МО, предварительная обработка данных. Применение методов хеометрики и МО для анализа жидких многокомпонентных сред (в частности, спектров флуоресценции и КР)	22
Выводы к главе 1	45
ГЛАВА 2.ВЫБОР И ОБОСНОВАНИЕ АЛГОРИТМОВ ДЛЯ ОБРАБОТКИ СПЕКТРАЛЬНЫХ ДАННЫХ. РАЗРАБОТКА ПРОГРАММНОГО ПАКЕТА.....	47
2.1 Формирование данных.....	48
2.2 Предварительная обработка спектров.....	49
2.3 Выбор алгоритмов.....	51
2.4 Обучение и валидация (тестирование) модели	52
Выводы к главе 2	56
ГЛАВА 3. ИССЛЕДОВАНИЕ ПРОИСХОЖДЕНИЯ АЭРОЗОЛЬНЫХ ЧАСТИЦ ЧЕРНОГО УГЛЕРОДА.....	57
3.1 Классификация частиц черного углерода по спектрам КР.....	57
3.2 Классификация частиц черного углерода по спектрам поглощения.....	62
Выводы к главе 3	70
ГЛАВА 4. КЛАССИФИКАЦИЯ АЛКОГОЛЬНОЙ ПРОДУКЦИИ ПО ГЕОГРАФИЧЕСКОМУ ПРОИСХОЖДЕНИЮ И ВОЗРАСТУ ВЫДЕРЖКИ	72

4.1 Характеризация коньяков и виноградных бренди по спектрам флуоресценции.....	73
4.2 Классификация бренди и коньячной продукции по географическому происхождению и возрасту выдержки с использованием КР-спектроскопии.....	81
Выводы к главе 4.....	87
ГЛАВА 5. ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ КОЛИЧЕСТВЕННОГО АНАЛИЗА МНОГОКОМПОНЕНТНЫХ СМЕСЕЙ ПО СПЕКТРАМ КР	89
5.1 Градуировка фотометрической шкалы в приборно-независимых единицах с помощью образца сравнения в виде пленки полистирола с добавлением серы и количественный анализ многокомпонентных смесей по спектрам КР	89
5.2 Градуировка фотометрической шкалы в приборно-независимых единицах с помощью образца сравнения в виде раствора циклогексана	100
ЗАКЛЮЧЕНИЕ	108
СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ.....	109
СПИСОК ЛИТЕРАТУРЫ.....	110
Приложение 1	123

ВВЕДЕНИЕ

Актуальность темы работы

Оптическая спектроскопия является одним из основных подходов инструментальной химии, применяемым при качественном и количественном анализе разнообразных объектов. Различные методы оптической спектроскопии широко используется при контроле продукции химической и фармацевтической промышленности, объектов окружающей среды, пищевых продуктов и продовольственного сырья, в клинической лабораторной диагностике, судебной медицине и криминалистике.

В последнее время возрастают требования к методам оптической спектроскопии, связанные с расширением их аналитических возможностей, т.е. способностью анализировать все более сложные объекты, повышением чувствительности и специфичности анализа.

Первая группа требований часто диктует усложнение аналитической аппаратуры и пробоподготовки, увеличение длительности анализа, вторая — их максимальное упрощение и сокращение времени анализа.

Вторая группа требований связана с быстротой и доступностью анализа, включая максимальное упрощение пробоподготовки и возможность проведения анализов во внелабораторных условиях, в том числе непосредственно на месте взятия пробы.

Для того, чтобы удовлетворить обе группы требований, нужно не увеличивать сложность аппаратуры, пробоподготовки и длительность анализа, а правильно выбирать наиболее информативные фрагменты оптических спектров и применять для их обработки методы хемометрики и машинного обучения (МО), позволяющие извлечь из этих фрагментов максимум достоверной информации. К настоящему времени такой подход развит лишь применительно к небольшому числу конкретных аналитических задач.

Таким образом, развитие методов оптической спектроскопии с обработкой данных на основе хемометрики и МО применительно к анализу объектов, представляющих научный и/или практический интерес, является актуальным и

востребованным. В диссертации такой подход реализован для абсорбционной, флуоресцентной и рамановской (комбинационного рассеивания, КР) спектроскопии на примере трех объектов – аэрозольных частиц в атмосферном воздухе, алкогольных напитков и многокомпонентных жидких смесей.

Цель работы

Целью данной работы является разработка и апробация методов оптического спектрального анализа жидких сред и аэрозолей с применением хемометрики и МО для решения задач качественного и количественного анализа на примере конкретных объектов.

Основные задачи работы

1. Разработка методов классификации аэрозольных частиц черного углерода по источникам их выбросов с помощью спектров поглощения и КР с обработкой данных на основе алгоритмов машинного обучения.
2. Разработка метода градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах с помощью образцов сравнения и использование этого метода для количественного анализа многокомпонентных смесей.
3. Разработка методов классификации образцов алкогольной продукции по географическому происхождению и возрасту выдержки по спектрам флуоресценции и КР с последующей обработкой данных методами МО и хемометрики.
4. Проведение экспериментальных исследований предложенных методов, оценка достоверности и точности получаемых результатов.
5. Выбор и обоснование алгоритмов для обработки оптических спектров методами хемометрики и машинного обучения, разработка ПО, реализующего эти алгоритмы.

Научная новизна работы:

1. Впервые построена модель МО, позволяющая классифицировать аэрозольные частицы черного углерода по источникам выбросов, используя спектры абсорбции.

2. Впервые предложен и реализован метод градуировки КР-спектрометров по фотометрической шкале в приборно-независимых единицах путем использования образцов сравнения из вещества, имеющего характерные линии КР.

3. Определены информативные фрагменты спектров флуоресценции и КР и построены модели МО для анализа образцов бренди и коньячной продукции, с дальнейшей обработки методами хемометрики и МО.

Практическая ценность и использование результатов работы

Модели классификации, созданные на основе алгоритма экстремального градиентного бустинга (ЭГБ), позволяют с высокой точностью определять происхождение частиц черного углерода в атмосфере, что имеет важное значение для экологического мониторинга и контроля загрязнения воздуха. Применение модели для классификации частиц черного углерода по источникам их происхождения на основе КР-спектров подтверждается актом внедрения от ФГБУ «ИГКЭ». При этом используется разработанное ПО «ИнтелСпектр», которое представляет значительную практическую ценность для специалистов аналитических лабораторий, предоставляя им удобный инструментарий, не требующий глубоких знаний в программировании или хемометрике. Благодаря реализации четырех основных стадий обработки данных (формирование набора спектров, их предварительная обработка, обучение модели и её валидация), программа стала универсальным средством для решения широкого спектра аналитических задач. Указанное ПО было использовано при обработке спектров оптического поглощения микрочастиц черного углерода в рамках НИОКР по теме «Создание анализаторов для измерения концентрации микрочастиц черного углерода в атмосферном воздухе». Применение данного ПО подтверждается актом внедрения от ФГБУ «ВНИИОФИ».

Созданные методики спектрального анализа алкогольной продукции обеспечивают быстрый и надежный скрининг образцов при проверке их подлинности, географического происхождения и возраста выдержки, что представляет практическую ценность, как для производителей, так и для

контролирующих органов. Достижение высокой точности классификации образцов на тестовом наборе с высокими значениями доверительной вероятности подтверждает надежность разработанных подходов.

Предложенный метод градуировки фотометрической шкалы КР-спектрометров, позволяющий определять интенсивности линий КР в приборно-независимых единицах, открывает новые возможности для стандартизации и метрологического обеспечения рамановских измерений. Это позволяет использовать градуировки, построенные на одном КР-спектрометре, для анализа спектров, измеренных на других КР-спектрометрах, что существенно расширяет возможности количественного анализа многокомпонентных смесей.

Разработанные методы и модели могут быть интегрированы в ПО различных аналитических приборов, включая аэталометры и спектрометры, для автоматической классификации и количественного анализа в режиме реального времени. Это повышает оперативность и достоверность получаемых результатов, делая их более доступными для широкого круга пользователей.

Вклад автора

Личный вклад автора состоит в выполнении основного объема теоретических и экспериментальных исследований, изложенных в диссертационной работе, разработке подходов и их апробации, анализе и формировании данных для их обработки методами машинного обучения и хемометрики, оформлении результатов в виде публикаций и научных докладов

Апробация работы

Основные материалы, представленные в диссертации, были доложены и обсуждены на следующих научно-технических конференциях:

1. Международная научная конференция «Стандартные образцы в измерениях и технологиях» (Россия, г. Екатеринбург, 13-16 сентября 2022 г.)
2. VI Международная научно-техническая конференция «Метрология физико-химических измерений» (Россия, Менделеево, 05–08 сентября 2023 г.)
3. Всероссийская научно-техническая конференция «Метрологическое обеспечение фотоники-2023» (Россия, г. Москва, 15-16 ноября 2023 г.)

4. VI международная научная конференция «Стандартные образцы в измерениях и технологиях» (Россия, г. Екатеринбург, 3-6 сентября 2024 г.)

5. Всероссийская научно-техническая конференция «Метрологическое обеспечение фотоники-2025» (Россия, г. Москва, 01-04 апреля 2025 г.)

Публикации

По теме диссертационной работы опубликовано 8 научных работ, в том числе 4 статьи в российских рецензируемых журналах, входящих в перечень ВАК [1-4]; 4 тезисов докладов на научно-технических конференциях, получено свидетельство о государственной регистрации программ для ЭВМ [5].

Структура и объем работы

Полный объем диссертации составляет 124 страниц, в том числе 18 рисунков, 26 таблиц и 1 приложение. Список литературы содержит 100 источников.

Работа включает введение, пять глав, заключение, список основных сокращений и обозначений, список используемой литературы.

Основные положения, выносимые на защиту

1. Построенная в диссертации модель МО на основе алгоритма ЭГБ позволяет осуществлять классификацию аэрозольных частиц черного углерода по источникам выбросов, используя наборы значений оптической плотности на 7 длинах волн в диапазоне от 370 до 950 нм, измеренных для частиц, собранных путем прокачки воздушной пробы через стекловолоконную мембрану. Доля достоверных предсказаний при классификации составляет около 89 % для тестовых наборов, не используемых при построении модели МО [1].

2. Использование образцов сравнения в виде циклогексана или пленки полистирола с добавлением серы позволяет провести градуировку КР-спектрометров по фотометрической шкале в приборно-независимых единицах, и дает возможность определять концентрации веществ по спектрам КР, используя градуировочный график, построенный на другом приборе или при других условиях измерения [2].

3. Модель МО на основе алгоритма ЭГБ позволяет осуществлять классификацию бренди и коньячной продукции по географическому происхождению и возрасту выдержки с доверительной вероятностью не менее 86 %, используя следующие информативные фрагменты оптических спектров [3-4]:

— эмиссия флуоресценции в диапазоне от 260 до 420 нм при возбуждении на длине волны 250 нм;

— эмиссия флуоресценции в диапазоне от 290 до 450 нм при возбуждении на длине волны 280 нм;

— синхронное сканирование в диапазоне длин волн от 235 до 450 нм при разности длин волн 50 нм;

— рамановское рассеяние в диапазоне от 1400 до 1500 см⁻¹ при возбуждении на длине волны 532 нм.

ГЛАВА 1. АНАЛИТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ

Исходя из поставленных во Введении задач, в настоящей главе проанализировано современное состояние задач и исследований для анализа разных веществ методами оптической спектроскопии с последующей обработкой данных методами МО и хеометрики. В подразделе 1.1 и 1.2 приведены теоретические основы флуоресцентной и КР-спектроскопии, обзор современного оборудования, различные физико-химические параметры, определяющие интенсивность и спектральные характеристики, а также проведен обзор актуальных современных методов измерения. В разделе 1.3 приведены определение хеометрики, основные типы решаемых задач, виды предварительной обработки данных, алгоритмов МО и хеометрики, особенности обучения и тестирования моделей, основные факторы, влияющие на оценку создаваемых моделей и оценку их качества, а также приведены примеры использования методов хеометрики и МО для обработки спектров флуоресценции и КР для различных объектов.

1.1 Методы флуоресцентной спектроскопии

Флуоресцентная спектроскопия является одним из наиболее чувствительных и информативных методов исследования структуры и свойств молекул [6]. В основе метода лежит явление флуоресценции – способности вещества излучать свет при переходе электронно-возбужденных частиц в основное состояние [7].

Процесс флуоресценции можно разделить на три основных этапа [8]:

1. Поглощение кванта света молекулой и переход ее в возбужденное состояние (происходит за $\sim 10^{-15}$ с);
2. Внутренняя конверсия и колебательная релаксация (10^{-14} – 10^{-11} с);
3. Излучательный переход в основное состояние с испусканием кванта флуоресценции (10^{-9} – 10^{-7} с).

Важной особенностью флуоресценции является то, что спектр испускания всегда сдвинут в длинноволновую область относительно спектра поглощения

(правило Стокса). Это связано с потерей части энергии при колебательной релаксации молекулы. Величина стоксова сдвига является важной характеристикой флуорофора (молекулы, способной флуоресцировать) и может давать информацию о его микроокружении и взаимодействиях с растворителем.

Квантовый выход флуоресценции (ϕ) – это отношение числа испущенных фотонов к числу поглощенных. Он является одной из важнейших характеристик флуорофора и может быть выражен через константы скоростей различных процессов

$$\phi = \frac{k_f}{(k_f + k_{nr})}, \quad (1)$$

где k_f – константа скорости флуоресценции; k_{nr} – сумма констант скоростей безызлучательных процессов [9].

Время жизни возбужденного состояния τ определяется как

$$\tau = \frac{1}{(k_f + k_{nr})}. \quad (2)$$

Эти параметры крайне чувствительны к окружению флуорофора и могут использоваться для исследования различных межмолекулярных взаимодействий.

Спектры флуоресценции обладают рядом фундаментальных свойств, которые необходимо учитывать при проведении спектрального анализа. Одним из таких свойств является независимость формы спектра испускания от длины волны возбуждения, известное как правило Каши [10]. Это явление объясняется тем, что избыточная колебательная энергия быстро рассеивается, и флуоресценция всегда происходит с нижнего колебательного уровня первого возбужденного электронного состояния.

Особое значение для аналитического применения имеет зависимость интенсивности флуоресценции от концентрации флуорофора. В области низких концентраций наблюдается линейная зависимость, описываемая законом Паркера [11]. При высоких концентрациях возникают эффекты внутреннего фильтра и концентрационного тушения, приводящие к отклонению от линейности.

Температура является одним из ключевых факторов, влияющих на флуоресценцию. Повышение температуры обычно приводит к уменьшению квантового выхода флуоресценции вследствие увеличения вероятности безызлучательной дезактивации возбужденного состояния. Этот эффект особенно важен при разработке количественных методик анализа и требует тщательного контроля температурного режима [12].

pH среды оказывает существенное влияние на спектральные характеристики многих флуорофоров, особенно тех, которые содержат ионизируемые группы. Изменение pH может приводить как к изменению интенсивности флуоресценции, так и к спектральным сдвигам, что используется при разработке pH-чувствительных флуоресцентных сенсоров [13].

Концентрация флуорофора также является критическим фактором, влияющим на флуоресценцию. При высоких концентрациях могут наблюдаться эффекты самотушения и образования эксимеров, что приводит к изменению спектральных характеристик [14].

Присутствие кислорода в растворе может значительно снижать интенсивность флуоресценции за счет динамического тушения. Этот эффект особенно важен при проведении количественных измерений и часто требует деаэрации образцов [14].

Агрегация молекул флуорофора может приводить к существенным изменениям спектральных характеристик. В некоторых случаях образование агрегатов может приводить к усилению флуоресценции, что находит применение при разработке новых флуоресцентных материалов [16].

Современная теория флуоресценции также включает рассмотрение эффектов, связанных с поляризацией излучения. Степень поляризации флуоресценции определяется ориентационной подвижностью флуорофора и может использоваться для изучения микровязкости среды, процессов молекулярного вращения и белок-лигандных взаимодействий [17].

Современные методы регистрации флуоресценции основаны на использовании высокочувствительных спектрофлуориметров, позволяющих

проводить измерения с высокой точностью и воспроизводимостью [18]. Базовая конструкция спектрофлуориметра включает источник возбуждающего излучения, монохроматоры для выделения длин волн возбуждения и эмиссии, кюветное отделение и детектор.

В качестве источников возбуждающего излучения традиционно используются ксеноновые лампы высокого давления, обеспечивающие непрерывный спектр в широком диапазоне длин волн (250–700 нм). В последние годы всё большее распространение получают светодиодные источники и лазеры, обладающие высокой монохроматичностью, стабильностью и длительным сроком службы.

Выделение требуемых длин волн осуществляется с помощью монохроматоров различных типов. Наиболее распространены дифракционные монохроматоры с голографическими решетками, обеспечивающие высокое спектральное разрешение и низкий уровень рассеянного света. В современных приборах часто используются двойные монохроматоры, позволяющие существенно снизить уровень паразитного излучения.

Детектирование флуоресцентного излучения осуществляется с помощью различных типов фотоприемников. Традиционно используются фотоэлектронные умножители (ФЭУ), обеспечивающие высокую чувствительность и широкий динамический диапазон. В современных приборах также применяются полупроводниковые детекторы – фотодиоды и приборы с зарядовой связью (ПЗС) – матрицы, позволяющие проводить многоканальную регистрацию спектров [19].

Особое внимание уделяется разработке многомодальных систем, сочетающих флуоресцентную спектроскопию с другими методами анализа. Например, комбинация флуоресцентной и абсорбционной спектроскопии позволяет получать дополнительную информацию о фотофизических процессах в исследуемых системах. Сочетание флуоресцентных измерений с методами электрохимии открывает новые возможности для изучения окислительно-восстановительных процессов [20].

Важным направлением развития инструментальных методов является миниатюризация приборов и создание портативных флуориметров. Такие устройства находят широкое применение в полевых исследованиях, экологическом мониторинге, клинической диагностике и т.д. [21].

Современные системы регистрации флуоресценции все чаще оснащаются микрофлюидными элементами, что позволяет проводить анализ чрезвычайно малых объемов образцов. Интеграция микрофлюидных чипов с флуоресцентными детекторами открывает новые возможности для высокопроизводительного скрининга и анализа единичных молекул [22].

Синхронная флуоресцентная спектроскопия является мощным аналитическим методом, который позволяет одновременно регистрировать спектры возбуждения и эмиссии при постоянной разности длин волн между ними. При синхронном сканировании монохроматоры возбуждения и эмиссии движутся одновременно, поддерживая постоянную разность длин волн ($\Delta\lambda$) между ними.

Современный флуоресцентный анализ представляет собой комплекс методических подходов, позволяющих решать широкий спектр аналитических задач. В основе этих методик лежит использование различных типов флуоресцентных зондов и меток, специфически взаимодействующих с определяемыми веществами.

Особый интерес представляют методики мультипараметрического флуоресцентного анализа, позволяющие одновременно регистрировать несколько характеристик флуоресценции. Комбинация различных параметров (интенсивность, время жизни, анизотропия) повышает информативность и достоверность анализа [23].

В области биоанализа широкое распространение получили методики на основе флуоресцентных наносенсоров. Такие сенсоры позволяют проводить измерения различных параметров (pH, концентрации ионов, метаболитов) непосредственно в живых клетках.

Развиваются методики количественного флуоресцентного анализа с использованием стандартных добавок и внутренних стандартов. Особое внимание

уделяется разработке процедур калибровки и оценки метрологических характеристик методик [24].

Особое внимание уделяется разработке методик мультиплексного анализа, позволяющих одновременно определять несколько аналитов. Для этого используются флуорофоры с различными спектральными характеристиками или квантовые точки.

В области экологического мониторинга развиваются методики определения загрязняющих веществ на основе флуоресцентных хемосенсоров [25]. Эти методики позволяют проводить экспрессное определение токсикантов в различных объектах окружающей среды.

Сравним флуоресцентную спектроскопию с абсорбционной спектроскопией УФ, видимого и ближнего ИК диапазона (от 190 до 1100 нм).

Абсорбционная спектроскопия основана на явлении поглощения света атомами и молекулами при определенной частоте или длине волны. Это происходит, когда фотоны света, проходя через образец, поглощаются, вызывая переход электронов на более высокие энергетические уровни. Ослабление интенсивности начального светового потока, проходящего через образец, описывается законом Бугера-Ламберта-Бера:

$$I = I_0 e^{-\alpha(\lambda)xc}, \quad (3)$$

где I_0 — начальная интенсивность света; I — интенсивность света на выходе; x — толщина образца; c — концентрация вещества; $\alpha(\lambda)$ — коэффициент ослабления (экстинкции).

В ослабление света, наряду с поглощением, вносит вклад также рассеяние на неоднородностях показателя преломления внутри образца. Этот закон позволяет по поглощению света определять концентрацию поглощающего вещества в образце. Флуоресцентная спектроскопия обладает более высокой чувствительностью по отношению к абсорбционной, это объясняется, в частности, тем, что сигнал флуоресценции измеряется от уровня низкого фона [26]. Поэтому для низких концентраций флуоресцентная спектроскопия является предпочтительной (разумеется, при наличии характерных линий

флуоресценции у аналита), а для относительно высоких, когда возможно насыщение флуоресценции, лучше подходит спектроскопия абсорбционная.

1.2 Методы спектроскопии КР (рамановской)

Эффект КР света является результатом неупругого взаимодействия фотонов с молекулами исследуемого вещества. В основе этого явления лежит модуляция поляризуемости молекулы колебательными движениями ядер, что приводит к появлению в спектре рассеяния новых спектральных линий, сдвинутых относительно частоты возбуждающего излучения [27].

Теоретические основы КР базируются на фундаментальных принципах взаимодействия электромагнитного излучения с веществом. Неупругое рассеяние света молекулами приводит к появлению в спектре новых линий, частоты которых отражают структуру энергетических уровней исследуемых молекул. Особое значение имеет связь интенсивности линий КР с изменением поляризуемости молекул при колебаниях, что делает метод комплементарным к ИК спектроскопии.

Процесс КР можно представить как последовательность следующих событий [28]:

1. Поглощение падающего фотона молекулой с переходом в виртуальное состояние;
2. Практически мгновенное (порядка 10^{-14} с) излучение рассеянного фотона;
3. Переход молекулы в новое колебательное состояние.

При рассеянии света могут наблюдаться два типа процессов:

- стоксово КР (с потерей энергии фотона);
- антистоксово КР (с увеличением энергии фотона).

Интенсивность линий КР определяется выражением

$$I = const \times N \times I_0 \times \nu^4 \times (\partial\alpha/\partial Q)^2, \quad (4)$$

где N – число рассеивающих молекул; I_0 – интенсивность возбуждающего излучения; ν – частота рассеянного излучения; $\partial\alpha/\partial Q$ – производная поляризуемости по нормальной координате колебания.

Важной характеристикой КР является поляризация рассеянного излучения. Степень поляризации рассеянного света определяется симметрией колебательных мод молекулы и может быть использована для получения дополнительной структурной информации [29].

Сечение КР, определяющее вероятность процесса, обычно составляет порядка 10^{-30} см², что существенно меньше сечения флуоресценции (от 10^{-16} до 10^{-20} см²). Это приводит к относительно низкой интенсивности сигнала КР и необходимости использования интенсивных источников возбуждения.

Правила отбора для КР определяются изменением поляризуемости молекулы при колебании

$$\partial\alpha/\partial Q \neq 0, \quad (5)$$

где α – тензор поляризуемости; Q – нормальная координата колебания.

В отличие от ИК-спектроскопии, где активны колебания с изменением дипольного момента, в КР-спектроскопии могут наблюдаться полносимметричные колебания, что делает эти методы взаимодополняющими [30].

Интенсивность и качество спектров КР зависят от множества физических и химических факторов. Понимание этих факторов критически важно для получения надежных и воспроизводимых результатов.

Длина волны возбуждающего излучения является одним из ключевых параметров, определяющих характеристики спектров. При уменьшении длины волны, интенсивность рассеяния возрастает пропорционально λ^{-4} , однако одновременно увеличивается вероятность возникновения флуоресценции, которая может маскировать сигнал КР [31].

Мощность лазерного излучения также существенно влияет на качество получаемых спектров. Хотя увеличение мощности приводит к росту интенсивности сигнала, необходимо учитывать возможность локального нагрева

образца и его фотохимической деструкции. Особенно критичен этот фактор при анализе биологических объектов и термически нестабильных соединений [32].

Физическое состояние образца также влияет на качество получаемых спектров. Для кристаллических веществ важную роль играет ориентация кристаллов относительно поляризации возбуждающего излучения. В случае жидкостей необходимо учитывать эффекты рассеяния и преломления света, а также возможность конвекции при локальном нагреве образца.

Концентрация анализируемого вещества является критическим параметром, особенно при количественном анализе [33]. При низких концентрациях основным ограничением является соотношение сигнал/шум, в то время как при высоких концентрациях могут проявляться эффекты внутреннего фильтра и межмолекулярных взаимодействий.

Наличие примесей и загрязнений может значительно осложнять анализ спектров КР. Особенно критичным является присутствие флуоресцирующих примесей, даже в следовых количествах. Для минимизации их влияния разработаны различные методические приемы, включая использование длинноволнового возбуждения и временное разрешение.

Оптические характеристики образца являются одними из определяющих факторов в формировании спектров КР. Среди них особое значение имеет показатель преломления среды, который существенно влияет на процессы взаимодействия излучения с веществом. При прохождении возбуждающего излучения через образец, показатель преломления определяет глубину проникновения света и эффективность его фокусировки. Это особенно важно при проведении микроскопических исследований, где требуется высокое пространственное разрешение [34]. Оптическая прозрачность образца играет ключевую роль в формировании аналитического сигнала. В случае прозрачных образцов возможно получение информации из их внутренних областей, однако необходимо учитывать эффекты многократного рассеяния, которые могут существенно исказить спектральную картину. Для непрозрачных образцов анализ

ограничивается приповерхностным слоем, что требует особого внимания к состоянию поверхности и выбору геометрии измерений.

Современные приборы для регистрации спектров КР представляют собой сложные оптико-электронные системы. Основными компонентами КР-спектрометра являются:

1. источник возбуждающего излучения;
2. оптическая система;
3. спектральный прибор;
4. система детектирования;
5. система сбора и обработки данных.

В качестве источников возбуждения в современных КР-спектрометрах наиболее часто используются лазеры различных типов:

- аргоновые лазеры (488; 514,5 нм);
- неодимовые лазеры (532 нм);
- гелий-неоновые лазеры (632,8 нм);
- диодные лазеры (785; 830 нм);
- титан-сапфировые лазеры (перестраиваемые).

Оптическая система включает элементы для фокусировки излучения на образец и сбора рассеянного света [35]. Особое внимание уделяется фильтрации рэлеевского рассеяния с помощью голографических режекторных фильтров или монохроматоров.

Спектральные приборы в современных КР-спектрометрах представлены несколькими основными типами:

1. одинарные монохроматоры – наиболее простые приборы, используемые для рутинного анализа;
2. двойные и тройные монохроматоры – обеспечивают лучшее подавление рассеянного света;
3. спектрографы с голографическими решетками – позволяют получать спектры с высоким разрешением;

4. Фурье-спектрометры КР – обеспечивают высокую точность определения частот.

Системы детектирования прошли значительную эволюцию [36]:

- ФЭУ – использовались в ранних приборах;
- диодные линейки – обеспечивают многоканальное детектирование;
- ПЗС-матрицы – современные детекторы с высокой квантовой эффективностью;
- электронно-умножающие ПЗС – для регистрации слабых сигналов;
- комплементарные металл-оксидные полупроводники – новейшие быстродействующие детекторы.

Существенный прогресс достигнут в области автоматизации пробоподготовки и измерений. Современные приборы оснащаются автоматическими системами ввода проб, проточными ячейками и автосемплерами, что позволяет существенно повысить производительность анализа и снизить влияние человеческого фактора на результаты измерений.

Методы спектроскопии КР представляют собой мощный инструмент современной аналитической химии, позволяющий получать уникальную информацию о молекулярной структуре и составе исследуемых веществ. Детальное рассмотрение различных аспектов КР-спектроскопии позволяет выделить несколько ключевых направлений развития этого метода.

В области инструментальных методов регистрации КР-спектров достигнут значительный прогресс, связанный с развитием лазерной техники, оптических систем и детекторов. Современные КР-спектрометры представляют собой сложные комплексы, включающие высокостабильные источники возбуждения, эффективные системы фильтрации рассеянного света и высокочувствительные детекторы. Особое место занимают конфокальные КР-микроскопы, позволяющие получать пространственное распределение химических компонентов с высоким разрешением.

Понимание факторов, влияющих на КР-спектры, имеет ключевое значение для получения надежных аналитических результатов. Среди этих факторов особое

место занимают длина волны и мощность возбуждающего излучения, температура образца, его физическое состояние и оптические характеристики. Учет этих факторов позволяет оптимизировать условия эксперимента и минимизировать влияние мешающих эффектов.

Таким образом, КР-спектроскопия представляет собой динамично развивающуюся область аналитической химии, где успешно сочетаются фундаментальные физические принципы и современные инструментальные решения. Дальнейшее развитие метода связано с совершенствованием приборной базы, разработкой новых методических подходов и применением современных методов анализа данных.

1.3 Методы хеометрики и МО, предварительная обработка данных. Применение методов хеометрики и МО для анализа жидких многокомпонентных сред (в частности, спектров флуоресценции и КР)

Хеометрика представляет собой научную область, в которой применяются математические, статистические и иные методы, опирающиеся на формальную логику, для обработки химических данных с целью извлечения максимально полезной информации. Хеометрическая обработка данных является неотъемлемой частью современного флуоресцентного и КР-анализа, позволяющей извлекать максимум информации из экспериментальных данных. Также для хеометрической обработки данных критически важным является предварительная обработка спектральных данных. Качество этой обработки напрямую определяет эффективность последующего хеометрического анализа и успешность применения методов МО [37].

При разработке аналитических методов важно понимать основные типы решаемых задач и их специфику. Рассмотрим три фундаментальных направления применения МО и хеометрики в этой области.

Классификация (качественный анализ) – устанавливает принадлежность образца к определенному классу по аналитическим сигналам спектрального происхождения с определенной доверительной вероятностью. Доверительная

вероятность для каждого алгоритма оценивается по формулам, специфичным для конкретного метода классификации [38], это разновидность дискриминантного анализа (DA). Пороговое значение доверительной вероятности для отнесения к определенному классу (иногда называемое порогом достоверности) определяется, исходя из специфики конкретной задачи. Каких-либо единых общепринятых правил для назначения этой величины в настоящее время не существует. Авторы работы [39] предлагают подход подбора наилучшей точности предсказания, исходя из результатов перебора пороговых доверительных вероятностей, выбирается наилучшая комбинация подбора. Примеры использования классификации в контексте анализа напитков:

- определение типа (происхождения);
- выявление подлинности продукта (подлинный или фальсификат).

Кластеризация – это задача разделения набора образцов на группы (кластеры) таким образом, чтобы объекты внутри одного кластера были больше похожи друг на друга, чем на объекты из других кластеров. В отличие от классификации, кластеризация является методом обучения без учителя, то есть не требует предварительного разделения данных на определенные группы. В основе кластеризации лежит концепция определения естественной структуры данных на основе меры сходства или расстояния между образцами. Различные алгоритмы кластеризации используют разные подходы к определению этого сходства и формированию кластеров.

Регрессия – это задача предсказания непрерывной числовой величины на основе характеристик образца. В анализе жидкостей это может включать:

- прогнозирование концентрации различных компонентов;
- предсказание физико-химических параметров (например, вязкости).

В случаях классификации и регрессии алгоритм обучается на наборе данных с известными значениями определяемых параметров (обучающая выборка), чтобы затем делать предсказания для новых образцов. При этом используются различные характеристики образцов, такие как спектральные данные, химический состав и т.д.

Для хемометрической обработки спектров флуоресценции и спектров КР одним из важнейших этапов является: выбор информативного диапазона спектральных данных, который зависит от проводимого анализа, и исследуемого вещества. Например, для КР-спектроскопии диапазон $750\text{--}950\text{ см}^{-1}$ – область проявления характеристических колебаний органических соединений [40].

Важным аспектом является также выявление и устранение выбросов – аномальных измерений, которые могут существенно повлиять на результаты анализа. Для этого применяются:

- метод Хотеллинга T^2 ;
- Q-статистика;
- метод ROBPCA [41];
- визуальный анализ спектров.

Также важное значение имеет корректная предварительная обработка данных. Самыми распространенными методами предварительной обработки являются:

- фильтр Савицкого-Голея;
- нормализация спектров (нормировка);
- коррекция по (вычитание) базовой линии;
- центрирование;
- масштабирование.

При анализе спектров КР наиболее эффективным методом предварительной обработки является фильтр Савицкого-Голея. Этот метод позволяет сохранить важные спектральные особенности сигнала, такие как высота и ширина пиков, что критично для последующего анализа компонентного состава веществ. Фильтр работает по принципу аппроксимации участков спектра полиномами небольших степеней методом наименьших квадратов в скользящем окне. Это позволяет эффективно удалять высокочастотный шум, сохраняя при этом форму спектральных линий. Параметрами фильтра Савицкого-Голея являются размер окна (обычно выбирается нечетным, например 11-15 точек), степень полинома

(как правило, 2-4 степени достаточно для большинства случаев) и порядок производной (0 для сглаживания, 1 или 2 для выделения особенностей спектра).

После применения фильтрации становится возможным более точное определение положений характеристических пиков различных компонентов вещества, таких как этанол, сложные эфиры, альдегиды и другие органические соединения. Преимущество данного метода заключается также в его вычислительной эффективности и простоте реализации в современных программных пакетах для спектрального анализа. Правильный выбор параметров фильтрации Савицкого-Голея является ключевым для получения качественных результатов при дальнейшем анализе спектральных данных.

Нормализация спектров позволяет привести данные к единому масштабу и компенсировать влияние различных экспериментальных факторов. В современной практике используются следующие подходы к нормализации:

- нормировка по площади под спектром – позволяет компенсировать различия в общей интенсивности спектров, обусловленные концентрацией образца;
- векторная нормализация – приведение данных к единичной длине вектора;
- стандартизация данных (z-score) – преобразование к нулевому среднему и единичной дисперсии;
- внутренняя нормировка на полосу растворителя [42].

Коррекция базовой линии – подход, направленный на устранение систематических смещений данных, что способствует более точной интерпретации основных сигналов. Настраивая базовую линию, аналитики имеют возможность повысить соотношение сигнал/шум, что является важным для выделения значимых элементов из сложных наборов данных. В спектроскопии необработанные данные зачастую содержат шум и дрейф базовой линии, которые могут затушевывать истинный сигнал. Внедрение коррекции базовой линии позволяет обеспечить более чистые данные для анализа, что, в свою очередь, приводит к более достоверным результатам.

Процедура центрирования данных заключается в том, чтобы привести все данные к среднему значению, равному нулю. Процедура масштабирования данных представляет собой приведение всех данных спектра в определённый диапазон с максимумом и минимумом интенсивности.

Для оценки качества предварительной обработки данных используются следующие критерии:

1. отношение сигнал/шум в спектрах;
2. воспроизводимость положения и интенсивности спектральных полос;
3. стабильность базовой линии.

При предварительной обработке спектров необходимо учитывать особенности конкретных типов образцов (пищевая продукция, органические соединения, неорганические соединения и т.д.).

Для обработки спектральных данных используются различные алгоритмы МО, Эти методы позволяют извлекать значимую информацию из сложных спектральных профилей и строить предсказательные модели.

Метод главных компонент (Principal Component Analysis, PCA) является одним из базовых методов хеометрики. PCA решает задачу снижения размерности данных с сохранением максимума полезной информации.

При анализе спектров метод PCA позволяет трансформировать многомерное пространство спектральных переменных в набор новых, некоррелированных переменных – главных компонентах [43]. Каждая главная компонента представляет собой линейную комбинацию исходных переменных, при этом первые компоненты описывают наибольшую долю вариации данных.

В случае спектрального анализа многокомпонентных жидких сред исходные данные могут включать сотни или тысячи значений спектральных сигналов, что создает избыточность данных и усложняет их интерпретацию. PCA позволяет сократить размерность до нескольких главных компонент (обычно 2-5), которые объясняют 90-95 % общей дисперсии данных. Это значительно упрощает визуализацию и последующий анализ. Важной особенностью PCA является то, что он сохраняет наиболее значимые спектральные особенности, отражающие

химический состав образцов. При этом шумы и случайные вариации, как правило, попадают в компоненты с меньшими собственными значениями, что позволяет эффективно отфильтровывать неинформативную часть данных. Например, если у нас есть ковариационная матрица признаков A , то при её спектральном разложении получаем

$$A * B_i = c_i * B_i, \quad (6)$$

где B_i – собственный вектор (направление главной компоненты), c_i – собственное значение (дисперсия данных конкретной компоненты).

То есть собственное значение для каждой компоненты показывает, долю дисперсии в компоненте. Кроме того, анализ главных компонент дает возможность выявить спектральные области, наиболее важные для различения образцов, что может быть использовано для понимания химических различий между разными типами анализируемых образцов, например, выдержанных винных дистиллятов [44].

Линейный дискриминантный анализ (Linear discriminat analysis, LDA) – это поиск линейных зависимостей между спектральными данными и разными классами. В (LDA) все классы имеют одинаковую корреляционную функцию (ковариационная матрица). Преимуществами (LDA) являются снижение размерности линейными комбинациями, высокая интерпретируемость результатов, устойчивость результатов и высокая скорость обучения и прогнозирования.

Квадратичный дискриминантный анализ (Quadratic discriminat analysis, QDA) – это поиск нелинейных зависимостей между спектральными данными и разными классами. В (QDA) все классы имеют разную корреляционную функцию. Преимуществами (QDA) являются индивидуальные ковариационные матрицы для каждого класса, учет уникальной структуры корреляции и высокая скорость обучения и прогнозирования

Метод k-ближайших соседей (k-nearest neighbors, KNN) – это непараметрический метод классификации, основанный на принципе отнесения объекта к тому классу, к которому принадлежит большинство из k ближайших к

нему объектов обучающей выборки. При анализе спектральных данных многокомпонентных жидких сред метод k -ближайших соседей позволяет эффективно классифицировать образцы на основе их спектральных характеристик, используя различные метрики расстояния (евклидово, манхэттенское, косинусное). Преимуществами метода являются простота реализации, отсутствие необходимости обучения модели и способность обрабатывать многоклассовые задачи. Однако эффективность метода k -ближайших соседей существенно зависит от выбора параметра k и используемой метрики расстояния [45].

Метод частичных наименьших квадратов (Partial Least Squares, PLS) – это статистический метод, устанавливающий связь между набором независимых переменных X (спектральные данные) и набором зависимых переменных Y (характеристики образцов, интенсивность). PLS особенно эффективен при анализе спектральных данных, где число переменных значительно превышает число наблюдений, а также при наличии мультиколлинеарности. Алгоритм PLS используется как для классификации (PLS-DA), так и для регрессии (PLS-R). PLS позволяет одновременно снижать размерность данных и строить предсказательные модели для количественного определения содержания компонентов или других характеристик [46].

Регрессия на главные компоненты (Principal Component Regression, PCR) – это двухэтапный метод, сочетающий PCA и множественную линейную регрессию. Сначала с помощью PCA снижается размерность данных и выделяются главные компоненты, затем эти компоненты используются как предикторы в множественной линейной регрессии. PCR особенно полезен при анализе спектральных данных, когда необходимо устранить мультиколлинеарность и шум, сохраняя при этом наиболее информативные данные спектров для построения регрессионных моделей [47].

Многослойный перцептрон (Multilayer Perceptron, MLP) – это искусственная нейронная сеть прямого распространения, состоящая из нескольких слоев нейронов: входного, скрытых и выходного. MLP обучается

методом обратного распространения ошибки и способен моделировать сложные нелинейные зависимости между спектральными данными и характеристиками анализируемых образцов. MLP может эффективно обрабатывать большие массивы данных и выявлять скрытые закономерности, которые трудно обнаружить традиционными статистическими методами. Основным недостатком является склонность к переобучению, что требует применения методов регуляризации и тщательного подбора архитектуры сети [48].

Экстремальный градиентный бустинг (ЭГБ) – это алгоритм МО, основанный на деревьях решений, отличающийся высокой эффективностью и точностью.

Метод последовательно строит ансамбль деревьев решений, где каждое новое дерево обучается на ошибках предыдущих деревьев. ЭГБ использует более сложную модель регуляризации, т.е. добавление дополнительных условий задачи, помогающих предотвратить переобучение. Переобучение модели означает идеальную точность предсказаний для обучающей выборки, но для предсказания тестовых образцов такая модель не сработает. В алгоритме ЭГБ вопрос переобучения решается методами регуляризации, чтобы уменьшить сложность моделей.

Преимущества метода ЭГБ включают:

- устойчивость к переобучению, повышенную устойчивость модели к шумовым воздействиям;
- высокую скорость обучения;
- способность работать с ограниченными наборами данных.

ЭГБ широко применяется в задачах классификации и регрессии, особенно в случаях, когда требуется высокая точность предсказаний при ограниченном объеме обучающих данных [49].

Базовые настройки алгоритма: количество моделей решений или количество итераций модели, глубина моделей (или дерева решений) и настройка количества итераций в ветке решений.

Пример глубины модели продемонстрирован на рисунке 1.

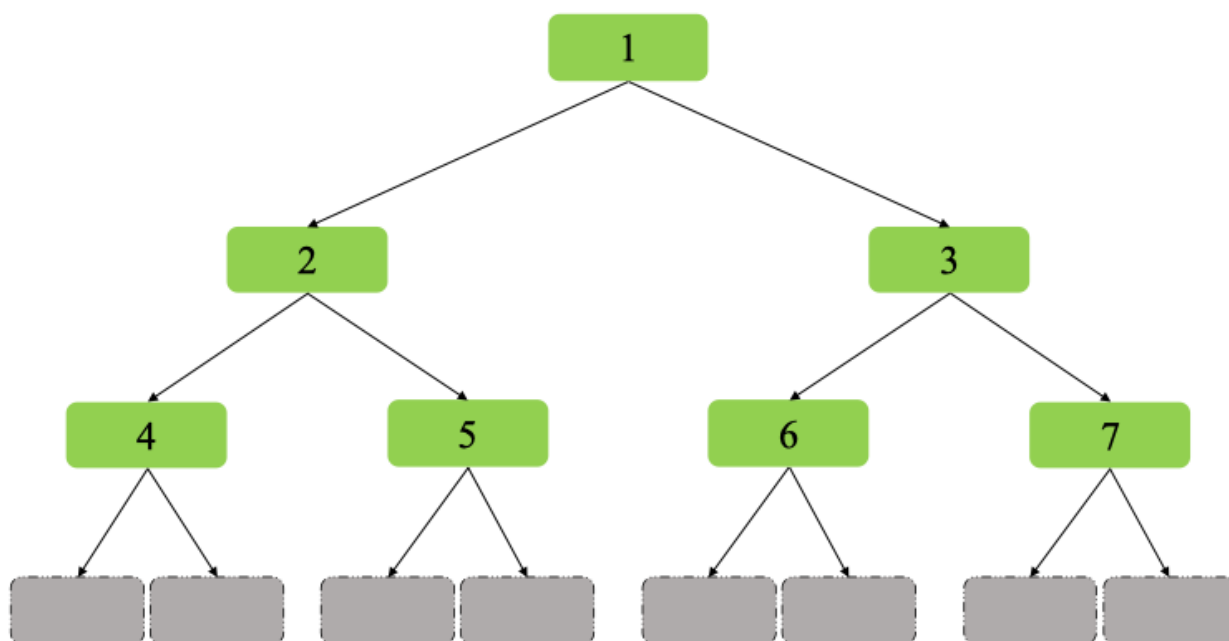


Рисунок 1 – Пример глубины модели или дерева решений

Скорость обучения — это коэффициент, по которому предотвращается переобучение модели $\alpha \in (0,1]$. Если расписать базовое уравнение в модифицированном виде, то оно будет иметь следующий вид

$$A_{n+1}(x) = r_1(x) + \alpha \times r_2(x) + \alpha \times r_3(x) + \dots + \alpha \times r_{n+1}(x). \quad (7)$$

Обычно параметр α определяется экспериментальным путем по входным данным.

Метод случайного леса (Random Forest, RF) — это ансамблевый метод МО, основанный на построении множества деревьев решений [50]. Каждое дерево в случайном лесу строится независимо с использованием случайной подвыборки из обучающего набора данных и случайного набора признаков. В случае анализа спектральных данных каждое дерево обучается на разных участках спектра и разных образцах из обучающей выборки. При классификации нового образца каждое дерево дает свой «голос», и окончательное решение принимается путем голосования большинства деревьев.

Данный метод обладает следующими преимуществами:

- встроенная оценка важности признаков, позволяющая выделить наиболее информативные участки спектров;

- устойчивость к переобучению за счет усреднения результатов множества деревьев решений;
- возможность эффективного распараллеливания вычислений;
- способность работать с нелинейными зависимостями;
- низкая чувствительность к выбросам в данных.

Анализ спектральных данных методом случайного леса особенно эффективен благодаря способности обрабатывать большое количество спектральных переменных и выявлять сложные нелинейные зависимости между спектральными характеристиками и целевыми параметрами. Метод успешно применяется для решения задач классификации и регрессии в спектральном анализе.

Метод опорных векторов (Support vector machine, SVM) также показал высокую эффективность при анализе спектральных данных. Это алгоритм МО, который создает разделяющую гиперплоскость для задач классификации или регрессии. Ключевая концепция SVM состоит в том, чтобы определить гиперплоскость, максимально увеличивающую расстояние между классами, что, в свою очередь, способствует лучшему обобщению на новых данных. Его основные преимущества при анализе [51]:

- способность работать с нелинейными зависимостями через различные ядерные функции;
- устойчивость к шуму в данных;
- хорошая обобщающая способность.

Алгоритм ЭГБ освещён подробнее, чем остальные, так как данный алгоритм новый и использовался в рамках данной работы чаще, чем другие алгоритмы.

При анализе многокомпонентных смесей выбор конкретного метода обработки спектральных данных зависит от ряда ключевых факторов. В работе Sarker I.H. [52] показано, что эффективность применения методов МО определяется несколькими основными параметрами, которые кратко рассматриваются ниже .

Требования к точности анализа также существенно влияют на выбор метода. При необходимости высокой точности классификации целесообразно использовать комбинацию различных спектральных методов.

Необходимость интерпретации результатов играет важную роль при выборе метода анализа. Некоторые методы МО, такие как нейронные сети, могут обеспечивать высокую точность классификации, но при этом работают как «черный ящик», затрудняя интерпретацию результатов. В случаях, когда важно понимать физико-химическую природу различий между образцами, предпочтительно использование более прозрачных методов, допускающих визуализацию данных, таких как РСА или линейный дискриминантный анализ.

Вычислительные ресурсы и временные затраты также являются важными факторами при выборе метода анализа. Для рутинного анализа большого количества образцов предпочтительны методы, требующие минимальных вычислительных затрат при сохранении приемлемой точности классификации.

При использовании методов МО особое внимание уделяется валидации моделей. В настоящее время используются следующие методы валидации:

1. Внутренняя валидация на обучающем наборе:
 - кросс-валидация с разделением на 5-10 блоков либо кросс-валидация «Оставить одного без внимания» (Leave One Out). Каждый обучающий набор создается путем взятия всех образцов, кроме одного, а проверочный набор – это оставленный образец;
 - проверка устойчивости результатов при разных разбиениях данных.
2. Внешняя валидация на независимом тестовом наборе:
 - оценка точности классификации и регрессии;
 - расчет доверительных вероятностей предсказаний (для классификации);
 - анализ матрицы ошибок либо расчёт погрешности предсказания для каждого образца для алгоритмов регрессии;
 - выявление систематических ошибок классификации и регрессии.
3. Проверка на «фальш-объектах»:

- анализ образцов, не относящихся ни к одному из заданных классов либо с ложным указанием концентрации;
- оценка способности модели выявлять нетипичные образцы;
- определение пороговых значений доверительной вероятности только для классификации.

При разработке методов МО важную роль играет правильный выбор метрик качества, т.е. показателей, используемых для оценки качества модели. В качестве таких показателей могут использоваться следующие.

Для задач классификации [53]:

- достоверность (ассигасу) – доля правильно классифицированных образцов;
- точность – доля истинно положительных среди всех положительных предсказаний;
- полнота – доля обнаруженных истинно положительных случаев;
- f1-мера – гармоническое среднее между точностью и полнотой;
- доверительная вероятность предсказаний.

Для задач регрессии:

- среднеквадратичная ошибка (MSE) – среднее квадратов разностей между предсказанными и фактическими значениями;
- корень из среднеквадратичной ошибки – квадратный корень из MSE, измеряется в тех же единицах, что и целевая переменная;
- средняя абсолютная ошибка – средняя абсолютных разностей между предсказанными и фактическими значениями;
- коэффициент детерминации (R^2) – доля дисперсии зависимой переменной, объясняемая моделью;
- средняя абсолютная процентная ошибка – средняя абсолютных процентных отклонений предсказаний от фактических значений;
- объясненная дисперсия – мера того, насколько модель учитывает вариацию данных.

Особенности конкретной аналитической задачи также определяют выбор метода. При этом необходимо учитывать:

1. специфику анализируемой продукции (сложность химического состава, стабильность характеристик во времени, влияние условий хранения, возможные способы фальсификации);
2. требования к достоверности результатов (необходимый уровень доверительной вероятности, допустимый процент ложноположительных результатов, требования к воспроизводимости, устойчивость к влиянию внешних факторов);
3. практические применения (простота использования метода, возможность автоматизации анализа, требования к квалификации персонала, экономическая эффективность).

Важным фактором является также возможность адаптации метода к изменяющимся условиям. Методы МО должны обеспечивать возможность:

- добавления новых классов образцов;
- учета сезонных вариаций характеристик;
- адаптации к изменениям в технологии производства;
- корректировки параметров при смене измерительного оборудования.

При выборе метода анализа необходимо также учитывать возможность его валидации и стандартизации для применения в рутинном анализе. Это включает:

- разработку стандартных операционных процедур;
- определение метрологических характеристик;
- оценку неопределенности результатов;
- проведение межлабораторных сличений [54].

Методы многомерного анализа данных играют ключевую роль в извлечении химической информации из спектров флюоресценции и КР. Наиболее широко используется PCA, позволяющий выявить основные источники вариативности в спектральных данных и снизить размерность данных без существенной потери информации [55].

Методы многомерной градуировки являются важнейшим инструментом количественного анализа по спектрам флуоресценции и КР. Наиболее широко применяются следующие подходы к построению градуировочных моделей.

PLS позволяет создавать надежные модели даже в случае сильно коррелированных спектральных переменных. Особую важность имеет выбор оптимального числа латентных переменных для предотвращения переобучения модели и обеспечения её робастности [56].

Дискриминантный анализ широко используется для решения задач классификации и установления подлинности различных материалов. Современные алгоритмы МО, включая нейронные сети и методы глубокого обучения, показывают высокую эффективность при анализе сложных спектральных данных [57].

При анализе выдержанных алкогольных напитков методами оптической спектроскопии важно учитывать взаимное влияние компонентов сложной смеси, эффекты среды (рН, ионная сила), межмолекулярные взаимодействия и процессы агрегации [58-59] и т.д. Например, при анализе спектров вин с дальнейшей обработкой методами МО, основное внимание уделяется определению подлинности, географического происхождения, возраста и качественных характеристик винной продукции. При анализе вин применяется комбинация спектральных методов с алгоритмами МО. Основные спектральные методы включают спектроскопию среднего инфракрасного диапазона (как правило, с использованием Фурье-спектрометров) [60], спектроскопию в ближней инфракрасной области, спектроскопию УФ-видимого диапазона, спектроскопию КР.

Перечисленные выше оптические спектры вин определяются ключевыми параметрами: содержание этанола, органических кислот (винной, яблочной, молочной), сахаров, полифенольных соединений и летучих ароматических компонентов. Для определения подлинности вин оценивается содержание характерных маркерных соединений и соотношение изотопов. Спектральные методы позволяют проводить быстрый скрининг образцов без сложной

пробоподготовки и, во многих случаях, без количественного определения содержания основных компонентов. В описанном процессе анализа вин используются методы многомерной классификации и регрессионного анализа. Авторы статьи применяют методы PCA, дискриминантный анализ, PLS и МО на основе искусственных нейронных сетей [61] для создания моделей классификации и количественного анализа.

Данный подход реализует концепцию неинвазивного анализа, где хемометрические алгоритмы обрабатывают спектральные «отпечатки пальцев» вин, позволяя классифицировать вина по сорту винограда, году урожая, региону производства. Это обеспечивает эффективную аутентификацию продукции и выявление фальсификатов. Разработанные модели классификации позволяют определять географическое происхождение вин с точностью до 95-98 %. Наряду с классификацией возможно использование методов регрессии и построение градуировочных моделей для количественного определения основных компонентов с точностью 0,1-0,2 % для этанола и 0,1-0,5 г/л для органических кислот и сахаров.

Сортовые особенности винограда создают специфические наборы полифенолов и ароматических соединений, формирующих уникальные спектральные характеристики вин. При разработке моделей МО учитывается сортовая специфичность, и применяются иерархические методы классификации для повышения точности анализа.

В результате применения этих методов проводится количественный анализ компонентов вин и их классификация по различным параметрам с высокой точностью. Регулярная валидация моделей с использованием независимых наборов образцов обеспечивает надежность результатов.

При анализе спектральных данных вин методы хемометрики позволяют решать следующие задачи:

- определение сортового состава [62];
- выявление года урожая [63];
- классификация по региону происхождения;

- оценка возраста выдержки.

Помимо вин, методы оптической спектроскопии применяются для анализа и других алкогольных напитков, в том числе пива, виски и коньячной продукции (бренди и коньяков).

Например, исследователи [64] применили хемометрическую обработку данных спектрального происхождения при анализе пива. Особенности, которые необходимо учитывать для этого.

- влияние процесса ферментации на спектральные характеристики;
- необходимость учета мутности образцов;
- зависимость спектров от типа используемого солода.

Особое внимание уделяется контролю процесса ферментации и определению качества исходного сырья. При анализе пива и контроле процесса ферментации применяется комбинация спектральных методов с алгоритмами МО. Основные спектральные методы включают спектроскопию в ближней инфракрасной области, спектроскопию КР и УФ-видимую спектроскопию.

В процессе ферментации контролируются ключевые параметры: содержание сахаров, этанола, аминокислот и других метаболитов. Спектральные данные регистрируются в режиме реального времени, что позволяет отслеживать динамику изменения состава. Для анализа качества исходного сырья оценивается содержание крахмала, белков и других компонентов в солоде и хмеле. В описанном процессе контроля ферментации используется многопараметрический регрессионный анализ временных рядов, интегрированный с алгоритмами многомерной градуировки. Система применяет PLS-регрессию для создания робастных прогностических моделей, связывающих спектральные характеристики с целевыми аналитическими параметрами.

Данный подход реализует концепцию процессной аналитической технологии, где хемометрические алгоритмы обрабатывают потоковые спектральные данные в реальном времени, обеспечивая непрерывный мониторинг биохимических трансформаций. Это позволяет с высокой точностью отслеживать

кинетику ферментативных процессов и оперативно корректировать технологические параметры.

Полученные результаты включают построение градуировочных моделей для количественного определения основных компонентов с точностью 0,1-0,5 % для этанола и 0,1-1,0 г/л для сахаров. Разработанные модели также позволяют прогнозировать конечные характеристики продукта на основе спектральных данных промежуточных стадий производства. Это дает возможность оптимизировать технологический процесс и обеспечивать стабильное качество готовой продукции.

Спектры КР и флуоресценции позволяют отслеживать изменения химического состава в процессе производства пива, включая образование и превращение различных органических соединений. Спектральные методы позволяют определять содержание этанола, остаточных сахаров, органических кислот и других соединений.

Для учета влияния мутности и взвешенных частиц применяются математические методы коррекции спектров, включая мультипликативную коррекцию рассеяния и стандартную нормальную вариацию. Это позволяет минимизировать влияние физических факторов на качество спектральных данных.

В результате применения этих методов определяются количественные содержания основных компонентов с точностью до 0,1-0,5 % для этанола и 0,1-1,0 г/л для сахаров и кислот. При этом важно проводить градуировку методик по стандартным образцам и периодически проверять их стабильность.

Особенности применения методов анализа различаются в зависимости от типа пива. Для светлых сортов наиболее информативными являются спектры флуоресценции в УФ области, тогда как для темных сортов более эффективно использование спектров КР в видимой области. Это связано с различным содержанием окрашенных соединений, в первую очередь продуктов, образующихся при обжарке солода [65].

Для виски особенно важным является определение возраста выдержки и типа используемых бочек, что может быть осуществлено на основе анализа спектров флуоресценции. При анализе виски методом флуоресцентной спектроскопии изучаются спектры возбуждения и эмиссии флуорофоров, которые экстрагируются из древесины бочек в процессе выдержки [66].

Основными флуоресцирующими соединениями являются фенольные компоненты, такие как танины, лигнины и их производные. Для обработки многомерных данных применяются хемометрические методы, включая параллельный факторный анализ и методы дискриминантного анализа [67].

Результаты показывают, что интенсивность флуоресценции и положение максимумов спектров коррелируют со возрастом выдержки виски. С увеличением времени выдержки наблюдается рост интенсивности определенных спектральных полос, что связано с накоплением компонентов древесины. Тип используемых бочек (например, из американского или европейского дуба) также влияет на спектральный профиль.

В работе [68] метод PLS-DA успешно применен для классификации бренди и винных дистиллятов с точностью более 95 %. Исследователи использовали PLS-DA для анализа спектральных данных. Этот метод позволяет одновременно проводить снижение размерности данных и строить классификационную модель. Целью работы было создание надежного аналитического метода для различения подлинных бренди от винных дистиллятов. Для этого были собраны спектральные данные различных образцов, которые затем использовались для построения и валидации модели PLS-DA. В результате исследования:

- достигнута высокая точность классификации – более 95 % как для обучающего, так и для тестового набора образцов;
- выявлены специфические спектральные области, наиболее информативные для различения типов напитков;
- построена робастная модель, способная корректно классифицировать новые образцы;

- определены характерные спектральные маркеры, позволяющие идентифицировать каждый тип напитка;
- подтверждена воспроизводимость результатов на независимой выборке образцов.

Разработанный метод продемонстрировал высокую эффективность для задач контроля подлинности бренди и может быть использован в практике аналитических лабораторий.

Авторы работы [69] продемонстрировали возможность применения спектров КР для определения торговой марки и географического происхождения фруктовых дистиллятов. Географическое происхождение фруктовых дистиллятов по спектрам КР, обработанным методом k-ближайших соседей, определено с достоверностью порядка 90,9 %.

Также спектральный анализ с хемометрической обработкой данных используется для анализа атмосферного воздуха и при экологическом контроле. Например, исследователи [70] применили спектроскопию КР для классификации частиц черного углерода в атмосферном воздухе по источникам выбросов:

- при этом учитывались структурные особенности углеродных частиц;
- влияние остаточных органических примесей на спектральные характеристики;
- зависимость спектров от распределения частиц по размерам.

Особое внимание уделяется определению степени графитизации и выявлению структурных дефектов. При исследовании образцов применялась комбинация КР-спектрометра и конфокального микроскопа LEICA. Полученные данные обрабатывались с помощью методов хемометрики для построения моделей и классификации образцов по структурным и физико-химическим характеристикам. Различные источники черного углерода, включая золу, биомассу и выбросы транспортных средств, а также различные образцы черного углерода со станций мониторинга качества воздуха, были проанализированы с помощью КР-спектроскопии. Образцы черного углерода, используемые в исследовании, были собраны из двух разных мест: городской среды (Турин,

Италия) и альпийской долины (Улькс, Италия). Все полученные спектры, содержащие характерные полосы G и D, аппроксимировались функцией с пяти параметрами. Наконец, тот же алгоритм, обученный на данных источников выбросов черного углерода, пытался связать каждый спектр с источником черного углерода. В частности, во всех рассматриваемых образцах обнаружено большое количество черного углерода из выхлопных газов дизельных двигателей автомобилей.

Также перспективным является применение методов МО и хемометрики в фармацевтическом анализе. Современные подходы к анализу лекарственных препаратов требуют высокоточных, быстрых и надежных методик, способных идентифицировать активные фармацевтические ингредиенты, вспомогательные вещества и возможные примеси. Спектральные методы в сочетании с алгоритмами МО позволяют проводить количественный анализ сложных фармацевтических композиций без необходимости их предварительного разделения, что значительно сокращает время анализа и повышает производительность лабораторий. Методы хемометрического анализа успешно применяются для контроля качества фармацевтического сырья, промежуточных продуктов синтеза и готовых лекарственных форм. В частности, они эффективны для выявления фальсифицированных лекарственных средств, что имеет огромное значение для обеспечения безопасности пациентов и борьбы с нелегальным оборотом фармацевтической продукции [71]. Наряду с фармацевтикой, методы оптической спектроскопии в сочетании с последующей обработкой данных методами хемометрики и МО применяется в клинической диагностике разных заболеваний.

Флуоресцентная спектроскопия в сочетании с алгоритмом МО случайного леса представляет неинвазивный подход для диагностики глюкозурии – состояния, характеризующегося избытком сахара в моче пациентов с диабетом. В этом исследовании изучалась способность данного метода различать образцы мочи диабетиков и здоровых контрольных групп. Спектры флуоресценции были получены из образцов мочи с использованием ксеноновой дуговой лампы,

излучающей свет в диапазоне длин волн от 200 до 950 нм, при этом стабильная флуоресцентная эмиссия наблюдалась при 450 нм при длине волны возбуждения 370 нм. Образцы здоровых людей анализировались в том же спектральном диапазоне для сравнения. Для выявления спектральных различий между здоровыми и инфицированными образцами были применены алгоритмы МО случайного леса и k-ближайших соседей. Эти алгоритмы автоматически распознают спектральные паттерны, связанные с диабетом, что позволяет прогнозировать неизвестные классификации на основе установленных образцов. PCA использовался для снижения размерности данных перед их подачей в алгоритмы случайного леса и k-ближайших соседей для классификации. Эффективность классификации модели оценивалась с помощью 10-кратной перекрестной проверки, в результате чего предложенная модель на основе случайного леса достигла точности 96 %. Этот результат позволяет предположить, что предлагаемый метод является перспективным для удобной и точной диагностики глюкозурии у пациентов с диабетом [72].

Коронавирус (SARS-CoV-2), который был впервые обнаружен в конце 2019 года и быстро распространился по многим странам мира за короткий период времени, обладал высокой контагиозностью и представлял значительную угрозу для глобальной общественной безопасности. Быстрое и эффективное определение того, инфицирован ли человек коронавирусом, являлось важным шагом в области здравоохранения. Авторы работы представили спектры КР с предварительной обработкой методом нормализации и сглаживающим фильтром. Предложенная модель обнаружения коронавируса на основе алгоритма ЭГБ и КР-спектров осуществляет классификацию с точностью 94 % [73].

Таким образом, применение методов МО и хемометрики для анализа оптических спектров приобретает все большее значение. Для их применения необходимо специализированное ПО. В настоящее время на рынке хемометрического ПО существует несколько типов решений. Универсальные пакеты, такие как PLS Toolbox [74] и CHEMFLOW [75], требуют от пользователей

навыков программирования. Другая программа, SOLO, ориентирована на специалистов с глубокими знаниями в области хемометрики [76].

SOLO, разработанный компанией Eigenvector Research, Inc. (США), представляет собой комплексное программное решение для хемометрического анализа. Этот коммерческий продукт отличается продуманным графическим интерфейсом и широким набором встроенных инструментов для анализа данных. SOLO особенно эффективен в руках специалистов, обладающих глубокими знаниями в области хемометрики, поскольку позволяет использовать сложные алгоритмы анализа без необходимости их программирования.

В свою очередь, CHEMFLOW представляет собой открытую платформу, предлагающую более гибкий подход к анализу данных. Основанная на веб-интерфейсе, эта система позволяет создавать собственные аналитические процессы и организовывать совместную работу исследователей. CHEMFLOW особенно привлекателен для организаций, где требуется как глубокое понимание хемометрики, так и навыки программирования.

Преимущества SOLO:

- интуитивно понятный графический интерфейс;
- готовые инструменты для различных видов анализа;
- надежная техническая поддержка;
- регулярные обновления и улучшения;
- обширная документация и учебные материалы.

Преимущества CHEMFLOW:

- открытый исходный код;
- гибкость в создании аналитических процессов;
- отсутствие лицензионных ограничений;
- возможность совместной работы через веб-интерфейс;
- возможность интеграции собственных алгоритмов.

Выбор между этими программными продуктами зависит от конкретных потребностей лаборатории, уровня подготовки специалистов и финансовых возможностей организации. SOLO может быть оптимальным выбором для

лабораторий, готовых инвестировать в готовое решение и имеющих специалистов в области хемометрики. CHEMFLOW больше подойдет исследовательским группам, заинтересованным в создании собственных методов анализа и готовым вкладывать ресурсы в развитие программистских компетенций.

PLS Toolbox представляет собой профессиональный программный пакет для хемометрического анализа данных, интегрированный в MATLAB. Пакет включает инструменты для предварительной обработки данных (фильтрация шумов, коррекция базовой линии, нормализация спектров) и различные методы анализа, такие как метод главных компонент, проекция на латентные структуры, многомерное разрешение кривых. Программа предоставляет инструменты для валидации результатов (кросс-валидация, бутстрэп-анализ) и разнообразные возможности визуализации через 2D и 3D графики. Благодаря интеграции с MATLAB пакет позволяет создавать автоматизированные процедуры анализа, что особенно ценно для лабораторий с большим потоком рутинных исследований. Однако для эффективного использования PLS Toolbox необходимы навыки программирования в среде MATLAB.

Существуют также узкоспециализированные решения, предназначенные для работы с конкретными типами спектральных данных, такие как программа BWIQ (ООО «Промэнерголаб») [77].

Проведенный анализ показывает, что выбор конкретного метода должен осуществляться с учетом многих факторов, включая объем доступных данных, требования к точности, необходимость интерпретации результатов и доступные вычислительные ресурсы. При этом важно обеспечить возможность валидации и стандартизации выбранного метода для его применения в рутинном анализе.

Особую ценность представляет возможность комбинирования различных спектральных методов и алгоритмов анализа данных. Современные методы МО, в частности ЭГБ, метод случайного леса и т.д., демонстрируют высокую эффективность для решения задач классификации и регрессии по различным характеристикам. Это позволяет не только повысить надежность классификации и

регрессии, но и получать взаимодополняющую информацию о составе и свойствах анализируемых образцов.

В целом, рассмотренные методы хеометрики и МО представляют собой мощный инструментальный метод для анализа многокомпонентных смесей, позволяющий решать широкий спектр практических задач в области контроля качества пищевой продукции, экологического мониторинга, анализа лекарств, выявления фальсификатов и лабораторной диагностики болезней. Дальнейшее развитие этих методов и их внедрение в практику аналитических лабораторий будет способствовать повышению эффективности.

Выводы к главе 1

Обработка спектральных данных на основе алгоритмов хеометрики и машинного обучения существенно расширяет аналитические возможности методов оптической спектроскопии, обеспечивая высокую селективность, чувствительность и производительность анализов без усложнения аппаратуры и простой пробоподготовки. Такой подход является весьма перспективным при использовании абсорбционной спектроскопии, флуоресцентной спектроскопии и КР-спектроскопии для анализа многокомпонентных жидких сред и аэрозолей.

Исходя из аналитического обзора литературы для корректного решения **подобных** задачи необходимы:

- Выбор вида (видов) оптических спектров и их информативных фрагментов;
- Подготовка образцов и проведение измерений оптических спектров;
- Предварительная обработка данных;
- Формирование обучающего и тестового наборов;
- Выбор алгоритмов машинного обучения, оптимальных для данной задачи;
- Построение модели машинного обучения и ее кросс-валидация;
- Валидация модели на тестовом наборе.

Стадии подготовки к анализу существенно различаются для разных аналитических задач. К настоящему времени, несмотря на перспективность такого подхода, методики (методы) на его основе разработаны лишь для относительно небольшого числа объектов анализа. Так же решение данных задач затрудняется, из-за ограниченности выбора специализированного ПО для обработки спектральных данных. Существующие программные пакеты либо являются достаточно сложными, требующими хорошего знания хемометрики, а в ряде случаев и программирования, либо ориентированы на обработку определенного вида спектров.

ГЛАВА 2. ВЫБОР И ОБОСНОВАНИЕ АЛГОРИТМОВ ДЛЯ ОБРАБОТКИ СПЕКТРАЛЬНЫХ ДАННЫХ. РАЗРАБОТКА ПРОГРАММНОГО ПАКЕТА

Разработанная программа «ИнтелСпектр» представляет собой специализированное ПО, реализующее современные алгоритмы хемометрики и МО для обработки спектральных данных.

«ИнтелСпектр» занимает промежуточное положение между перечисленными в главе 1.3 категориями ПО, предлагая удобный инструментарий для специалистов аналитических лабораторий, не требующий глубоких знаний в программировании или хемометрике.

Программа «ИнтелСпектр» разработана для работы в современных операционных системах семейства Windows. Для успешного функционирования программы требуется операционная система Windows версий 8, 10 или 11, установленная на жесткий диск персонального компьютера. ПО написано на языке Python с использованием современных библиотек для обработки данных и МО, включая `scikit-learn`, `xgboost`, `numpy` и `pandas` [78-81].

Программа «ИнтелСпектр» поддерживает работу с различными форматами данных, что обеспечивается использованием специализированных библиотек, таких как `pandas` для обработки табличных данных и `matplotlib` для визуализации результатов.

Входные данные программы включают:

- спектральные данные в формате `.txt`;
- информацию о классах или концентрациях для обучающих наборов;
- файлы с параметрами ранее обученных моделей.

Выходные данные включают матрицы соответствия, которые сохраняются в формате `.xlsx`, что обеспечивает совместимость с Microsoft Excel и другими табличными процессорами;

Для кросс-валидации в случае классификации матрицы соответствия имеют следующий вид. В первом столбце указывается наименование и исходный

класс образцов, в последующих столбцах указываются доверительные вероятности отнесения к классам. В случае тестовой валидации исходные классы неизвестны, и исходный класс не отображаются.

Для регрессии матрицы соответствия имеют следующий вид. В первом столбце указывается наименование образцов, во втором – исходная концентрация, указанная при формировании данных, в третьем – предсказанная концентрация. Для многомерной градуировки и валидации такой модели в последующих столбцах ситуация аналогична.

Градуировочные графики и другие визуальные результаты сохраняются в формате .png.

При обработке спектральных данных реализуются четыре стадии функционирования ПО.

2.1 Формирование данных

Начальной стадией работы с программой «ИнтелСпектр» является формирование массива входных данных, необходимых для построения и тестирования моделей МО. Формирование данных осуществляется в два этапа.

Первый этап включает создание обучающего набора. В интерактивном окне оператор выбирает данные, которые будут использоваться для обучения модели. Для каждого образца задаются:

- предикторы (независимые переменные) – значения спектральных данных. В случае спектрального анализа предикторами выступают значения оптической плотности или интенсивности излучения;
- отклики (зависимые переменные), которые различаются в зависимости от типа решаемой задачи: для задач классификации это название и номер класса, к которому относится образец, а для задач регрессии – названия определяемых компонентов и их концентрации в образце.

Второй этап включает формирование тестового набора. Для каждого образца, предназначенного для включения в тестовый набор, задаются спектральные данные.

Важно отметить, что образцы тестового набора не должны использоваться при обучении модели для обеспечения объективной оценки ее предсказательной способности.

2.2 Предварительная обработка спектров

Цель предварительной обработки – повышение информативности спектров путем устранения составляющих, не содержащих полезной информации об анализируемых веществах. В рамках предварительной обработки программа «ИнтелСпектр» поддерживает следующие процедуры:

- вычитание базовой линии;
- сглаживание с помощью фильтра Савицкого-Голея;
- центрирование;
- нормировка;
- масштабирование.

Эти процедуры обсуждаются в разделе 1.3.

В зависимости от особенностей задачи пользователь должен выбрать процедуры из этого набора, уточнить их параметры (например, размер кадра и степень полинома в фильтре Савицкого-Голея). На рисунках 2 и 3 в качестве примера продемонстрированы спектры КР для одного из образца. На рисунке 2 продемонстрирован спектр без фильтрации данных фильтром Савицкого-Голея, на рисунке 3 представлен спектр того же образца с фильтрацией данных фильтром Савицкого-Голея с размером кадра 75 и полиномом 2.

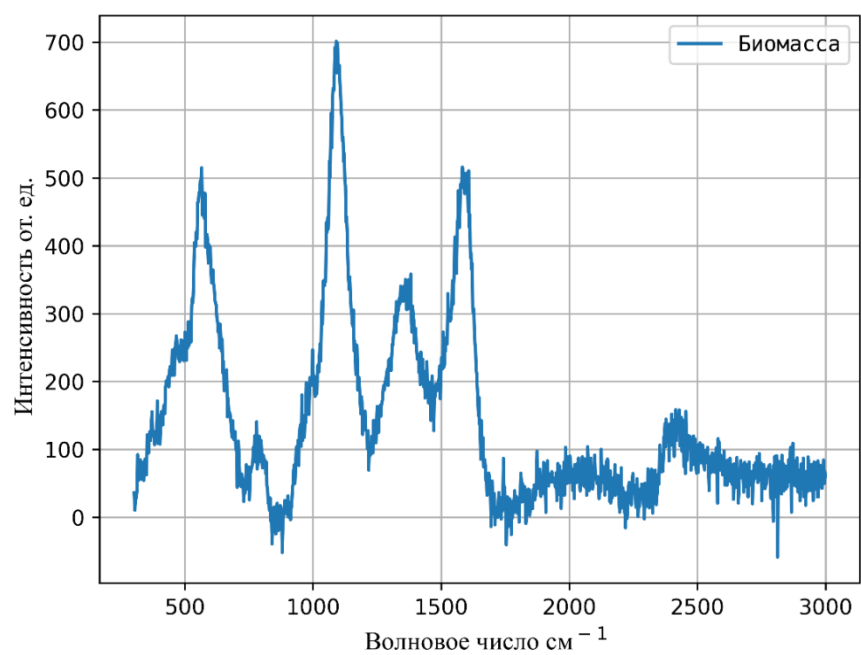


Рисунок 2 – Спектр без фильтрации

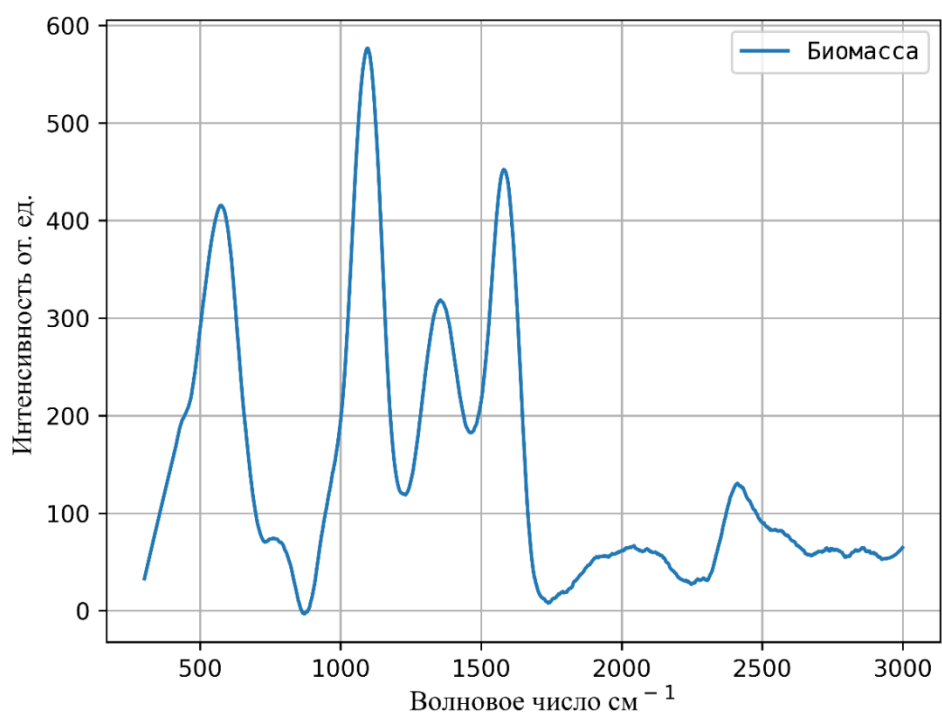


Рисунок 3 – Спектр с фильтрацией

2.3 Выбор алгоритмов

В таблице 1 представлены все алгоритмы МО и хемометрики, которые были реализованы в ПО, а также краткое упоминание преимущества алгоритмов, которые служили в качестве отличительных особенностей для реализации алгоритмов в ПО. Все преимущества алгоритмов более подробно были представлены в разделе 1.3. Исходя из всего вышесказанного, выбор того или иного алгоритма обработки данных обосновывается специфическими возможностями каждого из них. Отличительные особенности алгоритмов приведены в таблице 1.

Таблица 1 – Реализованные алгоритмы МО и хемометрики

Наименование алгоритма	Решаемые задачи		Отличительные особенности
	Классификация	Регрессия	
Экстремальный градиентный бустинг	+	+	Эффективность при ограниченном количестве обучающих образцов
Линейный дискриминантный анализ	+	—	Высокая скорость обучения и прогнозирования
Квадратичный дискриминантный анализ	+	—	Учет различных корреляционных (ковариационных) функций классов
Опорные векторы	+	+	Устойчивость к шуму в спектральных данных, хорошая обобщающая способность (высокая точность и устойчивость к переобучению)

Ближайшие соседи	+	+	Различные метрики расстояния между классами
Многослойный перцептрон	+	—	Выявление скрытых закономерностей
Случайный лес	+	+	Способность работать с нелинейными зависимостями между исходными и предсказываемыми параметрами
Анализ главных компонент	+	—	Снижение размерности данных, возможность их визуализации
Метод частичных наименьших квадратов	—	+	Устранение мультиколлинеарности (лучше, чем в PCR), высокая прогностическая способность
Регрессия на главных компонентах	—	+	Устранение мультиколлинеарности, устойчивость модели по отношению к шумам

2.4 Обучение и валидация (тестирование) модели

Построение модели проводится по фрагменту спектров после предварительной обработки. Выбирается наиболее информативный фрагмент (фрагменты) спектров, одинаковый в рамках одной задачи для всех образцов. Выбранные фрагменты спектров существенно различаются между собой для

каждого образца для корректного разделения на классы (в задачах классификации) или для разных концентраций компонентов (в задачах регрессии).

Для построения модели необходимо выбрать один из алгоритмов классификации или регрессии, после этого вводится команда на обучение модели и формируется обучающий набор из имеющихся фрагмента спектров. После этого необходимо провести перекрестную или кросс-валидацию модели, на спектрах образцов, которые были использованы при ее построении. Кросс-валидация проводится методом Leave One Out. По окончании перекрестной проверки строится матрица соответствия для классификации, а для регрессии выводится градуировочный график и таблица с исходными и предсказанными данными.

Затем проводится валидация на спектрах образцов тестового набора, которые не использовались при обучении модели. Предсказания должны сопоставляться с исходными данными.

Показателями надежности модели в случае классификации является процент достоверных предсказаний классов для образцов тестового набора, доверительная вероятность и минимальная доверительная вероятность.

В случае регрессии – среднеквадратичная разность концентраций между номинальными и предсказанными значениями концентраций для образцов тестового набора и квадрат коэффициента линейной корреляции (коэффициент детерминации) между этими значениями (R^2) [82]

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \text{SSR/SST}. \quad (8)$$

Формула коэффициента детерминации (R^2) состоит из двух основных компонентов:

где: SSR (Sum Squared Regression) – сумма квадратов регрессионных остатков: $\sum (y_i - \hat{y}_i)^2$; y_i – фактическое значение; $\hat{y}_i$ – предсказанное значение модели, эта разница показывает ошибку предсказания модели для каждого наблюдения; SST (Total Sum of Squares) – общая сумма квадратов: $\sum (y_i - \bar{y})^2$; y_i – фактическое

значение; $\bar{y}$ – среднее значение всех наблюдений, эта разница показывает отклонение каждого наблюдения от среднего значения.

Отношение SSR/SST показывает долю необъясненной дисперсии. Вычитая это отношение из 1, мы получаем долю дисперсии, которую объясняет модель. Чем ближе значение R^2 к 1, тем лучше модель описывает данные.

Эти показатели надежности и матрицы соответствия выводятся по завершении процедуры валидации.

Обученной и прошедшей валидацию модели предъявляются спектры образцов, подлежащих анализу. Результаты анализа в виде номера класса или концентраций компонентов в образце выводятся и сохраняются в электронном журнале. На рисунках 4 а и 4 б приведены примеры пользовательского окна для классификации и регрессии.

Назад

Выбор метода

- ☒ Экстремальный градиентный бустинг
- ☐ Линейный дискриминантный анализ
- ☐ Квадратичный дискриминантный анализ
- ☐ Метод опорных векторов
- ☐ Метод ближайших соседей
- ☐ Метод многослойного перцептрона
- ☐ Метод случайного леса
- ☐ Метод главных компонент

Предварительная обработка

- ☐ Фильтр Савицкого - Голца
- ☐ Нормализация данных
- ☐ Стандартизация данных (Scaling)
- ☐ Стандартный масштабатор (StandardScaler)
- ☐ Удаление базовой линии
- ☐ Масштабирование
- ☐ Центрирование

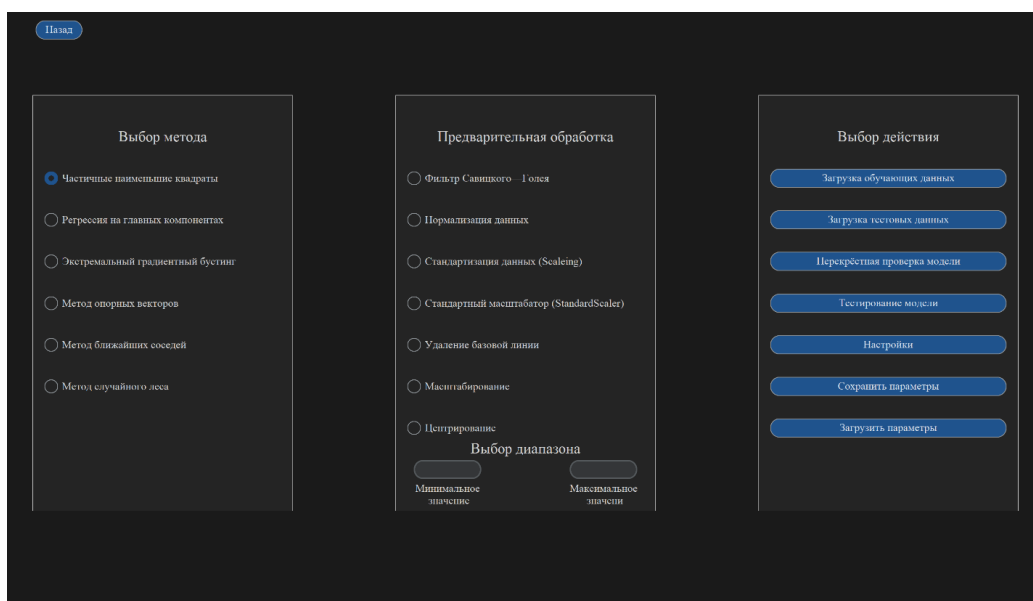
Выбор диапазона

Минимальное значение Максимальное значение

Выбор действия

- Загрузка обучающих данных
- Загрузка тестовых данных
- Перекрестная проверка модели
- Тестирование модели
- Настройки
- Сохранить параметры
- Загрузить параметры

а)



б)

а – пользовательское окно в режиме классификации; б – пользовательское окно в режиме регрессии

Рисунок 4 – Пользовательские окна

Сравнение ПО «ИнтелСпектр» с существующими решениями, такими как SOLO и CHEMFLOW, рассмотренными в разделе 1.3, позволяют выявить уникальные особенности разработанного ПО. В отличие от рассмотренных аналогов, «ИнтелСпектр» не требует от пользователя глубоких знаний в области программирования или хемометрики, что делает его доступным для широкого круга специалистов аналитических лабораторий.

Представленное ПО успешно решает задачи как качественного, так и количественного анализа с использованием спектральных данных, обеспечивая возможность сопоставления результатов и оценки их достоверности. Это делает его эффективным инструментом для практического применения в различных областях аналитической химии.

Разработанный программный пакет успешно реализует основные методы хемометрики и МО для обработки спектральных данных. Важными особенностями разработанной программы являются:

— возможность работы как в автономном режиме, так и в качестве встроенного модуля ПО аналитических спектральных приборов;

— успешная апробация на практических задачах, таких как качественный и количественный анализ объектов по спектрам КР и спектрам флуоресценции; в частности, ПО было использовано в исследованиях, проводившихся в ФГБУ «ВНИИОФИ» по хоздоговорной НИОКР, что подтверждено актом внедрения (Приложение 1).

Выводы к главе 2

В данной главе описан разработанный в рамках диссертационного исследования прикладное ПО «ИнтелСпектр» для обработки спектральных данных методами хеометрики и МО, которое представляет собой промежуточное решение между сложными универсальными пакетами и узкоспециализированным программным обеспечением.

Рассмотрены ключевые алгоритмы обработки данных, включая алгоритмы классификации и регрессии.

Проведено сравнение с существующими решениями, такими как SOLO и CNEMFLOW, что позволило выявить уникальные особенности и преимущества разработанного программного обеспечения.

В отличие от аналогов, ПО "ИнтелСпектр" легко в использовании, мобильно, требует относительно небольшого свободного пространства памяти твердотельного накопителя (около 600 Мегабайт). Пакет "ИнтелСпектр" может быть встроен в состав ПО аналитических приборов (в частности флуоресцентных спектрометров, КР-спектрометров, абсорбционных спектрометров), не требует от пользователя глубоких знаний в области программирования или хеометрики, что делает его доступным для широкого круга специалистов аналитических лабораторий.

Использование данного ПО при выполнении НИОКР в ФГБУ «ВНИИОФИ» подтверждено актом внедрения (Приложение 1).

ГЛАВА 3. ИССЛЕДОВАНИЕ ПРОИСХОЖДЕНИЯ АЭРОЗОЛЬНЫХ ЧАСТИЦ ЧЕРНОГО УГЛЕРОДА

Частицы черного углерода являются опасными загрязнителями атмосферы, вызывающими негативные изменения климата и наносящие вред здоровью людей. Необходим их постоянный контроль в воздухе и атмосферных осадках [83]. В Федеральной научно-технической программе в области экологического развития Российской Федерации и климатических изменений на 2021-2030 г.г., утвержденной постановлением Правительства РФ № 133 от 8 февраля 2022 г., предусматривается разработка системы мониторинга черного углерода в атмосферном воздухе на территории России.

Особую важность представляет определение происхождения частиц – необходимо различать частицы, появившиеся в результате лесных пожаров (биомасса) или сгорания углеводородов (автомобильное топливо, керосин).

Исследовалась возможность проведения подобной классификации частиц черного углерода по оптическим спектрам КР и абсорбции с обработкой данных методами МО.

С этой целью были проанализированы две партии образцов, одна из которых формировалась сбором в лабораторных условиях продуктов сгорания различных веществ, а вторая – сбором частиц путем продувания атмосферного воздуха через стекловолоконную мембрану.

3.1 Классификация частиц черного углерода по спектрам КР

Для классификации по спектрам КР использовались образцы, приготовленные в лабораторных условиях путем сбора на стеклах продуктов неполного сгорания различных объектов органического происхождения. Для проведения исследования сжигали образцы биомассы, керосина и пластика, продукты сгорания которых были осаждены на стекла. По происхождению частиц черного углерода выделены четыре группы:

- продукты сгорания биомассы;
- продукты сжигания керосина;

- продукты сжигания пластика ПЭТ;
- частицы технического углерода (ООО «Технотех», г. Йошкар Ола).

Измерения лабораторных образцов проводились на портативном раман-люминесцентном спектрометре «ИнСпектр R532» с длиной волны возбуждения 532 нм (производство ООО «Спектр-М», Россия, регистрационный номер в Федеральном информационном фонде по обеспечению единства измерений (далее – ФИФ ОЕИ) 69110-17). Каждый образец измерялся десять раз в разных точках, неравномерное оседание углерода позволило получить десять различных спектров из одного образца. В таблице 2 представлены метрологические характеристики прибора [84].

Таблица 2 – Метрологические и основные технические характеристики раман-люминесцентного портативного спектрометра «ИнСпектр» R532

Наименование характеристики	Значение характеристики		
	R532	Raport	R1064
Спектральный диапазон измерений волновых чисел, см^{-1}	от 380 до 3000		
Спектральное разрешение, см^{-1} , не более	25		
Пределы допускаемого значения относительного среднего квадратического отклонения (ОСКО) измерений волновых чисел, %, не более	± 1		
Пределы допускаемой относительной погрешности измерений волновых чисел, %, не более	± 1		
Наименование технической характеристики	Значение характеристики		
	R532	Raport	R1064
Спектральный диапазон показаний волновых чисел, см^{-1}	от 140 до 4000		от 380 до 3600
Длина волны лазера, нм	532		1064
Мощность излучения, мВт	от 20 до 30		до 400
Фокусное расстояние спектрометра, мм	75	50	75
Входная щель, мкм	20		30
Число штрихов голографической	1200 (1800*)	1800	300

дифракционной решетки, штрих/мм			
Условия эксплуатации: - температура окружающей среды, °С - относительная влажность, %, не более - атмосферное давление, мм рт. ст.	от +15 до +35 80 от 720 до 790		
Электропитание: а) от сети переменного тока - напряжением, В - частотой, Гц	от 100 до 240 от 50 до 60		
б) от аккумулятора	-	+	-
Габаритные размеры (длина×ширина×высота), мм, не более	230×146×60	240×230×160	230×171×60
Масса, кг, не более	1,6	2,1	1,7

Наибольшие отличия между спектрами КР образцов разных групп были выявлены в диапазоне волновых чисел от 750 до 2500 см^{-1} . Эти фрагменты спектров проходили предварительную обработку (фильтром Савицкого-Голея с размером кадра 75 и полиномом 2), далее строились модели хемометрики и МО (рисунок 5).

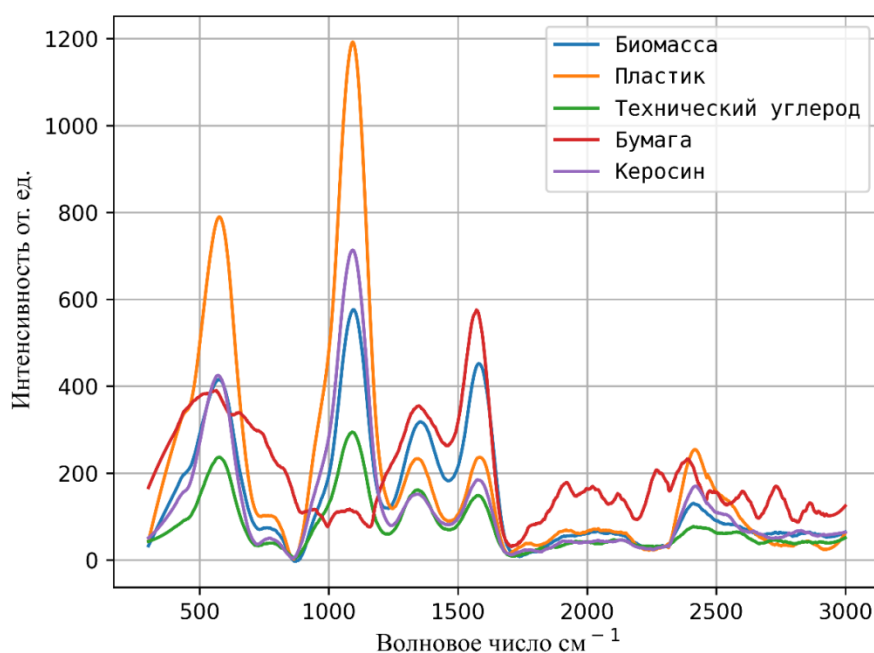


Рисунок 5 – Спектры образцов в интервале 300-3000 см^{-1}

Для разделения данных на химически осмысленные составляющие была проведена их визуализация с использованием алгоритма РСА. Полученные результаты показывают, что проекции спектров КР биомассы в плоскости главных компонент разделяются на две компактно расположенные группы (рисунок 6). Поэтому спектры продуктов сгорания биомассы целесообразно разделить на два различных класса.

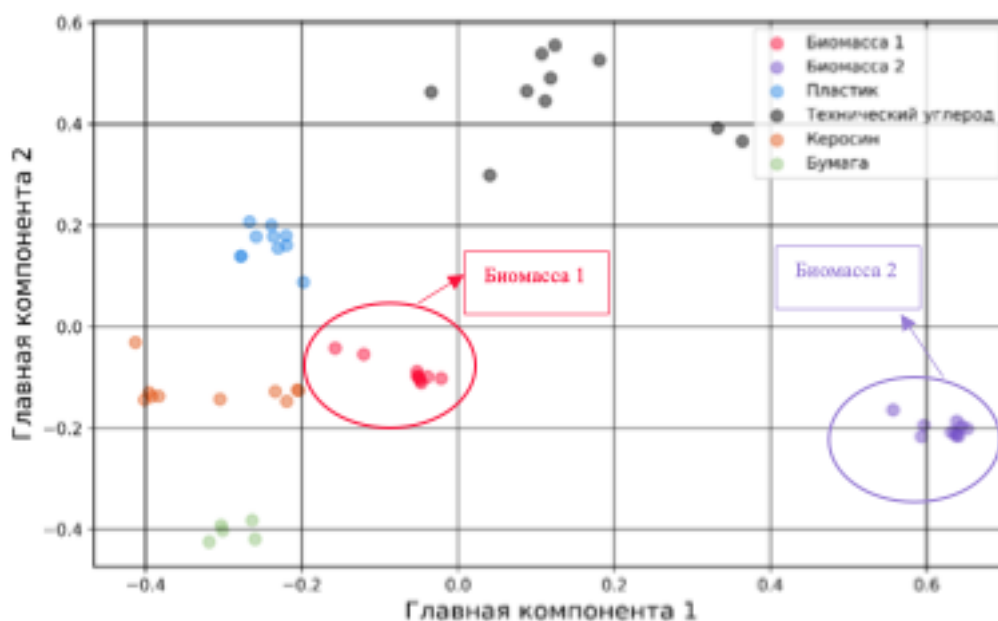


Рисунок 6 – Проекция спектров КР в плоскости главных компонент

Отмеченные различия в спектрах можно объяснить тем, что образцы биомассы получались путем сжигания в обожженной металлической бочке ветвей и стволов древесной растительности, преимущественно без листвы. Разные виды древесной растительности могут давать различающиеся между собой спектры, поэтому каждому классу может соответствовать свой вид сжигаемой растительности.

Важно отметить, что данная модель способна надежно распознавать продукты сгорания только тех типов биомассы, по которым производилось ее обучение. Нельзя исключать, что спектры продуктов сгорания принципиально других типов биомассы (например, травы) будут существенно отличаться от спектров указанных двух классов.

Спектры частиц технического углерода, полученного путем сжигания нефтепродуктов в специальных печах, продуктов сжигания пластика и продуктов сжигания керосина, выделены в самостоятельные классы. Предположительно, технический углерод образуется при более полном сгорании, поэтому он менее загрязнен органическими примесями по сравнению с другими источниками частиц, получаемыми при меньших температурах и (или) недостатке кислорода.

Таким образом, проведенная с использованием РСА визуализация спектральных данных, позволила выявить пять четко различающихся классов частиц черного углерода:

1. два класса продуктов сгорания биомассы;
2. частицы от сгорания керосина;
3. частицы технического углерода;
4. продукты сгорания пластика.

На основе такого разбиения была обучена модель классификации образцов частиц черного углерода. Для обучения модели сформирован набор из 10 информативных фрагментов КР-спектров каждого из пяти названных классов. Обучение производилось с использованием алгоритма ЭГБ.

Для оценки предсказательной способности (валидации) модели был сформирован тестовый набор из 10 спектров, полученных от тех же источников, но не использованных в процессе обучения. Дополнительно для проверки способности модели отличать посторонние образцы от образцов, относящихся к одному из заданных классов, в качестве «фальшь-объектов» были подготовлены образцы бумаги с зачерненной областью, при помощи лазерного принтера. В таблице 3 продемонстрированы результаты валидации по тестовому набору.

Таблица 3 – Результаты предсказания происхождения частиц черного углерода для тестового набора по КР-спектрам

Наименование образцов и исходный класс	Доверительная вероятность предсказания отнесения к классу				
	Биомасса-1	Биомасса-2	Технический углерод	Пластик	Керосин
Биомасса 1(1)	0,942	0,021	0,013	0,011	0,013
Биомасса 1(2)	0,925	0,021	0,013	0,028	0,013

Биомасса 2(1)	0,018	0,896	0,017	0,051	0,018
Биомасса 2(2)	0,016	0,904	0,016	0,047	0,016
Технический углерод (1)	0,015	0,019	0,939	0,012	0,015
Технический углерод (2)	0,015	0,017	0,935	0,018	0,015
Пластик (1)	0,016	0,013	0,017	0,939	0,016
Пластик (2)	0,038	0,030	0,038	0,856	0,038
Керосин (1)	0,014	0,013	0,014	0,011	0,948
Керосин (1)	0,014	0,013	0,014	0,011	0,948
Бумага (1)	0,007	0,009	0,458	0,021	0,504
Бумага (2)	0,007	0,010	0,465	0,006	0,512
Бумага (3)	0,007	0,010	0,465	0,006	0,512
Бумага (4)	0,008	0,010	0,465	0,006	0,512
Бумага (5)	0,008	0,010	0,465	0,006	0,512
Примечание. Зеленым цветом обозначены достоверные предсказания, а красным – ложные					

Результаты валидации показали, что для всех образцов частиц черного углерода доверительная вероятность достоверного распознавания класса составляет более 85 %, а вероятность любого из ошибочных распознаваний не превышает 4 %. При выбранном пороговом значении доверительной вероятности в 75 %, установленным исходя из многочисленных измерений и подбора критериев. Все 10 образцов тестового набора были достоверно классифицированы. При этом посторонние образцы (бумага) модель не относил ни к одному из пяти классов, так как наибольшее значение доверительной вероятности (51,2 %) оказалось существенно ниже порогового. Полученные результаты были использованы в работе [85] «Доклад о прогрессе развития сети мониторинга черного углерода в России в 2022-2023 гг. и перспективных планах», что подтверждается актом внедрения от ФГБУ «ИГКЭ» (Приложение 1).

3.2 Классификация частиц черного углерода по спектрам поглощения

Пробы атмосферного воздуха отбирались с помощью установки, состоящей из импактора, ротаметра со счетчиком газа и мембранного насоса, на фильтрующий материал из кварцевого стекловолокна фирмы Whatman с диаметром пор 0,2 мкм.

Отбор пробы воздуха осуществлялся на скорости не более 5 л/мин. Время отбора пробы не менее 20 минут. Расстояние до источника черного углерода не менее 0,5 метра. Каждый отбор проб выполнялся в двух повторах.

Измерения спектров поглощения мембран, на которых накапливались частицы черного углерода при прокачке атмосферного воздуха, проводились на двухлучевом спектрофотометре Lambda 950 (фирма Perkin-Elmer, регистрационный номер в ФИФ ОЕИ 46152-10), имеющем регистрационный номер 15947-04 в Государственном реестре средств измерений РФ. Прибор обеспечивает измерения в широком спектральном диапазоне (175-3300 нм) с высокой точностью (предел допускаемой абсолютной погрешности по фотометрической шкале $\pm 0,003$ Б при $D=1$ Б) и позволяет проводить измерения оптической плотности в диапазоне ± 4 Б. В таблице 4 приведены метрологические характеристики спектрофотометра Lambda 950 [86].

Таблица 4 – Метрологические характеристики спектрофотометра Lambda 950

Наименование характеристики	Значение характеристики
Спектральный диапазон, нм	175-3300
Пределы допускаемой абсолютной погрешности установки длины волны, нм, не более	
– для диапазона 200-780 нм	+0,0
– для диапазона 780-900 нм	+0,30
Воспроизводимость установки длины волны, нм, не более	
– для диапазона 200-780 нм	+0,005
– для диапазона 780-900 нм	+0,02
Пределы допускаемой абсолютной погрешности измерения оптической плотности, Б (по фильтру с оптической плотностью 1 Б)	+0,003

Измерялась оптическая плотность продуктов горения керосина, древесины и биомассы (сена).

Методика проведения измерений была разработана с учетом специфики исследуемых образцов. В опорный канал спектрофотометра помещалась чистая мембрана, служившая в качестве референтного образца, а в измерительный канал – мембрана с пятнами частиц черного углерода. Особое внимание уделялось позиционированию образца: мембрана в измерительном канале тщательно диафрагмировалась для обеспечения максимально полного прохождения светового пучка через область, содержащую частицы черного углерода.

Всего было исследовано 74 образца. В первой партии было 30 образцов, из которых для дальнейшего исследования были отобраны 22. Основным критерием отбора служил размер пятен с частицами черного углерода – они должны были быть достаточно большими для проведения корректных измерений. На рисунке 7 представлен характерный пример образца с четко выраженными пятнами, пригодного для спектрофотометрических измерений.



Рисунок 7 – Мембраны с пятнами частиц черного углерода с четко выраженными границами

Для иллюстрации критериев отбраковки на рисунке 8 показан пример образца, не соответствующего требованиям измерений из-за малого размера пятен (менее 3 мм).

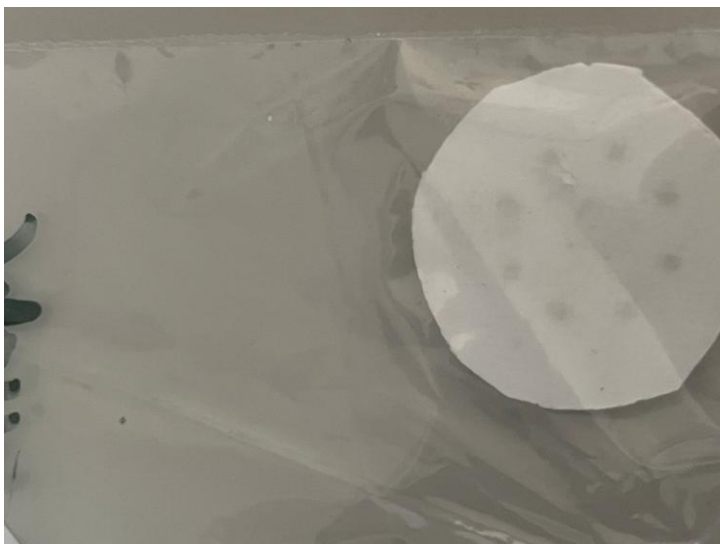


Рисунок 8 – Мембрана с нечеткими пятнами диаметром менее 3 мм

Измерения оптической плотности проводились для каждого образца на семи длинах волн (370, 470, 520, 590, 660, 880, 950 нм), которые соответствуют рабочим длинам волн разрабатываемого азталометра. Спектральная ширина щели при измерениях составляла 5 нм. Для повышения достоверности результатов на каждой длине волны проводилось по восемь измерений с последующим усреднением полученных значений.

Результаты измерений оптической плотности для всех исследованных образцов относительно чистой мембраны представлены в таблицах 5 и 6.

Таблица 5 – Результаты измерения оптической плотности стекловолоконных мембран с накопленными частицами черного углерода (первая партия)

Наименование образца	370 нм	470 нм	520 нм	590 нм	660 нм	880 нм	950 нм
К 1.1	1,2319	0,995	0,9221	0,841	0,7721	0,6091	0,5827
К 2.1	0,8033	0,5834	0,5201	0,4625	0,4196	0,4286	0,4307
К 2.2	0,4262	0,2982	0,2625	0,2299	0,2063	0,2642	0,2649
К 4.1	1,1809	0,8786	0,8111	0,7477	0,6955	0,3448	0,3399
К 4.2	0,6909	0,4981	0,4518	0,402	0,3645	0,373	0,3691
К 5.1	0,9607	0,6834	0,6001	0,5167	0,4564	0,4257	0,407
К 5.2	1,5467	1,1488	1,0355	0,9248	0,8394	0,6893	0,6489

Д 1.1	4,2285	2,3458	1,7504	1,0884	0,6253	0,1362	0,1264
Д 1.2	3,1215	1,4461	1,0406	0,6289	0,3532	0,1494	0,1339
Д 2.1	2,9198	1,32	1,014	0,7173	0,511	0,4438	0,3967
Д 2.2	1,5348	0,5265	0,3759	0,2502	0,1807	0,2704	0,2991
Д 3.2	1,7427	0,6461	0,4675	0,3142	0,2332	0,3682	0,4014
Д 4.1	2,3107	0,9964	0,7491	0,5162	0,3652	0,3995	0,3916
Д 4.2	1,9734	0,7741	0,573	0,3899	0,2876	0,4338	0,4493
Б 1.1	2,4195	1,2277	0,9081	0,6493	0,4931	0,3364	0,3264
Б 1.2	1,1998	0,5403	0,3993	0,2923	0,2351	0,4017	0,436
Б 2.1	3,0238	1,6924	1,2459	0,8601	0,6271	0,5192	0,5077
Б 2.2	3,9419	2,736	2,1461	1,5709	1,191	0,7634	0,7253
Б 3.1	4,7286	4,555	3,5597	2,5929	1,9721	1,5531	1,3273
Б 3.2	0,8762	0,2965	0,2038	0,1457	0,1197	0,2693	0,2308
Б 4.1	3,3177	1,9867	1,5573	1,1769	0,9324	0,6903	0,6809
Б 5.2	4,3158	3,3668	2,7048	2,0423	1,5969	1,0265	0,9758

Таблица 6 – Результаты измерения оптической плотности стекловолоконных мембран с накопленными частицами черного углерода (вторая партия)

Наименование образца	370 нм	470 нм	520 нм	590 нм	660 нм	880 нм	950 нм
К 1.1	2,776	2,262	2,160	2,049	1,955	1,707	1,680
К 3.1	4,273	3,499	3,357	3,212	3,087	2,949	2,613
К 4.3	1,842	1,597	1,504	1,390	1,291	1,081	1,062
К 5.2	3,565	3,505	3,397	3,266	3,142	2,800	9,253
К 5.1	1,561	0,687	0,543	0,407	0,326	0,236	0,219
К 2.1	3,402	2,852	2,745	2,649	2,568	2,441	2,281
Д 1.1	2,214	1,944	1,842	1,714	1,602	1,350	1,340
Д 1.2	4,907	2,655	2,323	2,013	1,786	1,449	1,332
Д 1.3	1,016	0,279	0,187	0,103	0,060	0,000	-0,009
Д 1.3	0,981	0,275	0,187	0,107	0,065	0,001	-0,010

Д 2.2	4,304	2,820	2,180	1,611	1,288	0,849	0,774
Д 3.1	1,174	0,354	0,232	0,126	0,071	0,002	-0,008
Д 3.2	4,294	3,183	2,539	1,821	1,387	0,746	0,623
Д 4.1	1,319	0,431	0,306	0,196	0,138	0,073	0,061
Д 4.2	6,496	3,425	2,946	2,343	1,942	1,319	1,176
Д 5.2	4,441	3,583	3,178	2,741	2,444	1,899	1,702
Д 6.1	3,398	1,769	1,436	1,185	1,042	0,896	0,835
Д 6.2	1,805	0,883	0,739	0,610	0,530	0,371	0,336
Д 6.3	1,429	0,652	0,543	0,449	0,389	0,269	0,243
Д 6.4	4,139	2,564	2,285	2,044	1,881	1,593	1,583
Д 7.1	3,608	2,264	1,548	0,919	0,600	0,322	0,276
Д 8.2	3,982	2,737	1,871	1,068	0,682	0,327	0,280
Д 9.1	3,443	1,996	1,397	0,835	0,517	0,181	0,135
Д 10.2	5,685	3,129	2,232	1,430	1,035	0,629	0,563
Д 11.2	3,913	2,498	2,051	1,582	1,279	0,901	0,850
Д 12.1	2,961	1,363	0,942	0,567	0,370	0,207	0,179
Д 12.1	3,110	1,454	1,008	0,608	0,397	0,225	0,195
Д 12.2	1,814	0,632	0,415	0,241	0,152	0,064	0,048
Д 12.3	1,359	0,404	0,275	0,179	0,129	0,064	0,053
Д 13.2	5,652	2,878	2,115	1,370	0,978	0,596	0,541
Д 1.1	0,135	0,139	0,626	1,090	1,752	2,346	4,009
Д 1.2	0,135	0,151	0,353	0,629	1,041	1,447	3,172
Д 3.2	0,398	0,369	0,233	0,315	0,468	0,647	1,742
Д 4.1	0,395	0,401	0,366	0,517	0,750	0,997	2,310
Д 4.2	0,473	0,435	0,288	0,391	0,573	0,774	1,974
Д 2.1	0,738	0,244	0,183	0,131	0,097	0,030	0,018
Д 2.1	0,439	0,449	0,511	0,717	1,015	1,320	2,904
Д 2.2	0,295	0,271	0,181	0,250	0,376	0,526	1,534
В 1.1	0,791	0,367	0,288	0,217	0,170	0,090	0,074

B 1.2	3,107	2,050	1,818	1,605	1,451	1,140	1,059
B 2.2	2,985	2,019	1,809	1,615	1,476	1,207	1,119
B 3.2	3,169	2,203	1,995	1,795	1,647	1,357	1,258
B 4.2	4,380	2,712	2,226	1,813	1,556	1,119	1,022
B 5.2	2,599	1,525	1,296	1,102	0,975	0,718	0,665
B 6.2	5,874	3,478	2,829	2,166	1,713	1,037	0,909
B 7.1	3,741	2,606	1,963	1,411	1,073	0,585	0,498
B 7.2	3,579	2,355	1,796	1,306	1,000	0,529	0,442
B 7.3	3,290	2,083	1,589	1,152	0,877	0,457	0,385
B 7.4	3,472	2,353	1,801	1,323	1,034	0,604	0,523
B 8.1	3,261	2,052	1,453	0,973	0,693	0,353	0,301
B 8.2	3,386	2,220	1,695	1,240	0,958	0,553	0,476
B 8.3	2,897	1,726	1,303	0,940	0,716	0,392	0,334
B 9.3	2,654	1,594	1,277	0,998	0,815	0,548	0,481
B 10.1	3,350	2,187	1,550	1,009	0,694	0,352	0,301
B 12.1	3,337	2,122	1,602	1,164	0,897	0,543	0,471
B 13.2	5,660	3,178	2,617	2,074	1,707	1,106	0,986
B 14.2	3,528	2,367	1,821	1,351	1,061	0,641	0,566
B 1.1	0,328	0,338	0,494	0,650	0,908	1,228	2,419
B 1.2	0,409	0,402	0,235	0,292	0,399	0,541	1,203
B 2.1	0,511	0,520	0,628	0,861	1,247	1,693	3,040
B 2.2	0,732	0,763	1,192	1,572	2,146	2,741	4,460
B 3.2	0,253	0,264	0,120	0,146	0,204	0,297	0,876
B 5.2	0,978	1,026	1,600	2,044	2,705	3,366	4,153
B 11.2	3,724	2,536	2,088	1,663	1,377	0,909	0,816
B 9.2	4,299	3,252	2,544	1,868	1,454	0,901	0,806
B 10.3	1,994	1,028	0,765	0,547	0,414	0,225	0,193
B 4.1	0,661	0,685	0,932	1,177	1,557	1,987	3,314

Полученные экспериментальные данные спектральных измерений частиц черного углерода различного происхождения были обработаны алгоритмом ЭГБ.

На первом этапе проводилась предварительная обработка данных, проводилась нормировка данных. После чего была проведена кросс-валидация модели методом Leave One Out. Обучающая выборка состояла из 65 образцов. При этом достоверность предсказания классов составила 78,5 %. Далее модели предъявлялись тестовые образцы, тестовая выборка состояла из 9 образцов. Результаты валидации по тестовому набору представлены в форме матрицы соответствия в таблице 7, в которой наглядно демонстрируется эффективность классификации для каждой группы образцов.

Таблица 7 – Результаты предсказания происхождения частиц черного углерода для тестового набора по спектрам абсорбции

Наименование образцов и исходный класс	Доверительная вероятность предсказания отнесения к классу		
	Керосин	Древесина	Биомасса
Древесина 2.1	0,004	0,994	0,003
Древесина 2.1	0,006	0,95	0,044
Древесина 2.2	0,005	0,952	0,043
Керосин 2.1	0,954	0,039	0,007
Керосин 4.1_old	0,948	0,021	0,031
Керосин 4.2_old	0,99	0,007	0,003
Биомасса 9.2	0,015	0,639	0,346
Биомасса 10.3	0,002	0,01	0,988
Биомасса 4.1	0,004	0,004	0,992
Примечание. Зеленым цветом обозначены достоверные предсказания, а красным – ложные			

Результаты исследования демонстрируют высокую эффективность разработанной модели классификации.

Данные результаты подтверждают перспективность использования методов МО и хемометрики в современной аналитической химии и показывают возможность их успешного применения для решения практических задач анализа.

Выводы к главе 3

В главе 3 была решена задача классификации частиц черного углерода по источникам выбросов с использованием методов абсорбционной и КР-спектроскопии с дальнейшей обработкой данных алгоритмом ЭГБ. На основе проведенных исследований можно сделать следующие основные выводы.

1. Проекция КР-спектров в плоскость главных компонент показывают, что спектры биомассы разделяются на две группы образцов, разных по происхождению.

2. Разработанная и прошедшая валидацию модель МО на основе алгоритма ЭГБ, демонстрирует высокую эффективность классификации:

- 10 достоверных предсказаний из 10 при тестовой валидации;
- распознавание фальш-образцов;
- доверительная вероятность достоверного распознавания более 89 %.

Применение данной модели подтверждается актом внедрения от ФГБУ «ИГКЭ».

3. Разработанная и апробированная методика спектральных измерений с использованием спектрофотометра Lambda 950, позволяет получать абсорбционные спектры частиц черного углерода различного происхождения в широком диапазоне значений оптической плотности. Методика включает тщательный контроль позиционирования образцов и учет возможных источников систематических погрешностей.

4. Разработанная и прошедшая валидацию модель МО на основе алгоритма ЭГБ, демонстрирует высокую эффективность классификации: 8 достоверных предсказаний из 9 при тестовой валидации, а также доверительную вероятность достоверного распознавания более 94 %.

5. Разработанные методы и алгоритмы будут интегрированы в ПО создаваемого аэталометра, что позволит проводить автоматическую классификацию частиц черного углерода по их происхождению в режиме реального времени. Аэталометр также работает на тех же длинах волн (370, 470, 520, 590, 660, 880, 950 нм), на которых проводились измерения оптической плотности.

Полученные результаты доказывают перспективность разработанного подхода для анализа происхождения частиц черного углерода в атмосферном воздухе.

ГЛАВА 4. КЛАССИФИКАЦИЯ АЛКОГОЛЬНОЙ ПРОДУКЦИИ ПО ГЕОГРАФИЧЕСКОМУ ПРОИСХОЖДЕНИЮ И ВОЗРАСТУ ВЫДЕРЖКИ

В области анализа алкогольной продукции особое значение имеет выбор методов исследования и обеспечение достоверности получаемых результатов. При формировании аналитических выборок критически важна точная информация об образцах, особенно для таких параметров, как географическое происхождение и возраст выдержки коньяков и бренди.

В наших исследованиях для классификации образцов бренди и коньячной продукции применяются независимо друг от друга различные спектральные методы:

- спектры эмиссии флуоресценции при различных длинах волн возбуждения;

- спектры синхронного сканирования с фиксированной разностью длин волн.

- спектры КР с возбуждением на длине волны 532 нм;

Эффективность методов МО во многом зависит от качества обучающей выборки. Для надежной классификации необходимо:

- использовать образцы с достоверно известными характеристиками всех классов в обучающей выборке;

- учитывать возможную вариабельность внутри классов;

- включать в обучающую выборку образцы от разных производителей;

- использовать информативные фрагменты спектров.

Важным является выбор оптимальных параметров алгоритмов МО. Для метода ЭГБ это включает:

- количество моделей решений или количество итераций модели (ознакомиться с алгоритмом можно в раздел 1.3);

- глубину моделей;

- скорость обучения;

- ограничение в количестве образов обучающей выборки.

4.1 Характеризация коньяков и виноградных бренди по спектрам флуоресценции

Виноградные бренди и коньяки содержат множество флуоресцирующих компонентов, включая фенольные соединения, ароматические альдегиды и другие вещества, формирующиеся в процессе выдержки в дубовых бочках. Состав и концентрация этих соединений существенно влияют на качество, органолептические свойства и подлинность продукции.

Для измерения спектров флуоресценции исследовались 43 образца виноградных бренди и коньяков разной выдержки от ведущих российских и зарубежных производителей.

Армянские бренди и коньяки были получены двумя путями: основная часть – непосредственно с завода-изготовителя в Армении или через магазины «duty free» с прямыми поставками от производителя, остальные – через крупные сетевые магазины.

Аналогично, большая часть дагестанских образцов была получена напрямую от производителей, меньшая – приобретена в крупных торговых сетях. При таком способе получения образцов их подлинность не вызывает сомнений.

При разбиении на классы по географическому происхождению учитывали, что в регионах Российской Федерации, кроме Дагестана, при производстве может быть использовано зарубежное сырье – коньячные дистилляты и виноград [87].

Важно отметить, что по законодательству Армении [88-89], в производстве армянских коньяков разрешено использовать только определенные сорта винограда, выращенные в Армении. Из дагестанских образцов три имели защищенное географическое указание «Дагестан»: «Старая крепость» (четырёхлетний), «коньяк пятилетний Дербент» и «КВ Дербент» (пятилетний). Так же дагестанские производители используют сырье только собственного происхождения и производства [90]. В таблице 8 продемонстрирована информация об образцах бренди и коньячной продукции.

Таблица 8 – Информация об образцах бренди и коньячной продукции

Страна	Производитель	Возраст (количество образцов)
Российская Федерация (кроме Дагестана)	ООО «ССБ», ООО «Винно-коньячный завод «КВС», ЗАО Ставропольский винно-коньячный завод, Московский Винно-коньячный завод «КиН», ООО КВКЗ (Коломенский винно-коньячный завод), АО «Татспиртпром», АО «Мосазервинзавод», ООО «ДЗАП «ОСТ-АЛКО», ЗАО винно-коньячный комбинат «РУСЬ», ООО «АЛВИСА», ВКЗ «Альянс-1892», ООО «СОРДИС», АО «Бахчисарайский комбинат Стройиндустрия», ООО «Завод Первомайский», ООО Завод марочных вин «Коктебель»	Меньше 3 лет (2), 3 года (8), 4 года (2), 5 лет (9), 7 лет (1)
Российская Федерация (Дагестан)	АО «Дербентский коньячный комбинат», АО «Кизлярский коньячный завод»	Меньше 3 лет (1), 3 года (1), 4 года (2), 5 лет (5)
Республика Армения	ЗАО «Ереванский Коньячный Завод», ООО «Коньячный завод «Агерак», ИП Гетапский винный завод, ЗАО «Веди-Алко», ООО «Арагатский Винный Завод», ООО Винно-коньячный дом «Шах назарян», ООО «Иджеванский винно-коньячный завод», «Арагат ЕКВВК», ООО «АРМЕНИЯ ВАЙН»	2 года (6), 5 лет (5), 15 лет (1), 20 лет (1)

Для регистрации спектров флуоресценции использовали спектрофотометр-флуориметр СФФ-2 «ФЛУОРАН» (производство ФГБУ «ВНИИОФИ», регистрационный номер в ФИФ ОЕИ 51314-12) с ортогональной схемой регистрации флуоресценции. В таблице 9 представлены метрологические характеристики прибора [91].

Таблица 9 – Метрологические характеристики спектрофотометр-флуориметр СФФ-2 «Флуоран»

Наименование характеристики	Значение характеристики
Спектральный диапазон, нм: – для режима абсорбции – для режима флуоресценции и рассеяния	190-1100 200-900
Спектральное разрешение монохроматора для режима абсорбции, нм, не более	1,0
Предел доверительной границы случайной погрешности установки длины волны при $P=0,95$, нм: – для режима абсорбции – для режима флуоресценции и рассеяния	0,2 0,2
Предел систематической составляющей погрешности установки длины волны, нм: – для режима абсорбции – для режима флуоресценции и рассеяния	0,3 0,7
Диапазон измерения СКНП, %	0-100
Предел доверительной границы случайной погрешности измерения СКНП при $P=0,95$ %	0,2
Пределы систематической составляющей погрешности измерения СКНП в спектральном диапазоне от 340 до 850 нм, %	$\pm 0,5$
Отклонение линии 100 % пропускания в режиме абсорбции в диапазоне от 200 до 900 нм, при спектральной ширине щели 2 нм скорости сканирования 240 нм/мин, %, не более	0,3
Уровень рассеянного излучения на длине волны 340 нм (NaNO_2), %, не более	0,05
Отношение Сигнал/Шум для полосы комбинационного (рамановского) рассеяния дистиллированной воды при длине волны возбуждающего излучение 361 нм и спектральной ширине щели 5 нм, времени регистрации 2 с, не менее	400

Были выбраны виды спектров флуоресценции и их фрагменты, имеющие характерные особенности, по совокупности которых предполагалось установление значимых различий между образцами бренди и коньяков различного географического происхождения:

1. спектры эмиссии в диапазоне длин волн от 290 до 450 нм при длине волны возбуждения 280 нм (коньяки обладают максимумом поглощения при 280 нм, коньячный максимум) [92];
2. спектры эмиссии в диапазоне длин волн от 260 до 420 нм при длине волны возбуждения 250 нм (коньяки обладают минимумом поглощения при 250 нм, коньячный минимум);
3. спектры синхронного сканирования в диапазоне длин волн от 235 до 450 нм при разнице длин волн испускания и возбуждения 50 нм.

Перед проведением измерений образцы бренди и коньяков предварительно разбавляли деионизованной водой в соотношении 1:29 для минимизации эффекта внутреннего фильтра, который мог бы привести к искажению получаемых спектров [93]. Оптимальное разбавление было подобрано экспериментально. Для каждого из образцов производилось по три измерения в условиях повторяемости.

На рисунках 9 и 10 продемонстрированы спектры эмиссии в диапазоне длин волн 290–450 нм при длине волны возбуждения 280 (коньячный максимум) и 260–420 нм при длине волны возбуждения 250 нм (коньячный минимум).

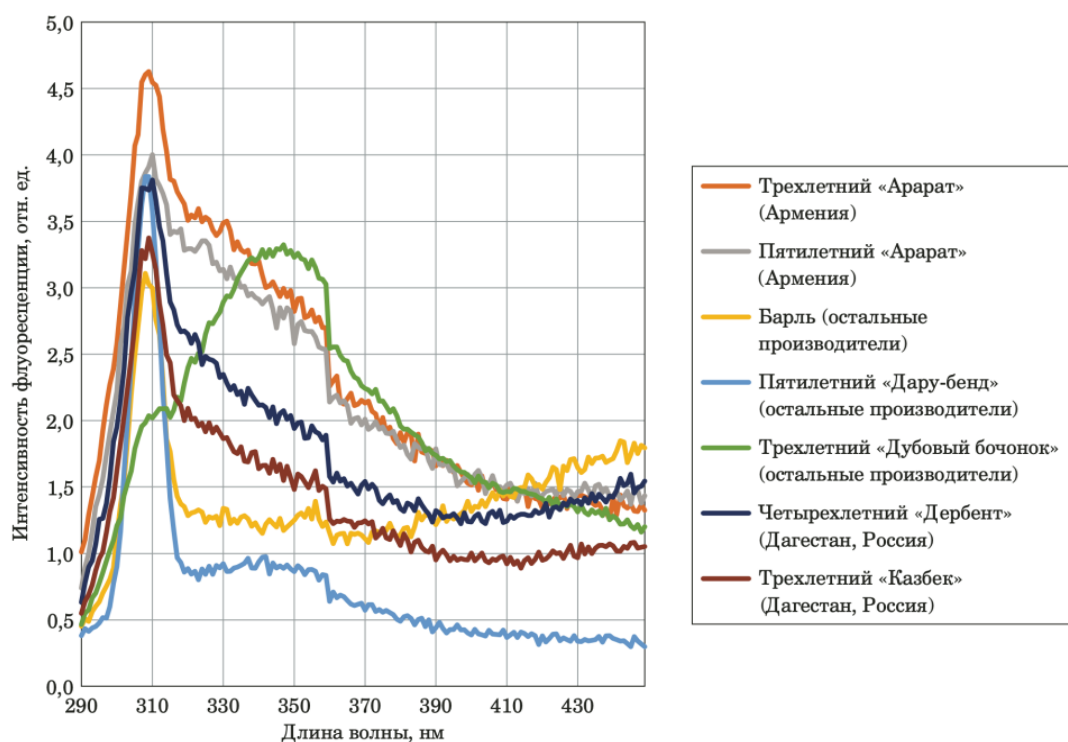


Рисунок 9 – Спектры эмиссии в диапазоне длин волн 290–450 нм при длине волны возбуждения 280 нм (коньячный максимум)

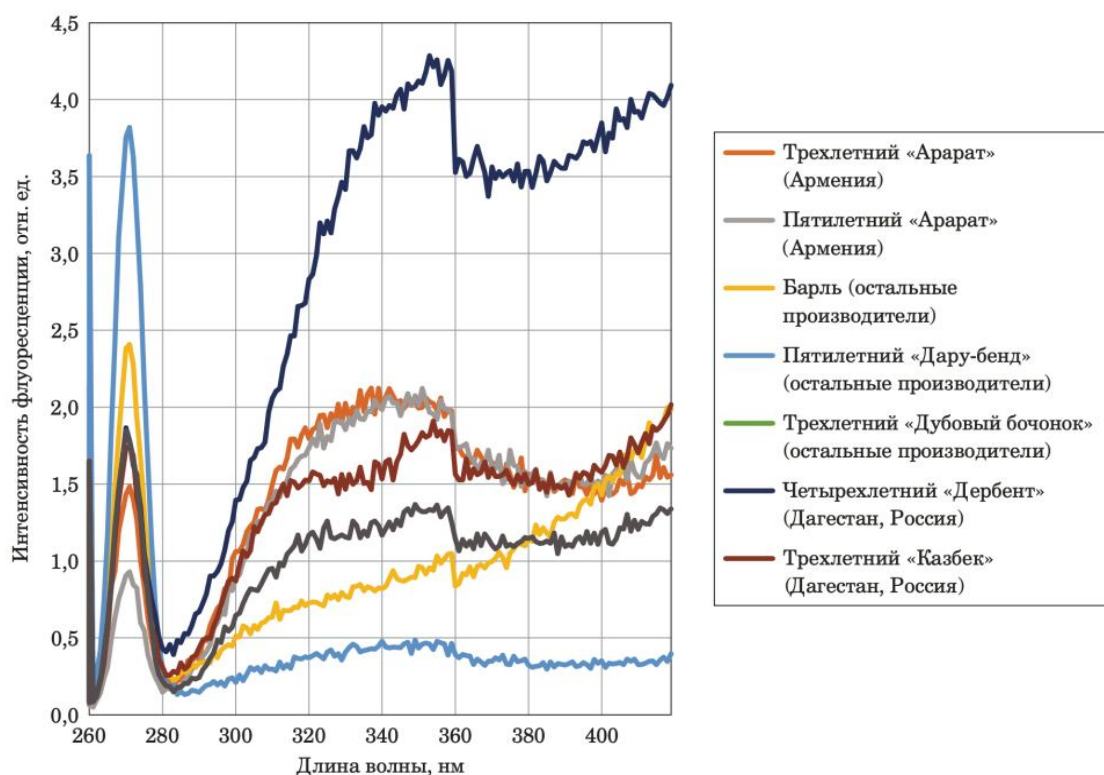


Рисунок 10 – Спектры эмиссии в диапазоне длин волн 260–420 нм при длине волны возбуждения 250 нм (коньячный минимум)

Метод синхронного сканирования позволяет получить более четкое разделение спектральных характеристик образцов, что особенно важно для их

классификации по географическому происхождению. На рисунке 11 продемонстрированы спектры синхронного сканирования тех же образцов в диапазоне длин волн 235-450 нм при разности длин волн возбуждения и эмиссии ($\Delta\lambda$) = 50 нм.

Обучение модели проводили на обучающей выборке, состоящей из 33 образцов (99 спектров) спектров бренди и коньяков различных марок. Достоверность классификации образцов с помощью обученной модели оценивали в два этапа.

На первом этапе проводилась кросс-валидация с использованием только образцов, входивших в обучающий набор. Достоверность предсказания при проверке обучающего набора для определения географического происхождения, состоящего из 33 образцов (99 спектров), составило 33 образца (99 спектров).

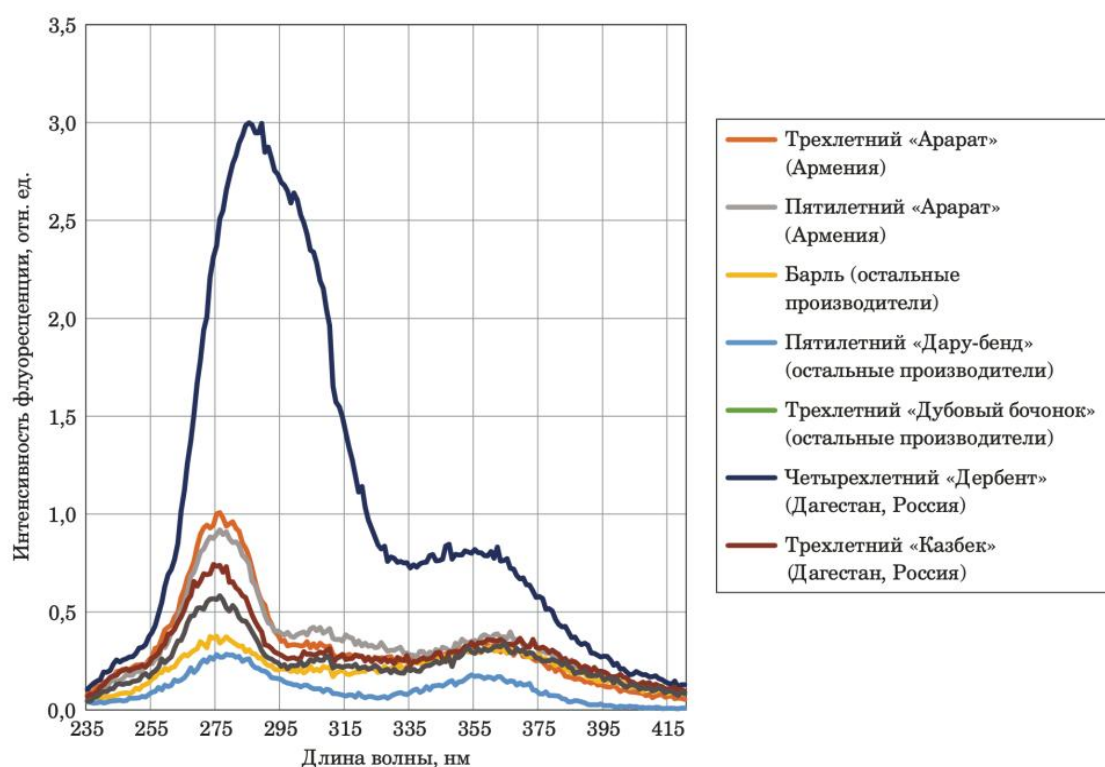


Рисунок 11 – Спектры синхронного сканирования в диапазоне длин волн 235–450 нм при разности длин волн возбуждения и эмиссии $\Delta\lambda = 50$ нм

На втором этапе модели, обученной ранее по 33 образцам (99 спектрам), предоставляли тестовый набор, состоящий из 10 образцов (10 спектров) различных напитков, которые не участвовали в обучении. Все образцы тестового набора были определены достоверно для моделей, построенных как по спектрам синхронного сканирования, так и по спектрам эмиссии при возбуждении на длинах волн конъюжного минимума и максимума.

В таблицах 10–12 продемонстрированы результаты классификации образцов тестового набора, полученные с использованием разных спектральных методов. В каждой таблице для каждого образца указаны предсказанное географическое происхождение и соответствующие доверительные вероятности. Достоверные результаты предсказаний выделены зеленым цветом.

Таблица 10 – Результаты предсказания географического происхождения для тестового набора по спектрам эмиссии на длине волны возбуждения конъюжного минимума (250 нм)

Наименование образцов и исходный класс	Доверительная вероятность предсказания географического происхождения		
	Армения	Регионы России, кроме Дагестана	Россия (Дагестан)
Армения «Арагат» трехлетний	0,979	0,01	0,011
Армения «Арвест» пятилетний	0,972	0,012	0,016
Армения «Митридат» трехлетний	0,973	0,012	0,014
Армения «Митридат» пятилетний	0,981	0,01	0,009
Россия «Киновский» трехлетний	0,017	0,971	0,012
Россия «Командирский»	0,009	0,978	0,013
Россия «Олд Баррель» пятилетний	0,003	0,986	0,01
Россия «Трофейный» четырехлетний	0,005	0,986	0,002
Россия (Дагестан) «Казбек» трехлетний	0,025	0,031	0,944
Россия (Дагестан) «Лезгинка» пятилетний	0,014	0,01	0,976

Таблица 11 – Результаты предсказания географического происхождения для тестового набора по спектрам эмиссии на длине волны возбуждения конъючного максимума (280 нм)

Наименование образцов и исходный класс	Доверительная вероятность предсказания географического происхождения		
	Армения	Регионы России, кроме Дагестана	Россия (Дагестан)
Армения «Арагат» трехлетний	0,969	0,014	0,017
Армения «Арвест» пятилетний	0,960	0,017	0,023
Армения «Митридат» трехлетний	0,975	0,01	0,015
Армения «Митридат» пятилетний	0,975	0,016	0,009
Россия «Киновский» трехлетний	0,011	0,978	0,012
Россия «Командирский»	0,008	0,978	0,014
Россия «Олд Баррель» пятилетний	0,01	0,975	0,015
Россия «Трофейный» четырехлетний	0,02	0,955	0,025
Россия (Дагестан) «Казбек» трехлетний	0,02	0,05	0,930
Россия (Дагестан) «Лезгинка» пятилетний	0,014	0,02	0,966

Таблица 12 – Результаты предсказания географического происхождения для тестового набора по спектрам синхронного сканирования со сдвигом длины волны 50 нм

Наименование образцов и исходный класс	Доверительная вероятность предсказания географического происхождения		
	Армения	Регионы России, кроме Дагестана	Россия (Дагестан)
Армения «Арагат» трехлетний	0,982	0,006	0,012
Армения «Арвест» пятилетний	0,976	0,01	0,014
Армения «Митридат» трехлетний	0,973	0,01	0,017
Армения «Митридат» пятилетний	0,966	0,014	0,02
Россия «Киновский» трехлетний	0,009	0,973	0,018
Россия «Командирский»	0,01	0,966	0,024
Россия «Олд Баррель» пятилетний	0,006	0,985	0,009
Россия «Трофейный» четырехлетний	0,006	0,982	0,012
Россия (Дагестан) «Казбек» трехлетний	0,03	0,036	0,934
Россия (Дагестан) «Лезгинка» пятилетний	0,014	0,015	0,971

Использованные спектральные методы (эмиссия при 250 нм, эмиссия при 280 нм и синхронное сканирование) с хемометрической обработкой

классифицировали все тестовые образцы достоверно с высокими значениями доверительной вероятности (96,6 % – 97,5 %). Это подтверждает эффективность разработанного подхода для определения географического происхождения коньяков и бренди.

4.2 Классификация бренди и коньячной продукции по географическому происхождению и возрасту выдержки с использованием КР-спектроскопии

Для измерения спектров КР исследовались 42 образца виноградных бренди и коньяков разной выдержки, те же образцы, о которых говорилось в прошлом подразделе.

Эксперименты проведены на раман-люминесцентном портативном спектрометре «ИнСпектр» R532 фирмы «Спектр-М» (Черноголовка, Россия), метрологические и технические характеристики прибора приведены в главе 3. Прибор работает на длине волны возбуждения 532 нм, мощность излучения лазера 20-30 мВт. В качестве наиболее информативного выбран диапазон рамановских сдвигов 1400-1500 см⁻¹.

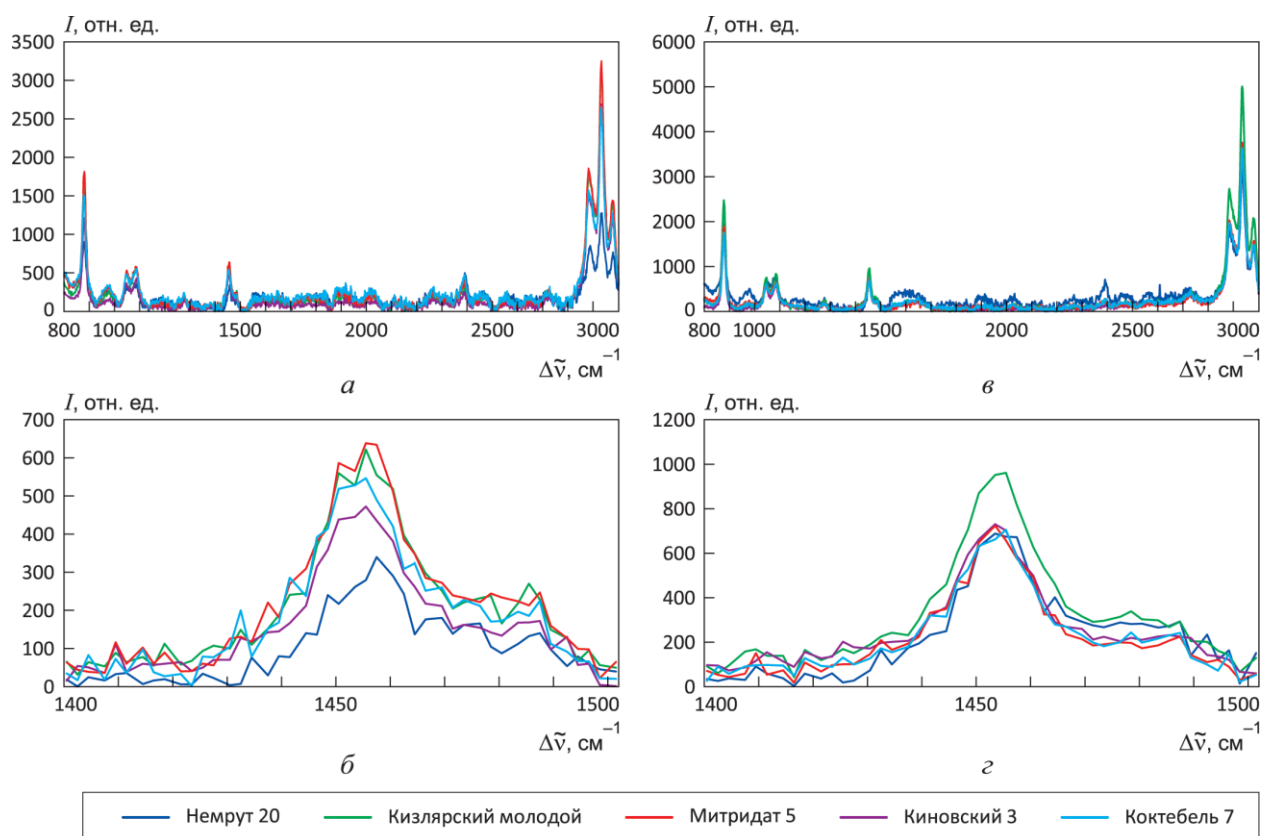
Проведены два вида измерений: с неразбавленными образцами и с образцами, разбавленными деионизованной водой в соотношении 1:2. Для каждого из образцов производилось по три измерения в условиях повторяемости. В таблице 13 представлена информация об образцах бренди и коньячной продукции.

По возрасту выдержки образцы также распределили на три класса: до 5 лет, от 5 до 7 лет и старше 7 лет. Распределение происходило с учетом особенностей образцов, чтобы создать три класса с достаточным количеством образцов для обучающего набора.

Таблица 13 – Информация об образцах бренди и коньячной продукции

Страна	Производитель	Возраст (количество образцов)
Республика Дагестан	АО «Дербентский коньячный комбинат», АО «Кизлярский коньячный завод»	Меньше 3 лет (1) 3 года (1) 4 года (2) 5 лет (4)
Республика Армения	ЗАО «Ереванский коньячный завод», ООО «Коньячный завод «Агерак», ИП Гетапский винный завод, ЗАО «Веди-Алко», ООО «Арагатский винный завод», ООО «Вино-коньячный дом «Шах назарян», ЗАО «Иджеванский винно-коньячный завод», «Арагат ЕКВВК», ООО «АРМЕНИЯ ВАЙН»	3 года (6) 5 лет (4) 15 лет (1) 20 лет (1)
Российская Федерация (кроме Дагестана)	ООО «ССБ», ООО «Винно-коньячный завод «КВС», ЗАО «Ставропольский винно-коньячный завод», АО «Московский винно-коньячный завод «КиН», ООО «Коломенский винно-коньячный завод», АО «Татспиртпром», АО «Мосазервинзавод», ООО «ДЗАП «ОСТ-АЛКО», ЗАО «Винно-коньячный комбинат «РУСЬ», ООО «АЛВИСА», ВКЗ «Альянс-1892», ООО «СОРДИС	Меньше 3 лет (2) 3 года (8) 4 года (2) 5 лет (9) 7 лет (1)

На рисунке 12 представлены спектры КР неразбавленных и разбавленных образцов коньячной продукции, а также характерный фрагмент диапазона, использованный для построения моделей.



а, в – обзорный спектр нескольких образцов во всём диапазоне $800\text{--}3000\text{см}^{-1}$;
 б, г – характерный фрагмент диапазона, использованный для построения моделей;
 I – интенсивность спектральных линий, $\Delta\tilde{\nu}$ – рамановский сдвиг (волновое число)
 Рисунок 12 – Спектры КР неразбавленных и разбавленных образцов коньячной
 продукции

Из измеренных КР-спектров 42 образцов сформированы обучающие и тестовые наборы:

Для классификации по географическому происхождению:

- обучающий набор – 32 образца (96 спектров);
- тестовый набор – 10 образцов (10 спектров).

Для классификации по возрасту выдержки:

- обучающий набор – 30 образцов (90 спектров);
- тестовый набор – 12 образцов (12 спектров).

Для обучения моделей применяли алгоритм ЭГБ. Для каждой модели производилась процедура кросс-валидации. Достоверность предсказания при проверке обучающего набора для определения географического происхождения, состоящего из 32 образцов (96 спектров), составила 32 образца (96 спектров). Для

модели по определению возраста выдержки достоверность предсказания составила 30 образцов (90 спектров).

В таблице 14 представлены результаты предсказаний географического происхождения для тестового набора из 10 образцов, не использованных при обучении модели. Достоверные предсказания выделены зеленым цветом, а ошибочные – красным.

Таблица 14 – Результаты предсказания географического происхождения для тестового набора по КР-спектрам

Исходные данные		Доверительная вероятность предсказания		
Регион происхождения	Наименование коньячной продукции	Республика Армения	Российская Федерация (кроме Дагестана)	Республика Дагестан
Неразбавленные образцы				
Республика Армения	Арарат 3	0,962	0,016	0,022
	Арвест 5	0,979	0,0109	0,0101
	Митридат 5	0,968	0,016	0,016
	Митридат 3	0,964	0,025	0,011
Российская Федерация (кроме Дагестана)	Олд Баррель 5	0,01	0,982	0,008
	Киновский 3	0,006	0,982	0,012
	Командирский 3	0,006	0,984	0,01
	Старейшина 3	0,008	0,984	0,008
Республика Дагестан	Казбек 3	0,025	0,01	0,965
	Лезгинка 5	0,02	0,014	0,966
Разбавленные образцы				
Республика Армения	Арарат 3	0,973	0,015	0,012
	Арвест 5	0,971	0,012	0,017
	Митридат 5	0,970	0,017	0,013
	Митридат 3	0,410	0,585	0,005
Российская Федерация (кроме Дагестана)	Олд Баррель 5	0,021	0,972	0,007
	Киновский 3	0,012	0,972	0,016
	Командирский 3	0,01	0,981	0,009
	Старейшина 3	0,008	0,986	0,006
Республика Дагестан	Казбек 3	0,013	0,047	0,940
	Лезгинка 5	0,013	0,023	0,964

Для неразбавленных образцов из Республики Армения («Арарат 3», «Арвест 5», «Митридат 5» и «Митридат 3») модель показала высокие доверительные вероятности достоверной классификации – от 96,2 % до 97,9 %.

Образцы из регионов России, кроме Дагестана («Олд Баррель 5», «Киновский 3», «Командирский 3» и «Старейшина 3»), были классифицированы с вероятностями от 98,2 % до 98,4 %. Образцы из Дагестана («Казбек 3» и «Лезгинка 5») были достоверно определены с вероятностями 96,5 % и 96,6 %, соответственно.

При анализе разбавленных образцов точность классификации по географическому происхождению несколько снизилась. Особенно заметно это для образца «Митридат 3» (Армения), который был ошибочно классифицирован как продукт из регионов России с вероятностью 58,5 %. Остальные армянские образцы были определены достоверно с вероятностями 97 % – 97,3 %. Российские образцы (кроме дагестанских) показали вероятности достоверной классификации 97,2 % – 98,6 %, а дагестанские – 94 % – 96,4 %.

Важно отметить, что в случае неразбавленных образцов модель не допустила ни одной ошибки классификации, в то время как для разбавленных образцов была допущена одна ошибка. Это свидетельствует о более высокой надежности анализа неразбавленных образцов.

В таблице 15 представлены результаты предсказаний возраста выдержки для тестового набора коньяков и бренди. Результаты представлены как для неразбавленных, так и для разбавленных образцов.

Таблица 15 – Результаты предсказания возраста выдержки для тестового набора по КР-спектрам

Наименование коньячной продукции (возраст выдержки)	Доверительная вероятность предсказания		
	меньше 5 лет	5–7 лет	больше 7 лет
Неразбавленные образцы			
Дару Бенд (до 3 лет)	0,980	0,013	0,007
Кизлярский молодой (до 3 лет)	0,983	0,01	0,007
Крымский статус (до 3 лет)	0,981	0,009	0,01
Арарат 3	0,980	0,011	0,009
Дудук 3	0,987	0,006	0,007
Казбек 4	0,980	0,008	0,012
Старый Кенигсберг 4	0,982	0,009	0,009
Баррель 5	0,006	0,983	0,011
Олд Баррель 5	0,007	0,985	0,008
Арарат 5	0,008	0,986	0,006
Арвест 5	0,009	0,984	0,007
Армянский коньяк 5	0,008	0,983	0,009
Коктебель 5	0,015	0,972	0,013
Немрут 20	0,053	0,080	0,867
Разбавленные образцы			
Дару Бенд (до 3 лет)	0,984	0,009	0,007
Кизлярский молодой (до 3 лет)	0,987	0,008	0,005
Крымский статус (до 3 лет)	0,985	0,006	0,009
Арарат 3	0,980	0,010	0,010
Дудук 3	0,983	0,010	0,007
Казбек 4	0,983	0,009	0,008
Старый Кенигсберг 4	0,982	0,006	0,012
Баррель 5	0,26	0,959	0,150
Олд Баррель 5	0,032	0,945	0,023
Арарат 5	0,019	0,970	0,011
Арвест 5	0,010	0,983	0,007
Армянский коньяк 5	0,015	0,973	0,012
Коктебель 5	0,022	0,967	0,011
Немрут 20	0,010	0,120	0,861

Для неразбавленных образцов с возрастом выдержки менее 5 лет модель показала высокие доверительные вероятности достоверной классификации – от 98 % до 98,7%. Все образцы этой группы были классифицированы верно.

Образцы с выдержкой 5–7 лет также были достоверно классифицированы с вероятностями от 97,2 % до 98,6 %. Особенно высокую точность модель показала для образцов «Арагат 5» (98,6 %) и «Олд Баррель 5» (98,5 %).

Для образца с длительным возрастом выдержки модель также достоверно определила категорию «больше 7 лет» с вероятностью 86,7 %.

При анализе разбавленных образцов точность классификации в целом сохранилась высокой, но с несколько меньшими значениями доверительных вероятностей. Для образцов с выдержкой до 5 лет вероятности составили 98 % – 98,7 %, для образцов с выдержкой 5–7 лет – 94,5 % – 98,3 %, а для образцов с выдержкой больше 7 лет – 86,1 %.

Выводы к главе 4.

В главе 4 представлены результаты разработки и применения методов классификации бренди и коньячной продукции по географическому происхождению и возрасту выдержки с использованием флуоресцентной и КР-спектроскопии. На основе проведенных исследований можно сделать следующие основные выводы:

1. Определены информативные фрагменты флуоресцентных и КР-спектров для классификации бренди и коньячной продукции по географическому происхождению:

— Эмиссия флуоресценции в диапазоне от 260 до 420 нм при возбуждении на длине волны 250 нм;

— Эмиссия флуоресценции в диапазоне от 290 до 450 нм при возбуждении на длине волны 280 нм;

— Синхронное сканирование в диапазоне длин волн от 235 до 450 нм при разности длин волн 50 нм;

— КР-спектроскопия в диапазоне от 1400 до 1500 см⁻¹ при возбуждении на длине волны 532 нм.

2. Модели МО, построенные по спектрам КР неразбавленных образцов, дают более достоверные предсказания возраста по сравнению с моделями, построенными по спектрам разбавленных образцов. Все образцы тестового набора были определены достоверно.

3. В результатах предсказаний обученной модели МО по распознаванию географического происхождения разбавленных образцов тестового набора присутствует только одно ошибочное предсказание, а для неразбавленных образцов ошибочных предсказаний нет.

4. Спектры флуоресценции и КР, обработанные с помощью выбранного алгоритма МО ЭГБ, позволили эффективно классифицировать образцы коньячной продукции по географическому происхождению и возрасту выдержки.

Реализованные методы представляют значительный интерес для экспресс-анализа (скрининга) образцов алкогольной продукции при проверке и подтверждении информации об их происхождении и возрасте выдержки, заявляемой производителями и продавцами.

ГЛАВА 5. ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ КОЛИЧЕСТВЕННОГО АНАЛИЗА МНОГОКОМПОНЕНТНЫХ СМЕСЕЙ ПО СПЕКТРАМ КР

КР-спектроскопия в сочетании с методами классификации успешно применяется для решения множества практических задач, некоторые из которых были обсуждены выше.

Представляет интерес применение КР-спектроскопии не только для качественного, но и для количественного анализа. Однако оно затрудняется из-за аппаратной зависимости интенсивности линий КР. В наших исследованиях для преодоления данной трудности был предложен метод градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах с помощью образцов сравнения, материал которых имеет характерный КР-спектр. Метод, который мы предлагаем, похож на подход, используемый в флуоресцентной спектроскопии, где используется так называемая относительная единица флуоресценции (ОЕФ), за которую принимается интенсивность флуоресценции раствора флуоресцеина натрия концентрацией 1 мг/дм^3 [94]. Мы рассматривали два варианта образцов сравнения с характерными КР-спектрами – на основе пленки полистирола с добавлением серы и на основе циклогексана (жидкости).

5.1 Градуировка фотометрической шкалы в приборно-независимых единицах с помощью образца сравнения в виде пленки полистирола с добавлением серы и количественный анализ многокомпонентных смесей по спектрам КР

В качестве одного из образцов сравнения была использована разработанная в ФГБУ «ВНИИОФИ» мера волновых чисел рамановских сдвигов, материал которой состоит из смеси полистирола и серы. Пленка помещена в специальную оправку. На рисунке 13 представлена фотография данной меры. Метрологические характеристики меры приведены в таблице 16 [95].



Рисунок 13 – Фотография меры волновых чисел рамановских сдвигов

Таблица 16 – Метрологические характеристики меры

Наименование характеристики	Значение
Диапазон волновых чисел рамановских сдвигов, см^{-1}	от 80 до 3000
Пределы допускаемой абсолютной погрешности измерения волновых чисел рамановских сдвигов, см^{-1}	$\pm 1,0$
Диапазон допускаемых значений волнового числа рамановского сдвига, см^{-1}	$85,1 \pm 10,0$
	$153,8 \pm 10,0$
	$219,1 \pm 10,0$
	$473,2 \pm 10,0$
	$620,9 \pm 10,0$
	$795,8 \pm 10,0$
	$1001,4 \pm 10,0$
	$1031,8 \pm 10,0$
	$1155,3 \pm 10,0$
	$1450,5 \pm 10,0$
	$1583,1 \pm 10,0$
	$1602,3 \pm 10,0$
	$2852,4 \pm 10,0$
	$2904,5 \pm 10,0$

Мера успешно прошла испытания в целях утверждения типа средства измерения и получила регистрационный номер 93847-24 в ФИФ ОЕИ, что подтверждает ее метрологическую надежность и соответствие требованиям.

Принцип использования образца сравнения для количественного анализа основан на том, что она позволяет измерять интенсивность линий КР в приборно-

независимых единицах. При измерениях на КР-спектрометрах зависимость интенсивности КР-линии I от концентрации анализируемого компонента C имеет следующий вид

$$I = I_0 \cdot \sigma \cdot A \cdot S \cdot h \cdot C, \quad (9)$$

где I_0 – интенсивность излучения, возбуждающего КР; σ – сечение КР для используемой спектральной линии; A – коэффициент, учитывающий особенности прибора, не зависящие от образца (мощность возбуждающего лазера, эффективность сбора рассеянного излучения, чувствительность фотоприемника и т.п.); S и h – соответственно поперечное сечение и толщина области рассеяния (конфокальной области); C – концентрация анализируемого компонента.

Параметры I_0 , A и S зависят только от характеристик и настроек прибора, параметры σ и C – только от свойств образца, а параметр h в случае объемного образца зависит от прибора (равен толщине его конфокальной области $h_{\text{конф}}$), в случае образца в виде пленки – ее толщине $h_{\text{пл}}$.

Численное значение интенсивности линии образца в предлагаемых единицах равно отношению регистрируемой прибором интенсивности линии КР исследуемого вещества к опорной интенсивности линии образца сравнения, расположенной вблизи используемой линии образца по шкале волновых чисел. Из формулы (9) следует, что при измерении спектров КР на разных приборах интенсивность линии КР исследуемого вещества в этих единицах $I_{\text{норм}}$ определяется выражением

$$I_{\text{норм}} = \frac{I_{\text{обр}}}{I_{\text{обр.срав.}}} = \frac{\sigma_{\text{обр}}}{\sigma_{\text{обр.срав.}} \cdot C_{\text{обр.срав.}}} \left(\frac{h_{\text{конф}}}{h_{\text{обр.срав.}}} \right) * C_{\text{обр}}, \quad (10)$$

где $I_{\text{норм}}$ – нормированная интенсивность; $I_{\text{обр}}$ – интенсивность образца; $I_{\text{мера}}$ – интенсивность меры, $h_{\text{конф}}$ – толщина конфокальной области; $h_{\text{мера}}$ – толщина меры; $C_{\text{мера}}$ – концентрация (доля) вещества в мере; $C_{\text{обр}}$ – концентрация (доля) вещества в образце.

Таким образом, интенсивность линии КР в нормированных единицах пропорциональна концентрации (доле) анализируемого компонента, а

единственным параметром прибора, от которого зависит эта интенсивность, является толщина его конфокальной области.

На основе этого принципа можно построить градуировочную зависимость интенсивности КР-линии, выраженной в предлагаемых единицах, от концентрации определяемого вещества. Такая зависимость может быть использована для определения концентраций веществ в анализируемых образцах по КР-спектрам, измеренным на других приборах или на том же приборе, но при измененных условиях.

При этом потребуется вводить поправку, учитывающую соотношение толщин конфокальных областей при построении градуировочной зависимости и измерениях в анализируемых образцах.

Для проведения количественного анализа использовались следующие химические реактивы производства «Sigma-Aldrich» (Германия):

- этанол (C_2H_5OH), CAS № 64-17-5, кат. № 493511, степень чистоты не менее 95,0 %;
- метанол (CH_3OH), CAS № 67-56-1, кат. № 34885, степень чистоты не менее 99,9 %;
- сахароза ($C_{12}H_{22}O_{11}$), CAS № 57-50-1, кат. № S0389, степень чистоты не менее 99,5 %;
- глюкоза кристаллическая ($C_6H_{12}O_6$), CAS № 50-99-7, кат. № G8270, степень чистоты не менее 99,5 %;
- фруктоза ($C_6H_{12}O_6$), CAS № 57-48-7, кат. № F0127, степень чистоты не менее 99 %.

Выбор данных реактивов обусловлен их широким применением в аналитической практике, а также возможностью создания модельных смесей различной сложности для проверки разрабатываемой методики.

Для обеспечения точности приготовления растворов с заданными концентрациями компонентов измерения массы используемых реагентов проводили на весах лабораторных аналитических марки Pioneer модификации

РА214С I класса точности, с дискретностью 0,1 мг (регистрационный номер в ФИФ ОЕИ 38796-08).

Для приготовления растворов использовали дозатор пипеточный одноканальный с переменным объемом от 100 до 1000 мкл ДПОП-1-100-1000 (регистрационный номер в ФИФ ОЕИ 37432-013). Предел допускаемой систематической составляющей основной относительной погрешности составляет $\pm$ (от 1,5 до 1,0) %. Предел допускаемого среднеквадратичного отклонения случайной составляющей относительной погрешности находится в диапазоне от 2,0 до 1,0 %.

Для построения надежной градуировочной зависимости для многокомпонентных смесей спиртов и сахаров были изготовлены растворы с объемными концентрациями, указанными в таблицах 17 и 18. Выбор именно этих концентраций обусловлен необходимостью охватить широкий диапазон содержания анализируемых компонентов.

Таблица 17 – Соотношение компонентов в образцах водных смесей этанола и метанола

№ образца	Количество этанола, об. %	Количество метанола, об. %	Количество воды, об. %
1	80	10	10
2	10	80	10
3	45	45	10
4	10	10	80
5	60	30	10
6	30	60	10
7	40	20	40

Таблица 18 – Соотношение компонентов в образцах водных смесей глюкозы, фруктозы и сахарозы

№ образца	Количество фруктозы, об. %	Количество глюкозы, об. %	Количество сахарозы, об. %
1	50	25	25
2	25	50	25
3	25	25	50
4	33,3	33,3	33,3
5	75	15	10
6	15	10	75
7	20	20	60
8	60	30	10
9	15	70	15
10	60	10	30
11	30	0	70
12	29	71	0
13	80	0	20
14	20	80	0

Измерения КР-спектров проводились на раман-люминесцентном портативном спектрометре «ИнСпектр» R532 фирмы «Спектр-М» (Черноголовка, Россия), описание метрологических и технических характеристик прибора приведены в разделе 3.1. Оптимальные параметры измерения были подобраны экспериментально для обеспечения высокого соотношения сигнал/шум при разумной продолжительности измерения и приведены в таблице 19 и 20.

Таблица 19 – Условия измерения КР-спектров образцов водных смесей спиртов

Наименование параметра	Значение параметра
Время накопления, мс	1000
Количество кадров/изм.	20
Количество последовательных измерений спектров	5

Таблица 20 – Условия измерения КР-спектров образцов водных смесей сахаров

Наименование параметра	Значение параметра
Время накопления, мс	1500
Количество кадров/изм.	20
Количество последовательных измерений спектров	5

На рисунках 14 и 15 продемонстрированы информативные фрагменты КР-спектров смеси спиртов и смеси сахаров (информативный диапазон спектра для смеси спиртов – от 100 до 3000 см^{-1} , а для смеси сахаров – от 117 до 1700 см^{-1}).

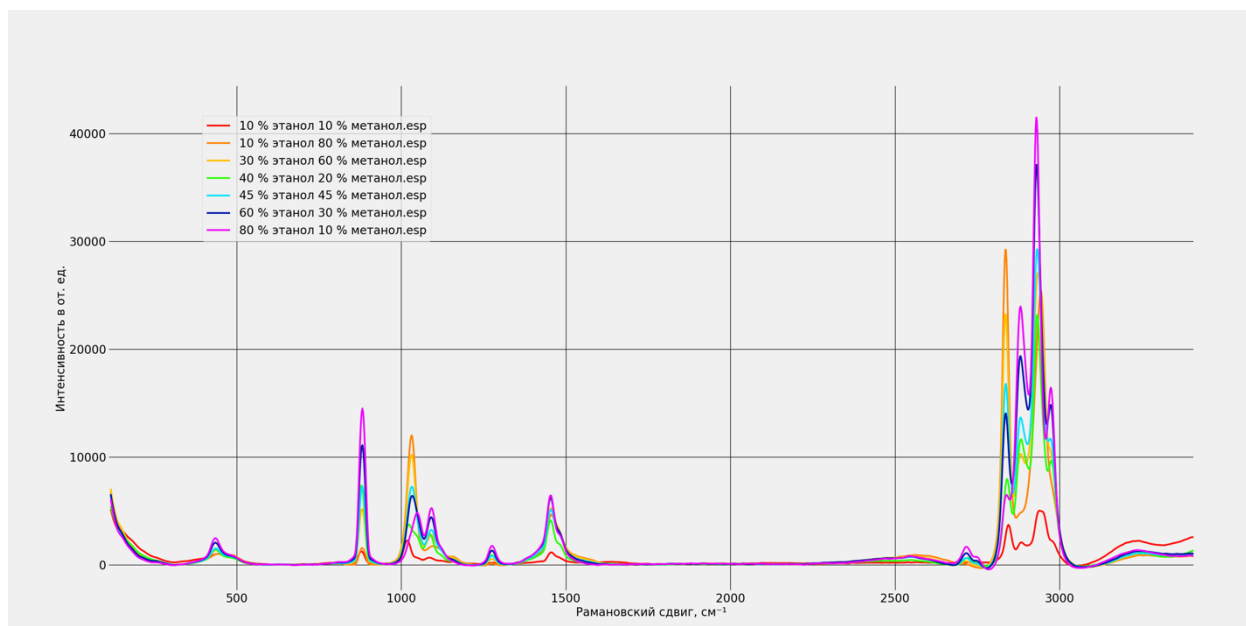


Рисунок 14 – Фрагменты КР-спектров растворов с разными объемными долями спиртов

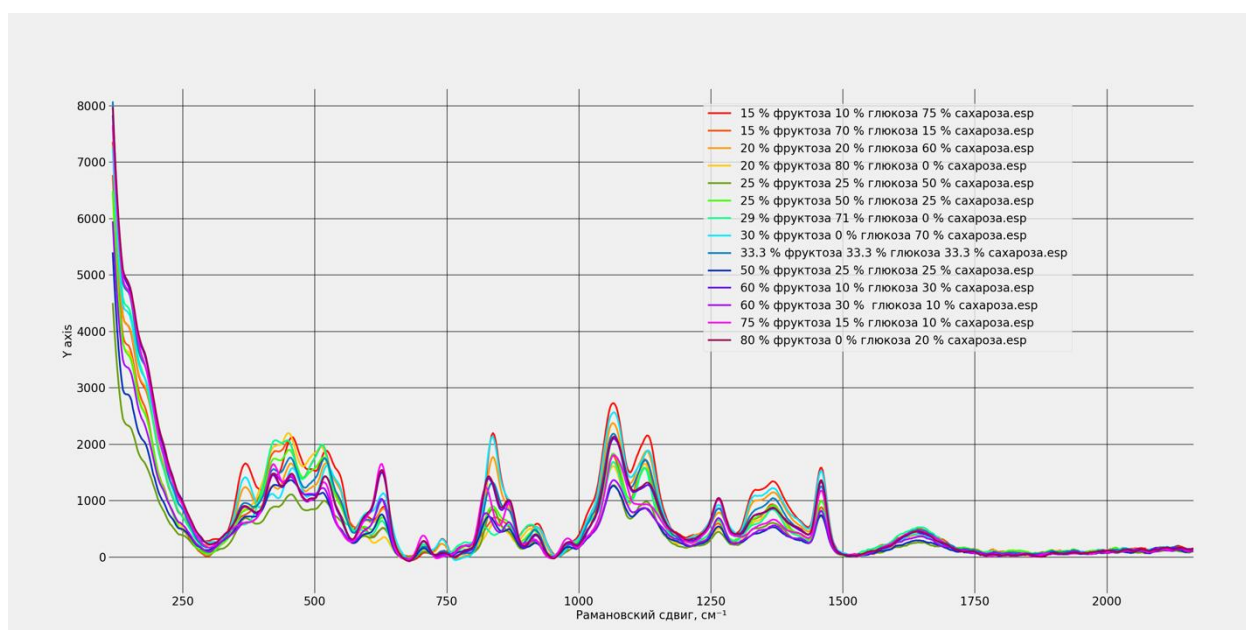


Рисунок 15 – Фрагменты КР-спектров растворов с разными объемными долями сахаров

Для каждого образца измеряли КР-спектры и вычисляли отношение интенсивностей характеристического пика веществ к интенсивности близкорасположенного пика меры (1002 см^{-1}) $I_{\text{норм}}$ по формуле (10).

На основе полученных данных строился градуировочный график в координатах «концентрация этанола – соотношение интенсивностей». Линейность полученной зависимости позволила в дальнейшем проводить определение концентрации неизвестного образца по одному образцу с известной концентрацией по формуле

$$c_x = \frac{c_1 * I_{\text{норм},x}}{I_{\text{норм},x}}, \quad (11)$$

где c_1 – известная объемная доля (%) этанола; $I_{\text{норм},1}$ – отношение интенсивностей пиков этанола и меры для образца с известной концентрацией; $I_{\text{норм},x}$ – отношение интенсивностей пиков этанола и меры для образца с неизвестной концентрацией.

Поскольку спектры КР, используемые для построения градуировочных графиков и для определения концентраций компонентов в неизвестных образцах, измерялись на одном и том же приборе, не было необходимости вводить поправки, связанные с толщиной конфокальной области.

Среднее значение определяемой объемной доли этанола рассчитывали по формуле

$$c_{x,\text{cp}} = \frac{\sum_i^N c_{x,i}}{N}, \quad (12)$$

где c_{xi} – рассчитанная объемная доля (%) этанола в i -м измерении, %, N – количество измерений.

Для оценки точности метода определяли относительное среднеквадратическое отклонение случайной составляющей погрешности определения концентрации по формуле

$$S = \frac{1}{c_{x,\text{cp}}} \sqrt{\frac{\sum_i^N (c_{x,i} - c_{\text{cp}})^2}{N - 1}}. \quad (13)$$

Относительную систематическую составляющую погрешности определения концентрации оценивали по формуле

$$\Delta = \frac{|c_{\text{cp}} - c_{\text{ном}}|}{c_{\text{ном}}} * 100, \quad (14)$$

где $c_{\text{ном}}$ – объемная доля (%) этанола в образце, определенная по процедуре приготовления.

Относительную погрешность приготовления растворов с использованием пипеточного дозатора рассчитывали по формуле

$$\Theta_{\partial} = \frac{\sqrt{\sum_{j=1}^M (\Theta_j)^2}}{V_{\partial}} * 100, \quad (15)$$

где Θ_j – погрешность j -го дозирования жидкости пипеточным дозатором, равная половине цены деления дозатора; M – число дозирования, необходимых для приготовления раствора; V – объем дозирования, мкл.

Для количественного анализа веществ в многокомпонентных смесях использовался хемометрический подход. Была построена многомерная градуировочная модель регрессии на основе алгоритма PLS-R.

По градуировочному набору спектров исследуемых смесей строилась многомерная градуировка, а затем производилась процедура кросс-валидации. Дальнейшая проверка градуировки проводилась с помощью набора спектров, которые не использовались в обучающем наборе, что позволяет получить более объективную оценку точности.

Все спектры, используемые при построении и проверке многомерной градуировки, прошли предварительную обработку фильтром Савицкого-Голея (размер окна – 15, степень полинома – 2) и процедуру нормировки. Фильтр Савицкого-Голея эффективно сглаживает спектральные данные, сохраняя при этом форму пиков. Как упоминалось выше, информативный фрагмент спектров для (смеси спиртов – от 100 до 3000 см^{-1} , а для смеси сахаров – от 117 до 1700 см^{-1}). Для обеих серий анализа было определено оптимальное количество главных компонент – 7, что обеспечивало наилучшее соотношение между точностью модели и её устойчивостью к переобучению.

Среднее значение определяемой объемной доли индивидуальных компонентов, относительное среднеквадратическое отклонение случайной составляющей погрешности определения объемной доли, относительную систематическую составляющую погрешности определения объемной доли и относительную погрешность приготовления растворов с использованием пипеточного дозатора вычисляли по формулам (11-15).

Относительную погрешность взвешивания определяли по формуле

$$\Theta_{\text{в}} = \frac{\theta_{\text{в}}}{m_{\text{н}}} * 100, \quad (16)$$

где $\theta_{\text{в}}$ – погрешность взвешивания, равная пределу допускаемой погрешности весов в эксплуатации; $m_{\text{н}}$ – масса навески.

Относительную погрешность приготовления растворов рассчитывали по формуле

$$\Theta = \sqrt{\Theta_{\text{в}} + \Theta_{\text{д}}}, \quad (17)$$

Относительную погрешность определения объемной доли компонентов в неизвестном образце при коэффициенте охвата $k = 2$ рассчитывали по формуле

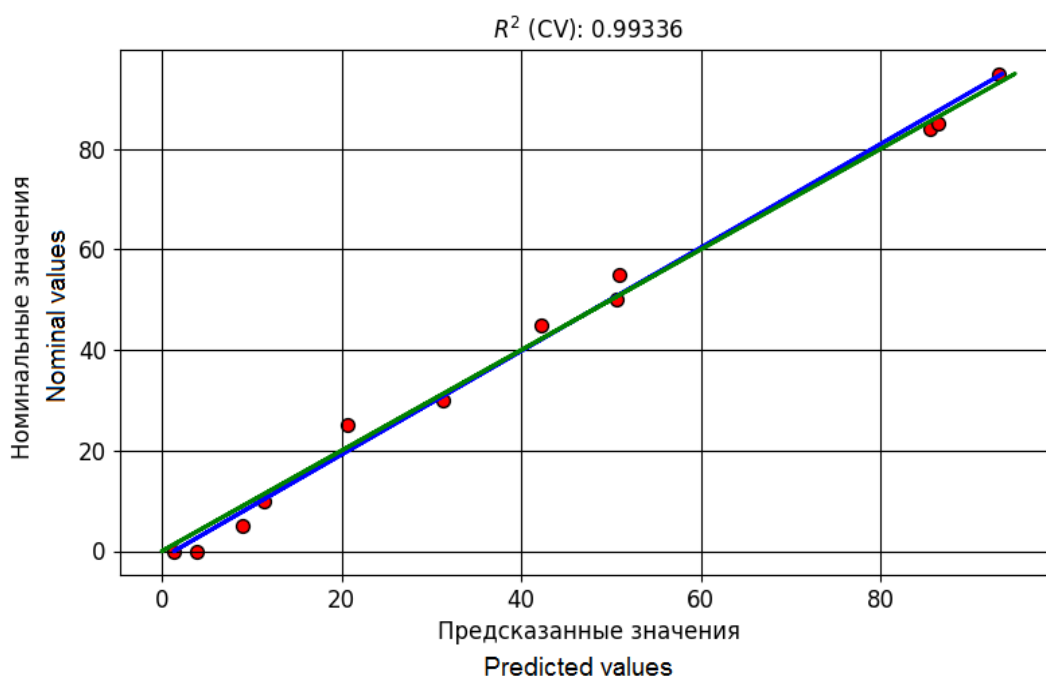
$$S_o = \sqrt{S^2 * t + \Delta^2 + \Theta^2}, \quad (18)$$

где t – коэффициент Стьюдента, соответствующий доверительной вероятности $P = 0,95$ и числу степеней свободы $n-1$, где n – число измерений.

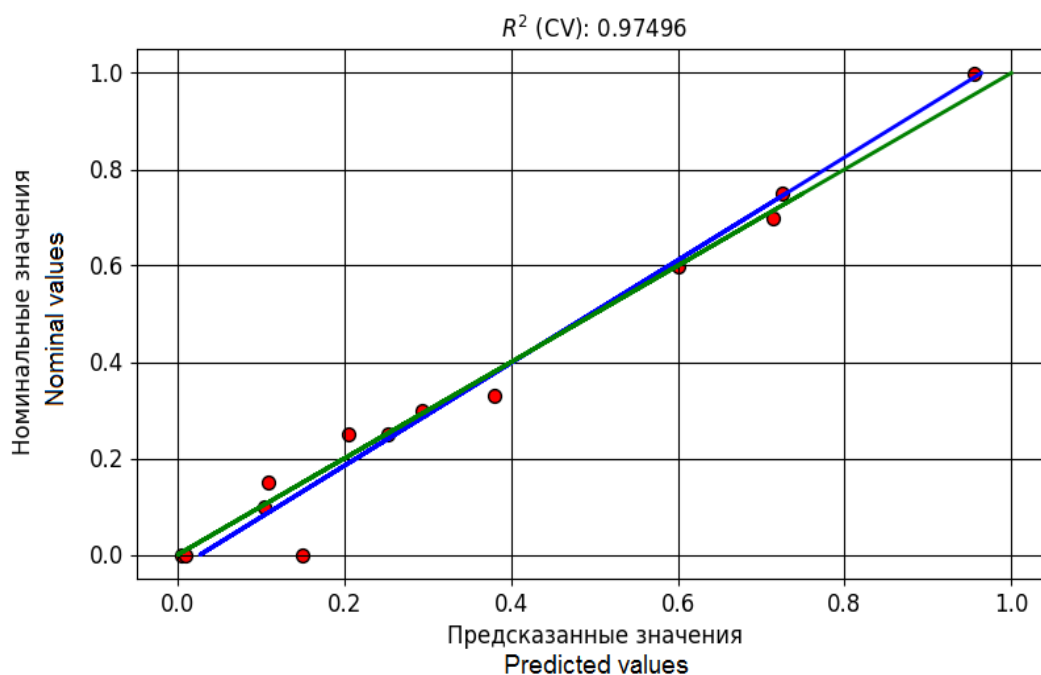
Такой подход позволяет учесть как случайные, так и систематические погрешности метода, давая более полную характеристику его точности.

Описанные выше операции производились с помощью программы обработки спектров веществ и материалов методами хемометрики и МО «ИнтелСпектр», разработанный диссертантом и описанной в главе 2 диссертации.

Результаты кросс-валидации представлены на рисунке 16.



a)



б)

а – для смеси спиртов; б – для смеси сахаров
Рисунок 16 – Результаты кросс-валидации

Графики демонстрируют хорошее соответствие между реальными (номинальными) и предсказанными значениями объемных долей в смесях, что подтверждает адекватность построенных моделей. Точки на графиках располагаются близко к диагональной линии, соответствующей идеальному соответствию предсказанных и реальных значений.

Далее разработанным моделям предъявлялись тестовые наборы данных. Результаты этой проверки для смеси спиртов представлены в таблице 21, а для смеси сахаров – в таблице 22.

Таблица 21 – Результаты предсказания объемных долей компонентов в смеси спиртов для тестового набора, по отнормированным КР-спектрам

Результаты предсказания, объемные доли %	Этанол	Метанол	Вода
Номинальная	50	30	20
Предсказанная	48,6	31,1	23,7
Относительная погрешность предсказания, %	5,7	4,2	8,1

Таблица 22 – Результаты предсказания объемных долей компонентов в смеси сахаров для тестового набора, по отнормированным КР-спектрам

Результаты предсказания, объемные доли %	Фруктоза	Глюкоза	Сахароза
Номинальная	10	75	15
Предсказанная	10,8	71,9	17
Относительная погрешность предсказания	9,1	7,1	14,3

Как видно из приведенных данных, с помощью градуировки, построенной по КР-спектрам смеси спиртов, концентрация отдельных компонентов определяется с относительной погрешностью, не превышающей 8,1 %. Это весьма хороший результат, учитывая сложность анализируемой системы и перекрывание спектральных линий компонентов. Многомерная градуировка, построенная по КР-спектрам смеси сахаров, позволяет определять концентрацию компонентов в неизвестном образце с относительной погрешностью, не превышающей 14,3 %. Более высокое значение погрешности для смеси сахаров объясняется большей сложностью этой системы (четыре компонента вместо трёх) и большим структурным сходством молекул сахаров, что приводит к значительному перекрыванию их спектральных линий.

Следует отметить, что для построения градуировки и измерения спектров КР анализируемых образцов разных приборов необходимо учитывать отношение толщин конфокальных областей этих приборов. Разработанный метод в дальнейшем будет усовершенствован за счет введения поправок на толщину конфокальной области, что позволит использовать градуировки, построенные с помощью одного прибора, для определения концентраций по спектрам КР, измеренным на других КР-спектрометрах или микроскопах. Это существенно расширит возможности метода и сделает его более универсальным.

5.2 Градуировка фотометрической шкалы в приборно-независимых единицах с помощью образца сравнения в виде раствора циклогексана

В данном подразделе рассмотрена возможность градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах при использовании в качестве образца сравнения разработанного недавно в ФГБУ

«ВНИИОФИ» стандартного образца характеристик спектра КР света циклогексана ГСО 12886-2025. Этот государственный стандартный образец (ГСО) представляет собой циклогексан с массовой долей основного компонента 99,9 %. При этом была использована характерная КР-линия при значении волнового числа $801,3 \text{ см}^{-1}$. Метрологические характеристики стандартного образца указаны в таблице 23 [96].

Таблица 23 – Метрологические характеристики стандартного образца циклогексана

Наименование характеристики	Значение характеристики
Диапазон волновых чисел рамановских сдвигов, см^{-1}	от 380 до 3000
Попускаемое значение абсолютной расширенной неопределенности аттестованного значения волновых чисел при $k = 2$ и $P = 0,95$, см^{-1}	$\pm 1,0$
Значения относительных интенсивностей сигналов (линий комбинационного рассеяния света) при длинах волн возбуждения 532, 633 и 785 нм, %	от 1,0 до 100,0
Допускаемое значение абсолютной расширенной неопределенности аттестованного значения относительных интенсивностей сигналов (линий комбинационного рассеяния света) при длинах волн возбуждения 532, 633 и 785 нм при $k = 2$ и $P = 0,95$, %, не более	1,0

В качестве пригодно-независимых единицы интенсивности КР предлагается принять отношение интенсивности линии КР анализируемого вещества к интенсивности ближайшей к нему по шкале волновых чисел линии циклогексана. При этом необходимо учитывать различие в спектральных чувствительностях сравниваемых приборов при волновых числах, соответствующих циклогексану и аналиту.

Для экспериментальной проверки предложенного подхода были сравнены измеренные на двух разных приборах интенсивности линии КР этанола (880 см^{-1}), нормированной на интенсивность ближайшей к ней линии циклогексана ($801,3 \text{ см}^{-1}$) в нескольких водно-спиртовых смесях.

Для данного эксперимента были приготовлены следующие смеси с объемной долей этанола (20 %, 30 %, 50 %, 70 % и 100 %).

Измерения КР-спектров проводились на двух приборах:

1. Раман-люминесцентный портативный спектрометр «ИнСпектр» R532 («Спектр-М», Россия), описание метрологических и технических характеристик которого приведено в разделе 3.1;

2. Портативно рамановский спектрометр RL 532 («ФОТОН-БИО», Россия), метрологические технические характеристики которого указаны в таблице 24.

Таблица 24 – Метрологические и технические характеристики портативного рамановского спектрометр «ФОТОН-БИО» RL 532

Наименование характеристики	Значение характеристики
Спектральный диапазон, см^{-1}	150-4000
Спектральное разрешение, см^{-1}	4-6
Предел допускаемого значения абсолютной погрешности измерения волнового числа, см^{-1}	2
Максимальная мощность лазерного излучения, мВт, не менее	35
Габаритные размеры, (Д×Ш×В), мм, не более	300×200×80
Масса, кг, не более	4
Условия эксплуатации температура окружающей среды, °C относительная влажность воздуха, % атмосферное давление, кПа	15-35 50-80 97-105
Потребляемая мощность, В·А, не более	80

Оптимальные параметры измерения были подобраны экспериментально для обеспечения высокого соотношения сигнал/шум при разумной продолжительности измерения. Для каждого образца проводилось пять параллельных измерений с их дальнейшим усреднением.

Для каждого усредненного спектра были вычислены отношения интенсивностей характеристического пика этанола к интенсивности близко расположенного характеристического пика стандартного образца циклогексана, а также отношения двух пиков циклогексана.

На рисунке 17 продемонстрированы усредненные фрагменты спектров стандартного образца циклогексана на двух приборах («ИнСпектр» R532 и «ФОТОН-БИО» RL 532). Видно, что отношения интенсивностей пиков при значениях волновых чисел $801,3 \text{ см}^{-1}$ и $1028,3 \text{ см}^{-1}$ существенно различаются для КР-спектров, измеренных на разных приборах, что обусловлено различием их спектральных чувствительностей.

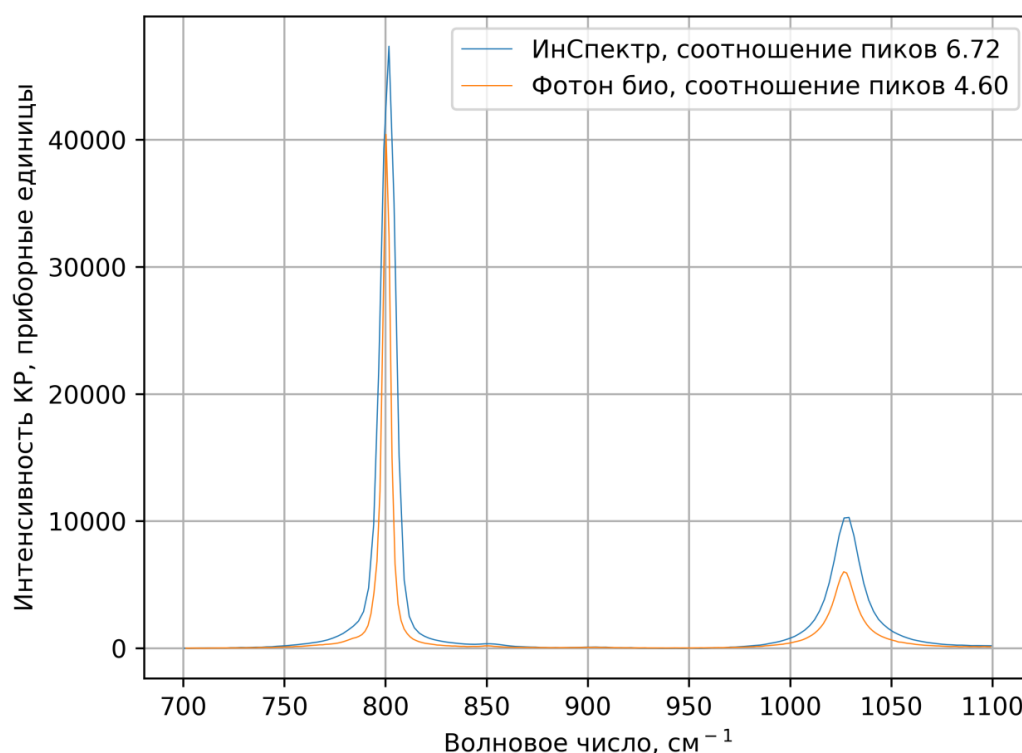


Рисунок 17 – Фрагменты КР-спектров циклогексана, измеренных на приборах «ИнСпектр» R532 и «ФОТОН-БИО» RL 532

Для сравнения в спектре этанола была выбрана линия 880 см^{-1} , интенсивность которой нормировалась на интенсивность линии $801,3 \text{ см}^{-1}$ циклогексана. Для оценки поправки на различие в отношениях спектральных чувствительностей приборов на этих длинах волн использовались спектры КР чистого этанола.

Необходимые для оценки значения интенсивностей линий КР приведены в таблице 25.

Таблица 25 – Данные для оценки различия в спектральных чувствительностях КР-спектрометров

Прибор	Этанол, 880 см ⁻¹	
	интенсивность в приборных единицах	интенсивность, нормированная на линию 801,3 см ⁻¹ циклогексана
«ИнСпектр» R 532	10104,06	0,213
«ФОТОН-БИО» RL 532	5636,39	0,139
отношение нормированных интенсивностей		1,532

Полученное по данным для чистого этанола отношение нормированных интенсивностей линии КР (880 см⁻¹) можно использовать в качестве коэффициента коррекции нормированных интенсивностей той же линии в спектрах водно-спиртовых смесей, измеренных на спектрометре «ФОТОН-БИО» RL 532, при их сопоставлении с соответствующими интенсивностями, измеренными на спектрометре «ИнСпектр» R 532. Результаты сопоставления приведены в таблице 26.

Таблица 26 – Сопоставление интенсивностей линии КР (880 см⁻¹) водно-спиртовых смесей, нормированных на интенсивность линии (801,3 см⁻¹) циклогексана, измеренных на спектрометрах «ФОТОН-БИО» RL 532 и «ИнСпектр» R 532

Объёмная доля спирта в водно-спиртовой смеси, %	Нормированная интенсивность по данным «ФОТОН-БИО» RL 532	Нормированная интенсивность по данным «ФОТОН-БИО», умноженная на коэффициент коррекции K=1,532	Нормированная интенсивность по данным «ИнСпектр» R 532	Относительная погрешность нормированных интенсивностей, %
20	0,021	0,032	0,036	5,88
30	0,034	0,052	0,060	7,14
50	0,064	0,098	0,099	0,51
70	0,083	0,127	0,140	4,87

Приведенная в последнем столбце таблицы относительная погрешность Δ оценки интенсивностей линий КР в единицах, нормированных на интенсивность ближайшей линии циклогексана, оценивалась по формуле

$$\Delta = \left| \frac{I_{\text{ФОТОН-БИО}} - I_{\text{ИнСпектр}}}{I_{\text{ФОТОН-БИО}} + I_{\text{ИнСпектр}}} \right| * 100. \quad (19)$$

В рассмотренном примере интенсивности линии КР аналита, нормированные на интенсивность ближайшей КР-линии циклогексана, мало зависят от особенностей приборов при условии коррекции различий в чувствительностях приборов на линии аналита.

Таким образом, нормирование на спектр циклогексана может быть использовано для градуировки фотометрической шкалы КР-спектрометров в аппаратно-независимых единицах. Примером практического использования такой градуировки являются градуировочные графики для водно-спиртовых смесей, построенные по данным, полученным на двух рассмотренных выше приборах, которые продемонстрированы на рисунке 18. Видно, что между этими графиками существуют незначительные отличия, и значения объемных долей этанола, определенные по этим графикам, различаются между собой не более чем на 10 %.

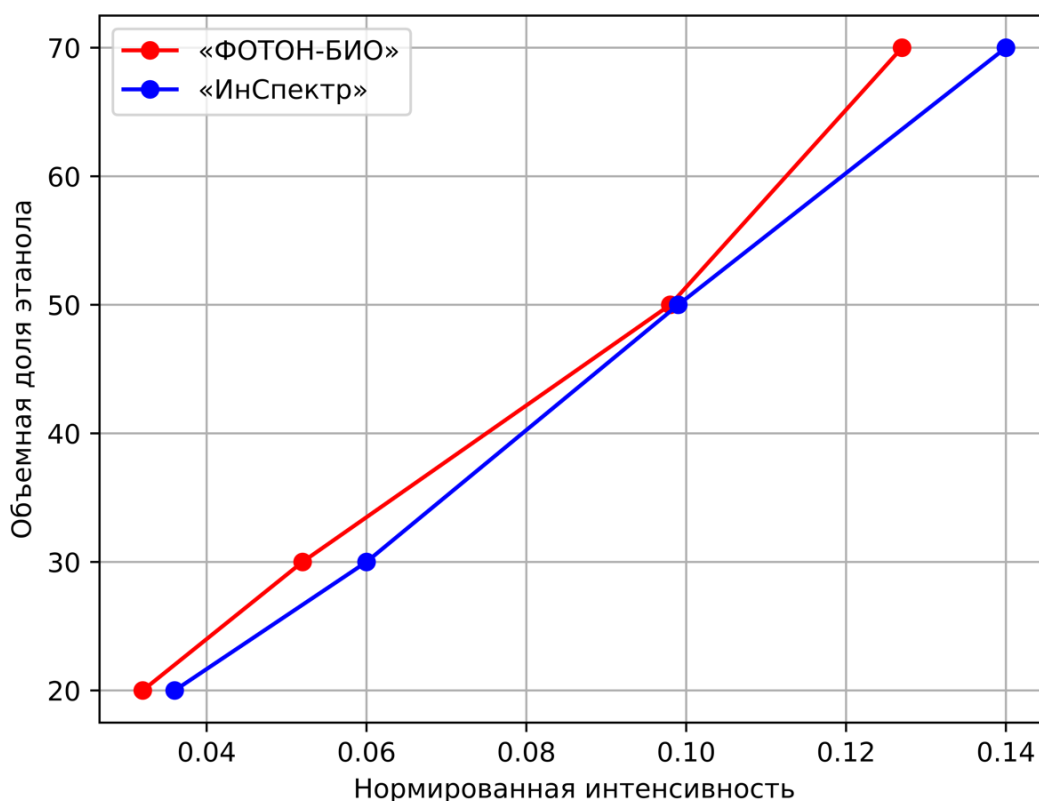


Рисунок 18 – Градуировочные графики для водно-спиртовых смесей, измеренных на приборах «ИнСпектр» R532 и «ФОТОН-БИО» RL 532

Выводы к главе 5

Полученные в главе 5 результаты имеют высокую практическую значимость, поскольку открывают новые возможности для применения КР-спектроскопии в количественном анализе.

Каждый из рассмотренных вариантов образцов сравнения для градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах имеет свои преимущества и недостатки. Использование меры волновых чисел рамановских сдвигов позволяет использовать один и тот же образец сравнения в рамках одной лаборатории в течение длительного времени, однако менее удобно при сопоставлении данных, полученных в разных лабораториях. Кроме того, может потребоваться введение поправок на различие толщин конфокальных объемов приборов. Использование циклогексана в качестве образца сравнения не требует введения таких поправок, однако предполагает определенную пробоподготовку с использованием токсичной жидкости. С другой стороны, наличие утвержденного ГСО характеристик спектра КР циклогексана дает возможность ввести приборно-независимые относительные единицы интенсивности КР с нормированием на линии циклогексана, аналогично упоминавшимся выше относительным единицам флуоресценции (ОЕФ). Такие единицы могут быть востребованы при количественном анализе с использованием спектров КР.

Разработанный метод градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах вносит вклад в обеспечение единства измерений в молекулярной оптической спектроскопии, давая возможность сопоставлять результаты измерений, полученные на разных приборах и в разных лабораториях.

Также результаты убедительно доказывают, что количественный анализ на основе КР-спектров с дальнейшей обработкой данных методом хемометрики эффективен. При этом метод хемометрики PLS существенно расширяет аналитические возможности КР-спектроскопии, позволяя определять

индивидуальные компоненты в сложных многокомпонентных смесях без их предварительного разбавления.

В целом, проведенное исследование демонстрирует перспективность применения КР-спектроскопии в сочетании с хемометрическими методами обработки данных для количественного анализа сложных многокомпонентных смесей.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований разработаны и усовершенствованы методы оптического спектрального анализа с хемометрической обработкой данных на примере конкретных объектов исследования. При этом были выработаны следующие научные и технические решения.

1. Разработан и апробирован метод классификации частиц черного углерода по источникам выбросов по данным об оптической плотности на различных длинах волн. Данную модель планируется включить в состав разрабатываемого специалистами ФГБУ «ВНИИОФИ» ПО для аэталометра.

2. Разработан и апробирован метод градуировки фотометрической шкалы КР-спектрометров в приборно-независимых единицах с помощью образцов сравнения, а также этот метод был использован для количественного анализа многокомпонентных смесей.

3. Разработан и апробирован метод классификации алкогольной продукции по географическому происхождению и возрасту выдержки на основе спектров флуоресценции и КР с последующей обработкой данных методом экстремального градиентного бустинга.

4. Проведены экспериментальные исследования предложенных методов, оценена достоверность и точность получаемых результатов.

5. Выбраны и обоснованы алгоритмы для обработки оптических спектров, на основе этих алгоритмов разработан и апробирован прикладной пакет ПО. Все результаты обработки спектров, приведенные в диссертации, получены с помощью данного ПО.

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

КР – Комбинационное рассеяние

ПО – Программное обеспечение

МО – Машинное обучение

PC (Principal Components) – Главные компоненты

PCA (Principal Components Analysis) – Метод главных компонент

LDA (Linear discriminat analysis) – Линейный дискриминантный анализ

QDA (Quadratic discriminant analysis) – Квадратичный дискриминантный анализ

PLS (Partial Least Squares) – Метод частичных наименьших квадратов

PLS-DA (Partial Least Squares Discriminant Analysis) – Метод частичных наименьших квадратов, дискриминантный анализ

PLS-R (Partial Least Squares Regression) – Метод частичных наименьших квадратов, регрессия

PCR (Principal Component Regression) – Регрессия на главные компоненты

ЭГБ – Экстремальный градиентный бустинг

MSE (mean squared error) – Среднеквадратичная ошибка

ФЭУ – Фотоэлектронные умножители

ПЗС (CCD, Charge-Coupled Device) – Прибор с зарядовой связью

ФИФ ОЕИ – Федеральный информационный фонд по обеспечению единства измерений

СПИСОК ЛИТЕРАТУРЫ

1. Саакян А. В., Левин А. Д. Программное обеспечение для обработки спектральных данных методами хеометрики и машинного обучения // Аналитика: научно-технический журнал. — 2024. — Т. 14, № 2. — С. 154–160. — DOI: <https://doi.org/10.22184/2227-572X.2024.14.2.154.160>.
2. Юшина А. А., Аленичев М. К., Саакян А. В., Левин А. Д. Мера волновых чисел рамановских сдвигов и возможности ее применения для количественного анализа // Эталоны. Стандартные образцы. — 2025. — Т. 21, № 1. — С. 22–37. — DOI: <https://doi.org/10.20915/2077-1177-2025-21-1-22-37>.
3. Саакян А. В., Аленичев М. К., Левин А. Д. Характеризация коньяков и виноградных бренди по спектрам флуоресценции, обработанным с помощью методов машинного обучения // Заводская лаборатория. Диагностика материалов. — 2023. — Т. 89, № 10. — С. 25–33. — DOI: <https://doi.org/10.26896/1028-6861-2023-89-10-25-33>.
4. Саакян А. В., Юшина А. А., Левин А. Д. Классификация бренди и коньячной продукции по географическому происхождению и сроку выдержки с использованием спектроскопии комбинационного рассеяния и машинного обучения // Измерительная техника. — 2023. — № 3. — С. 33–38. — DOI: <https://doi.org/10.32446/0368-1025it.2023-3-33-38>.
5. Свидетельство о государственной регистрации программы для ЭВМ «Программа обработки спектров веществ и материалов методами хеометрики и машинного обучения (ИнтелСпектр)» № 202468048 29.08.2024.
6. Лакович Д. Основы флуоресцентной спектроскопии. – М.: Мир, 1986. – 496 с.
7. Valeur B., Berberan-Santos M. N. Molecular fluorescence: principles and applications. – 2nd ed. – Weinheim: Wiley-VCH, 2013. – 592 p. DOI: 10.1002/9783527650002.

8. Gore M. (ed.). Spectrophotometry and spectrofluorimetry: a practical approach. – 2nd ed. – Oxford: OUP Oxford, 2000. – 368 p. – (Practical Approach Series; vol. 225). DOI: 10.1093/oso/9780199638130.001.0001.
9. Phillips D., Drake R. C., O'Connor D. V., Christensen R. L. Time correlated single-photon counting (TCSPC) using laser excitation // Instrumentation Science & Technology. – 1985. – Vol. 14, no. 3–4. – P. 267–292. DOI: 10.1080/10739148508543581.
10. Demchenko A. P. (ed.). Advanced Fluorescence Reporters in Chemistry and Biology III: Applications in Sensing and Imaging. – Berlin; Heidelberg: Springer, 2011. – 352 p. DOI: 10.1007/978-3-642-18035-4.
11. Crawford S. E., Ohodnicki P. R., Baltrus J. P. Materials for the photoluminescent sensing of rare earth elements: challenges and opportunities // Journal of Materials Chemistry C. – 2020. – Vol. 8, no. 24. – P. 7975–8006. DOI: 10.1039/D0TC00772A.
12. Kaur H., Nguyen K., Kumar P. Pressure and temperature dependence of fluorescence anisotropy of green fluorescent protein // RSC Advances. – 2022. – Vol. 12, no. 14. – P. 8647–8655. DOI: 10.1039/D2RA00430B.
13. Michel B. Y., Strakova K., Larsen E., Wenger J., Vestergaard M., Mokhir A., Verga D., Boutorine A. S., Nakanishi M., Saito Y., Shimizu T., Tor Y., Vester B., Teulade-Fichou M.-P. Probing of nucleic acid structures, dynamics, and interactions with environment-sensitive fluorescent labels // Frontiers in Chemistry. – 2020. – Vol. 8. – Art. 112. DOI: 10.3389/fchem.2020.00112.
14. Mei J., Leung N. L. C., Kwok R. T. K., Lam J. W. Y., Tang B. Z. Aggregation-induced emission: the whole is more brilliant than the parts // Advanced Materials. – 2014. – Vol. 26, no. 31. – P. 5429–5479. DOI: 10.1002/adma.201401356.
15. Tamarit F. C., Mirabello V., Pascu S. I. Oxygen sensing, hypoxia tracing and in vivo imaging with functional metalloprobes for the early detection of non-communicable diseases // Frontiers in Chemistry. – 2018. – Vol. 6. – Art. 27. DOI: 10.3389/fchem.2018.00027.

16. Zehra N., Creran B., Rotello V. M. AIE active polymers for biological applications // *Progress in Molecular Biology and Translational Science*. – 2021. – Vol. 185. – P. 137–177. DOI: 10.1016/bs.pmbts.2021.01.006.
17. Weber G. Polarization of the fluorescence of macromolecules. I. Theory and experimental method // *Biochemical Journal*. – 1952. – Vol. 51, no. 2. – P. 145–155. DOI: 10.1042/bj0510145.
18. Lakowicz J. R. Principles of fluorescence spectroscopy. – 3rd ed. – Boston, MA: Springer US, 2006. – 954 p. DOI: 10.1007/978-0-387-46312-4.
19. Croce A. C. Light and autofluorescence, multitasking features in living organisms // *Photochem.* – 2021. – Vol. 1, no. 2. – P. 67–124. DOI: 10.3390/photochem1020007.
20. Esfandiari N., Aliofkhazraei M. Advances in the determination of trace amounts of iron cations through electrochemical methods: a comprehensive review of principles and their limits of detection // *Talanta*. – 2024. – Vol. 277. – Art. 126365. DOI: 10.1016/j.talanta.2024.126365.
21. Shin Y. H., Gutierrez-Wing M. T., Choi J. W. Recent progress in portable fluorescence sensors // *Journal of The Electrochemical Society*. – 2021. – Vol. 168, no. 1. – Art. 017502. DOI: 10.1149/1945-7111/abd494.
22. Demello A. J. DNA amplification moves on // *Nature*. – 2003. – Vol. 422, no. 6927. – P. 28–29. DOI: 10.1038/422028a.
23. Esposito A., Venkitaraman A. R. Enhancing biochemical resolution by hyperdimensional imaging microscopy // *Biophysical Journal*. – 2019. – Vol. 116, no. 10. – P. 1815–1822. DOI: 10.1016/j.bpj.2019.04.009.
24. González A. G., Herrador M. Á. A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles // *TrAC Trends in Analytical Chemistry*. – 2007. – Vol. 26, no. 3. – P. 227–238. DOI: 10.1016/j.trac.2007.01.009.
25. Kwon N., Hu Y., Yoon J. Fluorescent chemosensors for various analytes including reactive oxygen species, biothiol, metal ions, and toxic gases // *ACS Omega*. – 2018. – Vol. 3, no. 10. – P. 13731–13751. DOI: 10.1021/acsomega.8b01875.

26. Шмидт В. Оптическая спектроскопия для химиков и биологов / В. Шмидт. – М. : Техносфера, 2007. – 368 с.
27. McCreery R. L. Raman spectroscopy for chemical analysis. – Hoboken, NJ: John Wiley & Sons, 2005. – 475 p. DOI: 10.1002/0471721646.
28. Ossikovski R., Nguyen T. T., Picardi G., Légaré F., Sorriso-Valvo L. Raman spectroscopy and polarization: selected case studies // *Comptes Rendus Physique*. – 2012. – Vol. 13, no. 8. – P. 837–852. DOI: 10.1016/j.crhy.2012.09.003.
29. Keresztury G., Chalmers J. M., Griffiths P. R. Raman spectroscopy: theory // *Handbook of Vibrational Spectroscopy* / eds. J. M. Chalmers, P. R. Griffiths. – Chichester: John Wiley & Sons, 2002. – Vol. 1. – P. 71–87. DOI: 10.1002/0470027320.s8901.
30. Moester M. J. B., van Driel T. B., Gualda E. J., Santos S. I. C. O., Otto C., Herek J. L., Eikema K. S. E., Boller K.-J., Lee C. K. Stimulated Raman scattering microscopy with long wavelengths for improved imaging depth // *Journal of Raman Spectroscopy*. – 2019. – Vol. 50, no. 9. – P. 1321–1328. DOI: 10.1002/jrs.5641.
31. Асеев В. А., Бабкина А. Н., Миронов Л. Ю., Нурыев Р. К. Спектроскопические методы исследования материалов фотоники. – СПб.: Университет ИТМО, 2021. – 97 с. ISBN 978-5-94836-220-5.
32. Vulchi R., Morgunov V., Junjuri R., Bocklitz T. Artifacts and anomalies in Raman spectroscopy: a review on origins and correction procedures // *Molecules*. – 2024. – Vol. 29, no. 19. – P. 4748. DOI: 10.3390/molecules29194748.
33. Pelletier M. J. Quantitative analysis using Raman spectrometry // *Applied Spectroscopy*. – 2003. – Vol. 57, no. 1. – P. 20A–42A. DOI: 10.1366/000370203321165142.
34. Pezzotti G. Raman spectroscopy in cell biology and microbiology // *Journal of Raman Spectroscopy*. – 2021. – Vol. 52, no. 12. – P. 2348–2443. DOI: 10.1002/jrs.6204.
35. Ferraro J. R. *Introductory Raman Spectroscopy*. – 2nd ed. – Amsterdam: Elsevier, 2003. – 434 p. DOI: 10.1016/B978-012254105-6/50000-0.

36. Pitt G. D., Bérard M., Bowden M., Bowley R. M., Everall N. J., Gee W. J., Matousek P., Parker A. W., Priestnall C. M., Wright M. Engineering aspects and applications of the new Raman instrumentation // *IEE Proceedings – Science, Measurement and Technology*. – 2005. – Vol. 152, no. 6. – P. 241–318. DOI: 10.1049/ip-smt:20045065.
37. Houhou R., Bocklitz T. Trends in artificial intelligence, machine learning, and chemometrics applied to chemical data // *Analytical Science Advances*. – 2021. – Vol. 2, no. 3–4. – P. 128–141. DOI: 10.1002/ansa.202000140.
38. Кутаевских А. В., Муромцев Д. И., Кирсанова О. В. Классические методы машинного обучения: учебное пособие. – СПб.: Университет ИТМО, 2022. – 53 с. ISBN 978-5-4497-0269-2.
39. Esposito C., De Pietro G., De Santis A., Moscato F., Sperlí G. GHOST: adjusting the decision threshold to handle imbalanced data in machine learning // *Journal of Chemical Information and Modeling*. – 2021. – Vol. 61, no. 6. – P. 2623–2640. DOI: 10.1021/acs.jcim.0c01312.
40. Кострицкий С. М., Бухарин М. А., Коркишко Ю. Н., Федоров В. А., Скрябин Н. Н., Севостьянов О. Г., Чиркова И. М. Исследование изменений структуры кристаллов LiNbO₃ при фемтосекундной записи световодов // *Письма в ЖЭТФ*. – 2013. – Т. 98, № 2. – С. 106–110. DOI: 10.7868/S0370274X13140015.
41. Hubert M., Rousseeuw P. J., Vanden Branden K. ROBPCA: a new approach to robust principal component analysis // *Technometrics*. – 2005. – Vol. 47, no. 1. – P. 64–79. DOI: 10.1198/004017004000000563.
42. Sádecká J., Tóthová J., Májek P. Classification of brandies and wine distillates using front face fluorescence spectroscopy // *Food Chemistry*. – 2009. – Vol. 117, no. 3. – P. 491–498. DOI: 10.1016/j.foodchem.2009.04.026.
43. Maadooliat M., Huang J. Z., Hu J. Integrating data transformation in principal components analysis // *Journal of Computational and Graphical Statistics*. – 2015. – Vol. 24, no. 1. – P. 84–103. DOI: 10.1080/10618600.2013.842176.
44. Sádecká J., Tóthová J., Májek P., Čurlej J., Sádecký M. Comparison of different fluorescence techniques in brandy classification by region of production //

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. – 2019. – Vol. 216. – P. 125–135. DOI: 10.1016/j.saa.2019.02.040.

45. Suyal M., Goyal P. A review on analysis of k-nearest neighbor classification machine learning algorithms based on supervised learning // International Journal of Engineering Trends and Technology. – 2022. – Vol. 70, no. 7. – P. 43–48. DOI: 10.14445/22315381/IJETT-V70I7P207.

46. Kumar K. Partial least square (PLS) analysis: most favorite tool in chemometrics to build a calibration model // Resonance. – 2021. – Vol. 26, no. 3. – P. 429–442. DOI: 10.1007/s12045-021-1139-6.

47. Родионова О. Е., Померанцев А. Л. Хемометрика в аналитической химии [Электронный ресурс]: <https://files.znu.edu.ua/files/Bibliobooks/Sinyaeva/0032933.pdf> (дата обращения: 14.9.2025)

48. Piccioni Costa L., Guerreiro M., Puchta E., Tadano Y., Antonini Alves T., Kaster M., Siqueira H. Multilayer Perceptron // Introduction to Computational Intelligence / ed. by F. Gomide, W. Pedrycz. – IEEE Computational Intelligence Society Open Book, 2023. – Ch. 9. – P. 105–122. DOI: 10.1109/9781665491786.ch9.

49. Piccioni Costa L., Guerreiro M., Puchta E., Tadano Y., Antonini Alves T., Kaster M., Siqueira H. Multilayer Perceptron [Электронный ресурс] // Introduction to Computational Intelligence / ed. by F. Gomide, W. Pedrycz. – IEEE Computational Intelligence Society Open Book, 2023. – Ch. 9. – P. 105–122. DOI: 10.1109/9781665491786.ch9: <https://neptune.ai/blog/xgboost-everything-you-need-to-know> (дата обращения: 14.9.2025)

50. Salman H. A., Kalakech A., Steiti A. Random forest algorithm overview // Babylonian Journal of Machine Learning. – 2024. – Vol. 2024. – P. 69–79. DOI: 10.58496/BJML/2024/007.

51. Evgeniou T., Pontil M. Support vector machines: theory and applications // Advanced Course on Artificial Intelligence / ed. by P. Brazdil. – Berlin; Heidelberg: Springer, 1999. – P. 249–257. – (Lecture Notes in Computer Science; vol. 2049). DOI: 10.1007/3-540-44673-7_12.

52. Sarker I. H. Machine learning: algorithms, real-world applications and research directions // SN Computer Science. – 2021. – Vol. 2, no. 3. – Art. 160. DOI: 10.1007/s42979-021-00592-x.
53. Miller C., Portlock T., Nyaga D. M., O’Sullivan J. M. A review of model evaluation metrics for machine learning in genetics and genomics // Frontiers in Bioinformatics. – 2024. – Vol. 4. – Art. 1457619. DOI: 10.3389/fbinf.2024.1457619.
54. Basalekou M., Kyraleou M., Kallithraka S. Authentication of wine and other alcohol-based beverages – future global scenario // Future Foods. – Amsterdam: Academic Press, 2022. – P. 669–695. DOI: 10.1016/B978-0-12-822715-8.00017-9.
55. Jolliffe I. T., Cadima J. Principal component analysis: a review and recent developments // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2016. – Vol. 374, no. 2065. – Art. 20150202. DOI: 10.1098/rsta.2015.0202.
56. Barton B., Bartick E. G., Bobiak J. A., Bocklitz T. W., Deegan R. D., Emmons E. D., Guo J., Joester D., Kodali A. K., Lavine B. K., McLaughlin G., Miller A. D., Phillips M. C., Podstawka E., Santos M. C. D., Sikirzhytski V., Treado P. J., Trevisan J., Valentine N. B., Zampronio C. G., Zhao Y. Chemometrics for Raman spectroscopy harmonization // Applied Spectroscopy. – 2022. – Vol. 76, no. 9. – P. 1021–1041. DOI: 10.1177/00037028221089943.
57. Luo R., Popp J., Bocklitz T. Deep learning for Raman spectroscopy: a review // Analytica. – 2022. – Vol. 3, no. 3. – P. 287–301. DOI: 10.3390/analytica3030020.
58. Ельчищева Ю. Б. Спектрофотометрические методы анализа: учеб. пособие [Электронный ресурс]. – Пермь: ПГНИУ, 2023. – 188 с: <http://www.psu.ru/files/docs/science/books/uchebnie-posobiya/Elchishcheva-Spektrofotometricheskie-metody-analiza.pdf> (дата обращения: 14.9.2025)
59. Ranaweera R. K. R., Gilmore A. M., Capone D. L., Bastian S. E. P., Jeffery D. W. A review of wine authentication using spectroscopic approaches in combination with chemometrics // Molecules. – 2021. – Vol. 26, no. 14. – Art. 4334. DOI: 10.3390/molecules26144334.

60. Murru C., Vlahopoulou G., Milella A., Arca P., Sanna G. Artificial neural network and attenuated total reflectance-Fourier transform infrared spectroscopy to identify the chemical variables related to ripeness and variety classification of grapes for Protected Designation of Origin wine production // *Computers and Electronics in Agriculture*. – 2019. – Vol. 164. – Art. 104922. DOI: 10.1016/j.compag.2019.104922.
61. Cosme F., Nunes J. M., Filipe-Ribeiro L., Reis S., Moreira A. S. P., Malheiro A. C., Coimbra M. A. Authentication of Douro DO monovarietal red wines based on anthocyanin profile: comparison of partial least squares–discriminant analysis, decision trees and artificial neural networks // *Food Control*. – 2021. – Vol. 125. – Art. 107979. DOI: 10.1016/j.foodcont.2021.107979.
62. Coetzee P. P., Van Jaarsveld F. P., Vanhaecke F. Intraregional classification of wine via ICP-MS elemental fingerprinting // *Food Chemistry*. – 2014. – Vol. 164. – P. 485–492. DOI: 10.1016/j.foodchem.2014.05.061.
63. Dufour E., Letort A., Laguet A., Lebecque A., Serra J. N. Investigation of variety, typicality and vintage of French and German wines using front-face fluorescence spectroscopy // *Analytica Chimica Acta*. – 2006. – Vol. 563, no. 1–2. – P. 292–299. DOI: 10.1016/j.aca.2005.11.083.
64. Viejo C., Fuentes S., Torrico D., et al. Assessment of beer quality based on foamability and chemical composition using computer vision algorithms, near infrared spectroscopy and machine learning algorithms // *Journal of the Science of Food and Agriculture*. – 2018. – Vol. 98, No. 2. – P. 618–627. – DOI: 10.1002/jsfa.8506.
65. Basalekou M., Kyraleou M., Kallithraka S. Authentication of wine and other alcohol-based beverages – future global scenario // In: *Future Foods: Global Trends, Opportunities, and Sustainability Challenges* / eds. D. Barjolle, J. Lengyel, S. Vecchio. – Amsterdam: Elsevier, 2022. – P. 669–695. DOI: 10.1016/B978-0-323-91001-9.00028-1.
66. Cádecká J., Uričková V., Jakubíková M. Fluorescence Spectroscopy for the Analysis of Spirit Drinks // *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences* / ed. by Mark T. Stauffer. – London:

IntechOpen, 2016. – DOI: 10.5772/64002. – [Электронный ресурс]: <https://www.intechopen.com/chapters/51194> (дата обращения: 14.9.2025)

67. Lenhardt L., Bro R., Zeković I., Dramićanin T., Dramićanin M. D. Fluorescence spectroscopy coupled with PARAFAC and PLS-DA for characterization and classification of honey // *Food Chemistry*. – 2015. – Vol. 175. – P. 284–291. DOI: 10.1016/j.foodchem.2014.11.127.

68. Sádecká J., Uričková V. Fluorescence spectroscopy for the analysis of spirit // In: *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences* / ed. M. Uddin. – Rijeka: InTech, 2016. – P. 339–357. DOI: 10.5772/64293.

69. Berghian-Grosan C., Magdas D. A. Influence of storage factors on wine quality: spectroscopic analysis // *Scientific Reports*. – 2020. – Vol. 10. – Art. 21152. DOI: 10.1038/s41598-020-78159-8

70. rudi, L., Giardino, M., Janner, D., Pognant, F., Matera, F., Sacco, M., Bellopede, R. Black carbon characterization with Raman spectroscopy and machine learning techniques: first results for urban and rural areas // *Proceedings of the International Conference on Environmental Science and Technology (CEST2023)*. – 2023. DOI: 10.30955/gnc2023.00088.

71. Smith R. K., Jones D. A. Chemometrics in pharmaceutical analysis // *Journal of Pharmaceutical and Biomedical Analysis*. – 1988. – Vol. 6, no. 7. – P. 585–590. DOI: 10.1016/0731-7085(88)80067-0

72. Sen S., Shi K., Li S., Zhang L., Zhang S. Utilizing machine learning algorithms for precise discrimination of glycosuria in diabetes mellitus using fluorescence spectroscopy // *Analytica Chimica Acta*. – 2024. – Vol. 1234. – P. 1–9. DOI: 10.1016/j.aca.2024.04.048.

73. Zeng W., Wang Q., Xia Z., Li Z., Qu H. Application of XGBoost algorithm in the detection of SARS-CoV-2 using Raman spectroscopy // *Journal of Physics: Conference Series*. – 2021. – Vol. 1775, no. 1. – Art. 012007. DOI: 10.1088/1742-6596/1775/1/012007.

74. Eigenvector Research, Inc. PLS_Toolbox: Software for chemometrics and machine learning in MATLAB [Электронный ресурс]. – Manson, WA: Eigenvector Research, Inc., 2025: <https://eigenvector.com/software/pls-toolbox/> (дата обращения: 14.9.2025)
75. ChemHouse. ChemProject: software for chemometrics and machine learning [Электронный ресурс]. – Montpellier: ChemHouse, 2025: <https://www.chemproject.org/> (дата обращения: 14.9.2025)
76. Eigenvector Research, Inc. Solo: software for chemometrics and machine learning [Электронный ресурс]. – Manson, WA: Eigenvector Research, Inc., 2025: <https://eigenvector.com/software/solo/> (дата обращения: 14.9.2025)
77. B&W Tek, Inc. BWIQ: программное обеспечение для химометрического анализа спектральных данных [Электронный ресурс]. – Ньюарк, штат Делавэр: B&W Tek, Inc., 2025: <https://www.cz1.ru/catalog/spektr/bwtek/bwiq> (дата обращения: 14.9.2025)
78. Scikit-learn: machine learning in Python [Электронный ресурс]: <https://scikit-learn.org/stable/> (дата обращения: 14.9.2025)
79. XGBoost Documentation. XGBoost Python Package [Электронный ресурс]: https://xgboost.readthedocs.io/en/stable/python/python_intro.html (дата обращения: 14.9.2025)
80. Pandas Development Team. pandas: data analysis library for Python [Электронный ресурс]: <https://pandas.pydata.org> (дата обращения: 14.9.2025)
81. Matplotlib Development Team. Matplotlib: data visualization library for Python [Электронный ресурс]: <https://matplotlib.org/> (дата обращения: 14.9.2025)
82. Newcastle University. Coefficient of Determination (R-squared) [Электронный ресурс]: <https://www.mas.ncl.ac.uk/ask/numeracy-maths-statistics/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html> (дата обращения: 14.9.2025)
83. Liu F., Yon J., Fuentes A., Lobo P., Smallwood G. J., Corbin J. C. Review of recent literature on the light absorption properties of black carbon: Refractive index,

mass absorption cross section, and absorption functi // Aerosol Science and Technology. – 2020. – Vol. 54, no. 1. – P. 33–51. DOI: 10.1080/02786826.2019.1676878.

84. Федеральный информационный фонд по обеспечению единства измерений. Реестр средств измерений РФ № 69110-17: «Экспресс-анализаторы рамановские портативные «ИнСпектр». Описание типа средства измерения. [Электронный ресурс]: <https://fgis.gost.ru/fundmetrology/cm/mits/cd7d59d6-c490-b5b4-9468-eccc171b51ef> (дата обращения: 14.9.2025)

85. ФГБУ «Институт глобального климата и экологии имени академика Ю. А. Израэля». Доклад о прогрессе развития сети мониторинга чёрного углерода в России в 2022–2023 гг. и перспективных планах [Электронный ресурс]. — М.: ФГБУ «ИГКЭ им. акад. Ю. А. Израэля», 2023 http://downloads.igce.ru/for_main_site/Doklad.pdf (дата обращения: 14.9.2025).

86. Федеральный информационный фонд по обеспечению единства измерений, номер 46152-10: двухлучевой спектрофотометр Lambda 950. [Электронный ресурс]: <https://fgis.gost.ru/fundmetrology/cm/mits/7d536234-f154-26c8-06cc-27d65562684e> (дата обращения: 14.9.2025)

87. АО «Ставропольский винно-коньячный завод» [Электронный ресурс]: <http://www.stavvinprom.com/index.php/predpriyatiya/item/174> (дата обращения: 14.9.2025)

88. Закон Республики Армения «О алкогольных напитках на основе виноградного сырья», Глава 5, стр. 24–25 [Электронный ресурс]: <http://www.parliament.am/legislation.php?sel=show&ID=3348&lang=rus#5> (дата обращения: 14.9.2025)

89. Armenian Legal Information System [Электронный ресурс]: <https://www.arlis.am/DocumentView.aspx?DocID=77443> (дата обращения: 14.9.2025)

90. Кизлярский коньячный завод [Электронный ресурс]: <https://kizlyar-cognac.ru/производство> (дата обращения: 14.9.2025)

91. ВНИИОФИ. Спектрофотометр-флуориметр СФФ-2 «ФЛУОРАН». Реестр средств измерений РФ. № 51314-12. [Электронный ресурс]:

<https://fgis.gost.ru/fundmetrology/cm/mits/75d49668-42a5-e97b-d19f-059f83cc35f8>

(дата обращения: 14.9.2025)

92. Оселедцева И. В. Научное обоснование и развитие методологии контроля качества коньячных дистиллятов и коньяков: автореф. дис. ... д-ра техн. наук. – Краснодар, 2017. – 48 с. [Электронный ресурс]:

https://www.kubansad.ru/media/uploads/files/dissovet/avtoreferat/avto_oseledceva.pdf

(дата обращения: 14.9.2025)

93. Ranaweera R. K. R., Gilmore A. M., Capone D. L., Bastian S. E. P., Jeffery D. W. Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modeling // Food Chemistry. – 2021. – Vol. 335. – Art. 127592. DOI: 10.1016/j.foodchem.2020.127592.

94. Чугунова М. М. Метрологическое обеспечение флуоресцентного анализа в России и за рубежом // Метрология. – 2020. – № 4. – С. 52–68. DOI: 10.32446/0132-4713.2020-4-52-68.

95. Федеральный информационный фонд по обеспечению единства измерений, номер 93847-24: меры волновых чисел рамановских сдвигов. [Электронный ресурс]: <https://fgis.gost.ru/fundmetrology/cm/mits/15e75641-f288-118e-7703-0263eab3d1ce> (дата обращения: 14.9.2025)

96. ОЕИ-Аналитика. Сведения об утверждённом типе стандартных образцов: ГСО 12886-2025 «СО характеристик спектра комбинационного рассеяния света циклогексана». [Электронный ресурс]: <https://fgis.gost.ru/fundmetrology/registry/19/items/1426470> (дата обращения: 14.9.2025)

97. Юшина А. А., Аленичев М. К., Саакян А. В., Левин А. Д. Мера волновых чисел рамановских сдвигов и возможности ее применения для количественного анализа. VI Тезисы докладов. Международная научная конференция «Стандартные образцы в измерениях и технологиях» г. Екатеринбург, 2024, с 194-195.

98. Саакян А.В., Аленичев М.К., Левин А.Д. Классификация алкогольной продукции по спектрам флуоресценции с обработкой данных методом машинного

обучения. Тезисы докладов. V Международная научная конференция «Стандартные образцы в измерениях и технологиях» г. Екатеринбург, 2022, с 130

99. Саакян А. В., Юшина А. А., Левин А. Д. Классификация бренди и коньячной продукции по географическому происхождению и сроку выдержки с использованием спектроскопии комбинационного рассеяния и машинного обучения // Измерительная техника. — 2023. — № 3. — С. 33–38. — DOI: <https://doi.org/10.32446/0368-1025it.2023-3-33-38>.

100. Саакян А.В., Левин А.Д. Программный пакет для обработки спектральных данных методами хеометрики и машинного обучения. Тезисы докладов. VI Международная научная конференция «Метрология физико-химических измерений» г. Менделеево, 2023, с 75-78.

Приложение 1



УТВЕРЖДАЮ

Директор ФГБУ «ВНИИОФИ»

К. Ф. - М. Н.

И.С. Филимонов

«13» октября 2025 г.

АКТ

внедрения в Федеральном государственном бюджетном учреждении
«Всероссийский научно-исследовательский институт
оптико-физических измерений» (ФГБУ «ВНИИОФИ»)
результатов диссертации А.В. Саакяна
*«Развитие методов абсорбционной, флуоресцентной и рамановской
спектроскопии с хемометрической обработкой данных для анализа жидких
сред и аэрозолей»*,
представленной на соискание ученой степени кандидата технических наук

Настоящим актом подтверждается, что в ФГБУ «ВНИИОФИ» внедрен разработанный в диссертации А.В. Саакяна прикладной пакет программного обеспечения «Интел-Спектр», предназначенный для обработки спектральных данных методами хемометрики и машинного обучения. Указанный пакет был использован при обработке спектров оптического поглощения микрочастиц черного углерода, попадающих в воздух при неполном сгорании различных органических веществ. Измерения и обработка этих спектров проводились в рамках НИОКР по теме «Создание анализаторов для измерения концентрации микрочастиц черного углерода в атмосферном воздухе», выполняемой по договору с Федеральным государственным бюджетным учреждением «Институт глобального климата и экологии имени академика Ю.А. Израэля» (ФГБУ «ИГКЭ»), № 0373100106824000002 от 28 мая 2024 г.

Зам. начальника лаборатории аналитической
спектроскопии и метрологии наночастиц
ФГБУ «ВНИИОФИ», к.т.н.

В.И. Нагасев



УТВЕРЖДАЮ

Директор ФГБУ «ИГКЭ»

А.А. Романовская

«23» октября 2025 г.

АКТ

внедрения в Федеральном государственном бюджетном учреждении
«Институт глобального климата и экологии имени академика Ю.А. Изразля»
результатов диссертации А.В. Саакяна
«Развитие методов абсорбционной, флуоресцентной и рамановской спектроскопии с
хеометрической обработкой данных для анализа жидких сред и аэрозолей»,
представленной на соискание ученой степени кандидата технических наук

Настоящим актом подтверждается, что в работах ФГБУ «ИГКЭ» по развитию сети мониторинга черного углерода в России использовались результаты исследований, полученные в диссертации А.В. Саакяна, по определению преобладающего источника черного углерода в пробах атмосферного воздуха по спектрам комбинационного рассеяния света с обработкой данных с помощью алгоритмов машинного обучения.

В дальнейшем, результаты этих работ позволят разработать методику определения преобладающего источника эмиссии черного углерода для разных территорий и климатических зон, что в свою очередь важно для верификации климатических моделей.

Зав. отделом ФГБУ «ИГКЭ»

Е.А. Позднякова