

Федеральное государственное бюджетное учреждение науки
Ордена Ленина и Ордена Октябрьской Революции Институт геохимии и
аналитической химии им. В.И. Вернадского Российской академии наук
(ГЕОХИ РАН)

ПРИНЯТО

Ученым советом ГЕОХИ РАН

Протокол № 8 от 22 октября 2014 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Многомерные данные в химическом анализе»

Направления подготовки: 04.06.01 - Химические науки

(указывается код и наименование направления подготовки)

Направленности (профили) подготовки: 02.00.02 «Аналитическая химия»

(наименование направленности подготовки)

Квалификация: Исследователь. Преподаватель-исследователь

Форма обучения: очная

Вид итогового контроля: Зачет

(Зачет/Дифференцированный зачет/Экзамен)

Москва 2016

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 04.06.01 Химические науки (уровень кадров высшей квалификации). Приказ Минобрнауки РФ №869 от 30 июля 2014г. (зарегистрирован в Минюсте России 20 августа 2014г., регистрационный № 33718).

Рабочая программа адаптирована для аспирантов ГЕОХИ РАН на основе программы доцента Гармаша А.В.

1. Цели и задачи освоения дисциплины

Целью изучения дисциплины «Многомерные данные в химическом анализе» является освоение аспирантами фундаментальных знаний в области хемометрики, изучение основных методов обработки многомерных данных химического анализа.

Задачи дисциплины:

- Формирование базовых знаний в области хемометрики как научной дисциплины, интегрирующей математическую и химико-аналитическую подготовку химиков и обеспечивающей методологические основы современных инновационных сфер деятельности;
- Обучение студентов основам хемометрических методов обработки многомерных данных и применения их в химическом анализе;
- Формирование подходов к выбору хемометрических методов, адекватных поставленной прикладной задаче

2. Место дисциплины (модуля) в структуре ООП

Дисциплина «Многомерные данные в химическом анализе» входит в часть ООП и относится к дисциплинам по выбору обучающихся, которые направлены на подготовку к сдаче кандидатского экзамена по специальности.

3. Требования к предварительной подготовке обучающегося:

Знать: основные методы научно-исследовательской работы.

Уметь: выделять и систематизировать основные идеи в научных публикациях; критически оценивать информацию вне зависимости от источника.

Владеть: навыками поиска, анализа и систематизации информации по теме исследования; навыками выбора методов и средств решения задач исследования.

Для освоения дисциплины «Многомерные данные в химическом анализе» требуются знания и умения, которые должны быть освоены до начала освоения данной дисциплины: «Аналитическая химия», «Линейная алгебра», «Элементы прикладной математической статистики», «Основы хемометрики».

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

Код компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине
УК-1	способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях	Знать: методы критического анализа и оценки современных научных достижений, а также методы генерирования новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях. Уметь: анализировать альтернативные варианты решения исследовательских и

		практических задач. Владеть: навыками анализа методологических проблем, возникающих при решении исследовательских и практических задач, в том числе в междисциплинарных областях.
ОПК-1	способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	Знать: современные способы использования информационно-коммуникационных технологий в выбранной сфере деятельности Уметь: выбирать и применять в профессиональной деятельности экспериментальные и расчетно-теоретические методы исследования.
ПК-1	способность и готовность к самостоятельному проведению научно-исследовательской работы и получению научных результатов в области аналитической химии	Знать: современное состояние науки в области аналитической химии. Владеть: методами планирования, подготовки, проведения научно-исследовательской работы по направленности (02.00.02 Аналитическая химия).

4. Структура 5. Структура и содержание дисциплины

Общая трудоёмкость дисциплины составляет 3 зачётных единицы, 108 часов

Вид работы	Всего
Общая трудоёмкость	108
Аудиторная работа:	18
лекции	18
семинары	-
Самостоятельная работа	90
Вид итогового контроля	зачет

Содержание дисциплины, виды учебных занятий и формы их проведения

№ раздела	Наименование раздела	Кол. часов
		Лекции
1	Многомерные данные в химическом анализе и основные операции с ними	4
2	Количественный анализ многокомпонентных систем	6
3	Методы многомерной классификации и идентификации в химическом анализе	6

4	Методы многомерного разрешения кривых	2
	Самостоятельная работа	90
	ИТОГО	108
Вид итогового контроля -Зачет		

Лекции

№ раздела.	Наименование раздела	Содержание лекции
1	Многомерные данные в химическом анализе и основные операции с ними	<p>Предмет хемометрики, ее задачи. Полезная информация и шум. Многомерные данные в химическом анализе. Аналитический сигнал, аналитический признак, аналитическая позиция. Представление спектральных, хроматографических и других данных в векторной форме. Матрицы "объект-свойство". Программное обеспечение для работы с многомерными данными. Основные приемы работы с многомерными данными в среде Excel.</p> <p>Линейные преобразования матриц данных: сдвиг (центрирование), масштабирование, поворот, проекция. Их назначение, алгебраическая и геометрическая интерпретация. Пространство главных компонент, матрицы счетов и нагрузок, их свойства, сингулярные числа, формальный и эффективный ранг матрицы.</p> <p>Линейные преобразования матриц многомерных аналитических данных. Представление матриц данных в пространстве главных компонент Расчет матриц счетов и нагрузок, изучение их свойств. Расчет сингулярных чисел матрицы, нахождение ее эффективного ранга</p>
2	Количественный анализ многокомпонентных систем	<p>Многомерный регрессионный анализ. Математическая формулировка задач градуировки и расчета концентраций в матричной форме. Метод Фирордта и метод множественной линейной регрессии. Метод наименьших квадратов. Прямая и обратная градуировки. Анализ многокомпонентных систем методами Фирордта и множественной линейной регрессии.</p> <p>Проекционные методы регрессионного анализа. Проекция на главные компоненты (PCA), проекция на скрытые структуры (PLS1, PLS2). Их использование в анализе многокомпонентных систем. Анализ многокомпонентных систем с использованием проекционных методов.</p> <p>Метрологические аспекты многокомпонентного</p>

		<p>анализа. Описательная и предсказательная способность регрессионной модели. Метод "введено-найдено" в многокомпонентном анализе. Градуировочная и проверочная выборки. Перекрестная проверка на достоверность. Среднеквадратичная погрешность. Оптимизация регрессионной модели.</p> <p>Оценка описательной и предсказательной способности регрессионной модели. Изучение влияния погрешности исходных данных на погрешность результатов. Стратегия выбора оптимальной регрессионной модели и ее параметров для анализа многокомпонентных систем</p>
3	<p>Методы многомерной классификации и идентификации в химическом анализе</p>	<p>Основные принципы многомерной классификации. Классификационные признаки, принципы их выбора. Расстояние от неизвестного объекта до известного класса как главный классификационный критерий. Способы предварительного преобразования пространства признаков и способы вычисления расстояний в нем. Расстояние "метрическое" (Минковского) и "статистическое" (Махаланобиса). Графическое представление результатов классификации. Кластерный анализ, дендрограммы.</p> <p>Выбор признаков, преобразование данных, вычисление расстояний в методах многомерной классификации, графическое представление результатов.</p> <p>Основные методы многомерной классификации: линейный дискриминантный анализ, SIMCA, PLS-классификация, метод К ближайших соседей. Оптимизация классификационной модели. Классификация химических объектов с использованием различных методов и оценка качества классификации</p>
4	<p>Методы многомерного разрешения кривых</p>	<p>Постановка задачи разрешения кривых. Регрессионный анализ с ограничениями. Метод чередующихся наименьших квадратов (ALS), итерационного целевого (ITFA), эволюционного (EFA) и оконного (WFA) факторного анализа</p> <p>Постановка задачи разрешения кривых. Регрессионный анализ с ограничениями. Метод чередующихся наименьших квадратов (ALS), итерационного целевого (ITFA), эволюционного (EFA) и оконного (WFA) факторного анализа</p>

6. Учебно-методическое обеспечение самостоятельной работы аспирантов

Самостоятельная работа подкрепляется учебно-методическим и информационным обеспечением, включающим учебники, учебно-методические пособия, конспекты лекций. Основные виды самостоятельной работы: в читальном зале библиотеки, на рабочем месте с доступом к ресурсам Интернет по IP-адресам.

Основной контроль знаний осуществляется в процессе участия в обсуждениях, ответов на вопросы и др.).

7. Учебно-методическое и информационное обеспечение дисциплины

Основная литература

1. Померанцев А.Л.;Хеометрика в Excel: учебное пособие; ;Томск;Издательство ТПУ;2014; ;
2. Эбсенсен К. Анализ многомерных данных. Избранные главы Родионова О.Е. Черноголовка ИПХФ РАН 2005
3. Аналитическая химия. Проблемы и подходы. Т. 2. Гл. 12 Кельнер Р., Мерме Ж.-М., Отто М., Видмер Г.М. Москва Мир, АСТ 2004
4. Родионова О.Е, Померанцев А.Л. Хеометрика: достижения и перспективы Москва 2006 Успехи химии Т.75. № 4. С.302-321

Основная литература:

1. Ким Дж., Мюллер Ч. и др. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика.1989.
2. Шараф М.А., Иллман Д.Л., Ковальски Б.Р. Хеометрика. Л: Химия. 1989.
3. Brereton R.G. Chemometrics. Data Analysis for the Laboratory and Chemical Plant

Интернет-ресурсы:

Доступ к электронным научным информационным ресурсам осуществляется как в читальном зале библиотеки ГЕОХИ РАН, так и на рабочем месте по IP-адресам ГЕОХИ РАН:

1. <http://rcs.chemometrics.ru> <http://www.chem.msu.ru/rus/elibrary/welcome.html>
2. <http://apps.webofknowledge.com/> -База данных Web of Science
3. <http://www.scopus.com/> - База данных SCOPUS
4. <http://www.sciencedirect.com/> - Журналы издательства Elsevier - Freedom Collection
5. <http://link.springer.com>- Журналы издательства Springer
6. <http://www.springerprotocols.com/> - Журналы издательства Springer Journals и SpringerProtocols
7. <http://e-library.ru> Научная электронная библиотека eLIBRARY.RU
8. <http://www.benran.ru/> с компьютеров библиотеки ГЕОХИ РАН открыт доступ к электронным ресурсам Библиотеки по естественным наукам РАН.

7. Образовательные технологии

Кроме очных лекций проводятся практические занятия на базе аналитических лабораторий ГЕОХИ, используется самостоятельная работа и выполнение расчетных заданий. При осуществлении образовательного процесса по дисциплине используются следующие информационно-телекоммуникационные технологии:

- Персональный компьютер (или ноутбук) с необходимым программным обеспечением для работы устройства, а также для демонстрации презентаций MS PowerPoint.

8. Материально-техническое обеспечение дисциплины

8.1. Для освоения программы обучения и для выполнения научно-исследовательских работ по теме диссертации каждому аспиранту предоставлено индивидуальное рабочее место, оборудованное приточно-вытяжной вентиляцией, водопроводом, водоотведением, воздуховодом. Аспиранты имеют возможность использовать материально-технические средства лабораторий, в которых выполняют квалификационные и диссертационные работы (оргтехника, включая персональный компьютер, реактивы, расходные материалы, лабораторная посуда, измерительное оборудование).

8.2. Лекционная аудитория, оборудованная проекционным оборудованием и доступом в сеть «Интернет».

9. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

9.1. Форма контроля знаний:

9.1.1. Текущий контроль знаний в течение всего курса осуществляется с помощью контрольных вопросов.

9.1.2. Зачет по дисциплине в конце курса обучения.

9.2.3. Оценочные средства (в виде устных вопросов)

9.2.3.1. Перечень примерных вопросов для текущего контроля знаний:

Контрольные вопросы:

1. В чем заключается метод Фирордта для анализа многокомпонентных систем?
2. В чем сходство и в чем различие метода Фирордта и метода множественной линейной регрессии?
3. В чем сущность метода наименьших квадратов?
4. Что такое прямая и обратная градуировка?
5. Как в матричной форме записываются уравнения, лежащие в основе анализа многокомпонентных систем методом множественной линейной регрессии, при использовании прямой градуировки? Обратной градуировки? Как решить эти уравнения?
6. Каковы ограничения на число аналитических позиций при прямой и обратной градуировке? В чем причина этих ограничений? С какой целью?
7. В чем сущность проекционных методов регрессионного анализа?
8. Каковы преимущества проекционных методов регрессионного анализа перед методом множественной линейной регрессии?
4. В чем сходство и различие методов проекции на главные компоненты и на скрытые структуры?

5. В чем сходство и различие в свойствах матриц счетов и нагрузок в этих методах?
6. В чем различие методов PLS1 и PLS2?
7. Какой способ градуировки - прямой или обратный - целесообразнее применять в проекционных методах? Почему?
8. В чем сходство и различие в свойствах матриц счетов и нагрузок при проекции на главные компоненты и на скрытые структуры?
9. Как записать в матричном виде уравнения для градуировки и анализа многокомпонентных систем при использовании методов PCA, PLS1, PLS2? Как решить эти уравнения?
10. Какую предварительную обработку данных проводят при анализе многокомпонентных систем проекционными методами? С какой целью?
11. Как выбрать число счетов и нагрузок при анализе многокомпонентных систем проекционными методами?
9. Что такое описательная и предсказательная способность регрессионной модели? Как их можно охарактеризовать количественно?
10. Как влияют на описательную и предсказательную способность модели число ее параметров?
11. Что такое градуировочная и проверочная выборки?
5. В чем заключается метод перекрестной проверки на достоверность?
6. Из каких основных этапов состоит процесс оптимизации регрессионной модели для анализа многокомпонентных систем?
12. В чем состоит основной принцип многомерной классификации?
13. Что такое обучающая и контрольная выборки, каковы требования к ним?
14. Каковы требования к аналитическим признакам, используемым для классификации химических объектов?
15. Какие предварительные преобразования данных и с какой целью применяют в многомерной классификации?
16. Каковы основные способы вычисления расстояний между объектами в пространстве аналитических признаков?
7. В чем состоит принцип группировки данных в кластеры? Как можно графически представить результаты такой группировки?
17. Какую предварительную обработку данных проводят при анализе многокомпонентных систем методами Фирордта и множественной линейной регрессии. В чем состоит задача разрешения кривых?
18. Что такое спектральный и концентрационный профиль?
19. В чем заключается неопределенность задачи разрешения кривых? Каким способом можно устранить эту неопределенность?
20. Какие основные виды ограничений используют при решении задачи разрешения кривых? Как эти ограничения формулируются математически?
21. Охарактеризуйте суть методов чередующихся наименьших квадратов, итерационного целевого, эволюционного и оконного факторного анализа.

9.2.3.2. Перечень вопросов к зачету:

1. Многомерные данные в химическом анализе и основные операции с ними

1. Предмет и задачи хемометрики. Полезная информация и шум. Структура данных. Математические модели данных.
2. Аналитический сигнал, аналитический признак, аналитическая позиция. Представление аналитических данных в векторной и матричной форме.
3. Линейные преобразования матриц данных: сдвиг (центрирование), масштабирование, поворот, проекция. Их назначение, алгебраическая запись, геометрическая интерпретация.
4. Пространство главных компонент. Матрицы счетов и нагрузок, их свойства. Сингулярные числа, формальный и эффективный ранг матрицы. Критерии определения эффективного ранга матрицы.

2. Количественный анализ многокомпонентных систем

1. Многомерный регрессионный анализ. Формулировка задач градуировки и расчета концентраций в матричной форме. Метод Фирордта и метод множественной линейной регрессии. Метод наименьших квадратов. Прямая и обратная градуировка.
2. Проекционные методы регрессионного анализа. Проекция на главные компоненты, проекция на скрытые структуры, их использование в регрессионном анализе. Анализ многокомпонентных систем с использованием проекционных методов регрессионного анализа.
3. Описательная и предсказательная способность регрессионной модели. Градуировочная и проверочная выборки. Перекрестная проверка на достоверность. Среднеквадратичная погрешность. Оптимизация регрессионной модели.

3. Многомерная классификация в химическом анализе

1. Постановка задач классификации и идентификации. Основные принципы многомерной классификации. Классификационные признаки, принципы их выбора. Применение корреляционного анализа для выбора классификационных признаков. Пространство классификационных признаков.
2. Расстояние от неизвестного объекта до известного класса как главный классификационный критерий. Способы предварительного преобразования пространства признаков и способы вычисления расстояний в нем. Расстояния "метрическое" (Минковского) и "статистическое" (Махаланобиса). Критерий Стьюдента как простейший пример использования расстояния Махаланобиса. Графическое представление (визуализация) результатов классификации.
3. Группировка объектов в классы. Кластерный анализ. Дендрограммы и способы их построения.
4. Основные методы многомерной классификации. Линейный дискриминантный анализ, SIMCA, PLS-классификация, метод К ближайших соседей: способы вычисления расстояний, классификационные критерии.
5. Оптимизация классификационной модели. Обучающая и контрольная выборки.
6. Применение многомерной классификации для контроля качества образцов, установления их чистоты, источника происхождения.

4. Многомерное разрешение кривых

1. Постановка задачи разрешения кривых. "Спектральный" и "концентрационный" профили.
2. Неоднозначность решения задачи разрешения кривых и способы ее преодоления. Регрессионный анализ с ограничениями. Основные виды

ограничений, используемые в задачах разрешения кривых: неотрицательность, унимодальность, монотонность. Их математическая формулировка.

3. Методы чередующихся наименьших квадратов, итерационный целевой, эволюционный и оконный факторный анализ: алгоритмы, основные формулы.
4. Применение методов разрешения кривых в спектроскопии, хроматографии, термодинамических и кинетических исследованиях.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ листа изменений или наименование элемента приложения	№ протокола Ученого Совета (секции Уч. Совета)	Дата заседания Ученого Совета (секции Уч. Совета)	Всего листов в документе	Подпись зам. директора по научной работе
	Рабочая программа обсуждена и принята на заседании Ученого совета ГЕОХИ РАН, протокол № 8	22 октября 2014 года	11	
--	Обновленный текст программы принят на заседании Ученого совета РАН, Протокол №11(8)	28 декабря 2016 года	11	