

Харьковский национальный университет
имени В.Н. Каразина

на правах рукописи

УДК 543.08+543.061+543.054

Пантелеймонов Антон Виталиевич

НОВЫЕ ХЕМОМЕТРИЧЕСКИЕ СПОСОБЫ ОБРАБОТКИ ДАННЫХ
АНАЛИТИЧЕСКОГО ЭКСПЕРИМЕНТА

02.00.02 – аналитическая химия

Диссертация

на соискание ученой степени кандидата химических наук

Научный руководитель
Холин Юрий Валентинович,
доктор химических наук,
профессор

Харьков – 2008

СОДЕРЖАНИЕ

Список обозначений.....	4
Введение.....	5
0.1. Хемометрия и ее значение для аналитической химии	5
0.2. Методы хемометрии	9
0.3. Выбор объектов и обоснование задач исследования. Актуальность и цель работы. Научная и практическая ценность результатов.....	12
Раздел 1. Метрологические характеристики методик качественного анализа с бинарным откликом	19
1.1. Качественный химический анализ и его место в аналитической химии.....	19
1.2. Метрологические характеристики методик анализа с бинарным откликом.....	22
1.3. Построение кривых эффективности	31
1.3.1. Оценка адекватности моделей. Расчет ковариационных матриц подгоночных параметров.....	46
1.3.2. Назначение статистических весов.....	51
1.4. О возможности подбора кривых эффективности на основе системы функций плотности Пирсона.....	66
1.5. Робастный алгоритм расчета подгоночных параметров кривых эффективности	75
1.6. Выводы к разделу 1	80
Раздел 2. Идентификация соединений – применение хемометрических подходов	81
2.1. Принципы компьютерной идентификации соединений.....	81
2.2. Априорный подход к идентификации аналитов.....	85
2.3. Алгоритм идентификации аналитов с использованием теории нечетких множеств	92
2.4. Испытание алгоритма идентификации аналитов	95
2.4.1. Идентификация по данным газовой хроматографии.....	96
2.4.2. Идентификация по данным ИК-спектроскопии.....	98

2.4.3. Идентификация по данным УФ-спектроскопии	103
2.5. Выводы к разделу 2	109
Раздел 3. Химико-аналитические характеристики твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов	110
3.1. Гибридные органо-кремнеземные материалы в задачах разделения и концентрирования.....	110
3.2. Задача моделирования равновесий по данным количественного физико-химического анализа.....	111
3.3. Модели для описания сорбционных равновесий	116
3.4. Алгоритм расчета сорбционной емкости и констант сорбционных равновесий на основе теорий нечетких множеств и робастного оценивания.....	125
3.5. Испытание разработанного метода при исследовании модельной системы	127
3.6. Определение химико-аналитических характеристик материалов на основе силсесквиоксан хлоридных полимеров.....	130
3.6.1. Моделирование адсорбционных свойств материала SiPy^+Cl^-	130
3.6.2. Моделирование адсорбционных свойств $\text{SiPic}^+\text{Cl}^-$	146
3.7. Выводы к разделу 3	154
Выводы	155
Список публикаций автора по теме работы	157
Благодарности.....	160
Список использованных источников	161

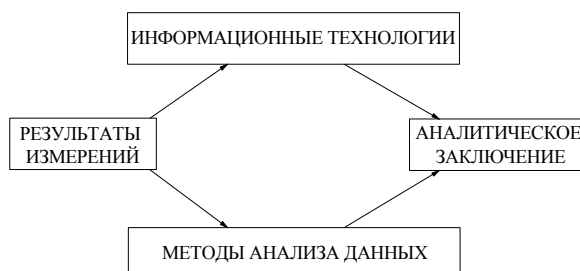
СПИСОК ОБОЗНАЧЕНИЙ

A	адсорбция
b	допустимый размах данных
c	концентрация аналита
d_E	Евклидово расстояние
M, N	число экспериментальных точек
P	вероятность
r	коэффициент корреляции
R	достоверность
R	степень извлечения
s	стандартное отклонение
t	общая концентрация компонента
t_Q	эффективная сорбционная емкость
w	статистический вес
α	вероятность ошибки I рода
β	вероятность ошибки II рода
β	константа устойчивости
$\chi^2_{f, \alpha}$	100 α -процентная точка распределения χ^2 для f степеней свободы и уровня значимости α
μ	функция принадлежности
θ	вектор неизвестных параметров

ВВЕДЕНИЕ

0.1. Хемометрия и ее значение для аналитической химии

В аналитической химии академик Ю.А. Золотов выделяет блоки методов и объектов анализа; аналитов; отдельных областей анализа (неразрушающий, непрерывный, дистанционный, локальный, вещественный анализ и др.); средств для осуществления анализа; общих, в том числе общетеоретических, аспектов и направлений (метрология анализа, теория пробоотбора, принципы и пути автоматизации и др.) [1]. На современном этапе развития аналитической химии в блоке методов анализа все большее значение приобретают инструментальные методы, а в общетеоретическом блоке – средства обеспечения качества измерений химического состава. Инструментальные методы предоставляют аналитику большие числовые массивы, которые необходимо хранить, сравнивать с имеющимися в базах данных аналогами, подвергать обработке, опираясь на методы информатики и теории анализа данных. Обеспечение качества измерений выдвигает на авансцену аналитической химии метрологическую проблематику, в связи с чем особо актуальными становятся проблемы расширения объема, повышения точности и достоверности информации, извлекаемой из результатов измерений. Очевидно, что при решении этой задачи также не обойтись без интенсивного применения компьютерно ориентированных математических методов. Не будет большим преувеличением сказать, что в конце XX – начале XXI века лейтмотивом применения самых разных аналитических методов стало преобразование массивов результатов измерений в аналитические выводы с применением теории анализа данных и информационных технологий:



В условиях, когда объекты химического анализа весьма различаются, а теоретические основы используемых методов зачастую принадлежат различным научным дисциплинам, внедренные в аналитическую химию информационные технологии и методы теории анализа данных остаются, наряду с теорией пробоотбора и пробоподготовки, едва ли не единственными элементами, объединяющими разрозненные разделы аналитической химии и предотвращающими ее дезинтеграцию.

Разработкой статистических и математических методов обработки массивов аналитических данных занимается хемометрия¹. Рассматривая историю хемометрии, авторы обзора [2] указывают, что она «зародилась и длительное время развивалась внутри аналитической химии». Первая научная статья, в названии которой фигурировало слово «хемометрия», была опубликована в 1975 г. Брюсом Ковальски [3]. Начиная с 1977 г. ведущий аналитический журнал издательства «Elsevier», «Analytica Chimica Acta», регулярно публикует оригинальные и обзорные работы по хемометрии; с 1980 г. издаваемый Американским химическим обществом журнал «Analytical Chemistry» заменил название обзорного раздела «Statistical and Mathematical Methods in Analytical Chemistry» на «Chemometrics». Хотя в последние годы хемометрию иногда рассматривают как самостоятельную дисциплину, преодолевшую в своем развитии границы аналитической химии, авторитетные аналитики по-прежнему полагают, что хемометрия остается «важнейшим разделом аналитической химии» [4], «аналитики до сих пор остаются главными пользователями хемометрических методов» [2]. Авторы недавно опубликованного учебника [5] включили хемометрические разделы в общетеоретическую часть курса аналитической химии; целиком хемометрии посвящено учебное пособие проф. Б.М. Марьянова [6]; соответствующие курсы стали частью подготовки химиков-аналитиков (например, на химическом факультете МГУ им. М.В. Ломоносова для студентов-аналитиков читается курс «Введение в хемометрику и химическую метрологию»).

¹ В русскоязычной литературе часто используют также термин «хемометрика», но мы согласны с проф. Ю.В. Грановским (МГУ им. М.В. Ломоносова), что более удачным названием является «хемометрия».

Таким образом, есть все основания рассматривать хемометрию как раздел теоретических основ современной аналитической химии.

Определений хемометрии известно множество. Согласно Сванте Волду [7], одному из отцов-основателей хемометрии, это «наука о том, как получить химически важную информацию из химических данных, как организовать и представить эту информацию, и как получить данные, содержащие такую информацию». Российское хемометрическое общество полагает [8], что «хемометрия – это научная дисциплина, находящаяся на стыке химии и математики, предметом которой являются математические методы исследования химических данных». В учебнике [9] приводится очень широкое определение: «Хемометрия – химическая дисциплина, применяющая математику, статистику и формальную логику для (а) построения или отбора оптимальных измерительных процедур; (б) извлечения максимально полной химической информации путем анализа химических данных и (в) получения знаний о химической системе». Встречаются и толкования хемометрии, стирающие, по сути, различие между ней и химической метрологией, например, такое: «Хемометрия – химическая дисциплина, которая позволяет извлекать химически значимую информацию путем оптимизации экспериментов, обработки данных, градуировки, контроля качества и организации аналитического процесса» [10].

Хемометрия тесно связана с теорией анализа данных (data mining). Последняя развилась из прикладной математической статистики. Под «анализом данных» понимают любые процедуры обработки данных, которые невозможно полностью формализовать. Методологически анализ данных близок к неопозитивистским концепциям, таким, как «бутстреп» [11], согласно которой результаты изучения механизма явлений представляет не одна модель, а их множество [12]. При этом возрастает значение формулировки цели исследования и неформального (содержательного) анализа соответствия решения поставленной цели: «Анализ данных изыскивает различные приемы для наиболее полного использования эндогенной информации (что вообще характерно для любых

статистических методов), но вместе с тем он постоянно нацелен на максимальное использование информации внешней» [13].

Следует упомянуть и о глубокой связи хемометрии и химической метрологии. Некоторые направления хемометрии развивались еще в 50-е годы XX в. в Харьковском университете под руководством проф. Н.П. Комаря [2]. Сам Н.П. Комарь, разумеется, не пользовался термином «хемометрия»; он выступил за преобразование аналитической химии в химическую метрологию как науку об измерении химической формы движения материи. Термин «измерение» связывал эту дисциплину с общей метрологией, причем химическую метрологию Н.П. Комарь не представлял без систематического применения математических методов. Идеи Н.П. Комаря близки к актуальным подходам к анализу химических данных, когда в центр исследования ставится максимально полное извлечение скрытой в результатах измерений содержательной (в частности, необходимой для принятия решений) информации на основе применения адекватных математических средств.

Как указывает А.Б. Бланк [14], «программа работ в области химической метрологии намечает основные пути обеспечения достоверности аналитических измерений... Хемометрика исследует некоторые (но не все) способы решения указанных задач и развивает информационный аспект химического анализа. Взаимодополняющий характер методов и задач химической метрологии и хемометрики делает перспективным объединение их концепций и подходов в интересах аналитической химии...».

Современная химическая метрология [15-17], оперирующая фундаментальными понятиями единства (прослеживаемости, *traceability*) измерений и неопределенности (*uncertainty*) результатов анализа [18], не в состоянии решить свои задачи без обращения к хемометрическим методам, поскольку важен не столько результат анализа сам по себе, а результат содержательный, на основе которого потребитель может принимать решения [19] («the information provided by chemical measurements must be reliable if it is to form the basis of important decision-making processes») [20].

Тридцатилетний опыт показал, что благодаря использованию хемометрических подходов «очень часто традиционные аналитические методы, требующие больших затрат труда, времени, уникального оборудования, дорогих реактивов, могут быть заменены на косвенные методы, которые гораздо быстрее и дешевле» [2]. Первые общепризнанные достижения хемометрии относились к анализу многомерных спектральных данных [21], а в настоящее время она успешно применяется в самых разных областях, в частности, для идентификации, классификации и контроля качества пищевых продуктов [22-26], классификации здоровых и больных тканей [27, 28], фармацевтических препаратов [29], хроматографических фаз [30], органических растворителей [31, 32], вредных отходов [33], бензинов [34], для классификации и измерения сходства химических структур [35], при моделировании механизмов химических реакций, в том числе процессов гомогенного и гетерогенного катализа [36, 37], при создании электронного носа и электронного языка [38-40], аналитическом мониторинге технологических процессов и контроле сырья [41, 42].

0.2. Методы хемометрии

Исходными данными для хемометрического исследования являются массивы результатов измерений. Чаще всего обработке подвергают массивы данных, содержащих многомерные отклики высокой размерности, лишь в минимальной степени, как и положено в теории анализа данных, привлекая предположения о статистических характеристиках результатов измерений, раскрывая структуру данных «не ссылаясь на протокол эксперимента или исходные гипотезы» [43]. Традиционными задачами являются группировка и классификация химических объектов, выявление взаимосвязей между различными переменными, градуировка, в том числе многомерная и др. [43].

Хемометрия владеет многими методами, существенно облегчающими решение традиционных задач и делающими возможным решение задач, ранее непосильных для аналитиков. Находят применение как давно известные методы

(например, линейный и нелинейный метод наименьших квадратов (МНК), корреляционный, кластерный анализ), так и их современные модификации и новые разработки. Примерами могут служить анализ главных компонент (PCA), проекция на латентные структуры (PLS), факторный анализ, разложение по сингулярным значениям (SVD), дискриминантный анализ, нелинейное итерационное проецирование при помощи чередующихся наименьших квадратов (NIPALS), робастная непрерывная регрессия, классификация при помощи искусственных нейронных сетей, системы искусственного интеллекта и многие другие процедуры [43]. Разработка и внедрение новых методов и алгоритмов – отнюдь не прихоть исследователей. С одной стороны, это ответ на возникающие практические запросы. С другой стороны, это следствие логики развития хемометрии. «Хемометрика тесно связана с математикой и, в особенности, с математической статистикой, откуда она черпает свои идеи» [2]. Но подходы и методы, известные в математике, должны быть адаптированы к обработке химических данных, должна быть проверена их эффективность именно как хемометрических средств, а это неформальная и зачастую довольно непростая задача. Кроме того, на основе известных математических идей специалисты в области хемометрии изобретают и новые методы, причем «делают они это так быстро, что математики ... не успевают не только раскритиковать их за это, но и просто понять, что же происходит в этой хемометрике» [2].

Таким образом, расширение арсенала хемометрических методов следует признать актуальной задачей, но при этом следует убедиться в целесообразности нового средства анализа данных для решения определенного круга химических задач.

Наше внимание привлек тот факт, что хемометрия сравнительно в малой степени и лишь с недавнего времени обращается к аппарату такой мощной концепции, как теория нечетких множеств (fuzzy sets theory). Чаще всего она использовалась для построения алгоритмов нечеткой многомерной классификации и использования нечеткой логики при принятии решений [44-57].

В основу теории нечетких множеств положено понятие субъективной вероятности (possibility), далекое от статистического толкования. Поскольку для хемометрии характерно стремление в минимальной степени опираться на статистические гипотезы, привлечение теории нечетких множеств представляется естественным. Важным является также то соображение, что на основе теории нечетких множеств удастся построить робастные алгоритмы, устойчивые к наличию в данных так называемых «грубых промахов» (outliers) [58-61].

Теория нечетких множеств [62] оперирует с нечеткими числами – нечеткими переменными, определенными на числовой оси. Нечеткое число определяется как нечеткое множество A на множестве действительных чисел R с функцией принадлежности $\mu_A(x) \in [0,1]$, где x – действительное число, т.е. $x \in R$.

Представление результатов измерений в виде нечетких чисел производится с помощью процедуры «фаззификации» («fuzzyfication») обычных чисел. Простейший способ перевода обычного числа в нечеткое иллюстрирует рис. 0.1.

Здесь x – действительное число (результат измерения), x_L и x_R – границы нечеткости числа, μ – функция принадлежности, являющаяся мерой истинности высказывания «результат измерения равен x » ($0 \leq \mu \leq 1$). Степень близости нечетких чисел оценивают с помощью операции пересечения [62]. На рис. 0.2 нечеткие числа x_1 и x_2 принадлежат одному множеству с $\mu = 0.83$.

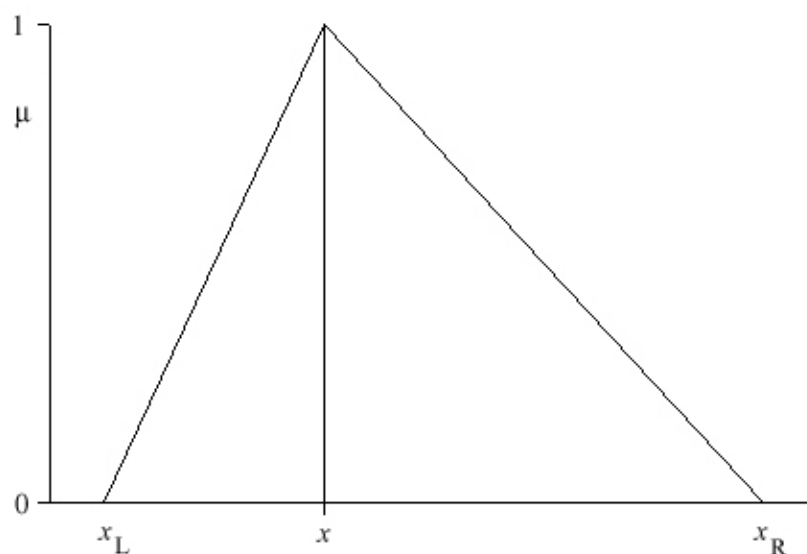


Рис. 0.1. Представление числа x в виде нечеткого.

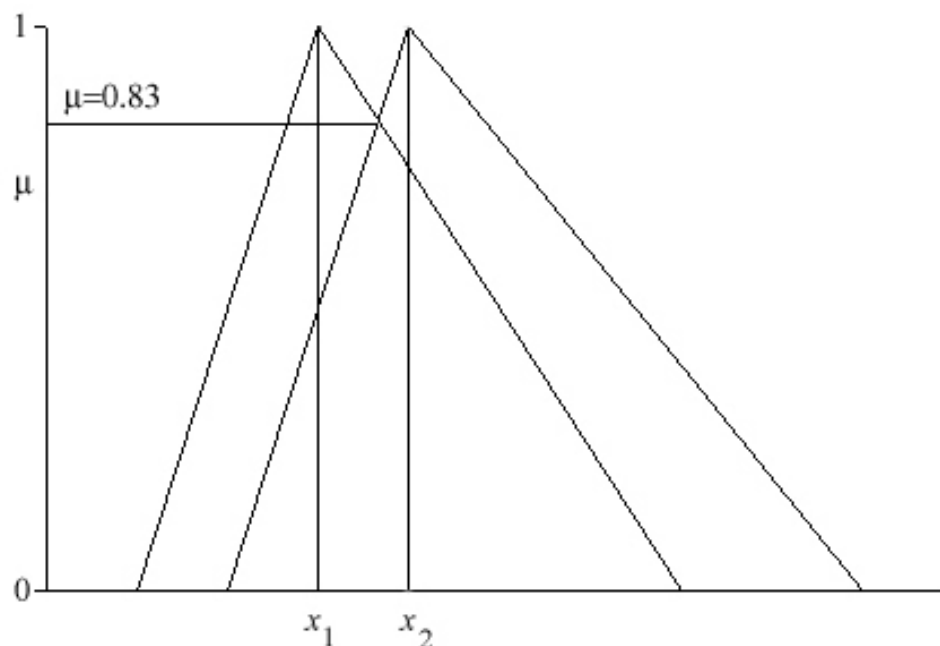


Рис. 0.2. Оценка степени принадлежности двух нечетких чисел одному множеству.

0.3. Выбор объектов и обоснование задач исследования. Актуальность и цель работы. Научная и практическая ценность результатов

В настоящей работе была поставлена задача разработки хемометрических процедур на основе использования теории нечетких множеств для решения ряда химико-аналитических задач.

Круг задач, для решения которых планировалось применить разработанные методы, включал обнаружение и идентификацию аналитов (качественный анализ) и характеристику свойств органо-кремнеземных материалов как сорбентов или ионообменников по данным количественного физико-химического анализа (КФХА).

Область качественного анализа была выбрана в связи с недостаточным развитием метрологического обеспечения этого вида анализа и потребностью в надежных вычислительных средствах для решения задач обнаружения и идентификации.

Целесообразность внедрения хемометрических средств на основе теории нечетких множеств в решение задач КФХА была обусловлена тем, что ранее использовавшиеся методы не позволяли извлечь из первичных экспериментальных данных объективную информацию о характеристиках сорбентов (ионообменников) в условиях, когда емкость материала является лишь одной из искомым характеристик, а положенные в основу расчетных процедур гипотезы о статистических свойствах результатов измерений заведомо нарушались.

Одно из затруднений связано с тем, что концентрация **активных** сорбционных или ионообменных центров может значительно отличаться от концентрации, найденной по данным элементного анализа. Даже в случае идеальной сорбции одновременное определение емкости материалов и константы сорбционного равновесия (β) по изотерме сорбции представляет собой нетривиальную задачу, если устойчивость сорбционных комплексов не очень высока ($\beta < 10^3 - 10^4$ л моль⁻¹): для того, чтобы занять все активные центры, необходимы большие избытки сорбата, практически недостижимые экспериментально. Как результат, прямое определение емкости по участкам насыщения на изотермах сорбции дает ненадежные оценки [63]. В случае неидеальной сорбции (ионного обмена) приходится также количественно описывать эффекты кооперативности или энергетической неоднородности, что представляет собой непростую в методологическом и вычислительном отношении задачу. Наконец, следует учитывать, что традиционно получаемые с помощью МНК оценки искомым параметров могут не обладать оптимальными статистическими свойствами, если нарушается гипотеза о нормальном распределении погрешностей результатов измерений.

Таким образом, **актуальность работы обусловлена потребностью в разработке новых хемометрических процедур анализа химико-аналитических данных, в особенности для решения задач обнаружения и идентификации в качественном анализе и получения объективных**

характеристик твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов.

Связь работы с научными программами, планами, темами. Работа выполнена в соответствии с научно-исследовательскими работами кафедры химического материаловедения ХНУ имени В.Н. Каразина «Керування хімічними рівновагами в гетерогенних та мікрогетерогенних середовищах, перспективних для тестових та гібридних методів аналізу» (№ государственной регистрации 0103U004212); «Інформаційні технології та комп'ютерні засоби прогнозування властивостей хімічних речовин» (№ государственной регистрации 0105U002844), согласно координационным планам Научного Совета НАН Украины по проблеме «Неорганическая химия» и Дополнительному соглашению к Договору о сотрудничестве между Харьковским национальным университетом имени В.Н. Каразина и Университетом штата Сан-Пауло в Кампинасе (Бразилия).

Цель работы: разработка хемометрических процедур на основе использования теории нечетких множеств для решения задач обнаружения и идентификации в качественном анализе и определения химико-аналитических характеристик твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов.

Для достижения поставленных целей необходимо было решить следующие задачи:

- обосновать набор метрологических характеристик методик обнаружения с бинарным откликом;
- сформировать ограниченный набор функций для представления кривых эффективности;
- обосновать расчетные процедуры определения параметров кривых эффективности и нахождения на этой основе метрологических характеристик методик обнаружения;
- разработать алгоритм идентификации аналитов по многомерным массивам их физико-химических признаков, измеренных в условиях, отличных от условий измерения свойств эталонов;

- разработать метод одновременного определения химико-аналитических характеристик твердофазных аналитических реагентов (эффективной сорбционной емкости, констант сорбционных равновесий) на основе органо-кремнеземных гибридных материалов.

Объект исследования: обнаружение и идентификация аналитов по данным одно- и многооткликковых экспериментов; химико-аналитические характеристики твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов.

Предмет исследования: метрологические характеристики методик обнаружения с бинарным откликом; вычислительные процедуры расчета параметров кривых эффективности; критерии идентификации аналитов по данным многооткликкового эксперимента при вариабельности условий измерения; методы одновременного определения нескольких химико-аналитических характеристик твердофазных аналитических реагентов по данным количественного физико-химического анализа.

Методы исследования: теория нечетких множеств, робастное оценивание, регрессионный анализ, физико-химический анализ.

Научная новизна полученных результатов:

1. Обосновано, что в качестве метрологических характеристик методик обнаружения с бинарным откликом следует использовать:
 - интервал ненадежности;
 - предел обнаружения – концентрацию аналита, при превышении которой вероятность ошибки I рода (ложного заключения об отсутствии аналита) меньше 1%;
 - недостоверность – вероятность ошибки I рода.
2. На основе анализа данных для 27 методик обнаружения показано, что в качестве функций, описывающих кривые эффективности, целесообразно использовать функции логистического и экспоненциального распределений.

3. Показано, что оценки параметров кривых эффективности, найденные нелинейным МНК, мало зависят от способа назначения статистических весов.
4. Установлено, что при расчетах допустимо использовать приближение, согласно которому частоты обнаружения аналита в интервале ненадежности являются равноточными случайными величинами, т.е. использовать невзвешенный МНК.
5. На основе применения теории нечетких множеств разработан робастный алгоритм оценки параметров кривых эффективности. Показано, что робастные оценки практически совпадают с эвристическими МНК-оценками. Тем самым обосновано использование МНК в качестве вычислительного средства для решения задач обнаружения аналитов.
6. На основе теории нечетких множеств предложен критерий идентификации аналитов по многооткликовым данным, малочувствительный к варьированию гипотез о статистических свойствах результатов измерений.
7. Разработан алгоритм, предоставляющий объективные оценки химико-аналитических характеристик сорбентов на основе органо-кремнеземных гибридных материалов в условиях нарушения предпосылок МНК и отсутствия информации о распределении экспериментальных погрешностей.
8. Определены химико-аналитические характеристики двух новых твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов.

Практическое значение полученных результатов. Формирование перечня обоснованных метрологических характеристик методик обнаружения с бинарным откликом развивает метрологию скрининга, позволяя на единой основе разрабатывать и аттестовать новые методики анализа. Выработанные рекомендации по обработке зависимостей частот обнаружения аналитов от их концентрации для нахождения параметров кривых эффективности существенно упрощают нахождение метрологических характеристик методик обнаружения.

Уточнение метрологических характеристик 27 методик обнаружения обеспечивает повышение качества анализа. Разработанный в работе алгоритм позволяет идентифицировать аналиты по данным многооткликowego эксперимента при различии условий измерения их свойств и свойств эталонов. Предложенный метод оценивания эффективной сорбционной емкости и констант равновесий процессов сорбции отличается большей объективностью предоставляемой информации, что обеспечивает надежное прогнозирование условий концентрирования и разделения аналитов. Определение химико-аналитических характеристик новых твердофазных аналитических реагентов с закрепленными группами 3-*n*-пропил(4-метилпиридина) и 3-*n*-пропилпиридина позволяет расчетным путем находить условия концентрирования и разделения солей металлов.

Личный вклад автора. Анализ литературных данных, реализация, испытание и применение вычислительных алгоритмов и количественный физико-химический анализ протолитических равновесий глицина выполнены автором самостоятельно. Постановка цели и задач исследования, формулировка выводов, обобщение результатов работы выполнялись совместно с научным руководителем проф. Ю.В. Холиным. Рекомендации по метрологии бинарного тестирования обсуждались с научным руководителем, доц. Е.А. Решетняк и доц. Н.А. Никитиной (Харьковский национальный университет имени В.Н. Каразина). Экспериментальные данные о зависимости частот обнаружения аналитов от их концентрации (12 систем) предоставлены доц. Е.А. Решетняк и доц. Н.А. Никитиной, изотермы сорбции солей металлов гибридными материалами с закрепленными группами 3-*n*-пропил(4-метилпиридина) и 3-*n*-пропилпиридина – проф. Й. Гушикемом и д-ром Э. Магоссо (Институт химии Университета штата Сан-Пауло в Кампинасе).

Апробация результатов диссертации. Основные результаты работы были доложены на международных, украинских и российских научных конференциях: IV и V Всеукраинских конференциях студентов и аспирантов «Сучасні проблеми хімії» (Киев, 2003, 2004); Сессии Научного Совета НАН Украины по проблеме

«Аналитическая химия» (Харьков, 2007); Международной конференции «Analytical Chemistry and Chemical Analysis» (Киев, 2005); Международной конференции «Modern Physical Chemistry for Advanced Materials» (Харьков, 2007); Международной конференции «4th Black Sea Basin Conference on Analytical Chemistry» (Солнчев Бряг, Болгария, 2007); XV Международной конференции студентов, аспирантов и молодых ученых по фундаментальным наукам «Ломоносов – 2008» (Москва, 2008).

Публикации. По материалам диссертации опубликованы: 1 монография, 5 статей и 7 тезисов докладов на научных конференциях.

РАЗДЕЛ 1

МЕТРОЛОГИЧЕСКИЕ ХАРАКТЕРИСТИКИ МЕТОДИК КАЧЕСТВЕННОГО АНАЛИЗА С БИНАРНЫМ ОТКЛИКОМ

1.1. Качественный химический анализ и его место в аналитической химии

Наряду с впечатляющим развитием методов количественного анализа, непрерывно расширяется и область применения качественного анализа. В качественном анализе все шире используются методики обнаружения с бинарным откликом (аналит обнаружен / аналит не обнаружен, ДА / НЕТ), не предусматривающие, как правило, использования дорогостоящего оборудования, доставки проб в лабораторию, привлечения высококвалифицированного персонала. Методики с бинарным откликом необходимы для быстрого обнаружения загрязнителей и токсикантов в объектах окружающей среды и промышленных отходах, продуктах питания, потребительских товарах [64-70], скрининга продуктов химических реакций (в частности, при комбинаторном синтезе или контроле за реакциями химического модифицирования твердых поверхностей [70-73]), экспресс-анализа медицинских препаратов и витаминов, допинг-контроля, выявления препаратов с возможным наркотическим и психотропным действием, медицинской диагностики [74-78]. Обнаружение может выполняться с помощью методик с органолептической (в большинстве случаев визуальной) регистрацией аналитического сигнала, непосредственно дающих заключение о присутствии или отсутствии аналита в пробе (к таким методикам относятся, например, широко распространенные «spot tests»). Широко применяются и методики, использующие для измерения сигнала инструментальные методы и позволяющие установить, присутствует ли аналит в пробе в концентрации, превышающей заранее установленный порог. Оба типа методик относят к процедурам скрининга, находящимся в основании пирамиды

прослеживаемости измерений в качественном анализе («traceability pyramid», рис. 1.1) [79].



Рис. 1.1. Система прослеживаемости измерений в качественном анализе.

Аналитическим сигналом может быть возникновение или изменение окраски при хромогенной реакции, возникновение либо тушение флуоресценции, образование осадка, появление запаха, ускорение или ингибирование ферментативных реакций, изменения в жизнедеятельности организмов и др. [74, 80-82]. В качестве сред для проведения реакций обнаружения используются реагентные индикаторные бумаги, пенополиуретан, желатиновые пленки, модифицированные кремнеземы, индикаторные трубки [66-68, 75, 82-84]. Основные требования к используемым реакциям и методикам изложены в работах [82, 85]:

- селективность по отношению к одному веществу или классу веществ;
- низкий предел обнаружения;
- простота регистрации и устойчивость аналитического сигнала, контрастность хромогенной реакции;
- экспрессность (в частности, высокая скорость хромогенной реакции);
- устойчивость аналитических форм при хранении.

При решении аналитических задач скрининг разделяет массив образцов на группы, содержащие и не содержащие аналит (одно вещество или класс веществ) в концентрации, превышающей некоторое заранее установленное значение. В литературе используются термины: «specification limit» – допускаемое значение,

«preset threshold», «threshold value», «threshold limit», «cut-off value» – пороговое значение, «maximum permissible limit», «maximum permitted level», «permitted level» – предельно допустимый уровень и др. [64, 86-90]. Ложный отрицательный результат, как правило, гораздо важнее ложного положительного, поскольку за заключением об открытии аналита следует количественное определение, а пробы, не показавшие наличия аналита, количественному анализу не подвергаются (рис. 1.2) [84, 86].

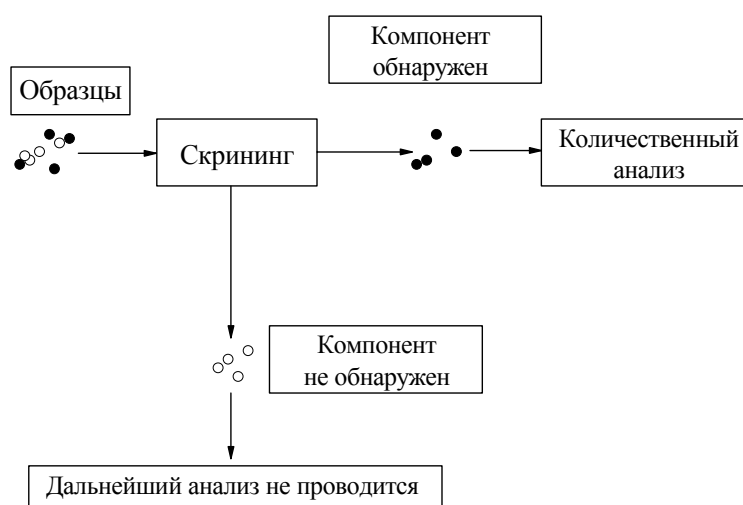


Рис. 1.2. Схема скрининга. ● – образцы, в которых аналит обнаружен, ○ – в которых аналит не обнаружен.

В отличие от количественного анализа, метрологии качественного анализа, в частности, метрологии аналитических методик с бинарным откликом, до недавнего времени уделялось явно недостаточное внимание. Поскольку очевидно существенное отличие метрологических основ качественного и количественного анализа, представляется, что перенос подходов, развитых для количественного определения аналитов, на их качественное обнаружение или вовсе невозможен, или, по крайней мере, является нетривиальной задачей. Лишь в последние годы активно публикуются работы по метрологии качественного анализа (см., например, [87-92]). Растущий интерес к этой области аналитической химии стимулировал выполнение проекта MEQUALAN [79], метрологии качественного анализа был посвящен отдельный номер журнала «Trends in Analytical Chemistry»

(2005, V. 24, No 6). Уместно заметить, что зарубежные исследователи зачастую заново приходят к результатам, публиковавшимся отечественными исследователями, начиная с 50-х годов XX века (укажем, например, на работы Н.П. Комаря, А.Б. Бланка, Р.П. Панталера, Ю.А. Золотова, В.Г. Амелина, А.А. Бугаевского и других ученых [65, 75, 84, 85, 93]).

К настоящему времени для метрологии качественного анализа с бинарным откликом характерна несогласованность терминологии и подходов, а обеспечение надежности расчетных процедур оставалось за рамками ведущейся дискуссии. Данный раздел представляет собой попытку добиться некоторой ясности в терминологии этой области химической метрологии, сократить перечень используемых метрологических характеристик до разумного минимума и изучить эффективность и устойчивость вычислительных процедур, обеспечивающих определение этих характеристик.

1.2. Метрологические характеристики методик анализа с бинарным откликом

В метрологии количественного анализа ключевым является понятие неопределенности [94]. Согласно [95, 96], неопределенность – это «параметр, связанный с результатом измерения и характеризующий разброс значений, которые с достаточным основанием могут быть приписаны измеряемой величине. Этим параметром может быть, например, стандартное отклонение (или кратное ему число) или ширина доверительного интервала». Для методик анализа с бинарным откликом концепция неопределенности, очевидно, неприменима. Вместо нее предложено использовать «недостоверность»¹ (unreliability) [79, 98]; при высокой достоверности доля ошибочных заключений о присутствии или

¹ При переводе термина «unreliability» как «недостоверность» учитывали определение [97], согласно которому «методика анализа вещества [материала] ... – документированная совокупность операций и правил, выполнение которых обеспечивает получение результата анализа вещества [материала] с установленными характеристиками погрешности [неопределенностью] или – для методик качественного анализа – установленной достоверностью».

отсутствии аналита низка [99]. Ряд метрологических характеристик методик анализа с бинарным откликом приведен в табл. 1.1.

Таблица 1.1.

**Некоторые характеристики методик качественного анализа
с бинарным откликом**

Характеристика, английский термин и обозначение		Определение, формула
1	2	3
Положительный результат	Positive, P	Наличие аналитического сигнала
Отрицательный результат	Negative, N	Отсутствие аналитического сигнала
Правильный положительный результат	True positive, TP	Положительный результат при заведомом присутствии аналита в пробе
Правильный отрицательный результат	True negative, TN	Отрицательный результат при заведомом отсутствии аналита в пробе
Ложный положительный результат	False positive, FP	Положительный результат при заведомом отсутствии аналита в пробе
Ложный отрицательный результат	False negative, FN	Отрицательный результат при заведомом присутствии аналита в пробе
Число правильных положительных результатов	N_{TP}	
Число ложных отрицательных результатов	N_{FN}	

1	2	3
Число ложных положительных результатов	N_{FP}	
Число правильных отрицательных результатов	N_{TN}	
Вероятность ошибки I рода	α	Вероятность ложного заключения об отсутствии аналита A
Вероятность ошибки II рода	β	Вероятность ложного заключения о присутствии аналита A
Достоверность (надежность)	Reliability, R	$R = 100 - \alpha - \beta$, где α и β выражены в %
Частота ложных отрицательных результатов	False negative rate, <i>FNR</i> , €	$\text{€} = N_{FN} / (N_{FN} + N_{TP})$
Частота ложных положительных результатов	False positive rate, <i>FPR</i> , §	$\text{§} = N_{FP} / (N_{FP} + N_{TN})$
Частота правильных положительных результатов (чувствительность)	True positive rate, <i>TPR</i> (sensitivity)	$TPR = N_{TP} / (N_{TP} + N_{FN})$
Частота правильных отрицательных результатов (специфичность)	True negative rate, <i>TNR</i> (specificity)	$TNR = N_{TN} / (N_{TN} + N_{FP})$

<i>1</i>	<i>2</i>	<i>3</i>
Прогностичность положительного результата	Positive predictive value, <i>PPV</i>	$PPV = N_{TP} / (N_{TP} + N_{FP})$
Прогностичность отрицательного результата	Negative predictive value, <i>NPV</i>	$NPV = N_{TN} / (N_{TN} + N_{FN})$
Эффективность	Efficiency, <i>E</i>	$E = (N_{TP} + N_{TN}) / (N_{TP} + N_{TN} + N_{FP} + N_{FN})$
Индекс правдоподобия	Likelihood ratio, <i>LR</i>	$LR = (1 - FNR) / FPR$
Индекс Юдена	Youden index, <i>YI</i>	$YI = 100(TPR + TNR - 1)$
Предел обнаружения	<i>LOD</i> , minimum detectable (true) value (IUPAC), minimum detectable net concentration (ISO), detection limit or limit of detection (IUPAC, EURACHEM, British Pharmacopoeia, United States Pharmacopoeia)	
Предельное разбавление, минимальное разбавление		«Минимальная концентрация раствора, при которой реакция дает заметный результат» [93]
Предел идентификации, открываемый минимум		Концентрация, при превышении которой вероятность обнаружения аналита превышает 50% [93]

1	2	3
Уверенно открываемый минимум (абсолютная чувствительность)		«Минимальное количество открываемого иона, которое дает во всех случаях вполне отчетливый положительный эффект реакции при оптимальных условиях ее проведения» [100]

Заключение об отсутствии аналита в пробе представляет собой результат испытания статистической гипотезы H_0 , формулируемой, согласно [99], следующим образом:

- нулевая гипотеза H_0 : образец содержит аналит А;
- альтернативная гипотеза $\overline{H_0}$: образец не содержит А.

При такой формулировке H_0 ложное заключение об отсутствии аналита А является ошибкой I рода. Ошибочное принятие гипотезы H_0 представляет собой ошибку II рода. Согласно рекомендациям [101], выражение «ошибка первого рода» следует употреблять для обозначения той из двух возможных ошибок, которой важнее избежать¹.

Согласно [64, 102], достоверность (R , %) можно определить как

$$R = 100 - \alpha - \beta, \quad (1.1)$$

где α и β – вероятности ошибок I и II рода (%), соответственно. Для практического применения наиболее ценны методики с низкой вероятностью ошибки I рода, низкая вероятность ошибок II рода также желательна.

Следуя рекомендациям [79, 99], достоверность методики с бинарным откликом можно оценить следующим образом. Для набора образцов, заведомо

¹ Альтернативная формулировка гипотез – H_0 : аналит А отсутствует и $\overline{H_0}$: аналит А присутствует – нецелесообразна, так как в этом случае наиболее значимая ошибка не является ошибкой I рода. Ложное заключение о присутствии аналита считают более значимым лишь в редких ситуациях, например, при контроле спортсменов на применение допинга [87].

содержащих аналит **A**, по результатам испытаний найти число правильных положительных (N_{TP}) и ложных отрицательных результатов (N_{FN}), а по ним – частоту ложных отрицательных результатов

$$\epsilon = N_{FN} / (N_{FN} + N_{TP}). \quad (1.2)$$

Величина ϵ представляет выборочную оценку вероятности ошибки I рода α .

Для набора образцов, заведомо не содержащих аналит **A** (точнее, содержащих лишь следовые количества **A**), определить число правильных отрицательных (N_{TN}) и ложных положительных (N_{FP}) результатов, а по этим данным – частоту ложных положительных результатов

$$\beta = N_{FP} / (N_{FP} + N_{TN}). \quad (1.3)$$

Величина β дает выборочную оценку вероятности ошибки II рода β .

Зная N_{FN} , N_{TP} , N_{FP} и N_{TN} , легко рассчитать, помимо ϵ и β , и другие метрологические характеристики (табл. 1.1). Впрочем, большинство из них не дает какой-либо новой информации о методике обнаружения и введены, скорее всего, лишь в стремлении провести параллели с метрологией количественного анализа, в котором применяют большой набор характеристик (предел обнаружения, предел определения, чувствительность, селективность и т. д.).

Поскольку частоты ложных результатов – случайные биномиально распределенные величины, число испытаний, необходимое для определения α или β с заданной точностью, легко рассчитать. Например, для оценки $\alpha = 5\%$ с относительной погрешностью 20% при доверительной вероятности 95% нужно выполнить 1825 испытаний [99]. Чем выше достоверность методики, тем больше испытаний требуется для получения сколько-нибудь достоверной численной оценки достоверности.

Для оценки достоверности методик с бинарным откликом представляется более перспективным другой путь, основанный на изучении зависимости вероятности обнаружения аналита (P) от его концентрации (c) (рис. 1.3). Такие зависимости называют «кривыми эффективности» (в англоязычной литературе используют термины «performance characteristic curve», «probability-concentration

graph», «probability graph», «sensitivity curve» [70, 74, 79, 86, 89, 103-105])¹.

Очевидно, что вероятность ошибки I рода

$$\alpha = 100 - P(c). \quad (1.4)$$

Первым исследование кривых эффективности предпринял Н.П. Комарь [93], и этот подход получил значительное распространение в работах отечественных исследователей (см., например, [65, 67, 68, 75, 108]).

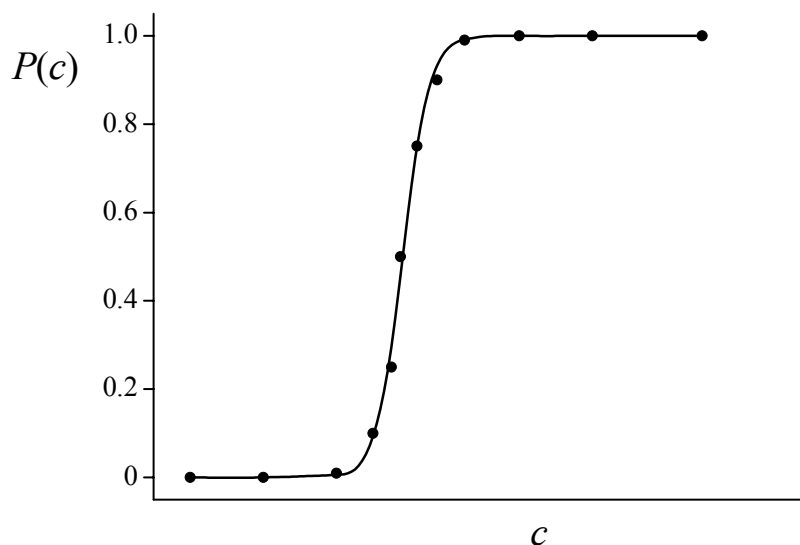


Рис. 1.3. Кривая эффективности.

В идеальном случае кривая эффективности представляет собой единичную смещенную функцию Хевисайда

$$P(c) = \begin{cases} 0, & c < c^* ; \\ 1, & c > c^* , \end{cases} \quad (1.5)$$

где c^* – некоторая пороговая концентрация. В этом случае методика с бинарным откликом всегда приводит к обнаружению аналита при $c > c^*$ (область нулевой

¹ Бросается в глаза аналогия между кривыми эффективности и кривыми «доза-эффект» («dose-response curve»), отражающим биологический ответ системы на дозу (концентрацию) действующего соединения [106, 107]. Частный случай кривых «доза-эффект», «кривые выживания», выражают зависимость доли погибших объектов (организмов, клеток и т. д.) от дозы яда или радиации. Как и кривые эффективности, многие кривые «доза-эффект» возрастают от 0 до 1.

вероятности ошибки I рода) и к выводу о его отсутствии при $c < c^*$ (область нулевой вероятности ошибки II рода).

На практике пороговая концентрация размывается в **интервал ненадежности (Δc)**. Интервал ненадежности – интервал концентраций аналита, в которой для части идентичных проб приходят к положительному заключению о присутствии аналита, а для части – к отрицательному. В этом случае вместо пороговой концентрации c^* приходится обсуждать левую и правую границы интервала ненадежности. На существование такого интервала обратил внимание еще 100 лет назад Ф. Эмих [109], а детальные исследования были начаты Н.П. Комарем [93]. В англоязычной литературе указания на важнейшую роль интервала ненадежности появились почти 50 лет спустя [110].

Согласно [99, 104], границами интервала ненадежности (нижней и верхней пороговыми концентрациями) предлагается считать такие концентрации $c_{5\%}$ и $c_{95\%}$, для которых вероятности обнаружения аналита больше 5% и меньше 95%, соответственно (концентрацию $c_{95\%}$ часто называют пределом обнаружения).

Учитывая многообразие предлагаемых метрологических характеристик, неустойчивость и, в ряде случаев, противоречивость используемой терминологии, актуально формирование ограниченного перечня характеристик с точным их определением. При этом важно исключить неверное понимание терминов.

Ключевой метрологической характеристикой методики с бинарным откликом мы предлагаем считать **интервал ненадежности Δc** . Поскольку при скрининге более важно избежать ошибки I рода, **в качестве верхней границы целесообразно принять $c_{99\%}$ (при концентрациях, превышающих $c_{99\%}$, $\alpha < 1\%$), а нижним пороговым значением можно считать $c_{5\%}$** .

Особо следует обсудить принципиальное для аналитической химии понятие предела обнаружения [111-114]. Согласно [97], пределом обнаружения называется «наименьшее содержание аналита, при котором он может быть обнаружен по данной методике анализа вещества или материала с заданной доверительной вероятностью» (ср. с определением ИЮПАК: «the minimum single result which, with a stated probability, can be distinguished from a suitable blank value» [111]).

Иными словами, пределом обнаружения считают минимальное содержание аналита, которое можно обнаружить, но не определить количественно по данной методике. Значение предела обнаружения находят, основываясь на теории проверки статистических гипотез и оценке вероятности ложных положительных и ложных отрицательных выводов [112].

Хотя определение, детализированное для количественного анализа («за предел обнаружения обычно принимают содержание аналита, равное сумме результата холостого опыта и его стандартного отклонения, умноженного на коэффициент, соответствующий заданной доверительной вероятности (например, для $P = 0.99$, $K = 3$)» [97]), нельзя соотнести с характеристиками методик с бинарным откликом, ничто не мешает пользоваться более общим определением и для методик обнаружения с бинарным откликом **пределом обнаружения считать правую границу интервала ненадежности, т.е. концентрацию аналита, при превышении которой вероятность ошибки I рода меньше 1%.**

Учитывая, что из нескольких методик для практического применения предпочтительнее та, для которой интервал ненадежности меньше, полезной **вспомогательной характеристикой может служить относительная ширина интервала ненадежности [83] $\Delta c / c_{99\%}$.**

Нельзя не согласиться с предложением [79, 115] отказаться в метрологии методик с бинарным откликом от понятия «неопределенность» и заменить его «недостоверностью». Вместе с тем, с учетом различной значимости ошибок I и II рода и существования интервала ненадежности, выражение для достоверности (R), включающее вероятности обеих ошибок (уравнение (1.1)), представляется неудачным. Мы предлагаем **«недостоверность» связывать только с вероятностью более значимой ошибки I рода.** В этом случае понятие достоверности вполне соответствует прикладной аналитической задаче, а выражение для R приобретает исключительно простой вид:

$$R_{\text{обнаружения}} = 100 \cdot P(c), \% \quad (1.6)$$

Недостоверность в таком случае равна $100 - R$.

1.3. Построение кривых эффективности

На практике кривую эффективности подбирают как функцию, аппроксимирующую эмпирическую зависимость частот обнаружения $P^{\text{эмп}}$ от c . Для получения такой зависимости выбирают несколько (M) концентраций c_i внутри интервала ненадежности, для каждой из них проводят N_i испытаний, фиксируют число положительных результатов (n_i) и вычисляют эмпирическую частоту обнаружения аналита

$$P_i^{\text{эмп}} = \frac{n_i}{N_i}. \quad (1.7)$$

Для аппроксимации зависимости $P^{\text{эмп}}$ от c можно, в принципе, использовать любую неубывающую функцию, ограниченную 0 и 1. Такими свойствами обладают, например, интегральные функции распределения случайных величин [116].

Снижению трудозатрат и повышению надежности определения метрологических характеристик по кривым эффективности помогло бы ограничение набора выражений для описания зависимостей $P(c)$ лишь несколькими обоснованно рекомендованными к применению функциями.

Возможный путь решения этой задачи, по крайней мере, для методик с визуальной индикацией, связан с теоретическим анализом различий в цветовосприятии наблюдателей в зависимости от условий проведения хромогенных реакций. К сожалению, в настоящее время доступны лишь отрывочные данные [82, 117, 118], недостаточные для построения работоспособных моделей, и лишь недавно появились указания на то, что вид зависимости $P(c)$ в большей мере связан с типом хромогенной реакции, чем со средой, в которой реакция проводится [83, 119].

Другой подход требует проанализировать на значительном массиве методик обнаружения с бинарным откликом применение различных функций, описывающих кривые эффективности, и рекомендовать для практического использования те, которые в большинстве случаев адекватно воспроизводят

зависимости $P^{\text{эмп}}$ от c . При этом для расчета подгоночных параметров аппроксимирующих функций необходимы надежные вычислительные процедуры.

Наиболее часто зависимость $P^{\text{эмп}}$ от c описывали функцией нормального распределения, применяли также функции логнормального, экспоненциального, логистического распределений, распределения Вейбулла [65, 75, 83, 119-125]. Высказывались мнения, что различные функции близки по качеству аппроксимации экспериментальных данных [99], а для практического применения можно ограничиться использованием только функции нормального распределения [75]. Подгоночные параметры (параметры функций распределения) определяли как приближенными графическими методами (являющимися, фактически, методами пробит-анализа, см. ниже) [65, 67, 122-125], так и с использованием нелинейного МНК [68, 119, 121].

В родственной задаче описания кривых «доза-эффект» зависимость $P^{\text{эмп}}$ от c часто аппроксимируют функциями логнормального (автоматически отсекает бессмысленную область отрицательных доз) и логистического распределений (обе функции описывают S -образные кривые) [106, 126-128], а также функцией распределения Вейбулла [129], способной описывать и выгнутые кривые, встречающиеся на практике [130].

Именно при решении задач описания кривых «доза-эффект» наиболее часто начал использоваться метод пробит-анализа (пробит-метод)¹. Пробит-анализ – один из распространенных методов математической статистики, применяемый в биологических и медицинских исследованиях для обработки S -образных кривых зависимости встречаемости учитываемого в альтернативной форме эффекта (т.е. в форме «есть эффект – нет эффекта», например «есть гибель – нет гибели», «есть фармакологический эффект – нет фармакологического эффекта») при воздействии каких-либо агентов (физических, химических, лекарственных) от величины их дозы.

¹ Термин «пробит» происходит от англ. probability unit – вероятностная единица. Пробит-анализ – количественная оценка экспериментальных данных, основанная на изучении зависимости между дозами (чаще их логарифмами) и пробитами, соответствующими наблюдавшимся эффектам.

Сущность этого метода заключается в том, что он позволяет провести линеаризацию S -образной кривой, т.е. ее преобразование в прямую линию, которая может быть обработана относительно простыми методами, используемыми для анализа линейной зависимости.

При использовании метода пробит-анализа подобное преобразование обычно осуществляется, следующим образом. По оси ординат откладывается не отклик $P^{\text{эмп}}$ наблюдаемого эффекта (процент гибели организмов при проведении токсикологических исследований, процент положительных ответов при решении рассматриваемой задачи установления метрологических характеристик методик качественного анализа и др.), а значения статистической функции $\Psi(P^{\text{эмп}})$ (Ψ – функция, обратная функции распределения), или пробитов, соответствующие различным уровням проявления эффекта $P^{\text{эмп}}$. По оси абсцисс в этом случае откладывается предиктор модели – $pred$ (в медицинских исследованиях чаще всего предиктором выступает не доза D , а ее десятичный логарифм $\lg D$, в общем же случае вид предиктора не лимитирован). Таким образом, полученные экспериментальные данные представляются в координатах « $f(pred) - \Psi(P^{\text{эмп}})$ ». При этом значения $f(pred)$ рассчитываются, а значения функции $\Psi(P^{\text{эмп}})$, соответствующие различным уровням эффекта $P^{\text{эмп}}$, определяются по специальным таблицам [106].

Метод пробит-анализа использовался [65, 67, 122-125] для оценки подгоночных параметров кривых эффективности. В табл. 1.2 приведены использовавшиеся в этих работах функции и соответствующие им предикторы, математические выражения для $P(c)$ этих и ряда других функций указаны в табл. 1.2.

После представления кривой эффективности в линейных координатах, процедура оценки ее параметров не вызывает затруднений. Для этого достаточно воспользоваться линейным МНК и рассчитать параметры уравнения

$$\Psi(P^{\text{эмп}}) = k \cdot f(pred) + t, \quad (1.8)$$

где $\Psi(P^{\text{эмп}})$ является пробитом (в общем понимании), а k и t – коэффициенты уравнения прямой.

Таблица 1.2.

Функции распределения, использовавшиеся для определения подгоночных параметров кривых эффективности пробит-методом

Распределение	Выражение для вычисления вероятности	Подгоночные параметры
Логнормальное	$P(c) = \frac{1}{s\sqrt{2\pi}} \int_0^c \exp\left(-\frac{1}{2}\left(\frac{\ln(x/\bar{c})}{s}\right)^2\right) dx$	$\bar{c} > 0, s$
Нормальное	$P(c) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^c \exp\left(-\frac{1}{2}\left(\frac{\bar{c}-x}{s}\right)^2\right) dx$	$\bar{c} > 0, s$
Экспоненциальное	$P(c) = 1 - \exp\left(-\frac{c-a}{b}\right)$	$a > 0, b > 0$

Можно также, определить коэффициент линейной корреляции r , который выступает мерой адекватности модели.

$$r = \frac{\sum_i (y_i - \bar{y}) \cdot (\xi_i - \bar{\xi})}{\sqrt{\sum_i (y_i - \bar{y})^2 \cdot (\xi_i - \bar{\xi})^2}}, \quad (1.9)$$

где y_i и \bar{y} – экспериментальные точки ($\Psi(P^{\text{эмп}})$ и среднее из набора $\Psi(P^{\text{эмп}})$ по i экспериментальным точкам), ξ_i и $\bar{\xi}$ – соответствующие значения, оцененные в рамках модели.

Вычисляя коэффициенты корреляции линейных моделей по формуле (1.9), представляется возможным выбрать из набора проверяемых моделей ту, которая лучше других воспроизводит экспериментальные данные.

При всей наглядности графических методов, к которым можно отнести и пробит-метод, их можно рекомендовать, скорее, для предварительного исследования гипотез о распределениях: во-первых, далеко не все возможные распределения $P^{\text{эмп}}(c)$ допускают получение линейных зависимостей; во-вторых,

потребность в объективной информации о метрологических характеристиках методик тестового анализа предполагает использование численных статистических оценок. Кроме того, при расчете параметров законов распределения на основе применения МНК к линеаризованным зависимостям нарушаются основные предпосылки МНК.

Следует упомянуть и упрощенный способ построения аппроксимирующей функции $P(c)$, предложенный по итогам выполнения проекта MEQUALAN [79, 104]. Пользуясь этим способом, кривую эффективности представляли ломаной линией, принимая, что в области ненадежности зависимость $P^{\text{эмп}}$ линейно зависит от c (рис. 1.4). Хотя этот способ построения аппроксимирующей функции привлекает не слишком большим числом необходимых испытаний, рекомендовать его для применения невозможно. Действительно, описание в интервале ненадежности зависимости $P^{\text{эмп}}$ от c прямой линией лишает суждения о поведении функции $P(c)$ в окрестности концов интервала ненадежности всякой достоверности. Между тем, окрестности «плечей» (области, где вероятности положительного отклика близки к 0 и 1) наиболее важны для определения формы кривой эффективности и нахождения границ интервала ненадежности [103].

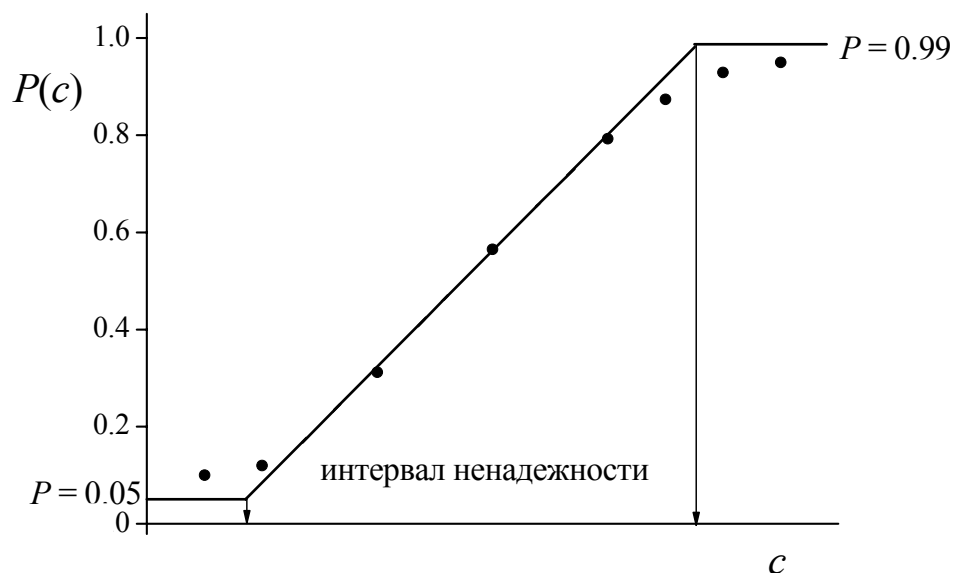


Рис. 1.4. Представление кривой эффективности ломаной линией согласно рекомендациям [79, 104].

В настоящей работе для аппроксимации большого набора зависимостей $P^{\text{эмп}}$ от c испытан расширенный набор функций (табл. 1.4); подгоночные параметры определены надежными численными методами; приведено сравнение качества аппроксимации, погрешности расчета параметров и трудоемкости расчетов; выработаны рекомендации по ограничению набора аппроксимирующих функций.

Таблица 1.4.

**Функции распределения,
испытанные при аппроксимации зависимости $P^{\text{эмп}}$ от c**

Распределение	Выражение для $P(c)$, область определения, подгоночные параметры
Экстремального значения первого рода (Вейбулла)	$P(c) = 1 - \exp\left(-\left(\frac{c-a}{b}\right)^k\right), a < c < \infty, a > 0, b > 0, k$
Лапласа	$P(c) = \frac{1}{2k} \int_{-\infty}^c \exp\left(-\left \frac{\bar{c}-x}{k}\right \right) dx, -\infty < c < \infty, \bar{c} > 0, k$
Логистическое	$P(c) = \frac{1}{1 + \exp\left(-\frac{c-k}{t}\right)}, k < c < \infty, k > 0, t > 0$
Логнормальное	$P(c) = \frac{1}{s\sqrt{2\pi}} \int_0^c \exp\left(-\frac{1}{2}\left(\frac{\ln(x/\bar{c})}{s}\right)^2\right) dx, 0 < c < \infty, \bar{c} > 0, s$
Нормальное	$P(c) = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^c \exp\left(-\frac{1}{2}\left(\frac{\bar{c}-x}{s}\right)^2\right) dx, -\infty < c < \infty, \bar{c} > 0, s$
Пуассона *	$P(Q) = \sum_{j=0}^Q \frac{(\bar{c})^j \exp(-\bar{c})}{j!}, Q \geq 0, \bar{c} > 0$
Экспоненциальное	$P(c) = 1 - \exp\left(-\frac{c-a}{b}\right), a < c < \infty, a > 0, b > 0$

* Поскольку распределение Пуассона описывает распределение дискретных случайных величин, интервал ненадежности Δc разбивали на M подынтервалов равной длины и аппроксимировали функцией $P(i)$ частоты $P_i^{\text{эмп}}$, $i = 1, 2, \dots, M$.

Сведения о методиках обнаружения аналитов и соответствующие ссылки на литературу представлены в табл. 1.5. Хромогенные реакции проводились в растворах, на фильтровальной бумаге, реагентных индикаторных бумагах (РИБ), пенополиуретане (ППУ), поверхности геля метилкремниевой кислоты (МКК), фторопластовых пластинах, в желатиновых пленках; исследовались реакции различного типа с участием широкого набора реагентов и аналитов.

Таблица 1.5.

Характеристики исходных данных

№	Аналит / реагент	Носитель	№ таблицы с исходными данными	Ссылка
1	2	3	4	5
1.	ДХА	Al ₂ O ₃	1.6	[119]
2.	ΣМ / РИБ ⁽¹⁾	Бумага	1.7	[68]
3.	ΣМ / РИБ ⁽²⁾		1.8	[68]
4.	Fe ²⁺ / РИБ		1.9	[131]
5.	NO ₂ ⁻ / РИБ		1.10	[131]
6.	Ph ₄ B ⁻ / C ₂₂ H ₂₉ N ₂ Cl ⁽³⁾		1.11	[132]
7.	Cl ₂ / Тиокетон Михлера		1.12	[122]
8.	Fe ^{2,3+} / Фенилфлуорон		1.13	[123]
9.	Co ³⁺ / 2-нитрозо-1-нафтол		1.14	[124]
	ΣМ / ПАР		Желатиновая пленка	1.15
10.	Co ²⁺ / Нитрозо-R-соль	1.16		[131]
11.	ΣМ / ПАР	МКК	1.17	[131]

1	2	3	4	5
12.	Co^{2+} / 2-нитрозо-1-нафтол	ППУ	1.18	[131]
13.	Co^{2+} / SCN^-		1.19	[131]
14.	NO_2^-		1.20	[131]
15.	ΣM / ПАР ⁽⁴⁾		1.21	[131]
16.	ΣM / ПАР ⁽⁵⁾		1.22	[131]
	Ni^{2+} / $\text{C}_2\text{H}_4\text{N}_2\text{S}_2$ ⁽⁶⁾	Раствор	1.23	[67]
17.	Co^{2+} / 2-нитрозо-1-нафтол		1.24	[131]
18.	ΣM / ПАР		1.25	[131]
19.	Fe^{3+} / KSCN		1.26	[93]
20.	Fe^{2+} / $\text{C}_8\text{H}_{12}\text{O}_2$ ⁽⁷⁾		1.27	[119]
21.	Na^+ / $\text{K}_2\text{C}_5\text{O}_5$		1.28	[121]
22.	K^+ / $\text{Na}_2\text{C}_5\text{O}_5$		1.29	[121]
23.	Li^+ / $\text{K}_2\text{C}_5\text{O}_5$	1.30	[65]	
24.	ΣM / КО	Ксерогель	1.31	[133]
25.	Cu^{2+} / КЦ		1.32	[133]

ДХА – дихлоранилин, ΣM – сумма ионов металлов, ПАР – 4-(2-пиридилазо)резорцин, КО – ксиленоловый оранжевый, КЦ – кальцеин.

⁽¹⁾ режим «с концентрированием», ⁽²⁾ режим «без концентрирования», ⁽³⁾ краситель катионный розовый 2С, ⁽⁴⁾ комплексы сорбируются из раствора, ⁽⁵⁾ ПАР закреплен на таблетках ППУ, ⁽⁶⁾ рубеановодородная кислота, ⁽⁷⁾ димедон

Таблица 1.6.

**Экспериментальные данные
для тест-системы No 1**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{ЭМП}}$
30	9	73	0.123
40	17	73	0.233
60	50	92	0.543
80	67	92	0.728
100	78	92	0.848
150	89	92	0.967

Таблица 1.8.

**Экспериментальные данные
для тест-системы No 3**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{ЭКСП}}$
2.00	30	316	0.095
2.25	82	300	0.273
2.50	163	316	0.516
2.75	205	300	0.683
3.00	259	300	0.863
3.25	271	300	0.903
3.50	308	316	0.975
3.75	296	300	0.987

Таблица 1.7.

**Экспериментальные данные
для тест-системы No 2**

№	$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{ЭМП}}$
1	$3.7\cdot 10^{-3}$	34	300	0.113
2	$3.8\cdot 10^{-3}$	78	280	0.279
3	$3.9\cdot 10^{-3}$	132	300	0.440
4	$4.0\cdot 10^{-3}$	181	280	0.646
5	$4.1\cdot 10^{-3}$	240	300	0.800
6	$4.2\cdot 10^{-3}$	249	280	0.889
7	$4.3\cdot 10^{-3}$	275	300	0.917

Таблица 1.9.

**Экспериментальные данные
для тест-системы No 4**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{ЭКСП}}$
0.0056	26	100	0.260
0.0084	34	100	0.340
0.0112	49	100	0.490
0.0140	53	100	0.530
0.0168	60	100	0.600
0.0196	65	100	0.650
0.0223	74	100	0.740
0.0251	81	100	0.810
0.0279	85	100	0.850

Таблица 1.10.

**Экспериментальные данные
для тест-системы No 5**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.05	22	152	0.146
0.10	29	152	0.191
0.15	46	152	0.301
0.20	62	152	0.408
0.25	85	152	0.559
0.30	100	152	0.655
0.35	118	152	0.779
0.40	129	152	0.846
0.45	138	152	0.908
0.50	146	152	0.959

Таблица 1.12.

**Экспериментальные данные
для тест-системы No 7**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.03	7	66	0.106
0.04	15	66	0.227
0.05	38	66	0.576
0.06	43	66	0.652
0.07	55	66	0.833
0.08	63	66	0.955
0.10	65	66	0.985

Таблица 1.11.

**Экспериментальные данные
для тест-системы No 6**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
2.90	2	98	0.020
3.05	1	42	0.024
3.20	5	88	0.057
3.70	8	99	0.081
4.10	21	98	0.214
4.80	15	42	0.357
6.40	57	75	0.760
8.0	41	42	0.976

Таблица 1.13.

**Экспериментальные данные
для тест-системы No 8**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.05	2	69	0.029
0.10	15	69	0.217
0.15	39	69	0.565
0.20	63	69	0.913
0.25	68	69	0.985

Таблица 1.14.

**Экспериментальные данные
для тест-системы No 9**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.10	3	40	0.075
0.20	8	40	0.200
0.30	25	40	0.625
0.40	37	40	0.925
0.50	39	40	0.975

Таблица 1.16.

**Экспериментальные данные
для тест-системы No 11**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.103	15	100	0.150
0.118	26	100	0.260
0.133	36	100	0.360
0.147	46	100	0.460
0.162	60	100	0.600
0.177	67	100	0.670
0.192	77	100	0.770
0.206	86	100	0.860
0.221	94	100	0.940

Таблица 1.15.

**Экспериментальные данные
для тест-системы No 10**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.15	21	100	0.210
0.16	28	100	0.280
0.17	42	100	0.420
0.18	52	100	0.520
0.19	66	100	0.660
0.20	76	100	0.760
0.21	78	100	0.780
0.22	86	100	0.860
0.23	94	100	0.940

Таблица 1.17.

**Экспериментальные данные
для тест-системы No 12**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.30	30	100	0.300
0.32	44	100	0.440
0.34	52	100	0.520
0.36	62	100	0.620
0.38	74	100	0.740
0.40	87	100	0.870
0.42	94	100	0.940

Таблица 1.18.

**Экспериментальные данные
для тест-системы No 13**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.477	21	100	0.210
0.483	42	100	0.420
0.489	43	100	0.430
0.495	39	100	0.390
0.501	67	100	0.670
0.507	65	100	0.650
0.513	69	100	0.690

Таблица 1.20.

**Экспериментальные данные
для тест-системы No 15**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.01	11	291	0.037
0.02	45	293	0.154
0.03	90	308	0.292
0.04	140	292	0.479
0.05	200	299	0.669
0.06	264	292	0.904
0.07	284	312	0.910
0.08	255	263	0.970
0.09	289	302	0.957
0.10	305	318	0.959

Таблица 1.19.

**Экспериментальные данные
для тест-системы No 14**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.020	18	100	0.180
0.022	39	100	0.390
0.024	46	100	0.460
0.026	62	100	0.620
0.028	90	100	0.900
0.030	93	100	0.930
0.032	99	100	0.990

Таблица 1.21.

**Экспериментальные данные
для тест-системы No 16**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.0125	22	100	0.220
0.0150	35	100	0.350
0.0175	43	100	0.430
0.0200	54	100	0.540
0.0225	65	100	0.650
0.0250	79	100	0.790
0.0275	82	100	0.820
0.0300	86	100	0.860
0.0325	94	100	0.940

Таблица 1.22.

**Экспериментальные данные
для тест-системы No 17**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.0030	22	100	0.220
0.0050	36	100	0.360
0.0070	48	100	0.480
0.0090	67	100	0.670
0.0110	78	100	0.780
0.0130	92	100	0.920

Таблица 1.24.

**Экспериментальные данные
для тест-системы No 19**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.0884	8	100	0.080
0.118	24	100	0.240
0.147	53	100	0.530
0.177	71	100	0.710
0.206	73	100	0.730
0.236	90	100	0.900
0.265	96	100	0.960

Таблица 1.23.

**Экспериментальные данные
для тест-системы No 18**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.005	3	30	0.100
0.007	5	30	0.167
0.010	9	30	0.300
0.013	15	30	0.500
0.015	18	30	0.600
0.017	23	30	0.767
0.020	28	30	0.933
0.030	29	30	0.967

Таблица 1.25.

**Экспериментальные данные
для тест-системы No 20**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.032	24	100	0.240
0.036	34	100	0.340
0.040	47	100	0.470
0.044	65	100	0.650
0.048	74	100	0.740
0.052	87	100	0.870
0.056	94	100	0.940

Таблица 1.26.

**Экспериментальные данные
для тест-системы No 21**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.022	3	24	0.125
0.032	2	25	0.080
0.045	5	27	0.185
0.054	10	25	0.400
0.065	6	27	0.222
0.086	8	25	0.320
0.108	19	27	0.703
0.162	22	26	0.846

Таблица 1.28.

**Экспериментальные данные
для тест-системы No 23**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
17.97	75	212	0.354
18.97	157	346	0.454
20.00	221	346	0.639
21.00	232	346	0.671
22.00	255	346	0.737
23.00	288	346	0.832
24.00	280	346	0.809
25.00	309	346	0.893
26.00	129	134	0.963

Таблица 1.27.

**Экспериментальные данные
для тест-системы No 22**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
4.40	7	50	0.140
4.50	11	50	0.220
4.60	20	50	0.400
4.70	29	50	0.580
4.80	44	50	0.880

Таблица 1.29.

**Экспериментальные данные
для тест-системы No 24**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эмп}}$
13.5	1	40	0.025
14.0	3	40	0.075
14.5	8	20	0.400
15.0	9	20	0.450
15.5	11	20	0.550
16.0	44	81	0.543
16.5	12	20	0.600
17.0	55	80	0.688
17.5	32	40	0.800
18.0	69	81	0.852

Таблица 1.30.

**Экспериментальные данные
для тест-системы No 25**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
55.00	32	228	0.140
57.50	68	228	0.298
60.00	108	228	0.474
62.50	147	226	0.650
65.00	188	228	0.825
67.50	200	221	0.905
70.00	127	134	0.948
72.50	65	68	0.956

Таблица 1.32.

**Экспериментальные данные
для тест-системы No 27**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.05	12	100	0.120
0.10	20	100	0.200
0.20	47	100	0.470
0.40	67	100	0.670
0.60	87	100	0.870
0.80	93	100	0.930

Таблица 1.31.

**Экспериментальные данные
для тест-системы No 26**

$c_i, \text{мг}\cdot\text{л}^{-1}$	n_i	N_i	$P_i^{\text{эксп}}$
0.60	20	100	0.200
0.80	67	100	0.670
0.90	87	100	0.870
1.00	93	100	0.930
1.20	99	100	0.990

При расчете подгоночных параметров аппроксимирующих функций следует учитывать, что положительный («ДА, анализ обнаружен») и отрицательный («НЕТ, анализ не обнаружен») результаты анализа – это исходы случайного эксперимента с известными вероятностями «успеха» (P) и «неудачи» ($1 - P$). Если допустить, что единственной переменной, определяющей вероятность отклика «ДА» является концентрация аналита c , то число положительных результатов n_i в N_i испытаниях – случайная величина, подчиняющаяся закону биномиального распределения:

$$P(n_i, N_i) = \frac{n_i!(N_i - n_i)!}{N_i!} P(c_i)^{n_i} [1 - P(c_i)]^{(N_i - n_i)}. \quad (1.10)$$

Функция правдоподобия – вероятность получения совокупности всех независимых результатов – это произведение величин (1.10) для разных концентраций c_i :

$$L(|\theta\rangle, |n_i\rangle) = \prod_{i=1}^M P(n_i, N_i) = \prod_{i=1}^M \frac{n_i!(N_i - n_i)!}{N_i!} P(c_i)^{n_i} [1 - P(c_i)]^{(N_i - n_i)}, \quad (1.11)$$

где $|\theta\rangle$ – вектор параметров функции $P(c_i)$. Применяя метод максимума правдоподобия [128], $|\theta\rangle$ находили [121] как оценки, обращающие функцию (1.11) в максимум. Вычисление параметров кривых эффективности методом максимума правдоподобия не получило широкого распространения, возможно, вследствие довольно значительных вычислительных трудностей.

1.3.1. Оценка адекватности моделей. Расчет ковариационных матриц подгоночных параметров. В настоящей работе $|\theta\rangle$ находили взвешенным нелинейным МНК [134, 135] как оценки, обращающие в минимум функционал

$$\chi_{\text{эксп}}^2 = \sum_{i=1}^M \xi_i^2, \quad (1.12)$$

где взвешенные невязки

$$\xi_i = \frac{P_i^{\text{эмп}} - P_i}{s(P_i)}, \quad (1.13)$$

\hat{P}_i – рассчитанное значение частоты для концентрации c_i , $s(P_i)$ – стандартное отклонение P_i .

МНК-оценки являются оценками максимального правдоподобия при нормальном распределении n_i [128]. В рассматриваемой задаче они приближаются к оценкам максимального правдоподобия с ростом числа испытаний и являются асимптотически состоятельными, несмещенными и эффективными [136, 137].

Минимизацию критерия (1.12) проводили численно методами деформируемого многогранника и скорейшего спуска [138, 139].

Метод деформируемого многогранника (симплекса) является методом оптимизации нулевого порядка (не требует вычисления частных производных по искомым параметрам) и оперирует набором из $p+1$ точек $x_1, x_2, \dots, x_p, x_{p+1}$, которые упорядочены таким образом, что для соответствующих точек выполняются неравенства

$$U(x_{p+1}) \geq U(x_p) \geq \dots \geq U(x_1), \quad (1.14)$$

где U – оптимизируемый функционал.

Эти точки интерпретируются как вершины многогранника в p -мерном пространстве. При $p=2$ многогранник является треугольником, при $p=3$ – пирамидой и т.д. На каждой итерации текущий многогранник заменяется новым: «худшая» вершина x_{p+1} отбрасывается и вместо нее в набор вводится «более подходящая» вершина. На рис. 1.5 наглядно показана работа метода при $p=2$.

Вначале выбирают 3 случайных точки A, B и C , в которых определяют значение оптимизируемого критерия U . Пусть $U(A) > U(B) > U(C)$. Для создания нового симплекса необходимо выполнить следующую процедуру:

1. Рассчитать значение $U(E)$, где E – отражение точки A , имеющей наибольшее (в случае поиска минимума) значение оптимизируемой функции.
2. При $U(E) < U(C)$ направление минимизации признается удачным и рассчитывается значение точки $U(J)$, где J – еще один шаг вдоль вектора AE .

При $U(J) \leq U(C)$ новый симплекс – BCJ . Перейти к шагу 5.

При $U(J) > U(C)$ новый симплекс – BCE . Перейти к шагу 5.

3. При $U(B) \geq U(E) \geq U(C)$ новый симплекс – BCE . Перейти к шагу 5.

4. При $U(E) > U(B)$ делается заключение о том, что симплекс необходимо подвергнуть сжатию:

при $U(E) > U(A)$ рассчитать значение $U(G)$. Новый симплекс – BCG . Перейти к шагу 5;

при $U(A) \geq U(E) \geq U(B)$ рассчитать значение $U(H)$. Новый симплекс – BCH . Перейти к шагу 5.

5. Перейти к построению нового симплекса.

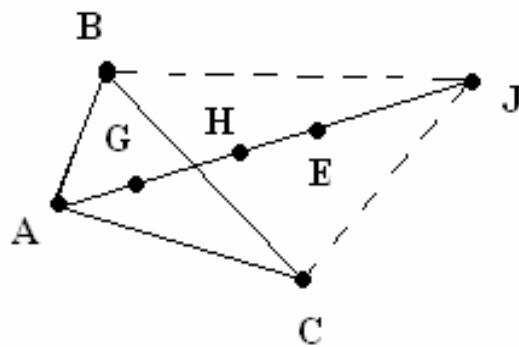


Рис. 1.5. Работа симплекс-метода при $p = 2$.

Критерием останова расчетов в этом методе является величина модуля разности значений оптимизируемой функции в вершинах симплекса, не превосходящая заданный порог (например, $1 \cdot 10^{-4}$).

Метод скорейшего спуска относится к методам оптимизации первого порядка (требуется численное или аналитическое вычисление производных по искомым параметрам оптимизируемой функции). Алгоритм метода основан на том, что экстремум функции наблюдается в точке, где частные производные по параметрам этой функции равны нулю.

Процесс оптимизации начинается с вектора начальных приближений $|\theta_0\rangle$, размерностью $(z \times z)$, где z – число подгоночных параметров в оптимизируемой функции. От вектора начальных приближений необходимо двигаться в направлении скорейшего изменения значения оптимизируемой функции; оно

совпадает с вектором, берущим начало из $|\theta_0\rangle$, координаты которого пропорциональны соответствующим частным производным в этой точке. Таким образом, значение функции перемещается в точку

$$|\theta_1\rangle = \left| \theta_0 - k \frac{\partial U}{\partial \theta_i} \right\rangle \quad (1.15)$$

Число k выбирается таким образом, чтобы координаты вектора $|\theta_1\rangle$ соответствовали наименьшему (в случае задачи поиска минимума) значению функции на всем луче, идущем в избранном направлении. Одновременно контролируются значения частных производных. Далее эта процедура повторяется, что приводит к улучшенным оценкам точки минимума $|\theta_2\rangle$, $|\theta_3\rangle$ и т.д.

Критерием останова расчетов в этом случае является одновременное значение модулей частных производных по всем параметрам, не превосходящее заданный порог (например, $1 \cdot 10^{-5}$).

Качество аппроксимации $P_i^{\text{ЭМП}}$ оценивали с помощью нескольких статистических критериев. При использовании критерия хи-квадрат аппроксимацию признавали адекватной, если выполнялось неравенство

$$\chi_{\text{эсп}}^2 < \chi_{f,5\%}^2, \quad (1.16)$$

где $\chi_{f,5\%}^2$ – 5%-ная точка распределения χ^2 с f степенями свободы ($f = M - z$, z – число подгоночных параметров функции).

Применяли также λ -критерий Колмогорова-Смирнова [140] и находили значение статистики

$$\lambda = \max_i \left| P_i^{\text{ЭМП}} - F_i \right| \cdot \sqrt{M}. \quad (1.17)$$

Аппроксимацию признавали адекватной, если вероятность $P(\lambda)$, найденная для наблюдаемого λ , превышала пороговое значение 5%.

В качестве критериев адекватности аппроксимации использовали также среднее выборочное значение взвешенных невязок $\bar{\xi}$ и среднее выборочное

значение модулей взвешенных невязок $|\bar{\xi}|$. Если распределение ξ_i подчиняется закону стандартного нормального распределения, математические ожидания составляют [141, 142]:

$$E(\bar{\xi}) = 0, \quad E(|\bar{\xi}|) = \sqrt{2/\pi} \approx 0.8. \quad (1.18)$$

Для адекватной модели выборочные оценки $\bar{\xi}$ и $|\bar{\xi}|$ близки к своим математическим ожиданиям.

Также рассчитывалось значение остаточной дисперсии

$$s_0^2 = \frac{1}{M-z} \sum_{i=1}^M \xi_i^2 = \frac{1}{M-z} \chi_{\text{эксп}}^2, \quad (1.19)$$

где z – число подгоночных параметров модели. Модель считают адекватной при выполнении неравенства (1.16).

Зная величину (1.19) и матрицу Гессе производных $H = \left\| \frac{\partial^2 U}{\partial \theta_i \partial \theta_j} \right\|$, размером $(z \times z)$, можно вычислить ковариационную матрицу подгоночных параметров

$$D(\theta^*) = s_0^2 [0.5H]^{-1} \quad (1.20)$$

Ковариационная матрица вектора подгоночных параметров $|\theta^*\rangle$ состоит из оценок дисперсий и ковариаций его компонентов

$$D(\theta^*) = \begin{pmatrix} s^2(\theta_1) & \text{cov}(\theta_1\theta_2) & \cdots & \text{cov}(\theta_1\theta_z) \\ \text{cov}(\theta_2\theta_1) & s^2(\theta_2) & \cdots & \text{cov}(\theta_2\theta_z) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\theta_z\theta_1) & \text{cov}(\theta_z\theta_2) & \cdots & s^2(\theta_z) \end{pmatrix} \quad (1.21)$$

Зная матрицу $D(\theta^*)$ можно вычислить общие коэффициента корреляции s_{ij} :

$$s_{ij} = \frac{D_{ij}}{D_{ii}D_{jj}} \quad (1.22)$$

При значениях s_{ij} , близких к ± 1 , модель переопределена.

1.3.2. Назначение статистических весов. Определенную проблему представляет назначение статистических весов $w_i = \frac{1}{s^2(P_i^{\text{эмп}})}$. Рассмотрим два подхода.

Наиболее обоснованной является модель, учитывающая, что частоты P_i , равно как и количества «успехов» n_i в N_i испытаниях, – случайные биномиально распределенные величины. Тогда дисперсии частот обнаружения

$$s^2(P_i) = \frac{P_i(1-P_i)}{N_i}, \quad (1.23)$$

что приводит к формуле для весов

$$w_i = \frac{N_i}{P_i^{\text{эмп}}(1-P_i^{\text{эмп}})}. \quad (1.24)$$

Другой способ назначения весов предполагает, что испытания по определению эмпирической частоты обнаружения аналита при его концентрации c_i повторяют J раз, проводя в каждой серии $(N_i)_j$ испытаний. Общее число испытаний для концентрации c_i

$$N_i^{\text{общ}} = \sum_{j=1}^J (N_i)_j. \quad (1.25)$$

Зная число «успехов» $(n_i)_j$ в каждой из серий испытаний, вычисляют среднее значение частоты обнаружения:

$$\overline{P_i^{\text{эмп}}} = \frac{1}{N_i^{\text{общ}}} \sum_{j=1}^J (n_i)_j. \quad (1.26)$$

Можно найти также частоты обнаружения в каждой из J серий:

$$(P_i^{\text{эмп}})_j = \frac{(n_i)_j}{(N_i)_j} \quad (1.27)$$

и, считая их равнозначными, оценить стандартное отклонение $\overline{P_i^{\text{эмп}}}$ как

$$s(\overline{P_i^{\text{ЭМП}}}) = \left(\frac{1}{J \cdot (J-1)} \sum_{j=1}^J \left[(P_i^{\text{ЭМП}})_j - \overline{P_i^{\text{ЭМП}}} \right]^2 \right)^{1/2}. \quad (1.28)$$

Для некоторых тест-систем частоты обнаружения аналитов были определены в нескольких сериях наблюдений. Как показали расчеты, оценки стандартных отклонений $s(P_i^{\text{ЭМП}})$, найденные по формуле (1.28), не слишком сильно отличаются от значений, рассчитанных по формуле (1.23), причем эти различия не оказывают существенного влияния на выбор вида аппроксимирующей функции и мало сказываются на оценках подгоночных параметров. Иллюстрацией может служить обработка данных, относящихся к тест системе No 5 (табл. 1.33 – 1.35).

Таблица 1.33.

**Экспериментальные частоты обнаружения NO_2^-
с использованием РИБ- NO_2^- -Тест (тест-система No 5)**

$c,$ $\text{мг} \cdot \text{л}^{-1}$	Число измерений в серии								
	17	10	17	14	19	23	20	16	16
	Частота обнаружения $P_i^{\text{ЭМП}}$ в серии								
0.05	0.18	0.40	0.06	0.13	0.05	0.23	0.15	0.06	0.13
0.10	0.29	0.30	0.12	0.20	0.10	0.23	0.20	0.13	0.19
0.15	0.41	0.20	0.29	0.33	0.15	0.32	0.40	0.25	0.31
0.20	0.59	0.60	0.24	0.40	0.21	0.32	0.60	0.37	0.44
0.25	0.35	0.80	0.30	0.53	0.63	0.68	0.70	0.50	0.56
0.30	0.53	0.70	0.41	0.66	0.84	0.71	0.75	0.56	0.69
0.35	0.88	0.90	0.59	0.73	0.75	0.95	0.75	0.63	0.81
0.40	1.00	0.90	0.71	0.80	0.70	1.00	0.80	0.75	0.94
0.45	0.82	1.00	0.88	0.87	1.00	0.86	0.90	0.88	1.00
0.50	1.00	0.94	0.90	0.93	1.00	0.90	1.00	0.96	1.00

Таблица 1.34.

Частоты обнаружения и их стандартные отклонения для тест-системы No 5

$c,$ мг·л ⁻¹	$\overline{P_i^{\text{ЭМП}}}$	$s(P_i)$	
		расчет по формуле (1.28)	расчет по формуле (1.23)
0.05	0.146	0.037	0.029
0.10	0.191	0.024	0.032
0.15	0.301	0.029	0.037
0.20	0.408	0.051	0.040
0.25	0.559	0.054	0.040
0.30	0.655	0.043	0.039
0.35	0.779	0.040	0.034
0.40	0.846	0.040	0.029
0.45	0.908	0.023	0.024
0.50	0.959	0.014	0.016

Таблица 1.35.

**Значения параметров аппроксимирующих функций,
статистики χ^2 и интервалы ненадежности для тест-системы No 5**

Функция распределения	Расчет $s(P_i)$ по формуле (1.28)			Расчет $s(P_i)$ по формуле (1.23)		
	Параметры	$\chi^2_{\text{эксп}}$	Δc , мг/л	Параметры	$\chi^2_{\text{эксп}}$	Δc , мг/л
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
Экстремаль- ного значения первого рода (Вейбулла)	$a = 7.0 \cdot 10^{-2}$ $b = 0.18$ $k = 1.13$	25.0	$8.3 \cdot 10^{-2} -$ 0.77	$a = -0.16$ $b = 0.44$ $k = 2.77$	1.9	$-6.9 \cdot 10^{-3} -$ 0.61
Лапласа	$a = 0.23$ $b = 0.13$	5.4	$-5.7 \cdot 10^{-2} -$ 0.66	$a = 0.23$ $b = 0.13$	5.3	$-5.5 \cdot 10^{-2} -$ 0.66
Логистичес- кого	$k = 0.23$ $t = 9.3 \cdot 10^{-2}$	1.7	$-4.2 \cdot 10^{-2} -$ 0.66	$k = 0.23$ $t = 9.4 \cdot 10^{-2}$	1.7	$-4.6 \cdot 10^{-2} -$ 0.66

1	2	3	4	5	6	7
Логнормаль- ного	$\bar{c} = 0.20$ $s = 0.64$	34.1	$8.0 \cdot 10^{-2} -$ 0.77	$\bar{c} = 0.21$ $s = 0.63$	40.8	$7.8 \cdot 10^{-2} -$ 0.77
Нормального	$\bar{c} = 0.23$ $s = 0.16$	1.0	$-1.2 \cdot 10^{-2} -$ 0.56	$\bar{c} = 0.23$ $s = 0.16$	1.1	$-1.6 \cdot 10^{-2} -$ 0.56
Пуассона	$\bar{c} = 0.11$	323	$7.1 \cdot 10^{-3} -$ 0.76	$\bar{c} = 0.10$	291	$7.0 \cdot 10^{-3} -$ 0.76
Экспонен- циального	$a = 7.4 \cdot 10^{-2}$ $b = 0.17$	29.2	$8.3 \cdot 10^{-2} -$ 0.86	$a = 7.7 \cdot 10^{-2}$ $b = 0.17$	28.0	$8.6 \cdot 10^{-2} -$ 0.88

5%-ные точки распределения χ^2 : $\chi_{f=6}^2 = 12.6$, $\chi_{f=7}^2 = 14.1$, $\chi_{f=8}^2 = 15.5$ [141].

При использовании функции распределения Вейбулла оценки параметров показали низкую устойчивость к варьированию способа назначения статистических весов. Функция распределения Пуассона не смогла даже грубо воспроизвести зависимость $P_i^{\text{ЭМП}}$ от c ; для остальных функций существенного влияния способа назначения весов на значения подгоночных параметров не обнаружено.

Обнаружение нечувствительности выбора функций, аппроксимирующих зависимости $P_i^{\text{ЭМП}}$ от c , и метрологических характеристик методик обнаружения к варьированию статистических весов породило вопрос: нельзя ли их назначение кардинально упростить?

Мы испытали еще два способа назначения весов, основанные на примитивных моделях погрешностей P_i . В одной модели принимали, что все $P_i^{\text{ЭМП}}$ имеют одинаковые стандартные отклонения $s(P_i^{\text{ЭМП}}) = 0.02$. Тогда статистические веса, назначаемые как

$$w_i = \frac{1}{s(P_i^{\text{ЭМП}})^2} \quad (1.29)$$

тоже одинаковы.

В другой модели допускали, что одинаковы относительные стандартные отклонения $s_r(P_i^{\text{эмп}}) = 0.05$, а веса назначали как

$$w_i = \frac{1}{[P_i^{\text{эмп}} \cdot s_r(P_i^{\text{эмп}})]^2}. \quad (1.30)$$

Поскольку далеко не для всех систем, перечисленных в табл. 1.6 – 1.32, эмпирические частоты обнаружения аналитов определяли в нескольких сериях испытаний, да и сам этот подход пока не вошел в широкую практику, в дальнейшем приводим результаты расчетов, в которых стандартные отклонения $s(P_i^{\text{эмп}})$ оценивали по формуле (1.23). Если учесть, что при высоких и низких значениях $P_i^{\text{эмп}}$ стандартные отклонения (1.23) могут оказаться заниженными по отношению к стандартному отклонению генеральной совокупности [143], назначение статистических весов с применением указанного подхода может приводить к завышенным значениям статистики $\chi^2_{\text{экс}}$ и излишне пессимистической оценке адекватности аппроксимации $P_i^{\text{эмп}}$.

В табл. 1.36 и 1.37 сопоставлены результаты расчетов. Для каждой из тест-систем находили подгоночные параметры той из функций (логистического или экспоненциального распределений), которая была выбрана для аппроксимации зависимости $P_i^{\text{эмп}}$ от c при оценке стандартных отклонений по формуле (1.23).

Легко видеть, что при назначении весов по формулам (1.23) и (1.29) параметры аппроксимирующих функций и границы интервалов ненадежности очень близки, а значения статистик $\chi^2_{\text{экс}}$ являются величинами одного порядка. А это означает, что простейшая возможная модель, признающая $P_i^{\text{эмп}}$ равноточными, может использоваться, наряду с моделью, учитывающей биномиальное распределение частот обнаружения, при подборе параметров кривых эффективности. Поскольку в первом случае веса одинаковы, расчеты существенно упрощаются, а перечень программных продуктов, которые можно использовать для вычислений, расширяется.

Сопоставление оценок параметров аппроксимирующих функций при разных способах назначения статистических весов

No / парам етры	Параметры функции распределения			Значения $\chi^2_{\text{эксп}}$			$\chi^2_{f, 0.05}$
	Формула для w_i			Формула для w_i			
	(1.23)	(1.29)	(1.30)	(1.23)	(1.29)	(1.30)	
<i>l</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
1							
<i>a</i>	26.0	26.1	23.7	2.9	2.8	5.5	9.5
<i>b</i>	40.6	42.1	50.7				
2							
<i>k</i>	$3.9 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	6.9	1.1	1.7	11.1
<i>t</i>	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$				
3							
<i>k</i>	2.53	2.52	2.51	9.4	15.8	1.8	12.6
<i>t</i>	0.27	0.27	0.23				
4							
<i>a</i>	$1.4 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$7.7 \cdot 10^{-4}$	3.2	2.2	1.1	14.1
<i>b</i>	$1.5 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$				
5							
<i>a</i>	$7.7 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$9.9 \cdot 10^{-3}$	28.0	15.1	14.4	15.5
<i>b</i>	0.17	0.23	0.33				
6							
<i>k</i>	5.38	5.35	5.04	5.1	2.4	19.4	12.6
<i>t</i>	0.74	0.81	0.56				
7							
<i>k</i>	$5.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-2}$	4.7	7.2	4.0	11.1
<i>t</i>	$1.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$9.9 \cdot 10^{-3}$				

Продолж. табл. 1.36.

<i>l</i>	2	3	4	5	6	7	8
8							
<i>k</i>	0.14	0.14	0.14	0.8	1.3	4.6	7.8
<i>t</i>	$2.7 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$2.5 \cdot 10^{-2}$				
9							
<i>k</i>	0.27	0.27	0.29	0.9	1.3	7.4	7.8
<i>t</i>	$5.7 \cdot 10^{-2}$	$5.5 \cdot 10^{-2}$	$7.1 \cdot 10^{-2}$				
10							
<i>k</i>	0.18	0.18	0.18	2.5	1.4	0.6	14.1
<i>t</i>	$2.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$				
11							
<i>k</i>	0.15	0.15	0.15	4.6	1.1	0.9	14.1
<i>t</i>	$2.7 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$	$2.9 \cdot 10^{-2}$				
12							
<i>k</i>	0.33	0.33	0.33	4.1	2.8	1.0	11.1
<i>t</i>	$3.6 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$				
13							
<i>a</i>	0.47	0.47	0.47	12.8	14.9	12.4	$\chi_{5, 0.01}^2 =$ 15.1
<i>b</i>	$3.9 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$				
14							
<i>k</i>	$2.4 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	13.1	9.0	4.8	$\chi_{5, 0.01}^2 =$ 15.1
<i>t</i>	$2.2 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$				
15							
<i>k</i>	$4.1 \cdot 10^{-2}$	$4.1 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	4.4	2.4	5.9	15.5
<i>t</i>	$1.1 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$				
16							
<i>k</i>	$1.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	2.2	1.2	0.7	14.1
<i>t</i>	$5.4 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$5.3 \cdot 10^{-3}$				

Продолж. табл. 1.36.

<i>l</i>	2	3	4	5	6	7	8
17							
<i>a</i>	$1.9 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	13.8	11.5	5.0	9.5
<i>b</i>	$5.8 \cdot 10^{-3}$	$6.4 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$				
18							
<i>k</i>	$1.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	9.0	2.3	23.2	14.1
<i>t</i>	$3.4 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$				
19							$\chi_{5, 0.01}^2 =$
<i>k</i>	0.16	0.15	0.15	12.8	9.8	6.4	15.1
<i>t</i>	$3.4 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$	$2.6 \cdot 10^{-2}$				
20							
<i>k</i>	$4.0 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	1.6	1.1	0.6	11.1
<i>t</i>	$6.4 \cdot 10^{-3}$	$6.6 \cdot 10^{-3}$	$6.9 \cdot 10^{-3}$				
21							$\chi_{6, 0.01}^2 =$
<i>a</i>	$1.9 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	13.8	39.0	63.0	16.8
<i>b</i>	0.10	0.10	0.18				
22							
<i>k</i>	4.63	4.64	4.65	3.1	7.7	4.7	7.8
<i>t</i>	0.11	0.11	0.13				
23							
<i>k</i>	19.2	19.2	19.3	19.2	3.5	1.2	14.1
<i>t</i>	2.6	2.5	2.4				
24							
<i>a</i>	13.6	13.5	13.4	7.0	9.8	55.1	15.1
<i>b</i>	2.68	2.76	4.27				
25							
<i>k</i>	60.4	60.4	60.3	1.9	0.5	0.5	12.6
<i>t</i>	3.2	3.2	3.0				

<i>l</i>	2	3	4	5	6	7	8
26							
<i>k</i>	0.73	0.73	0.73	0.6	0.4	0.1	7.8
<i>t</i>	$9.6 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$				
27							
<i>a</i>	$1.7 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$7.9 \cdot 10^{-3}$	2.6	2.7	4.3	9.5
<i>b</i>	0.31	0.31	0.35				

Таблица 1.37.

**Сопоставление границ интервалов ненадежности при разных способах
назначения статистических весов**

No	Δc , мг/л		
	Формула для w_i		
	(1.23)	(1.29)	(1.30)
<i>l</i>	2	3	4
1	23.9 – 213	23.9 – 220	23.7 – 257
2	$3.5 \cdot 10^{-3} - 4.6 \cdot 10^{-3}$	$3.6 \cdot 10^{-3} - 4.6 \cdot 10^{-3}$	$3.5 \cdot 10^{-3} - 4.6 \cdot 10^{-3}$
3	1.75 – 3.77	1.70 – 3.75	1.83 – 3.57
4	$2.2 \cdot 10^{-3} - 7.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-3} - 7.5 \cdot 10^{-2}$	$1.6 \cdot 10^{-3} - 7.9 \cdot 10^{-2}$
5	$8.6 \cdot 10^{-2} - 0.88$	$5.1 \cdot 10^{-2} - 1.10$	$2.6 \cdot 10^{-2} - 1.48$
6	3.21 – 8.78	2.94 – 8.99	3.38 – 7.62
7	$2.1 \cdot 10^{-2} - 9.6 \cdot 10^{-2}$	$1.9 \cdot 10^{-2} - 0.10$	$2.2 \cdot 10^{-2} - 9.7 \cdot 10^{-2}$
8	$6.1 \cdot 10^{-2} - 0.26$	$5.8 \cdot 10^{-2} - 0.27$	$6.2 \cdot 10^{-2} - 0.26$
9	0.10 – 0.53	0.11 – 0.53	$8.1 \cdot 10^{-2} - 0.62$
10	0.12 – 0.27	0.12 – 0.27	0.12 – 0.27
11	$7.0 \cdot 10^{-2} - 0.27$	$6.2 \cdot 10^{-2} - 0.29$	$6.5 \cdot 10^{-2} - 0.28$
12	0.22 – 0.50	0.22 – 0.51	0.24 – 0.46

<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
13	0.47 – 0.65	0.47 – 0.65	0.47 – 0.66
14	$1.7 \cdot 10^{-2} - 3.4 \cdot 10^{-2}$	$1.5 \cdot 10^{-2} - 3.5 \cdot 10^{-2}$	$1.6 \cdot 10^{-2} - 3.6 \cdot 10^{-2}$
15	$1.1 \cdot 10^{-2} - 8.6 \cdot 10^{-2}$	$8.6 \cdot 10^{-3} - 9.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-2} - 8.3 \cdot 10^{-2}$
16	$3.4 \cdot 10^{-3} - 4.3 \cdot 10^{-2}$	$3.1 \cdot 10^{-3} - 4.4 \cdot 10^{-2}$	$3.4 \cdot 10^{-3} - 4.3 \cdot 10^{-2}$
17	$2.2 \cdot 10^{-3} - 2.9 \cdot 10^{-2}$	$2.1 \cdot 10^{-3} - 3.1 \cdot 10^{-2}$	$1.6 \cdot 10^{-3} - 3.7 \cdot 10^{-2}$
18	$4.2 \cdot 10^{-3} - 2.7 \cdot 10^{-2}$	$3.0 \cdot 10^{-3} - 2.9 \cdot 10^{-2}$	$4.3 \cdot 10^{-3} - 2.4 \cdot 10^{-2}$
19	$6.0 \cdot 10^{-2} - 0.32$	$4.7 \cdot 10^{-2} - 0.31$	$7.3 \cdot 10^{-2} - 0.27$
20	$2.1 \cdot 10^{-2} - 6.9 \cdot 10^{-2}$	$2.1 \cdot 10^{-2} - 7.0 \cdot 10^{-2}$	$2.0 \cdot 10^{-2} - 7.2 \cdot 10^{-2}$
21	$2.4 \cdot 10^{-2} - 0.48$	$2.4 \cdot 10^{-2} - 0.48$	$2.0 \cdot 10^{-2} - 0.84$
22	4.31 – 5.14	4.32 – 5.15	4.27 – 5.25
23	11.5 – 31.1	11.8 – 30.7	12.2 – 30.3
24	13.7 – 25.9	13.6 – 26.2	13.6 – 33.1
25	51.0 – 75.0	51.0 – 75.0	51.5 – 74.1
26	0.44 – 1.19	0.45 – 1.17	0.45 – 1.17
27	$3.3 \cdot 10^{-2} - 1.44$	$3.3 \cdot 10^{-2} - 1.44$	$2.6 \cdot 10^{-2} - 1.62$

При назначении весов по формуле (1.30) значения подгоночных параметров аппроксимирующих функций для ряда тест-систем меняются на 30-200% (особенно часто в случае функции экспоненциального распределения), значительным является и различие между оценками пределов обнаружения. Следует полагать, что модель, предполагающая одинаковые относительные погрешности $P_i^{\text{эмп}}$ далека от реальности.

Функция распределения Пуассона с одним подгоночным параметром оказалась в большинстве случаев неспособной адекватно описать зависимости $P_i^{\text{эмп}}(c)$ и была исключена из набора функций, перспективных для построения кривых эффективности.

Наиболее часто наилучшее воспроизведение $P_i^{\text{эмп}}$ достигалось при выборе в качестве аппроксимирующей функции распределения Вейбулла. В то же время, для ряда систем подгоночные параметры не удавалось определить с приемлемой точностью (см. табл. 1.38) из-за их высокой коррелированности, что является признаком избыточности модели. Вследствие этого функция распределения Вейбулла также была исключена из перечня.

Таблица 1.38.

Значения статистики χ^2 , ширина интервала ненадежности и параметры аппроксимирующих функций для тест-системы No 20 из табл. 1.5

Функция распределения	Вектор параметров	Ковариационная матрица	$\chi^2_{\text{эксп}}$
Экстремального значения первого рода (Вейбулла)	$\begin{pmatrix} a = 3.0 \cdot 10^{-3} \\ b = 4.7 \cdot 10^{-2} \\ k = 4.5 \end{pmatrix}$	$\begin{pmatrix} 0.34 & & \\ -0.34 & 0.35 & \\ 1.5 \cdot 10^{-3} & -1.0 \cdot 10^{-2} & 8.8 \cdot 10^{-3} \end{pmatrix}$	0.6
Лапласа	$\begin{pmatrix} a = 4.0 \cdot 10^{-2} \\ b = 9.5 \cdot 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 9.7 \cdot 10^{-6} & \\ -2.9 \cdot 10^{-6} & 5.8 \cdot 10^{-6} \end{pmatrix}$	3.8
Логистического	$\begin{pmatrix} k = 4.0 \cdot 10^{-2} \\ t = 6.4 \cdot 10^{-3} \end{pmatrix}$	$\begin{pmatrix} 1.4 \cdot 10^{-3} & \\ -4.3 \cdot 10^{-4} & 1.4 \cdot 10^{-4} \end{pmatrix}$	1.6
Логнормального	$\begin{pmatrix} \bar{c} = 4.4 \cdot 10^{-2} \\ s = 0.25 \end{pmatrix}$	$\begin{pmatrix} 1.6 \cdot 10^{-4} & \\ -9.0 \cdot 10^{-5} & 2.6 \cdot 10^{-4} \end{pmatrix}$	3.8
Нормального	$\begin{pmatrix} \bar{c} = 4.0 \cdot 10^{-2} \\ s = 1.1 \cdot 10^{-2} \end{pmatrix}$	$\begin{pmatrix} 7.6 \cdot 10^{-4} & \\ -3.6 \cdot 10^{-4} & 1.4 \cdot 10^{-3} \end{pmatrix}$	1.0
Пуассона	$\bar{c} = 4.1 \cdot 10^{-2}$	$1.0 \cdot 10^{-3}$	40
Экспоненциального	$\begin{pmatrix} a = 3.0 \cdot 10^{-2} \\ b = 1.2 \cdot 10^{-2} \end{pmatrix}$	$\begin{pmatrix} 1.3 \cdot 10^{-2} & \\ -1.2 \cdot 10^{-2} & 2.1 \cdot 10^{-2} \end{pmatrix}$	16

5%-ные точки распределения χ^2 : $\chi^2_{f=4} = 9.5$, $\chi^2_{f=5} = 11.1$, $\chi^2_{f=6} = 12.6$ [141].

Для большинства изученных систем приемлемое качество описания зависимостей $P_i^{\text{эмп}}(c)$ достигнуто с использованием, помимо функции распределения Вейбулла, функций логистического и нормального распределений. Поскольку при использовании функций логистического и нормального распределений получаем практически совпадающие кривые эффективности, но расчеты с использованием функции нормального распределения более трудоемки, при построении кривых эффективности достаточно испытывать лишь функцию логистического распределения, а функцию нормального распределения можно исключить из перечня функций, рекомендуемых для аппроксимации зависимостей $P_i^{\text{эмп}}(c)$.

К полезным заключениям привел анализ находимых интервалов ненадежности. Показателен пример обработки данных о частотах обнаружения ионов Fe^{2+} с помощью реагентной индикаторной бумаги (тест-система No 4, табл. 1.39). Для большинства испытанных функций распределения значения $c_{5\%}$ отрицательны, хотя при этом и достигалось адекватное воспроизведение $P_i^{\text{эмп}}$. Такие ситуации возникали, преимущественно, в тех случаях, когда эмпирические частоты $P_i^{\text{эмп}}$ были сосредоточены вне «плеч» кривых эффективности (рис. 1.6).

Имеющие физический смысл оценки $c_{5\%}$ и адекватное воспроизведение $P_i^{\text{эмп}}$ получены при использовании функций логнормального и экспоненциального распределений. В первом случае это обеспечено тем, что область определения функции логнормального распределения – положительные c , а во втором – тем обстоятельством, что функция экспоненциального распределения не имеет точек перегиба. Мы рекомендуем последнюю для включения в перечень, поскольку в ситуациях, аналогичных встретившейся при обработке данных для тест-системы No 4, она полезна для построения кривой эффективности.

Таким образом, при построении кривых эффективности целесообразно испытывать всего две функции – логистического и экспоненциального распределений, а для определения метрологических характеристик методик использовать ту из них, которая лучше описывает эмпирическую

зависимость $P_i^{\text{эмп}}$ и приводит к значению нижней границы интервала ненадежности, имеющему физический смысл.

Таблица 1.39.

**Значения статистики χ^2 и интервалы ненадежности
для тест-системы No 4 из табл. 1.5**

Функция распределения	$\chi_{\text{эксп}}^2$	Δc , мг/л
Экстремального значения первого рода (Вейбулла)	1.6	$-2.7 \cdot 10^{-3} - 5.1 \cdot 10^{-2}$
Лапласа	2.9	$-1.5 \cdot 10^{-3} - 5.6 \cdot 10^{-2}$
Логистического	2.0	$-1.1 \cdot 10^{-3} - 5.2 \cdot 10^{-2}$
Логнормального	5.0	$2.7 \cdot 10^{-3} - 8.6 \cdot 10^{-2}$
Нормального	2.0	$-7.9 \cdot 10^{-3} - 4.2 \cdot 10^{-2}$
Пуассона	23	$1.7 \cdot 10^{-3} - 3.0 \cdot 10^{-2}$
Экспоненциального	3.2	$2.2 \cdot 10^{-3} - 7.1 \cdot 10^{-2}$

5%-ные точки распределения χ^2 : $\chi_{f=6}^2 = 12.6$, $\chi_{f=7}^2 = 14.1$, $\chi_{f=8}^2 = 15.5$ [141].

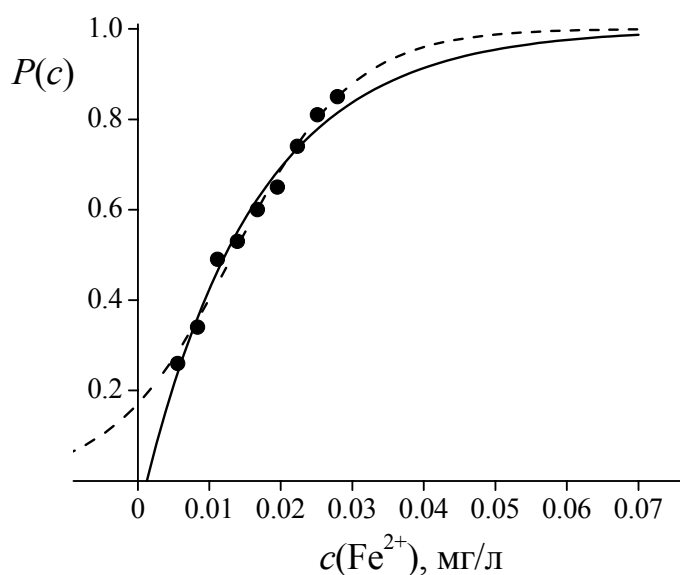


Рис. 1.6. Экспериментальные данные (•) для системы No 4 и их описание кривыми логистического (пунктир) и экспоненциального (сплошная линия) распределений.

Заметим, что при аппроксимации S -образных зависимостей «доза-эффект» также пришли к эмпирическому обобщению о целесообразности использования функции логистического распределения [144].

В табл. 1.40 приведены результаты определения границ интервалов ненадежности для изученных тест-систем (при подборе кривых эффективности использовались функции логистического и экспоненциального распределений).

С использованием критерия χ^2 сделан вывод об адекватном воспроизведении зависимостей $P_i^{\text{эмп}}(c)$ для 24 тест-систем из 27 ($\chi_{\text{эксп}}^2 < \chi_{f, 0.05}^2$ или, в худшем случае, $\chi_{\text{эксп}}^2 < \chi_{f, 0.01}^2$); остальные критерии указывают на адекватность описания для всех 27 систем.

Таблица 1.40.

Выбранные аппроксимирующие функции, значения статистик χ^2 и интервалы ненадежности для тест-систем из табл. 1.5

№	Функция распределения	$\chi_{\text{эксп}}^2 / f / \chi_{f, 0.05}^2$	Δc , мг/л
1	2	3	4
1	экспоненциального	2.9 / 4 / 9.5	23.9 – 213
2	логистического	6.9 / 5 / 11.1	$3.5 \cdot 10^{-3} - 4.6 \cdot 10^{-3}$
3	логистического	9.4 / 6 / 12.6	1.75 – 3.77
4	экспоненциального	3.2 / 7 / 14.1	$2.2 \cdot 10^{-3} - 7.1 \cdot 10^{-2}$
5	экспоненциального	23.9 / 8 / 15.5	$8.6 \cdot 10^{-2} - 0.88$
6	логистического	5.1 / 6 / 12.6	3.21 – 8.78
7	логистического	4.7 / 5 / 11.1	$2.1 \cdot 10^{-2} - 9.6 \cdot 10^{-2}$
8	логистического	0.8 / 3 / 7.8	$6.1 \cdot 10^{-2} - 0.26$
9	логистического	0.9 / 3 / 7.8	0.10 – 0.53

Продолж. табл. 1.40.

1	2	3	4
10	логистического	2.5 / 7 / 14.1	0.12 – 0.27
11	логистического	4.6 / 7 / 14.1	$7.0 \cdot 10^{-2}$ – 0.27
12	логистического	4.1 / 5 / 11.1	0.22 – 0.50
13	экспоненциального	12.8 / 5 / 11.1*	0.47 – 0.65
14	логистического	13.1 / 5 / 11.1*	$1.7 \cdot 10^{-2}$ – $3.4 \cdot 10^{-2}$
15	логистического	4.4 / 8 / 15.5	$1.1 \cdot 10^{-2}$ – $8.6 \cdot 10^{-2}$
16	логистического	2.2 / 7 / 14.1	$3.4 \cdot 10^{-3}$ – $4.3 \cdot 10^{-2}$
17	экспоненциального	13.8 / 4 / 9.5	$2.2 \cdot 10^{-3}$ – $2.9 \cdot 10^{-2}$
18	логистического	9.0 / 7 / 14.1	$4.2 \cdot 10^{-3}$ – $2.7 \cdot 10^{-2}$
19	логистического	12.8 / 5 / 11.1*	$6.0 \cdot 10^{-2}$ – 0.32
20	логистического	1.6 / 5 / 11.1	$2.1 \cdot 10^{-2}$ – $6.9 \cdot 10^{-2}$
21	экспоненциального	13.8 / 6 / 12.6**	$2.4 \cdot 10^{-2}$ – 0.48
22	логистического	3.1 / 3 / 7.8	4.31 – 5.14
23	логистического	19.2 / 7 / 14.1	11.5 – 31.1
24	экспоненциального	7.0 / 8 / 15.5	13.7 – 25.9
25	логистического	1.9 / 6 / 12.6	51.0 – 75.0
26	логистического	0.6 / 3 / 7.8	0.44 – 1.19
27	экспоненциального	2.6 / 4 / 9.5	$3.3 \cdot 10^{-2}$ – 1.44

* Адекватное (по критерию χ^2) описание зависимости $P_i^{\text{эмп}}(c)$ не достигнуто ни для одной из испытанных функций распределения.

** $\chi_{5, 0.01}^2 = 15.1$.

*** $\chi_{6, 0.01}^2 = 16.8$.

1.4. О возможности подбора кривых эффективности на основе системы функций плотности Пирсона

Легко представить ситуацию, когда функцию, способную адекватно описать зависимость $P_i^{\text{ЭМП}}(c)$, не включили в априорно сформированный набор аппроксимирующих функций. Привлекательными представляются подходы, свободные от этого недостатка.

Большинство непрерывных распределений можно достаточно полно охарактеризовать первыми четырьмя моментами, если они конечны. В методе моментов гистограмму наблюдений аппроксимируют кривой распределения, имеющего те же среднее (первый начальный момент), дисперсию, третий и четвертый центральные моменты), что и наблюдаемое распределение. Один из подходов, реализующих метод моментов, был предложен в 1894 г. Карлом Пирсоном; его подробное описание можно найти в старых руководствах по статистике (см., например, [145]), но и сегодня метод отнюдь не забыт. Кривые Пирсона различных типов используются для описания особенностей кристаллических решеток в кристаллографии, аппроксимации кривых распределения полимеров по молекулярным массам полимеров, полос поглощения в спектроскопии, в демографии и для решения других прикладных статистических задач [146-152].

Суть подхода состоит в том, что по результатам наблюдений находят выборочные оценки четырех моментов распределения, определяют тип подходящей кривой Пирсона и по оценкам моментов вычисляют неизвестные параметры искомой кривой.

Решение задачи основано на исследовании дифференциального уравнения Пирсона

$$\frac{dy'}{y'} = \frac{-x'dx'}{a_0x'^2 + a_1x' + a_2}. \quad (1.31)$$

При $x' = 0$ дифференциал dy' обращается в нуль, и, следовательно, экстремум кривой (мода) соответствует началу координат. При произвольном

смещении вводится преобразование $x' = x - m$, $y' = y$, где m – мода новой системы координат, а уравнение (1.31) принимает вид

$$\frac{dy}{y} = \frac{-(x-m)dx}{b_0x^2 + b_1x + b_2}, \quad (1.32)$$

или, в более удобной форме,

$$\frac{1}{y} \frac{dy}{dx} = -\frac{x-m}{b_0x^2 + b_1x + b_2}, \quad (1.33)$$

где $b_0 = a_0, b_1 = a_1 - 2a_0m, b_2 = a_2 - a_1m + a_0m^2$. Задача состоит в том, чтобы определить значения коэффициентов b_0, b_1 и b_2 через значения моментов кривой для выбранной системы координат.

Обозначив абсциссы крайних точек кривой распределения a и b и полагая в этих точках равенство ординат нулю, начальные моменты k -порядка вычисляют как

$$\mu'_k = \frac{\int_a^b x^k y dx}{\int_a^b y dx}, \quad (1.34)$$

Обозначив знаменатель выражения (1.33) $f(x)$, уравнение преобразуют (1.33) к виду

$$f(x)dy + (x-m)ydx = 0 \quad (1.35)$$

Обе части выражения (1.35) умножают на x^k и интегрируют по x в пределах от a до b . Интегрирование приводит к системе из четырех уравнений с четырьмя неизвестными (к коэффициентам b_0, b_1 и b_2 добавляется m), причем коэффициентами этих уравнений являются величины начальных моментов.

Вычисление неизвестных параметров упрощается, если начало координат поместить в центр. Тогда $\mu'_1 = 0$, а остальные начальные моменты совпадают с центральными.

$$\text{Вводя обозначения } \beta_1 = \frac{\mu_3^2}{\mu_2^3}, \beta_2 = \frac{\mu_4}{\mu_2^2}, r = \frac{6(\beta_2 - \beta_1 - 1)}{3\beta_1 - 2\beta_2 + 6},$$

получают окончательно:

$$b_0 = -\frac{1}{r-2}, \quad b_1 = \frac{(r+2)\mu_3}{2(r-2)\mu_2}, \quad (1.36)$$

$$b_2 = \frac{(r+1)\mu_2}{r-2}, \quad M = -b_1 = -\frac{(r+2)\mu_3}{2(r-2)\mu_2}. \quad (1.37)$$

Вычисленные значения b_0 , b_1 , b_2 и m вносят в уравнение (1.33), интегрированием которого находят уравнение, задающее функцию распределения. При этом, в зависимости от значений корней уравнения (по сути, от значения дискриминанта D многочлена $f(x)$), возникают различные функции плотности распределения. Дискриминант D имеет вид:

$$D = b_1^2 - 4b_0b_2 = b_1^2 \left(1 - \frac{4b_0b_2}{b_1^2} \right) = b_1^2 \left(1 - \frac{1}{k} \right), \quad (1.38)$$

где $\kappa = \frac{b_1^2}{4b_0b_2}$ – критерий Пирсона. Используя систему уравнений (1.36), (1.37),

выразим критерий Пирсона через значения моментов:

$$\kappa = -\frac{\beta_1(r+2)^2}{16(r+1)}. \quad (1.39)$$

Различные значения κ дают следующие выводы о корнях уравнения:

- если $\kappa < 0$, то $0 < b_1^2 < D$, уравнение (1.31) имеет вещественные корни различных знаков;
- если $0 < \kappa < 1$, то $D < 0$, уравнение (1.31) имеет комплексные корни;
- если $1 < \kappa < \infty$, то $0 < D < b_1^2$, уравнение (1.31) имеет вещественные корни одного знака.

Соответственно этим случаям различают главные типы кривых Пирсона – типы I, IV и VI. Кроме того, κ может принимать значения 0 , 1 , $\pm\infty$, что приводит к дополнительным (переходным) типам кривых. Вводя некоторые дополнительные условия, число переходных типов можно увеличить. Главные типы кривых Пирсона охарактеризованы в табл. 1.41.

Типы кривых Пирсона и их характеристики

Значение κ	Тип кривой	Свойства кривой
$\kappa < 0$	I (частный случай – бета-распределение I рода)	Область определения ограничена
$\kappa = 0, \beta_1 = 0, \beta_2 \neq 0$	II (частный случай – равномерное распределение)	Область определения ограничена, кривая симметрична
$\kappa = 0, \beta_1 = 0, \beta_2 = 3$	VII (частный случай – распределение Стьюдента)	Область определения неограничена, кривая симметрична
$0 < \kappa < 1$	IV	Область определения неограничена
$\kappa = 1$	V	Область определения неограничена
$1 < \kappa < \infty$	VI (частный случай – F -распределение)	Область определения неограничена
$\kappa = \infty$	III (частный случай – распределение χ^2)	Область определения неограничена

Для подбора кривой Пирсона, описывающей результаты бинарного тестирования, мы использовали алгоритм, включающий следующие шаги.

i. Разбить интервал концентраций аналита на $M-1$ подынтервалов и вычислить набор

$$x_i = (c_{i+1} - c_i) / 2 \quad (1.40)$$

ii. Вычислить набор $p_i^* = P_{i+1}^{\text{ЭМП}} - P_i^{\text{ЭМП}}$, нормировочный множитель

$$l = \frac{1}{\sum_i p_i^*} \text{ и набор}$$

$$p_i = lp_i^* \quad (1.41)$$

Набор $(x_i; p_i)$ соответствует дифференциальному виду кривой эффективности.

iii. Вычислить выборочные оценки начальных

$$\mu'_k = \frac{1}{M-1} \sum_1^{M-1} x_i^k p_i \quad (1.42)$$

и центральных моментов

$$\mu_k = \frac{1}{M-1} \sum_1^{M-1} p_i \cdot (x_i - \mu'_1)^k. \quad (1.43)$$

iv. Вычислить β_1, β_2 и r .

v. Вычислить значение критерия Пирсона k по формуле (1.39).

Расчеты, проведенные для 27 методик обнаружения с бинарным откликом показали, что на практике встречается лишь один тип кривых Пирсона – кривые типа I. Они соответствуют следующему случаю: корни многочлена $f(x)$ (обозначим их d_1 и d_2) действительны, конечны и отличаются друг от друга. Левая часть уравнения (1.33) является производной функции $\ln y$ по x , следовательно

$$\frac{d}{dx} \ln y = \frac{m_1}{d_1 + x} + \frac{m_2}{d_2 - x}, \quad (1.44)$$

где m_1 и m_2 – постоянные.

Таким образом, справедливо соотношение

$$\ln y = m_1 \ln(d_1 + x) + m_2 \ln(d_2 - x) + \text{const} \quad (1.45)$$

или

$$y = y_0 (d_1 + x)^{m_1} (d_2 - x)^{m_2}, \quad (1.46)$$

где y_0 – нормировочный множитель.

При возвращении к исходной системе координат, получаем:

$$y = y_0 (d_1 - \mu'_1 + x)^{m_1} (d_2 + \mu'_1 - x)^{m_2}. \quad (1.47)$$

Вводя вспомогательные коэффициенты $z_1 = m_1 + 1$, $z_2 = m_2 + 1$, $d = d_1 + d_2$,

$d_1 = z_1 \frac{d}{r}$, $d_2 = z_2 \frac{d}{r}$, вычисляют величины:

$$d^2 = 0.25\mu_2(\beta_1(r+2)^2 + 16(r+1)), \quad (1.48)$$

$$s = \frac{4r^2(r+1)}{\beta_1(r+2)^2 + 16(r+1)}. \quad (1.49)$$

Из условий $z_1 + z_2 = r$, $z_1 \cdot z_2 = s$ следует тождество

$$z^2 - rz + s = 0. \quad (1.50)$$

Алгоритм, таким образом, продолжается следующими шагами.

- vi. Вычислить величины d и s .
- vii. Найти корни уравнения (1.50).
- viii. Найти величины d_1 , d_2 , z_1 и z_2 .
- ix. С использованием итеративной процедуры численного интегрирования найти нормировочный множитель y_0 в уравнении (1.47).
- x. Проинтегрировать полученное уравнение и получить набор оценок вероятности обнаружения компонента в интервале ненадежности $\hat{\mu}_i$.
- xi. Оценить адекватность модели с использованием обычного набора критериев.

В качестве примера приведем результаты применения алгоритма для описания зависимости частоты обнаружения суммы ионов металлов от их концентрации по методике [68] (табл. 1.7). Набор $(x_i; p_i)$ приведен в табл. 1.42.

Были получены такие выборочные оценки центральных моментов: $\mu_1 = 3.94 \cdot 10^{-3}$, $\mu_2 = 2.0 \cdot 10^{-5}$, $\mu_3 = 7.2 \cdot 10^{-7}$, $\mu_4 = 8.4 \cdot 10^{-7}$, значения вспомогательных величин: $\beta_1 = 0.070$, $\beta_2 = 2.19$, $r = 3.69$ и критерия Пирсона: $\kappa = -0.030$. Следовательно, набор данных $(x_i; p_i)$ следует описывать кривой Пирсона типа I.

Уравнение дифференциальной кривой (1.47) имеет вид:

$$\hat{\mu}_i = 1.81 \cdot 10^9 \left\{ \left(-3.7 \cdot 10^{-3} + x_i \right)^{0.53} \cdot \left(4.3 \cdot 10^{-3} - x_i \right)^{1.16} \right\}. \quad (1.51)$$

В табл. 1.42 и на рис. 1.7 $p_i^{\text{ЭМП}}$ сопоставлены с оценками $\hat{\mu}_i$, рассчитанными по уравнению (1.51).

Таблица 1.42

Сопоставление $p_i^{\text{ЭМП}}$ с f_i

№	x_i	$p_i^{\text{ЭМП}}$	f_i
1	$3.75 \cdot 10^{-3}$	0.21	0.19
2	$3.85 \cdot 10^{-3}$	0.20	0.24
3	$3.95 \cdot 10^{-3}$	0.26	0.23
4	$4.05 \cdot 10^{-3}$	0.19	0.19
5	$4.15 \cdot 10^{-3}$	0.11	0.12
6	$4.25 \cdot 10^{-3}$	0.03	0.04

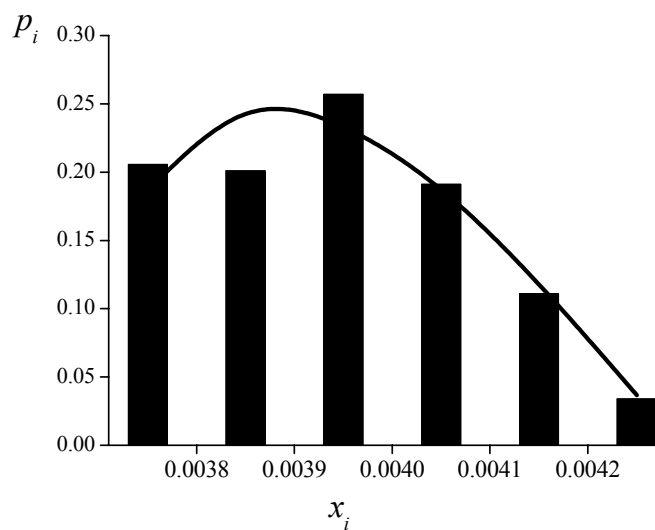


Рис. 1.7. Экспериментальные частоты обнаружения суммы ионов металлов (гистограмма) и их аппроксимация кривой Пирсона типа I.

Численно интегрируя зависимость $f(x)$, построили кривую эффективности (рис. 1.8).

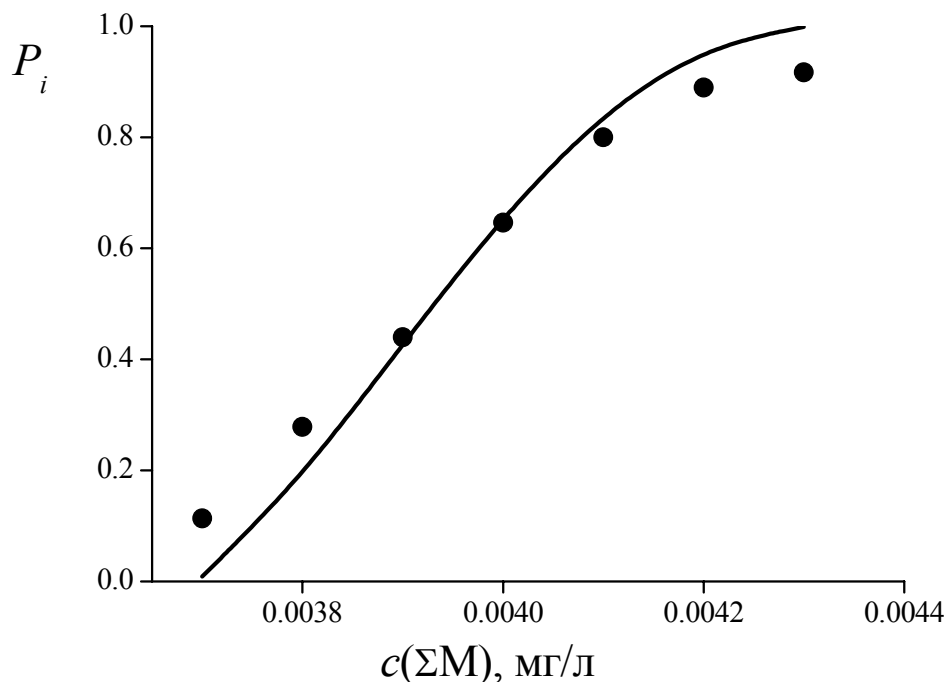


Рис. 1.8. Экспериментальные частоты обнаружения суммы ионов металлов (точки) и кривая эффективности, соответствующая кривой Пирсона типа I.

В табл. 1.43 представлены результаты расчетов для тест-систем из табл. 1.5. Испытанный способ обеспечил адекватную аппроксимацию зависимостей $P_i^{\text{эмп}}(c)$ лишь для двух систем. При этом значения пределов обнаружения в некоторых случаях отличались от ранее найденных (табл. 1.40) на 30, 50 и даже 200%. Очевидно, метод моментов не может быть рекомендован для установления метрологических характеристик методик скрининга. Систему функций Пирсона можно рекомендовать скорее для предварительной, грубой оценки метрологических характеристик методик с бинарным откликом.

По всей вероятности, неудача связана с налагаемым расчетной процедурой условием нормировки дифференциальной кривой эффективности (см. формулу (1.41)). Кроме того, выборочные моменты, найденные всего по нескольким значениям случайных величин, определяются с высокой погрешностью (выборочные моменты далеки от моментов генеральной совокупности).

Результаты использования метода моментов

№	$\chi^2_{\text{эксп}} / f / \chi^2_{f, 0.05}$	Δc , мг/л
1	9.77 / 2 / 9.2	34.5 – 162.4
2	42.79 / 3 / 11.3	$3.7 \cdot 10^{-3} - 4.3 \cdot 10^{-3}$
3	0.49 / 4 / 9.5	2.09 – 3.75
4	85.51 / 5 / 11.1	$1.4 \cdot 10^{-3} - 4.7 \cdot 10^{-2}$
5	37.6 / 6 / 16.8	$9.8 \cdot 10^{-2} - 0.50$
6	9.7 / 4 / 9.5	3.37 – 8.16
7	4.6 / 3 / 11.3	$3.6 \cdot 10^{-2} - 8.8 \cdot 10^{-2}$
8	1.7 / 1 / 6.6	$7.2 \cdot 10^{-2} - 0.25$
9	4.4 / 1 / 6.6	0.16 – 0.50
10	107.9 / 5 / 11.1	0.16 – 0.23
11	77.9 / 5 / 11.1	0.11 – 0.22
12	63.2 / 3 / 11.3	0.30 – 0.41
13	109.3 / 3 / 11.3	0.48 – 0.51
14	41.0 / 3 / 11.3	$2.0 \cdot 10^{-2} - 3.2 \cdot 10^{-2}$
15	53.3 / 6 / 16.8	$1.7 \cdot 10^{-2} - 0.10$
16	60.3 / 5 / 11.1	$1.3 \cdot 10^{-3} - 3.3 \cdot 10^{-2}$
17	17.3 / 2 / 9.2	$3.7 \cdot 10^{-3} - 1.3 \cdot 10^{-2}$
18	15.8 / 5 / 11.1	$6.1 \cdot 10^{-3} - 3.0 \cdot 10^{-2}$
19	99.9 / 3 / 11.3	0.11 – 0.28
20	33.5 / 3 / 11.3	$3.4 \cdot 10^{-2} - 5.6 \cdot 10^{-2}$
21	51.0 / 4 / 9.5	$3.5 \cdot 10^{-2} - 0.16$
22	6.5 / 1 / 6.6	4.55 – 4.80
23	201 / 5 / 11.1	18.7 – 25.5
24	23.5 / 6 / 16.8	14.2 – 17.7
25	60.6 / 4 / 9.5	55.0 – 73.0
26	13.1 / 1 / 6.6	0.66 – 1.20
27	75.0 / 2 / 9.2	$8.1 \cdot 10^{-2} - 0.80$

1.5. Робастный алгоритм расчета подгоночных параметров кривых эффективности

Оценки параметров кривых эффективности, $|\theta\rangle$, находимые при решении задачи нелинейного взвешенного МНК, являются асимптотически состоятельными, несмещенными и эффективными при нормальном распределении P_i [136, 137] или в случае, если плотность вероятности распределения P_i имеет хвосты более короткие, чем у нормального распределения [153]. В случае более длинных хвостов оценки $|\theta\rangle$ теряют свои оптимальные статистические свойства. О длине хвостов судят по коэффициенту эксцесса распределения (γ_2): если хвосты длиннее, чем у нормального распределения, $\gamma_2 > 0$.

Частоты P_i подчиняются биномиальному распределению с коэффициентом эксцесса

$$\gamma_2 = -\frac{6}{N_i} + \frac{1}{N_i P_i (1 - P_i)}. \quad (1.52)$$

Поскольку $\gamma_2 > 0$ при малых и больших P_i , оптимальные свойства МНК-оценок не гарантированы и параметры кривых эффективности целесообразно оценивать с помощью робастных методов, устойчивых к нарушению предпосылок МНК. Если робастные оценки $|\theta\rangle$ и метрологические характеристики методик окажутся близкими к оценкам, полученным на основе использования МНК, то от применения довольно трудоемких процедур робастного оценивания в практике обработки экспериментальных данных можно отказаться.

В работе [154] описан робастный алгоритм оценивания параметров градуировочных прямых на основе теории нечетких множеств [155, 156]. Возможность построения с использованием теории нечетких множеств оценок, обладающих свойствами робастности, была показана и в работах [157, 158].

Выполняя вычисления, точки из наборов $P_i^{\text{ЭМП}}(c)$ представляли как нечеткие числа (использованная процедура описана в [159]). С учетом рекомендаций работы [160] для заданного вида функции принадлежности рассчитывали суммарную функцию принадлежности функции, аппроксимирующей зависимость $P_i^{\text{ЭМП}}(c)$, как

$$\mu_{\text{sum}} = \text{card}(F) = \frac{1}{M} \sum_{i=1}^M \mu_i, \quad (1.53)$$

где $\text{card}(F)$ – мощность множества F функций принадлежности μ_i .

Подгоночные параметры аппроксимирующей функции находили как оценки, соответствующие максимуму суммарной функции принадлежности:

$$|\theta\rangle = \arg \max \left(\frac{1}{M} \sum_{i=1}^M \mu_i \right) = \arg \max(\text{card}(F)). \quad (1.54)$$

Проверили устойчивость оценок параметров функций экспоненциального и логистического распределений. Задавали колоколообразную форму функции принадлежности:

$$\mu_i = 1 - \left[\left(\frac{c_i - \epsilon_i}{\Delta c_i} \right)^2 - \left(\frac{P_i^{\text{ЭМП}} - \mathcal{R}_i}{\Delta P_i} \right)^2 \right], \quad (1.55)$$

где μ_i – функция принадлежности, соответствующая i -й экспериментальной точке $(c_i; P_i^{\text{ЭМП}})$, координаты $(\epsilon_i; \mathcal{R}_i)$ принадлежат точке на аппроксимирующей кривой, а Δc_i и ΔP_i – границы нечеткости значений концентрации и частоты обнаружения аналита, соответственно (см. рис. 1.9). Границы нечеткости задавали, исходя из условий выполнения эксперимента (с учетом погрешности концентраций аналита и стандартных отклонений $P_i^{\text{ЭМП}}$). Принимая относительную погрешность определения концентрации аналита (δ) равной 0.01, значения Δc_i находили как $\delta \cdot c_i$. Границам ΔP_i соответствовали значения стандартных отклонений, рассчитанные по формуле (1.23).

В табл. 1.44 сопоставлены результаты расчета параметров кривых эффективности и границ интервалов ненадежности методом максимизации

суммарной функции принадлежности μ_{sum} и с использованием МНК. **Найденные разными методами оценки параметров оказались близки, а максимальные отличия в значениях границ интервалов ненадежности не превышают 10%, что указывает на устойчивость МНК-оценок.**

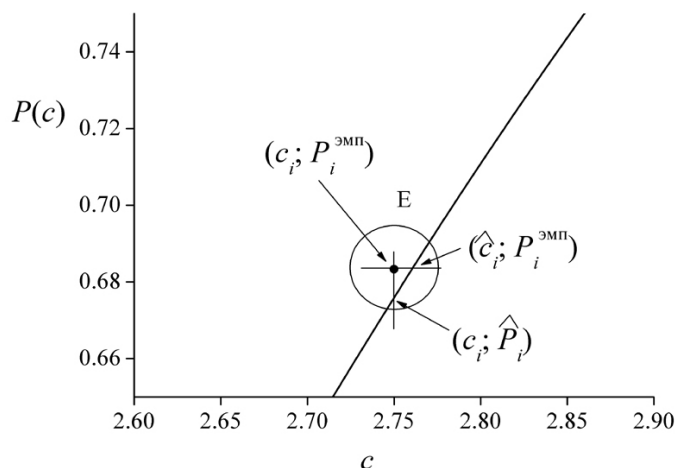


Рис. 1.9. Часть кривой эффективности. Полуосям эллипса E соответствуют границы нечеткости.

Таблица 1.44.

Сопоставление оценок параметров аппроксимирующих функций и интервалов ненадежности для тест-систем из табл. 1.5

№	Параметры выбранной функции распределения		Δc , мг/л	
	МНК-оценки	Робастные оценки	МНК-оценивание	Робастное оценивание
1	2	3	4	5
1	$a = 26.0 (0.3)$ $b = 40.6 (0.4)$	$a = 26.25 (0.03)$ $b = 41.25 (0.02)$	23.9 – 213	28.4 – 216
2	$k = 3.9 \cdot 10^{-3} (1 \cdot 10^{-5})$ $t = 1.3 \cdot 10^{-4} (1 \cdot 10^{-5})$	$k = 3.99 \cdot 10^{-3} (2 \cdot 10^{-4})$ $t = 1.04 \cdot 10^{-4} (7 \cdot 10^{-5})$	$3.5 \cdot 10^{-3} - 4.6 \cdot 10^{-3}$	$3.7 \cdot 10^{-3} - 4.5 \cdot 10^{-3}$
3	$k = 2.53 (2 \cdot 10^{-2})$ $t = 0.27 (3 \cdot 10^{-2})$	$k = 2.53 (2 \cdot 10^{-2})$ $t = 0.32 (1 \cdot 10^{-2})$	1.75 – 3.77	1.58 – 4.02

<i>l</i>	2	3	4	5
4	$a = 1.4 \cdot 10^{-3} (7 \cdot 10^{-4})$ $b = 1.5 \cdot 10^{-2} (1 \cdot 10^{-3})$	$a = 1.4 \cdot 10^{-3} (1 \cdot 10^{-4})$ $b = 1.6 \cdot 10^{-2} (1 \cdot 10^{-3})$	$2.2 \cdot 10^{-3} - 7.1 \cdot 10^{-2}$	$2.2 \cdot 10^{-3} - 7.3 \cdot 10^{-3}$
5	$a = 7.7 \cdot 10^{-2} (5 \cdot 10^{-3})$ $b = 0.17 (1 \cdot 10^{-2})$	$a = 9.0 \cdot 10^{-2} (2 \cdot 10^{-2})$ $b = 0.17 (3 \cdot 10^{-2})$	$8.6 \cdot 10^{-2} - 0.88$	$9.8 \cdot 10^{-2} - 0.87$
6	$k = 5.38 (0.10)$ $t = 0.74 (5 \cdot 10^{-2})$	$k = 5.35 (0.10)$ $t = 0.71 (0.06)$	$3.21 - 8.78$	$3.27 - 8.60$
7	$k = 5.1 \cdot 10^{-2} (1 \cdot 10^{-3})$ $t = 1.0 \cdot 10^{-2} (1 \cdot 10^{-3})$	$k = 5.0 \cdot 10^{-2} (5 \cdot 10^{-3})$ $t = 1.0 \cdot 10^{-2} (3 \cdot 10^{-3})$	$2.1 \cdot 10^{-2} - 9.6 \cdot 10^{-2}$	$2.1 \cdot 10^{-2} - 9.5 \cdot 10^{-2}$
8	$k = 0.14 (2 \cdot 10^{-3})$ $t = 2.7 \cdot 10^{-2} (1 \cdot 10^{-3})$	$k = 0.14 (7 \cdot 10^{-3})$ $t = 2.6 \cdot 10^{-2} (1 \cdot 10^{-3})$	$6.1 \cdot 10^{-2} - 0.26$	$6.5 \cdot 10^{-2} - 0.26$
9	$k = 0.27 (6 \cdot 10^{-3})$ $t = 5.7 \cdot 10^{-2} (4 \cdot 10^{-3})$	$k = 0.27 (1 \cdot 10^{-2})$ $t = 5.2 \cdot 10^{-2} (1 \cdot 10^{-2})$	$0.10 - 0.53$	$0.12 - 0.51$
10	$k = 0.18 (1 \cdot 10^{-3})$ $t = 2.1 \cdot 10^{-2} (1 \cdot 10^{-3})$	$k = 0.18 (1 \cdot 10^{-2})$ $t = 2.0 \cdot 10^{-2} (5 \cdot 10^{-3})$	$0.12 - 0.27$	$0.12 - 0.27$
11	$k = 0.15 (1 \cdot 10^{-3})$ $t = 2.7 \cdot 10^{-2} (1 \cdot 10^{-3})$	$k = 0.15 (1 \cdot 10^{-3})$ $t = 2.8 \cdot 10^{-2} (1 \cdot 10^{-3})$	$7.0 \cdot 10^{-2} - 0.27$	$7.0 \cdot 10^{-2} - 0.28$
12	$k = 0.33 (3 \cdot 10^{-3})$ $t = 3.6 \cdot 10^{-2} (3 \cdot 10^{-3})$	$k = 0.33 (3 \cdot 10^{-3})$ $t = 3.8 \cdot 10^{-2} (3 \cdot 10^{-3})$	$0.22 - 0.50$	$0.22 - 0.51$
13	$a = 0.47 (4 \cdot 10^{-3})$ $b = 3.9 \cdot 10^{-2} (7 \cdot 10^{-3})$	$a = 0.47 (4 \cdot 10^{-3})$ $b = 3.9 \cdot 10^{-2} (7 \cdot 10^{-3})$	$0.47 - 0.65$	$0.44 - 0.62$
14	$k = 2.4 \cdot 10^{-2} (4 \cdot 10^{-4})$ $t = 2.2 \cdot 10^{-3} (2 \cdot 10^{-4})$	$k = 2.4 \cdot 10^{-3} (1 \cdot 10^{-3})$ $t = 2.3 \cdot 10^{-3} (4 \cdot 10^{-4})$	$1.7 \cdot 10^{-2} - 3.4 \cdot 10^{-2}$	$1.7 \cdot 10^{-2} - 3.4 \cdot 10^{-2}$
15	$k = 4.1 \cdot 10^{-2} (1 \cdot 10^{-3})$ $t = 1.1 \cdot 10^{-2} (1 \cdot 10^{-3})$	$k = 4.1 \cdot 10^{-2} (4 \cdot 10^{-3})$ $t = 1.0 \cdot 10^{-2} (3 \cdot 10^{-3})$	$1.1 \cdot 10^{-2} - 8.6 \cdot 10^{-2}$	$1.1 \cdot 10^{-2} - 8.7 \cdot 10^{-2}$
16	$k = 1.9 \cdot 10^{-2} (2 \cdot 10^{-4})$ $t = 5.4 \cdot 10^{-3} (2 \cdot 10^{-4})$	$k = 1.9 \cdot 10^{-2} (2 \cdot 10^{-4})$ $t = 5.6 \cdot 10^{-3} (2 \cdot 10^{-4})$	$3.4 \cdot 10^{-3} - 4.3 \cdot 10^{-2}$	$2.7 \cdot 10^{-3} - 4.5 \cdot 10^{-2}$

<i>l</i>	2	3	4	5
17	$a = 1.9 \cdot 10^{-3} (5 \cdot 10^{-4})$ $b = 5.8 \cdot 10^{-3} (8 \cdot 10^{-4})$	$a = 2.5 \cdot 10^{-3} (5 \cdot 10^{-4})$ $b = 5.8 \cdot 10^{-3} (7 \cdot 10^{-4})$	$2.2 \cdot 10^{-3} - 2.9 \cdot 10^{-2}$	$2.7 \cdot 10^{-3} - 2.9 \cdot 10^{-2}$
18	$k = 1.3 \cdot 10^{-2} (2 \cdot 10^{-4})$ $t = 3.4 \cdot 10^{-3} (2 \cdot 10^{-4})$	$k = 1.3 \cdot 10^{-4} (2 \cdot 10^{-4})$ $t = 3.7 \cdot 10^{-3} (2 \cdot 10^{-4})$	$4.2 \cdot 10^{-3} - 2.7 \cdot 10^{-2}$	$2.3 \cdot 10^{-3} - 3.0 \cdot 10^{-2}$
19	$k = 0.16 (5 \cdot 10^{-3})$ $t = 3.4 \cdot 10^{-2} (4 \cdot 10^{-3})$	$k = 0.16 (5 \cdot 10^{-3})$ $t = 3.3 \cdot 10^{-2} (5 \cdot 10^{-3})$	$6.0 \cdot 10^{-2} - 0.32$	$6.2 \cdot 10^{-2} - 0.31$
20	$k = 4.0 \cdot 10^{-2} (3 \cdot 10^{-4})$ $t = 6.4 \cdot 10^{-3} (3 \cdot 10^{-4})$	$k = 4.0 \cdot 10^{-2} (3 \cdot 10^{-4})$ $t = 6.0 \cdot 10^{-3} (3 \cdot 10^{-4})$	$2.1 \cdot 10^{-2} - 6.9 \cdot 10^{-2}$	$2.3 \cdot 10^{-2} - 6.7 \cdot 10^{-2}$
21	$a = 1.9 \cdot 10^{-2} (1 \cdot 10^{-2})$ $b = 0.10 (2 \cdot 10^{-2})$	$a = 3.0 \cdot 10^{-2} (1 \cdot 10^{-2})$ $b = 7.0 \cdot 10^{-2} (2 \cdot 10^{-2})$	$2.4 \cdot 10^{-2} - 0.48$	$3.4 \cdot 10^{-2} - 0.35$
22	$k = 4.63 (0.01)$ $t = 0.11 (1 \cdot 10^{-2})$	$k = 4.62 (1 \cdot 10^{-2})$ $t = 0.10 (1 \cdot 10^{-2})$	$4.31 - 5.14$	$4.33 - 5.10$
23	$k = 19.2 (0.1)$ $t = 2.6 (0.1)$	$k = 19.3 (0.1)$ $t = 2.5 (0.1)$	$11.5 - 31.1$	$11.9 - 31.0$
24	$a = 13.6 (6 \cdot 10^{-2})$ $b = 2.68 (0.17)$	$a = 13.40 (0.1)$ $b = 2.55 (0.20)$	$13.7 - 25.9$	$13.5 - 25.2$
25	$k = 60.4 (1.0)$ $t = 3.2 (0.5)$	$k = 60.4 (1.0)$ $t = 3.2 (0.5)$	$51.0 - 75.0$	$51.1 - 75.0$
26	$k = 0.73 (5 \cdot 10^{-3})$ $t = 9.6 \cdot 10^{-2} (3 \cdot 10^{-3})$	$k = 0.73 (5 \cdot 10^{-3})$ $t = 0.10 (3 \cdot 10^{-3})$	$0.44 - 1.19$	$0.44 - 1.18$
27	$a = 1.7 \cdot 10^{-2} (1 \cdot 10^{-2})$ $b = 0.31 (5 \cdot 10^{-2})$	$a = 1.2 \cdot 10^{-2} (1 \cdot 10^{-2})$ $b = 0.29 (3 \cdot 10^{-2})$	$3.3 \cdot 10^{-2} - 1.44$	$2.8 \cdot 10^{-2} - 1.36$

1.6. Выводы к разделу 1

1. В качестве метрологических характеристик методик скрининга с бинарным откликом достаточно использовать интервал ненадежности и предел обнаружения – концентрацию аналита, при превышении которой вероятность ошибки I рода (ложного заключения об отсутствии аналита) меньше 1%. «Недостоверность», определяемая как вероятность ошибки I рода, выступает как аналог «неопределенности» результатов количественного анализа.
2. Метрологические характеристики целесообразно находить по кривым эффективности, выражающим концентрационную зависимость вероятности обнаружения аналита в интервале ненадежности. Рекомендовано ограничить вид возможных кривых эффективности двумя функциями – логистического и экспоненциального распределений. Метод моментов Пирсона, свободный от априорного назначения вида кривой эффективности, неприменим из-за большой погрешности эмпирических частот обнаружения аналита в интервале ненадежности.
3. Показано, что оценки параметров кривых эффективности мало зависят от способа назначения статистических весов и при расчетах допустимо использовать приближение, согласно которому частоты обнаружения аналита в интервале ненадежности являются равноточными случайными величинами.
4. Робастные оценки параметров кривых эффективности, полученные с применением аппарата теории нечетких множеств, и соответствующие значения метрологических характеристик близки к оценкам, рассчитанным с использованием нелинейного МНК. Это свидетельствует о возможности при построении кривых эффективности ограничиться более простым вычислительным средством – МНК.

РАЗДЕЛ 2

ИДЕНТИФИКАЦИЯ СОЕДИНЕНИЙ – ПРИМЕНЕНИЕ ХЕМОМЕТРИЧЕСКИХ ПОДХОДОВ

2.1. Принципы компьютерной идентификации соединений

Часто в задачах химического анализа необходимо сделать вывод об идентификации объекта на основании анализа многооткликowego массива данных (результаты спектральных или хроматографических измерений). Методология решения такой задачи близка к подходам, применяемым для формирования выводов об обнаружении / необнаружении аналита в методиках обнаружения с бинарным откликом.

Один из основных подходов к идентификации основан на оценивании сходства характеристик (например, спектров) аналита и эталона. Возможна ситуация, когда условия проведения эксперимента по измерению характеристик эталона и аналита не совпадают.

Отечественная история применения ЭВМ для экспрессного качественного спектрального анализа началась в 70-е годы прошлого века с создания информационно-поисковых систем (ИПС) [161]. Использование средств вычислительной техники в автоматизированных системах регистрации и обработки сигналов, в совокупности с использованием крупных баз данных и ИПС привело к зарождению нового раздела аналитической химии – аналитической химии, основанной на применении компьютеров [162] (computer based analytical chemistry, COBAC). Методы COBAC ориентированы на получение достоверной качественной и количественной характеристики исследуемого материала путем машинной обработки совокупности аналитических сигналов [102].

Достоверность идентификации соединения по массивам его характеристик зависит не только от факта присутствия эталонного массива (например, спектра этого же соединения в определенной базе данных). В работе [102] отмечено, что на результат идентификации аналита по спектральным данным оказывают влияние способ представления исходных данных, алгоритм поиска и способ сопоставления спектра.

Одним из ключевых условий успешной идентификации является запись эталонного спектра и спектра сравнения в тождественных или максимально близких экспериментальных условиях. Однако даже в условиях записи спектров одного вещества на одном и том же оборудовании и с соблюдением одной методики, различия в интенсивности и положении пиков все же возможны. Максимальная воспроизводимость пиков наблюдается для спектроскопии ЯМР, минимальная – для масс-спектрометрии. Для оценки воспроизводимости информации, получаемой из баз данных, используют методы математической статистики.

Можно выделить такие ситуации, типичные при поиске соединения в базах данных:

- эталонный спектр исследуемого соединения присутствует в базе данных, условия его регистрации близки к условиям регистрации исследуемого спектра;
- эталонный спектр исследуемого соединения присутствует в базе данных, но условия его получения существенно отличны от условий регистрации исследуемого спектра;
- эталонный спектр исследуемого соединения имеется в базе данных, но обладает низким качеством (записан с ошибками либо не для индивидуального соединения);
- анализируемая проба представляет собой смесь соединений и в базе данных содержится достоверная информация обо всех составляющих этой смеси;

- анализируемая проба – индивидуальное соединение, но вместо эталонных спектров в базе данных имеются спектры его структурных аналогов (гомологов, изомеров);
- в базе данных отсутствуют как спектры составляющих анализируемой смеси, так и ее структурных аналогов.

Первые три ситуации ведут к однозначной достоверной идентификации соединения; четвертая – к идентификации компонентов пробы; пятая – к установлению особенностей строения соединения; шансы на успешное решение шестой задачи практически отсутствуют [163].

Выделяют такие этапы поиска:

- ввод запроса и обработка информации;
- предварительный отбор эталонных спектров;
- сопоставление спектра пробы с отобранными эталонными в рамках используемого алгоритма поиска;
- формирование ответа на запрос путем вывода ранжированного списка эталонных спектров, наиболее схожих с исследуемым.

Сопоставление спектра пробы с эталоном может проводиться с помощью нескольких алгоритмов поиска – прямого, обратного, или их комбинации.

Спектр пробы при прямом поиске поочередно сопоставляют с эталонными, при этом для расчета степени совпадения для всех эталонов учитывают все сигналы, присутствующие в спектре пробы. Степень совпадения снижается при любом несоответствии признаков, например, при появлении в спектре пробы «лишних» пиков по сравнению со спектром эталона.

В обратном поиске эталонные спектры сопоставляют со спектром пробы. При этом признаки, отсутствующие у проверяемого эталона, не учитывают, а лишние пики не снижают степень совпадения пробы с эталоном (сигнал присутствия проверяемого компонента [164]).

Процесс прямого поиска проходит быстрее и дает более однозначные результаты для чистых веществ. Однако при идентификации недостаточно чистых соединений в спектре пробы проявляются пики примесей. В этом случае результат прямого поиска ухудшается, и вещество может быть идентифицировано с применением алгоритма обратного поиска. На рис. 2.1 проиллюстрирована ситуация сопоставления спектра вещества, содержащего примеси, со спектром из базы данных (эталонным).

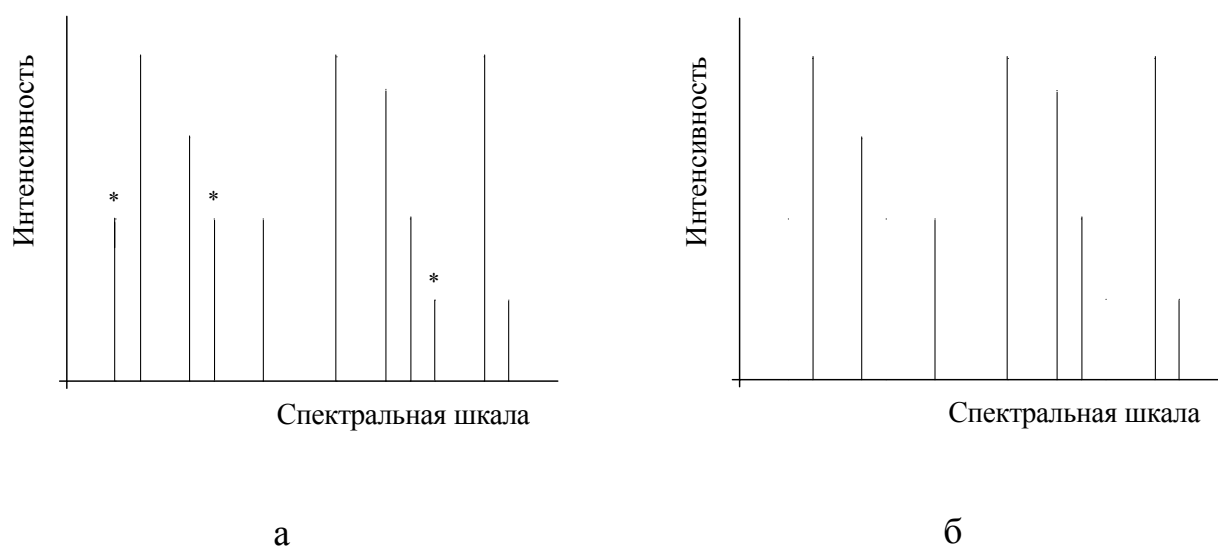


Рис. 2.1. Гипотетический спектр пробы (а)
и эталонный спектр из базы данных (б).

На рис. 2.1 звездочкой отмечены сигналы примеси. Представим, что спектр пробы (рис. 2.1а) сравнивают со спектром эталона (рис. 2.1б), подсчитывая число пиков, совпадающих по положению. Тогда в случае прямого поиска окажется, что спектр пробы не полностью совпадает со спектром эталона (8 пиков совпали, а 3 нет), в то время как для обратного поиска будет сделано заключение о том, что все пики эталона присутствуют в спектре пробы, т.е. спектры совпали полностью. Если предположить, что в эталонных спектрах содержится информация об индивидуальных веществах, а спектр пробы содержит примеси, то алгоритм обратного поиска обладает явным преимуществом.

Отметим также, что в литературе довольно подробно описаны как сами алгоритмы прямого [165], обратного [166] и комбинированного [167] поиска, так и достоинства и недостатки каждого из них [168, 169].

Многие современные методы идентификации веществ по результатам спектроскопических и / или хроматографических измерений предусматривают сравнение многомерных массивов данных, характеризующих «эталон» («образец») и исследуемое вещество. Как правило, такие массивы представляют собой таблично заданные зависимости отклика (откликов) от предикторов, описывающих условия выполнения измерений, необходимых для идентификации. Для принятия решений: «исследуемое вещество совпадает с эталоном», «исследуемое вещество отличается от эталона», «определенное суждение невозможно», исследуют количественные критерии сходства эталона и исследуемого вещества (критерии предусматривают оценку близости откликов эталона и аналита при одинаковых значениях предикторов).

Следующий этап при принятии решения о совпадении / несовпадении образца и исследуемого вещества – количественная оценка достоверности (другой термин, используемый в литературе – надежность), характеризующей принятое решение (аналог процесса проверки гипотез). На этом этапе необходимо оценить как общую (суммарную) достоверность идентификации, так и ее составляющие – вероятности ошибок I рода (необнаружение присутствующего в пробе аналита) и II рода (ложная идентификация вещества, отсутствующего в пробе). Как и в задаче обнаружения, более значимую ошибку (например, неотожествление аналита с опасным веществом) целесообразно рассматривать как ошибку I рода.

2.2. Априорный подход к идентификации аналитов

Расстоянием между объектами a и b называют [170] величину d_{ab} , удовлетворяющую следующим требованиям:

1. $d_{ab} > 0, d_{aa} = 0$;
2. $d_{ab} = d_{ba}$;
3. $d_{ab} + d_{bc} \geq d_{ac}$.

Мерой близости m_{ab} называют величину, имеющую предел и возрастающую по мере приближения объектов друг к другу. В [171] мере близости приписывают такие характеристики:

1. m_{ab} непрерывна, т.е. малому изменению положения точек a и b соответствует малое изменение меры близости;
2. $m_{ab} = m_{ba}$;
3. $0 \leq m_{ab} \leq 1$.

Переход от расстояний к мерам близости довольно прост:

$$m_{ab} = \frac{1}{1 + d_{ab}} \quad (2.1)$$

Предлагаемые в литературе расстояния рекомендуется выбирать, исходя из поставленной перед исследователем задачи и априорно (либо апостериорно) известных особенностей данных [170]. Например, при коррелированности признаков рекомендуют использовать метрики, вычисляемые с использованием ковариационных или обратных им матриц (расстояние Махаланобиса). Известно, однако, что при коэффициентах корреляции близких к 1, ковариационная матрица приближается к вырожденной форме, что может привести к вычислению расстояний, далеких от их геометрических интерпретаций.

В целом, на наш взгляд, в задачах идентификации соединений достаточно использовать привычное и широко используемое Евклидово расстояние:

$$d_E = \sqrt{\sum_i (a(x_i) - e(x_i))^2}, \quad (2.2)$$

где $a(x_i)$ и $e(x_i)$ – величины откликов измерительной системы для аналита и эталона при аналитической позиции x_i .

Очевидна аналогия между d_E^2 и статистикой $\chi_{\text{эксп}}^2$. Используя для вычисления расстояний формулу (2.2) или соответствующие им меры близости, можно оценить достоверность идентификации, сравнивая полученные экспериментальные критерии с процентными точками распределения соответствующих статистик.

В настоящее время рассматривают суммарную надежность (достоверность) идентификации аналита [102, 172]:

$$R = 1 - \alpha - \beta, \quad (2.3)$$

где α и β – вероятности ошибок I и II рода, соответственно.

Как и при оценивании достоверности обнаружения, в задаче идентификации соединений используют два подхода – статистический и априорный.

Методология статистического расчета α и β во многом схожа с описанными алгоритмами оценки соответствующих характеристик для методик визуального тестового анализа с бинарным откликом (см. раздел 1). Используют N образцов известного состава, из которых N_1 не содержат X , а N_2 – содержат X . С учетом критериев идентификации рассчитывают частоту ложной идентификации X (false positive rate, FPR) и частоту необнаружения X , когда он присутствует (false negative rate, FNR). Далее рассчитывают оценки α и β . Выделяют [173] такие достоинства и недостатки статистических методов оценки составляющих достоверности:

Достоинства	Недостатки
– универсальны, просты;	– трудоемки и длительны;
– хорошо отработаны в клиническом анализе;	– требуют либо наличия множества эталонов известного состава, либо наличия референтной методики
– применимы для достоверного определения пределов обнаружения;	качественного анализа.

– не требуют информации о характере распределения экспериментальных данных и о факторах, ведущих к идентификационным ошибкам.

Априорный путь оценки ошибок I и II рода состоит в следующем. Рассмотрим случай, когда экспериментальные пики (на хроматограмме или в спектре) однозначно соотнесены с характеристиками эталона из базы данных, т.е. имеются сведения о характеристиках целевого анализа (рассматриваемого как кандидат на идентификацию), и о характеристиках его ближайших соседей в пробе, рассматриваемых как источники ложной идентификации (отнесение ближайших соседей к эталонной характеристике означает ложную идентификацию целевого анализа).

Принимают определенную модель, описывающую экспериментальные спектры (хроматограммы). Так, в работе [102] рассматривают следующую модель:

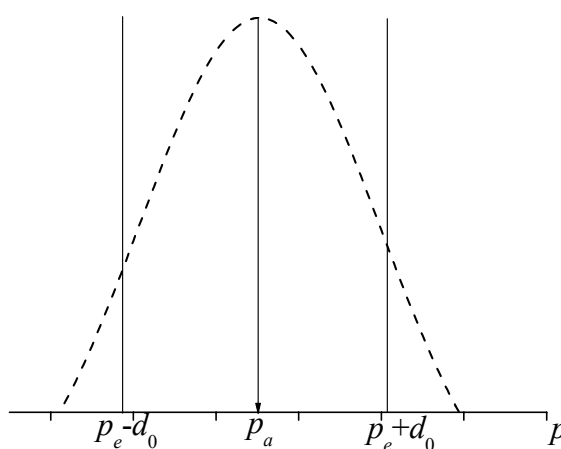
- концентрации всех компонентов пробы (a) выше, чем их пределы обнаружения по данным условиям измерения;
- все пики спектра хорошо разрешены;
- положение пика анализа p_a – нормально распределенная случайная величина с известным (заранее оцененным) стандартным отклонением σ ;
- значения σ для всех пиков близки;
- математическое ожидание положения пика p_a совпадает с константой эталона p_e в базе данных (систематических расхождений между измеренными и табличными значениями характеристик положения пика нет);
- БД включает значения p_e для всех предполагаемых компонентов пробы, причем измерены они в условиях, абсолютно идентичных

экспериментальным; кроме того, БД не содержит совпадающих p_e для разных a .

В рамках модели считают, что ошибки обоих родов – и пропуск аналита, и ложная идентификация, – вызваны случайным сдвигом пиков пробы. Вывод о совпадении пиков делают при соблюдении условия

$$|p_a - p_e| < d_0, \quad (2.4)$$

где критерий d_0 в простейшем случае одинаков для всех пиков и не зависит от состава пробы (рис. 2.2).



б

Рис. 2.2. Определение ошибки I рода в априорном подходе [163].

Оценки вероятностей ошибок I и II родов рассчитывают с применением функции Лапласа (Φ).

Вероятность пропуска сигнала, т.е. случайного выхода пика p_a за пределы $(p_e - d_0; p_e + d_0)$, равна

$$\alpha = 1 - 2\Phi\left(\frac{d_0}{\sigma}\right) \quad (2.5)$$

Проявление пика другого компонента a_i в интервале $(p_e - d_0; p_e + d_0)$ означает ложную идентификацию аналита (рис. 2.3).

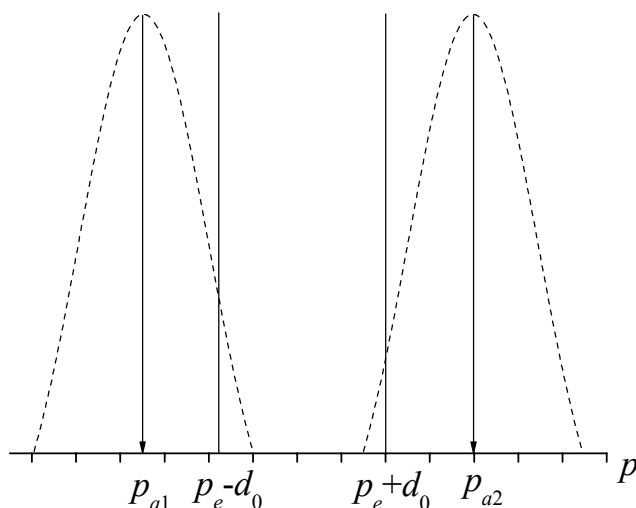


Рис. 2.3. Определение ошибки II рода в априорном подходе [163].

Вероятность ошибки II рода

$$\beta = 1 - \Phi\left(\frac{\Delta_1 - d_0}{\sigma}\right) - \Phi\left(\frac{\Delta_2 - d_0}{\sigma}\right), \quad (2.6)$$

где $\Delta_i = |p_{a_i} - p_e|$ – расстояние от пика в базе данных, который идентифицируется как пик пробы до «ближайших соседей» пика пробы.

Если $\Delta_1 = \Delta_2 = \Delta$, формула (2.6) упрощается:

$$\beta = 1 - 2\Phi\left(\frac{\Delta - d_0}{\sigma}\right) \quad (2.7)$$

Указывают такие достоинства и недостатки априорного подхода к оценке составляющих достоверности идентификации в качественном спектроскопическом (хроматографическом) анализе [173]:

Достоинства

- позволяют быстро судить о достоверности идентификации;
- пригодны для оптимизации методики анализа и подбора

Недостатки

- не универсальны;
- зависят от выбора модели, то есть субъективны;
- позволяют оценить лишь нижний

критериев идентификации; предел достоверности;
 – не требуют стандартных образцов – алгоритмы оценки должны
 состава или референтных методик; создаваться отдельно для каждого
 метода.
 – пригодны для оптимизации
 работы систем компьютерной
 идентификации любого типа.

В работе [163] показано, что на оценки α и β существенно влияет величина d_0 , задаваемая пользователем базы данных, исходя из условий эксперимента и особенностей идентифицируемых соединений (например, в случае идентификации опасных веществ – наркотических или взрывчатых препаратов – критерий d_0 следует увеличивать).

Как при оценке суммарной достоверности идентификации, так и при оценке сходства объектов на основе вычисления расстояний, исследователь принимает решение об идентичности аналита и эталона или о принадлежности аналита определенному классу веществ, руководствуясь критическими (граничными) значениями критериев, которые рассчитывают на основе предположений об известном (включая и величины параметров) законе распределения экспериментальных погрешностей.

Трудности идентификации состоят, главным образом, в невозможности абсолютно точно контролировать условия измерений и в отсутствие априорной информации о распределении экспериментальных погрешностей и форме пиков. Кроме того, критерий «суммарная достоверность идентификации» не является универсальным (например, процедура его применения для обработки данных оптических методов не разработана). Предложить один универсальный критерий сходства объектов невозможно, но особый интерес вызывают робастные критерии, не слишком чувствительные к гипотезам о статистических характеристиках экспериментальных погрешностей.

В настоящей работе для количественной оценки сходства аналита и эталона на основе сравнения положений их пиков привлечена теория нечетких множеств.

2.3. Алгоритм идентификации аналитов с использованием теории нечетких множеств

С данными, характеризующими положения пиков (волновые числа, времена удерживания, индексы Ковача и т.п.), обращаемся как с одномерными нечеткими числами.

Вместо вычисления на основе модельных представлений вероятности неидентификации (пропуска) аналита при идентификации аналитов рассматриваем принадлежность пиков аналита и эталона одному соединению ($\mu_{a,e}$, рис. 2.4). Непринадлежность $\overline{\mu_{a,e}}$ пиков аналита и эталона одному соединению – это дополнение к нечеткому множеству

$$\overline{\mu_{a,e}} = 1 - \mu_{a,e}. \quad (2.8)$$

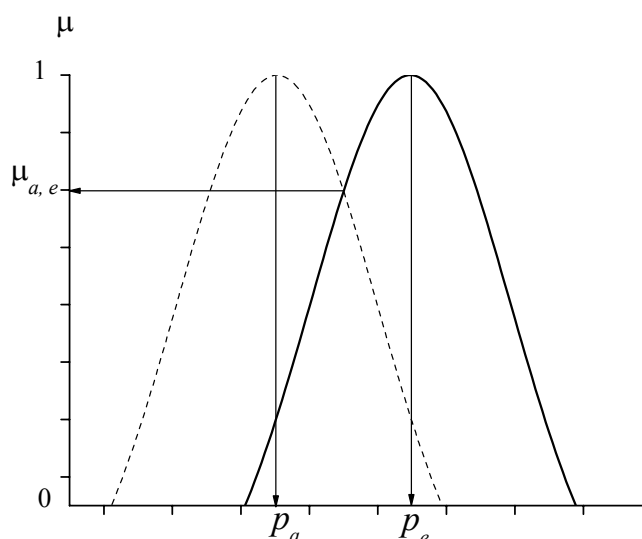


Рис. 2.4. Принадлежность пиков аналита (пунктир) пику эталона (сплошная линия) одному соединению.

Аналогом вероятности ошибки II рода выступает суммарная степень принадлежности пиков, находящихся по соседству с пиком аналита (при его отсутствии в пробе), пику эталона (рис. 2.5).

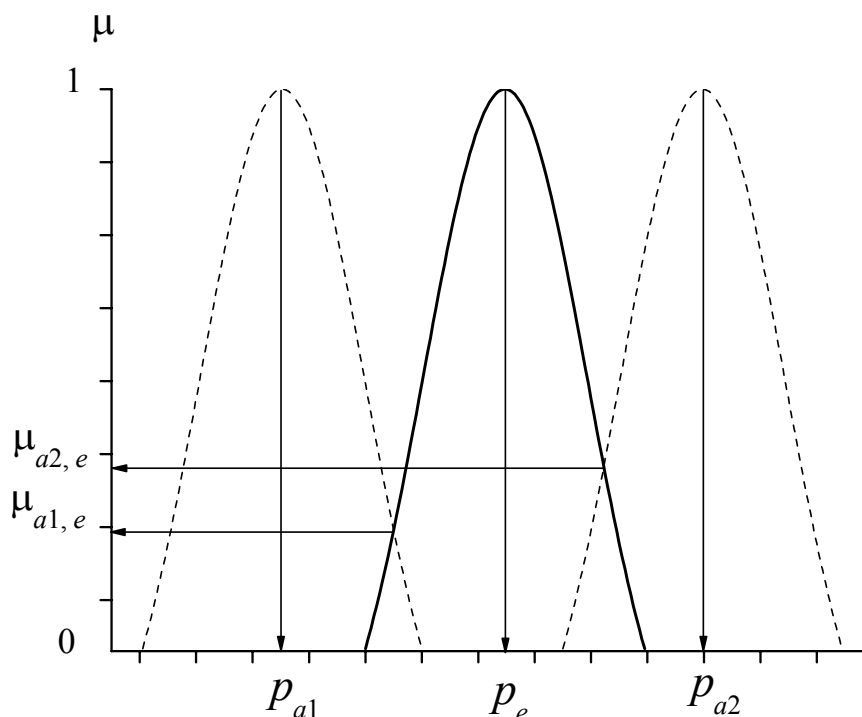


Рис. 2.5. Степень принадлежности пику эталона пиков веществ, находящихся по соседству с пиком аналита.

Предлагается следующий алгоритм идентификации аналитов. Пусть для эталона наблюдается N пиков.

1. Задать допустимый размах данных b_i (обычно значения b_i можно считать одинаковыми для всех измерений).
2. Найти границы нечеткости пиков аналита и эталона $p_{zi} \pm 0.5b_i$ ($z - a$ или e).
3. Фаззифицировать данные p_{zi} , введя функции принадлежности пиков μ_i , $i = 1, 2, \dots, N$. Задавали функции принадлежности двух типов (рис. 2.6) – с линейными ветвями (Симпсона)

$$\mu_i^L(x) = \begin{cases} \frac{x - p_i + 0.5b_i}{0.5b_i} & \text{при } x \leq p_i \\ \frac{x - p_i - 0.5b_i}{-0.5b_i} & \text{при } x > p_i \end{cases} \quad (2.9)$$

и колоколообразную

$$\mu_i^G(x) = \exp\left(-0.5 \left[\frac{x - p_i}{s_i}\right]^2\right), \quad (2.10)$$

где $x \in [p_i - 0.5b_i; p_i + 0.5b_i]$, s_i – параметр, задающий размах функции принадлежности.

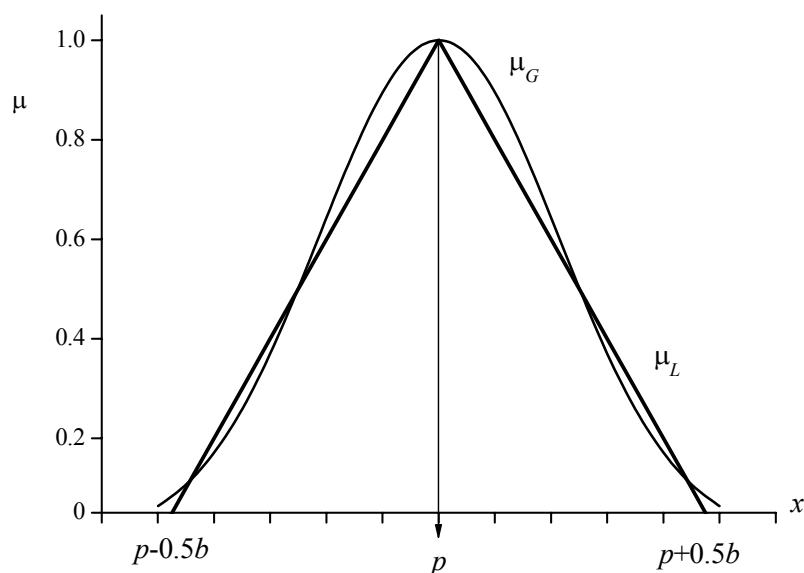


Рис. 2.6. Задаваемые функции принадлежности.

Известно [174], что стандартное отклонение плотности распределения Гаусса (s) связано с полной шириной на половине высоты ($b^{0.5}$) соотношением

$$b^{0.5} = 2s\sqrt{2 \ln 2} = 2.35s. \quad (2.11)$$

С учетом этого s_i задавали как

$$s_i = \frac{b_i}{2} \cdot \frac{1}{2.35} = \frac{b_i}{4.7} \quad (2.12)$$

- Для i -го пика эталона и соответствующего пика аналита вычислить функции их принадлежности одному соединению ($\mu_{a,e,i}$).

5. Найти суммарную мощность [155] множества F функций принадлежности $\mu_{a,e,i}$:

$$\mu_{\text{sum}} = \text{card}(F) = \frac{1}{N} \sum_{i=1}^N \mu_{aei}. \quad (2.13)$$

Мощность μ_{sum} выступает критерием сходства анализа с эталоном.

Теория нечетких множеств использует понятие вероятности (possibility), отличное от статистического (probability). Поэтому подход к оценке критического значения $(\mu_{\text{sum}})_{\text{кр}}$, при превышении которого делается вывод о совпадении анализа с эталоном, отличается от априорного, основанного на статистических моделях [102, 175]. Наиболее правильно устанавливать $(\mu_{\text{sum}})_{\text{кр}}$ на основе обобщения результатов обработки массивов модельных и реальных данных. При недостаточной эмпирической базе для таких обобщений в данной работе использовали значение $(\mu_{\text{sum}})_{\text{кр}} = 0.6$. Основанием для этого был тот факт, что при форме пиков, описываемой функциями (2.9) или (2.10) значение $(\mu_{\text{sum}})_{\text{кр}} = 0.6$ соответствует вероятности ошибки I рода $\sim 5\%$.

В отличие от априорного подхода, в нашем алгоритме необходимо задать лишь тип функции принадлежности и допустимый размах b . При этом важно убедиться, что результаты идентификации малочувствительны к варьированию вида функции принадлежности и оценке b . Информация о законе распределения экспериментальных погрешностей не используется.

2.4. Испытание алгоритма идентификации аналитов

Предлагаемый подход испытали, обрабатывая различные спектральные и хроматографические данные.

Результаты применения разработанного алгоритма для обработки спектральных (многооткликowych) данных сравнивали с результатами использования подхода, основанного на гипотезе о нормальном распределении погрешностей положений пиков. В последнем сходство анализа и эталона оценивали по Евклидовым расстояниям (2.2) и значениям статистик

$$\chi_{\text{эксп}}^2 = \frac{d_E^2}{s^2}. \quad (2.14)$$

где s – стандартное отклонение (2.12). Полученные значения $\chi_{\text{эксп}}^2$ и d_E сравнивали с 5%-ными точками распределения χ^2 для N степеней свободы.

2.4.1. Идентификация по данным газовой хроматографии. В работе [102] приведены результаты анализа бензина методом газовой хроматографии (табл. 2.1).

Таблица 2.1.

Времена удерживания (t) углеводородов методом газовой хроматографии

№	Углеводород	t_a , мин	t_e , мин
1	2,2-диметилбутан	13.597	13.587
2	2-метилпентан	15.697	15.694
3	2,3-диметилгексан	36.393	36.409
4	1,2-диметилциклогексан	40.727	40.719
5	1,2,4-триметилциклогексан	47.312	47.305
6	3-метилоктан	49.637	49.613

Идексы a и e относятся к результатам конкретного анализа и эталонным характеристикам углеводородов, соответственно.

Для допустимых размахов b

$$b = 0.005 \cdot t_e \quad (2.15)$$

и

$$b = 0.015 \cdot t_e \quad (2.16)$$

для каждого из 6 углеводородов рассчитали значения $\mu_{a,e,i}$ (табл. 2.2). Считали, что времена удерживания аналита и эталона совпадают, если $\mu_{a,e,i} > 0.6$. Важно

отметить, что варьирование вида функции принадлежности практически не влияет на значения $\mu_{a,e,i}$ и не сказывается на результатах идентификации.

В работе [102] указывается, что при ширине окна $0.5\% t_e$ «окно совпадений оказалось слишком узким, в ряде случаев наблюдался пропуск сигнала. При $d = 1.5\% t$ и более широких окнах таких случаев не было». Предложенный в настоящей работе алгоритм обеспечивает надежное отождествление аналитов и эталонов даже при допустимом размахе пиков $0.5\% t_e$.

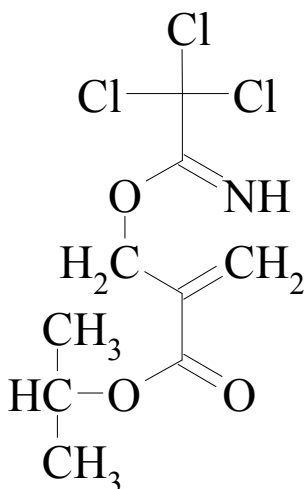
Таблица 2.2.

Результаты расчета $\mu_{a,e,i}$

№	Углеводород	Оценка b	
		(2.15)	(2.16)
1	2,2-диметилбутан	0.85*	0.95
		0.94	0.99
2	2-метилпентан	0.96	0.99
		0.99	1
3	2,3-диметилгексан	0.91	0.97
		0.98	1
4	1,2-диметилциклогексан	0.96	0.99
		1	1
5	1,2,4-триметилциклогексан	0.97	0.99
		1	1
6	3-метилоктан	0.90	0.97
		0.97	1

* В верхней строке приведены значения $\mu_{a,e,i}$ для функции принадлежности типа (2.9), в нижней – (2.10).

2.4.2. Идентификация по данным ИК-спектроскопии. В работе [176] приведены характеристики ИК-спектров индивидуального трет-бутил 2-[[[(2,2,2-трихлорэтанимидоил)окси]метил]акрилата



и его растворов в ряде растворителей (табл. 2.3). Частоты максимумов полос поглощения соединения приведены в табл. 2.4.

Расстояния между аналитами и эталоном значительно меняются при варьировании растворителей, тогда как суммарные функции принадлежности, напротив, меняются сравнительно мало.

В качестве примера в табл. 2.5 приведены результаты обработки данных табл. 2.4 для допустимого размаха пиков $b = 6 \text{ см}^{-1}$. На рис. 2.7 - 2.9 показано изменение d_E , $\chi_{\text{ЭКСП}}^2$ и μ_{sum} в зависимости от условий измерения спектров (растворителя). При оценке совпадения спектров эталона и спектров аналитов по значениям $\chi_{\text{ЭКСП}}^2$ при заданной вероятности ошибки I рода 5% вывод о тождестве аналитов с эталоном сделан в 11 случаях из 16 (рис. 2.7, 2.8). Наш алгоритм обеспечивает правильную идентификацию аналитов во всех 16 случаях.

Важно отметить, что при варьировании b в интервале $1-10 \text{ см}^{-1}$ результаты идентификации с помощью нашего метода сохраняют силу, тогда как выводы о тождестве или отличии аналитов и эталона, сделанные на основе исследования Евклидовых расстояний сильно зависят от предположения о ширине окна.

Растворители, влияние которых на свойства трет-бутил 2-[[[(2,2,2-трихлорэтанимидоил)окси]метил]акрилата исследовалось в работе [175]

№	Растворитель	Акцепторное число	Диэлектрическая проницаемость
1.	Гексан	0	1.9
2.	Циклогексан	0	2
3.	Диэтиловый эфир	3.9	4.3
4.	Дисульфид углерода	5.1	2.6
5.	Тетрахлорметан	8.6	2.2
6.	Тетрагидрофуран	8	7.4
7.	Диоксан	10.8	2.2
8.	1,2-дихлорэтан	16.7	10.4
9.	Ацетонитрил	19.3	35.9
10.	Диметилсульфоксид	19.3	46.5
11.	Дихлорметан	20.4	9.1
12.	Хлороформ	23.1	4.8
13.	2-бутанол	32	15.8
14.	2-пропанол	33.5	18.3
15.	Этанол	37.1	25.1
16.	Метанол	41.3	33.6

Таблица 2.4.

Максимумы полос поглощения, см⁻¹

	V _{NH}	V _{C=O}									V _{C=N}			V _{C=C}
1	3355.1	1738.7	1732.5	1726.7	–	1722.2	1716.4	1707.1	1701.8	–	1673.7	1670.3	1665	1645.6
2	3353.4	1738.7	1732.4	1728	–	1721.8	1715.9	1707.2	1701.7	–	1673.9	1670.2	1664.9	1645.3
3	3353.7	1738.6	1732.2	1728	–	1722.1	1714	1708	1701.9	–	1674	1670.6	1665	1645.4
4	3346.1	1738.4	1731.3	1726.7	–	1720.6	1714.6	1708.5	1702.1	1697.8	1673.3	1667.9	1665.2	1642.9
5	3350.8	1738.6	1732.1	1727.8	1725.1	1719.8	1712.6	1707.9	1702	1697.1	1673.7	1669.6	1664.6	1645.3
6	3348.4	1738	1732.3	1727.5	1723.3	1720.1	–	1707.9	1701.8	1696.8	1674.1	1671	1664.7	1645.6
7	*	1738	1732.5	1727.4	1723	1719.5	–	1707.7	1702	1697	1674	1670.9	1664.8	*
8	3341.4	1738.8	1732.7	1727.9	1722	1717.3	1712.1	1707.1	1702	1697.3	1674	1669	1664.9	1645.1
9	3165	–	–	–	1721.9	1718	1712.1	1707.9	1701.9	1697.1	1674	1670.7	1664.7	1647.6
10	3341.7	1738.6	1732.6	–	1722.8	1716.1	1712.5	1707.5	1702	1697.2	1673.8	1671.1	1664.7	1645.2
11	3343.7	1738.7	1732.5	1728	1722.1	1717.1	1712.4	1707	1702.1	1697.6	1673.8	1668.8	1664.8	1645
12	3353.2	1738.7	1732.6	1728	1722	1716.5	1713.1	1706.9	1702.1	1697.5	1673.6	1668.5	1664.5	1645
13	*	1739	1732	1728.2	1722.5	1714.8	–	1707.8	1701.8	1697.2	1673.9	1669.2	1665	1645.1
14	*	1738.5	1732.2	1727.9	1722.1	1714.9	–	1707.4	1702	1696.8	1674	1669.8	1664.9	1645.4
15	*	1738.7	1732.3	1728	0	1714	–	1707.8	1702.3	1697.2	1674.1	1670.2	1664.9	1645.2
16	*	1738.5	1732.4	1727.8	1723.8	1713.8	–	1707.5	1701.8	1696.9	1674.2	1670.9	1664.6	1645.3
17	3348.7	1738.5	1732.2	1727.9	1724	1719.7	1712	1707.3	1702.3	1697	1674	1669.7	1664.7	1645.2

Номера строк соответствуют номерам растворителей в табл. 2.3. № 17 соответствует ИК-спектру индивидуального вещества (эталоны).

* неопределяемые полосы.

Характеристики близости аналитов и эталона ($b = 6 \text{ см}^{-1}$)

Растворитель	d_E	$\chi_{\text{эксп}}^2$	μ_{sum}	
			выражение для μ_i	
			(2.9)	(2.10)
1	8.31	42.4	0.77	0.81
2	6.52	26.1	0.82	0.85
3	6.03	22.3	0.83	0.87
4	5.29	17.2	0.78	0.86
5	2.57	4.1	0.93	0.97
6	1.92	2.3	0.93	0.98
7	1.82	2.0	0.93	0.98
8	8.01	39.4	0.84	0.88
9	184	$2 \cdot 10^4$	0.78	0.82
10	8.12	40.5	0.82	0.85
11	6.09	22.7	0.85	0.88
12	6.17	23.3	0.83	0.87
13	5.25	16.9	0.87	0.91
14	5.18	16.5	0.89	0.91
15	5.76	20.3	0.88	0.91
16	6.06	22.5	0.88	0.91

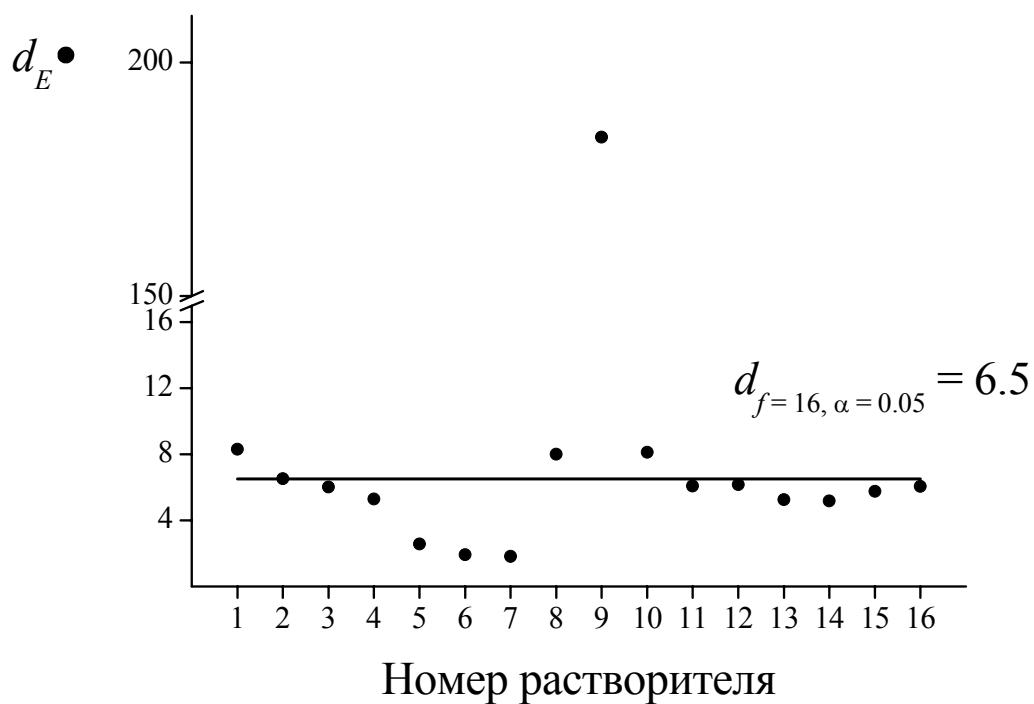


Рис. 2.7. Изменение Евклидова расстояния в зависимости от условий измерения спектров (расчеты проведены для $b = 6 \text{ см}^{-1}$).

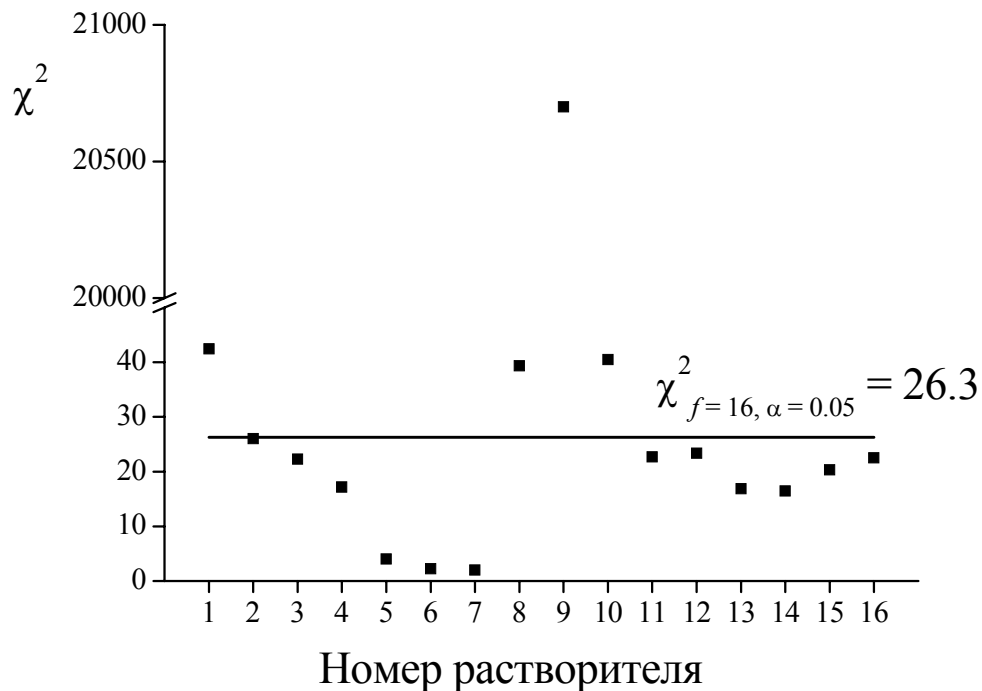


Рис. 2.8. Изменение статистик $\chi^2_{\text{эксп}}$ в зависимости от условий измерения спектров (расчеты проведены для $b = 6 \text{ см}^{-1}$).

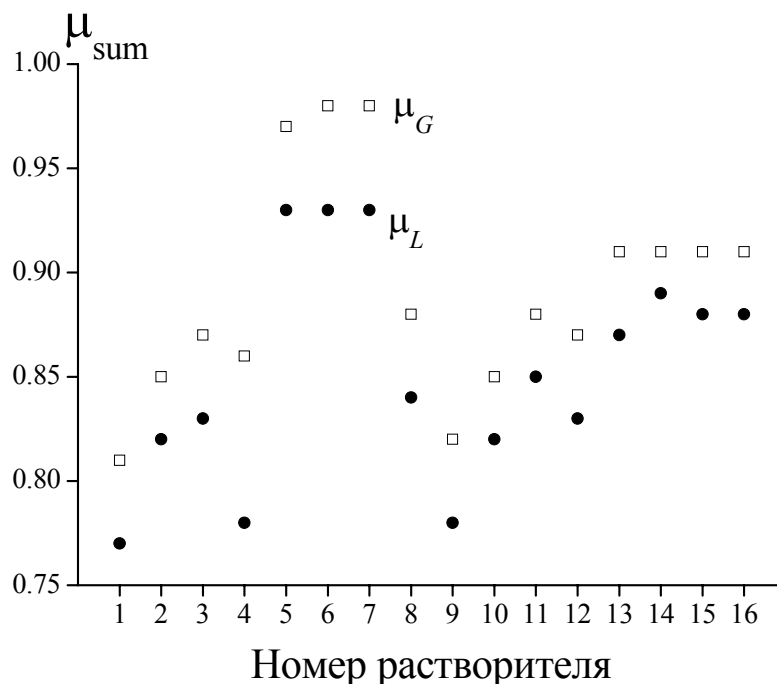
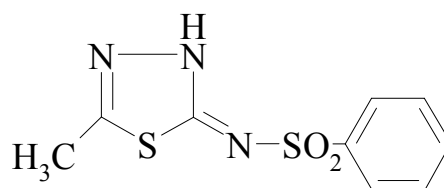


Рис. 2.9. Изменение суммарных функций принадлежности в зависимости от экспериментальных условий (расчеты проведены для $b = 6 \text{ см}^{-1}$).

2.4.3. Идентификация по данным УФ-спектроскопии. На рис. 2.10 представлены зависимости логарифмов молярных коэффициентов поглощения 2-бензол-сульфонимидо-5-метилтиадиазолина



в этаноле (1), растворе 0.1 М NaOH (2) и растворе 0.1 М HCl (3) от длин волн [177]. Спектры оцифровали, начиная с положения максимума полосы поглощения с шагом 2 нм в обе стороны по оси длин волн. Получили три матрицы размером (2×33). С учетом процедуры подготовки данных, оценивая сходство спектров по значению μ_{sum} , процедуре «фазификации» подвергали как отклики ($\lg \varepsilon$), так и предикторы (λ).

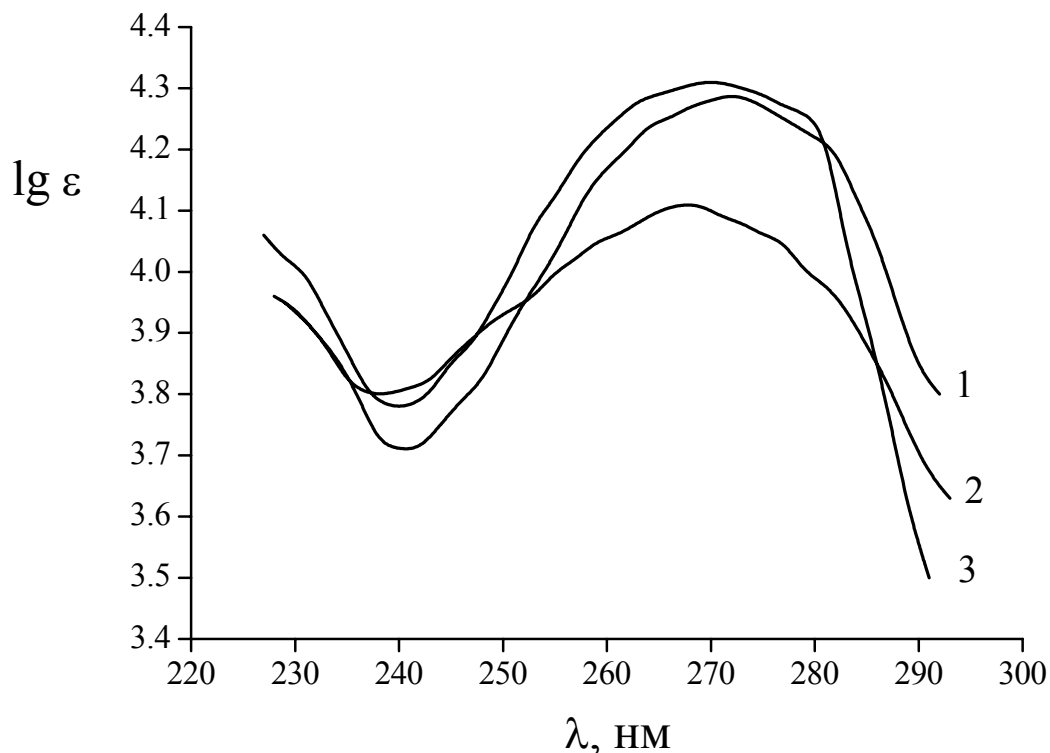


Рис. 2.10. Спектры поглощения 2-бензолсульфонимидо-5-метилтиадиазолина.

Критерий μ_{sum} рассчитывали как суммарную функцию принадлежности вывода об идентичности компонентов (μ_{sum}) двум нечетким множествам $\mu_{\text{sum, отклики}}$ и $\mu_{\text{sum, предикторы}}$ [155]

$$\mu_{\text{sum}} = \mu_{\text{sum, отклики}} + \mu_{\text{sum, предикторы}} - \mu_{\text{sum, отклики}} \cdot \mu_{\text{sum, предикторы}} \quad (2.17)$$

Расчет μ_{sum} выполняли для функций принадлежности двух видов – с двумя линейными ветвями (2.9) и колоколообразной (2.10). Кроме того, вычисляли значения Евклидовых расстояний и статистик $\chi^2_{\text{эксп}}$. Результаты расчетов для различных значений $b_{\lg \varepsilon}$ приведены в табл. 2.6 и на рис. 2.11-2.17.

**Евклидовы расстояния и меры близости для спектров поглощения
2-бензолсульфонимида-5-метилтиадиазолина в различных
экспериментальных условиях**

$d_E; m_E$		Номер спектра		
		1	2	3
Номер спектра	1	0; 1		
	2	0.85; 0.58	0; 1	
	3	0.54; 0.78	0.87; 0.57	0; 1

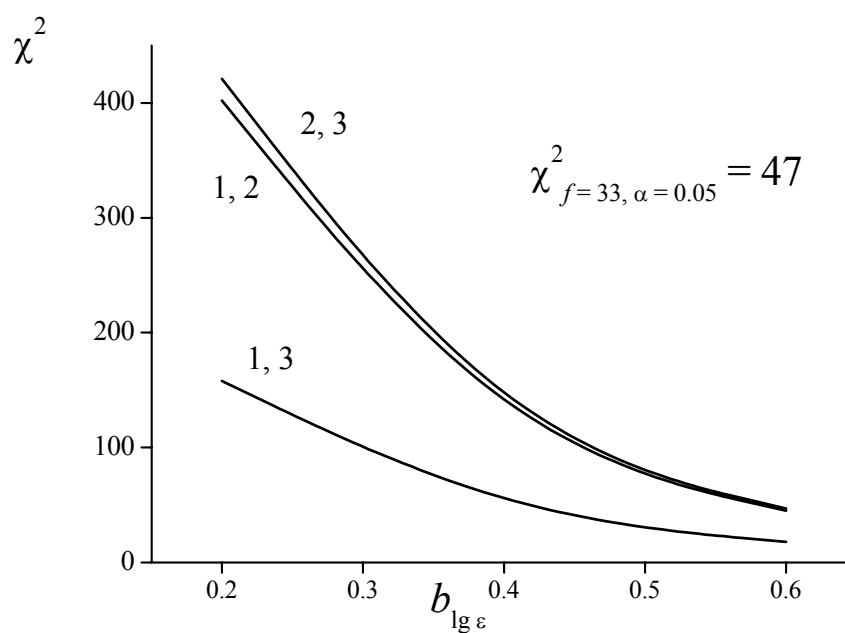


Рис. 2.11. Зависимость $\chi^2_{\text{эксп}}$ от $b_{\lg \varepsilon}$.

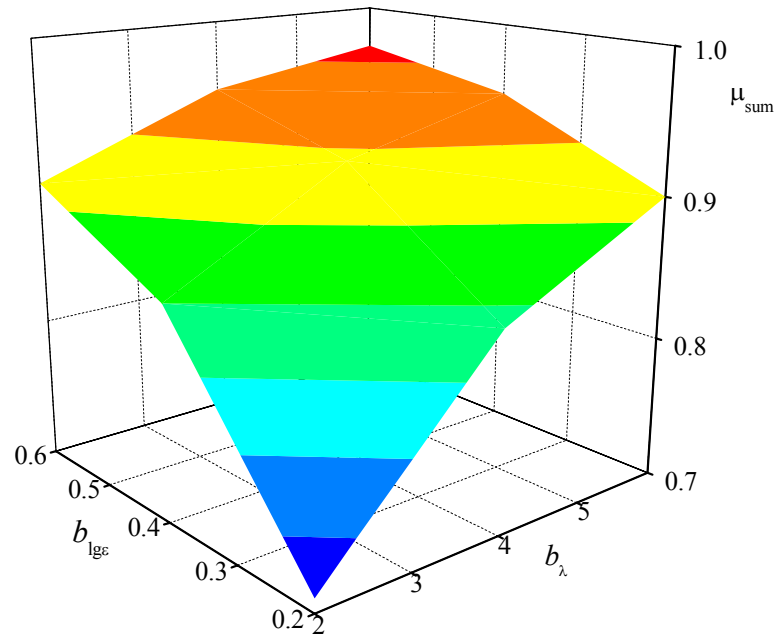


Рис. 2.12. Значения $b_{\text{lg}\varepsilon}$ и b_{λ} , при которых $\mu_{\text{sum}} > 0.6$ для спектров 1 и 2 при функциях принадлежности типа (2.9).

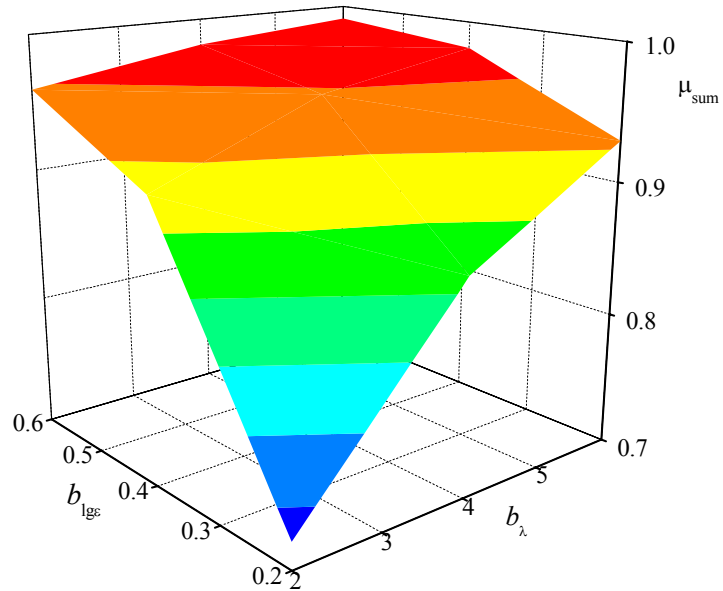


Рис. 2.13. Значения $b_{\text{lg}\varepsilon}$ и b_{λ} , при которых $\mu_{\text{sum}} > 0.6$ для спектров 1 и 2 при функциях принадлежности типа (2.10).

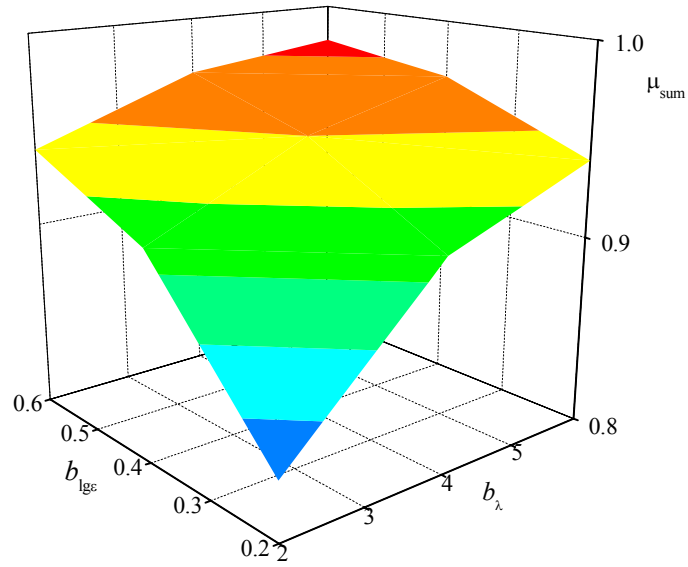


Рис. 2.14. Значения $b_{\lg \varepsilon}$ и b_{λ} , при которых $\mu_{\text{sum}} > 0.6$ для спектров 1 и 3 при функциях принадлежности типа (2.9).

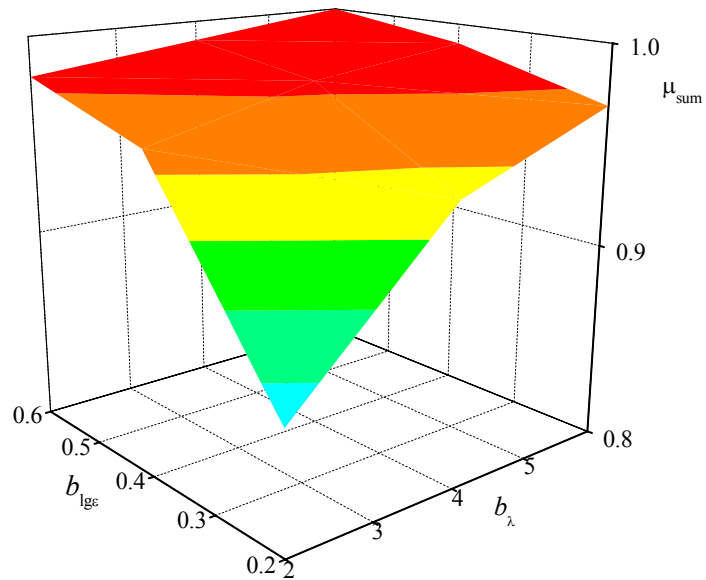


Рис. 2.15. Значения $b_{\lg \varepsilon}$ и b_{λ} , при которых $\mu_{\text{sum}} > 0.6$ для спектров 1 и 3 при функциях принадлежности типа (2.10).

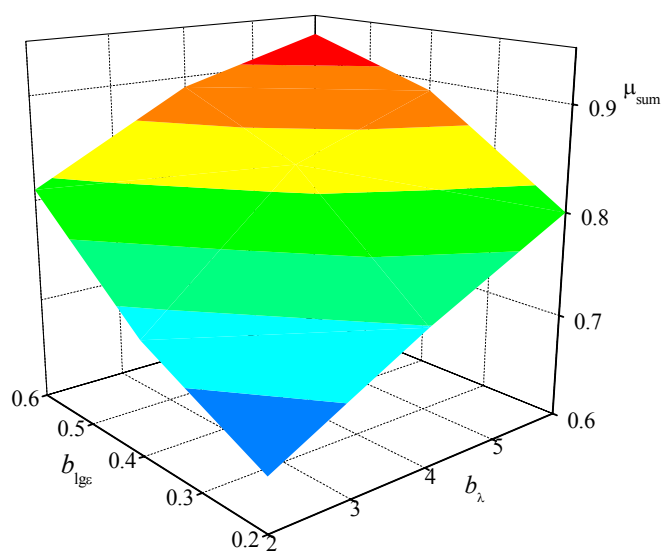


Рис. 2.16. Значения $b_{lg\epsilon}$ и b_λ , при которых $\mu_{sum} > 0.6$ для спектров 2 и 3 при функциях принадлежности типа (2.9).

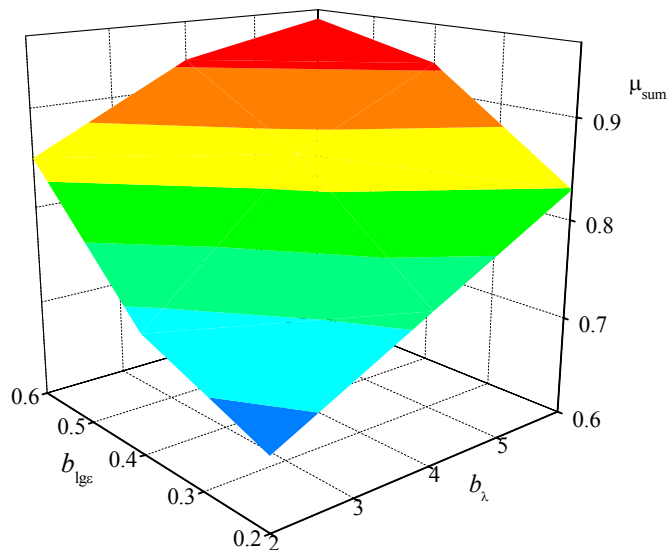


Рис. 2.17. Значения $b_{lg\epsilon}$ и b_λ , при которых $\mu_{sum} > 0.6$ для спектров 2 и 3 при функциях принадлежности типа (2.10).

Анализ значений статистик $\chi_{\text{эксп}}^2$ показывает, что выводы о совпадении спектров 1-3 определяются заданными значениями ширины окна, с использованием которых вычисляются критические значения $\chi_{f,\alpha}^2$. Так, вывод об их идентичности спектров 1 и 3 правомерен при $b_{\lg \varepsilon} > 0.43$, что соответствует стандартному отклонению $\sigma_{\lg \varepsilon} > 0.09$ (относительная погрешность измерения $\lg \varepsilon \sim 2\%$). Аналогично, пары спектров 1, 2 и 2, 3 можно признать совпадающими, если принять относительную погрешность измерения молярных коэффициентов поглощения $> 3\%$.

Наш алгоритм, допускающий введение, помимо $b_{\lg \varepsilon}$, и допустимого размаха b_λ , дает гораздо более отчетливый вывод о близости спектров 1 и 3 и о существенном отличии спектров 1, 2 и спектров 2, 3. Таким образом, использование нечетких критериев делает процедуру идентификации более убедительной и адаптирующейся к данным.

2.5. Выводы к разделу 2

1. Для расчета критерия сходства аналита с эталоном на основе теории нечетких множеств необходимо указывать лишь допустимый размах положений пиков для аналита и эталона. По сравнению с априорным подходом уменьшается число параметров, контролирующих работу алгоритма идентификации и основанных на модельных представлениях. Привлечение гипотез о законе распределения экспериментальных погрешностей не требуется.
2. Адаптивность алгоритма нечеткой идентификации к данным увеличивается, если придавать нечеткость не только положениям пиков, но и величинам поглощения.
3. Выводы о тождестве аналита и эталона, сделанные на основе применения разработанного алгоритма, устойчивы как к варьированию вида задаваемых функций принадлежности, так и к предположениям о допустимом размахе измеряемых свойств.

РАЗДЕЛ 3

ХИМИКО-АНАЛИТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ТВЕРДОФАЗНЫХ АНАЛИТИЧЕСКИХ РЕАГЕНТОВ НА ОСНОВЕ ОРГАНО- КРЕМНЕЗЕМНЫХ ГИБРИДНЫХ МАТЕРИАЛОВ

3.1. Гибридные органо-кремнеземные материалы в задачах разделения и концентрирования

Гибридные органо-кремнеземные материалы обычно используют в виде порошков и пленок, закрепленных на твердой подложке и применяют как твердофазные аналитические реагенты при пробоподготовке (например, для концентрирования и разделения ионов металлов, при сорбции из водных [178, 179] и неводных [180] растворов, ионообменного разделения ионов [181] и др.), а также как реагенты для комбинированных и тестовых методов анализа [182-195].

Если закрепленный аналитический реагент при взаимодействии с аналитом образует окрашенный или флуоресцирующий продукт, твердофазный аналитический реагент можно использовать для гибридного сорбционного спектрофотометрического или флуориметрического обнаружения или определения аналита.

Выбор оптимальных условий использования твердофазного аналитического реагента, прогнозирование селективности разделения значительно упрощаются, если известны такие химико-аналитические характеристики, как эффективная сорбционная емкость, стехиометрический состав продуктов взаимодействия органических реагентов с аналитом, параметры сорбционных равновесий.

Указанные характеристики определяют по изотермам сорбции аналита исследуемым материалом с привлечением содержательных физико-химических моделей и численных методов их параметрической идентификации. При решении этой задачи возможны осложнения, связанные с неидеальностью процесса сорбции и необходимостью совместного определения сорбционной емкости и параметров сорбционных равновесий.

В настоящем разделе описан хемометрический подход, объединяющий робастное оценивание по Хьюберу с алгоритмами теории нечетких множеств и предоставляющий объективную информацию о химико-аналитических характеристиках твердофазных аналитических реагентов. Неидеальность сорбционных равновесий учитывается с помощью модели полидентатного связывания. Подход, работоспособность которого подтверждена при исследовании модельной системы, успешно применен для определения характеристик двух новых гибридных материалов.

3.2. Задача моделирования равновесий по данным количественного физико-химического анализа

Моделирование равновесий в растворах – хорошо изученная проблема, для решения которой развиты как обоснованные содержательные модели, так и многочисленные расчетные методы. В области сорбционных и ионообменных процессов ситуация менее благоприятна. Связано это, во-первых, с тем, что в условиях эксплуатации концентрация активных центров материала может существенно отличаться от концентрации активных групп, определенной элементным анализом. Вследствие этого удельную концентрацию активных центров (эффективную емкость материала) следует определять вместе с составом и константами устойчивости иммобилизованных комплексов, что значительно усложняет процесс расчетов и построения моделей. Во-вторых, на сорбционные и ионообменные процессы могут оказывать влияние эффекты, отсутствующие для равновесий в растворах.

Таким образом, для выявления и количественного описания энергетической неоднородности закрепленных реагентов, латеральных взаимодействий и других усложняющих процесс моделирования факторов, необходимо применять специальные методы.

Существуют два принципиально отличных друг от друга способа, пригодных для определения состава и констант устойчивости комплексов.

Первый – идентификация всех присутствующих в исследуемой системе реагентов и продуктов реакции и определение равновесных концентраций каждого из них. Классическим примером такого подхода для равновесий комплексообразования в растворах служит работа Н. Бьеррума [196], в которой описано выделение из равновесной системы 6 комплексов состава $[\text{Cr}(\text{SCN})_i]^{(3-i)+}$, $i = 1, 2, \dots, 6$, и измерения их выходов. После этого определение констант устойчивости не представляет проблемы. Используя современные методы исследований (ЯМР, масс-спектропия), можно идентифицировать сосуществующие комплексы в смесях самого различного состава [197]. Однако на практике используют, преимущественно, другой, универсальный способ определения состава и устойчивости комплексов – количественный физико-химический анализ (КФХА).

КФХА – раздел физико-химического анализа, в котором одновременно со стехиометрическим составом продуктов реакций определяют их термодинамические, и, возможно, другие физико-химические характеристики [198]. КФХА является основным способом определения термодинамических характеристик реакций, протекающих в разных средах – истинных и организованных растворах, расплавах, на поверхности твердых тел. Несмотря на долгую историю развития и применения, формирование надежных стратегий использования КФХА и наличие программно реализованных вычислительных средств, проблематика количественного физико-химического анализа не теряет актуальности. Это связано, во-первых, с распространением КФХА на новые типы объектов, значение которых постоянно растет, а области применения расширяются. Речь идет, в частности, об исследовании свойств органо-минеральных гибридных материалов, объединяющих неорганический каркас и органический модификатор. Во-вторых, развитие хемометрии позволяет существенно повысить надежность численных оценок, получаемых в результате использования КФХА, и предопределяет внедрение в практику новых расчетных процедур.

Изучая равновесные системы методами КФХА, находят число сортов, стехиометрический состав, константы устойчивости и другие физико-химические параметры химических форм. Первичными данными КФХА являются зависимости состав – свойство, где под «составом» следует понимать количества вещества (или начальные концентрации) реагентов в исследуемой системе, а под «свойством» – экспериментально измеренную характеристику равновесной системы (в случае исследования сорбционных равновесий – величины адсорбции либо коэффициента распределения) [198, 199].

Построение модели комплексообразования в равновесной системе состоит из следующих шагов:

1. Планирование и осуществление эксперимента – получение первичных экспериментальных данных КФХА – зависимости величин измеряемого свойства равновесной системы (A) от ее начального состава (зависимостей «состав – свойство»).
2. Структурная идентификация модели – выбор такого вида функции в уравнении зависимости «состав – свойство», что имеющие физический смысл параметры, например константы равновесия, будут подгоночными параметрами модели.
3. Параметрическая идентификация модели – определение подгоночных параметров из условия «наилучшей» аппроксимации зависимости «состав – свойство» моделью.
4. Проверка статистической адекватности модели.
5. Верификация модели или дискриминация нескольких адекватных моделей.

Дискриминация предусматривает выбор на основе теоретических соображений или (и) результатов дополнительных экспериментов одной модели из нескольких адекватных. При верификации наилучшими аргументами в пользу модели признают ее способность описать данные КФХА, полученные независимым экспериментальным методом или в другой области плана эксперимента, а также независимое подтверждение состава присутствующих в системе химических форм структурно чувствительными методами исследования.

Следует отметить, что перечисленные этапы КФХА лишь в идеальных случаях являются последовательными. На практике часто приходится возвращаться к тому или иному этапу (например, в случае получения неадекватных моделей, т.е. невозможности получить приемлемый результат в п. 4, следует вернуться к шагу 3, а иногда к шагу 2 или даже к шагу 1).

Таким образом, построение модели представляет собой сложную комплексную задачу, решение которой требует на разных этапах привлечения методов химической термодинамики и хемометрики. В частности, необходимо обращаться к теориям планирования эксперимента, робастного оценивания, решения математически некорректных задач, использовать статистические и нестатистические методы проверки адекватности и др.

Остановимся подробнее на некоторых этапах КФХА.

Основные принципы этапа планирования эксперимента для получения первичных данных КФХА подробно описаны в работах [200, 201]. При планировании и проведении эксперимента возникают определенные трудности. Во-первых, далеко не всегда удается использовать оптимальный экспериментальный метод. Во-вторых, границы диапазона концентраций реагентов, в котором подгоночные параметры определяются с максимальной точностью, можно оценить, исследуя уже готовую модель. В силу этих причин актуальность приобретает задача построения модели на основе зависимостей «состав – свойство», полученных по, возможно, неоптимальному плану. Тогда одним из результатов исследования будет являться выработка рекомендаций по изменению плана эксперимента.

На этапе структурной идентификации модели необходимо определить вид функции в зависимости «состав – свойство» с помощью трех групп уравнений [198]:

- уравнений связи измеряемого свойства с равновесным составом;
- уравнений материального баланса;
- уравнений закона действия масс (ЗДМ).

Основные экспериментальные методы исследования равновесий позволяют измерять линейные комбинации равновесных концентраций химических форм. Тогда измеряемое свойство

$$A = \sum_{i=1}^S \alpha_i [L_i]_k, \quad i = 1, 2, \dots, S, \quad (3.1)$$

где L_i – реагенты, S – их количество, $[L_i]$ – равновесная концентрация L_i , α_i – фактор интенсивности реагента L_i . При формулировке условий материального баланса удобно реакции в системе записать в канонической форме:

$$\sum_{j=1}^Y \nu_{ij} B_j = L_i, \quad (3.2)$$

где ν_{ij} – стехиометрические коэффициенты, B_j – подмножество реагентов, называемое независимыми компонентами. Независимые компоненты B_j друг в друга при реакциях не переходят и, следовательно, инвариантами системы являются количества вещества компонентов. Если реакции происходят в растворах, объем которых при этом не меняется, инвариантами являются и общие (аналитические) концентрации компонентов:

$$t_j = \sum_{i=1}^S \nu_{ij} \cdot c(L_i) = \sum_{i=1}^S \nu_{ij} \cdot [L_i], \quad j = 1, 2, \dots, Y, \quad (3.3)$$

где $c(L_i)$ – начальные (известные по условиям смешивания реагентов) концентрации L_i . Остается записать уравнения ЗДМ или, что эквивалентно, выражения для химических потенциалов реагентов. Для реакций в неполярных растворителях и водных растворах с добавками избытка фонового электролита можно принять, что коэффициенты активности химических форм при реакциях остаются неизменными, и записывать уравнения ЗДМ в виде:

$$[L_i] = \exp \left\{ \ln \beta_i + \sum_{j=1}^Y \nu_{ij} \ln [B_j] \right\}, \quad (3.4)$$

где $[B_j]$ – равновесные концентрации независимых компонентов, β_i – концентрационная константа устойчивости химической формы L_i .

Модели со структурой, задаваемой уравнениями (3.1 – 3.4), применяются для моделирования равновесий в растворах и различных системах, связанных с растворами (система раствор-осадок, экстракционные равновесия и др.).

Однако в случае возникновения эффектов, заведомо отсутствующих в растворах, необходима модификация уравнений закона действия масс, учитывающая эти эффекты. Необходимость в этом возникает, например, при учете влияния энергетической неоднородности поверхности на сорбционные равновесия [198, 202-206].

Решение задачи параметрической идентификации затрудняется математически некорректным характером самой задачи: малые (даже на уровне экспериментального шума) возмущения входных данных могут вызывать большие возмущения решений и приводить к множеству моделей, воспроизводящих экспериментальные данные в пределах их погрешностей.

3.3. Модели для описания сорбционных равновесий

Подходы к структурной и параметрической идентификации моделей разного типа сильно отличаются [202, 207, 208]. В настоящем подразделе обсуждается построение моделей, описывающих сорбционные равновесия.

Исходными данными для моделирования являются изотермы сорбции, а регистрируемой экспериментально величиной – адсорбция

$$A = \frac{(t(M) - [M]) \cdot V}{m}, \quad (3.5)$$

где $t(M)$ – общая концентрация сорбата M , моль/л, $[M]$ – его равновесная концентрация, моль/л, V – объем раствора, л, m – масса навески сорбента, г. Значения $t(M)$, V и m известны по условиям смешивания реагентов, значение $[M]$ измеряют после достижения системой равновесия.

Ключевой химико-аналитической характеристикой материала является его эффективная сорбционная (ионообменная) емкость (t_Q , моль/г) [209]. Часто ее приходится определять по изотермам сорбции.

Это связано с тем, что t_Q – удельная концентрация активных групп – может значительно отличаться от теоретической емкости, найденной по данным элементного анализа. Эффективная емкость зависит от природы сорбата и растворителя, рН и ионной силы раствора и других факторов [209-211].

В работе [209] выделяются следующие типы емкостей (табл. 3.1).

Таблица 3.1.

Типы ионообменных / сорбционных емкостей

Тип	Определение	Примечание
Максимальная (теоретическая) емкость	Удельная концентрация активных групп	Постоянна, используется как характеристика материала
Эффективная емкость	Удельная концентрация активных групп	Обычно меньше максимальной емкости. Зависит от экспериментальных условий (рН, растворитель, природа сорбата и др.)
Полезная емкость	Характеристика, которую используют, если состояние равновесия не достигнуто	Зависит от экспериментальных условий (степень ионного обмена и др.)
Динамическая ионообменная емкость	Используется как характеристика колоночных процессов	Зависит от условий процесса (скорость потока, концентрация растворов и др.)

Обсуждая свойства ионообменных материалов, Ф. Гельферрих указывает [209]: «с практической точки зрения, число ионов, способных к обмену более

важно, нежели число ионогенных групп. Например, группы слабой кислоты или слабого основания могут быть не полностью ионизированы, следовательно, частично неэффективными».

Таким образом, хотя теоретическую емкость определить не слишком сложно, проведя элементный анализ материала, эффективная емкость – более важная характеристика, хотя ее определение во многих случаях затруднено.

Рассмотрим сначала случай идеальной адсорбции сорбата M на поверхности сорбента с активными центрами \bar{Q} . Связывание M описывает реакция



где β – константа сорбционного равновесия. Чертой обозначены реагенты и продукты на поверхности. При идеальной адсорбции все центры \bar{Q} энергетически однородны, латеральные взаимодействия отсутствуют. Константа равновесия

$$\beta = \frac{[\overline{MQ}]}{[M] \cdot [\bar{Q}]} \quad (3.7)$$

в таком случае не зависит от степени заполнения поверхности частицами сорбата, адсорбция

$$A = [\overline{MQ}] = \beta \cdot [M] \cdot [\bar{Q}], \quad (3.8)$$

сорбционная емкость

$$t(\bar{Q}) = [\bar{Q}] + [\overline{MQ}] = [\bar{Q}](1 + \beta[M]). \quad (3.9)$$

Зная A и $[M]$, легко найти коэффициент распределения сорбата между фазами

$$D = \frac{A}{[M]}. \quad (3.10)$$

Легко показать [198], что

$$\frac{1}{D} = \frac{1}{\beta \cdot t_Q} + \frac{1}{t_Q} \cdot [M], \quad (3.11)$$

т.е. зависимость $1/D$ от $[M]$ описывается уравнением прямой с тангенсом угла наклона $a = 1/t_Q$ и свободным членом $b = 1/\beta \cdot t_Q$. Применяя линейный МНК, оценивают коэффициенты a и b , а по их значениям – подгоночные параметры модели – сорбционную емкость t_Q и константу сорбционного равновесия β .

Этот простой метод не лишен ограничений, поскольку предполагает образование единственного вида комплексов и идеальный характер сорбционного равновесия.

Вследствие математической некорректности задачи параметрической идентификации и необходимости определять как дискретные, так и непрерывные переменные, создание некой универсальной стратегии построения моделей представляется невозможным. Наиболее успешными оказываются стратегии моделирования «снизу-вверх» – от неадекватной исходной модели, через более адекватные, к одной или нескольким приемлемым.

Отклонение сорбции от идеальности наблюдали для огромного числа систем, от физической сорбции углеводов на силикагеле [212] до связывания кислорода гемоглобином [213]. Гибридные органо-минеральные материалы – новый объект исследования, но и для сорбционных процессов на их поверхности уже зафиксированы эффекты неидеальности [214, 215]. В этом случае зависимость $1/D$ от $[M]$ является нелинейной, и описанный простой способ определения t_Q и β неприменим. Попытка оценить t_Q и β по участкам насыщения на изотерме адсорбции не всегда приводит к успеху. Объяснить это можно следующим образом: если образующиеся комплексы относительно нестабильны ($\lg \beta < 4$), для занятия всех адсорбционных центров необходим большой избыток соли, практически недостижимый экспериментально, поэтому прямое определение t_Q и β из изотерм адсорбции дает нестабильные результаты.

Чаще всего отклонения от идеальности обсуждаются в терминах эффектов «энергетической неоднородности поверхности» (которую рассматривают как имманентную характеристику материала) или «кооперативности». В случае

эффектов кооперативности сродство центров связывания \overline{Q} к сорбату M в процессе сорбции меняется.

Для описания неидеальной адсорбции предназначены многие модели [216]. В настоящей работе использовали модель полидентатного связывания [217], интерпретирующую неидеальность как проявление эффектов кооперативности. В этой модели реакционная поверхность материала рассматривается как совокупность центров $\overline{Q_Z}$, каждый из которых содержит Z активных групп Q , удельная концентрация центров $\overline{Q_Z}$ при этом равняется $\frac{tQ}{Z}$. Связывание частиц сорбата центрами $\overline{Q_Z}$ рассматривается как ступенчатый процесс, а равновесие описывается Z значениями констант равновесия $\beta_i^{(Z)}$, где i – номер ступени. Пример модели для $Z = 4$ приведен на рис. 3.1.

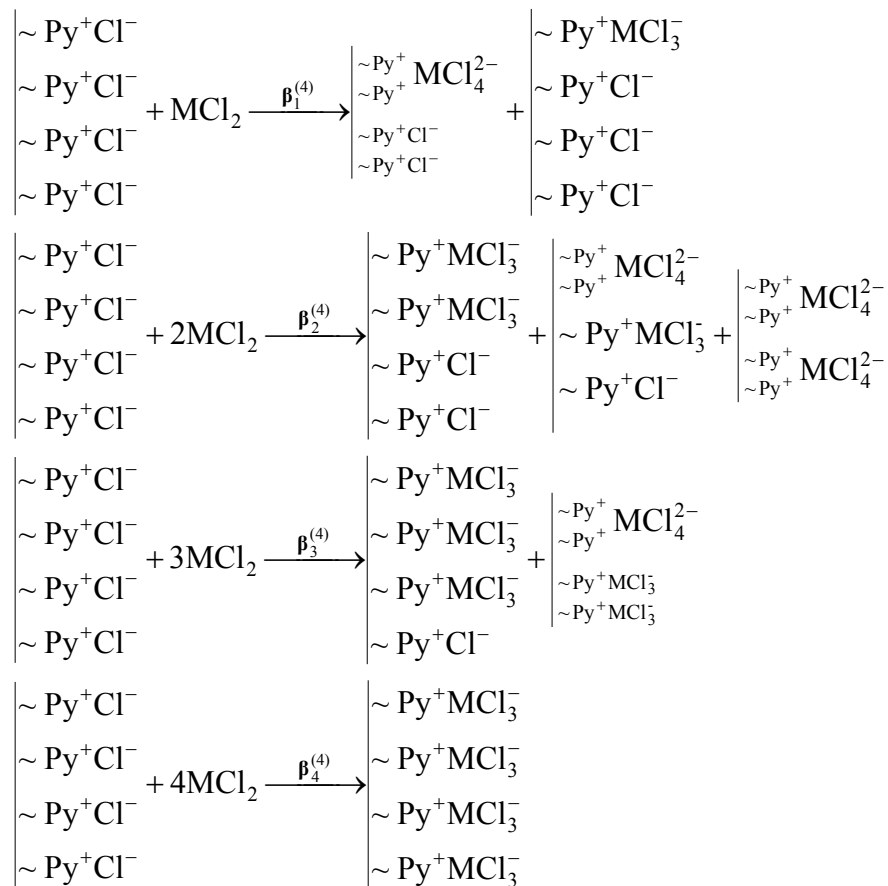


Рис. 3.1. Описание связывания частиц сорбата MCl_2 активными группами поверхности материала SiPy^+Cl^- в модели тетрадентатного связывания.

При соблюдении неравенства

$$K_i^{(Z)} < K_{i-1}^{(Z)} < \dots < K_1^{(Z)} \quad (3.12)$$

эффекты кооперативности не проявляются в адсорбционных процессах, либо наблюдается отрицательная кооперативность. В случае отклонения значений ступенчатых констант равновесия от неравенства (3.12) имеет место положительная кооперативность, т.е. рост сродства сорбционных центров к частицам сорбата по мере заполнения поверхности.

Построение модели начинается с малого значения Z (например, $Z = 2$). Соответствующие оценки $\beta_i^{(Z)}$ рассчитывают оптимизацией определенного функционала и оценивают адекватность построенной модели (см. ниже). Если модель не адекватна, значение Z увеличивают и повторяют процедуру расчета $\beta_i^{(Z)}$ до построения адекватной модели.

Расчет констант равновесия $\beta_i^{(Z)}$ осуществляют численно. В качестве оценки истинных значений параметров принимают такой вектор $\left| \beta_i^{(Z)*} \right\rangle$, который обращает в минимум выбранную критериальную функцию U , характеризующую качество аппроксимации экспериментальных данных моделью:

$$\left| \beta_i^{(Z)*} \right\rangle = \arg \min U \left(\left| \beta_i^{(Z)*} \right\rangle \right). \quad (3.13)$$

Со времен отхода в середине 60-х годов XX в. от графических методов параметрической идентификации и вплоть до конца 90-х годов у химиков не возникало вопроса о виде критерия U : во всех алгоритмах, реализованных в виде компьютерных программ, его задавали как взвешенную сумму квадратов невязок [200, 201, 218], что соответствует нахождению оценок параметров модели с помощью МНК [134, 135]. Как было упомянуто ранее, при соблюдении предпосылок МНК, оценки этого метода являются асимптотически эффективными, несмещенными и состоятельными.

Рассчитывая параметры модели полидентатного связывания, функционал U назначают как

$$U(\langle \beta_i^{(Z)} \rangle) = \sum_{k=1}^N w_k (\mathcal{A}_k - A_k^{\text{эксп}})^2, \quad (3.14)$$

где \mathcal{A}_k – значение адсорбции в k -й точке, оцененное в рамках модели полидентатного связывания, $A_k^{\text{эксп}}$ – экспериментальное значение адсорбции в k -й точке, k – номер точки изотермы сорбции, N – их число, w_k – статистический вес k -го измерения,

$$w_k = (s_r \cdot A_k^{\text{эксп}})^{-2}, \quad (3.15)$$

где s_r – относительная погрешность k -го измерения.

В работе [219] показано, что во многих случаях гипотеза о нормальном распределении экспериментальных погрешностей не подтверждается, а реальные массивы данных всегда на 5–10% засорены «промахами». В таком случае рекомендуется оценивать параметры модели робастными методами, например, методом наименьших модулей (МНМ).

Известно, что условия формирования обоих предельных законов распределения для экспериментатора практически неразличимы [220], а МНК и МНМ являются равно обоснованными.

Найти компромисс позволяют M -оценки Хьюбера, сочетающие свойства МНК и МНМ. M -оценки Хьюбера обосновываются методом максимума правдоподобия, если распределение погрешностей подчиняется «модели грубых промахов», когда свойства A_k измерены с погрешностями ε , плотность распределения которых задается формулой

$$\rho(\varepsilon) = \frac{[(100 - \delta) \cdot \tau(0, \sigma_{\text{Gauss}}) + \delta \cdot h(\varepsilon)]}{100}, \quad (3.16)$$

где $\tau(0, \sigma_{\text{Gauss}})$ – плотность нормального распределения с нулевым средним и стандартным отклонением σ_{Gauss} , $h(\varepsilon)$ – плотность вероятности грубых промахов (плотность распределения с длинными хвостами, в качестве которого может выступать и плотность распределения Лапласа), δ – интенсивность промахов, (%).

Функция потерь для M -оценок Хьюбера принимает вид

$$\rho(\xi_i) = \begin{cases} 0.5 \cdot \xi_{lk}^2 & \text{при } |\xi_{lk}| \leq c_{out} \\ c_{out} \cdot |\xi_{lk}| - 0.5 \cdot c_{out}^2 & \text{при } |\xi_{lk}| > c_{out} \end{cases}, \quad (3.17)$$

где константа c_{out} зависит от значения δ . Ни априорно, ни апостериорно интенсивность грубых промахов определить невозможно. Анализ данных с помощью M -оценок Хьюбера обладает повышенной гибкостью по сравнению с МНК: появляется новая варьируемая переменная – интенсивность промахов δ . Значения $\delta = 0$ отвечает МНК, а $\delta = 100\%$ – МНМ. Обработка данных при задании разных δ в интервале $0 \leq \delta \leq 100\%$ позволяет найти решение, малочувствительное к отступлению от гипотезы о нормальном распределении экспериментальных погрешностей. Испытание различных гипотез о статистических свойствах измерений (варьирование δ) – одно из характерных для хемометрии и теории анализа данных средств, адаптирующих процедуру параметрической идентификации к особенностям обрабатываемых данных.

Помимо точечных оценок параметров, соответствующих минимуму критерия U , находят их ковариационную матрицу размером $z \times z$.

Алгоритмы нахождения таких оценок описаны в литературе [198]. В настоящей работе использован метод, реализованный в качестве надстройки к табличному процессору Microsoft Excel CLINP 2.1 [198].

Адекватность воспроизведения моделью экспериментальных данных проверяют с помощью статистических критериев, важнейшим из которых является критерий χ^2 . Модель считают адекватной, если выполняется неравенство

$$\chi_{\text{эксп}}^2 = \sum_{k=1}^N w_k \left(\bar{A}_k - A_k^{\text{эксп}} \right)^2 < \chi_{f,\alpha}^2, \text{ где } \chi_{f,\alpha}^2 - 100\alpha\text{-процентная точка распределения}$$

χ^2 для f степеней свободы. В случае M -оценивания критерий модифицируют, корректируя число степеней свободы: $f = (N - z) \cdot \{1 + 0.5 \gamma_2 \cdot (N - z)/N\}^{-1}$.

Отметим еще одну особенность находимых оценок. Точечные оценки $\left\langle \beta_i^{(z)*} \right\rangle$, соответствующие минимуму критериального функционала, имеют не слишком высокую ценность вследствие их возможной коррелированности.

Вместо точечных оценок правильнее определять доверительную область R , в которой вероятность нахождения $\left| \beta_i^{(Z)} \right\rangle_{\text{true}}$ составляет

$$P\left(\left| \beta_i^{(Z)} \right\rangle_{\text{true}} \in R\right) = \gamma, \quad (3.18)$$

где γ – некоторая задаваемая вероятность (например, 0.95). Зная $\left| \beta_i^{(Z)*} \right\rangle$ и их ковариационную матрицу $D\left(\left| \beta_i^{(Z)*} \right\rangle\right)$, можно построить z -мерный эллипсоид, ограничивающий доверительную область R с приближенной вероятностью $100\gamma\%$. Его центр размещен в точке $\left| \beta_i^{(Z)*} \right\rangle$, а границы задаются условием [221, 222]

$$\left(\left| \beta_i^{(Z)*} \right\rangle - \left| \beta_i^{(Z)} \right\rangle\right) \times D\left(\left| \beta_i^{(Z)*} \right\rangle\right) \times \left(\left| \beta_i^{(Z)*} \right\rangle - \left| \beta_i^{(Z)} \right\rangle\right)^T = z \cdot F(1 - \gamma, z, f), \quad (3.19)$$

где $F(1 - \gamma, z, f)$ – γ -процентная точка распределения Фишера с f степенями свободы.

При моделировании равновесий в растворах начальные количества вещества (или общие концентрации реагентов) считают известными точно. Лишь в отдельных работах [223, 224] рассматривалась задача одновременного расчета начальных концентраций реагентов и констант равновесия.

Очевидно, именно эта задача аналогична обсуждаемой проблеме нахождения подгоночных параметров модели t_Q и $\left| \beta_i^{(Z)} \right\rangle$ по изотерме адсорбции.

Ранее предложенным алгоритмам характерна низкая численная устойчивость, вызванная высокой коррелированностью подгоночных параметров модели, и они не нашли широкого применения. Кроме того, к недостаткам этих алгоритмов следует отнести их ориентацию исключительно на МНК.

3.4. Алгоритм расчета сорбционной емкости и констант сорбционных равновесий на основе теорий нечетких множеств и робастного оценивания

Стремясь разделить шаги, на которых рассчитывают t_Q и β_i , мы предлагаем подход, использующий аппарат теории нечетких множеств и робастное оценивание по Хьюберу [225]. На первом этапе находим интервал возможного варьирования t_Q (все t_Q из этого интервала допускают построение моделей, адекватных эксперименту). На втором этапе массив адекватных моделей сводится к одной окончательной модели с единственным набором параметров $\left| \beta_i^{(z) \text{ final}}; t_Q^{\text{ final}} \right\rangle$.

Перед началом расчетов любые значения t_Q , не превышающие теоретическую емкость T , рассматриваем как допустимые. Моделирование начинаем с $t_Q = T$ и выполняем для меньших значений, равномерно распределенных в интервале $0 < t_Q < T$. Для каждого из испытываемых t_Q при интенсивности грубых промахов δ вычисляем M -оценки z неизвестных $\lg \beta_i^{(z)} \left(\left| \lg \beta_i^{(z)} \right\rangle \right)$. На основании этих расчетов определяем интервал возможного варьирования t_Q , а всем испытанным t_Q в соответствие ставим доверительные эллипсоиды $R_{\left| \lg \beta_i^{(z)} \right\rangle}$ и значения статистик $\chi_{\text{экср}}^2$. Переход от массива наборов параметров $\theta = \left| \lg \beta_i^{(z)}; t_Q \right\rangle$ к единственному набору $\left| \lg \beta_i^{(z) \text{ final}}; t_Q^{\text{ final}} \right\rangle$ простым усреднением полученных оценок представляется крайне нежелательным, поскольку эта процедура в рамках метода максимума правдоподобия базируется на гипотезе о нормальном распределении независимых случайных величин. Очевидно, что реальная ситуация далека от указанной гипотетической. Более того, поскольку наборы параметров $\theta = \left| \lg \beta_i^{(z)}; t_Q \right\rangle$ получены обработкой одних и тех же экспериментальных данных, методы, основанные на концепции статистической вероятности, могут привлекаться лишь как эвристические средства.

С наборами параметров θ обращались как с многомерными нечеткими числами [226, 227]. Единого способа задания функции принадлежности на наборе четких чисел (фаззификация), как и единого алгоритма выполнения обратной операции – перехода от нечеткого числа к четкому (дефаззификация) – не существует.

Процедура дефаззификации в теории нечетких множеств аналогична нахождению характеристик положения (математического ожидания, моды, медианы) случайных величин в теории вероятности. Простейшим способом выполнения процедуры дефаззификации является выбор четкого числа, соответствующего максимуму функции принадлежности.

Конструируя алгоритм, руководствовались следующими соображениями. Во-первых, для каждой модели со значениями статистики $\chi_{\text{эксп}}^2$ должны быть ассоциированы доверительные области R_θ (не точечные оценки θ^*). Во-вторых, приемлемыми следует считать все модели, адекватные эксперименту (и соответствующие этим моделям наборы параметров), но модели, обеспечивающие меньшие значения $\chi_{\text{эксп}}^2$ выглядят более привлекательно, и им следует приписать большие значение функций принадлежности.

Наконец, представляется естественным требовать, чтобы окончательные оценки параметров θ^{final} мало зависели от способа задания функции принадлежности.

Для каждой адекватной модели функцию принадлежности задавали двумя способами:

$$\mu_v^I = G \cdot \exp(-\chi_{\text{эксп},v}^2), \quad v = 1, 2, \dots, V, \quad (3.20)$$

где v – номер модели (набора искомых параметров), V – общее количество рассмотренных моделей, G – нормировочный множитель, обеспечивающий $\mu = 1$ для модели с наименьшим значением $\chi_{\text{эксп}}^2$ ($\chi_{\text{эксп min}}^2$), и

$$\mu_v^{II} = \exp\left(-0.5(\chi_v^2 - \chi_{\text{эксп min}}^2)^2\right). \quad (3.21)$$

Доверительные области $R_{\theta,v}$ для всех моделей представляли наборами параметров θ_{vn} , $n = 1, 2, \dots, N$, где N – число наборов параметров. Каждый набор

θ_{vn} включал значение t_Q , отвечающее v -й модели, и z значений $\lg \beta_i^{(z)}$. Последние были равномерно распределены внутри эллипсоида, границы которого устанавливает соотношение (3.22). Для компонентов множества θ_{vn} функции принадлежности задавали как

$$\mu_{vn, \lg \beta_i^{(z)}} = \begin{cases} \frac{1}{N} \mu_v, & \text{если } \beta_i^{(z)} \text{ определен,} \\ 0, & \text{если } \beta_i^{(z)} \text{ избыточный,} \end{cases} \quad (3.22)$$

$$\mu_{vn, t_Q} = \frac{1}{N} \mu_v.$$

Конечные оценки θ^{final} находили, используя метод центроидной дефазификации [227]:

$$\theta_y^{\text{final}} = \frac{\sum_{v,n} \theta_{vn,y} \cdot \mu_{vn,y}}{\sum_{v,n} \mu_{vn,y}}, \quad y = 1, 2, \dots, Y, \quad Y = z + 1. \quad (3.23)$$

Предлагаемая процедура отдельного расчета $\lg \beta_i^{(z)}$ и t_Q не меняет возможную высокую коррелированность параметров, но обеспечивает численную устойчивость вычислений, а объединение в θ^{final} характеристик всех найденных адекватных моделей делает окончательные результаты параметрической идентификации более объективными.

Алгоритм реализован в качестве процедуры в MatLab.

3.5. Испытание разработанного метода при исследовании модельной системы

Работоспособность предложенного алгоритма проверили при исследовании равновесий в водном растворе глицина. Выполнили эксперимент по рН-метрическому титрованию при $(20 \pm 2)^\circ\text{C}$ 20 мл раствора с начальными концентрациями глицина (HL) $4.455 \cdot 10^{-2}$ моль·л⁻¹ и HCl $5.80 \cdot 10^{-3}$ моль·л⁻¹ (ионная сила раствора 0.006 моль·л⁻¹) раствором NaOH с концентрацией 0.0900 моль·л⁻¹ (концентрация определена по бифталату калия). Принимали, что стандартные отклонения рН составляют 0.01. Кривая титрования приведена на рис. 3.2. По

результатам титрования находили МНК- и МНМ-оценки логарифмов констант равновесия β_1 ($L^- + H^+ = HL$) и β_2 ($HL + H^+ = H_2L^+$) (табл. 3.2). Модели адекватны эксперименту: для МНК-оценивания $\chi_{\text{эксп}}^2 = 6.8 < \chi_{23, 0.05}^2 = 35.2$; для МНМ-оценивания $\chi_{\text{эксп}}^2 = 15.5 < \chi_{3, 0.05}^2 = 35.2$.

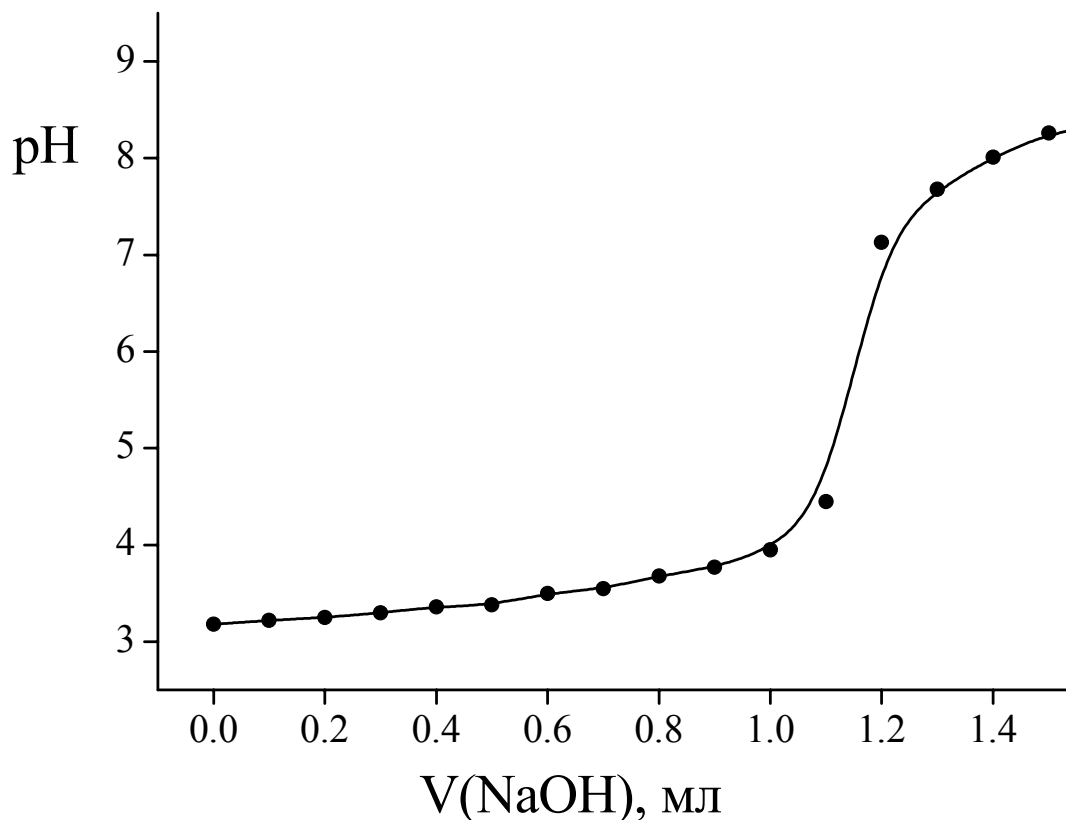


Рис. 3.2. Кривая титрования подкисленного раствора глицина.

При испытании предложенного алгоритма считали, что начальная концентрация HCl ($c(\text{HCl})$) подлежит определению наряду с $\lg \beta_1$ и $\lg \beta_2$. Сначала нашли интервал возможных $c(\text{HCl})$ (для значений $c(\text{HCl})$ в этом интервале удастся построить адекватные модели, рис. 3.3). Для каждого испытанного значения $c(\text{HCl})$ находили, варьируя гипотезу об интенсивности грубых промахов δ , М-оценки $\lg \beta_1$ и $\lg \beta_2$ и их ковариационную матрицу. От набора оценок параметров, их ковариационных матриц и соответствующих $\chi_{\text{эксп}}^2$, переходили к единственной конечной оценке $\lg \beta_1$ и $\lg \beta_2$.

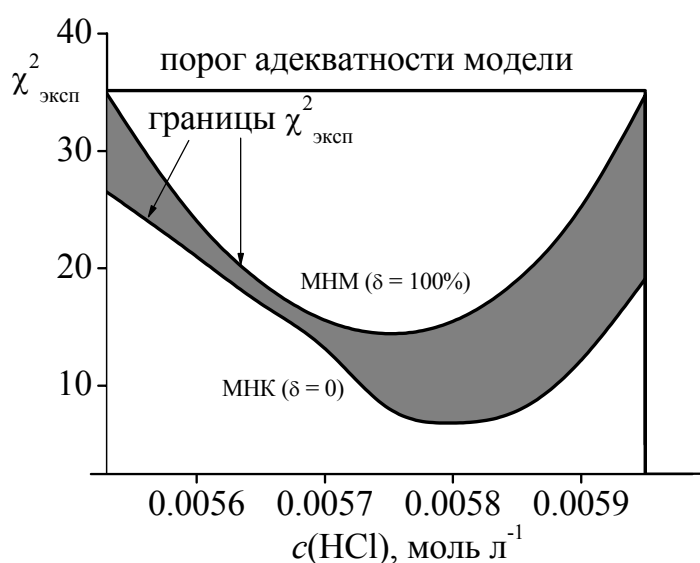


Рис. 3.3. Зависимость $\chi^2_{\text{эксп}}$ от $c(\text{HCl})$.

Полученные оценки $\lg \beta_1$ и $\lg \beta_2$ (табл. 3.2) практически совпадают для обоих видов задаваемых функций принадлежности. На рис. 3.4 представлена зависимость $\lg \beta_1$ и $\lg \beta_2$ от $c(\text{HCl})$ и обозначены конечные оценки искомых параметров. Важно, что полученные оценки в пределах погрешности определения совпадают с МНК- и МНМ-оценками, и близки к значениям, приводимым в [228] для 20°C и ионной силы раствора < 0.01 моль·л $^{-1}$.

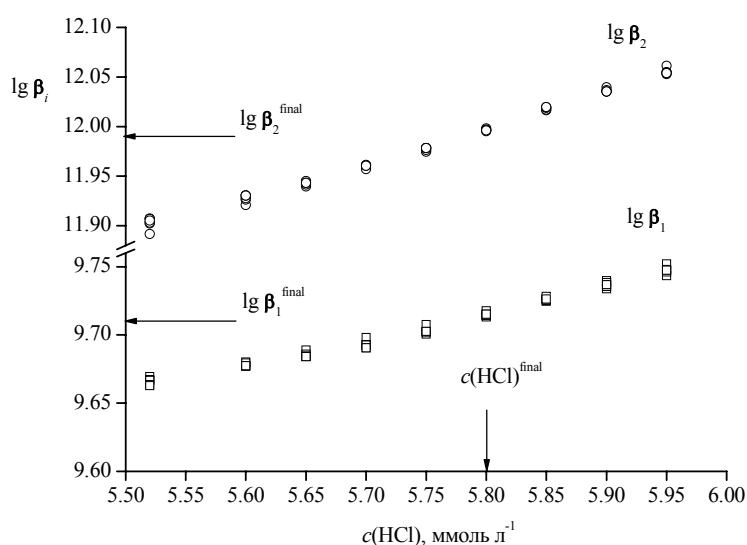


Рис. 3.4. Зависимость оценок $\lg \beta_1$ (□) и $\lg \beta_2$ (○) от оценки $c(\text{HCl})$.

**Сопоставление оценок параметров равновесия,
найденных различными методами ***

	МНК- оценки	МНМ- оценки	Объединенные оценки		Литературные данные [228]
			Формула для функции принадлежности		
			(3.20)	(3.21)	
$\lg \beta_1$	9.68 (0.01)	9.66 (0.01)	9.71 (0.01)	9.71 (0.01)	9.61
$\lg \beta_2$	11.92 (0.01)	11.93 (0.01)	11.99 (0.01)	11.99 (0.01)	11.96
$c(\text{HCl}),$ моль / л	не уточнялось	не уточнялось	0.0058	0.0058	—

* Здесь и далее в скобках приведены стандартные отклонения параметров.

3.6. Определение химико-аналитических характеристик материалов на основе силсесквиоксан хлоридных полимеров

Подход, описанный в п. 3.4, использовали для определения химико-аналитических характеристик двух новых органо-кремнеземных материалов – силикагеля, покрытого 3-*n*-пропилпиридиний силсесквиоксан хлоридным полимером (SiPy^+Cl^-) [214] и силикагеля, покрытого 3-*n*-пропил(4-метилпиридиний) силсесквиоксан хлоридным полимером ($\text{SiPic}^+\text{Cl}^-$) [215]. Эти материалы могут применяться для извлечения солей металлов из растворов.

3.6.1. Моделирование адсорбционных свойств материала SiPy^+Cl^- . На рис. 3.5 приведена структура 3-*n*-пропилпиридиний силсесквиоксан хлоридного

полимера, условия его синтеза и процедура доказательства строения описаны в работе [214].

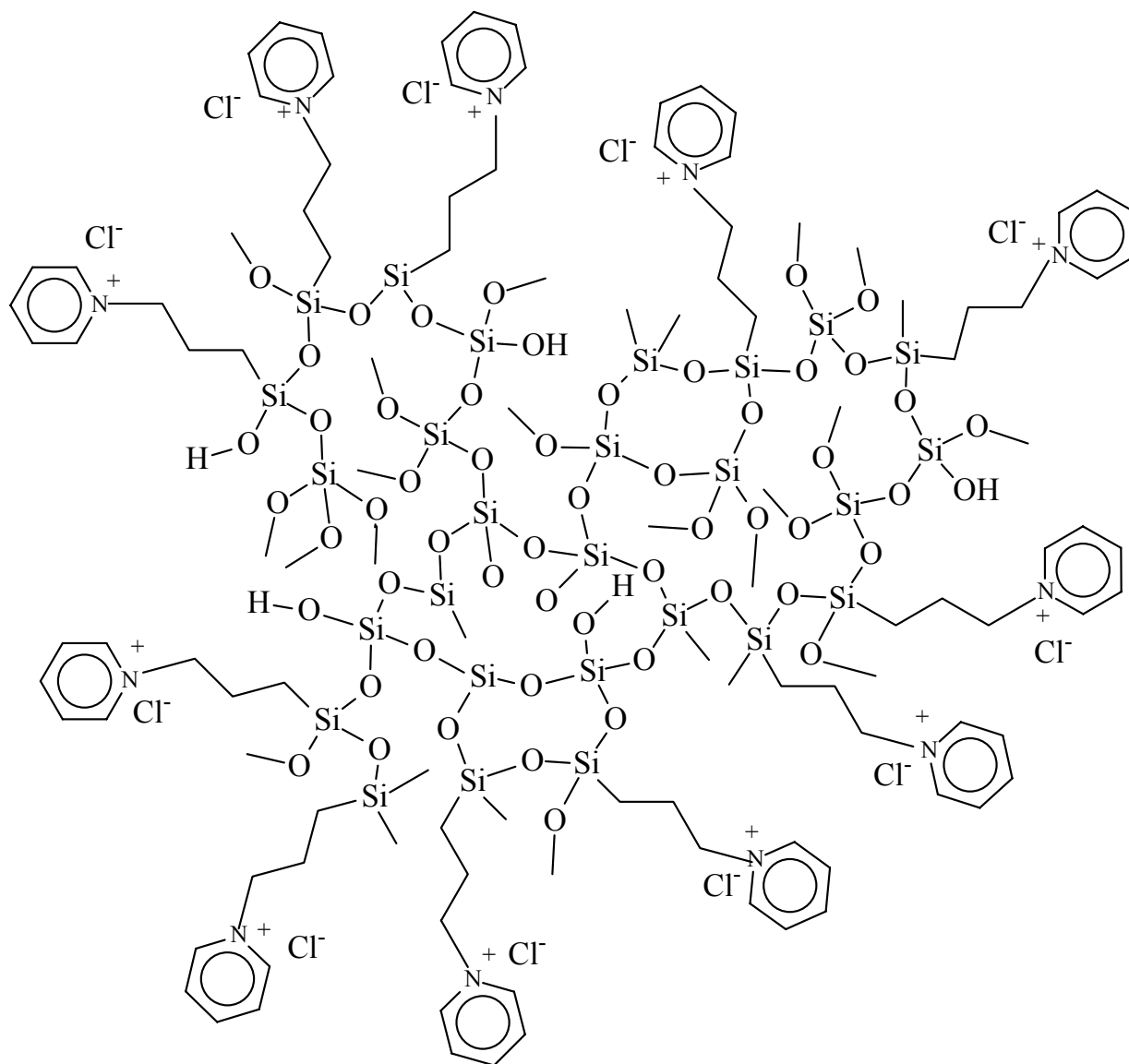


Рис. 3.5. Структура материала SiPy^+Cl^- .

В работе [214] приведены данные об адсорбции ZnCl_2 , CdCl_2 , HgCl_2 , CuCl_2 и FeCl_3 из этанольных растворов на поверхности материала SiPy^+Cl^- (табл. 3.3). Во всех экспериментах объемы растворов солей металлов составляли 50.0 мл.

Изотермы адсорбции MCl_2 на $SiPy^+Cl^-$

№	$m, 10^{-2}$	$n_i, 10^{-5}$	$n^e, 10^{-5}$	$[MCl_x], 10^{-4}$	$A, 10^{-4}$
1	2	3	4	5	6
$M = Zn, x = 2$					
1	5.16	1.50	1.42	2.48	0.151
2	5.06	2.40	2.30	4.60	0.201
3	5.03	3.00	2.88	5.77	0.231
4	5.10	3.30	3.17	6.34	0.251
5	5.01	3.75	3.59	7.19	0.311
6	5.13	4.00	3.80	7.60	0.390
7	5.10	4.50	4.29	8.60	0.421
8	5.04	5.10	4.88	9.77	0.430
9	5.04	6.00	5.76	11.5	0.480
10	5.05	7.50	7.19	14.4	0.612
11	5.13	8.50	8.09	16.2	0.801
12	5.03	9.50	9.00	18.0	1.00
13	5.09	10.5	9.89	19.8	1.20
$M = Cd, x = 2$					
1	5.17	1.50	1.42	2.84	0.151
2	5.25	2.40	2.30	4.59	0.200
3	5.07	3.00	2.88	5.76	0.241
4	5.11	3.30	3.17	6.33	0.260
5	5.21	3.75	3.58	7.17	0.320
6	5.17	4.00	3.81	7.62	0.371
7	5.18	4.50	4.27	8.53	0.452
8	5.11	5.10	4.85	9.71	0.481
9	5.12	6.00	5.74	11.5	0.512
10	5.23	7.50	7.03	14.1	0.900

Продолж. табл. 3.3.

<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
11	5.17	8.50	7.88	15.8	1.20
12	5.27	9.50	8.71	17.4	1.50
13	5.28	10.5	9.34	18.7	2.20
$M = Hg, x = 2$					
1	4.95	1.50	1.35	2.70	0.301
2	5.12	2.40	2.14	4.29	0.500
3	4.97	3.00	2.65	5.30	0.700
4	4.99	3.30	2.87	5.73	0.872
5	5.07	3.75	3.29	6.59	0.901
6	4.94	4.00	3.53	7.06	0.951
7	5.03	4.50	4.00	8.00	0.994
8	5.10	5.10	4.54	9.08	1.10
9	5.09	7.50	6.63	13.3	1.70
10	5.11	8.50	7.22	14.4	2.50
11	4.97	9.50	7.81	15.6	3.40
12	4.98	10.5	8.15	16.3	4.72
$M = Cu, x = 2$					
1	5.15	1.50	1.30	2.59	0.396
2	5.04	2.40	2.14	4.28	0.514
3	5.08	3.00	2.70	5.40	0.590
4	5.15	3.30	2.96	5.91	0.668
5	5.15	3.75	3.34	6.69	0.788
6	5.05	4.00	3.60	7.21	0.792
7	5.25	4.50	4.07	8.13	0.825
8	5.05	5.10	4.66	9.31	0.877
9	5.17	6.00	5.53	11.1	0.911

Продолж. табл. 3.3.

<i>l</i>	2	3	4	5	6
10	5.26	7.50	6.98	14.0	0.996
11	5.26	8.50	7.92	15.8	1.10
12	5.31	9.50	8.85	17.7	1.23
13	5.36	10.5	9.79	20.0	1.33
M = Fe, x = 3					
1	5.14	1.50	0.873	1.74	1.22
2	5.00	1.80	1.13	2.25	1.35
3	5.00	2.40	1.50	2.99	1.81
4	5.28	3.00	1.86	3.73	2.15
5	5.34	3.30	2.02	4.04	2.39
6	5.43	3.75	2.33	4.67	2.61
7	5.29	4.00	2.45	4.90	2.93
8	5.41	4.50	2.82	5.65	3.10
9	5.37	5.10	3.37	6.74	3.22
10	5.32	6.00	4.08	8.17	3.60
11	5.17	7.50	5.12	10.2	4.60
12	5.11	8.50	5.54	11.1	5.80
13	5.30	9.50	6.21	12.4	6.20
14	5.35	10.5	6.97	13.9	6.60

Обозначения: здесь и далее m – навеска сорбента, г; n_i – начальное количество вещества MCl_x в растворе, моль; n^e – количество вещества MCl_x в растворе после достижения равновесия, моль; $[MCl_x] = \frac{n^e}{50 \cdot 10^{-3}}$ – равновесная концентрация MCl_x , моль·л⁻¹, $A = \frac{n_i - n^e}{m}$ – адсорбция, моль·г⁻¹.

Рассмотрим построение модели на примере исследования равновесий адсорбции $CdCl_2$.

При попытке воспользоваться моделью идеальной сорбции были получены такие результаты (рис. 3.6):

$$\frac{1}{D} = 27 - 8460[\text{CdCl}_2], \text{ коэффициент корреляции } r = 0.61. \quad (3.25)$$

Очевидно, что модель идеальной сорбции неприменима.

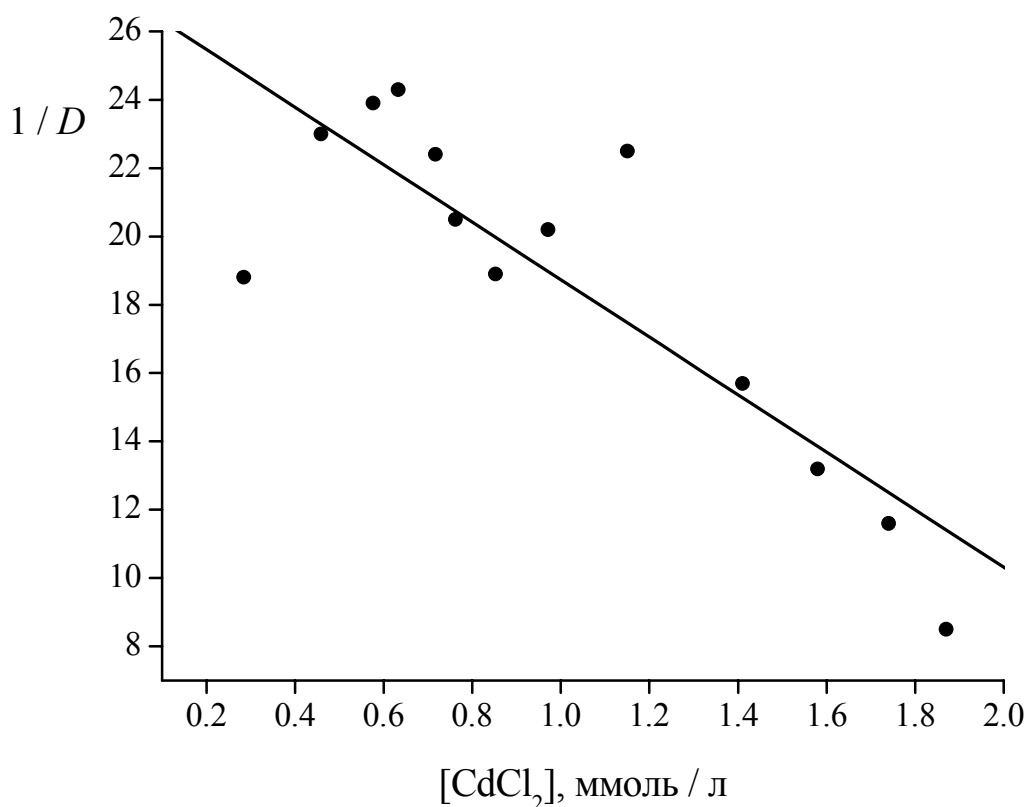


Рис. 3.6. Испытание модели идеальной адсорбции для описания равновесия адсорбции CdCl_2 на поверхности SiPy^+Cl^- .

Для учета неидеального характера адсорбции воспользовались моделью полидентатного связывания. Построение модели начинали с $Z = 2$ (модель бидентатного связывания). Для всех возможных значений эффективной сорбционной емкости подобрать адекватную модель невозможно (рис. 3.7).

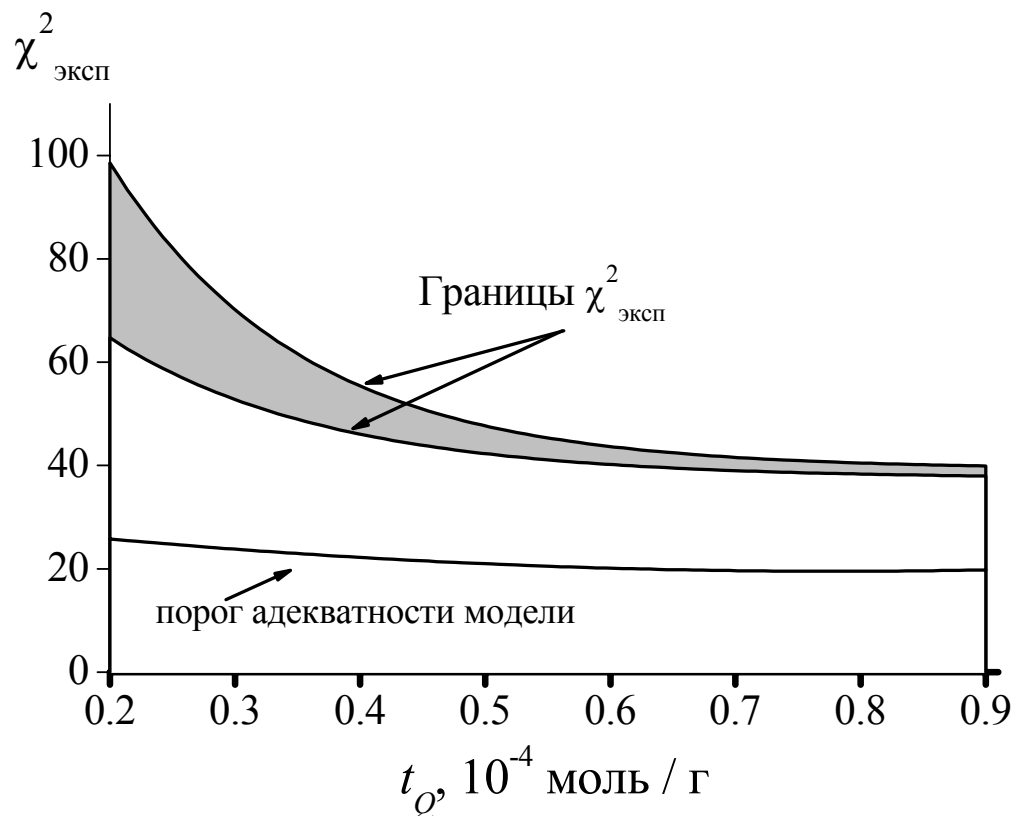


Рис. 3.7. Зависимость $\chi^2_{\text{эксп}}$ от t_Q в рамках модели бидентатного связывания.

Модель тридентатного связывания также оказалась неадекватной. Адекватную модель удалось построить лишь в рамках модели тетрадентатного связывания (рис. 3.1, табл. 3.4, рис. 3.8-3.12).

Частицы в правой части уравнений рис. 3.1 стехиометрически неразличимы, т.е. равновесные концентрации продуктов $\overline{Q_4MCl_2}$, $\overline{Q_4(MCl_2)_2}$ и $\overline{Q_4(MCl_2)_3}$ представляют собой суммы концентраций соответствующих частиц. При этом из четырех химических форм, существование которых допускает модель тетрадентатного связывания, для описания адсорбции $CdCl_2$ формы $\overline{Q_4(CdCl_2)_2}$ и $\overline{Q_4(CdCl_2)_3}$ оказались избыточными (по данным эксперимента с приемлемой точностью можно определить лишь $\lg \beta_1^{(4)}$ и $\lg \beta_4^{(4)}$).

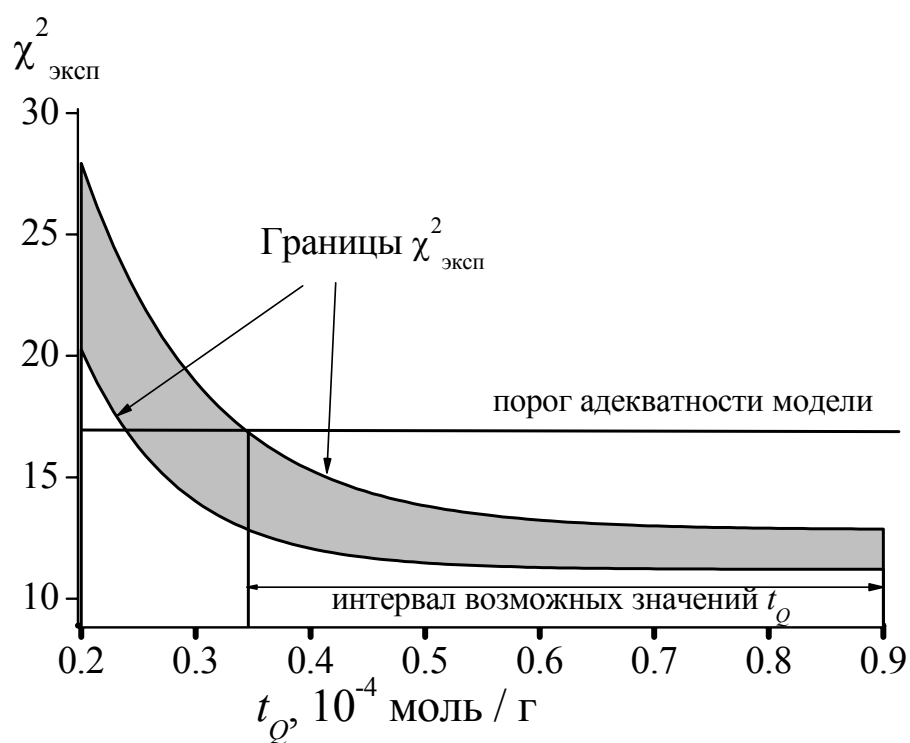


Рис. 3.8. Зависимость $\chi^2_{\text{эксп}}$ от t_Q (модель тетраденатного связывания).

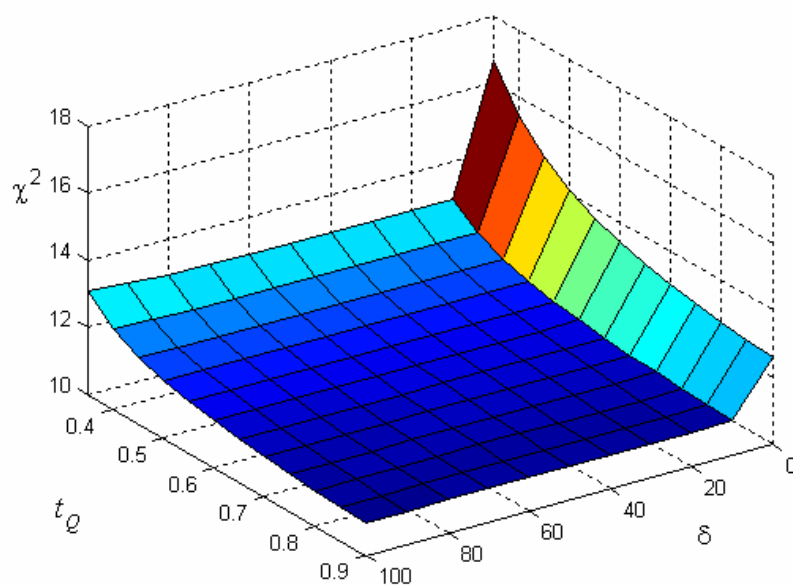


Рис. 3.9. Зависимость $\chi^2_{\text{эксп}}$ от t_Q , ммоль/Г, и δ (модель тетраденатного связывания).

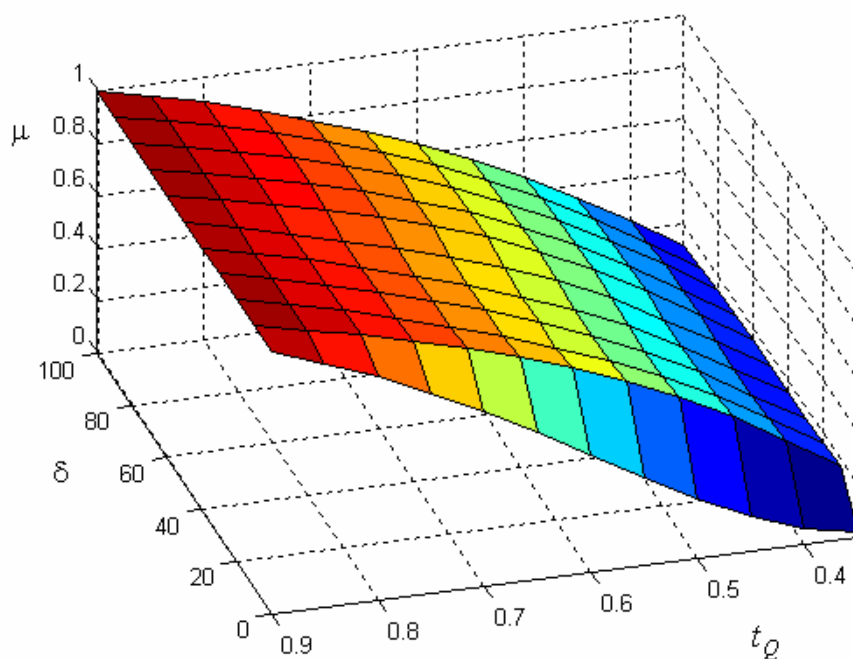


Рис. 3.10. Изменение функции принадлежности (3.20) в зависимости от t_Q , ммоль/г, и δ (модель тетрадентатного связывания).

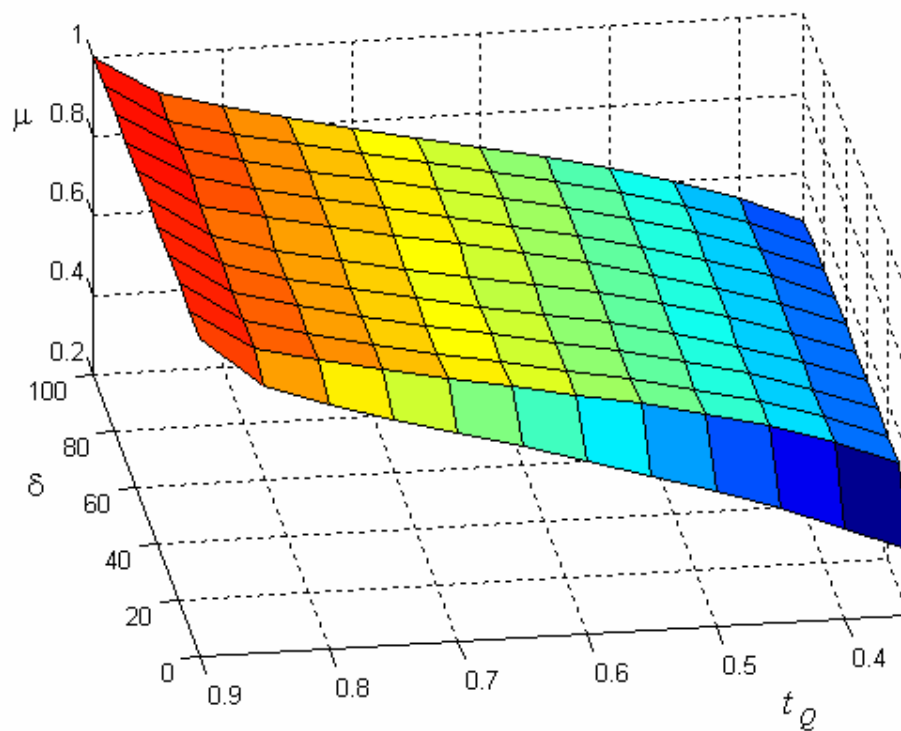


Рис. 3.11. Изменение функции принадлежности (3.21) в зависимости от t_Q , ммоль/г, и δ (модель тетрадентатного связывания).

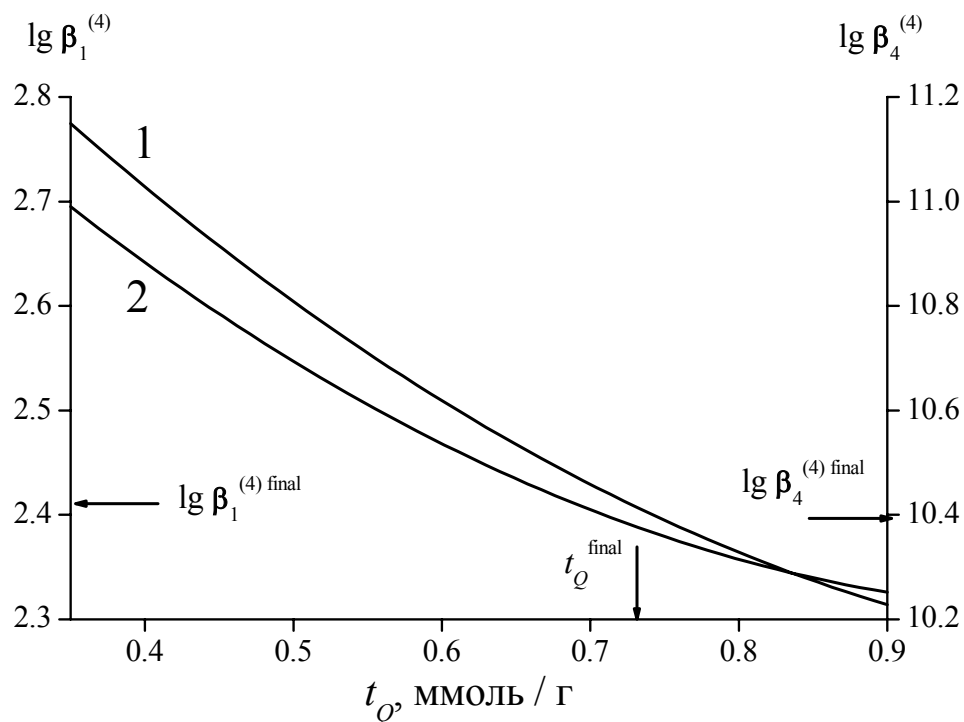


Рис. 3.12. Зависимость оценок $\lg \beta_1^{(4)}$ и $\lg \beta_4^{(4)}$ от оценки t_Q .

Найдя конечные параметры модели, рассчитывали степени образования комплексов $\overline{Q_Z(\text{CdCl}_2)_i}$ как

$$\alpha_i = \frac{\overline{Q_Z(\text{CdCl}_2)_i}}{\sum_{i=0}^Z \overline{Q_Z(\text{CdCl}_2)_i}} \cdot 100\%, \quad (3.26)$$

где i – номер ступени комплексообразования.

На рис. 3.13 и 3.14 показаны степени образования комплексов $\overline{Q_4(\text{CdCl}_2)_i}$ и экспериментальные и рассчитанные значения адсорбции.

На рис. 3.15 сопоставлены экспериментальные и модельные значения адсорбции для остальных хлоридов металлов, а на рис. 3.16–3.19 приведены степени образования комплексов соответствующих металлов.

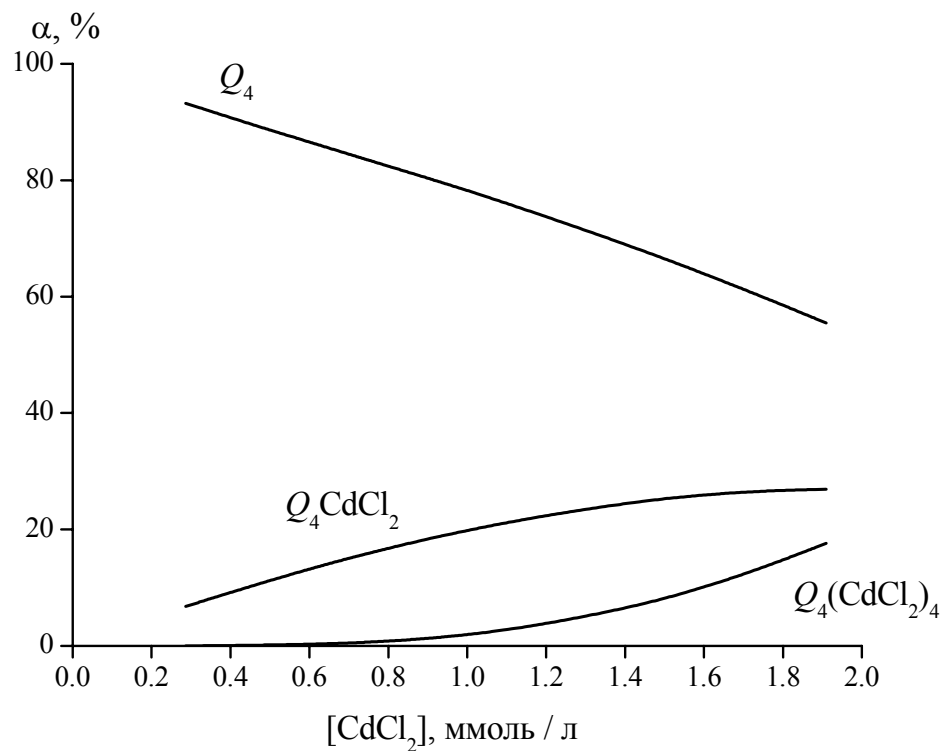


Рис. 3.13. Степени образования химических форм в системе раствор $\text{CdCl}_2 - \text{SiPy}^+\text{Cl}^-$.

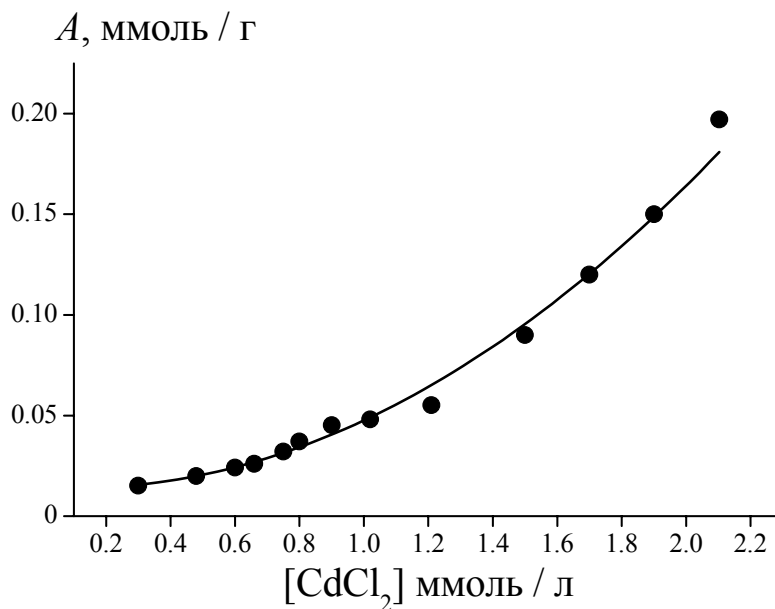


Рис. 3.14. Экспериментальная (•) и рассчитанная (линия) изотермы сорбции CdCl_2 материалом SiPy^+Cl^- .

Рассчитали также зависимость степени извлечения (3.27) хлоридов каждого из металлов из 50 мл раствора в зависимости от общей концентрации соли и навески сорбента (рис. 3.15).

$$R = \frac{t(\text{MCl}_x) - [\text{MCl}_x]}{t(\text{MCl}_x)}, \quad (3.27)$$

где $t(\text{MCl}_x)$ – общая концентрация хлорида металла, $[\text{MCl}_x]$ – равновесная концентрация соли.

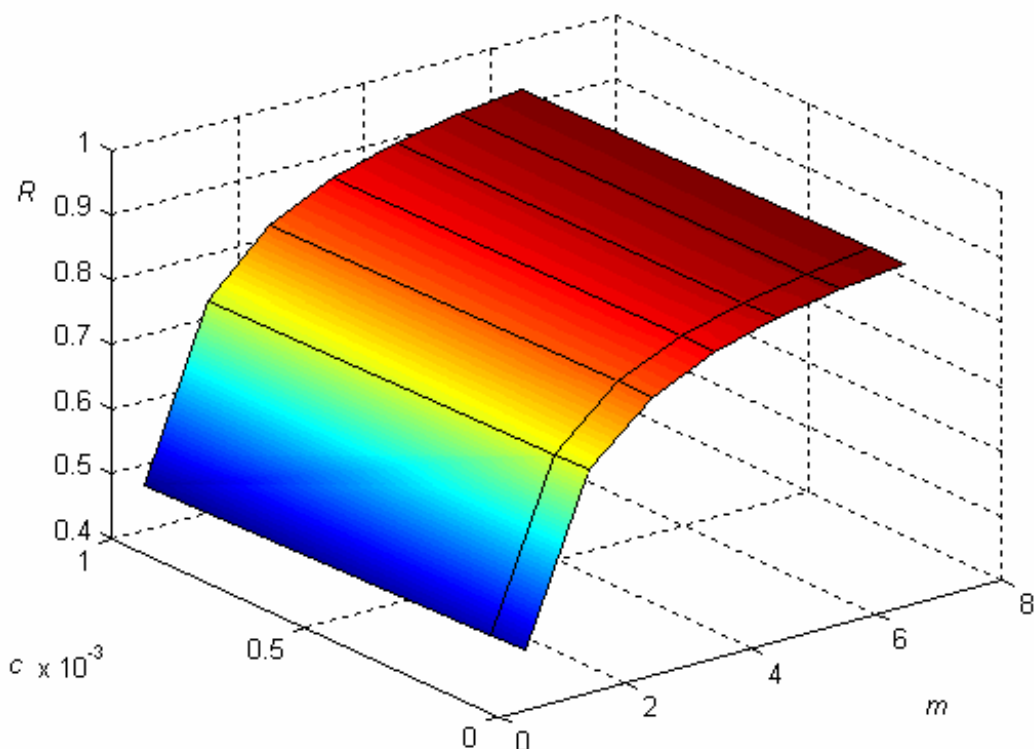


Рис. 3.15. Зависимость степени извлечения CdCl_2 из этанольного раствора от массы сорбента и начальной концентрации соли.

В табл. 3.5 приведены массы навесок сорбентов, необходимые для 99%-го извлечения хлоридов металлов из 50 мл этанольных растворов при использовании SiPy^+Cl^- в качестве сорбента.

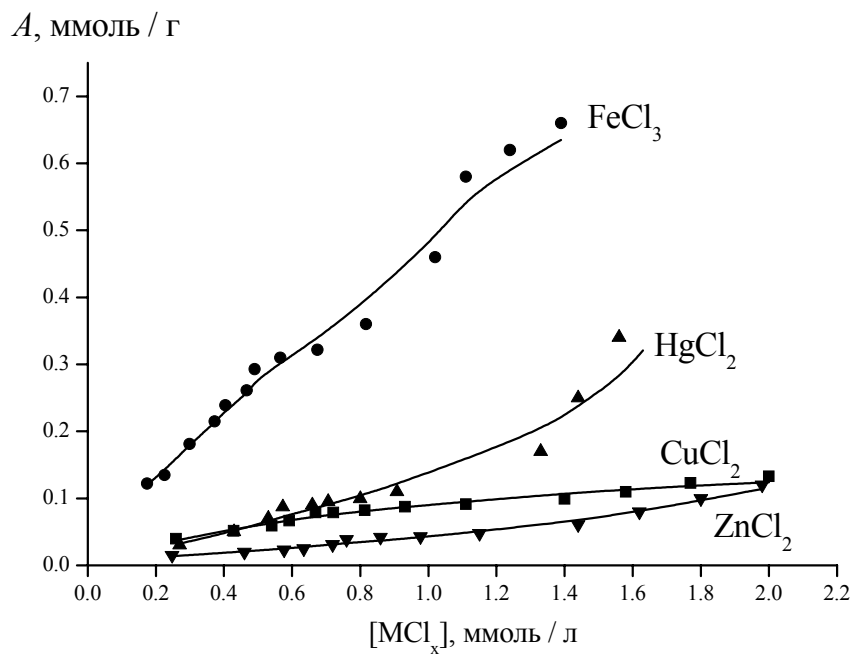


Рис. 3.16. Экспериментальные (точки) и модельные (линии) изотермы адсорбции MCl_x на $SiPy^+Cl^-$.

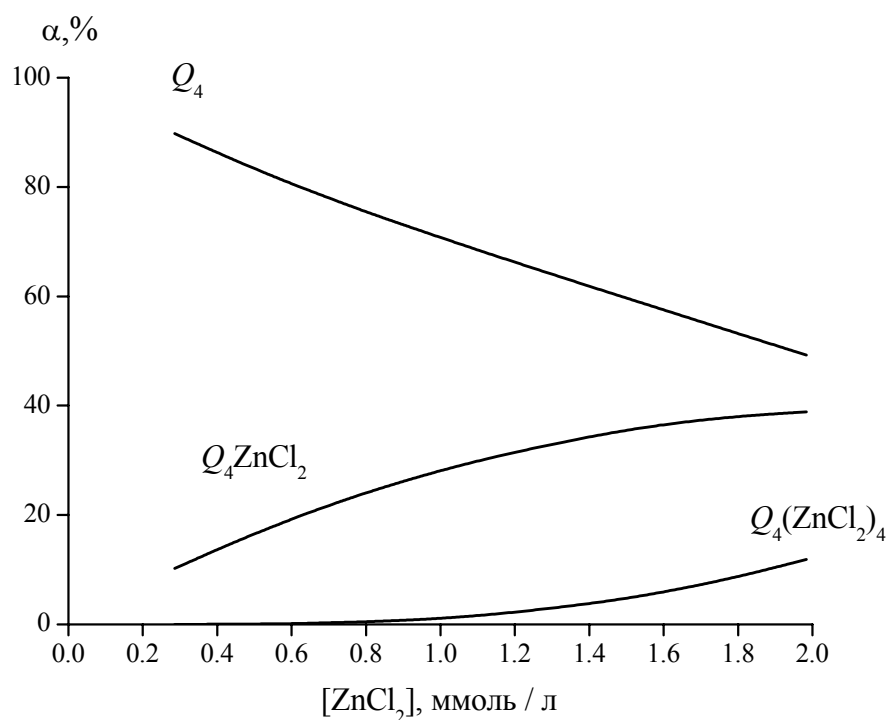


Рис. 3.17. Степени образования химических форм в системе раствор $ZnCl_2-SiPy^+Cl^-$.

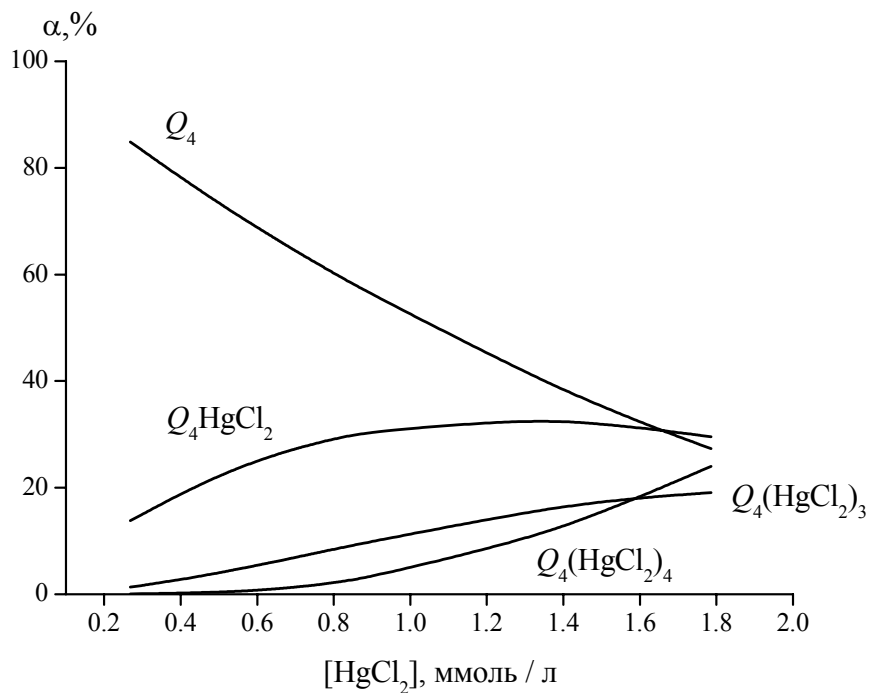


Рис. 3.18. Степени образования химических форм в системе раствор $\text{HgCl}_2\text{-SiPy}^+\text{Cl}^-$.

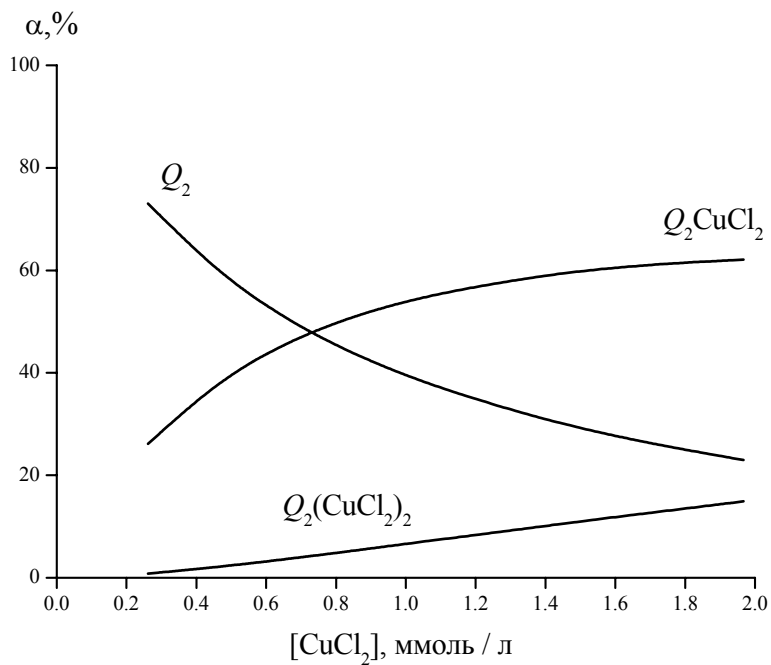


Рис. 3.19. Степени образования химических форм в системе раствор $\text{CuCl}_2\text{-SiPy}^+\text{Cl}^-$.

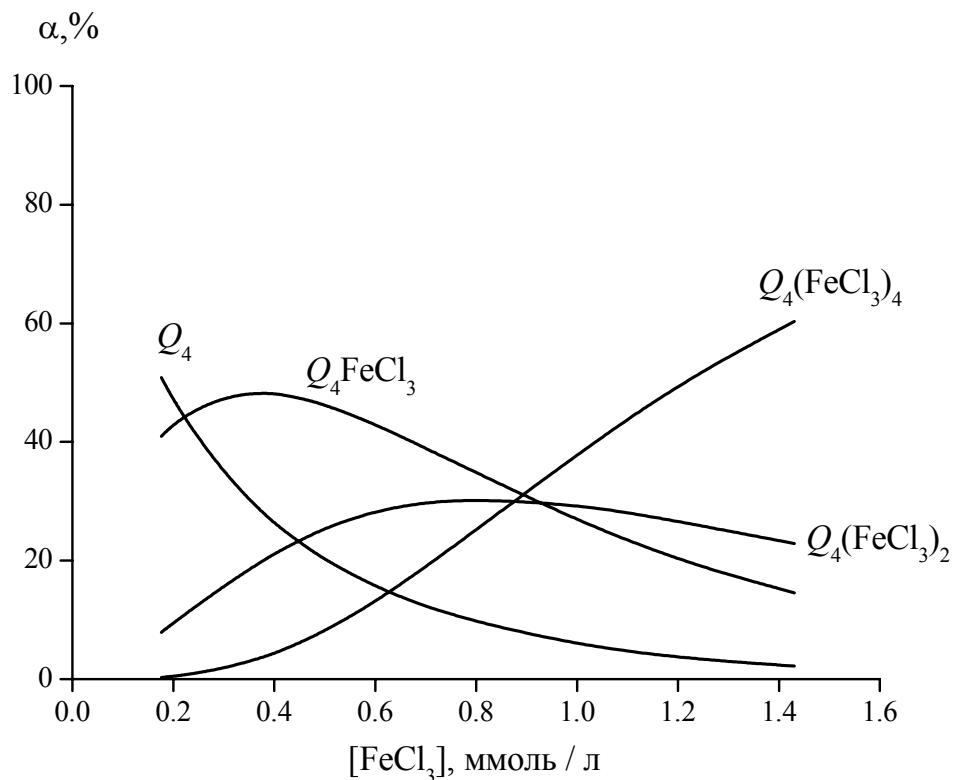


Рис. 3.20. Степени образования химических форм в системе раствор $\text{FeCl}_3\text{--SiPy}^+\text{Cl}^-$.

Таблица 3.4.

Характеристики материала SiPy^+Cl^-

	Сорбат	ZnCl_2	CdCl_2	HgCl_2	CuCl_2	FeCl_3
Параметры модели	$\lg \beta_1^{(4)}$	2.65 (0.03)	2.42 (0.04)	2.8 (0.1)	3.14** (0.03)	3.7 (0.1)
	$\lg \beta_2^{(4)}$	*	*	5.4 (0.3)	5.20** (0.05)	6.7 (0.1)
	$\lg \beta_4^{(4)}$	10.23 (0.05)	10.40 (0.06)	11.0 (0.2)	—	12.8 (0.1)
	t_{Q_2} , ммоль / г	0.53	0.74	0.77	0.27	0.83

Избыточными параметрами для всех моделей тетрадентатного связывания являются $\lg \beta_3^{(4)}$, для моделей, помеченных *, избыточными параметрами являются также $\lg \beta_2^{(4)}$.

** Параметры модели бидентатного связывания.

Анализ табл. 3.4 позволяет сделать вывод о проявлении эффектов отрицательной кооперативности при адсорбции хлорида меди ($\lg K_1 > \lg K_2$). При адсорбции остальных хлоридов наблюдается положительная кооперативность из-за отсутствия одного (для FeCl_3 и HgCl_2) или двух (в случае ZnCl_2 и CdCl_2) промежуточных комплексов. Другими словами, присоединение одной или двух частиц хлоридов металлов к адсорбционному центру \overline{Q}_4 облегчает образование следующих комплексов.

Таблица 3.5.

Массы навесок m сорбента SiPy^+Cl^- , необходимые для 95%-го извлечения MCl_x с общей концентрацией t

Соль	t , моль/л	m , г
ZnCl_2	$1 \cdot 10^{-3}$	14.8
	$1 \cdot 10^{-4}$	15
	$1 \cdot 10^{-5}$	15.3
CdCl_2	$1 \cdot 10^{-3}$	17
	$1 \cdot 10^{-4}$	19
	$1 \cdot 10^{-5}$	20
HgCl_2	$1 \cdot 10^{-3}$	6.5
	$1 \cdot 10^{-4}$	7.1
	$1 \cdot 10^{-5}$	7.6
CuCl_2	$1 \cdot 10^{-3}$	5
	$1 \cdot 10^{-4}$	6
	$1 \cdot 10^{-5}$	6.5
FeCl_3	$1 \cdot 10^{-3}$	0.9
	$1 \cdot 10^{-4}$	1
	$1 \cdot 10^{-5}$	1.2

3.6.2. Моделирование адсорбционных свойств SiPic⁺Cl⁻. В работе [215] приведены изотермы адсорбции хлоридов металлов MCl₂ (M = Zn, Cd, Hg) из водных растворов на поверхности гибридного органо-минерального материала SiPic⁺Cl⁻. Условия синтеза полимера, его физические свойства и доказательство строения приведены там же.

Структура материала приведена на рис. 3.21. Теоретическая сорбционная емкость, оцененная как концентрация хлорид-ионов, удаляемых из фазы сорбента при обработке раствором AgNO₃, составляла 1.46±0.01 ммоль·г⁻¹ [215].

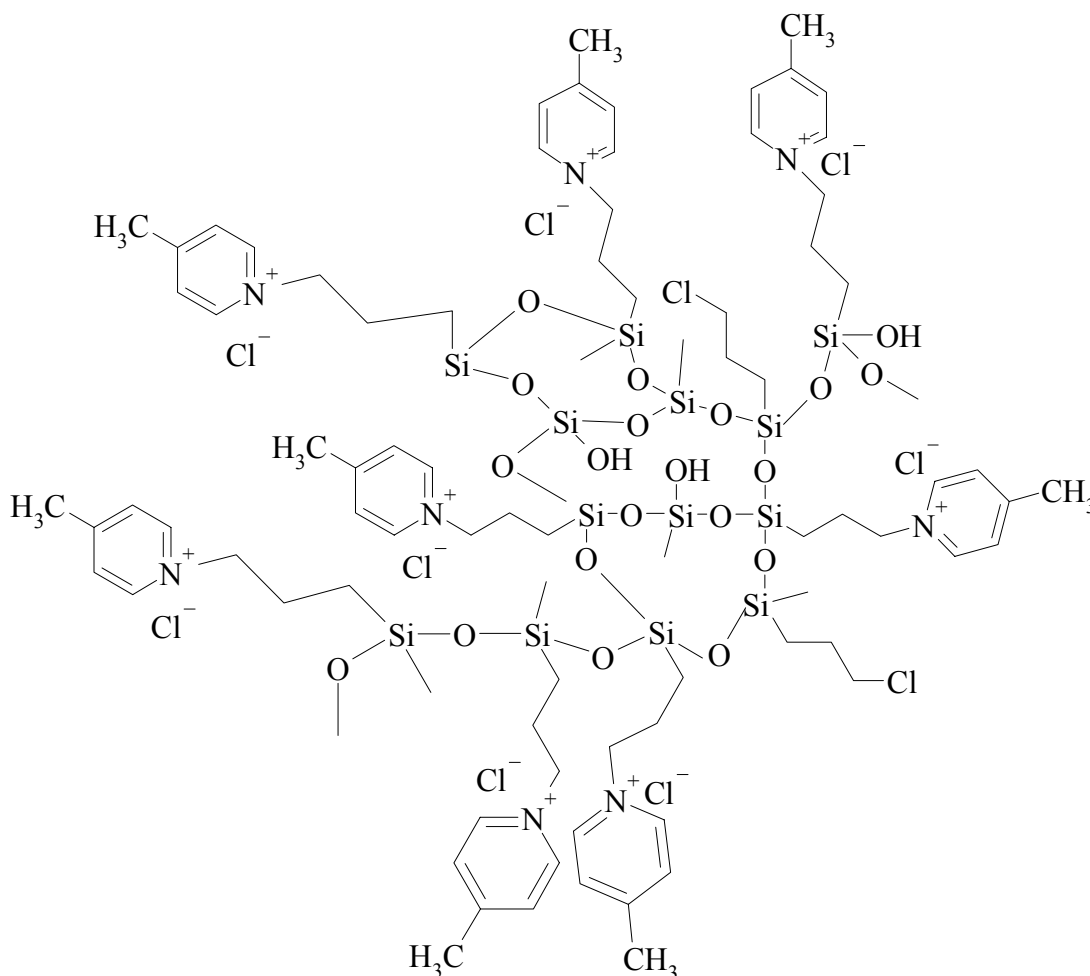


Рис. 3.21. Структура 3-*n*-пропил(4-метилпиридиний) силсесквиоксан хлоридного полимера.

Изотермы адсорбции приведены в табл. 3.6. Во всех экспериментах объемы растворов составляли 50.0 мл.

Изотермы адсорбции MCl_2 на $SiPic^+Cl^-$

№	$m \cdot 10^{-2}$	$n_i \cdot 10^{-5}$	$n^e \cdot 10^{-5}$	$[MCl_2] \cdot 10^{-4}$	$A \cdot 10^{-4}$
1	2	3	4	5	6
M = Zn					
1	5.01	1.39	0.51	1.03	1.74
2	5.02	2.85	1.73	3.46	2.22
3	5.00	4.23	2.95	5.90	2.56
4	5.00	5.62	4.36	8.72	2.51
5	4.99	7.00	5.77	1.15	2.47
6	5.02	7.69	6.41	1.28	2.55
7	5.04	8.46	7.05	1.41	2.80
8	5.01	9.23	8.08	1.62	2.30
M = Cd					
1	4.98	0.51	0.38	0.77	0.24
2	5.03	1.01	0.64	1.28	0.74
3	5.01	2.02	1.41	2.82	1.22
4	4.98	3.04	2.44	4.87	1.20
5	4.99	4.05	3.21	6.41	1.69
6	4.99	5.06	4.36	8.72	1.40
7	4.98	6.07	5.13	10.26	1.89
8	5.04	7.08	6.03	12.07	2.10
9	5.05	7.59	6.54	13.08	2.08
M = Hg					
1	5.04	0.50	0.38	0.77	0.23
2	5.02	0.10	0.38	0.77	1.23
3	5.04	0.20	0.38	0.77	3.22
4	5.03	0.40	0.96	1.92	6.08

1	2	3	4	5	6
5	5.01	0.60	2.50	5.00	7.04
6	5.04	0.80	3.85	7.69	8.31
7	5.04	1.00	5.96	11.92	8.09
8	4.99	1.21	7.50	15.00	9.11
9	4.99	1.51	9.42	18.85	11.30
10	5.01	1.71	11.35	22.70	11.42

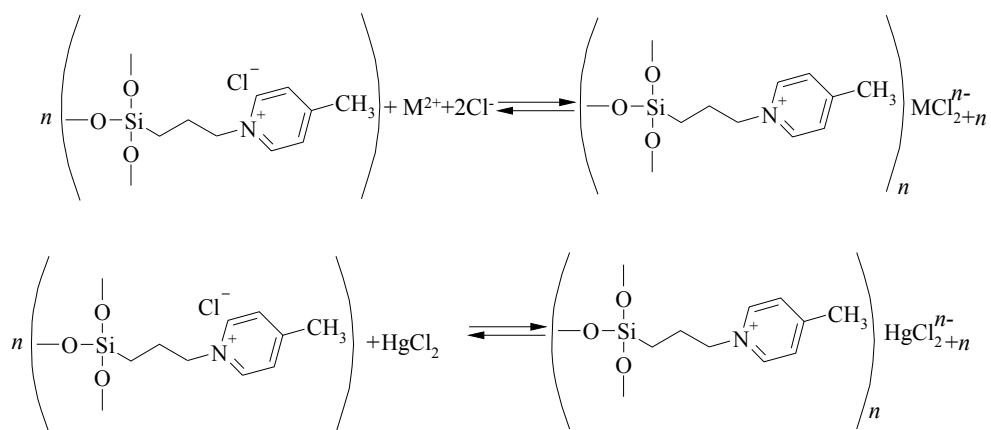
Обозначения: m – навеска сорбента, г; n_i – начальное количество вещества MCl_2 в растворе, моль; n^e – количество вещества MCl_2 в растворе после достижения

равновесия, моль; $[MCl_2] = \frac{n^e}{50 \cdot 10^{-3}}$ – равновесная концентрация MCl_2 , моль/л;

$$A = \frac{n_i - n^e}{m} \text{ – адсорбция, моль/г.}$$

Адекватное воспроизведение изотерм адсорбции было достигнуто с использованием модели бидентатного связывания. Так как поверхность материала должна оставаться нейтрально заряженной, фиксация солей металлов возможна только путем образования анионных комплексов состава MCl_{2+n}^{n-} , компенсирующих положительный заряд 3-*n*-пропил(4-метилпиридиний) силсеквиоксан-катионов (Pr^{3+}).

Типичное координационное число d^{10} -металлов равно четырем, следовательно, можно предположить образование анионных комплексов MCl_3^- и MCl_4^{2-} . Необходимо также учесть состояние хлоридов металлов в водном растворе: в то время как $ZnCl_2$ и $CdCl_2$ в водном растворе диссоциируют с преимущественным образованием гидратированных катионов $M(H_2O)_n^{2+}$, $HgCl_2$ практически не подвергается диссоциации. Таким образом, процесс адсорбции хлоридов металлов можно изобразить следующей схемой.



Отметим, что анионные комплексы состава MCl_4^{2-} фиксируются двумя закрепленными Pic^+ -катионами, в то время как MCl_3^- -анионы – только с одним (рис. 3.22, 3.23).

В табл. 3.7 представлены модели равновесий адсорбции MCl_2 из водных растворов на поверхности материала $\text{SiPic}^+\text{Cl}^-$. В табл. 3.8 приведены массы сорбента, необходимые для 90%-го извлечения солей металлов из водных растворов, а на рис. 3.24–3.27 – изотермы адсорбции и степени образования присутствующих в системах химических форм. Можно заключить, что на поверхности доминируют химические формы: $\left\{ \begin{array}{l} \sim \text{Pic}^+ \\ \sim \text{Pic}^+ \end{array} \right. [\text{MCl}_4]^{2-}$ для Zn(II) ,



Таблица 3.7.

Характеристики материала SiPic⁺Cl⁻

Параметры модели	Сорбат	ZnCl ₂	CdCl ₂	HgCl ₂
	$\lg \beta_1^{(2)}$	11.8 (0.2)	11.2 (0.1)	–
	$\lg \beta_2^{(2)}$	18.8 (0.4)	19.6 (0.2)	6.9 (0.3)
	t_Q , ммоль / г	0.48	0.24	0.96

Таблица 3.8.

**Массы навесок m сорбента SiPic⁺Cl⁻, необходимые для 95%-го извлечения
MCl₂ с общей концентрацией t**

Соль	t , моль/л	m , г
ZnCl ₂	$1 \cdot 10^{-3}$	7
	$1 \cdot 10^{-4}$	12
	$1 \cdot 10^{-5}$	18
CdCl ₂	$1 \cdot 10^{-3}$	8
	$1 \cdot 10^{-4}$	15
	$1 \cdot 10^{-5}$	17
HgCl ₂	$1 \cdot 10^{-3}$	7
	$1 \cdot 10^{-4}$	15
	$1 \cdot 10^{-5}$	21

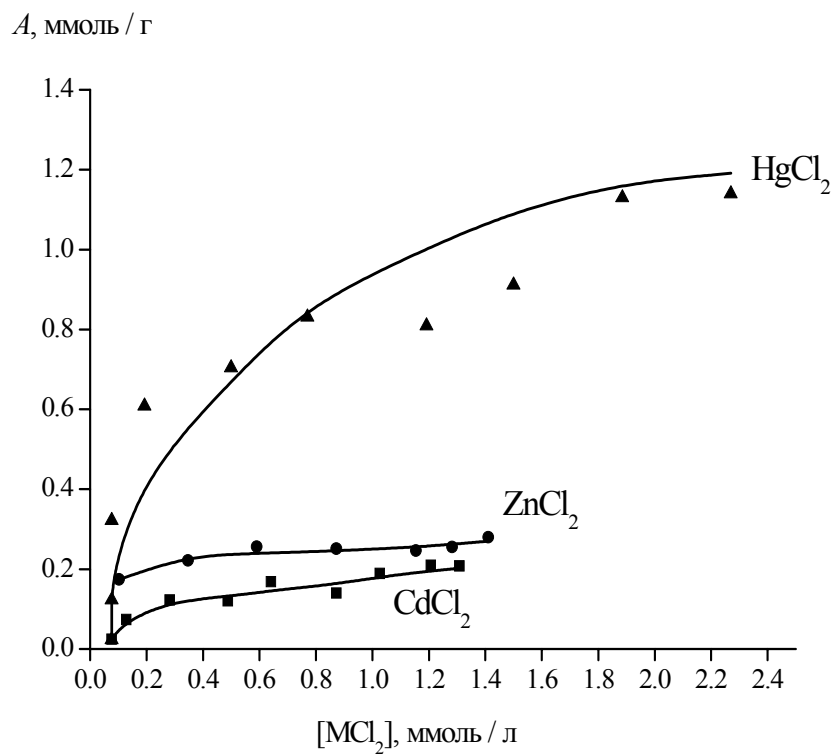


Рис. 3.24. Экспериментальные (точки) и модельные (линии) изотермы адсорбции MCl_2 .

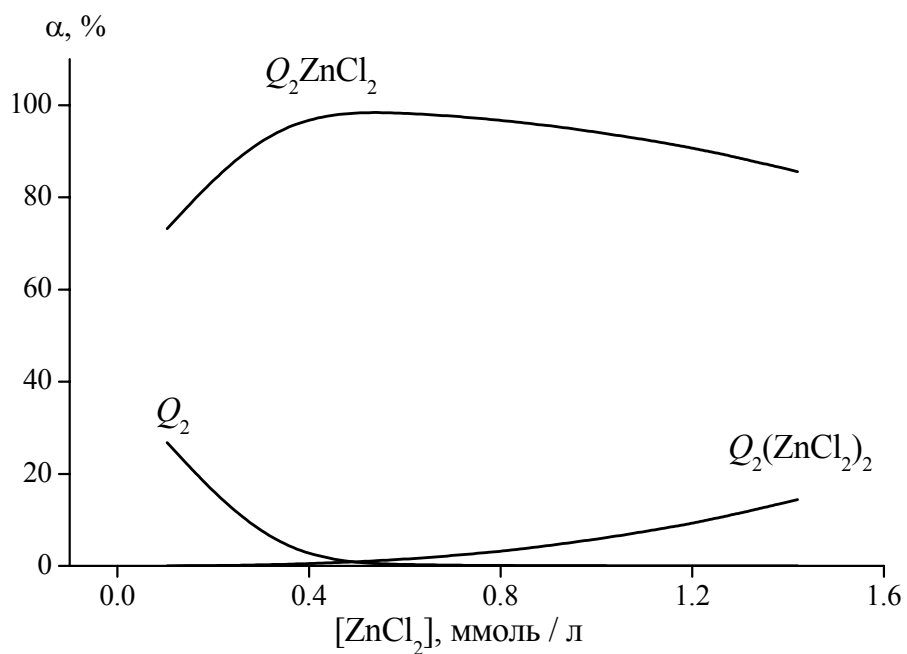


Рис. 3.25. Степени образования химических форм в системе раствор $ZnCl_2-SiPic^+Cl^-$.

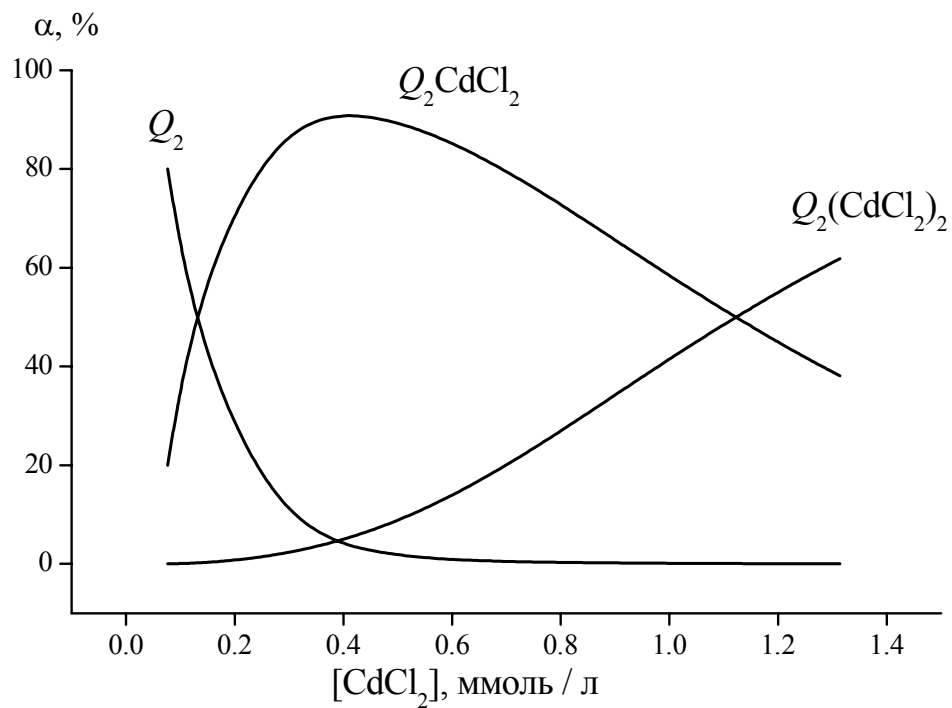


Рис. 3.26. Степени образования химических форм в системе раствор $\text{ZnCl}_2\text{-SiPic}^+\text{Cl}^-$.

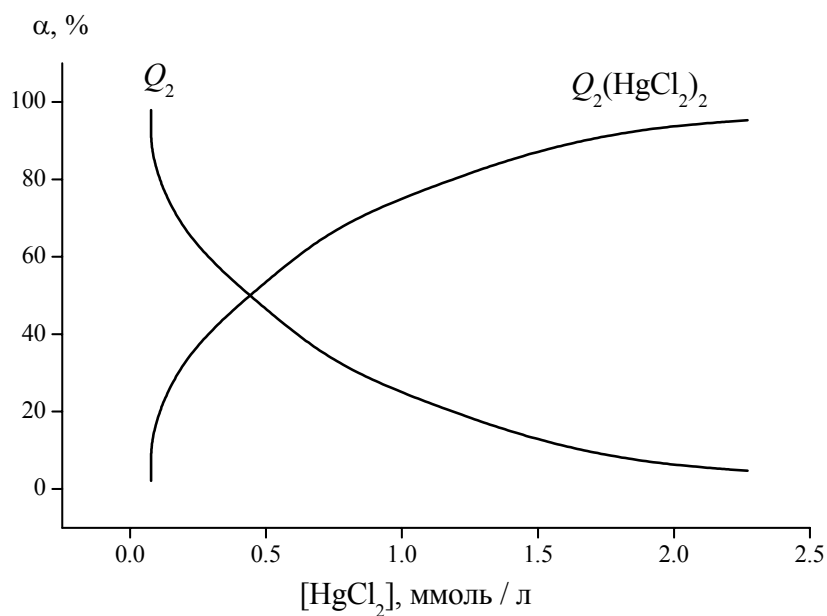


Рис. 3.27. Степени образования химических форм в системе раствор $\text{HgCl}_2\text{-SiPic}^+\text{Cl}^-$.

3.7. Выводы к разделу 3

1. Показана эффективность новой расчетной процедуры, объединяющей принципы робастного оценивания и методы теории нечетких множеств, для одновременного оценивания эффективной сорбционной емкости и констант сорбционных равновесий как ключевых химико-аналитических характеристик гибридных органо-кремнеземных материалов.

2. Моделирование протолитических равновесий в водных растворах глицина по данным рН-метрического титрования подтвердило возможность одновременного оценивания констант равновесия и начальных концентраций реагентов в задачах КФХА с помощью разработанного алгоритма.

3. В построенных моделях равновесий адсорбции хлоридов металлов новыми органо-кремнеземными материалами выявлены и охарактеризованы эффекты отрицательной и положительной кооперативности, определяющие неидеальный характер адсорбции.

4. Установлено, что новые органо-кремнеземные материалы могут использоваться для извлечения хлоридов металлов с полнотой 90-99% из водных и этанольных растворов.

ВЫВОДЫ

В работе решена актуальная задача развития хемометрических методов обработки данных, что позволяет повысить достоверность результатов качественного химического анализа, и обеспечивает надежное оценивание химико-аналитических характеристик твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов.

1. Показано, что важнейшими метрологическими характеристиками методик обнаружения с бинарным откликом следует считать находимые по кривым эффективности предел обнаружения аналита и интервал ненадежности; для представления кривых эффективности достаточно ограничиться лишь двумя функциями – логистического и экспоненциального распределений.
2. На основе теории нечетких множеств создан робастный алгоритм оценивания параметров кривых эффективности по эмпирическим зависимостям частот обнаружения от концентрации аналитов.
3. С использованием упомянутого алгоритма для обработки 27 массивов данных о зависимостях частот обнаружения аналитов от их концентраций установлено, что робастные оценки параметров кривых эффективности и метрологические характеристики методик обнаружения практически не отличаются от оценок, полученных на основе применения нелинейного взвешенного МНК.
4. Способы назначения статистических весов по моделям погрешностей, предполагающим равноточность частот обнаружения или же их биномиальное распределение, приводят к близким оценкам подгоночных параметров. Таким образом, обоснована возможность применения невзвешенного нелинейного МНК для оценивания метрологических характеристик методик обнаружения с бинарным откликом.
5. Разработан алгоритм идентификации соединений по массивам их спектральных или хроматографических характеристик, основанный на использовании аппарата теории нечетких множеств. Алгоритм минимально

использует априорные допущения о свойствах экспериментальных данных и обеспечивает достаточную устойчивость выводов о тождественности аналита и эталона к варьированию условий эксперимента.

6. Использование новой расчетной процедуры при решении задачи обработки данных количественного физико-химического анализа позволяет преодолеть трудности, связанные с одновременным определением химико-аналитических характеристик органо-кремнеземных гибридных материалов (эффективной сорбционной емкости и констант сорбционных равновесий).
7. Впервые определены химико-аналитические характеристики двух новых твердофазных аналитических реагентов на основе органо-кремнеземных гибридных материалов и определены условия их использования для извлечения ряда солей металлов из водных и этанольных растворов.

СПИСОК ПУБЛИКАЦИЙ АВТОРА ПО ТЕМЕ РАБОТЫ

1. Метрологические характеристики методик обнаружения с бинарным откликом : монография / [Ю.В. Холин, Н.А. Никитина, **А.В. Пантелеймонов** и др.] – Харьков: Тимченко, 2008. – 128 с. Рекомендована к печати Ученым Советом Харьковского национального университета имени В.Н. Каразина, протокол № 7 от 27.06.2008 р.

Автором обоснованы сокращенный список метрологических характеристик методик обнаружения с бинарным откликом и выбор функций логистического и экспоненциального распределений для представления кривых эффективности, сопоставлены способы назначения статистических весов при вычислении параметров кривых эффективности, испытан метод кривых Пирсона, разработан робастный алгоритм оценивания параметров на основе теории нечетких множеств, оценены метрологические характеристики 27 методик обнаружения аналитов; автор участвовал в обсуждении всех результатов и написании монографии.

2. Островская В. М. Тест-метод определения суммы металлов реактивной индикаторной бумагой и его метрологические характеристики / В. М. Островская, Е. А. Решетняк, Н. А. Никитина, **А. В. Пантелеймонов**, Ю. В. Холин // Журн. аналит. химии. – 2004. – Т. 59, № 10. – С. 995-1002.

Автор реализовал алгоритм оценки метрологических характеристик систем обнаружения с бинарным откликом, участвовал в обсуждении результатов и подготовке публикации.

3. Пантелеймонов А. В. К разработке нечетких критериев идентификации в качественном химическом анализе / **А. В. Пантелеймонов**, Ю. В. Холин // Вісник Харківського національного університету. – 2005. – № 669. Хімія. Вип. 13(36). – С. 128-133.
4. Magosso H. A. Synthesis, characterization and metal adsorption properties of the new ion exchanger polymer 3-n-propyl(4-methylpyridinium) silsesquioxane

chloride / H. A. Magosso, **A. V. Panteleimonov**, Y. V. Kholin, Y. Gushikem // J. Colloid and Interf. Sci. – 2006. – V. 303. – P. 18-24.

Автор провел моделирование сорбционных свойств материала, определил его химико-аналитические характеристики, принял участие в обсуждении результатов и подготовке публикации.

5. Lucho A. M. S. Simulation of adsorption equilibria on hybrid materials: Binding of metal chlorides with 3-n-propylpyridiniumsilsesquioxane Chloride Ion Exchanger / A. M. S. Lucho, **A. V. Panteleimonov**, Y. V. Kholin, Y. Gushikem // J. Colloid and Interf. Sci. – 2007. – V. 310. – P. 47-56.

Автор реализовал алгоритм одновременной оценки химико-аналитических характеристик материала, провел моделирование сорбционных равновесий, принял участие в обсуждении результатов и подготовке публикации.

6. Холин Ю. В. Некоторые методические и вычислительные проблемы количественного физико-химического анализа / Ю. В. Холин, С. А. Мерный, Д. С. Коняев, **А. В. Пантелеймонов** // В кн.: Научное наследие Н. А. Измайлова и актуальные проблемы физической химии. – Харків: ХНУ імені В. Н. Каразіна, 2007. – С. 525-553.

Автор испытал разработанный алгоритм одновременного оценивания констант равновесия и начальных (общих) концентраций реагентов в задачах физико-химического анализа для исследования протолитических равновесий в водном растворе глицина, принял участие в обсуждении результатов и подготовке публикации.

7. Пантелеймонов А. В. Математические методы проверки гипотез в тестовом анализе / **А. В. Пантелеймонов**, Е. А. Решетняк, Ю. В. Холин // Тези доп. Четвертої Всеукр. конф. студентів та аспірантів «Сучасні проблеми хімії». – Київ. – 2003. – С. 13-14.

Автор реализовал алгоритм оценки параметров кривых эффективности нелинейным МНК, выступил с устным докладом.

8. Пантелеймонов А. В. Усовершенствованный метод исследования метрологических характеристик методик визуального тестового анализа /

- А. В. Пантелеймонов**, Ю. В. Холин // Тези доп. П'ятої Всеукр. конф. студентів та аспірантів «Сучасні проблеми хімії». – Київ. – 2004. – С. 151.
9. Panteleimonov A. V. Application of fuzzy sets theory to solving tasks of multicomponent qualitative analysis / **A. V. Panteleimonov**, Yu. V. Kholin. // Book of abstracts International conference «AC&CA». – Kyiv. – 2005. – P. 48.
10. Пантелеймонов А. В. Вычислительные аспекты определения метрологических характеристик методик визуального тестового анализа / **А. В. Пантелеймонов**, Ю. В. Холин // Матеріали Сесії Наукової Ради НАНУ з проблеми «Аналітична хімія». – Харків. – 2007. – С. 32.
11. Kholin Yu. Chemometric tools at building models of equilibria on surfaces of hybrid silica-organic materials / Yu. Kholin, **A. Panteleimonov** // Book of abstracts International conference «Modern Physical Chemistry for Advanced Materials». – Kharkiv. – 2007. – P. 304-306.
12. Kholin Yu. Metrological characteristics of visual test analysis: chemometric consideration and new computational tools / Yu. Kholin, **A. Panteleimonov** // Book of abstracts 4th Black Sea Basin Conference on Analytical Chemistry. – Sunny Beach, Bulgaria. – 2007. – P141.
13. Пантелеймонов А. В. Скрининг с помощью тест-систем: метрологические характеристики и вычислительные аспекты их определения / А. В. Пантелеймонов // Материалы XV Международной конференции студентов, аспирантов и молодых ученых по фундаментальным наукам «Ломоносов – 2008». Химия. М.: МГУ имени М. В. Ломоносова. – 2008. – С. 60.

БЛАГОДАРНОСТИ

Автор выражает искреннюю признательность научному руководителю проф. Ю.В. Холину за всестороннюю поддержку и помощь.

Выполнение работы существенно стимулировали научные дискуссии с проф. А.И. Коробовым, доц. В.В. Ивановым и асс. Д.И. Ляхом, которым автор выражает искреннюю признательность.

Автор благодарен проф. А.А. Бугаевскому, проф. Л.П. Логиновой, доц. Е.А. Решетняк и доц. Н.А. Никитиной за плодотворное сотрудничество в области качественного химического анализа.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Золотов Ю.А. Химический анализ и аналитический контроль в различных областях науки, техники и производства // Рос. хим. ж. (Ж. Рос. хим. об-ва им. Д.И. Менделеева). – 2002. – Т. XLVI, № 4. – С. 8-10.
2. Родионова О.Е. Хемометрика: достижения и перспективы / О.Е. Родионова, А.Л. Померанцев // Успехи химии. – 2006. – Т. 75, № 4. – С. 302-321.
3. Kowalski B.R. Chemometrics: views and propositions // J. Chem. Inf. Comp. Sci. – 1975. – V. 15. – P. 201–203.
4. Аналитическая химия. Химические методы анализа / Под ред. д-ра хим. наук проф. О.М. Петрухина. – М.: Высшая школа, 1993. – 400 с.
5. Кельнер Р., Мерме Ж.-М., Отто М., Видмер Г.М. Аналитическая химия. Проблемы и подходы: Пер. с англ. Т. 1–2. – М.: Мир, 2004. – 608 с.
6. Марьянов Б.М. Избранные главы хемометрики. Учебное пособие для вузов по спец. 0112001 – аналитическая химия. – Томск: Изд-во Томского ун-та, 2004. – 164 с.
7. Wold S. Chemometrics: what do we mean with it, and what do we want from it? // Chemom. Intell. Lab. Syst. – 1995. – V. 30, No 1. – P. 109-115.
8. <http://rcs.chph.ras.ru/rcschar.htm>
9. Massart D.L., Vandeginste B.G.M., Buydens L.M.C., de Jong S., Lewi P.J., Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics: Part A. Amsterdam: Elsevier, 1997. – 886 p.
10. Department of Chemometrics, Institute of Chemistry, the University of Silesia, <http://www.chemometria.us.edu.pl>
11. Bradley E. A leisurely look at the bootstrap, the jackknife and the cross-validation // American statistician. – 1983. – V. 37. – P. 36-48.
12. Налимов В.В. Анализ трудностей, связанных с построением нелинейных по параметрам моделей в задачах химической кинетики // Зав. лаб. – 1978. – No 3. – С. 325-331.
13. <http://www.ecsocman.edu.ru/db/msg/6534.html>

14. Бланк А.Б. Аналитическая химия в исследовании и производстве неорганических функциональных материалов. – Харьков: «Институт монокристаллов», 2005. – 352 с.
15. Metrology in Chemistry: considerations, approaches and developments on the applicability of methods of ‘higher order’ // *Comptes Rendus Physique*. – 2004. – V. 5. – P. 907–920.
16. Love J.L. Chemical metrology, chemistry and the uncertainty of chemical measurements // *Accred Qual Assur.* – 2002. – V. 7. P. 95-100.
17. A. Marschal. Chemical metrology – QUID? / Marschal A., Andrieux T., Compagnon P.A., Fabre H. // *Accred Qual Assur.* – 2002. – V. 7. P. 42-49.
18. EUROCHEM/CITAC Guide «Quantifying uncertainty in analytical measurements», 2nd ed., 2000. – 120 p.; Руководство ЕВРАХИМ/СИТАК «Количественное описание неопределенности в аналитических измерениях», 2-е изд. / Пер. с англ. С.-Пб, 2002. – 141 с.
19. Valcárcel M. Analytical chemistry at the interface between metrology and problem solving / M. Valcárcel, B. Lendl // *Trends in Analytical Chemistry*. – 2004. – V. 23, No 8. – P. 527-534.
20. Practical guide to chemometrics. 2nd ed. / Ed. Paul Gemperline. – Boca Raton-London-New York: Taylor & Francis Group, 2006. – 520 p.
21. Geladi P. Chemometrics in spectroscopy. Part 1. Classical chemometrics // *Spectrochimica Acta Part B* – 2003. – V. 58. – P. 767-782.
22. Karoui R. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products / R. Karoui, J. De Baerdemaeker // *Food Chemistry*. – 2007. – V. 102. – P. 621-640.
23. Arvanitoyannis I.S. A Review on Tomato Authenticity: Quality Control Methods in Conjunction with Multivariate Analysis (Chemometrics) / I.S. Arvanitoyannis, O.B. Vaitis // *Critical Reviews in Food Science and Nutrition*. – 2007. – V. 47. – P. 675-699.
24. Granitto P.M. Modern data mining tools in descriptive sensory analysis: A case study with a Random forest approach / P.M. Granitto, F. Gasperi, F. Biasioli,

- E. Trainotti, C. Furlanello // *Food Quality and Preference*. – 2007. – V. 18. P. 681-689.
25. Setiawan M.A. Partial correlation metric based classifier for food product characterization / M.A. Setiawan, R.K. Rao, S. Lakshminarayanan // *J. of Food Engineering* – 2009. – V. 90. P. 146-152.
26. Urbano M. Ultraviolet-visible spectroscopy and pattern recognition methods for differentiation and classification of wines / M. Urbano, M.D.L. de Castro, P.M. Perez, J. Garcia-Olmo, M.A. Gomez-Nieto // *Food Chemistry* – 2006. – V. 97. – P. 166-175.
27. Bonnier F. Detection of pathological aortic tissues by infrared multispectral imaging and chemometrics / F. Bonnier, D. Bertrand, S. Rubin, L. Ventéo, M. Pluot, B. Baehrel, M. Manfait, G.D. Sockalingum // *Analyst*. – 2008. – V. 133. – P. 784-790.
28. Zhang Z. Classification of cancer patients based on elemental contents of serums using bidirectional associative memory networks / Z. Zhang, H. Zhuo, S. Liu, P. de B. Harrington // *Analytica Chimica Acta*. – 2001. – V. 436. – P. 281-291.
29. Komst Ł. The Basis Function Regression in pharmaceutical analysis. Theory and example of application / Ł. Komst, R. Skibinski, M. Paryło, K. Dudek // *J. of Pharmaceutical and Biomedical Analysis*. – 2008. – V. 47. – P. 659-669.
30. West C. Characterisation of stationary phases in subcritical fluid chromatography with the zolvation parameter model III. Polar stationary phases / C. West, E. Lesellier // *J. of Chromatography A*. – 2006. – V. 1110. – P. 200-213.
31. de Juan A. Solvent classification based on solvatochromic parameters: a comparison with the Snyder approach / A. de Juan, G. Fonrodona, E. Casassas // *Trends Anal. Chem.* – 1997. – V. 16, No 1. – P. 52-62.
32. Gramatica P. Classification of organic solvents and modelling of their physico-chemical properties by chemometric methods using different sets of molecular descriptors / P. Gramatica, N. Navas, R. Todeschini // *Trends Anal. Chem.* – 1999. – V. 18, No 7. – P. 461-471.

33. Musee N. New methodology for hazardous waste classification using fuzzy set theory. Part I. Knowledge acquisition / N. Musee, L. Lorenzen, C. Aldrich // *J. of Hazardous Materials*. – 2008. – V. 154. P. 1040-1051.
34. Doblea P. Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks / P. Doblea, M. Sandercock, E. Du Pasquier, P. Petocz, C. Roux, M. Dawson // *Forensic Science International*. – 2003. – V. 132. P. 26-39.
35. Scsibrany H. Clustering and similarity of chemical structures represented by binary substructure descriptors / H. Scsibrany, M. Karlovits, W. Demuth, F. Müller, K. Varmuza // *Chemometrics and Intelligent Laboratory Systems*. – 2003. – V. 67. P. 95-108.
36. Bendtsen A.B. Chemometric analysis of a detailed chemical reaction mechanism for methane oxidation / A.B. Bendtsen, P. Glarborg, K. Dam-Johansen // *Chemometrics and Intelligent Laboratory Systems*. – 1998. – V. 44. P. 353-361.
37. Rothenberg G. Data mining in catalysis: Separating knowledge from garbage // *Catalysis Today*. – 2008. – V. 137. P. 2-10.
38. Vlasov Yu. Non-selective chemical sensors in analytical chemistry: from “electronic nose” to “electronic tongue” / Yu. Vlasov, A. Legin // *Fresenius J. Anal. Chem.* – 1998. – V. 361. – P. 255-260.
39. Winqvist F. An electronic tongue based on voltammetry / F. Winqvist, P. Wide, I. Lundstrom // *Anal. Chim. Acta*. – 1997. – V. 357. – P. 21-31.
40. Akhmetshin A. Use of four-factorial design in ion-selective potentiometry for analysis of multi-ionic solutions / A. Akhmetshin, V. Baranovsky, A. Akhmetshina // *Fresenius J. Anal. Chem.* – 1998. – V. 361. – P. 282-284.
41. Workman J., jr. Process Analytical Chemistry / J. Workman, jr., M. Koch, D. Veltkamp // *Anal. Chem.* – 2005. V. 77. – P. 3789-3806.
42. Seasholtz M.B. Making money with chemometrics // *Chemometrics and Intelligent Laboratory Systems*. – 1999. V. 45. P. 55-63.
43. Mutihac L. Mining in chemometrics / L. Mutihac, R. Mutihac // *Anal. Chim. Acta*. – 2008. – V. 612. P. 1-18.

44. Linusson A. Fuzzy clustering of 627 alcohols, guided by a strategy for cluster analysis of chemical compounds for combinatorial chemistry / A. Linusson, S. Wold, B. Norden // *Chemometrics and Intelligent Laboratory Systems*. – 1998. – V. 44. – P. 213-227.
45. Marengo E. Linear discriminant hierarchical clustering: A modeling and cross-validable divisive clustering method / E. Marengo, R. Todeschini // *Chemometrics and Intelligent Laboratory Systems*. – 1993. – V. 19, No 1. – P. 43-51.
46. Ros F. Molecular descriptor selection combining genetic algorithms and fuzzy logic: application to database mining procedures / F. Ros, M. Pintore, J.R. Chretien // *Chemometrics and Intelligent Laboratory Systems*. – 2002. – V. 63. – P. 15-26.
47. Pop H.F. A fuzzy classification of the chemical elements / H.F. Pop, C. Sârbu, O. Horowitz, D. Dumitrescu // *J. Chem. Inf. Comput. Sci.* – 1996. – V. 36. – P. 465-482.
48. Sârbu C. Principal component analysis versus fuzzy principal component analysis. A case study: the quality of danube water (1985–1996) / C. Sârbu, H.F. Pop // *Talanta*. – 2005. – V. 65. – P. 1215-1220.
49. Pop H.F. The Fuzzy Hierarchical Cross-Clustering Algorithm. Improvements and Comparative Study / H.F. Pop, C. Sârbu // *J. Chem. Inf. Comput. Sci.* – 1997. – V. 37. – P. 510-516.
50. Sârbu C. Fuzzy clustering analysis of the first 10 MEIC chemicals / C. Sârbu, H.F. Pop // *Chemosphere*. – 2000. – V. 40. P. 513-520.
51. Marsili-Libelli S. Control of SBR switching by fuzzy pattern recognition // *Water Research*. – 2006. – V. 40. P. 1095-1107.
52. Holliday J.D. Clustering Files of Chemical Structures Using the Fuzzy k-Means Clustering Method / J.D. Holliday, S.L. Rodgers, P. Willett // *J. Chem. Inf. Comput. Sci.* – 2004. – V. 44. – P. 894-902.

53. Iliev B. A fuzzy technique for food- and water quality assessment with an electronic tongue / B. Iliev, M. Lindquist, L. Robertsson, P. Wide // *Fuzzy Sets and Systems*. – 2006. – V. 157. – P. 1155-1168.
54. Guler C. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data / C. Guler, G.D. Thyne, J.E. McCray, A.K. Turner // *Hydrogeology Journal*. – 2002. – V. 10, No 4. – P. 455-474.
55. Rassokhin D.N. Nonlinear mapping of massive data sets by fuzzy clustering and neural networks / D.N. Rassokhin, V.S. Lobanov, D.K. Agrafiotis // *J. Comp. Chem.* – 2001. – V. 22, No 4. – P. 373-386.
56. Chen Z.-P. Fuzzy linear discriminant analysis for chemical data sets / Z.-P. Chen, J.-H. Jiang, Y. Li, Y.-Z. Liang, R.-Q. Yu // *Chemometrics and Intelligent Laboratory Systems*. – 1999. – V. 45. – P. 295-302.
57. Musee N. New methodology for hazardous waste classification using fuzzy set theory. Part I. Knowledge acquisition / N. Musee, L. Lorenzen, C. Aldrich // *J. Hazard. Materials*. – 2008. – V. 154. – P. 1040-1051.
58. Pop H.F. A new fuzzy regression algorithm / H.F. Pop, C. Sârbu // *Anal. Chem.* – 1999. – V. 68, No 5. – P. 771-778.
59. Sârbu C. Fuzzy robust estimation of central location / C. Sârbu, H.F. Pop // *Talanta*. – 2001. – V. 54. – P. 125-130.
60. Changa Y.-H.O. Fuzzy regression methods – a comparative assessment / Y.-H.O. Changa, B.M. Ayyub // *Fuzzy Sets and Systems*. – 2001. – V. 119. – P. 187-203.
61. Nasrabadi E. LP-based approach to outliers detection in fuzzy regression analysis / E. Nasrabadi, S. Mehdi Hashemi, A.N. Mehdi Ghatee // *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems*. – 2007. – V. 15, No. 4. – P. 441-456.
62. Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе / Классификация и кластер. М.: Мир, 1980. – 389 с. С. 208-247.

63. Gushikem Y. Determination of the constants of affinity of FeCl_3 , CuCl_2 , and ZnCl_2 for a nitrogen-containing organosilane bonded on Al_2O_3 -cellulose acetate hybrid material surface from ethanol solution / Y. Gushikem, A.M. Lazarin, R. Landers, Y. Kholin // *J. Colloid Interface Sci.* – 2002. – V. 254. – P. 31-38.
64. Cárdenas S. Analytical features in qualitative analysis / S. Cárdenas, M. Valcárcel // *Trends Anal. Chem.* – 2005. – V. 24, No 6. – P. 477-487.
65. Kravchenko M.S. Estimation of detection limits of test methods for water analysis / M.S. Kravchenko, M.S. Fumarova, A.A. Bugaevski // *Intern. J. Environ. Anal. Chem.* – 1988. – V. 33. – P. 257-267.
66. Иванов В.М. Тест-методы в аналитической химии. Обнаружение и определение кобальта иммобилизованным 1-(2-пиридилазо)-2-нафтолом / В.М. Иванов, С.А. Морозко, С.В. Качин // *Журн. аналит. химии.* – 1994. – Т. 49, No 8. – С. 857-861.
67. Атаманиченко Я.А. Тест-метод определения никеля (II) в питьевой воде / Я.А. Атаманиченко, Р.П. Панталер, А.Б. Бланк // *Методы и объекты хим. анализа.* – 2007. – Т. 2, No 1. – С. 35-39.
68. Островская В.М. Тест-метод определения суммы металлов реактивной индикаторной бумагой и его метрологические характеристики / В.М. Островская, Е.А. Решетняк, Н.А. Никитина, А.В. Пантелеймонов, Ю.В. Холин // *Журн. аналит. химии.* – 2004. – Т. 59, No 10. – С. 995-1002.
69. Mizuguchi H. Highly sensitive colour change system within slight differences in metal ion concentrations based on homo-binuclear complex formation equilibrium for visual threshold detection of trace metal ions / H. Mizuguchi, H. Atsumi, K. Hashimoto, Y. Shimada, Y. Kudo, M. Endo, F. Yokota, J. Shida, T. Yotsuyanagi // *Anal. Chim. Acta.* – 2004. – V. 527. – P. 131-138.
70. Plata M.R. Development and validation strategies for qualitative spot tests: application to nitrite control in waters / M.R. Plata, N. Pérez-Cejuela, Rodríguez J., Ríos Á. // *Anal. Chim. Acta.* – 2005. – V. 537. – P. 223-230.

71. Veselova I.A. Visual determination of mercury (II) using horseradish peroxidase immobilized on polyurethane foam / I.A. Veselova, T.N. Shekhovtsova // *Anal. Chim. Acta.* – 1999. – V. 392. – P. 151-158.
72. Fake R.E. Rapid detection of hydroxyl groups on solid-phase / R.E. Fake, A. Routledge // *Tetrahedron Let.* – 2004. – V. 45. – P. 8925-8926.
73. Shannon S.K. Colorimetric monitoring of solid-phase aldehydes using 2,4-Dinitrophenylhydrazine / S.K. Shannon, G. Barany // *J. Comb. Chem.* – 2004. – V. 6. – P. 165-170.
74. León-Ruiz V. Validation of a screening method for the simultaneous identification of fat-soluble and water-soluble vitamins (A, E, B₁, B₂ and B₆) in an aqueous micellar medium of hexadecyltrimethylammonium chloride / V. León-Ruiz, S. Vera, M.P. San Andrés // *Anal. Bional. Chem.* – 2005. – V. 381. – P. 1568-1575.
75. Панталер Р.П. Тест-методи аналізу об'єктів оточуючої середовища, технологічних розчинів, препаратів наркотичного і психотропного дії / Р.П. Панталер, Т.І. Івкова, Л.В. Гудзенко, О.В. Гайдук, А.Б. Бланк // *Вісник Харківського національного університету імені В.Н. Каразіна.* – 2001. – № 532. – Хімія. Вип. 7 (30). – С. 31-40.
76. Evans M.I. Principles of screening / M.I. Evans, R.S. Galen, D.W. Britt // *Seminars in Perinatology.* – 2005. – V. 29. – P. 364-366.
77. Forsum U. The impact of qualitative analysis in laboratory medicine / U. Forsum, H.O. Hallander, A. Kallner, D. Karlsson // *Trends Anal. Chem.* – 2005. – V. 24, No. 6. – P. 546-555.
78. Notarianni L.J. False positives and negatives in routine testing for drugs of abuse / L.J. Notarianni, D. Belk, A.J. Collins // *Lancet.* – 1995. – V. 345. – P. 1115.
79. Ríos A. Quality assurance of qualitative analysis in the framework of the European project “MEQUALAN” / A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman,

- B. Neidhart, R.W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel // *Accred. Qual. Assur.* – 2003. – V. 8. – P. 68-77.
80. Вершинин В.И. Визуальный эффект качественной реакции как аналитический признак для компьютерной идентификации ионов / В.И. Вершинин, О.В. Соколова // *Вестн. Омск. гос. ун-та.* – 1999. Вып. 3. – С. 61-63.
81. Saritha K. Qualitative detection of selenium in fortified soil and water samples by a paper chromatographic – carboxyl esterase inhibition technique / K. Saritha, N.V. Nanda Kumar // *J. of Chromatography A.* – 2001. – V. 919. – P. 223-228.
82. Островская В.М., Запорожец О.А., Будников Г.К., Чернавская Н.М. Вода. Индикаторные системы. – М.: ФГУП ВТИИ, 2002. – 265 с.
83. Решетняк Е.А. Предел обнаружения в тест-методах анализа с визуальной индикацией. Влияющие факторы / Е.А. Решетняк, Н.А. Никитина, Л.П. Логинова, В.М. Островская // *Журн. аналит. химии.* – 2005. – Т. 60, No 10. – С. 1102-1109.
84. Zolotov Yu.A. Test methods for extra-laboratory analysis / Yu.A. Zolotov, V.M. Ivanov, V.G. Amelin // *Trends Anal. Chem.* – 2002. – V. 21, No 4. – P. 302-319.
85. Золотов Ю.А., Иванов В.М., Амелин В.Г. Химические тест-методы анализа. – М.: Едиториал УРСС, 2002. – 304 с.
86. Trullols E. Validation of qualitative analytical methods / E. Trullols, I. Ruisánchez, F.X. Rius // *Trends Anal. Chem.* – 2004. – V. 23, No 2. – P. 137-145.
87. Faber N.M. On the calculation of decision limits in doping control / N.M. Faber, R. Boqué // *Accred. Quality Assur.* – 2006. – V. 11, No 10. – P. 536-538
88. Pulido A. Uncertainty of results in routine qualitative analysis / A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius // *Trends Anal. Chem.* – 2003. – V. 22, No 10. – P. 647-654.
89. Ríos A. Reliability of binary analytical response / A. Ríos, H. Téllez // *Trends Anal. Chem.* – 2005. – V. 24, No 6. – P. 509-515.

90. Lendl B. Advancing from unsupervised, single variable-based to supervised, multivariate-based methods: A challenge for qualitative analysis / B. Lendl, B. Karlberg // Trends Anal. Chem. – 2005. – V. 24, No. 6. – P. 488-492.
91. Milman B.L. Identification of chemical compounds // Trends Anal. Chem.– 2005. – V. 24, No 6. – P. 493-508.
92. Pulido A. Estimating the uncertainty of binary test results to assess their compliance with regulatory limits / A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius // Anal. Chim. Acta. – 2002. – V. 455. – P. 267-275.
93. Комарь Н.П. Основы качественного химического анализа. I. Ионные равновесия. – Харьков: Изд-во Харьковского университета, 1955. – 448 с.
94. Руководство ЕВРАХИМ / СИТАК. Количественное описание неопределенности в аналитических измерениях. 2-е издание. Перевод с англ. Р.Л. Кадиса, Г.Р. Нежиховского, В.Б. Сими́на под общей редакцией Л.А. Конопелько – Санкт-Петербург: ВНИИМ им. Д.И. Менделеева, 2002. – 149 с.
95. International vocabulary of basic and general terms in metrology. ISO, Geneva, 1993.
96. Рекомендации по метрологии Р 50.2.038-2004. Измерения однократные прямые. Оценивание погрешностей и неопределенности результата измерений. М.: ИПК «Издательство стандартов», 2004.
97. Национальный стандарт Российской Федерации ГОСТ Р 52361-2005 «Контроль объекта аналитический. Термины и определения». М.: ИПК «Издательство стандартов», 2005.
98. Valcárcel M., Barceló D., Cárdenas S. et al. Metrology of qualitative chemical analysis. Luxembourg: Office for Official Publications of the European Communities, 2002. – 166 p.
99. Мильман Б.Л. Неопределенность результатов качественного химического анализа. Общие положения и бинарные тест-методы / Б.Л. Мильман, Л.А. Конопелько // Журн. аналит. химии. – 2004. – Т. 59, No 12. – С. 1244-1258.

100. Васильев В.В., Григорьев Н.Н. Практическое руководство по качественному химическому полумикроанализу. – Л.: Изд-во ЛГУ, 1966. – 264 с.
101. Нейман Ю. Вводный курс теории вероятностей и математической статистики. – М.: Наука, 1968. – 448 с.
102. Вершинин В.И., Дерендяев Б.Г., Лебедев К.С. Компьютерная идентификация органических соединений. – М.: Академкнига, 2002. – 197 с.
103. Ellison S.L.R. Characterizing the performance of qualitative analytical methods: Statistics and terminology / S.L.R. Ellison, T. Fearn // Trends Anal. Chem. – 2005. – V. 24, No 6. – P. 468-476.
104. Simonet B.M. Unreliability of screening methods / B.M. Simonet, A. Ríos, M. Valcárcel // Anal. Chim. Acta. – 2004. – V. 516. – P. 67-74.
105. Simonet B.M. Quality control in qualitative analysis // Trends Anal. Chem. – 2005. – V. 24, No 6. – P. 525-531.
106. Finney D.J. Probit analysis. 3-rd ed. Cambridge: Cambridge University Press, 1980. – 333 p.
107. Лошадкин Н.А. Пробит-метод в оценке эффектов физиологически активных веществ при низких уровнях воздействия / Н.А. Лошадкин, В.Д. Гладких, В.А. Голденков, А.Н. Сеницын, Л.В. Дарьина, Л.П. Буланова // Рос. хим. журн. – 2002. – Т. 46, No 6. – С. 63-67.
108. Решетняк Е.А. О достоверной оценке метрологических характеристик тестового анализа / Е.А. Решетняк, Н.А. Никитина, Ю.В. Холин, Н.В. Светлова, В.М. Островская // Вісник Харківського національного університету імені В.Н. Каразіна. – 2003. – No 596. – Хімія. Вип. 10 (33). – С. 90-98.
109. Emich F. Uber Mikrochemie, mit besonderer Berucksichtigung der Arbeiten von H. Behrens // Berichte. – 1910. – V. 43, No 1. – P. 10-13.
110. Song R. Field screening test methods: performance criteria and performance characteristics / R. Song, P.C. Schlecht, K. Ashley // J. Hazard. Mater. – 2001. – V. 83. – P. 29-39.

111. IUPAC Compendium of Chemical Terminology, Electronic version, <http://goldbook.iupac.org/D01629.html>.
112. Currie L.A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995) // *Pure Appl. Chem.* – 1995. – V. 67, No. 10. – P. 1699-1723.
113. Currie L.A. Detection and quantification limits: origins and historical overview // *Anal. Chim. Acta.* – 1999. – V. 391, No 2. – P. 127-134.
114. Currie L.A. Detection and quantification limits: basic concepts, international harmonization, and outstanding (“low-level”) issues // *Appl. Radiation and Isotopes.* – 2004. – V. 61, No 2-3. – P. 145-149
115. Aguilera E. Robustness in qualitative analysis: a practical approach / E. Aguilera, R. Lucena, S. Cardenas, M. Valcarcel, E. Trullols, I. Ruisanchez // *Trends Anal. Chem.* – 2006. – V. 25, No. 6. – P. 621-627.
116. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. – М.: Статистика, 1980. – 95 с.
117. Комарь Н.П. Особенности и возможности колориметрического и спектрофотометрического анализа // *Труды комиссии по аналитической химии.* – 1958. – VIII (XI). – С. 21-51.
118. Джадд Д. Цвет в науке и технике. – М.: Мир, 1978. – 592 с.
119. Решетняк Е.А. Влияние материала носителя на предел обнаружения при визуальном тестировании / Е.А. Решетняк, Н.А. Никитина, В.Э. Кармазина // *Вісник Харківського національного університету імені В.Н. Каразіна.* – 2007. – No 770. – Хімія. Вип. 15 (38). – С. 119-124.
120. Бугаевский А.А. Графический анализ функций распределения вероятностей, связанных с пределом обнаружения / А.А. Бугаевский, М.С. Кравченко // *Журн. аналит. химии.* – 1983. – Т. 38, No 1. – С. 17-21.
121. Бугаевский А.А. О вероятностных моделях чувствительности качественных химических реакций / А.А. Бугаевский, В.О. Круглов, М.С. Кравченко // *Зав. лаб.* – 1976. – Т. 42, No 1. – С. 68-70.

122. Панталер Р.П. Экспрессное полуколичественное определение остаточного активного хлора в питьевой воде с помощью индикаторной бумаги / Р.П. Панталер, Л.А. Егорова, Л.И. Авраменко, А.Б. Бланк // Журн. аналит. химии. – 1996. – Т. 51, No 5. – С. 521-524.
123. Панталер Р.П. Экспрессное полуколичественное определение железа в питьевой воде с помощью индикаторной бумаги / Р.П. Панталер, А.К. Тимченко, Л.И. Авраменко, А.Б. Бланк // Журн. аналит. химии. – 1997. – Т. 52, No 4. – С. 384-386.
124. Панталер Р.П. Экспресс-тест для определения кобальта в воде / Р.П. Панталер, А.К. Тимченко, Л.И. Авраменко, А.Б. Бланк // Журн. аналит. химии. – 1998. – Т. 53, No 5. – С. 529-532.
125. Юрченко О.И. Оценка предела обнаружения ионов железа (II) по реакции с димедоном / О.И. Юрченко, А.А. Бугаевский, Т.А. Бланк // Вісник Харківського національного університету імені В.Н. Каразіна. – 2003. – No 596. – Хімія. Вип. 10 (33). – С. 106-109.
126. Throne J.E. Probit analysis: assessing goodness of fit based in backtransformation and residuals / J.E. Throne, D.K. Weaver, J.E. Baker // J. Econom. Entomol. – 1995. – V. 88, No 5. – P. 1513-1516.
127. Seefeldt S.S. Log-logistic analysis of herbicide dose-response relationships / S.S. Seefeldt, J.E. Jensen, E.P. Fuerst // Weed Technology – 1995. – V. 9. – P. 218-227.
128. Справочник по прикладной статистике. Т. 1. Пер. с англ. / Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. – М.: Финансы и статистика, 1989. – 510 с. (Гл. 6. Метод максимального правдоподобия).
129. Fowles J.R. The use of benchmark dose methodology with acute inhalation lethality data / J.R. Fowles, G.V. Alexeeff, D. Dodge // Regulatory Toxicology and Pharmacology. – 1999. – V. 29, No 3. – P. 262-278.
130. Кукса А.И. Использование ГИС для системного анализа экобиомедицинских данных в контексте оценивания рисков. Часть 1: Общая

- характеристика предметной области и возникающих проблем // Проблемы программирования. – 2003. – № 4. – С. 72-81.
131. Нікітіна Н.О. Тест-методи хімічного аналізу з візуальною індикацією: метрологічне забезпечення, нові тест-системи: Дис. ... канд. хим. наук: 02.00.02 / ХНУ имени В.Н. Каразина. – Харьков, 2006. – 140 с.
 132. Бугаевский А.А., Зухер Хуари, Кравченко В.А. Исследование предела обнаружения капельной реакции на тетрафенилборат-ион. – Харьков, ХГУ, 1985. – 12 с. Деп. ОНИИТЭХИМ г. Черкассы 25.09.1985, № 2301.
 133. Холин Ю.В. Органо–кремнеземные материалы с иммобилизованными ксиленоловым оранжевым и кальцеином: получение, физико-химические свойства, обнаружение ионов металлов / Ю.В. Холин, С.В. Корнеев, И.В. Христенко, F. Pissetti, Y. Gushikem // Методы и объекты хим. анализа. – 2008. – Т. 3, № 1. – С. 64-74.
 134. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн. Кн. 1. – М.: Финансы и статистика, 1986. – 366 с.
 135. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн. Кн. 2. – М.: Финансы и статистика, 1987. – 351 с.
 136. Уилкс С. Математическая статистика. – М.: Наука, 1967. – 632 с.
 137. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятности и математической статистики. – М.: Наука, 1965. – 512 с.
 138. Поллард Дж. Справочник по вычислительным методам статистики. – М.: Финансы и статистика, 1982. – 344 с.
 139. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. – М.: Мир, 1985. – 509 с.
 140. Налимов В.В. Применение математической статистики в анализе вещества. М.: Гос. изд-во физ.-мат. л-ры, 1960. – 432 с.
 141. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. – 416 с.
 142. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: исследование зависимостей. – М.: Финансы и статистика, 1985. – 487 с.

143. Ван дер Варден Б.Л. Математическая статистика. – М.: Изд-во иностр. л-ры, 1960. – 434 с.
144. Басиль Е.Е. Анализ и выбор модели зависимости «доза-эффект» при оценке загрязнения атмосферы / Е.Е. Басиль, С.А. Изотов // Труды Одесского политехнического университета. – 1998. – Вып. 2 (6). http://www.library.ospu.odessa.ua/online/periodic/opu_98_2/index.htm#Ecolog.
145. Лахтин Л.К. Кривые распределения и построение для них интерполяционных формул по способам Пирсона и Брунса. – М., 1922. – 152 с.
146. Prevéy P.S. The use of Pearson VII distribution functions in X-ray diffraction residual stress measurement // *Adv. in X-Ray Analysis*. – 1986. – V. 29. – P. 103-111.
147. Loh S. Pearson Type I approximation of human mortality curve / S Loh, P. Krishnan // *Population Reprints*. – 1986. – No. 82. – P. 429-432.
148. Zografos K. On maximum entropy characterization of Pearson's type II and VII multivariate distributions // *J. Multivariate Analysis*. – 1999. – V. 71. – P. 67-75.
149. Nagahara Y. The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters // *Stat. Prob. Lett.* – 1999. – V. 43. – P. 251-264.
150. Gutierrez-Jaimez R. On the efficiency property of the linear estimators in the bivariate Pearson type VII distribution / R. Gutierrez-Jaimez, J.D. Jimenez-Lopez // *J. Stat. Comp. and Simul.* 72 (12): 937-948 2002.
151. Aleksandrov M.S. Approximation of structureless bands in the electron spectra of molecules using Pearson curves / M.S. Aleksandrov, V.S. Ivanov, V.B. Sovkov // *J. Appl. Spectrosc.* – 1990. – V. 53, No 2. – P. 827-830.
152. Bogdanic G. Pearson-type I distribution function for polydisperse polymer systems. Molar mass distribution / G. Bogdanic, L. Jakab // *J. Chem. Inf. Comput. Sci.* – 2003. – V. 43. – P. 880-884.
153. Мудров В.И., Кушко В.Л. Методы обработки измерений. Квазиправдоподобные оценки. – М.: Радио и связь, 1983. – 304 с.
154. Bandemer H. Fuzzy theory in analytical chemistry / H. Bandemer, M. Otto // *Mikrochim. Acta.* – 1986. – No. 2. – P. 93-124.

155. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – М.: Наука, 1981. – 208 с.
156. Кофман А. Введение в теорию нечетких множеств. – М.: Радио и связь, 1982. – 432 с.
157. Sarbu C. Fuzzy robust estimation of central location / C. Sarbu, H.F. Pop // *Talanta*. – 2002. – V. 54, No 1. – P. 125-130.
158. Wu J.-W. Weighted quasi-likelihood estimation based on fuzzy clustering analysis method and dimension reduction technique / J.-W. Wu, T.-R. Tsai // *Fuzzy Sets and Systems*. – 2002. – V. 128. P. 353-364.
159. Холин Ю.В. Метрологические характеристики методик обнаружения с бинарным откликом / Ю.В. Холин, Н.А. Никитина, А.В. Пантелеймонов, Е.А. Решетняк, А.А. Бугаевский, Л.П. Логинова – Харьков: Тимченко, 2008. – 128 с.
160. Назаренко А.Ю. Применение теории нечетких множеств для обработки результатов анализа / А.Ю. Назаренко, В.В. Сухан, Н.А. Назаренко // *Зав. лаб.* – 1991. – Т. 57, No 10. – С. 63-65.
161. Дробышев Ю.П., Коптюг В.А. // *Вестн. АН СССР*, 1970. – № 8. – С. 75-79.
162. Основы аналитической химии: В 2 кн. / Под ред. Ю.А. Золотова. М.: Высш. шк., 1999. Т. 2. С. 431.
163. Vershinin V.I. A priori method of evaluating uncertainties in qualitative chromatographic analysis: Probabilistic approach // *Accred. Qual. Assur.* — 2004. – V. 9. – P. 415–418.
164. Вершинин В.И. Методология компьютерной идентификации веществ с применением информационно-поисковых систем // *Журн. аналит. химии.* – 2000. – Т. 55. – С. 468-476.
165. Mc Lafferty F.W. Interpretation of mass-spectra. Sausalito: University Science Book, 1993. 371 p.
166. Mc Lafferty F.W. Probability based matching system using a large collection of reference mass spectra // *Anal. Chem.* – 1976. – V. 48. – P. 1362-1368.

167. Stauffer D.B. Probability-based-matching algorithm with forward searching capabilities for matching unknown mass spectra of mixtures / D.B. Stauffer, F.W. McLafferty, R.D. Ellis, D.W. Peterson // *Anal. Chem.* – 1985. – V. 57. – P. 1056-1060.
168. Harrington P.B. Closure effects on infrared spectral library search performance / P.B. Harrington, T.L. Isenhour // *Appl. Spectroscopy* – 1987. – V. 41. – P. 1298-1301.
169. Clerc J.T. Performance analysis of infrared library search systems / J.T. Clerc, E. Pretch, M. Zürcher // *Mikrochim. Acta.* – 1986. – V. 89. – P. 217-242.
170. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
171. Шусторович А.М. Об адекватных парных мерах сходства в задачах распознавания образов с разнородными признаками // Вопросы обработки информации при проектировании систем. – Новосибирск: ИМ СО АН СССР, 1977. – С. 147-152.
172. Вершинин В.И. Критерии совпадения пиков в качественном хроматографическом анализе. Учет воспроизводимости характеристик удерживания / В.И. Вершинин, В.А. Топчий, И.И. Медведовская // *Журн. аналит. химии.* – 2001. – Т. 56, № 4. – С. 367-373.
173. www.chemometrics.ru/materials/presentations/analytica-2005/vershinin.ppt
174. Тейлор Дж. Введение в теорию ошибок. – М.: Мир, 1985. – 272 с.
175. Мильман Б.Л. Введение в химическую идентификацию. – СПб.: ВВМ, 2008. – 180 с.
176. Conti C. FT-IR of trichloroacetoimidates in different solvent system / C. Conti, R. Galeazzi, E. Giorgini, G. Tosi // *J. Mol. Struct.* – 2005. – Vol. 744-747. P. 417-423.
177. Большаков Г.Ф., Ватаго В.С., Агрест Ф.Б. Ультрафиолетовые спектры гетероорганических соединений. Л.: Химия, 1969. – 504 с.

178. Neagu V. Synthesis of new chelating ion exchange resin with hydroxamic and amidoxime groups and study of its metal binding property / V. Neagu, I. Bunia, I. Plesca, M. Popa // *J. Appl. Polym. Sci.* – 2003. – V. 88, No 13. – P. 2956-2962.
179. Henry W.D. Preparation and characterization of a new class of polymeric ligand exchangers for selective removal of trace contaminants from water / W.D. Henry, D. Zhao, A.K. SenGupta, C. Lange // *React. Funct. Polym.* – 2004. – V. 60. – P. 109-120.
180. Rao G.N. Chelating ion-exchange resin membrane sensor for nickel(II) ions / G.N. Rao, S. Srivastava, S.K. Srivastava, M. Singh // *Talanta.* – 1996. – V. 43, No 10. – P. 1821-1825.
181. Borgo C.A. Aluminium phosphate dispersed on a cellulose acetate fiber surface. Preparation, characterization and application in Li^+ , Na^+ and K^+ separation / C.A. Borgo, Y. Gushikem, Y. Kholin // *Anal. Chim. Acta.* – 2003. – V. 477, No 2. – P. 305-313.
182. Zaporozhets O.A. Xylenol orange adsorbed on silica surface as a solid phase reagent for lead determination using diffuse reflectance spectroscopy / O.A. Zaporozhets, L.Ye. Tsyukalo // *Talanta.* – 2002. – V. 58. – P. 861-868.
183. Zaporozhets O.A. Solid-phase reagents for the determination of anionic surfactants in water / O.A. Zaporozhets, O.Yu. Nadzhafova, V.V. Verba, S.A. Dolenko, T.Ye. Keda, V.V. Sukhan // *Analyst.* – 1998. – V. 123. – P. 1583-1586.
184. Запорожец О.А. Имобилизация аналитических реагентов на поверхности носителей / О.А. Запорожец, О.М. Гавер, В.В. Сухан // *Успехи химии.* – 1997. – Т. 66, № 7. – С. 703-712.
185. Запорожець О.А. Адсорбовані на кремнеземах органічні реагенти у комбінованих спектроскопічних і тест-методах аналізу // Автореф. дис. ... докт. хім. наук. Київ, 2003. – 34 С.
186. Запорожец О.А. Сорбционно-спектроскопическое и тест-определение Cu(II) с помощью иммобилизованного на силикагеле дитизоната цинка /

- О.А. Запорожец, Т.Е. Кеда, И.М. Богославец // Химия и технология воды. – 2005. – Т. 22, № 6. – С. 436-445.
187. Запорожец О.А. Визуальный тест-метод определения селена(IV) иммобилизованным на кремнеземе индигокармином / О.А. Запорожец, С.Л. Билоконь // Журн. аналит. химии. – 2007. – Т. 62, № 2. – С. 208-212.
188. Запорожец О.А. Иммобилизованный на кремнеземе 1-(4-адамантил-2-тиазолилазо)-2-нафтол в анализе сосуществующих форм меди в природных водах / О.А. Запорожец, Р.П. Линник, О.Б. Воловенко, Т.М. Радзиевская // Методы и объекты хим. анализа. – 2007. – Т. 2, № 1. – С. 40-50.
189. Запорожець О.А. Твердофазно-спектрофотометричне визначення нікелю та цинку іммобілізованим дитизоном / О.А. Запорожець, Л.С. Зінько, Т.Є. Кеда, К.П. Левченко, І.В. Притика // Методы и объекты хим. анализа. – 2007. – Т. 2, № 1. – С. 62-69.
190. Zaporozhets O.A. Determination of fluoride and oxalate using the indicator reaction of Zr(IV) with methylthymol blue adsorbed on silica gel / O.A. Zaporozhets, L.Ye. Tsyukalo // Anal. Chim. Acta. – 2007. – V. 597, No 1. – P. 171-177.
191. Моросанова Е.И. Использование модифицированных кремнезёмов для сорбционного концентрирования, разделения и определения неорганических и органических соединений // Лабораторный журнал. – 2002. – № 2(2). – С. 32-38.
192. Моросанова Е.И. Ксерогели, модифицированные 4-(2-пиридилазо)-резорцином и хлоридом цетилпиридиния: твердофазно-спектрофотометрическое определение цинка в растворах / Е.И. Моросанова, Ж.М. Азарова, Ю.А. Золотов // Журн. аналит. химии. – 2002. – Т. 57, № 5. – С. 499-503.
193. Azarova Z.M. Xerogels modified with 1-(2-pyridylazo)-2-naphtol and dimethylglyaxime: indicator tubes for determining nickel / Z.M. Azarova, E.I. Morosanova, Y.A. Zolotov // J. Anal. Chem. – 2000. – V. 55, No 7. – P. 641-644.

194. Золотов Ю.А. Сорбционное концентрирование микрокомпонентов для целей химического анализа / Ю.А. Золотов, Г.А. Цизин, Е.И. Моросанова, С.Г. Дмитриенко // Успехи химии. – 2005. – Т. 74, № 1. – С. 41-66.
195. Золотов Ю.А., Цизин Г.А., Дмитриенко С.Г., Моросанова Е.И. Сорбционное концентрирование микрокомпонентов из растворов: применение в неорганическом анализе. М.: Наука, 2007. – 320 с.
196. Россоти Ф., Россоти Х. Определение констант устойчивости и других констант равновесия в растворах: Пер. с англ. – М.: Мир, 1965. – 564 с.
197. Суворов А.В. Модельные представления при описании равновесных систем. // Математические методы в химической термодинамике. – Новосибирск: Наука, 1982. – С. 100-107.
198. Холин Ю.В. Количественный физико-химический анализ комплексообразования в растворах и на поверхности химически-модифицированных кремнезёмов: содержательные модели, математические методы и их приложения. – Харьков: Фолио, 2000. – 288 с.
199. Аносов В.Я., Озерова Н.И., Фиалков Ю.Я. Основы физико-химического анализа. – М.: Наука, 1976. – 504 с.
200. Бек М., Надьпал И. Исследование комплексообразования новейшими методами. – М.: Мир, 1989. – 413 с.
201. Meloun M., Havel J., Hogfeldt E. Computation of solution equilibria: a guide to methods in potentiometry, extraction and spectrophotometry. – Chichester: Ellis Horwood, 1988. – 298 p.
202. Джейкок М., Парфит Д.А. Химия поверхности раздела фаз: Пер. с англ. – М.: Мир, 1984. – 269 с.
203. Рогинский С.З. Адсорбция и катализ на неоднородных поверхностях. – М-Л.: Изд-во АН СССР, 1948. – 643 с.
204. Jaroniec M., Madey R. Physical Adsorption on Heterogeneous Solids. – Amsterdam, e.a.: Elsevier, 1988. – 354 p.

205. Koopal L.K. Ion adsorption on mineral oxide surface // Adsorption on New and Modified Inorganic Sorbents. / Ed. by A.Dabrowsky, V.A.Tertykh. – Amsterdam e.a.: Elsevier, 1996. – P. 757-796.
206. Адамсон А.А. Физическая химия поверхностей: Пер. с англ. – М.: Мир, 1979. – 568 с.
207. Kholin Yu. Energetic heterogeneity of sorbents: numerical calculation of affinity distributions / Yu. Kholin, S. Myerniy // Вісник Харківського національного університету імені В.Н. Каразіна. – 2004. – No 626. – Хімія. Вип. 11 (34). – С. 351-366.
208. Khoroshevskiy Y. A mathematical simulation of H⁺ ion chemisorption by anilinepropylsilica xerogels / Y. Khoroshevskiy, S. Korneev, S. Myerniy, Y.V. Kholin, F.A. Pavan, J. Schifino, T.M.H. Costa, E.V. Benvenuti // J. Colloid Interface Sci. – 2005. – V. 284, No 2. – P. 424-431.
209. Helfferich F. Ion Exchange. – N.Y.: Dover, 1995. – 624 p.
210. Inglezakis V.J. The concept of “capacity” in zeolite ion-exchange systems // J. Colloid Interface Sci. – 2005. – V. 281, No 1. – P. 68-79.
211. Vaghetti J.C.P. Influence of the synthesis conditions on the characteristics and metal adsorption properties of the 3-(1,4-phenylenediamine)propylsilica xerogel / J.C.P. Vaghetti, J.L. Brasil, T.M. Costa, E.V. Benvenuti // Eclat. Quim. – 2005. – V. 30, No 4. – P. 43-49.
212. <http://www.anchem.ru/chemanalysis/2005/020-036-full.asp>
213. Bordbar A.-K. Analysis of oxygen binding by hemoglobin on the basis of mean intrinsic thermodynamic quantities / A.-K. Bordbar, S. Habib-Allah Mousavi, H. Dazhampanah // Acta Biochimica Polonica. – 2006. – V. 53, No 3. – P. 536-568.
214. Lucho A.M.S. Simulation of adsorption equilibria on hybrid materials: Binding of metal chlorides with 3-n-propylpyridiniumsilsesquioxane chloride ion exchanger / A.M.S. Lucho, A.V. Panteleimonov, Y.V. Kholin, Y. Gushikem // J. Colloid Interface Sci. – 2007. – V. 310. – P. 47-56.

215. Magosso H.A. Synthesis, characterization and metal adsorption properties of the new ion exchanger polymer 3-n-propyl(4-methylpyridinium) silsesquioxane chloride / H.A. Magosso, A.V. Panteleimonov, Y.V. Kholin, Y. Gushikem // *J. Colloid Interface Sci.* – 2006. – V. 303. – P. 18-24.
216. Мамлеев В.Ш., Золотарев П.П., Гладышев П.П. Неоднородность сорбентов: (феноменологические модели). – Алма-Ата: Наука, 1989. – 288 с.
217. Strauss U.P. Analysis of ionization equilibriums of polyacids in terms of species population distributions. Examination of a "two-state" conformational transition / U.P. Strauss, B.W. Barbieri, G. Wong // *J. Phys. Chem.* – 1979. – V. 83. – P. 2840-2843.
218. Meloun M., Havel J. Computation of solution equilibria. 1. Spectrophotometry. 1984. 184 p. 2. Potentiometry. 1985. 145 p. – Brno: Univerzita J.E. Purkyně-Prirodovědecká Faculta.
219. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // *Зав. лаб.* – 1991. – Т. 57, No 7. – С. 64-66.
220. Корнилов А.Н. Метод наименьших квадратов и метод наименьших // *Математические методы в химической термодинамике.* – Новосибирск: Наука, 1982. – С. 164-167.
221. Уилкс С. Математическая статистика. – М.: Наука, 1967. – 632 с.
222. Налимов В.В. Применение математической статистики при анализе вещества. – М.: Гос. изд-во физ.-мат. лит-ры, 1960. – 430 с.
223. Havel J. Multiparametric curve fitting. IX / J. Havel, M. Meloun // *Talanta.* – 1986. – V. 33, No 5. – P. 435-441.
224. Brauner P. High-speed computers as a supplement to graphical methods. 9. Adjustment for systematic experimental errors and other "group parameters" in LETAGROP: Applications to potentiometric titrations. / P. Brauner, L.G. Sillén, R. Whiteker // *Arkiv Kemi.* – 1969. – V. 31. – P. 365-376.
225. Huber P. Robust statistical procedures, in: CBMS-NSF Regional conf. series in appl. mathematics, SIAM, Philadelphia, 1996.

226. Otto M. Fuzzy theory explained // Chemometrics and Intelligent Lab. Systems. – 1988. – V. 4. – P. 101 120.
227. Ross T.J. Fuzzy Logic with Engineering Applications. – N.Y.: McGraw-Hill, 1995. – 268 p.
228. IUPAC Stability Constants Database.
<http://www.acadsoft.co.uk/download/download.htm>