

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

На правах рукописи

Богомолов Андрей Юрьевич

РАЗВИТИЕ МУЛЬТИСЕНСОРНОГО ПОДХОДА
В ОПТИЧЕСКОМ СПЕКТРАЛЬНОМ АНАЛИЗЕ

02.00.02 – Аналитическая химия

ДИССЕРТАЦИЯ

На соискание учёной степени

доктора химических наук

Научный консультант:

доктор химических наук,
профессор Д.О. Кирсанов

Самара – 2020

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	7
1. ОПТИЧЕСКИЕ МУЛЬТИСЕНСОРНЫЕ СИСТЕМЫ В АНАЛИТИЧЕСКОЙ СПЕКТРОСКОПИИ (ОБЗОР ЛИТЕРАТУРЫ)	17
1.1 Современные направления в развитии оптической спектроскопии..	17
1.1.1 Лабораторная спектроскопия	17
1.1.2 Спектроскопия в анализе технологических процессов	18
1.1.3 Материалы, технические решения и методы исследования	20
1.1.4 Анализ спектральных данных	26
1.1.5 Децентрализация и персонализация анализа	27
1.2 Оптические мультисенсорные системы	29
1.2.1 Определение ОМС	29
1.2.2 Отличительные особенности	30
1.2.3 ОМС и одноканальные сенсоры	34
1.2.4 Архитектура и принцип работы ОМС	36
1.2.5 Конструкционные элементы ОМС	39
1.2.6 Классификация ОМС	40
1.3 Анализ спектральных данных и данных ОМС	43
1.3.1 Хемометрика как метод анализа спектральных данных	43
1.3.2 Предварительная обработка данных	46
1.3.3 Отбор переменных и объектов	48
1.3.4 Разведочный факторный анализ данных	51
1.3.5 Регрессионный анализ данных и многомерная градуировка	53
1.3.6 Дискриминантный анализ данных	56
1.3.7 Проверка градуировочной модели	57
1.3.8 Стратегии моделирования	59
1.3.9 Планирование градуировочного эксперимента	62
1.3.10 Хемометрика процессов	64
ЗАКЛЮЧЕНИЕ К ГЛАВЕ 1	66

2.	МЕТОДОЛОГИЯ РАЗРАБОТКИ ОМС	68
2.1	Цикл разработки ОМС	68
2.1.1	Подход к разработке	68
2.1.2	Основные этапы и задачи разработки	69
2.2	Моделирование информационных каналов ОМС	72
2.2.1	Выбор метода и полноспектральные измерения.....	72
2.2.2	Конфигурация оптических каналов.....	75
2.2.3	Алгоритм оптимизации.....	78
2.2.4	Интервальный метод отбора переменных при одновременной оптимизации предобработки данных.....	80
2.2.5	Техника измерения.....	80
2.3	Сбор и низкоуровневая обработка данных ОМС	83
2.3.1	Сбор данных и их свойства	83
2.3.2	Нормирование мультисенсорных данных	84
2.3.3	Обеспечение прецизионности измерения.....	87
2.3.4	Выбор эталонного образца	89
2.3.5	Безэталонный количественный анализ	90
2.4	Построение и валидация моделей в процессе разработки ОМС.....	94
2.4.1	Предобработка мультисенсорных данных.....	94
2.4.2	Особенности построения градуировочных моделей для ОМС	96
2.4.3	Многоуровневая валидация модели	97
2.4.4	«TRT cloud» – комплекс программ для анализа многомерных данных и разработки ОМС «в облаке».....	101
	ЗАКЛЮЧЕНИЕ К ГЛАВЕ 2	103
3.	ПЛАНИРОВАНИЕ ГРАДУИРОВОЧНОГО ЭКСПЕРИМЕНТА.....	105
3.1	Введение в планирование многокомпонентного градуировочного эксперимента	105
3.1.1	Понятия и определения.....	105
3.1.2	ПЭ как оптимизационная проблема	106

3.1.3	Оценка результата ПЭ.....	106
3.2	Требования к набору градуировочных образцов.....	108
3.2.1	Некоррелированность факторов	108
3.2.2	Равномерность заполнения.....	109
3.2.3	Покрытие экспериментального пространства.....	110
3.2.4	Встроенная проверка.....	112
3.2.5	Интерпретируемость	113
3.2.6	Дополнительные характеристики	113
3.3	Диагональный дизайн многокомпонентного градуировочного эксперимента	114
3.3.1	Основные допущения и правила конструирования.....	114
3.3.2	Диагональный дизайн градуировки для двух компонентов	116
3.3.3	Обобщение ДД для градуировки трёх и более компонентов	123
3.3.4	Расширенные схемы ДД	130
3.4	Примеры специальных дизайнов градуировочного эксперимента для различных приложений	133
3.4.1	Планирование градуировочного эксперимента для определения жира и белка в молоке	133
3.4.2	Дизайн градуировочного эксперимента в АКТП.....	138
	ЗАКЛЮЧЕНИЕ К ГЛАВЕ 3	143
4.	ОМС В АНАЛИТИЧЕСКОМ КОНТРОЛЕ ПРОЦЕССОВ.....	145
4.1	Аналитическое пространство и траектория процесса.....	145
4.1.1	Основные понятия и определения	145
4.1.2	Дискретизация сбора данных.....	148
4.1.3	Визуализация траекторий.....	150
4.1.4	Проблема избыточности аналитического пространства	154
4.2	Разрешение и анализ траекторий.....	156
4.2.1	Методология анализа процессов.....	156
4.2.2	Качество траектории	159
4.3	Применение траекторий процесса.....	161

4.3.1	Анализ процесса в реальном времени.....	163
4.3.2	Мониторинг процессов.....	163
4.3.3	Контроль процесса.....	165
4.3.4	Понимание процесса.....	165
4.3.5	Оптимизация процесса.....	166
ЗАКЛЮЧЕНИЕ К ГЛАВЕ 4.....		167
5.	ПРИМЕРЫ РАЗРАБОТКИ И ПРИМЕНЕНИЯ ОМС.....	168
5.1	Определение жира и общего белка в молоке.....	168
5.1.1	Оптический анализ молока на содержание жира и белка.....	168
5.1.2	Обоснование спектрального метода определения молочного жира и белка в области 400–1100 нм.....	173
5.1.3	Построение рабочих моделей.....	184
5.1.4	Перенос модели методом коррекции наклона и отсечения.....	186
5.1.5	Разработка светодиодного анализатора.....	188
5.1.6	Пространственно-разрешённая спектроскопия и безэталонный метод измерения в анализе молока.....	193
5.2	Контроль качества фармацевтических продуктов.....	200
5.2.1	Анализ процессов фармацевтических производств.....	200
5.2.2	Построение точной градуировочной модели для онлайн-мониторинга содержания воды в грануляте.....	203
5.2.3	Онлайн-мониторинг массовой доли воды и толщины покрытия в процессах изготовления пеллет.....	209
5.2.4	Онлайн-прогноз профилей растворения АФИ из БИК-спектров пеллет в процессе их изготовления.....	219
5.3	Онлайн-мониторинг ферментации <i>Saccharomyces cerevisiae</i>	223
5.3.1	Мониторинг биотехнологических процессов.....	223
5.3.2	Разработка мультисенсорной системы в ИК-области.....	226
5.3.3	Мониторинг процесса ферментации посредством ИК- и БИК-спектроскопии.....	236
5.3.4	Флуориметрический мониторинг процесса ферментации.....	239

5.3.5 Разрешение траектории процесса ферментации из данных 2D-флуориметрии.....	243
5.4 Оптические методы диагностики рака почки	246
5.4.1 Методы, проблемы и задачи онкологической диагностики	247
5.4.2 Образцы	249
5.4.3 Разработка и тестирование светодиодного сенсора для диагностики рака почки	250
5.4.4 Эффект комбинации флуоресцентной и ИК-спектроскопии.....	255
5.5 Экологический мониторинг почв	262
ЗАКЛЮЧЕНИЕ ГЛАВЕ 5	265
ЗАКЛЮЧЕНИЕ	267
ВЫВОДЫ.....	268
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	270
СПИСОК ЛИТЕРАТУРЫ	273
ПРИЛОЖЕНИЕ А. ГЕНЕРАТОР СХЕМ ДИАГОНАЛЬНОГО ДИЗАЙНА НА ЯЗЫКЕ «МАТЛАВ» С ПРИМЕРАМИ	302
ПРИЛОЖЕНИЕ Б. СТАНДАРТНЫЕ ОБРАЗЦЫ	304
ПРИЛОЖЕНИЕ В. ПРАКТИЧЕСКИЕ РАЗРАБОТКИ АНАЛИТИЧЕСКИХ УСТРОЙСТВ С ИСПОЛЬЗОВАНИЕМ МУЛЬТИСЕНСОРНОГО ПОДХОДА..	305

ВВЕДЕНИЕ

Востребованность химического анализа в современном мире стремительно растёт. Высокие требования, предъявляемые к эффективности производства, качеству продукции и безопасности окружающей среды стимулируют развитие инструментальных аналитических методов. Научные и технические достижения открывают всё новые возможности для внедрения аналитического контроля в различные области человеческой деятельности: промышленность, науку, медицину, экологию и даже в личную среду обитания, в быт.

Анализ сложных многокомпонентных смесей перестаёт быть прерогативой специализированных лабораторий. Для решения современных аналитических задач необходимы новые устройства, способные контролировать состояние анализируемого объекта быстро и на месте – на производственной линии или в полевых условиях, без постоянного отбора проб. Результат анализа должен выдаваться без задержки, зачастую в реальном времени. При необходимости анализатор должен производить десятки и сотни измерений в секунду.

Оптические методы являются одним из ведущих направлений в развитии новой инструментальной базы современного химического анализа. В наше время эта область переживает бурное развитие, во многом связанное с техническими достижениями фотоники, а именно с совершенствованием и появлением новых систем детектирования, источников света, оптики и светопроводящих материалов. Оптический анализ обладает рядом практических достоинств, включая высокую информативность, недеструктивность, адаптируемость к различным объектам и средам. Немаловажной предпосылкой роста популярности оптических, прежде всего спектроскопических, методов является совершенствование информационных технологий: компьютеров, приборной электроники и современных методов анализа многомерных данных (*хеометрики*).

Традиционная оптическая спектроскопия исторически развивалась как универсальное аналитическое решение и была ориентирована на лабораторные измерения предварительно отобранных проб разнообразного состава. По этой причине она оказалась плохо приспособлена к решению ряда современных

аналитических задач, таких как мониторинг процессов, экспресс-анализ и полевые исследования. Невысокая пропускная способность лабораторной спектроскопии не отвечает растущему числу анализируемых проб и подлежащих контролю объектов. Технологическая сложность универсальных спектрометров обуславливает их стационарность и высокую стоимость, тем самым препятствуя широкому распространению методов оптической спектроскопии за пределы лабораторий научно-исследовательских учреждений, крупных предприятий и аналитических центров.

Часть задач оперативного аналитического контроля различных образцов и сред, особенно в индустриальной сфере, успешно решается при помощи простых одноканальных сенсоров, в том числе оптических. Одноканальный оптический (фотометрический) сенсор определяет концентрацию компонента через его корреляцию с какой-либо измеряемой характеристикой образца, например с оптической плотностью на выбранной длине волны. Одноканальные сенсоры хорошо подходят для решения узкоспециализированных задач, таких как определение концентрации взвешенного компонента по мутности среды или анализ образцов, в которых только определяемый компонент даёт оптический отклик. То есть применимость фотометрии ограничена требованием отсутствия сторонних факторов, влияющих на измерение. При количественном анализе смесей, где аналитические сигналы различных компонентов существенно перекрываются, к каковым относится подавляющее большинство реальных образцов, информативность одноканального сенсора зачастую становится недостаточной и не может обеспечить требуемой точности анализа.

В последние годы наблюдается развитие оптической спектроскопии в сторону специализированных аналитических устройств, разрабатываемых под конкретное практическое приложение. Эти устройства, получившие в настоящей работе название *оптические мультисенсорные системы (ОМС)*, занимают промежуточное положение между одноканальными фотометрическими сенсорами и универсальными лабораторными спектрометрами. В то же время они обладают рядом ярких отличительных особенностей, позволяющих выделить их в отдельный

новый класс оптических анализаторов. ОМС работают в широком спектральном диапазоне и предназначены для решения разнообразных аналитических задач, связанных с определением как индивидуальных веществ, так и обобщённых показателей химического состава. При этом используется небольшое число сенсорных каналов, например регистрирующих суммарное поглощение на определённых интервалах длин волн, а недостаток селективности компенсируется применением математического моделирования как на этапе оптимизации каналов, так и при анализе данных измерений. Научные исследования в этой области и примеры действующих ОМС показывают их огромный потенциал, реализация которого, как ожидается, устраним разрыв между сегодняшней растущей потребностью в аналитическом контроле и ограниченными возможностями традиционных методов, более того, приведёт к формированию новых областей применения химического анализа.

Дальнейшая систематическая разработка и широкое внедрение оптических мультисенсорных систем требует решения ряда научных, технических и методологических задач, чем и обусловлена актуальность настоящей работы. Прежде всего необходимо сформулировать общие подходы к созданию оптических мультисенсорных систем для конкретного практического приложения и разработать методы оптимизации ОМС на основе данных предварительно спланированного эксперимента. Для действующих ОМС необходимо разработать новые и усовершенствовать существующие аналитические методики с тем, чтобы расширить области применения и улучшить аналитические характеристики подобных систем, чем и обусловлена **актуальность** настоящей работы.

Цель работы состоит в развитии научных основ мультисенсорного подхода в оптическом спектральном анализе и в создании оптических мультисенсорных систем для широкого круга аналитических приложений.

Для достижения поставленной цели требовалось решить следующие **задачи**:

- 1) Систематизировать подходы к разработке оптических мультисенсорных систем как отдельного класса специализированных спектроскопических анализаторов низкой селективности и их оптимизации под конкретное

практическое приложение при определении как отдельных компонентов смесей, так и обобщённых показателей химического состава.

- 2) Разработать алгоритмы и программное обеспечение (ПО) для оптимизации каналов ОМС, для централизованного хранения и анализа больших объёмов данных, а также удалённого применения прогностических моделей.
- 3) Систематизировать требования, предъявляемые к обучающим и проверочным образцам в рамках планирования многокомпонентного градуировочного эксперимента, и предложить эффективный экспериментальный план (дизайн), позволяющий строить достаточно точные градуировочные модели на минимальном наборе образцов с учётом решаемых при разработке и использовании ОМС задач.
- 4) Разработать методологию применения мультисенсорного подхода в аналитическом контроле технологических процессов (АКТП) для создания ОМС, способных производить непрерывный мониторинг процессов и онлайн-анализ качества продукции в ходе производства.
- 5) Расширить аналитические возможности и область практического применения ОМС, проиллюстрировать их потенциал примерами разработки действующих систем и соответствующих методик анализа.

Научная новизна исследования заключается в формулировании и решении ряда задач, составляющих научную основу создания оптических мультисенсорных систем – специализированных аналитических устройств низкой селективности.

Разработан новый подход к оптимизации ОМС для конкретного аналитического приложения на этапе технической разработки устройства на базе адаптированного генетического алгоритма.

Предложен новый централизованный подход к математическому обеспечению полного цикла создания и использования ОМС.

Систематизированы общие требования к планированию градуировочного эксперимента для одновременного определения нескольких компонентов. Сформулированные теоретические положения расширяют возможности методологии количественного анализа сложных смесей и дают исследователям

эффективные критерии оценки качества градуировочного набора, что особенно важно при использовании хемометрических методов. Разработан диагональный дизайн – новое семейство высокоэффективных схем, интегрирующих обучающие и проверочные образцы в одном наборе.

Предложена концепция *траектории процесса* в пространстве аналитических переменных, дающая методологический инструментарий для более эффективного спектрального мониторинга, контроля и оптимизации технологических процессов.

Сделан ряд рекомендаций по улучшению метрологических характеристик методик анализа в различных областях, в частности, в пищевой и фармацевтической промышленности, а также в медицине и экологии, в том числе для онлайн-мониторинга процессов, экспресс-анализа и полевых измерений, за счёт использования предложенных новых подходов к получению, предобработке и анализу данных.

Практическая значимость. Предложены оптические мультисенсорные системы, обладающие небольшим размером, портативностью, лёгкой встраиваемостью в производственную линию при значительном снижении стоимости анализа (иногда на несколько порядков по сравнению с лабораторной спектроскопией). Это позволяет расширить круг сложных объектов, пригодных для аналитического контроля на местах, как показано на примере натурального молока, твёрдых лекарственных форм, среды процесса ферментации, образцов биологической ткани и загрязнённых почв.

Предложенный улучшенный алгоритм расчёта каналов ОМС для конкретного приложения многократно сокращает время вычислений, что позволяет использовать большее число параметров в оптимизации, повышая точность анализа. Подход может также использоваться как обобщённый интервальный метод улучшения хемометрических моделей.

Разработанный централизованный подход к математическому обеспечению ОМС реализован «в облаке» в не имеющем аналогов комплексе программ «TPT cloud» (tptcloud.com), работающем через стандартный интернет-браузер и

позволяющем оптимизировать каналы ОМС под приложение, а также строить, улучшать и использовать прогностические модели.

Предложенный диагональный дизайн многокомпонентного градуировочного эксперимента облегчает процесс создания ОМС, предоставляя разработчикам удобные схемы построения обучающего и проверочного наборов образцов, что показано на следующих примерах: сравнение нескольких прототипов инфракрасной (ИК) ОМС для одновременного определения этанола и различных углеводов в среде идущего процесса ферментации, разработка ОМС для анализа молока на содержание жира и общего белка в области 400–1100 нм, а также разработка методики зондирования почв на содержание углеводов нефти и воды.

При помощи разработанных ОМС и соответствующих методик определения компонентов сложных смесей решён ряд аналитических задач полевого, экспресс- и онлайн-анализа. Получены следующие результаты:

- снижена погрешность (среднеквадратичное отклонение) онлайн-определения массовой доли воды в диапазоне 0–10% в ходе сушки порошкообразных фармацевтических и пищевых продуктов до 0.1% (в три раза по сравнению с существующими подходами);
- построена модель, которая из спектров в ближней инфракрасной (БИК) области, снятых онлайн в ходе процесса изготовления препарата целевой доставки (пеллет), прогнозирует будущую кинетику высвобождения ацетаминофена после приёма лекарства пациентом;
- показано улучшение аналитических характеристик соответствующих методик при дополнении флуориметрии данными измерений в ИК- и БИК-области, что даёт более надёжную диагностику рака почки (снижение ложных отнесений) и повышение точности определения биомассы в ходе ферментации (снижение погрешности на четверть и более);
- разработана улучшенная методика зондирования почв на содержание углеводов нефти методом ИК-спектроскопии нарушенного полного

внутреннего отражения (НПВО) в условиях широкой вариабельности влажности объекта.

Разработаны прототипы:

- низкоселективных компактных светодиодных ОМС-анализаторов для определения содержания жира и общего белка в натуральном молоке при измерении (в том числе пространственно-разрешённом) в видимой и коротковолновой БИК-области в режимах пропускания и диффузного отражения с погрешностью менее 0.1% по обоим компонентам, то есть создающих альтернативу стандартным лабораторным методам ИК- и БИК-спектроскопии;
- ИК ОМС с оптоволоконным зондом, позволяющей определять содержание глюкозы, фруктозы и этилового спирта в ходе процесса ферментации дрожжей с погрешностями менее 10 г/л по этим компонентам при их совместном определении в практически используемом диапазоне концентраций;
- светодиодной ОМС и соответствующие методики для обнаружения границы опухоли при раке почки, что может быть использовано как дополнительный тест в ходе хирургической операции.

Положения, выносимые на защиту:

- 1) Методология создания специализированных ОМС для химического анализа различных объектов сложного состава, в особенности при полевом, экспресс- и онлайн-анализе, позволяет проводить специализацию оптических каналов мультисенсорной системы (определение их необходимого числа и свойств) под конкретное практическое приложение на полных спектрах спланированного набора образцов при помощи улучшенного оптимизационного алгоритма.
- 2) Новый централизованный подход, реализованный «в облаке» в разработанном комплексе программ, предоставляет математическое обеспечение полного цикла создания и использования ОМС, включая

оптимизацию каналов разрабатываемых систем, а также построение и использование прогностических моделей.

- 3) Диагональный дизайн представляет собой новый подход к планированию многокомпонентного градуировочного эксперимента, учитывающий особенности данных спектральных методов и ОМС, применимый к определению питательной ценности молока, определению основных компонентов ферментационной среды и к экологическому мониторингу почв.
- 4) Концепция траектории процесса в АКТП систематизирует подход к онлайн-анализу технологических процессов производств, а именно сбору спектральных и ОМС-данных, построению прогностических моделей и использованию результатов моделирования для принятия управляющих решений в ходе процесса.
- 5) Разработанные рекомендации по созданию методик на основе ОМС обеспечивают улучшенные аналитические характеристики при решении ряда практических задач: анализа молока на содержание жира и белка, онлайн-мониторинга процесса ферментации дрожжей, онлайн-мониторинга различных процессов производства твёрдых фармацевтических форм, выявления границ злокачественной опухоли и определения нефтепродуктов в почвах.

Достоверность полученных в ходе исследования результатов обеспечивается достаточным количеством экспериментальных наблюдений, проведённых современными методами исследования, которые соответствуют поставленным в работе целям и задачам, использованием модельных образцов известного состава и референтных образцов, проанализированных аттестованными методами, а также результатами практической апробации. Научные положения, выводы и рекомендации, подкреплены убедительными фактическими данными, наглядно представленными в приведённых таблицах и рисунках. Обработка и статистический анализ данных, а также интерпретация полученных результатов проведены с использованием современного программного обеспечения.

Апробация работы. Материалы диссертации представлены на международных научных конференциях: Winter Symposium on Chemometrics (WSC) (г. Саратов, 2020; г. Санкт-Петербург, 2018; г. Самара, 2016; г. Томск, 2014; д. Дракино, 2012; г. Санкт-Петербург, 2010; г. Казань, 2008; г. Самара, 2006, г. Черноголовка, 2005; Пушкинские Горы, 2004; г. Белокуриха, 2003; все – в РФ), Advances in Process Analytics and Control Technology (АРАСТ) (г. Честер, 2019; г. Манчестер, 2015; г. Глазго, 2009; все – в Великобритании), Kolloquium für Prozessanalytik (г. Ганновер, ФРГ, 2018; г. Вена, Австрия, 2015), Centre for Process Analytics and Control Technology (CPACT) Research Day (г. Глазго, Великобритания, 2018), European Conference on Process Analytics and Control Technology (EuroРАСТ) (г. Потсдам, ФРГ, 2017; г. Барселона, Испания, 2014), Conference on Chemometrics in Analytical Chemistry (САС) (г. Барселона, Испания, 2016; г. Будапешт, Венгрия, 2012; г. Антверпен, Бельгия, 2010; г. Монпелье, Франция, 2008; г. Сиэтл, США, 2002), International Society of Clinical Spectroscopy (SPEС) (г. Монреаль, Канада, 2016), Royal Society of Chemistry (RSC) Faraday Discussion (г. Кембридж, Великобритания, 2016), Scientific Exchange (SCIEХ) (г. Провиденс, США, 2015), International Diffuse Reflectance Conference (IDRC) (г. Чамберсбург, США, 2014), International Conference on Near Infrared Spectroscopy (г. Ла Грант Мотт, Франция, 2013), Dairy Industry Association of Australia (DIAA) Victoria State Conference (г. Мельбурн, Австралия, 2013), Forum on milk quality and safety (г. Пекин, КНР, 2012), Process Analytical Technology (РАТ) Conference (г. Гейдельберг, ФРГ, 2009), ProzessNet-Jahrestagung (г. Мангейм, ФРГ, 2009), Scandinavian Symposium on Chemometrics (г. Лоен, Норвегия, 2009), International Forum on Process Analytical Technology (IFРАС) (г. Балтимор, США, 2009), РАТ workshop by GEA (г. Бубендорф, Швейцария, 2008), Quality by Design and Process Analytical Technology in a Good Manufacturing Practice (GMP) Framework (г. Галле (Заале), ФРГ, 2008), Pittsburgh Conference (PITTCOΝ) (г. Чикаго, 2007; г. Орландо, 2005; г. Чикаго, 2004; г. Орландо, 2003; г. Новый Орлеан, 2001; все – в США), AnalyticaExpo (г. Москва, РФ, 2005), Federation of Analytical Chemistry and

Spectroscopy Societies (FACSS) (г. Портленд, США, 2004), Conference on Hyphenated Techniques in Chromatography (г. Брюгге, Бельгия, 2004).

Разработки в области оптических мультисенсорных систем апробированы в России и за рубежом: на предприятии «Молочный комбинат «Самаралакто» (г. Самара, РФ), в Самарском государственном техническом университете (г. Самара, РФ), на предприятии «Schwarzwaldmilch» (г. Фрайбург, ФРГ), на предприятии «GEA Pharma Systems» (г. Бубендорф, Швейцария и г. Воммельгем, Бельгия), в госучреждении «Bundeskriminalamt» (г. Висбаден, ФРГ) и в других организациях.

Лицензии на программу «TRT cloud» были приобретены Санкт-Петербургским государственным университетом (г. Санкт-Петербург, РФ), фирмой «art photonics» (г. Берлин, ФРГ), фирмой «Mestrelab Research» (г. Сантьяго-де-Компостела, Испания) и другими организациями.

Публикации. По результатам работы опубликованы 32 статьи, 4 главы в монографиях, 1 патент на изобретение и 2 авторских свидетельства.

Личный вклад автора¹ в работы, опубликованные в соавторстве с другими исследователями, заключался в постановке цели исследования и решаемых в нём задач, выборе объектов исследования, разработке подходов к анализу экспериментальных данных, участии в проведении экспериментов и математической обработке данных, обобщении полученных результатов, а также в формулировании научных положений и выводов.

Структура и объём работы. Диссертационная работа состоит из введения, пяти глав, списка литературы и трёх приложений. Работа изложена на 307 страницах машинописного текста (включая приложения), содержит 78 рисунков и 24 таблицы. Библиографический список включает 301 источник.

¹ В 2016 г. автор награждён золотой медалью Российского хемометрического общества «За выдающийся вклад».

1. ОПТИЧЕСКИЕ МУЛЬТИСЕНСОРНЫЕ СИСТЕМЫ В АНАЛИТИЧЕСКОЙ СПЕКТРОСКОПИИ (ОБЗОР ЛИТЕРАТУРЫ)

Вступительная глава рассматривает современное состояние науки и техники, а также тенденции развития оптического спектрального анализа. Введено понятие оптической мультисенсорной системы и показаны её основные отличия от традиционной лабораторной спектроскопии.

1.1 Современные направления в развитии оптической спектроскопии

1.1.1 Лабораторная спектроскопия

Оптическая спектроскопия как метод анализа, основанный на измерении поглощённого, рассеянного или испущенного веществом света, начинается с работ Кирхгофа, Бунзена и Ангстрема [1,2], и её история насчитывает более 150 лет. В наше время она является одним из наиболее распространённых физических инструментальных методов аналитической химии.

Традиционным вектором развития молекулярной спектроскопии является совершенствование приборной базы с точки зрения повышения эффективности количественного и качественного анализа широкого круга объектов. Универсальность оптического анализатора достигается максимизацией спектрального и динамического диапазонов, разрешения, точности, то есть *прецизионности и правильности* измерения, автоматизацией, компьютеризацией и другими показателями качества, обуславливающими техническое усложнение прибора, а следовательно, его удорожание и рост габаритов. Важным фактором, диктующим конструкционную сложность промышленно производимых спектрометров, является стандартизация – обеспечение идентичности спектров одного и того же образца, полученных на различных приборах одного типа. Стандартизация обеспечивает достоверность анализа, но требует дополнительных технических решений, усложняя разработку, производство и обслуживание прибора. Вопросами технической стандартизации в спектроскопии занимается ряд международных и национальных организаций, выпускающих нормативные документы и стандартные образцы. К ним относятся: действующий в России и

странах СНГ «Межгосударственный совет по стандартизации, метрологии и сертификации», международная «ASTM International», американские «National Institute of Standards» («NIST») и «US Pharmacopeia Standards» («UPS»), а также другие организации.

Не удивительно, что спектроскопия, возникшая в научной среде, до сих пор прочно ассоциируется с лабораторным анализом предварительно отобранных проб (*образцов*), уникальным оборудованием и квалифицированным персоналом. Неизбежная дороговизна традиционного спектрального анализа делает его прерогативой централизованных лабораторий больших организаций и предприятий. Например, лаборатории наиболее крупных молокозаводов мира оснащены инфракрасным (ИК) спектральным анализатором «MilkScan» фирмы «FOSS» [3], по стоимости сравнимым с элитным автомобилем. В то же время мелкие и средние производители используют более трудоёмкие и зачастую менее эффективные, но экономически приемлемые методы анализа продукции.

В последние два-три десятилетия в оптической спектроскопии начала активно развиваться альтернатива традиционному анализу. Наметившиеся изменения связаны с тем, что господствующий на сегодняшний день лабораторный подход перестаёт справляться с растущими потребностями общества в химическом анализе. Это связано с постоянным ростом числа образцов и контролируемых параметров в существующих областях деятельности, таких как контроль качества производимой продукции, экологический мониторинг и медицина, а также с появлением новых областей и форм анализа, таких как полевой и экспресс-анализ. Методы оптической спектроскопии способны ответить на эти вызовы.

1.1.2 Спектроскопия в анализе технологических процессов

Наметившийся поворот во многом диктуется индустриальными нуждами, связанными с ужесточением требований к качеству продукции, прежде всего в пищевой и фармацевтической отраслях. Это привело к появлению самостоятельной научно-прикладной дисциплины, получившей название *аналитический контроль технологических процессов* (АКТП) [4].

Целью АКТП является реализация предложенной Юраном (*англ. Juran*)

концепции встроенного качества (*англ.* quality by design, QbD) [5] – перехода от проверки качества произведённого продукта к мониторингу самого процесса производства с тем чтобы как можно раньше обнаружить и по возможности предотвратить сбой [6,7]. Помимо технологических процессов как таковых, концепция Юрана распространяется и на используемые аналитические методы.

Достижение этой цели требует изменения самого подхода к химическому анализу производственных образцов с тем, чтобы значительно сократить время получения результата по сравнению с традиционным лабораторным анализом. Мониторинг идущего процесса должен происходить в реальном времени или с незначительной задержкой, позволяя принимать оперативные управляющие решения. Такой мониторинг называют *непрерывным*, а аналитические измерения принято разделять на три группы [8-10]:

- «у линии» (*англ.* at-line) – образец изымается и немедленно анализируется оператором в непосредственной близости от места *пробоотбора*;
- «на линии» (*англ.* on-line) – образец автоматически выводится из среды процесса без его физической изоляции (например, при помощи обводной трубы), анализируется, а затем возвращается обратно;
- «в линии» (*англ.* in-line) – анализ происходит непосредственно в среде процесса.

Задержка результата анализа снижается в порядке перечисления и для измерения «в линии» практически отсутствует. С точки зрения реализации и скорости анализа варианты «на линии» и «в линии» очень близки, и для них в дальнейшем будет использован общий термин *онлайн-измерение*.

Таким образом, для онлайн-анализа требуются быстрые, химически информативные, недеструктивные и не требующие пробоподготовки методы. Полное соответствие этим требованиям сделало спектроскопию ультрафиолетовой (УФ), видимой (Вид) и ближней ИК (БИК) областей основным инструментом АКТП. Как следствие, начали активно развиваться быстрые и компактные *процесс-спектрометры*, пригодные к встраиванию в производственную линию, и *процесс-интерфейсы* – зонды, пропускные ячейки и другие приспособления для снятия

онлайн-спектров.

Основные требования, предъявляемые к процесс-анализаторам – это скорость измерения, возможность подключения зондов и других процесс-интерфейсов, компактность, устойчивость к механическим воздействиям. Поскольку установка компьютера в производственном помещении, как правило, не допускается, важным качеством анализатора является автономность в обработке получаемых данных. Преследуемая в АКТП цель повышения уровня контроля процессов требует многократного увеличения числа точек онлайн-контроля и, следовательно, анализаторов. В этих условиях снижение стоимости устройства становится весьма актуальным. В то же время высокие требования к параметрам, которые важны для универсальных лабораторных спектрометров (разрешению, охватываемой спектральной области, точности, в том числе, прецизионности и т. д.), в процесс-спектрометрах могут быть снижены до минимального уровня, необходимого для обеспечения контроля качества продукции конкретного производства. Изменение требований, предъявляемых к процесс-спектрометрам, задаёт новое направление развития оптической спектроскопии, существенно отличающееся от традиционного лабораторного. Важнейшим дискуссионным вопросом в развитии процесс-спектроскопии является возможность отказа от стандартизации анализатора за счёт переноса этой функции в область обработки и анализа данных, что будет обсуждено в разделах 1.2.2 и 1.2.6.

Заданный АКТП импульс уже привёл и продолжает приводить к появлению новых материалов, технических решений и математического аппарата для инструментального оптического анализа. Благодаря этому в анализе процессов начинают применяться методы, считавшиеся ранее чисто лабораторными, например ИК-спектроскопия ($400\text{--}4000\text{ см}^{-1}$) и спектроскопия комбинационного рассеяния (КР), называемая также рамановской [11,12].

1.1.3 Материалы, технические решения и методы исследования

На развитие современной оптической спектроскопии в наибольшей степени повлияли технические достижения в создании новых оптоволоконных материалов и зондов, источников света и детекторов, а также методов получения и обработки

данных.

Зонд «Lighthouse Probe™» («LHP») является ярким примером процесс-интерфейса, разработанного с учётом особенностей непрерывного анализа на производстве [13]. Это – погружной зонд для УФ/Вид- и БИК-спектрального анализа, изготовленный из сертифицированных в фармацевтическом и пищевом производстве материалов. Зонд позволяет снимать спектры диффузного отражения в среде различных процессов обработки порошков, в частности *грануляции*, нанесения покрытия *пеллет*² и сушки. Такие процессы обычно проводятся в «кипящем» (*псевдооживленном*) слое – в реакторе, обеспечивающем непрерывную циркуляцию взвешенных частиц в подаваемом снизу воздушном потоке. Высокое качество и информативность спектров обеспечивается 360°-ным обзором среды через 7 оптических окон, радиально расположенных по бокам зонда. Каждое окно при этом одновременно служит как для облучения образца, так и для приёма диффузно отражённого света, передавая его на удалённый детектор при помощи оптоволоконного кабеля. Оптические компоненты зонда позволяют проводить измерения в полной УФ/Вид/БИК-области. «LHP» может быть оснащён системой автоматической очистки для удаления налипшего материала. При этом рабочая часть зонда выводится из среды идущего процесса и омывается струёй воды. Перед возвращением зонда в процесс происходит его сушка и перекалибровка относительно встроенного стандарта (рис. 1-1).

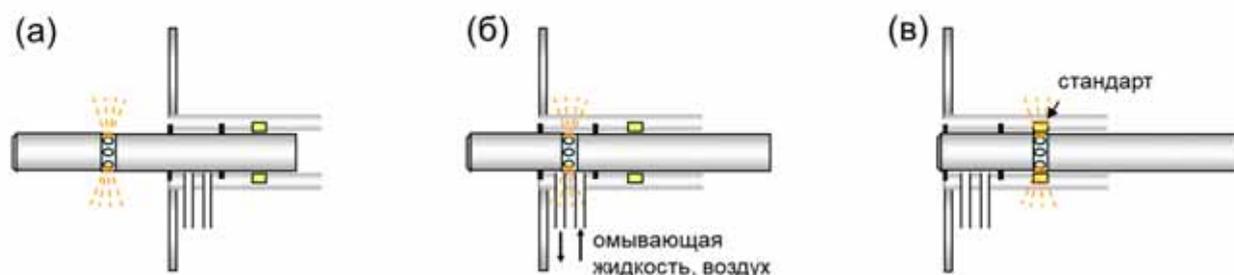


Рисунок 1-1. Принцип работы зонда «LHP»: (а) измерение; (б) очистка и последующая сушка и (в) перекалибровка.

² Пеллеты (*англ. pellets*) – фармацевтическая форма, сферические гранулы диаметром около 1 мм. Обычно используются для заполнения лекарственных капсул.

Новые светопроводящие материалы расширили рабочую область световодов. Современные оптоволокна покрывают практически полный оптический диапазон излучения. Удалось решить проблему «непрозрачности» ИК-области. Новые халькогенидные ИК (ХИК-)волокна на основе сульфида мышьяка (As_2S_3) пропускают электромагнитное излучение на интервале длин волн 1.5–6 мкм и имеют в данной области наивысшее пропускание среди известных оптоволоконных материалов [14]. В то же время бескислородные поликристаллические ИК (ПИК-)волокна, изготовленные из твёрдых растворов галогенидов серебра $AgCl_{1-x}Br_x$ ($0 < x < 1$) [15], обладают необходимой прозрачностью от 3 до 18 мкм, то есть в богатой химической информацией области «отпечатков пальцев», называемой так из-за попадающих в неё полос фундаментальных частот поглощения многих функциональных групп.

Благодаря новым светопроводящим материалам всё большее применение в АКТП находят ИК-зонды, использующие эффект нарушенного полного внутреннего отражения (НПВО; *англ.* attenuated total reflection, ATR). НПВО-зонды прекрасно подходят для ИК-анализа различных сред, в том числе в АКТП. Полный оптический путь при НПВО-измерении соответствует глубине проникновения переменного поля затухающих колебаний (*англ.* evanescent field) используемого кристалла (0.5–2 мкм), помноженной на число внутренних отражений, определяемое геометрией НПВО-элемента. Такой короткий оптический путь часто выгоден, учитывая, что коэффициенты экстинкции веществ в ИК-области гораздо выше, чем, например, в БИК. Наиболее распространённые НПВО-кристаллы производятся из ZrO_2 , $ZnSe$, Ge , Si и алмаза. Форма кристалла (призма, конус, полусфера, многогранная поверхность и т. д.) может варьироваться в зависимости от типа зонда и природы образца [16]. В течение двух последних десятилетий на базе ХИК- и ПИК-волокон были разработаны различные НПВО-зонды для ИК-анализа. Эти зонды, в паре с ИК-спектрофотометрами, были успешно протестированы в различных аналитических приложениях, таких как непрерывный мониторинг процессов и анализ биологических тканей [17-22].

Настоящий прорыв в области источников света происходит благодаря

расширению возможностей светодиодов – охватываемой ими спектральной области от УФ до ИК и диапазона интенсивностей до десятков Вт (в импульсном режиме), а также за счёт их существенного удешевления. Благодаря этим качествам, они всё чаще используются в оптическом анализе [23]. Будучи по спектральным свойствам близки к монохроматическим источникам излучения, мощные светодиоды могут заменить лазеры, например для возбуждения флуоресценции [24, с. 342] и даже КР-эффекта [25]. Срок службы стандартных светодиодов исчисляется тысячами часов с сохранением стабильной интенсивности и спектра излучения, что делает их гораздо более практичными, чем традиционно используемые в спектроскопии лампы. Новаторской, ещё мало разработанной областью исследования является использование в спектроскопии набора светодиодов на разных длинах волн вместо традиционного белого источника света в комбинации с физической дисперсией или математическим преобразованием, как в Фурье-спектроскопии. В этом случае спектрометр можно освободить от наиболее дорогостоящих комплектующих. В то же время реализация данного подхода потребует решения ряда проблем, например связанных с достаточно широкими спектрами излучения светодиодов, с их вариабельностью и т. д. Существуют первые примеры исследовательских устройств, использующих светодиоды для спектрального анализа [26-28]. Светодиоды обладают серьёзными преимуществами в качестве источников света для спектроскопии по сравнению с традиционными лампами. Они обладают более высоким коэффициентом полезного действия и способны давать гораздо более высокую яркость. Однако широкое использование светодиодных анализаторов всё ещё ограничено недостатком научной базы и практического опыта их разработки.

Среди новых систем детектирования стоит отметить малобюджетные пироэлектрические детекторы и миниатюрные спектрометры высокого разрешения на их основе [29-31]. Основой пироэлектрических детекторов является линейка светочувствительных элементов, покрытых плёнкой пироэлектрического материала – цирконат-титанат свинца, которые генерируют сигнал напряжения при изменении температуры. Пироэлектрические детекторы производятся для

определённых, достаточно узких спектральных участков в БИК- или ИК-области. Характерным примером микроспектрометра является настраиваемый интерферометрический сенсор (НИС) Фабри-Перо (*франц.* Fabry-Pérot), рассчитанный на регулируемый относительно узкий спектральный интервал с использованием технологии микроэлектромеханических систем (МЭМС; *англ.* microelectromechanical systems, MEMS) [32]. В качестве примера можно привести изделие фирмы «InfraTec GmbH» (г. Дрезден, ФРГ) [33]. НИС является продуктом микромашиностроения и состоит из движущегося и неподвижного зеркал, нанесённых на плёнку толщиной 300 мкм. Противоотражательные слои, оптимизированные для определённого интервала длин волн, например 8–10 мкм, наносятся на задние стороны обоих зеркал и подавляют отражения ими. Спектральное пропускание полученного оптического фильтра (функция фильтрации) характеризуется узкополосной резонансной кривой. Максимальное пропускание наблюдается при резонансной длине волны и может быть настроено путём электростатического изменения расстояния между зеркалами. Разрешающая способность НИС Фабри-Перо определяется отражающей способностью зеркал, их качеством и потерями из-за поглощения. При 50%-ном пропускании фильтра она обычно составляет порядка 200 нм. Пропущенное излучение детектируется пироэлектрическим детектором.

Стремительно развивается производство мини- и микроспектрометров для спектральных областей УФ-Вид и БИК, снижается их стоимость, что позволяет разрабатывать ОМС на их основе. В качестве примеров можно привести микроспектрометры «Qmini» германской фирмы «rgb photonics» [34], продукцию «Spectral Engines» (Финляндия) [35], а также «MicroNIR PAT-U» фирмы «VIAVI Solutions» (Италия) [36]. Текущие достижения оптических технологий создают предпосылки для расширения области применения спектроскопии среднего ИК и для её возрастающего использования в промышленных и полевых приложениях.

Изменения претерпевают и подходы к спектральному анализу. Комбинирование двух и более различных методов спектроскопии становится заметной тенденцией. Например, комбинации методов используют для того, чтобы

охватить различные аспекты процесса в АКТП [37-40]. Такое сочетание называют *комплементарным*. Другая ситуация возникает, когда спектральные данные различных источников соединяют в одной модели, чтобы улучшить аналитические возможности комбинированного метода, например точность определения анализируемого компонента. Однако желаемый *синергетический эффект* (когда комбинация работает лучше, чем каждый индивидуальный метод в отдельности) далеко не всегда сопровождается объединением. Например, комбинация ИК- и КР-спектроскопии не принесла никакого выигрыша [39], а при совместном использовании данных КР- и БИК-спектроскопии в работе [40] была повышена точность определения лишь одного из двух анализируемых параметров. Пока не существует общепринятого теоретического объяснения синергии при слиянии данных, которое позволило бы получать предсказуемый результат. Тем не менее, был разработан ряд алгоритмов эффективного анализа объединённых данных, включающих блоки разных аналитических методов или стадий технологического процесса [41-45].

Частным случаем дополненных данных являются *трёхмодальные* [46] (*англ. three-way*) данные и данные более высокой модальности. Такие данные получаются в результате работы определённых методов, таких как двумерная флуоресценция [47], пространственно-разрешённые измерения и комбинированные методы [48]. К последним относятся, например, высокоэффективная жидкостная хроматография с УФ-детектором и термогравиметрический анализ с ИК-анализатором продуктов. Единичным измерением в таких методах является матрица данных. В двумерной флуориметрии она представлена *спектром возбуждения-эмиссии* (*англ. excitation-emission matrix*). Набор измерений (в частности градуировочных образцов) образует в этом случае трёхмерный массив, подлежащий анализу специальными методами. Другим распространённым примером таких данных является пространственно-разрешённая спектроскопия, использующая набор (от нескольких до нескольких миллионов (как в спектральных изображениях [49]) спектральных каналов, имеющих различное пространственное расположение. Пространственная связь между каналами в таких методах также важна, она может

использоваться в моделировании или для представления его результатов. Анализ таких данных требует специальных методов, одним из которых является метод *развёртывания*, использующий преобразование трёхмерного массива данных в двумерный посредством стыковки составляющих его матриц в одну [50]. Трёхмодальные данные могут использоваться для разработки ОМС, как это показано в разделе 2.2.1.

1.1.4 Анализ спектральных данных

Эволюция методов анализа данных тесно связана с развитием приборной базы оптической спектроскопии. За появлением новых аналитических методов и устройств, к примеру, двумерной флуориметрии или спектральных изображений следует развитие математического аппарата и ПО для эффективного анализа новых типов данных. Но есть и противоположные примеры, когда анализ данных играл решающую роль в судьбе спектрального аналитического метода. Блестящей иллюстрацией служит регрессионный метод проекции на латентные структуры (ПЛС, *англ.* partial least-squares, PLS), ныне широко используемый для многомерной градуировки на спектральных данных. Его публикация в 1982 году [51] и последующее развитие и популяризация в многочисленных исследованиях [52,53] превратили давно известную БИК-спектроскопию в один из наиболее востребованных инструментов как для лабораторного анализа, так и в АКТП. До появления алгоритма ПЛС-регрессии БИК-спектры считались малопригодными для количественного анализа из-за сильного наложения характеристических полос и интенсивного фонового сигнала.

Извлечение химической информации, содержащейся в спектральных данных реальных образцов, часто осложнено наложением аналитических сигналов компонентов, неомогенностью и турбулентностью среды (особенно для данных процесса), шумом, приборными артефактами и другими факторами. Поэтому оптимальный анализ данных спектроскопии требует разработки специальных математических методов и алгоритмов. Эта прикладная дисциплина, получившая название *анализ многомерных данных* или хемометрика, является важной частью большинства современной аналитических методов на основе оптической

спектроскопии.

Сегодня хемометрика широко востребована и переживает этап активного развития. Ежегодно публикуются сотни статей, использующих методы анализа многомерных данных, в более чем 50-ти научных журналах, включая такие наименования как «Analytical Chemistry», «Analyst», «Analytica Chimica Acta», «Sensor and Actuators B: Chemical», «Talanta» и др. [46,54]. Два журнала непосредственно посвящены проблемам хемометрики: «Journal of Chemometrics» и «Chemometrics and Intelligent Laboratory Systems». Наиболее обсуждаемыми на сегодня темами в хемометрике являются: анализ пространственно-разрешённых данных и спектральных изображений, анализ данных многомодальных методов, анализ больших объёмов данных, хемометрика процессов, спектральное разделение смесей, факторный анализ данных, разработка робастных методов градуировки и классификации, построение глобальных моделей, поддержка и обновление практически работающих моделей и многие другие. Рассмотрению основных методов хемометрики в связи с настоящим исследованием посвящён раздел 1.3.

1.1.5 Децентрализация и персонализация анализа

Появление процесс-анализаторов и АКТП изменило сложившуюся десятилетиями структуру традиционного анализа. Основным следствием этого изменения является децентрализация. Место химика-аналитика занимает оператор процесса, который получает результат анализа на рабочем месте, без отправки образцов в общую лабораторию. Возникнув на производстве, эта тенденция начинает распространяться на другие сферы человеческой деятельности. Химический анализ сохраняет свои важнейшие функции в традиционных областях, таких как анализ пищевых продуктов, лекарств и медицинских образцов. Однако наряду с этим появляются новые приложения, социальная и экономическая необходимость которых может быть существенно ниже, например увеличение эффективности бизнеса или качества жизни отдельного человека. То есть анализ начинает применяться там, где раньше – по причине труднодоступности – приходилось обходиться без него.

Этот процесс только начинается. На сегодня существует масса потенциальных приложений, которые при господствующей на сегодня методологии химического анализа ещё не охвачены. Все чаще возникает потребность в анализе различных обобщенных показателей химического состава, таких как питательная ценность продукта, общая масса добавленного ингредиента, аутентичность изделия и многие другие. Нарастает использование оптических спектральных методов в областях, не считавшихся ранее сферой химического анализа, таких как диагностическое исследование биологических тканей, мониторинг среды процесса, сортировка материалов, удалённый анализ и т. д.

Потенциал для расширения круга анализируемых объектов заключается в дальнейшей «демократизации» анализа за счёт удешевления, миниатюризации и автономизации аналитических устройств. Флагманом такого развития является оптическая спектроскопия. Значительное повышение доступности оптических анализаторов создаст новые области применения и потребительские рынки. Малобюджетный анализ может охватить уровень малого и среднего бизнеса, стать инструментом для потребительских инициативных обществ по контролю за качеством продукции и окружающей среды, проникнуть в школьное образование и занять многие другие ниши. Экстраполируя эту тенденцию в будущее, легко представить себе новую категорию бытовых и индивидуальных спектральных анализаторов. Примером таких устройств в сфере персональной диагностики уже являются индивидуальные оптические пульсоксиметры, доступные в аптечных сетях [55]. Областью активной разработки являются неинвазивные индивидуальные глюкометры [56]. В целом наблюдается изменение отношения к химическому анализу как к экспертной области, и возникают предпосылки для её сначала более широкого, а в перспективе – и массового использования.

Персонализация анализа существенно облегчается современным уровнем компьютеризации, широкой доступностью мобильных телефонов и устройств, оснащённых мощным процессором и даже оптическим детектором – встроенной камерой. Ярким примером реализации оптических анализаторов на базе смартфонов является созданный в Санкт-Петербургском государственном

университете (СПбГУ) прототип миниатюрного анализатора содержания креатинина в моче [57]. Обзор подобных устройств дан в работе [58].

Промышленная реализация персональных оптических анализаторов безусловно связана с решением ряда технических проблем, с неизбежным компромиссом между ценой анализатора и качеством данных. В этой связи всё большее значение приобретает развитие новых подходов к разработке спектральных анализаторов и методов анализа выдаваемых ими данных, а также соответствующее ПО.

1.2 Оптические мультисенсорные системы

1.2.1 Определение ОМС

В настоящей работе вводится понятие оптической мультисенсорной системы как нового направления в развитии аналитической спектроскопии и нового класса спектрометрических устройств.

ОМС – это аналитическое устройство, включающее набор из двух и более оптических сенсоров (сенсорных каналов), оптимизированных под определённое приложение. Под термином «сенсор» здесь и далее в работе понимается *химический сенсор* по номенклатуре ИЮПАК [59]. В качестве информационных каналов ОМС выступают интегральные оптические сигналы на определённых интервалах длин волн, которые в общем случае могут перекрываться. Недостаток селективности отдельных каналов компенсируется на стадии обработки данных применением многомерной градуировки и других алгоритмов хемометрики.

Необходимость введения нового термина для спектрометрического по своей сути устройства обусловлена его концептуальными отличиями как от традиционной лабораторной, так и от современной процесс-спектроскопии, с которыми прочно ассоциируется термин «спектрометр». Кроме того, терминология отражает новую ориентацию оптической спектроскопии на разработку специализированных анализаторов в противовес универсальным.

Термин *мультисенсорная система* был введён в электрохимии для устройств типа «электронный нос» [60] и прочно закрепился в работах Ю.Г. Власова и

основанной им группы «электронного языка» СПбГУ [1,61]. Потенциометрический «электронный язык» тоже основан на оптимизации массива неспецифических сенсоров, обладающих различной чувствительностью к компонентам исследуемого образца (перекрестной чувствительностью) [62]. Таким образом, ОМС и «электронный язык» концептуально сходны, а различие связано с принципом действия составляющих их сенсоров, что и обосновывает предложенную терминологию.

Ярким примером природной ОМС является глаз. Воспринимаемое нами множество оттенков является результатом обработки сигналов трёх перекрестно-чувствительных оптических сенсоров низкой селективности [63]. Эволюция оптимизировала спектры чувствительности отдельных сенсоров глаза под ориентирование в среде обитания, поэтому цветовое восприятие у различных видов животных и человека не совпадает [64].

В данном терминологическом контексте оптические мультисенсорные системы можно охарактеризовать как устройства типа «электронный глаз».

1.2.2 Отличительные особенности

Являясь воплощением нового – мультисенсорного – подхода в оптическом спектральном анализе, ОМС обладают рядом отличительных особенностей, позволяющих выделить их в отдельный класс аналитических устройств. Изначальная специализация под определённое приложение, заложенная в конструкцию ОМС, является их главным концептуальным отличием от других спектрометров. Существует и ряд нестрогих особенностей ОМС, отличающих их от традиционных спектрометров, которые следует рассматривать как ориентировочные, отражающие тенденции в их разработке и использовании (табл. 1-1).

Специализация ОМС заключается в использовании минимально необходимого числа составляющих спектр *оптических каналов* на подобранных исходя из приложения участках длин волн вместо сотен каналов спектрометра, равномерно распределённых во всём технически доступном диапазоне для обеспечения высокой *разрешающей способности* [24, с. 91] и универсальности

анализа. В спектрах высокого разрешения переменные считаются монохроматическими или разрешёнными, то есть соответствующими узкому интервалу длин волн, не перекрывающемуся с интервалами соседей. На практике это не всегда так, но разрешённость каналов является весьма желательной. В ОМС же спектральные интервалы каналов в общем случае широки, они могут отстоять друг друга или, напротив, сильно перекрываться, так что сам термин «разрешающая способность» становится не всегда применимым. Высокое спектральное разрешение, важное для идентификации компонентов, не является определяющим в задачах количественного анализа. Напротив, возможность использования наиболее информативных длин волн и участков спектра позволяет избегать областей шума и постороннего поглощения. Уменьшение спектрального разрешения имеет и положительную сторону, приводя к снижению уровня шума, что благоприятно сказывается на точности (прецизионности и правильности) анализа. Было показано, что использование хемометрики позволяет в значительной мере скомпенсировать потерю разрешения при переходе от лабораторного ИК-анализа к мультисенсорным технологиям [65].

Селективность ОМС к анализируемым веществам, когда на отдельных оптических каналах наблюдается чистый сигнал компонента, возможна, но в общем случае не ожидается и не преследуется как цель. Например, использование в ОМС светодиодов с их относительно широкой полосой излучения, ширина которой к тому же быстро растёт с длиной волны максимума [66], заведомо предполагает низкую селективность анализатора. Как и в случае электрохимических сенсоров, тут можно говорить о перекрёстно-чувствительных каналах [62]. При этом анализ автоматически подразумевает использование инструментария хемометрики.

Скорость измерения важна для онлайн-анализа, поэтому у процесс-спектрометров она в целом выше, чем у лабораторных, и как правило составляет несколько миллисекунд, например в спектрометрах с диодно-матричным детектором (ДМД; *англ.* diode-array detection, DAD) [67]. Отказ от использования дисперсии света и движущихся элементов в ОМС некоторых типов (раздел 1.2.6)

создаёт потенциал для увеличения скорости анализа на несколько порядков, что востребовано некоторыми приложениями, такими как поштучная сортировка зерновых культур по результатам индивидуального спектрального анализа каждого зерна «на лету» [68].

Таблица 1-1. Сравнение типичных параметров ОМС и традиционной оптической спектроскопии.

№	Параметр анализатора	Спектрометр	ОМС
<i>Потребительские качества</i>			
1	Область применения	универсальный	специализированный
2	Автономность	нет	возможна полная
3	Габариты	настольный	миниатюрный
4	Портативность	стационарный	переносной
5	Место проведения анализа	специализир. лаборатория	произвольное
6	Подготовка пользователя	специалист	не требуется
7	Стоимость	выше	ниже
8	Техническое обслуживание	требуется регулярно	минимальное/не требуется
<i>Технические характеристики</i>			
9	Селективность	высокая	низкая
10	Число каналов	сотни	единицы
11	Спектральная шкала	равномерная	индивидуальная
12	Спектральная область	максимально широкая	индивидуальная
13	Оптические каналы	разрешённые	могут перекрываться
14	Разрешающая способность	максимальная (избыточная)	не всегда применимо
15	Скорость измерения	до миллисекунд	до микросекунд
16	Стандартизация прибора	аппаратная	программная
<i>Математическое и программное обеспечение</i>			
17	Использование хемометрики	желательно	обязательно
18	ПО и компьютер	внешние	встроенные
19	Прогностическая модель	локальная	глобальная

Проблема возможного отсутствия в ОМС аппаратной (то есть заложенной в конструкцию прибора, как в лабораторных спектрометрах) стандартизации решается применением математических методов. Так, идентичность результатов анализа одного и того же образца двумя разными ОМС-анализаторами, несмотря на возможные различия измерений нестандартизированными приборами, может быть достигнута использованием методов *переноса модели* [69-71].

Таким образом, использование хемометрики становится практически обязательным для построения прогностических моделей на данных ОМС. В случае

массового использования ОМС для какого-либо общего приложения, например в анализе популярных пищевых продуктов, оснащение каждого устройства собственной (локальной) градуировочной моделью может быть нецелесообразно. Альтернативой является построение универсальных моделей, называемых также глобальными [72-74]. Глобальная модель может быть размещена «в облаке» для использования распределёнными сенсорными системами. Математическое и программное обеспечение традиционного спектрометра, как правило, включает блок обработки данных, выдающий результат анализа (*прогноз*) с использованием заранее установленной хемометрической модели и располагается на соединённом с прибором компьютере. ОМС же, напротив, ориентируется на полную автономию и использует микрокомпьютерную технику, встроенную в корпус системы, которая рассчитывает и выдаёт результат.

Существенное упрощение конструкции с сохранением спектроскопического принципа сбора данных в ОМС приводит к многократному снижению стоимости и габаритов анализатора. Если рассмотреть в качестве примера БИК-спектрофотометры, то стоимость прибора при переходе к ОМС может снижаться в сотни раз, а вес и габариты – в десятки. Автономность ОМС достигается также отказом от постоянного подключения к внешнему компьютеру и необязательностью сетевого электропитания. Необходимую для работы устройства мощность можно обеспечить за счёт аккумуляторов или батарей. Автономность и малогабаритность ОМС определяют их портативность – одно из наиболее важных потребительских качеств мультисенсорных систем, которое может создавать новые области их применения. Доступность ОМС в совокупности с их миниатюрностью позволяет создавать распределённые сети анализаторов, использующие множество устройств для обеспечения более полного контроля. Это может быть установка сенсоров в различных узлах производственной линии или пространственное покрытие территорий для экологического мониторинга.

Важным преимуществом ОМС является высокая аппаратная адаптивность под новое приложение. Для технической перенастройки оптимального набора длин волн достаточно произвести замену элементов, определяющих оптические

свойства каналов системы, например светодиодов или фильтров. Это создаёт возможности для модульного дизайна ОМС, в котором узкая специализация парадоксально сочетается с универсальностью.

В целом оптический мультисенсорный анализ предлагает бóльшую гибкость по сравнению со спектроскопическим. Конструкционное разнообразие, связанное с возможностью использования современных светодиодов, зондов, световодов и детекторов, позволяет адаптировать анализ к самым разным образцам и условиям, а также проводить его на различном расстоянии от объекта.

1.2.3 ОМС и одноканальные сенсоры

Простые (одноканальные) фотометрические сенсоры широко используются в химическом анализе. Например, окрашенный компонент в истинном растворе может быть определён по поглощению в видимой области, а взвешенный компонент коллоидного раствора – по рассеянию света, как в методах турбидиметрии и нефелометрии [24, с. 186]. В каждом из этих примеров часто используется единичная длина волны, подобранная для обеспечения селективности анализа. При недостатке селективности используют методы разделения или химическую модификацию образца, облегчающую фотометрический анализ, например перевод аналита в окрашенный комплекс. Эти лабораторные методы анализа могут быть трудоёмки и не всегда дают желаемую точность, особенно в сложных смесях.

Основная проблема простой фотометрии заключается в том, что информации, поставляемой единственным каналом оптического сенсора, часто бывает недостаточно для анализа сложных систем. Естественным решением в таком случае является добавление дополнительного информационного канала (или каналов), то есть переход от фотометрии к спектрофотометрии. Рис. 1-2 схематически показывает преимущество двух каналов в задаче по дискриминации двух классов объектов. Линейная модель на плоскости полностью разделяет две группы объектов, не разделённых на каждом из каналов в отдельности (рис. 1-2 а). В более сложном случае может быть использована нелинейная дискриминация (рис. 1-2 б).

Практической иллюстрацией преимуществ многоканального измерения может быть оптический сенсор на содержание воды в порошках, основанный на интенсивной полосе поглощения воды в области 1940 нм [75]. В качестве источника излучения в нём используется светодиод, испускающий на этой длине волны, а диффузно отражённый порошком свет улавливается и передаётся на фотометрический детектор. Попытка создать градуировочную модель для содержания воды по показаниям такого сенсора приведёт к довольно низкой точности из-за сильных случайных колебаний регистрируемого сигнала, связанными с различным состоянием поверхности и морфологической неомогенностью порошка (так называемый эффект рассеяния) [76].

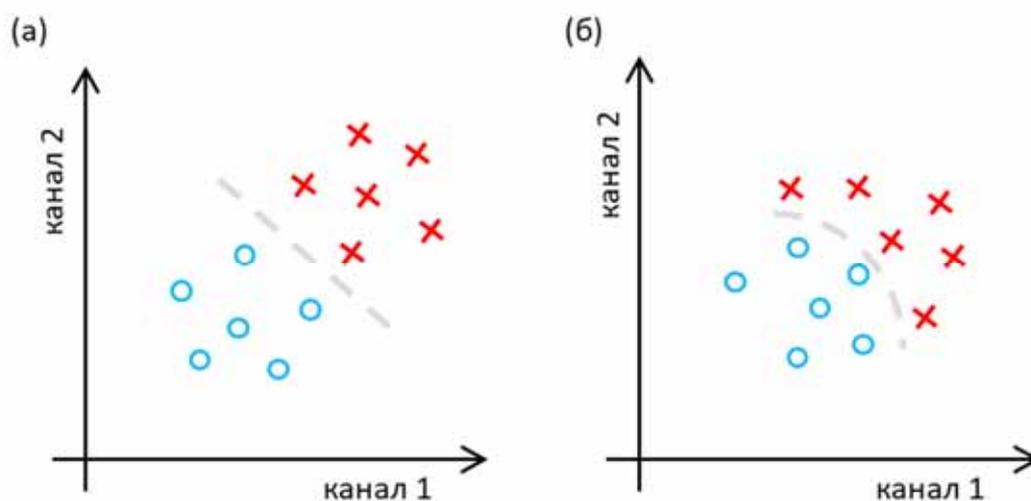


Рисунок 1-2. Схематическая иллюстрация дискриминантного анализа данных двухканальной мультисенсорной системы: (а) линейное и (б) нелинейное разделение классов на плоскости, составленной измерениями двух информационных каналов.

Поскольку интенсивность рассеяния относительно большими частицами порошка практически не зависит от длины волны в БИК-области, его можно учесть, добавив второй светодиод на 1300 нм. Детектируя сигналы попеременно работающих источников, можно скомпенсировать эффект рассеяния путём вычитания второго сигнала из первого. Для дальнейшего объяснения принципов создания ОМС представим, что на полосу воды накладывается полоса другого компонента, концентрация которого в исследуемом порошке может меняться. Для компенсации этого фактора в сенсорную систему необходимо ввести ещё один светодиод. Если мешающий компонент неизвестен, необходимую длину волны (и

другие параметры спектра излучения) третьего светодиода придётся определить экспериментально. Без полного разделения смеси это можно сделать, используя полные спектры набора анализируемых образцов, составленного по определённому плану. Но даже зная спектры чистых компонентов исследуемой смеси, умозрительно определить наилучшие светодиоды для создаваемой мультисенсорной системы едва ли возможно. Их общее число и оптические характеристики – предмет математической оптимизационной задачи, подходы к решению которой будут рассмотрены в главе 2.

Добавление третьего и последующих информационных каналов в оптимизированную ОМС открывает возможность использования многих алгоритмов хемометрики, как для предобработки данных, так и для построения прогностических моделей. Следует подчеркнуть, что число каналов ОМС определяется именно информационными каналами, то есть независимыми источниками информации, составляющими переменные в выходных данных, доступных для дальнейшего математического анализа. Техническое наличие множества оптических сенсоров не является определяющим признаком ОМС. К примеру, если из всех оптических сенсоров системы используется только один или если отклики сенсоров усредняются в один, то анализатор будет одноканальным, независимо от его технической сложности. Даже спектрометр высокого разрешения, отградуированный на определение концентрации по единичной длине волны или по интегралу пика поглощения, функционирует как простой одноканальный сенсор.

1.2.4 Архитектура и принцип работы ОМС

Мультисенсорную систему можно представить как устройство для сбора, передачи и преобразования информации, то есть в терминах *информационного потока* (рис. 1-3). Представленная схема в равной степени описывает информационный поток любого метода спектроскопии. В данном контексте она удобна для иллюстрации обсуждаемых далее подходов к разработке ОМС.

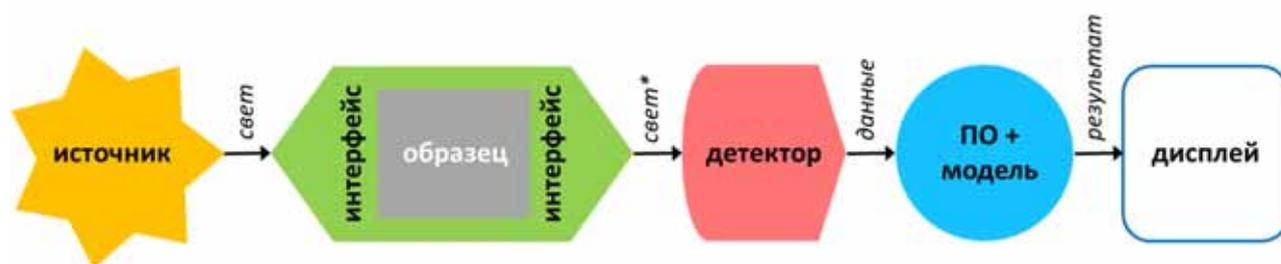


Рисунок 1-3. Схема информационного потока ОМС («*» обозначает преобразование).

Исходным носителем информации является электромагнитное излучение оптического спектра – свет, испускаемый источником или набором источников. В состав источника могут входить оптические устройства, преобразующие белый свет в монохроматический, например оптические фильтры или монохроматор. В ходе анализа источник может работать в непрерывном или импульсном режиме.

Сбор информации происходит при взаимодействии с образцом, в результате чего излучение претерпевает изменение спектрального состава. *Интерфейс измерения* – это набор оптико-механических приспособлений, обеспечивающих сбор релевантной информации при взаимодействии облучающего света с образцом за счёт преобразования светового потока до и после образца, а также формы самого образца в момент анализа. Интерфейс измерения может быть стандартным или специально созданным под определённую задачу. Интерфейс отвечает за *геометрию измерения* – длину оптического пути, угол измерения, расстояние до объекта и другие параметры, влияющие на распространение света в образце, а также за способ приёма изменённого светового потока, который затем поступает на *детектор*. Примерами часто используемых интерфейсов измерения являются кювета, пропускная ячейка, зонд и оптическое окно [77]. К более специфическим относятся гониометр (устройство, позволяющие изменять угол наблюдения), фотометрический шар Ульбрихта, а также более сложные устройства, такие как спектральная камера или конфокальный микроскоп. Анализ может проходить при контакте образца с элементами выбранного интерфейса или на расстоянии.

Стандартными атрибутами интерфейсов измерения являются линзы, НПВО-кристаллы, зеркала и оптические фильтры. Так, фильтр может быть установлен между образцом и детектором для отсека ненужных спектральных областей или линий, таких как сигнал возбуждающего лазера при КР-спектроскопическом или

люминесцентном анализе [24, с. 88].

При прохождении детектора носитель информации меняется. Свет сначала преобразуется в электрический аналоговый сигнал, который затем оцифровывается (в схеме на рис. 1-3 аналогово-цифровой преобразователь относится к детектору). Данные информационных каналов на выходе из детектора являются *аналитическим сигналом* ОМС.

Завершающим шагом в информационном потоке ОМС является прогноз. Прогноз – применение предварительно построенной математической (*прогностической*) модели к новому измерению для преобразования аналитического сигнала в значение концентрации (или обобщённого показателя химического состава), принадлежность к классу или ответ «да/нет», то есть в конечный результат качественного или количественного анализа. Например, если целью анализа является определение концентрации компонента образца, то нужна *градуировочная* модель. Таким образом, прогностическая модель является обязательным элементом информационного потока. Используемые в ОМС хемометрические модели являются по своей природе цифровыми, то есть требуют наличия ЭВМ.

Компьютер является важнейшим элементом ОМС. Установленное на нём ПО несёт функции управления, в том числе отвечает за прогноз и выдачу результата. Эти функции могут быть перенесены на внешнюю ЭВМ, которая обменивается данными с мультисенсорной системой по кабелю или беспроводным способом. В автономных ОМС используют встроенный компьютер с собственным дисплеем на корпусе анализатора, который одновременно служит и для управления устройством, и для отображения результата прогноза. В некоторых случаях результат может отображаться при помощи цветовых сигналов, например, зелёный – «да», красный – «нет», жёлтый – «спорно», белый – «ошибка измерения».

Существуют важные частные случаи ОМС, когда схема на рис. 1-3 может быть упрощена. Так, правая половина схемы после образца может заменяться человеческим восприятием – глазом и мозгом (аналоги «детектора» и

«компьютера»). Например, цвет видимой флуоресценции освещённого лазером образца показывает наличие определённого химического компонента, а интенсивность излучения – его концентрацию. В другом случае источником света может быть естественная, например солнечная, радиация. Более того, источник света сам может являться образцом, например, при спектральном анализе излучающих объектов – ламп, жидкокристаллических дисплеев и т. д.

1.2.5 Конструкционные элементы ОМС

В оптической спектроскопии существует широкий выбор конструкционных элементов, используемых в качестве источников света, интерфейсов измерения и детекторов. Выбор элементов при разработке ОМС зависит от природы образца, используемого спектрального метода, оптимальной геометрии измерения, а также от специфических требований к анализатору, таких как скорость измерения, стоимость, встраиваемость и размер. Современные технические элементы, используемые для конструирования ОМС, а также типы используемого ПО и компьютеров сведены в табл. 1-2.

Таблица 1-2. Элементы конструкции ОМС.

Блок	Конструкционный элемент
Источник	светодиод [26-28,75,78,79,80], лазерный диод [81], лазер, фильтр [82,83], дифракционная решётка [65], монохроматор [81], призма, лампа [77, с. 74]: накаливания [65], галогенная [84,85], дейтериевая, ртутная
Интерфейс	кювета [85], зонд на пропускание [65], зонд диффузного отражения (ДО) [86], НПВО-зонд [65,87,88], сменный НПВО-элемент [65], пропускная ячейка [84], световод [65,78,79,85,89,90], оптическое окно [91], линза, фильтр [81,88], диафрагма [85], шар Ульбрихта [86], гониометр, монохроматор [81]
Детектор/ Спектрометр	фотодиод [75,78], ДМД [85,86], ПЗС [89], ММС [84], пироэлектрический [65], микроспектрометр [91], НИС [65]
ПО для прогноза	стандартное, собственной разработки [78,91], интерфейс прикладного программирования (ИПП)
Компьютер и дисплей	персональный компьютер, микрокомпьютер [75], микроконтроллер [78]

Примечание: ссылки дают примеры использования различных конструкционных элементов в действующих ОМС или в процессе их разработки.

При всем их разнообразии различные источники излучения и системы детектирования образуют устойчивые пары, технические характеристики которых, как правило, и определяют разбиение анализаторов на спектральные области и методы, такие как УФ, Вид, БИК и ИК. Интерфейсы измерения более универсальны и часто применимы с различными спектрометрами и видами молекулярной спектроскопии. Так, кварцевая кювета может быть использована для регистрации любых оптических спектров, кроме ИК.

Для хранения модели и прогноза в автономных ОМС обычно используются современные компактные микрокомпьютеры. Для управления устройством всё чаще используются микроконтроллеры, оснащённые необходимым программным обеспечением. ПО для прогноза может быть установлено как на встроенном, так и на внешнем компьютере и быть стандартным (доступным коммерческим продуктом) или собственной разработки. Последнее может быть необходимо, например, при создании высокооптимизированного сенсора. Альтернативным подходом к прогнозу является использование интерфейса прикладного программирования (ИПП) [92], которое может упростить устройство за счёт перенесения всех расчётов на удалённый сервер («в облако»), но соединение с моделью потребует доступа к сети Интернет.

Список конструктивных элементов в табл. 1-2, разумеется, не исчерпывающий. Он включает лишь основные элементы, использовавшиеся в практических приложениях настоящей работы и рассмотренные в литературных источниках.

1.2.6 Классификация ОМС

Классификация ОМС может быть основана на различных критериях. Она может базироваться на физическом принципе, заложенном в их основу. Основные методы оптической спектроскопии с точки зрения используемого физического эффекта приведены на рис. 1-4, где также выделены методы, наиболее часто используемые в современных ОМС.

С точки зрения геометрии измерения в ОМС может быть реализовано три основных *режима измерения* в зависимости от угла детектирования по отношению

к облучению: на пропускание (180°), на отражение (0°) и на рассеяние (под другим углом, например 90° в нефелометрии). Исключение составляет метод НПВО, представляющий отдельную категорию. Терминология этого деления условна, поскольку в каждом из указанных режимов изменение светового потока при прохождении через образец может быть комбинацией эффектов поглощения, рассеяния, люминесценции и КР.



Рисунок 1-4. Физические эффекты, используемые в оптической спектроскопии и в ОМС (жирный шрифт).

С точки зрения конструкции ОМС можно разделить на три типа в зависимости от того, какой узел информационного потока (рис. 1-3) отвечает за специализацию спектрального анализатора под приложение: источник, детектор или ПО (табл. 1-3).

Таблица 1-3. Классификация ОМС по конструкционному принципу.

Тип №	Узел специализации	Элементы и методы формирования каналов
1	Источник	светодиоды, лазерные диоды, лазеры, фильтры
2	Детектор	пирозлектрический, ПЗС, ДМД
3	ПО	выбор переменных или интервалов, усреднение

Примечание: ПЗС – прибор с зарядовой связью (англ. charge-coupled device, CCD)

Светодиодные ОМС (тип 1) представляют одно из наиболее привлекательных современных направлений разработки, совмещающая в себе простоту конструкции с широтой возможностей [23]. Сходная конструкция с источником излучения широкого спектра и автоматически сменяемыми фильтрами реализована, например, в анализаторах фирмы «NDC» (г. Дейтон, США) [82] и в ОМС для двумерной флуоресцентной спектроскопии «BioView», оснащённой

зондом для анализа процессов [83, с. 29]. Вместо светодиодов могут использоваться и мощные лазерные диоды, например во флуоресцентных ОМС [81]. Настраиваемые лазеры могут стать основой высокоинформативных и точных флуоресцентных и рамановских ОМС, особенно если основное препятствие к их широкому использованию – цена – будет снижаться.

ОМС второго типа определяются использованием специализированных детекторов в определённом узком спектральном диапазоне, который подбирается в зависимости от приложения. К таким детекторам прежде всего относятся миниатюрные пироэлектрические детекторы [29,30] для БИК- и ИК-областей, например НИС типа Fabry-Perót [33]. Однако, основой специализированных детекторов может служить и оптимизированный массив фотодиодов типа ДМД или ПЗС с индивидуально разработанным на уровне электроники способом считывания спектральной информации.

Третий тип ОМС чаще всего возникает при использовании в качестве основы готовых мини- и микроспектрометров, разрабатываемых в коммерческих или исследовательских целях. В последние годы наблюдается повышенный интерес к изготовлению «демократичных» спектроскопических анализаторов, в основном, для Вид- и БИК-областей. Дебю и др. [57] показали, что спектрометр, изготовленный из подручных материалов – старого CD-диска, зеркальца и пустой упаковки, может быть успешно использован для определения креатинина в моче. В качестве детектора и компьютера при этом используется камера смартфона. Существуют и другие примеры использования готовых бытовых устройств типа фотоаппарата или смартфона как основы импровизированных спектральных анализаторов [58,90]. Такой подход является проявлением запроса на индивидуализацию анализа, отмеченную в разделе 1.1.5. Спецификой спектрометров такого типа (будем называть их упрощёнными спектрометрами), помимо максимального облегчения конструкции за счёт отказа от встроенного источника света и других элементов, является недостаток стандартизации, что характерно и для процесс-спектрометров (раздел 1.1.2). Это касается не только штучных исследовательских прототипов, но и более качественных (и не столь

дешёвых) коммерческих марок, таких как продукция фирм «rgb photonics» [34] или «Spectral Engine» [35]. Упрощённые спектрометры являются, по сути, полуфабрикатами, и их использование в практическом анализе сопряжено с достраиванием недостающих компонентов, таких как ПО для *низкоуровневой* (до сохранения в памяти компьютера) обработки данных, восполняющей недостаток стандартизации, в том числе неравномерности и избыточности шага сбора спектральных данных. Именно на этапе низкоуровневой обработки данных происходит выбор переменных и их усреднённых интервалов, а значит, может быть реализована специализация ОМС третьего типа (табл. 1-3), то есть её программная оптимизация под определённое приложение. В качестве примера можно привести сенсор фирмы «Blue Ocean Nova» (ФРГ) для онлайн-определения воды в порошкообразной пищевой добавке в ходе её производства (детали не разглашаются по требованию производителя) [91].

Принципы разработки ОМС всех типов подробнее рассмотрены в главе 2.

1.3 Анализ спектральных данных и данных ОМС

1.3.1 Хемометрика как метод анализа спектральных данных

Классическим считается определение хемометрики, данное Массартом: «Хемометрика – это химическая дисциплина, применяющая математические, статистические и другие методы, основанные на формальной логике, для построения или отбора оптимальных методов измерения и планов эксперимента, а также для извлечения наиболее важной информации при анализе экспериментальных данных» [93].

Методы хемометрики используются во многих областях, но именно в аналитической химии, и прежде всего в спектроскопии они нашли наиболее широкое распространение [54]. По мнению основателей Российского хемометрического общества А.Л. Померанцева и О.Е. Родионовой, начавшееся в 1980-х годах широкое распространение многомерного анализа данных стало научным событием парадигматического масштаба, особенно в химии и прикладной математике [94].

В оптической спектроскопии существует связь между селективностью метода и важностью использования методов анализа многомерных данных (рис. 1-5). Так, в линейчатых спектрах высокоселективных методов, таких как атомно-абсорбционная спектрометрия, существуют чистые отклики анализируемого компонента, что позволяет использовать соответствующие переменные для построения достаточно точных одномерных градуировочных и других моделей. Использование хемометрики в этом случае приветствуется, но является скорее факультативным. Спектры средней ИК-области, как и КР-спектры, содержат хорошо выраженные полосы поглощения. Но из-за их большого числа вероятность наложения пиков велика, и использование многомерного подхода рекомендуется в большинстве практических задач. Применение хемометрики становится критичным для точности анализа, основанного на БИК-спектрах, где «чистые» переменные практически отсутствуют из-за большой ширины полос, а также возможного влияния эффектов поглощения воды и рассеяния. Особый случай очень низкой селективности представлен количественным анализом спектральных данных, основанным на небольших различиях форм размытых пиков и профилей рассеяния компонентов смеси [95]. Пример такого анализа дан в разделе 5.1, где был разработан метод определения жира и общего белка в молоке в видимой и коротковолновой (КВ)БИК-области спектра. Анализ данных заведомо не может компенсировать лишь полное отсутствие селективности метода, то есть его неспособность распознать анализируемый компонент в спектрах смесей.

Традиционные спектральные данные характеризуются *коллинеарностью* – наличием большого числа коррелирующих между собой переменных. Коллинеарность является, с одной стороны, преимуществом спектральных данных, предоставляя «дублирующие» измерения на соседних длинах волн, а значит и статистически более достоверную информацию. С другой стороны, извлечение этой информации представляет собой нетривиальную задачу с учётом сложности реальных образцов, имеющих как правило неизвестный или не полностью известный состав.

Селективность	Метод спектроскопии	Использование хемометрики
 очень высокая	атомно-абсорбционная	желательно
 высокая	ИК, КР	важно
 низкая	УФ/Вид и БИК, флуориметрия	очень важно
 очень низкая	методы, основанные на рассеянии света	критично
 отсутствует	ошибочный метод	не работает

Рисунок 1-5. Селективность метода оптической спектроскопии и важность использования хемометрики.

Математической основой анализа многомерных спектральных данных является проекция – используемое многими методами хемометрики алгебраическое разложение исходной матрицы данных \mathbf{X} ($n \times m$), где n – число объектов (образцов) в строках, а m – число переменных в столбцах, на две комплементарные составляющие: матрицу счетов \mathbf{T} ($n \times a$) и матрицу нагрузок \mathbf{P} ($m \times a$) в соответствии с уравнением (1.1):

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^a \mathbf{t}_i \mathbf{p}_i + \mathbf{E} \quad (1.1)$$

где $a \ll n, m$ – число латентных переменных (ЛП) модели, а \mathbf{t}_i ($n \times 1$) и \mathbf{p}_i ($m \times 1$) – векторы, составляющие матрицы \mathbf{T} и \mathbf{P} соответственно, а \mathbf{E} ($n \times m$) – матрица остатков (ошибок). Факторная проекция схематически представлена на рис. 1-6.

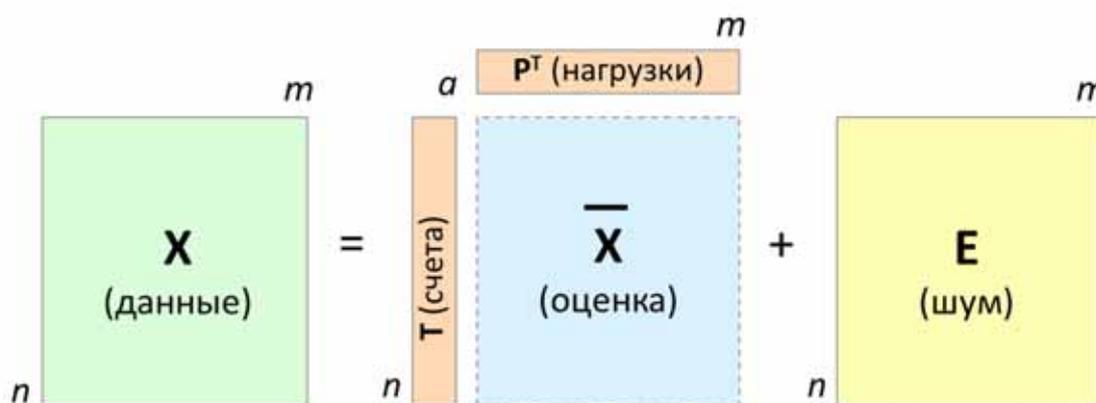


Рисунок 1-6. Схема факторной проекции многомерных данных.

В результате преобразования матрица данных \mathbf{X} проецируется на факторное

пространство \mathbf{T} меньшей мерности a , равной выбранному числу ЛП. Произведение матриц \mathbf{T} и \mathbf{P}^T , называемое оценкой ($\bar{\mathbf{X}}$), воспроизводит наиболее важную дисперсию в \mathbf{X} , оставляя нерелевантную информацию (шум) в матрице остатков \mathbf{E} . Проекция означает переход от анализа исходных переменных матрицы спектров \mathbf{X} (спектры как правило формируют строки матрицы – объекты, а длины волн соответствуют столбцам – переменным) к анализу латентных (абстрактных) переменных пространства, оптимизированного под решение соответствующей аналитической задачи. Методы, использующие преобразование данных по уравнению (1.1), называют *проекционными*.

Можно выделить три основных группы проекционных методов в зависимости от решаемых ими задач и, соответственно, от принципов построения факторных пространств: факторный анализ, многомерная градуировка и классификация (дискриминация). Однако, как следует из данного выше определения, область применения хемометрики ими не ограничивается и традиционно включает в себя вопросы эффективного сбора спектральных данных, планирования эксперимента, пробоотбора, переноса моделей, онлайн-анализа процессов и другие практические проблемы спектроскопии.

Данные ОМС обладают рядом особенностей, отличающих их от традиционных спектров. К ним относятся: малое количество переменных, их неравномерное распределение вдоль спектральной оси, недостаточная стандартизация и потенциально большой объём измерений. Это накладывает определённые ограничения на использование распространённых в спектроскопии алгоритмов анализа данных и создаёт потребность в разработке новых подходов (раздел 2.4).

1.3.2 Предварительная обработка данных

Предварительная обработка (*предобработка*) спектров, то есть их модификация перед анализом, проводится для улучшения качества и информативности данных. Целью предобработки является улучшение точности анализа данных за счёт удаления нерелевантной информации – шума, фонового сигнала (базовой линии), неинформативных областей спектра при максимальном

сохранении полезной информации. В терминах статистики *информация* – это дисперсия переменных. Предобработка может также служить улучшению визуализации полезной информации, например при использовании дифференцирования или функций преобразования данных, таких как линеаризация. Используемые в настоящей работе методы предобработки данных представлены в табл. 1-4.

Таблица 1-4. Основные методы предобработки оптических спектров.

Предобработка	Устраняемые эффекты	Алгоритмы
Взвешивание	разница шкал/ значимости переменных	автошкалирование [53, с. 76]
Нормировка	переменный объём образца	СНВ ^а [96, с. 124]
Коррекция рассеяния	«эффект рассеяния»	МКР ^б [97]
Коррекция базовой линии	фоновое поглощение в спектрах	КЛК ^в [98]
Сглаживание	шум	СГ ^г [99], МСС ^д [86]
Первая/ вторая производные	неразрешённость полос, базовая линия	СГ [99]
Отбор переменных и объектов	неинформативные переменные или объекты, выбросы	ГА ^е [100], ИПЛС ^ж [101], КАПВ ^з [102]

Примечание: ^а стандартизация нормировкой вариации (СНВ; *англ.* standard normal variate, SNV); ^б коррекция множественного рассеяния или множественная коррекция рассеяния (*англ.* multiplicative scatter correction, MSC); ^в кусочно-линейная коррекция базовой линии; ^г метод сглаживания и дифференцирования по Савицкому-Голею (Savitzky-Golay); ^д метод скользящего среднего; ^е генетический алгоритм (*англ.* genetic algorithm, GA); ^ж интервальный ПЛС (*англ.* interval PLS, iPLS); ^з метод конкурентной адаптивной перевзвешенной выборки (*англ.* competitive adaptive reweighted sampling, CARS).

Центрирование (*англ.* mean centering) – метод предобработки, состоящий в вычитании из каждой строки матрицы данных \mathbf{X} вектора средних значений по всем переменным \mathbf{X} . Этот специальный метод (не указан в табл. 1-4) практически всегда используется как заключительный шаг в последовательности предобработок перед основным анализом данных.

Рассмотренные методы часто используются в ходе разработки ОМС для предобработки исходных полных спектральных данных. Однако возможности модификации данных самих ОМС сильно ограничены в силу их особенностей (см. разделы 1.2.2 и 1.3.1). Так, из-за малого числа переменных и неравномерности шага

к ним могут быть применены лишь методы взвешивания, нормировки и коррекции рассеяния (табл. 1-4), причём методы СНВ и МКР имеют смысл лишь при числе каналов не менее 3-х. Для двухканальной ОМС доступны лишь арифметические действия со значениями на разных каналах, полученными в одном измерении, или независимое преобразование переменных в каждом из каналов, например взвешивание методом *автошкалирования* – делением каждого значения переменной на её стандартное отклонение.

1.3.3 Отбор переменных и объектов

Отбор наиболее качественных и информативных данных (как объектов, так и переменных) составляет отдельную группу методов предобработки. Отбор переменных особенно важен при разработке мультисенсорных систем, а отбор объектов – для анализа данных действующих ОМС.

Ценность переменных, составляющих спектральные данные, для хемометрической, в частности градуировочной, модели не одинакова. Помимо полезной информации, переменная может содержать высоко- и низкочастотный шум, отклики других – нерелевантных – компонентов смеси, а также приборные и экспериментальные артефакты. Этот шум неизбежно проецируется в пространство латентных переменных, снижая прогностическую точность модели, а наличие варьирующихся сигналов нерелевантных компонентов повышает её сложность и снижает надёжность, что в конечном счёте тоже сказывается на точности анализа. «Плохие» переменные могут появиться в данных в результате проведённой предобработки, например автошкалирования [53, с. 76], которое, усиливая слабые сигналы, одновременно повышает шум. Это может свести на нет позитивный эффект взвешивания или нормировки [87].

Эффективным средством для снижения негативного воздействия вышеописанных факторов на модель является отбор наиболее полезных (или исключение наименее полезных) для моделирования переменных, что было подтверждено многочисленными исследованиями [102-111]. Разработано большое число различных методов отбора переменных [112,113].

Зачастую отбор переменных производится экспертом на основании его

априорных знаний природы анализируемого образца и метода анализа. Например, по результатам визуальной оценки спектров исключаются области преобладающего высокочастотного шума, нулевого сигнала, заведомо нерелевантных сигналов, и напротив, для моделирования выбираются области полос поглощения анализируемого компонента. Методы «ручного» отбора просты и эффективны, однако не исключают ошибки, связанной с человеческим фактором или наличием в данных не выявляемой визуально скрытой информации.

Другая группа методов направлена на повышение достоверности отбора переменных за счёт использования различных характеристик, получаемых из самих данных. Один из наиболее простых и эффективных подходов заключается в удалении всех переменных ниже определённого порога, например, если её стандартное отклонение ниже 5%. Однако этот подход не способен различать «хорошие» и «плохие» переменные с высоким значением дисперсии. Эта проблема частично преодолевается использованием более сложных статистических характеристик переменных [114,115]. К этой же группе можно отнести методы, использующие для отбора регрессионные коэффициенты и другие параметры предварительных полных моделей, как в алгоритме КАПВ [102].

Третья группа методов применяет комбинаторное тестирование различных сочетаний переменных в моделировании и определяет оптимальное на основе полученной погрешности прогноза. Поиск решения методом простого исчерпывающего перебора для спектральных данных, как правило, невозможен даже с современной вычислительной техникой по причине слишком большого числа вариантов («комбинаторный взрыв»). Поэтому такой отбор переменных так или иначе связан с применением оптимизационных алгоритмов, ярким представителем которых является генетический алгоритм (ГА) – эффективный эвристический подход к направленной оптимизации, предложенный в 1970-е годы в работах Холланда [116]. Лежащие в основе ГА идеи заимствованы из процессов развития и естественного отбора биологических популяций. С ростом вычислительных мощностей ГА обретал всё большее распространение. В настоящее время он применяется к большинству задач высокой размерности и к

задачам, где сложно или невозможно математически выразить целевую функцию оптимизируемых параметров.

Оптимальный набор переменных для построения модели можно искать в виде интервалов – групп соседствующих переменных. Такой подход, прежде всего, отражает интервальную природу спектральных данных, связанную с высокой коррелированностью соседних переменных одной и той же спектральной полосы. С другой стороны, объединение переменных в группы существенно снижает сложность оптимизационной задачи. Поскольку единичная переменная также является частным случаем интервала, интервальный подход можно рассматривать как расширение и обобщение обычных методов отбора переменных.

Интервальный ПЛС (ИПЛС) [101] – один из наиболее эффективных алгоритмов этого типа, в том числе для анализа данных оптической спектроскопии [117]. На входе в ИПЛС задаётся ширина интервалов, которые делят спектр на отрезки – *бины*. Далее задаётся число интервалов для включения в ПЛС модель, что существенно сокращает число комбинаций. В ходе перебора строятся ПЛС-модели для выбранных комбинаций, от меньшего числа интервалов к большему, а поиск оптимального решения ведётся по принципу последовательного снижения ошибки моделирования. Алгоритм перебирает не все возможные комбинации интервалов; это создаёт риск пропустить глобальный оптимум и является основным недостатком ИПЛС, но существенно ускоряет расчёт. Другим недостатком метода является фиксированные (задаваемые пользователем) число интервалов и их ширина, что сильно ограничивает пространство решений.

Учитывая описанные преимущества и недостатки существующих методов, представляется актуальной разработка нового обобщённого метода отбора переменных, объединяющего эффективный оптимизационный алгоритм, такой как ГА, и преимущества интервального подхода (но без присущих ИПЛС ограничений), что явилось одной из задач настоящей работы.

Отбор переменных является основным инструментом оптимизации ОМС под выбранное приложение в ходе её разработки (раздел 2.2.3) и редко применяется в анализе данных мультисенсорных систем. Отбор объектов же, напротив, может

быть весьма полезен при анализе больших массивов данных, таких как данные мониторинга процессов. Наиболее типичными задачами отбора объектов являются исключение *выбросов* (необычных измерений), а также селекция обучающего и проверочного наборов, например, из больших исторических данных.

1.3.4 Разведочный факторный анализ данных

Разведочный (*англ.* exploratory) факторный анализ представлен в первую очередь методом главных компонент (МГК; *англ.* principal component analysis, PCA) [118] и сходными проекционными методами, направленными на углублённое изучение внутренней структуры данных в факторном пространстве. В отношении объектов это: выявление группировок и выбросов, анализ взаимосвязи объектов (например, кривой развития процесса во времени), анализ статистической влиятельности образцов и групп образцов. Те же свойства могут быть рассмотрены и изучены в отношении переменных. В последнем случае наибольший интерес представляет поиск наиболее влиятельных переменных и скрытых спектральных откликов [119]. В конечном счёте результатом такого разведочного анализа данных является улучшенное понимание исследуемых объектов и аналитического метода как такового. Для одних и тех же данных можно создать бесконечное число различных факторных пространств, используя известные или изобретая новые проекционные преобразования. Эффективность факторного пространства в конечном счёте определяется его практичностью для анализа конкретных данных.

Матрицы **T** и **P** (уравнение 1.1) позволяют выявить ценную информацию о внутренней структуре данных. Их интерпретация основана на том, что корреляция между двумя переменными или сходство двух образцов является функцией расстояния в пространстве ЛП. Графики пар счетов $t_i - t_j$, в особенности первых двух $t_1 - t_2$, описывающих основную дисперсию данных, называют «картой образцов», которая, в частности, выявляет группы объектов и выбросы. В свою очередь, соответствующие графики нагрузок («карты переменных») служат для визуализации корреляций между векторами переменных. Расстояние от начала координат до образца на графике счетов (или переменной на графике нагрузок) вдоль определённой ЛП отражает их важность в отношении этой ЛП. Сравнение

графиков счетов и соответствующих нагрузок помогает определить взаимосвязь между образцами и переменными в модели.

Особым случаем разведочного анализа данных является метод разделения кривых (МПК; *англ.* multivariate curve resolution, MCR), называемый также *автомодельным* [120], поскольку он способен проводить математическое разрешение компонентов смеси «автоматически», анализируя лишь матрицу данных X . Разложение X по уравнению (1.1) в случае МПК даёт спектры чистых компонентов и соответствующие им концентрационные профили (в матрицах P и T соответственно). Схема МПК временной серии спектров (пример взят из высокоэффективной жидкостной хроматографии с УФ-детектором) представлена на рис. 1-7.

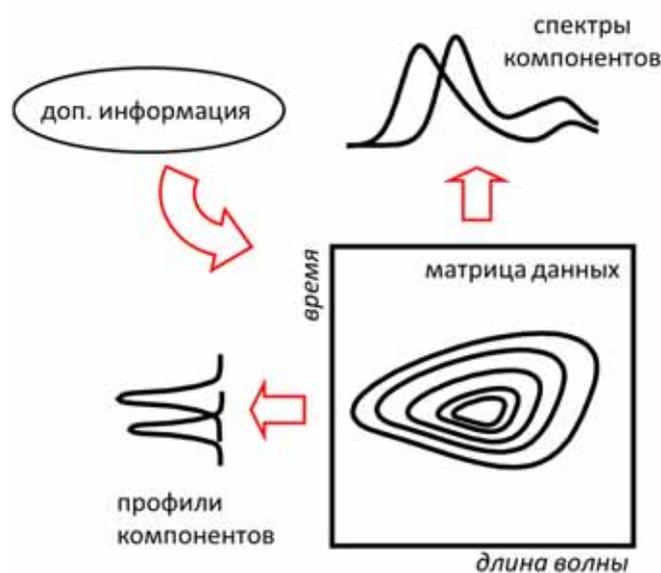


Рисунок 1-7. Схема МПК на примере временной серии оптических спектров.

С точки зрения математики МПК является так называемой обратной задачей (*англ.* reverse problem), имеющей в общем случае область решений, для сужения которой нужна дополнительная информация о данных. В МПК методом перемежающихся наименьших квадратов (ПНК-МПК; *англ.* alternating least squares MCR, ALS-MCR) – одним из наиболее используемых алгоритмов МПК, поиск решения проводится за счёт наложения граничных условий (*англ.* constraints), таких как требование отсутствия негативных значений [121]. Другой популярный алгоритм – СИМПЛИЗМА (*англ.* simple-to-use self-modeling mixture analysis,

SIMPLISMA) основан на поиске чистых переменных и пошаговом регрессионном разрешении пар «спектр-профиль» [122,123].

Метод МРК или, как его ещё называют, *спектральное разделение смесей* [124] имеет важные применения в различных практических областях, в том числе в АКТП [9]. Метод МРК может быть весьма полезен при разработке ОМС, в частности, для выделения спектров чистых компонентов из спектров смеси при определении оптимальной оптической конфигурации сенсорных каналов. Для анализа данных ОМС применение МРК ограничено определёнными ситуациями: в методе ПНК-МРК – наличием подходящих граничных условий, а в СИМПЛИЗМе – наличием селективных каналов. При определённых условиях МРК можно использовать для построения многомерных градуировок по ограниченному числу образцов – даже по трём, как это показано в работе [125].

1.3.5 Регрессионный анализ данных и многомерная градуировка

Градуировочные алгоритмы предназначены для построения регрессионной модели, задающей математическую связь между анализируемым свойством образца (например, концентрацией компонента) и его спектром. Для построения градуировочной модели создаётся *обучающий набор* образцов, измеренные спектры которого объединяются в матрицу \mathbf{X} (раздел 1.3.1). Известные значения градуируемого свойства, определённые независимым аналитическим (*референтным*) методом, образуют вектор \mathbf{y} , элементы которого соответствуют спектрам из \mathbf{X} . Конечной целью градуировки является последующий прогноз (предсказание) определяемого свойства в новом образце из его спектра с использованием построенной градуировочной модели. Одним из наиболее распространённых алгоритмов данной группы является регрессия методом ПЛС [126]. Успех алгоритма ПЛС объясняется его простотой и эффективностью в решении широкого круга практических задач, что делает его своеобразным эталоном. В соответствии законом Бугера-Ламберта-Бера [24, с. 144]:

$$\mathbf{A} = \boldsymbol{\varepsilon} \mathbf{C} \cdot l \quad (1.2)$$

где \mathbf{A} – матрица спектров в единицах оптической плотности, $\boldsymbol{\varepsilon}$ – матрица молярных коэффициентов поглощения компонентов смеси, \mathbf{C} – вектор их концентраций, а l –

толщина светопоглощающего слоя, спектральную матрицу можно считать представлением системы линейных уравнений. Эта система как правило является избыточной, то есть *ранг матрицы* данных (число линейно независимых уравнений) существенно ниже, чем число объектов – строк \mathbf{X} . Для решения избыточной системы уравнений используется линейный регрессионный анализ.

Результатом ПЛС-регрессии является нахождение вектора регрессионных коэффициентов \mathbf{b} ($m \times 1$), устанавливающего взаимосвязь между матрицей \mathbf{X} и вектором \mathbf{y} ($n \times 1$), который содержит известные значения определяемого свойства, например концентрации компонента по данным референтного метода:

$$\mathbf{y} = \mathbf{Xb} \quad (1.3)$$

Такая форма уравнения линейной регрессии называется инверсной, в отличие от классического уравнения (1.2). Главным преимуществом инверсных методов является то, что для построения градуировочной модели достаточно знать только вектор \mathbf{y} , участвующий в градуировке. С практической точки зрения это означает возможность построения градуировочных моделей по одному или нескольким выбранным компонентам смеси, не имея информации о природе, количестве и концентрациях других компонентов, несмотря на их возможное влияние на спектральные данные.

Модель \mathbf{b} в дальнейшем используется для прогноза неизвестной переменной \hat{y} для данных нового образца \mathbf{x}^{new} где \mathbf{x} – вектор аналитического измерения, представляющий образец, в нашем случае – его спектр или данные ОМС:

$$\hat{y} = \mathbf{x}^{new} \mathbf{b} = \sum_{i=1}^m x_i^{new} \mathbf{b}_i \quad (1.4)$$

где x_i^{new} – значение переменной i .

Факторное пространство в ПЛС также описывается уравнением (1.1). При этом ЛП выбираются таким образом, чтобы максимизировать ковариацию между проекцией данных \mathbf{X} и значениями \mathbf{y} . Результирующее новое пространство получается оптимизированным для решения регрессионной проблемы.

Заметим, что графическая интерпретация векторов счетов и нагрузок в факторном пространстве ПЛС-регрессии ничем не отличается от таковой в МГК и

часто используется для разведочного анализа и изучения внутренней структуры данных.

Существует модификация ПЛС2 алгоритма ПЛС, позволяющая найти регрессионные коэффициенты для нескольких зависимых переменных одновременно, используя одно и то же факторное пространство [96, с. 29]. В этом случае в левой части уравнения (1.3) будет фигурировать матрица \mathbf{Y} , столбцы которой содержат референтные значения для каждого из анализируемых компонентов, а вектор \mathbf{b} в правой части (1.3) заменится матрицей \mathbf{B} соответствующих регрессионных коэффициентов. На практике ПЛС2 используется редко. Теоретически ПЛС2 как альтернатива нескольким независимым ПЛС-моделям предназначен для ситуации, когда между концентрациями анализируемых компонентов есть корреляция. Однако исследований, дающих убедительное экспериментальное подтверждение преимуществ одной ПЛС2-регрессии перед отдельными ПЛС-моделями для каждого компонента, не найдено [9].

Разница между ПЛС и другими инверсными алгоритмами регрессионного анализа заключается в способе проекции данных по уравнению (1.1). С точки зрения целей настоящего исследования, выбор конкретного алгоритма анализа данных не является определяющим. Предпочтение, отдаваемое алгоритмам МГК и ПЛС при разработке ОМС, обусловлено высокой универсальностью этих проекционных инструментов и не исключает возможности их замены на другие линейные и нелинейные регрессионные методы.

Важным условием практического использования созданных градуировочных моделей является проверка соответствия новых данных (\mathbf{x}^{new}) модели, чтобы исключить заведомо неподходящие измерения, которые могут появиться в результате технической поломки или деградации прибора со временем, а также из-за человеческой ошибки (например, перепутанного образца). Диагностика тревожной ситуации производится на основании анализа тех или иных статистических критериев, характеризующих отличие нового спектра или измерения ОМС от данных модели. Часто используются расстояние до модели Q и расстояние до центра модели *Hotelling's* T^2 [127]. Возможны обобщённые

критерии, такие как отклонение *Deviation* в программе «The Unscrambler» норвежской фирмы «Camo Software» [128]. Более простым методом проверки является сравнение текущего и сохранённого измерений встроенного *эталонного образца*, называемого также образцом сравнения или контрольным, или обычного образца в стандартной ситуации (например, на той же стадии процесса). Критерии оценки новых данных на соответствие модели, а также функция расстояния, используемая для сравнения измерений, могут быть специально разработаны с учётом особенностей приложения и анализируемых данных, спектральных или ОМС [129].

1.3.6 Дискриминантный анализ данных

Задачи классификации и дискриминации образуют третью группу методов хемометрики. В общем виде классификация состоит в определении принадлежности (*отнесении*) нового образца к одной из известных категорий или групп (*классов*), причём возможны случаи полного отсутствия принадлежности к заданным классам или принадлежности к нескольким классам одновременно. Особый случай представляет собой одноклассовая классификация, определяющая принадлежность образца к единственно известному классу [130].

Дискриминация представляет собой специальный случай классификации объектов по двум (иногда более) известным заранее классам, когда вариант отсутствия принадлежности к обоим классам, строго говоря, не рассматривается [131, с. 233].

Рассмотрим подробнее один из методов дискриминантного анализа, а именно ПЛС-дискриминантный анализ (ПЛС-ДА; *англ.* PLS discriminant analysis, PLS-DA) [53, с. 572], который использовался в настоящем исследовании в процессе разработки ОМС. ПЛС-ДА использует регрессию для решения классификационной проблемы. Для построения модели принадлежность к классам кодируется присвоением значений 0 и 1 элементам вектора \mathbf{y} в уравнении (1.3), после чего ПЛС-регрессионная модель строится как обычно. Последующий прогноз принадлежности к классу с кодом «0» или «1» делается на основании того, лежит ли предсказанное по уравнению (1.4) значение \hat{y} ниже или выше порога 0.5

соответственно. Модель может быть «обучена» для работы с другим пороговым значением, которое в этом случае становится параметром модели. Для дискриминации между более чем двумя классами вместо ПЛС используется ПЛС2, а принадлежность к каждому классу кодируется отдельным столбцом матрицы Y .

ПЛС-ДА – хорошо разработанный метод, который часто используется для дискриминантного анализа спектральных данных, в том числе данных мультисенсорных систем [132].

1.3.7 Проверка градуировочной модели

Проверка (*валидация*) градуировочной или дискриминационной модели проводится с различными целями:

- 1) для определения оптимального числа ЛП при построении регрессионной модели [53, с. 122];
- 2) для проверки работоспособности метода анализа, например, для решения о принципиальной пригодности метода или о возможности использования модели в действующем анализаторе [88];
- 3) для сравнения моделей, полученных на разных данных или разными аналитическими методами, например в процессе разработки анализатора для сравнения с предыдущим прототипом [65];
- 4) для оценки достижимой точности прогноза новым аналитическим методом или в новом практическом приложении [53, с. 123].

В зависимости от решаемой задачи методы проверки модели могут различаться. Наиболее достоверной считается проверка специально спланированным независимым *тестовым* (или *проверочным*) набором, образцы которого используются только для валидации, то есть не участвуют в *обучающем наборе*, на котором строится модель. Проверка тестовым набором (ПТН) строго рекомендуется для оценки точности прогноза при построении «конечной» рабочей модели анализатора (цель 4 из вышеприведённого списка), несмотря на дополнительные затраты, связанные с получением независимого тестового набора. Для предварительного анализа и исследовательских задач (1–3) может быть применён алгоритм так называемой *перекрёстной проверки* (ПП; *англ. cross-*

validation) или валидация набором, отобранным (по каким-либо критериям или случайно) из полного набора имеющихся данных.

Для *перекрёстной проверки сегментами* (ППС) полный набор данных разбивается на n поднаборов (сегментов), и построение модели повторяется n раз с исключением одного из сегментов из обучающих данных на каждой итерации цикла. Исключённый сегмент при этом используется в качестве проверочного набора, а в качестве конечных показателей используется кумулятивная статистика по всем n проверкам. Существуют различные методы ПП в зависимости от способа разбиения полного набора данных на сегменты. Наиболее часто используемыми являются: *полная перекрёстная проверка* (ППП), если сегменты представлены единичными объектами, ПП последовательными блоками (сегменты образованы равновеликими группами последовательно идущих объектов матрицы данных, например «123123123» для 9-ти объектов), «венецианское жалюзи» (сегменты состоят из перемежающихся объектов: «111222333») и когда сегменты определяются аналитиком на основе структуры данных, например, соответствуют имеющемуся разбиению объектов на классы – ПП на классах (ППК).

В качестве основных характеристик точности градуировочной модели используются среднеквадратичное отклонение (СКО; *англ.* root mean-square error, *RMSE*) градуировки (СКОГ), прогноза с помощью тестового набора (СКОП) или по данным перекрёстной проверки (СКОПП), а также соответствующие значения коэффициента детерминации R^2 , которые вычисляются по уравнениям:

$$СКО(\Gamma, \Pi, \Pi\Pi) = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (1.5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \hat{y} = \frac{\sum_{i=1}^k y_i}{n} \quad (1.6)$$

где y_i и \hat{y}_i – известные и предсказанные значения, n – количество образцов.

Выбор оптимального числа ЛП для проекционных методов, использующих факторные проекции, является важнейшим аспектом построения точных регрессионных моделей. Как недооценка параметра a в уравнении (1.1)

(недоопределённая модель), так и его переоценка (переопределённая модель) могут существенно снижать *правильность* градуировки, внося систематическую погрешность. До сих пор определение оптимальной размерности факторного пространства модели остаётся экспертной областью. Это решение должно приниматься ответственным за моделирование аналитиком на основании знаний о природе образцов и аналитического сигнала и целого ряда факторов, таких как минимум или выход на плато кривых зависимостей $SKOP(P)$ от числа ЛП, расхождение $SKOG$ и $SKOP$, зашумлённость векторов регрессии и нагрузок, а также других статистических оценок [133]. До сих пор не удалось полностью алгоритмизировать нахождение оптимального числа ЛП, хотя такие попытки делаются, и есть успешные примеры решения проблемы для определённого круга задач [134].

Статистика проверки моделей ДА основана на подсчёте числа результатов отнесений разного типа: истинно положительных, ($ИП$; *англ.* true positives), ложно положительных ($ЛП$; *англ.* false negatives), называемых также ошибками I-го рода, истинно отрицательных ($ИО$; *англ.* false negatives) и ложно отрицательных ($ЛО$; *англ.* false negatives) – ошибок II-го рода. Из них рассчитываются значения чувствительности ($\%Ч$), селективности ($\%С$) и точности ($\%Т$):

$$\begin{aligned} \%Ч &= 100 \times ИП / (ИП + ЛО); \\ \%С &= 100 \times ИО / (ЛП + ИО); \text{ и} \\ \%Т &= 100 \times (ИП + ИО) / (ИП + ЛП + ИО + ЛО) \end{aligned} \quad (1.7)$$

Дополнительно могут использоваться значения Q^2 и DQ^2 [135], аналогичные R^2 градуировки, см. уравнение (1.6). В отличие от Q^2 , DQ^2 не выставляет значение прогноза на ближайшую границу области допустимых значений (от 0 до 1 в ПЛС-ДА) в случае выхода за неё истинных отнесений ($ИП$ или $ИО$).

1.3.8 Стратегии моделирования

Построение хемометрических, в частности градуировочных, моделей основано на статистическом подходе, при котором обучение модели происходит в рамках определённых условий. Такие модели создаются для решения конкретных прикладных задач, а диапазон моделирования и последующего прогноза

ограничивается заданными условиями.

В зависимости от охватываемого ими диапазона изменений различных факторов градуировочные модели можно условно разделить на *локальные* и *глобальные*. Такая терминология отражает скорее стратегический подход к моделированию, чем свойства создаваемых моделей. Наиболее распространёнными являются локальные модели, так как они позволяют избежать большого количества измерений за счёт сохранения нерелевантных факторов постоянными. Это облегчает построение градуировки, но и применимость локальных моделей будет ограничена выбранным диапазоном условий. Постоянство нерелевантных факторов зачастую трудно или невозможно обеспечить, особенно при анализе природных образцов. Если условия анализа меняются, необходимо строить новую модель или совершенствовать имеющуюся, с тем чтобы учесть изменившийся диапазон variability.

Модели, которые стремятся наиболее полно охватить всевозможную изменчивость образцов, называются глобальными. Эта изменчивость может быть учтена при помощи хорошо спланированного градуировочного эксперимента или с использованием достаточного количества реальных измерений, то есть обширных *исторических данных*. При анализе природных образцов использование глобального моделирования на исторических данных зачастую является единственным способом построения надёжной градуировки.

Преимуществом ПЛС-регрессии и других инверсных методов при их обучении на большом объёме исторических данных является то, что они учитывают все вариации в спектральных данных, даже если факт их наличия и природа неизвестны. Очень важно, чтобы модель охватывала не только настоящую, но и будущую variability искомым факторов. Это может быть достигнуто за счёт проведения анализа на большом количестве представительных образцов, которые покрывают весь возможный диапазон изменений. Глобальное моделирование было успешно применено в работах [72-74,91,136,137].

Использование стратегии глобального моделирования тесно связано с проблемой использования одной и той же модели на многих приборах, не

обязательно одной и той же серии, поскольку анализатор обычно постоянно совершенствуется. Это достигается применением техник *переноса модели* – адаптации существующей модели к новому прибору при помощи математических преобразований. Необходимость переноса локальной модели с прибора на прибор возникает при отсутствии стандартизации серийно производимых анализаторов [69] или при переносе лабораторной модели на производство (на другой анализатор или в другие условия) [138,139].

Обзор методов переноса градуировочных моделей для стандартизации БИК-спектрометров приведён в работе [70]. Прибор, на котором модель была создана, принято называть первичным (*англ. master*), а тот, на который модель переносится, называется при этом вторичным или рабочим (*англ. slave*). Для переноса, как правило, используют набор стандартов, спектры которых регистрируют как на первичном, так и на вторичном приборе.

В наиболее строгом методе покусочного прямого переноса (*англ. piece-wise direct standardization*) градуировочных моделей и его разновидностях по спектрам серии стандартов с первичного и вторичного приборов создаётся поправочная матрица [70], которая затем применяется к новым спектральным данным, измеренным на вторичном приборе. Метод прямого переноса модели является наиболее точным и универсальным, поскольку он принимает во внимание как возможный сдвиг шкалы длин волн, так и несоответствие интенсивностей спектров.

Перенос может также осуществляться за счёт корректировки самой градуировочной модели. По аналогии с классификацией подходов к решению регрессионной задачи (раздел 1.3.5) различают классический [69] и инверсный [138] перенос. Точность перенесённой модели сильно зависит от числа и состава использованных для переноса образцов.

Существуют упрощённые методы переноса модели, корректирующие не сами спектры вторичного прибора, а полученные на нём значения прогноза отградуированной величины [71]. Наиболее распространённым методом этого типа является метод коррекции наклона и отсека [139]. Этот метод прост в

реализации и хорошо зарекомендовал себя на практике [95]. Он основан на единственном допущении о линейности нескорректированных зависимостей «введено-найдено» (значений прогноза анализируемой величины от референтных), что часто наблюдается на практике. Соблюдение этого условия позволяет провести корректировку прогноза при помощи двух коэффициентов Δb_0 и Δb_1 , задающих отклонения величин наклона и отсечения смещённой зависимости «введено-найдено» от 0 и 1 соответственно:

$$y_{cor} = (y_{ref} + \Delta b_0) \times \Delta b_1 \quad (1.8)$$

где y_{ref} – оригинальный, а y_{cor} – скорректированный прогноз. Метод успешно работает в тех случаях, когда разница проста и систематична во всех измеренных на новом приборе данных, что и ожидается от ОМС.

Учитывая, что построение индивидуальных локальных моделей для множества недорогих серийно производимых мультисенсорных систем чаще всего нецелесообразно, а их аппаратная стандартизация зачастую невозможна, стратегия глобального моделирования с последующим переносом модели представляется основной в разработке ОМС.

1.3.9 Планирование градуировочного эксперимента

Теория планирования эксперимента (ПЭ) имеет вековую историю. Первой публикацией по ПЭ считается вышедший в 1753 г. трактат Джеймса Линда [140], применившего систематический подход в клинических исследованиях по лечению цинги среди британских моряков. Традиционными областями применения ПЭ в химии является выявление важности экспериментальных факторов и выбор оптимальных условий, например, для повышения выхода химической реакции, хроматографического разделения и др. [141].

Планирование эффективного градуировочного эксперимента для количественного анализа смесей представляет собой отдельную задачу, не решённую в полной мере существующей теорией ПЭ [142]. Задача одновременного определения нескольких компонентов на основе одного и того же спектра или другого многомерного измерения часто возникает при анализе качества пищевых продуктов и лекарственных средств, а также в медицинской диагностике, в

экологическом мониторинге, в АКТП и других областях (см. обзор в работе [143]). Реальные смеси часто не соответствуют условию закрытой системы, как это требуется в хорошо разработанной части ПЭ для анализа смесей [144].

Главной целью планирования многокомпонентной градуировки является повышение точности последующего прогноза, зависимость которой от набора обучающих образцов в спектральном анализе показана в работах [141,145,146]. Для решения этой задачи существующие подходы к ПЭ требуют дальнейшей доработки ввиду явных различий в целях анализа, природе образцов и особенностях моделирования. Одно из наиболее важных отличий при планировании градуировочного эксперимента относится к количеству необходимых *уровней*, на которых происходит варьирование градулируемых *факторов* – концентраций компонентов (раздел 3.1.1). Хорошо сбалансированный градуировочный набор должен состоять из достаточно большого числа образцов, равномерно распределённых по размерности каждого фактора, а следовательно, включать в себя достаточно много уровней. В многомерном регрессионном анализе число уровней должно значительно превышать число факторов [147]. Это требование делает принцип минимального числа уровней, используемый в классическом ПЭ, малоприменимым для количественного анализа смесей. Нехватка теории, методологии и ПО для планирования градуировочного эксперимента вынуждает исследователей конструировать собственные схемы или адаптировать методы традиционного ПЭ, даже если они далеки от оптимальных [143].

Публикации, посвящённые разработке многоуровневого ПЭ для одновременной градуировки двух или более компонентов, встречаются редко. Бреретон и др. обобщили традиционную двухуровневую схему с фильтрацией по Плакету-Бурману (*англ.* Plackett-Burman) до пяти [148], а затем и до семи уровней [149]. Этот подход был применён в работах [150-152]. Наиболее эффективным с точки зрения количества экспериментов является латинский гиперкуб образцов (ЛГО) [153], в котором на каждом уровне каждого фактора присутствует один и только один образец. Свойства ЛГО зависят от алгоритма заполнения.

Равномерный дизайн Фанга и Лина [154,155] исходит из максимальной

однородности заполнения ЛГО, определяемой функцией расхождения (*англ.* discrepancy function). Этот алгоритм требуют сложных вычислений, что делает метод пригодным для использования, в основном, в виде заранее просчитанных схем (таблиц) для разных практических случаев [156]. Недавно Кирсановым и др. был предложен упрощённый подход, основанный на однородности [157]. Опубликованные в литературе известные градуировочные ПЭ не рассматривают стратегии проверки и не включают встроенного тестового набора объектов, что является необходимым атрибутом практического моделирования.

В целом число образцов, требуемых для построения надёжной градуировочной модели, тем больше, чем меньше число переменных – каналов анализатора [147]. Поэтому использование методов ПЭ, позволяющих создавать экономичные обучающие и проверочные наборы, приобретает повышенную важность при построении градуировок для ОМС.

1.3.10 Хемометрика процессов

Понятие «процесс» (*лат.* processus) – течение, ход, продвижение в его абстрактном значении обычно связывают с поведением динамических систем, с их эволюционным развитием, то есть с последовательным переходом (в том числе целенаправленным) системы от одного состояния к другому, который можно описать посредством *траектории*. Более чёткие определения процесса возникают в применении к конкретной области деятельности: производственный процесс, технологический процесс, бизнес-процесс и т. д.

Приложения хемометрики к АКТП называют хемометрикой процессов (*англ.* process chemometrics) или многомерным анализом процессов (*англ.* multivariate process analysis) [9]. За более чем семидесятилетнюю историю своего развития хемометрика процессов накопила значительную теоретическую базу и большой практический опыт [158]. В наши дни этот раздел знаний испытывает необычайный подъем, сопровождающийся ежегодным ростом количества научных публикаций по теме анализа процессов, а также появлением во всем мире соответствующих образовательных программ и научных сообществ. Бурное развитие отрасли обусловлено научно-техническим прогрессом и сопутствующим ростом сложности

и наукоёмкости производств, где контроль становится инструментом эффективности. Немаловажным стимулом в развитии АКТП явился выпуск в 2004 г. регулирующей организации США «Food and Drug Administration» («FDA») инициативы, которая пересматривает подходы к анализу качества в пищевой и фармацевтической промышленности. Этот документ, получивший название технологии анализа процессов (*англ.* process analytical technology, PAT) [8] дал «зелёный свет» использованию современных методов хемометрики, официально признав её подходы и метрологические оценки.

Развитие инструментальных и математических методов АКТП в современном мире входит в период зрелости. Он характеризуется быстрым ростом числа практических приложений и их разнообразием. Как следствие, возникает необходимость методологических исследований. Методология АКТП касается разнообразных тем и призвана отвечать на многие практические вопросы. Какие задачи обычно решаются в ходе анализа процесса? Какие риски должны быть исключены или минимизированы [159]? Как найти необходимую или оптимальную комбинацию сенсоров [34,38,160]? Каков наилучший способ регистрации спектральных данных и референтных образцов [161,162]? Какие методы предварительной обработки данных и алгоритмы их анализа лучше соответствуют данной аналитической проблеме [163]? Как получить достаточно точную прогностическую модель процесса при минимальных затратах и с минимальными усилиями [164]? Как удостовериться в достаточном качестве данных и в надёжности модели [165,166]? Как преобразовать приобретённые знания о процессе в стратегию его контроля [166]? Эти и другие основополагающие вопросы должны быть заданы и получить свои ответы на стадии разработки процесса, ещё до начала сбора данных.

АКТП в наши дни решает амбициозную задачу перехода от контроля качества произведённой продукции к непрерывному мониторингу производственных процессов в реальном времени. Постоянно растёт количество объектов мониторинга, производств, технологических стадий, полевых и лабораторных измерений, переводя контроль процессов на новый качественный

уровень. Решение этих задач приводит к революционным изменениям в инструментальной аналитике, прежде всего, в оптических спектральных методах анализа. Благодаря новейшим разработкам в области фотоники, появлению новых источников света, детекторов, оптики и светопроводящих материалов становится возможным производить доступные, недорогие и легко тиражируемые мобильные анализаторы, способные работать в условиях идущего процесса, в полевых условиях и даже в быту. Все это в полной мере относится к оптическим мультисенсорным системам, являющимся техническим локомотивом АКТП. Их дальнейшее развитие требует решения различных научных, технических и методологических задач. В том числе, понадобятся новые методы работы с данными, природа которых меняется, а анализируемые объёмы стремительно растут.

ЗАКЛЮЧЕНИЕ К ГЛАВЕ 1

Произведённый обзор литературы даёт анализ современного состояния науки и техники в области разработки оптических мультисенсорных систем и позволяет наметить пути их дальнейшего развития. Показана важность создания на основе спектроскопии компактных и недорогих специализированных устройств, которые позволят значительно увеличить эффективность анализа объектов сложного состава в определении как индивидуальных компонентов, так и обобщенных химических показателей, а также расширить возможности экспресс-анализа, полевых измерений и онлайн-мониторинга технологических процессов. Это потребует решения ряда задач, составляющих основу научного подхода к разработке ОМС, начиная с выработки общей методологии их создания.

Необходимым условием создания высокоэффективных ОМС является использование современных разработок в области фотоники: источников излучения, детекторов, оптоволоконной техники и интерфейсов измерения. Важнейшую роль играют также математические методы анализа данных, и их совершенствование является предпосылкой дальнейшего успешного развития ОМС. Не в полной мере решена оптимизационная проблема расчёта оптических

свойств каналов мультисенсорных систем для конкретного приложения. Методы хемометрики и соответствующее программное обеспечение, хорошо зарекомендовавшие себя в анализе традиционных спектральных данных, могут потребовать адаптации к особенностям самих ОМС и использующих их новых практических приложений. При этом построение точных и надёжных градуировочных моделей и их проверка должны быть основаны на статистически представительных, но по возможности небольших, а значит, тщательно спланированных наборах образцов. Учитывая высокую востребованность мультисенсорных систем для аналитического контроля технологических процессов, разработкам ОМС и методов хемометрики в этой области следует уделить повышенное внимание.

2. МЕТОДОЛОГИЯ РАЗРАБОТКИ ОМС

Эта глава рассматривает теорию и практические аспекты разработки ОМС.

2.1 Цикл разработки ОМС

2.1.1 Подход к разработке

Оптическая мультисенсорная система изначально проектируется как специализированный анализатор, то есть оптимизируется под определённую аналитическую задачу. В предлагаемом в настоящем исследовании подходе ОМС создаётся как упрощённая альтернатива методу традиционной спектроскопии на основе лабораторного или процесс-спектрометра (будем называть исходный для создания ОМС спектроскопический метод *полносpekтральным*). Такой метод играет роль «золотого стандарта», с которым в процессе разработки сравниваются результаты тестирования прототипов и производственных моделей разрабатываемых устройств.

Основой упрощения является выбор наиболее подходящих для анализа спектральных интервалов (и исключение менее релевантных). Это достигается математической оптимизацией, проводимой на экспериментальных данных полносpekтрального метода (раздел 2.2) или на основе априорных знаний о природе образцов и об аналитическом методе, если полносpekтральные данные не доступны.

Дополнительной «платой» за упрощение, удешевление и миниатюризацию часто является отказ от технической (аппаратной) стандартизации анализатора, которая является существенным фактором, определяющим цену системы. Более того, использование в конструкции трудно стандартизируемых элементов, таких как световоды и светодиоды, делает аппаратную стандартизацию ОМС практически недостижимой. Стандартизация в этом случае переносится на ПО и осуществляется средствами переноса модели.

Помимо определения оптимальной конфигурации информационных каналов, в ходе разработки ОМС необходимо принять ряд конструктивных решений, таких как интерфейс и режим измерения, техника эталонного измерения

и выбор образца сравнения, стратегия построения модели, способ передачи данных и отображения результата, а также других решений, обеспечивающих максимальную информативность получаемых данных.

2.1.2 Основные этапы и задачи разработки

Разработка ОМС состоит из пяти этапов, каждый из которых включает определённую последовательность действий – задач, которые представлены в табл. 2-1. Каждый этап служит достижению определённой промежуточной или конечной цели, и его результат влияет на дальнейшие разработческие решения. Конечной целью представленного цикла разработки является промышленно производимый и работающий на практике ОМС-анализатор.

Таблица 2-1. Основные этапы и задачи в разработке ОМС.

№	Этап	Задачи
1	Проектирование	анализ требований, выбор спектрального метода, <u>полносpekтральные измерения, оптимизация каналов</u>
2	Создание прототипа	цикл: изготовление/ улучшение прототипа – <u>тестирование</u>
3	Производство	цикл: улучшение промышленной модели – <u>тестирование</u>
4	Ввод в эксплуатацию	установка (если требуется), цикл: <u>построение предварительной прогностической модели</u> – <u>тестирование, построение рабочей модели</u>
5	Техническая поддержка	диагностика и обслуживание устройства, <u>диагностика модели, улучшение модели</u> (локальной или глобальной)

Примечание: подчёркнутые задачи связаны с применением хемометрики.

Проектирование ОМС начинается с анализа требований его планируемого практического приложения, которые с учётом бюджета разработки определяют свойства разрабатываемого анализатора. Если предварительные измерения при помощи выбранного полносpekтрального метода доказывают работоспособность метода и целесообразность дальнейшей разработки, то этап проектирования завершается оптимизацией информационных каналов ОМС – их числа и рабочих спектральных интервалов, а также режима и интерфейса измерения (раздел 2.2.5). Результатом этого этапа является также оценка достижимой точности анализа, которую сравнивают с экспериментально установленной точностью полносpekтрального метода. Недостижение требуемой точности может быть

основанием для прекращения проекта.

Рассмотрим в качестве примера описанную в главе 5 разработку светодиодной ОМС для экспресс-анализа молока (табл. В-1, приложение В). Мотивацией к исследованию послужил существующий запрос на недорогой, компактный и портативный анализатор для определения обобщённых показателей питательной ценности натурального и переработанного молока: массовых долей содержащихся в нём суммарного жира и суммарного (общего) белка. Исходя из требований, для анализа была выбрана спектральная область Вид/КВБИК, разработан тестовый набор образцов и проведены полносpekтральные измерения, которые показали работоспособность метода при варьировании размеров жировых глобул, превышающем природное. С использованием интервального отбора переменных был найден оптимальный набор светодиодов для построения ОМС.

Целью второго этапа является создание хотя бы одного экземпляра действующей модели ОМС – *прототипа*, удовлетворяющего требованиям приложения. На этом этапе необходимо решить вопросы обеспечения требуемого качества данных, в частности прецизионности (*воспроизводимости*) измерения и выбора эталонного образца сравнения. Прототип должен быть испытан на серии стандартных или реальных образцов, разработанной с применением ПЭ. Доработка и последующее тестирование прототипа происходит до тех пор, пока он не будет удовлетворять заданным требованиям. Важно, чтобы методика испытаний давала возможность сравнивать модели прототипов между собой и с результатами полносpekтрального анализа, позволяя судить об улучшении или ухудшении версии анализатора. Как правило, рекомендуется использовать тот же набор тестовых образцов, что и на стадии проектирования. Тогда можно сравнивать результаты работы прототипа не только с результатами его разных версий, но и с полученными при проектировании результатами полносpekтрального эксперимента.

В примере с молоком анализ на содержание жира и белка может проводиться в режиме пропускания или диффузного отражения. Экспериментальные испытания с разработанным тестовым набором позволили сравнить различные прототипы и

геометрии измерения. Результаты показали более высокую точность определения на пропускание и позволили определить оптимальную длину оптического пути. В работе был предложен и успешно протестирован новый подход, использующий внутреннее нормирование или *безэталонный анализ* (раздел 2.3.5), не требующий расчёта теоретических спектров, как в описанном ранее одноимённом методе [167].

На третьем этапе происходит преобразование разработанного прототипа в готовое к работе и серийному производству устройство. В ходе оптимизации использованные в прототипе материалы и конструкционные решения могут быть заменены на более подходящие для производства и эксплуатации, а также на обеспечивающие более надёжную конструкцию устройства и снижающие его стоимость. Как и на втором этапе, улучшение промышленной модели состоит в циклическом тестировании каждой новой версии до достижения ею требуемого качества и цены с помощью тестового набора образцов. Так, в нашем примере анализатора молока потребуется подобрать светодиоды надлежащего качества, а также материалы зонда, регламентированные для пищевых производств и выдерживающие соответствующие условия стерилизации. Для тестирования может использоваться тот же спланированный набор, что и на предыдущих этапах разработки.

До начала практического использования ОМС необходимо оснастить рабочей прогностической (градуировочной или классификационной) моделью. В зависимости от выбранной стратегии моделирования (раздел 1.3.8) строится локальная или глобальная модель. При локальном подходе вначале часто используется предварительная «быстрая» модель, менее точная. Это бывает также необходимо, если требуется (до)обучение модели на исторических образцах в ходе использования анализатора, что часто имеет место в АКТП (раздел 4.2.1). Предварительная модель может быть получена с использованием экономного дизайна градуировочного эксперимента или же путём переноса лабораторной модели в новые условия. При использовании глобальной модели необходимо также решить проблему её переноса между устройствами. Так, возможность построения глобальной модели и её переноса была протестирована на рассматриваемой в

качестве примера светодиодной ОМС для анализа молока. Важную роль при этом сыграл тот же разработанный на первом этапе стандартный набор градуировочных образцов. Этап заканчивается созданием полноценного аналитического устройства, оснащённого рабочей моделью, дающей требуемую точность анализа.

В ходе практического использования ОМС может нуждаться в техническом обслуживании, в том числе в поддержке многомерной модели, что и составляет содержание последнего, пятого этапа разработки. Под действием изменений в техническом состоянии системы, условий эксплуатации, состава образцов, параметров процесса и других факторов точность прогноза может со временем снижаться. Программа поддержки модели ОМС, обеспечивающая постоянную исправность и точность системы, может включать следующие элементы: самотестирование качества спектров относительно внутреннего или внешнего эталона, тестирование новых данных на соответствие модели, взятие контрольных проб и их лабораторный анализ с последующей оценкой отклонения прогноза от требований и продолжительные измерения с целью выявления нежелательных временных отклонений: наличия приборного дрейфа, роста шума, изменения формы сигнала. При необходимости может понадобиться дообучение или полная замена модели.

Из табл. 2.1 видно, что методы многомерного анализа используются на всех этапах разработки ОМС.

2.2 Моделирование информационных каналов ОМС

2.2.1 Выбор метода и полноспектральные измерения

Изначально выбор спектрального метода или нескольких потенциально приемлемых методов осуществляется на основании имеющихся знаний о природе объектов анализа и на данных предварительных градуировочных экспериментов. Их целью является *технико-экономическое обоснование работоспособности* (англ. *feasibility study*) выбранного метода для решения поставленной задачи, а также достижимая точность прогноза. Статистические характеристики точности базового полноспектрального метода будут затем использованы как ориентир при

разработке ОМС. Достоверность результатов предварительных экспериментов и их сравнения с последующими результатами тестирования обеспечивается использованием одной и той же серии обучающих образцов, специально разработанной с использованием теории планирования градуировочного эксперимента (глава 3).

Если доступные спектральные методы оказываются недостаточно точными для решения поставленной аналитической задачи, информативность данных можно повысить, дополнив исходные спектры данными другой спектральной области, данными, снятыми с другой геометрией измерения, или даже данными другого спектрального метода [40,166,168,169]. *Дополненные* (англ. augmented) данные получаются за счёт объединения двух или более матриц так, чтобы они образовывали новую расширенную спектральную шкалу (по горизонтальной оси), которая при несоответствии единиц может быть условной. Строки объединённой матрицы при этом по-прежнему соответствуют измеряемым образцам, так что вторая (вертикальная) шкала данных является общей.

При разработке ОМС всегда желательно проверять возможность использования дополненных данных, прежде всего на основе методов, работающих в смежных спектральных областях, таких как УФ/Вид-, БИК- и ИК-спектроскопия. Это связано с тем, что технические ограничения (такие как спектр источника, спектр пропускания оптики и спектр чувствительности детектора), обуславливающие разбиение спектроскопии на методы и области, могут быть сняты при переходе к ОМС с иным принципом действия, например светодиодной.

Необходимым условием создания дополненных данных является общность набора образцов в объединяемых матрицах данных. Совместимые блоки данных могут быть получены одновременным или поочерёдным измерением одних и тех же созданных образцов или отобранных проб разными спектральными анализаторами. Это не представляет большой сложности в случае гомогенных образцов. При измерении же твёрдых негомогенных образцов разными зондами можно использовать разметку, обеспечивающую измерение в одних и тех же точках образца, как это делалось в работе [88] (раздел 5.4.2). При этом

предполагается, что методы можно будет объединить в одном зонде [170,171]. Взаимно дополняемые данные могут быть результатом одновременного наблюдения протекающего процесса различными онлайн-методами, как в работе [40], где процесс покрытия пеллет изучали посредством БИК- и КР-спектроскопии при помощи встроенных в реактор зондов системы «LHP» (раздел 5.2.3). Синхронизация данных мониторинга процесса, необходимая для их объединения, может происходить физически или математически. В первом случае точное совпадение времён и интервалов сбора данных обеспечивается технически – использованием синхронизирующего триггера и настроек регистрирующего спектры ПО. Математическая пост-синхронизация осуществляется путём интерполяции объектов объединяемых матриц данных к единой временной шкале, в качестве которой можно выбрать шкалу данных, снятых с более редким шагом по времени [40]. Большая разница во временах получения спектров объединяемых методов может потребовать усреднения нескольких спектров на определённом интервале – *биннинга* (англ. *binning*), определяемом более продолжительным временем регистрации. Это необходимо для обеспечения соответствия *динамических образцов* (материала процесса, представленного спектрами процесса) для каждой пары спектров взаимно дополняемых методов.

Для создания градуировочной модели необходимо также обеспечить синхронизацию объединённой временной шкалы со временами отбора референтных проб обучающего набора. Принципы синхронизации временных шкал в этом случае сходны. Математическая синхронизация данных может быть невозможной или некорректной, если методы получены с несовместимым шагом *дискретизации* сбора данных (раздел 4.1.2). Например, спектры обоих методов имеют большой и различный шаг, так что общая временная шкала не может быть выбрана, или она не совпадает со временами отбора референтных градуировочных проб.

Частным случаем дополненных данных является матрица, получаемая при анализе трёхмодальных данных методом развёртывания (раздел 1.1.3). Совместимость отдельных матриц по образцам обеспечивается в этом случае самой

исходной структурой данных.

Использование дополненных данных в разработке ОМС позволяет выбрать оптимальную конфигурацию оптических каналов с использованием всей объединённой спектральной шкалы с последующей её реализацией в одном устройстве. Этот подход показал свою эффективность в различных практических задачах, представленных в главе 5 (табл. 5-1 и табл. В-1 в приложении В).

2.2.2 Конфигурация оптических каналов

Прежде всего необходимо определить число и свойства информационных каналов как ключевые параметры системы, влияющие на информативность измерения и определяющие техническую реализацию проектируемой ОМС. Каждый информационный канал характеризуется спектральным интервалом, который можно задать двумя параметрами (положением границ или центральной длиной волны и шириной). Таким образом, установление оптимальной оптической конфигурации является многопараметрической оптимизационной задачей, сложность которой быстро возрастает с числом каналов.

На каждом шаге оптимизационного алгоритма строится многомерная градуировочная модель на поднаборе полноспектральных переменных, соответствующем текущей тестируемой конфигурации интервалов. В качестве минимизируемой целевой функции для поиска оптимальных интервалов могут использоваться статистические характеристики погрешности многомерной, например ПЛС-градуировочной, модели: $СКОГ$, $СКОП(П)$, R^2 и др. (раздел 1.3.7), а также их комбинации. Учитывая комбинаторную сложность задачи и интенсивность расчёта целевой функции, требуемый оптимизационный алгоритм должен быть эффективен в нахождении разумного локального, а в идеале – глобального оптимума. В настоящем исследовании был разработан подход на базе ГА, изложенный в разделе 1.3.3.

Важно отметить, что оптимальная конфигурация каналов не является независимой от методов анализа регистрируемых ими данных. Различным методам предобработки данных и построения прогностических моделей могут соответствовать различные по числу и свойствам каналов конструкции ОМС. В

связи с этим, выбор метода моделирования и оптимизация технических параметров должны производиться совместно на этапе проектирования ОМС.

Объективных ограничений свойств выбираемых интервалов нет, однако для ускорения расчёта на них могут быть наложены следующие опциональные условия:

(а) Фиксированное число каналов. Число интервалов выставляется оператором и не участвует в оптимизации с исключая из неё одного параметра. Это целесообразно, если оптимальное число каналов заранее известно или при наличии технических ограничений. Число каналов может быть оптимизировано «вручную», то есть тестированием наиболее подходящих вариантов. Максимально возможное число каналов всегда ограничено по техническим или иным соображениям.

(б) Фиксированные ширины. Ширины интервалов выставляются заранее (чаще всего равные) и не могут меняться в процессе оптимизации. Это сокращает число оптимизируемых параметров интервалов вдвое.

(в) Фиксированные положения интервалов. Изначально задаётся набор интервалов-кандидатов с фиксированными позициями, который может быть избыточным. Тогда оптимизации подлежат лишь число (выбирается поднабор интервалов-кандидатов) и ширины интервалов. При необходимости ширины могут быть также зафиксированы.

(г) Ограничение на сближение и перекрытие интервалов. Алгоритм может препятствовать сближению центров интервалов на расстояние менее заданного. Дополнительно можно установить ограничение на степень перекрытия соседних интервалов. В отсутствие этого ограничения схождение в процессе оптимизации двух и более исходных интервалов к одному и тому же или к сильно перекрывающимся интервалам может свидетельствовать о чрезмерном количестве каналов. Это можно использовать для оценки оптимального числа каналов в предварительных расчётах.

Не ограниченная условиями оптимизация оптической конфигурации каналов ОМС становится чересчур продолжительной уже для достаточно простых систем, даже с использованием эффективных алгоритмов и современной компьютерной

техники. Применение ограничений (а)-(г) и их комбинаций позволяет сократить время расчёта на порядки, до нескольких часов [172].

Для получения наиболее достоверной конфигурации мультисенсорной системы необходимо учитывать оптические характеристики используемых в ней элементов, делающие вклад измеренных полноспектральным методом переменных неравнозначным, зависящим от длины волны. К таким характеристикам относятся спектры испускания источника, поглощения интерфейса измерения и чувствительности детектора (рис. 1-3). Суперпозиция этих спектров даёт суммарную *функцию отклика*, которую следует использовать в процессе расчёта как весовую функцию для спектральных переменных на оптимизируемых интервалах. Взвешивание выполняется путём поэлементного умножения исходных спектральных интенсивностей на каждом интервале на соответствующий вектор весов, вычисленный с использованием данной функции для той же спектральной сетки. Весовые коэффициенты обычно не оптимизируются, а задаются заранее, хотя и могут рассчитываться на каждом шаге оптимизации, поскольку могут зависеть от спектральной длины волны текущего интервала.

Существует два варианта взвешивания. Если известна только теоретическая (аналитическая) функция отклика, она может использоваться для взвешивания переменных на всех оптимизируемых интервалах. Например, при моделировании светодиодных ОМС (тип 1 в табл. 1-3) вектор весов может рассчитываться по функции Гаусса [173]. Границы интервалов при этом можно определять условно, например ограничивать гауссовым стандартным отклонением (σ), взятым с определённым коэффициентом. Параметры функции могут быть зафиксированы или задействованы в оптимизации. Так, в работе [66] предложена эмпирическая полиномиальная функция, описывающая уширение экспериментальных спектров испускания светодиодов, аппроксимированных гауссианом, от длины волны.

Если же функция отклика, например спектры доступных светодиодов, известны априори, то оптимизация сводится к выбору наилучшего сочетания элементов сенсора (в приведённом примере – светодиодов) из базы данных. В этом случае опции (б)-(г) выше становятся неприменимыми, а веса определяются из

спектров самих конструктивных элементов. Обычно интенсивность (максимальное значение на интервале) функции отклика нормируется к единице, как это было сделано при проектировании ОМС в исследовании молока (раздел 5.1.5), или к индивидуальным значениям на интервалах. Однако если интенсивности отдельных оптических сенсорных каналов настраиваемы, как у светодиодов, то этот параметр может быть включён в оптимизацию вместе с параметрами интервалов.

2.2.3 Алгоритм оптимизации

В отдельных случаях для расчёта оптической конфигурации ОМС можно использовать полный перебор вариантов, например при сочетании ограничений (б) и (в) из раздела 2.2.2 или в случае подбора элементов (в том числе светодиодов) из имеющегося набора. Однако в большинстве случаев требуется применение оптимизационных алгоритмов. В настоящей работе предложен подход, основанный на генетическом алгоритме [115], адаптированном к оптимизации спектральных интервалов.

Блок-схема выбора интервалов представлена на рис. 2-1. Алгоритм включает в себя два цикла оптимизации. Внутренний цикл находит локальное оптимальное решение, основанное на так называемой эволюции популяции. Под популяцией в ГА подразумевается пакет решений, которые в начальном приближении, как правило, заполнены случайными числами. Хромосомы при этом представлены каналами ОМС. Внешний цикл оптимизации выполняет смену популяции, чтобы повысить вероятность нахождения глобального оптимума или просто найти лучшее решение. Количество итераций внешнего цикла, а следовательно, и качество решения ГА, зависит от времени, отведённого на вычисления.

На первом этапе внутреннего цикла полноспектральные данные редуцируются до переменных на текущих интервалах, определяемых текущими значениями оптимизируемых параметров (в первом цикле – их начальными приближениями). Затем на редуцированном наборе строится ПЛС-градуировка и рассчитывается целевая функция. В качестве последней в данном исследовании использовалось *СКОГ*. Использование более надёжных статистических оценок

погрешности, таких как *СКОПП*, во внутреннем цикле оптимизации сопряжено со значительным замедлением работы алгоритма. В качестве критерия сходимости внутреннего цикла (минимальное улучшение) была выбрана величина 10^{-5} .

Каждое новое решение, полученное в результате внутренней оптимизации, подвергается ПП, и её *СКОПП* сравнивается с решениями предыдущих популяций. Лучшее значение сохраняется для следующего сравнения. Изменение популяции повторяется до тех пор, пока не выполняются опциональные критерии остановки. Обычно это общее время расчёта. Итоговое решение принимается за оптимальное. Практический пример расчёта светодиодной ОМС для одновременного определения жира и общего белка в молоке рассмотрен в разделе 5.1.5.

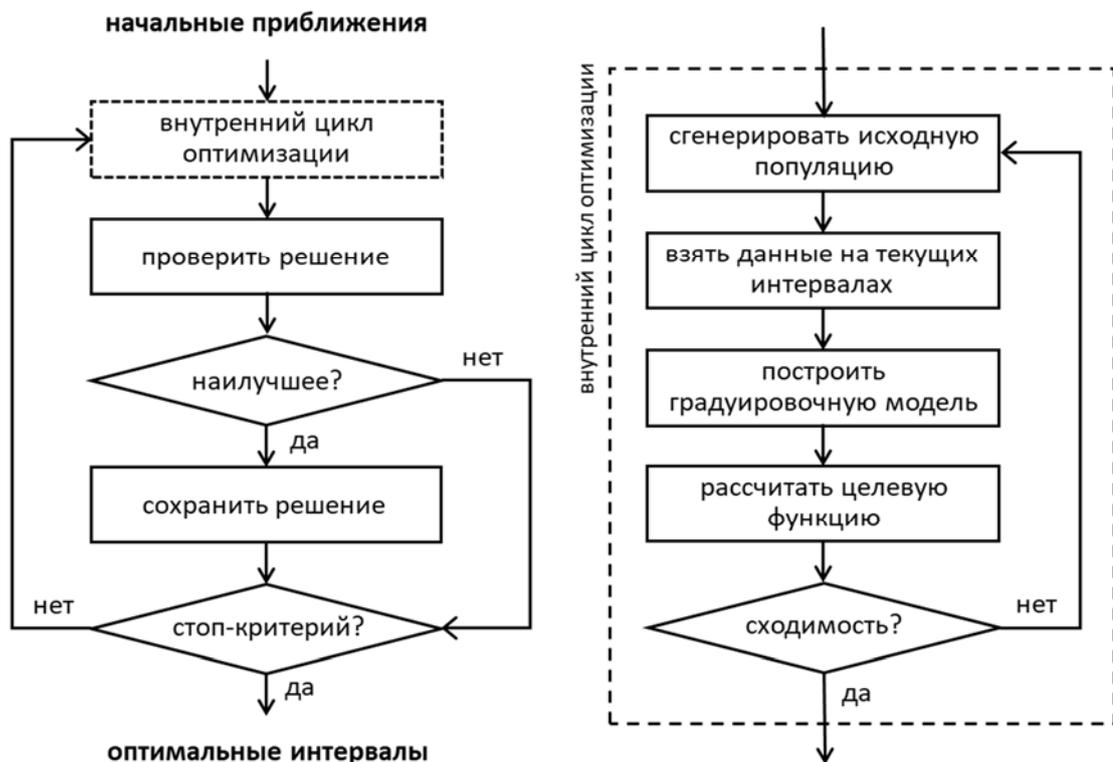


Рисунок 2-1. Блок-схема генетического алгоритма для оптимизации каналов ОМС [172].

Дополнительным или альтернативным методом при оптимизации информационных каналов ОМС может служить рассмотренный в разделе 1.3.4 разведочный анализ данных, в частности МРК [174]. Разрешение спектров индивидуальных (чистых) компонентов сложной смеси, такой как среда идущего процесса, может указать на информативные спектральные области пиков поглощения анализируемых веществ [40,81]. МРК и сходные методы могут также

быть использованы для оптимизации разделения в высокоэффективной газожидкостной хроматографии с ДМД [175], данные которой по своей природе аналогичны онлайн-спектрам процесса.

2.2.4 Интервальный метод отбора переменных при одновременной оптимизации предобработки данных

Описанный алгоритм оптимизации информационных каналов ОМС (раздел 2.2.3) представляет независимую ценность как интервальный метод отбора переменных для улучшения прогностических моделей. Настройки алгоритма позволяют сужать интервал до одной переменной, а также использовать усреднение спектральных интенсивностей на выбранных интервалах.

Нами было предложено важное расширение алгоритма [176], позволяющее включать в оптимизационный цикл (помимо свойств интервалов) параметры предобработки анализируемых спектральных данных, а именно метода и его опций, таких как порядок производной, ширина сглаживающего окна и т. д. Предобработка данных, как и отбор переменных (интервалов), является мощным средством улучшения точности прогностических моделей. Однако применение этих подходов одного за другим не гарантирует успеха, поскольку наилучшая в сочетании с одним набором переменных предобработка может выдавать неоптимальный результат для другого. Стратегической основой предложенного подхода является совместная оптимизация параметров интервалов и предобработки в одном цикле. Его эффективность была продемонстрирована на конкурсе по хемометрике³, где найденные с его помощью решения получили второй приз среди экспертов [176] и первый – в студенческой номинации [95].

2.2.5 Техника измерения

Вслед за определением оптической конфигурации каналов оптимизируется

³ Конкурс по хемометрике (*англ.* chemometrics shootout) проводился на международной конференции «International Diffuse Reflectance Conference (IDRC)» в 2014 г. в г. Чамберсбург (США). По условиям конкурса участники могли использовать любые математические и программные средства для получения наилучшей модели. Погрешность прогноза при этом проверялась организаторами с использованием тестового набора, неизвестного участникам. Автор занял на конкурсе второе место.

техника измерения, включающая интерфейс и геометрию измерения (раздел 1.2.4), а также другие конструкционно изменяемые технические параметры оптических каналов, такие как интенсивность излучения, продолжительность и частота импульса светодиода и др. Эти параметры во многом определяются свойствами измеряемого образца – его агрегатным состоянием и физико-химическими свойствами, такими как гомогенность, мутность, сыпучесть и общая оптическая плотность. Также следует принимать во внимание форму и состояние образца, в том числе доступный для измерения объём, характер поверхности, стационарный это или перемещаемый образец, динамичную среду процесса и т. д.

Важнейшей характеристикой спектрального метода является *эффективный объём* представляемого спектром динамического образца – количество анализируемого вещества, участвующего во взаимодействии с детектируемым светом. Этот объём называется эффективным, поскольку точное его определение едва ли возможно. Для приблизительной оценки эффективного объёма при анализе твёрдых веществ методами ДО (когда это наиболее актуально) можно использовать такие характеристики как диаметр пятна, глубина проникновения и скорость процесса (для онлайн-измерений) [40]. Соответствие эффективного объёма целям анализа является важнейшим условием, проверяемым при разработке анализаторов, особенно онлайн. Например, при онлайн-мониторинге гомогенности гранулята в ходе смешивания при изготовлении таблеток этот объём не должен быть ни чересчур большим (риск пропустить негомогенность), ни чересчур малым (измерения отражают поликристаллическую структуру гранулята). Оптимальный спектрально измеряемый объём гранулята в приведённом примере примерно соответствует размеру изготавливаемой из него впоследствии таблетки. Эффективный объём в АКТП называют также *шкалой подробности* (англ. scrutiny scale), с которой осуществляется наблюдение за процессом. Для градуировки также очень важно, чтобы эффективный объём спектрального измерения соответствовал объёму соответствующей ему пробы, отбираемой для референтного анализа.

При анализе прозрачных жидкостей хорошо подходит измерение на пропускание в УФ/Вид- и БИК-областях, а выбранная толщина поглощающего

слоя может варьироваться от 0.5 до 10 мм и более, в зависимости от рабочей области спектра, силы источника, оптической плотности образца и порога обнаружения анализируемого компонента. Однако на пропускание можно измерять и сильно рассеивающие образцы, такие как молоко и даже твёрдые объекты, такие как таблетки [177]. Отдельную проблему при этом представляет выбор образца сравнения для пересчёта в единицы оптической плотности [84,85]. В то же время молоко и другие дисперсионные среды можно анализировать в режиме диффузного отражения под углом 0° [72,178] или на рассеяние под другими углами [178,179]. При использовании высокоинформативного метода ИК-спектроскопии в сочетании с НПВО-зондом обычно предполагается гомогенность среды образца, но эта конфигурация может быть использована и в её отсутствие, если детектируемый при этом сигнал отражает важный обобщённый показатель химического состава образца. Так, ИК-спектрометр с оптоволоконным зондом НПВО использовался для диагностического анализа биологической ткани [27,78] и даже для анализа почвы на содержание воды в сочетании с определённой обработкой данных [87]. Благодаря своей «близорукости» НПВО-метод также пригоден для избирательного анализа только растворённых компонентов жидких негомогенных сред, таких как среда биотехнологического процесса [65]. Эти примеры наглядно показывают, что выбор оптимальной техники измерения – экспертная задача, требующая участия аналитика (или аналитиков) со знаниями химии, физики и спектроскопии, а также методов анализа многомерных данных.

Оптимизация техники измерения предполагает наличие работающего анализатора и проводится на этапе создания и улучшения прототипа, первая конфигурация которого создаётся на основании теории и математических симуляций. Дальнейшая оптимизация прототипа производится в ходе пошагового улучшения и тестирования на стандартной серии образцов после каждого конструкционного изменения. Аналогичный подход, включающий цикл «улучшение-тестирование», используется на следующем этапе разработки – при оптимизации конструкции анализатора для серийного производства. При этом заданная конфигурация каналов и техника измерения меняться уже не должны.

2.3 Сбор и низкоуровневая обработка данных ОМС

2.3.1 Сбор данных и их свойства

В целом процедура сбора данных ОМС та же, что и для полноспектральных данных в традиционной спектроскопии. Она состоит в регистрации вектора оцифрованных значений на выходе из детектора в схеме на рис. 1-3. Поскольку термин «спектр» в случае ОМС едва ли применим, будем называть этот вектор *измерением*. Необработанное (сырое) измерение, как и в большинстве методов оптической спектроскопии, представлено натуральными числами в диапазоне от 0 до верхней границы, определяемой выбранной разрядностью двоичного представления числа. Например, наиболее распространённое представление числа в 16-ти битах соответствует десятичному диапазону интенсивностей от 0 до 65535. Эти абсолютные значения интенсивности имеют единицы *счёта фотонов* (англ. counts). Абсолютная интенсивность также иногда выражается в процентах от верхней границы.

Чем больше двоичных разрядов используется для представления абсолютной интенсивности измерения, тем больше *динамический диапазон* анализатора. Динамический диапазон является по сути разрешением аналитического сигнала по интенсивности. Спектроскописты часто используют этот термин в качественном аспекте: говорят, что у анализатора большой, средний или малый динамический диапазон. Большой динамический диапазон на практике означает способность оптического метода различить малые различия спектральных интенсивностей. Это может быть особенно полезно, если в полученном наборе данных значимыми являются сигналы и низкой и высокой интенсивности, относящиеся, например, к разным компонентам смеси. В этом случае проблему малых разностей нельзя решить подбором настроек измерения.

Оптимальным для сбора данных является диапазон абсолютной интенсивности (в максимуме измерения) от 15% до 85%. Значения ниже 15% сильно подвержены воздействию различного рода шума, а при значениях выше 85% отклик детектора близок к насыщению и может страдать от нелинейности, создавая к тому же риск зашкала детектора для непредвиденно «светлых» образцов.

Параметры измерения должны быть подобраны таким образом, чтобы вся серия измерений по возможности лежала внутри указанного диапазона. Это достигается выбором соответствующей техники измерения (раздел 2.2.5) и *времени измерения* (накопления).

2.3.2 Нормирование мультисенсорных данных

Как и в случае спектральных данных, измерения мультисенсорных систем можно выразить в относительных единицах, таких как *пропускание* (T) или *оптическая плотность* (A), определяемых уравнением (2.1). Это преобразование подчёркивает разницу между измеряемым и эталонным образцами, отражающую изменение спектрального состава света при взаимодействии с анализируемым веществом. Эталонный образец называют также образцом сравнения, а саму процедуру преобразования спектра – *нормированием*. Распространёнными эталонными образцами являются вода, чистый растворитель, воздух и специальные рассеивающие (и при этом слабо поглощающие) материалы.

$$A = -\log_{10} T; \quad T = \frac{I}{I_0}; \quad I = C - D; \quad I_0 = C_0 - D \quad (2.1)$$

где D – темновой сигнал детектора, C и C_0 – нескорректированные абсолютные интенсивности спектров образца и эталонного образца соответственно; T и A – интенсивность в единицах пропускания и оптической плотности соответственно (при измерении в режиме диффузного отражения или рассеяния вместо T обычно используют символ R).

Нормирование в спектроскопии решает три задачи. Относительные единицы служат визуализации результата взаимодействия света с образцом, представляя его в виде традиционного спектра с характерными полосами поглощения, несущими химическую информацию. Это особенно важно для качественного спектроскопического анализа. С точки же зрения количественного анализа основная цель нормирования спектра – представить его в единицах оптической плотности, наиболее подходящих для построения градуировочных моделей. Кроме того, нормирование необходимо для того, чтобы сделать измерение независимым или минимально зависимым от нестабильности прибора, например колебаний

интенсивности источника и различных дрейфов.

В таких методах как флуоресцентная и КР-спектроскопия, основанных на измерении испускания, нормирование данных по уравнению (2.1) невозможно. Количественный анализ в этом случае осуществляется с использованием абсолютных интенсивностей, которые в общем случае не предполагают линейной зависимости от концентрации в силу физики явления и сильного влияния среды на слабый регистрируемый сигнал [180]. Приборная нестабильность может быть скомпенсирована математически – на этапе предобработки данных и при построении многомерной градуировочной модели. Однако в случае ОМС возможности коррекции приборных ошибок сильно ограничены малым количеством переменных (раздел 1.3.2). В настоящем исследовании предложен новый подход, использующий внутреннее эталонное измерение – так называемый безэталонный анализ, представленный в разделе 2.3.5.

В количественном анализе методами УФ/Вид-, БИК- и ИК-спектроскопии традиционно используют единицы оптической плотности, предполагающие линейную зависимость от концентрации по уравнению (1.2). Верхняя граница A обусловлена техническими характеристиками анализатора, и, как правило, не превышает трёх. Оптимальным с точки зрения минимизации шума и максимизации линейности и считают диапазон $0 \leq A \leq 2$. В идеальных условиях, к которым относится, пожалуй, лишь анализ сильно разбавленных растворов на пропускание (в качестве образца сравнения берётся чистый растворитель), A принимает неотрицательные значения, а T принадлежит диапазону $[0 \ 1]$.

Однако современная практика количественного анализа имеет дело с далеко не идеальными образцами, оптический анализ которых часто осложнён рассеянием, динамикой процесса (как при онлайн-измерении), сложностью подбора эталонного образца (раздел 2.3.4) и другими факторами [8]. Так, в данных становятся возможным существенное смещение A в область отрицательных значений, а концентрационная зависимость оптической плотности может отклоняться от линейности. Эти отклонения особенно характерны для ОМС в силу самой природы анализируемых ими объектов и их конструкционного

разнообразия, в том числе инновационных измерительных интерфейсов и современных эталонных образцов. Классический спектр поглощения в единицах оптической плотности, с нулевой базовой линией и линейной зависимостью сигнала от концентраций компонентов уже не является абсолютным требованием в современной оптической спектроскопии. Тем более это относится к измерениям ОМС, к которым сами понятия базовой линии и пика полосы поглощения не всегда применимы, а учёт функции отклика может быть частью математического аппарата обработки и анализа данных.

Таким образом, от измерений ОМС в общем случае не ожидается соблюдения закона линейности отклика по уравнению Бугера-Ламберта-Бера (1.2), а классические относительные единицы теряют свой физический смысл (так, отрицательная оптическая плотность означала бы усиление прошедшего через образец луча). Однако для анализа многомерных данных методами хемометрики первостепенное значение имеет не диапазон интенсивностей, а дисперсия спектральных переменных, и отрицательные оптические плотности в данных никак не влияют на прогностические свойства результирующей градуировочной модели. Как следствие, для построения градуировочных моделей могут быть использованы практически любые единицы интенсивности, а критерием выбора единиц становится минимизация погрешности прогноза с использованием результирующей модели.

Так, градуировочная модель, построенная на БИК-спектрах в единицах T оказалась заметно лучше модели для A [176]. Эмпирический подход, всё чаще применяемый на практике [8], становится возможным благодаря способности алгоритмов многомерной градуировки, в частности ПЛС регрессии, компенсировать нелинейность данных за счёт дополнительных ЛП, как показано в работах [8,181]. Метод проб и ошибок при выборе единиц представления спектров, методов предобработки данных и построения моделей в ещё большей степени свойственен при использовании ОМС. В таком подходе статистическое сравнение моделей становится важным инструментом принятия решений, а значит, возрастают требования к качеству тестового набора и надёжности методов

валидации.

Считая проблему нелинейности преодолимой средствами хемометрики, данные гипотетического анализатора, обладающего абсолютной стабильностью, не нуждаются в нормировании. В отдельных типах ОМС и приложениях такая стабильность может быть достигнута благодаря конструкционной простоте ОМС.

2.3.3 Обеспечение прецизионности измерения

Воспроизводимость или прецизионность измерения является одним из важнейших показателей качества данных. Очевидно, что высокая прецизионность необходима при повторных измерениях одним и тем же анализатором. Воспроизводимость этого уровня во многом обеспечивается стабильностью работы прибора и, по возможности, нормированием данных (раздел 2.3.2).

Если в построении и использовании градуировочной модели задействованы несколько однотипных (а иногда и различных) устройств, прецизионность измерений обеспечивается их *стандартизацией*. Когда аппаратная стандартизация невозможна, как это часто бывает в ОМС, она становится функцией ПО и осуществляется посредством переноса модели – коррекции полученных данных (прямая стандартизация) или значений прогноза уже после использования градуировочной модели. Стандартизация особенно важна для построения глобальных моделей прогноза на данных множества распределённых анализаторов, как в случае ОМС широкого применения (раздел 1.3.8). Прямая стандартизация требует разработки специального набора стандартных образцов – своего для каждого приложения ОМС. Как показано нами в работе [182], перенос модели методом коррекции наклона и отсечения (раздел 1.3.8) позволяет избежать сложной процедуры полной стандартизации и переносить модель с одного анализатора на другой с использованием нескольких обычных образцов молока.

Для методов, работающих в единицах абсолютной интенсивности, как флуоресцентная и КР-спектроскопия, существуют специальные методы стандартизации, позволяющие унифицировать полученные на различных приборах данные при помощи образцов специальных сертифицированных стандартных чистых веществ [183,184]. Однако эта стандартизация не защищает измерение от

ошибки, связанной с нестабильностью работы прибора, чаще всего – с временным дрейфом. Нескомпенсированные эффекты такого рода проецируются на многомерное пространство модели и ухудшают правильность прогноза. Поэтому при создании ОМС, не предполагающих нормирование данных, особое внимание следует уделять проблеме обеспечения стабильности измерения.

В отдельных случаях плохая прецизионность измерения может быть связана со свойствами образца, например при измерениях порошков и гранул [40,86,185]. Этот вид невоспроизводимости не может быть устранён средствами нормирования и стандартизации. Для его компенсации исследователи традиционно применяют методы коррекции рассеяния (раздел 1.3.2). Однако в настоящем исследовании было показано, что эти методы не всегда оптимальны, поскольку могут удалять из данных полезную информацию, содержащуюся в фоновой интенсивности. Предложен альтернативный метод предобработки таких данных, основанный на сглаживании переменных. Он будет подробнее рассмотрен в разделе 2.4.1. На примере онлайн-мониторинга концентрации воды в грануляте (раздел 5.2.2) в ходе его фармацевтического производства показано, что при использовании предложенного подхода можно достичь существенного повышения точности анализа.

К стандартизации следует также отнести приведение спектральных данных к универсальной спектральной шкале путём интерполяции значений интенсивности к определённой равномерной сетке по оси длин волн. Такая стандартизация часто используется в спектрометрах с ДМД, где невозможно обеспечить точное воспроизведение длин волн спектральных каналов. В УФ/Вид- и БИК-областях часто используется сетка целых значений в нанометрах (нм). Такая шкала удобна в использовании и достаточно подробна для целей количественного анализа [85]. Равномерная спектральная шкала облегчает разработку ОМС, например упрощает оптимизационный расчёт оптической конфигурации каналов ОМС (раздел 2.2.2). Что касается самих ОМС, интерполяция спектральной шкалы не применима к первому типу систем (табл. 1-3) с несколькими оптимизированными каналами, но может быть использована во втором [65] и третьем [91] типах.

2.3.4 Выбор эталонного образца

Важнейшим свойством эталонного образца является его *тиражируемость* – возможность получения любого числа дубликатов с идентичными оптическими свойствами. Неприемлемо использование уникальных, невозпроизводимых эталонов. Изменение свойств такого эталона, использованного при градуировке, делает новые данные несовместимыми, и проблема может быть решена только заменой или переносом модели.

Другим важным качеством эталонного образца является *устойчивость* к физическим, химическим или механическим воздействиям, возникающим в процессе измерения, а также отсутствие значимых временных изменений, старения.

Немаловажна также *доступность* эталона с точки зрения его стоимости, времени изготовления (поставки), а также лёгкости использования и замены. Доступность эталона становится важнейшим качеством в онлайн-анализе, к примеру, при помощи встроенного в трубу или реакторе зонда, когда стандартным образцом может быть только одна из естественных для выбранного процесса сред, таких как воздух, вода или промывочная жидкость. Примером неоптимального с точки зрения воспроизводимости эталонного образца (и всё же встречающегося на практике из-за его доступности) является стартовая среда при запуске процесса.

Для сложных объектов и условий анализа эталонный образец в его классическом понимании – как матрица образца без аналита – бывает недоступен. В качестве примера можно привести анализ молока методами рассеяния (раздел 5.1.2) или онлайн-мониторинг остаточной воды в грануляте в процессе его сушки через внедрённый в реактор зонд (раздел 5.2.3). В таких сложных случаях выбор эталона для количественного анализа часто индивидуален и диктуется достижением оптимального измерения, а именно, близостью абсолютных интенсивностей спектра сравнения и спектров анализируемой серии образцов. В идеале все они должны оставаться в оптимальном диапазоне 15–85% (раздел 2.3.1), а максимальные интенсивности не должны сильно различаться. Поскольку это не всегда возможно, в практических данных ОМС можно встретить значения

оптической плотности до -1 и ниже [186].

С ростом популярности методов спектроскопии диффузного отражения возрастает потребность в твёрдых эталонных материалах и образцах. Твёрдые эталоны выполняются в виде различных форм – таблетка, кольцо, насадка, вставка, в зависимости от конструкции интерфейса измерения. Для повышения точности измерения и скорости анализа эталонный образец может быть встроен в ОМС, а снятие спектра сравнения – автоматизировано [13].

В упомянутом выше примере анализ молока на содержание суммарных жира и белка проводился в режиме на пропускание в области 400–1100 нм. Из-за сильного рассеяния среды использование обычного для анализа жидкостей эталона – воды невозможно, поскольку разница интенсивностей света прошедшего через одинаковые по толщине слои молока и воды слишком велика. Проблема была решена использованием в качестве эталона оптического фильтра, равномерно ослабляющего излучение в рабочем диапазоне на 4 единицы оптической плотности [85]. В другом примере зонд «LHP» (раздел 1.1.3), использовавшийся для онлайн-мониторинга содержания воды в грануляте (раздел 5.2.2), был оснащён встроенным эталонным образцом, представляющим собой кольцо из стандартного материала на основе тефлона, опоясывающее рабочую часть зонда, и автоматикой для его применения без контакта со средой идущего процесса [13].

Приведённые данные показывают, что эталонный образец, зачастую воспринимаемый как стандартный аксессуар для лабораторной спектроскопии, должен быть частью разработки ОМС и вместе с ней оптимизироваться под целевое приложение.

2.3.5 Безэталонный количественный анализ

В отдельных случаях нормирование спектральных данных можно проводить с помощью внутреннего спектра (или канала) сравнения. Возможность такого *безэталонного* измерения должна обеспечиваться как спектральными свойствами образца, так и конструкцией анализатора. Простейшим примером безэталонного анализа является двухканальная ОМС для определения воды в порошках [75], описанная в разделе 1.2.3. Связанная с рассеянием нестабильность измерения на

полосе поглощения воды (I_{1940}) в этом случае компенсируется за счёт фонового сигнала на «пустой» длине волны 1300 нм (I_{1300}). В логарифмической форме эта процедура соответствует уравнению (2.1), где роль сигнала сравнения играет I_{1300} :

$$A = -\lg(I_{1940}/I_{1300}) \quad (2.2)$$

В работе [89] нами был разработан метод безэталонного определения жира и белка в молоке методом пространственно-разрешённой спектроскопии диффузного отражения. Пространственное разрешение в реализованном прототипе достигалось использованием специально разработанного оптоволоконного зонда, на плоскую торцевую поверхность которого в линию выведены восемь световодов кварцевого волокна, один из которых, крайний, был соединён с источником излучения (галогенная лампа) и служил для освещения образца, а остальные – для детектирования (рис. 2-2). Измерение проводили при прямом контакте рабочей поверхности зонда с кварцевой кюветой, содержащей образец молока. При данной геометрии измерения и с учётом дисперсности среды влиянием взаимодействия излучения со стенками кюветы и его потерями можно пренебречь, то есть считать объём образца бесконечным. Переключение пространственных каналов производилось при помощи мультипликатора, поочерёдно соединяющего их со спектрофотометром (детали метода см. в работе [89]).

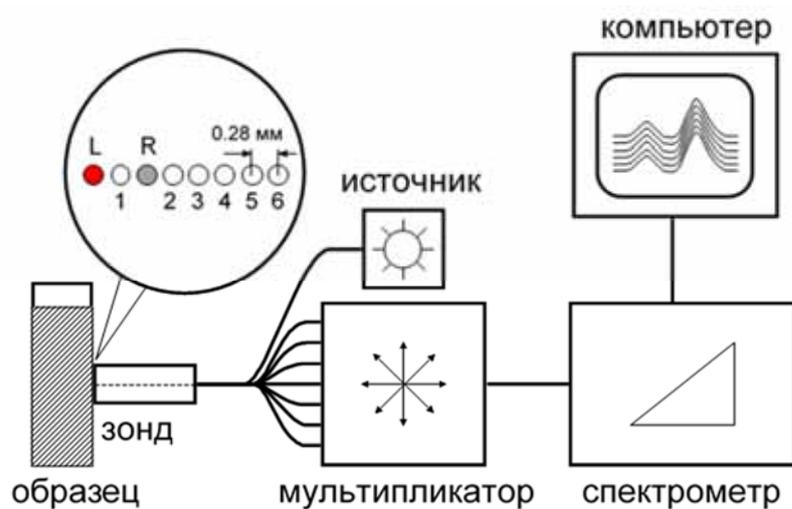


Рисунок 2-2. Экспериментальная установка для пространственно-разрешённых спектроскопических измерений. Расположение каналов на рабочей поверхности зонда обозначено цифрами от 1 до 6; L и R – каналы освещения и сравнения соответственно.

Для представления регистрируемых спектров в единицах оптической

плотности, согласно предложенному методу роль спектра сравнения I_0 в уравнении (2.1) отводится одному из детектирующих каналов. Предварительные измерения показали, что детектируемый сигнал канала C_1 , ближайшего к источнику света, имеет низкую по сравнению с другими каналами прецизионность (воспроизводимость детектируемого значения). Это было объяснено влиянием зеркального отражения от поверхности кюветы (бликами). Поскольку стабильность I_0 имеет огромное значение для точности анализа, было решено использовать в качестве канала сравнения следующее волокно, оставив C_1 в качестве измерительного канала. Полученные спектральные данные образовали трёхмерный массив нормированных спектров: 96 (измерений) \times 6 (каналов) \times 596 (спектральных переменных).

Внутренний спектр сравнения, получаемый на том же самом образце, успешно решает обе задачи спектрального нормирования для количественного анализа (раздел 2.3.2): представление спектров в относительных единицах оптической плотности и компенсация возможной нестабильности прибора, в частности, источника. Предложенный безэталонный метод и основанная на нём ОМС для пространственно-разрешённого анализа рассеивающих сред, в том числе молока, были запатентованы [79]. Апробация метода описана в разделе 5.1.6. Стандартные твердотельные материалы, традиционно используемые в качестве эталонных образцов диффузного отражения в Вид/КВБИК-области, не вполне подходят для спектроскопии оптически плотных жидких сред. В частности, в анализе молока выбор подходящего спектра сравнения представляет собой серьёзную экспериментальную проблему. Эти осложнения делают безэталонные методы анализа высоко востребованными.

Внутреннее нормирование ещё более актуально для флуоресцентной спектроскопии, где относительные единицы и эталонные образцы обычно не используются. Для компенсации различных факторов, негативно влияющих на стабильность интенсивности измеряемого эффекта люминесценции, таких как поглощение, рассеяние, неомогенность среды, перепады интенсивности источника, может использоваться интенсивность полосы возбуждающего

монохроматического излучения лазера или светодиода. Для этого оптический фильтр, обычно служащий для полного отсечения длины волны источника, не используется или подбирается таким образом, чтобы остаточный пик возбуждающего излучения имел интенсивность, сравнимую с интенсивностью флуоресценции. Этот метод, предложенный в нашей работе [81], был успешно применён для нормирования двумерного спектра флуоресценции хлорофилла А, представленного на рис. 2-3а спектрами испускания при разных длинах волн возбуждающего излучения – так называемым спектром возбуждения-эмиссии (СВЭ). Каждый из спектров флуоресценции матрицы СВЭ, полученный на 17-ти длинах волн возбуждения, был поделён на интенсивность в максимуме соответствующей полосы возбуждения. При этом, помимо компенсации уже перечисленных негативных эффектов, и происходит корректировка всего двумерного спектра по оси длин волн возбуждающего излучения. Использованный в этом исследовании источник-монохроматор имел сильные перепады интенсивности в зависимости от рабочей длины волны, что неизбежно сказывалось на форме регистрируемого спектра, в том числе на соотношении интенсивностей двумерных пиков флуоресцирующих компонентов.

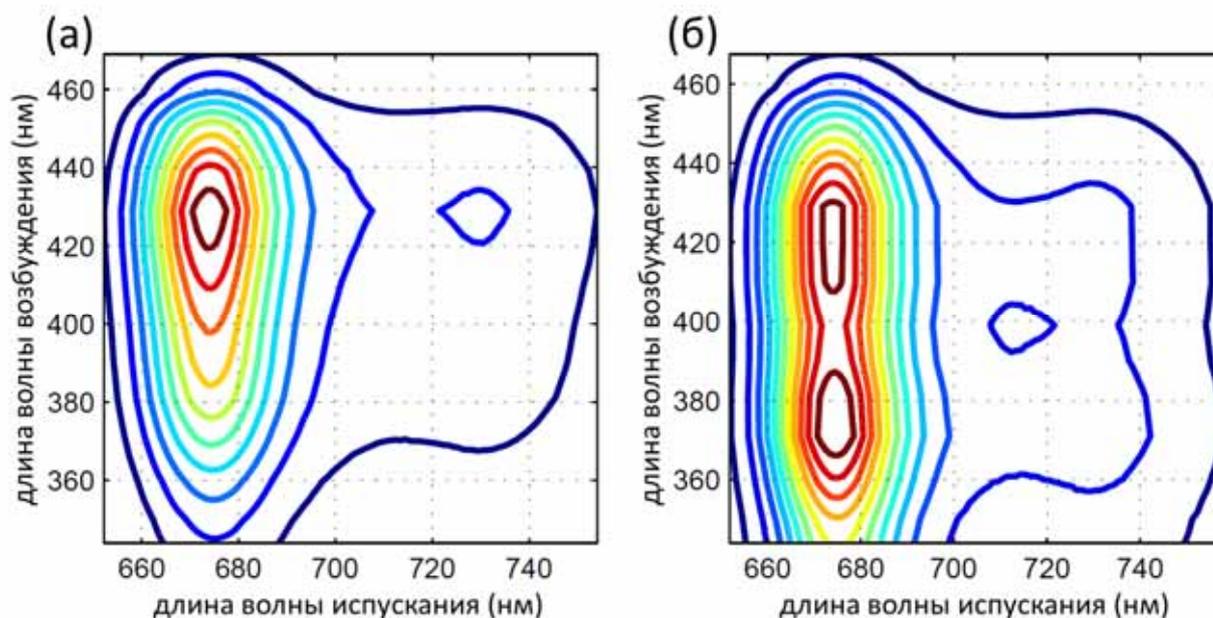


Рисунок 2-3. Спектр возбуждения-эмиссии флуоресценции водного раствора хлорофилла А при концентрации $4 \cdot 10^{-2}$ г/л (а) для исходных и (б) нормированных по интенсивности данных [81].

На рис. 2-3 показаны исходный спектр и результат произведённой

корректировки. Внутреннее нормирование привело к разрешению двух пиков флуоресценции, что прекрасно согласуется с литературными данными [187,188].

2.4 Построение и валидация моделей в процессе разработки ОМС

2.4.1 Предобработка мультисенсорных данных

Малое число переменных и их неравномерное распределение вдоль спектральной оси делает большинство методов предобработки неприменимыми или неэффективными для ОМС (раздел 1.3.2). Два наиболее эффективных и часто применяемых в спектроскопии метода улучшения данных перед анализом – спектральное сглаживание и взятие производной также не могут быть применены к данным большинства ОМС. Даже для систем второго типа (табл. 1-3), данные которой содержат узкие спектральные интервалы, применение традиционных методов сглаживания, таких как СГ, не рекомендуется из-за влияния краевых эффектов [99].

Это ограничение может быть преодолено при оптическом спектральном анализе временных серий, относящихся к наблюдению протекающего процесса: технологического или природного. Сглаживание и производная при этом могут быть применены не в традиционном аспекте – к спектральному вектору, а к векторам переменных матрицы данных, то есть во временном домене. Отметим, что такой способ сглаживания полностью согласуется с природой приборного шума – случайными колебаниями показаний каждого оптического канала во времени, не зависящими от других каналов. Использование такого подхода является абсолютно корректным и в случае полноспектральных данных, образующих связную временную серию. Связность в данном контексте означает, что данные снимаются с достаточно мелким шагом по времени, то есть интервалы между соседними измерениями намного короче, чем значимые события в наблюдаемом процессе.

Эффективность временного сглаживания можно проиллюстрировать на следующем примере. Качество спектров диффузного отражения, полученных при пробном запуске ОМС типа «упрощенный спектрометр» (тип 3 в табл. 1-3) в

производственную линию для онлайн-мониторинга концентрации воды в порошкообразной пищевой добавке (детали не разглашаются по требованию производителя), оказалось низким из-за сильного шума [91]. Сглаживание спектров не могло быть применено, поскольку данные содержали только 40 спектральных переменных (рис. 2-4). К тому же анализ усложнялся неомогенностью и перепадами плотности проходящего перед оптическим окном ОМС продукта. Разведочные методы анализа, применённые на сырых данных, не выявили внутренней структуры, которая могла бы быть полезной для построения градуировки.

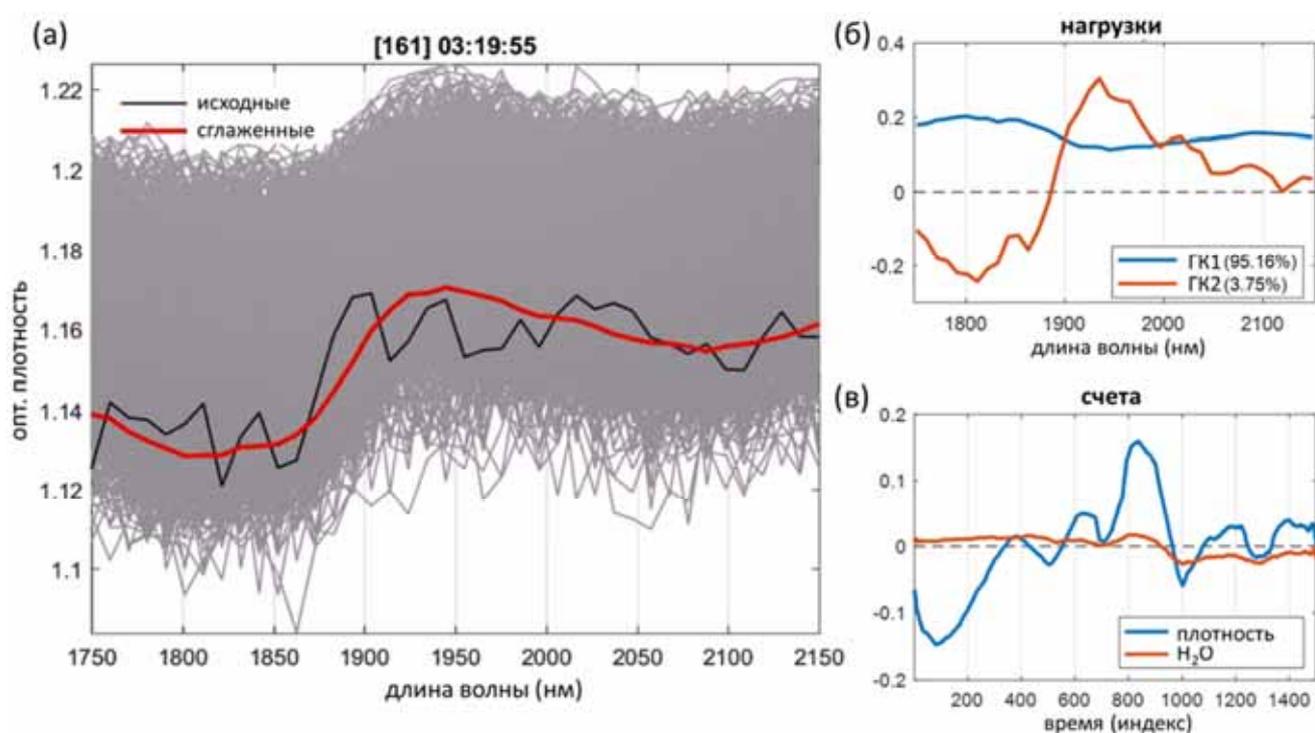


Рисунок 2-4. (а) Спектральные данные процесса: все (серый цвет), исходный спектр 161 (чёрный), спектр 161 после сглаживания данных во временном домене методом движущегося среднего с окном 101 точка (красный); (б) МГК-нагрузки и счета, полученные для сглаженных данных.

Однако после сглаживания переменных во временном домене МГК-нагрузки на ГК2 показали спектральные черты, характерные для поглощения воды (рис. 2-4 б), что свидетельствует о возможности её количественного определения данным методом. При этом ГК1, по всей видимости, отражает наблюдаемые перепады спектрального фона, реагирующего на плотность потока порошка (эффект рассеяния). В этом предварительном исследовании градуировка не проводилась из-за отсутствия данных референтного метода о содержании воды. В

приведённом примере применённая предобработка данных сыграла решающую роль в обосновании пригодности метода и в принятии решения о его дальнейшей разработке.

Следует учитывать, что использование сглаживания вдоль временной оси связано с задержкой в выдаче результата анализа, равной половине сглаживающего окна (аналогичный параметр есть и в других методах сглаживания, например в алгоритме СГ). Ширина окна подбирается исходя из динамики процесса – общего времени и скорости протекания важных для мониторинга событий. В приведённом выше примере ширина окна в 101 точку соответствовала промежутку времени 0.4 с, что заведомо приемлемо для мониторинга данного процесса. Подробнее проблема оптимизации ширины сглаживающего окна рассмотрена в разделе 5.2.2 и в работе [86].

Аналогично сглаживанию во временном домене может быть взята производная, а также применены другие методы предобработки. Этот подход представляет интерес не только для ОМС, но и в анализе полноспектральных данных, особенно при исследовании процессов. Влияние различных предобработок во временном домене на точность работы градуировочных и других прогностических моделей изучено недостаточно, и это должно стать предметом дальнейших исследований.

При использовании данных в абсолютных единицах, что более характерно для ОМС, чем для традиционных спектральных методов, в качестве предобработки может быть целесообразно преобразование измеренных значений какой-либо функцией, например логарифмирование. Функция может быть подобрана эмпирически с целью повышения точности результата анализа.

2.4.2 Особенности построения градуировочных моделей для ОМС

Данные ОМС, особенно типов 1 и возможно 3 (табл. 1-3), отличаются относительно малым числом переменных в матрице данных X , а также ожидаемым отсутствием существенных корреляций между ними, благодаря проведённой оптимизации. Эти свойства роднят данные ОМС с несектральными аналитическими данными, в которых число переменных тоже может быть

существенно сокращено без значимых потерь для результата анализа [189,190]. Уменьшение числа переменных в данных не влияет на применимость ПЛС-регрессии и большинства других алгоритмов многомерной градуировки, но создаёт предпосылки для использования классического метода множественной линейной регрессии (МЛР), широко применявшегося до появления проекционных методов [53, с. 125]. МЛР не использует факторную проекцию, решая регрессионное уравнение (1.3) напрямую из \mathbf{X} и \mathbf{y} :

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad (2.3)$$

По имеющимся данным, МЛР на предварительно отобранных спектральных переменных даёт отличные результаты [191,192], а его преимуществом является простота и меньшая расчётная нагрузка, что может быть существенным преимуществом для автономных сенсоров, использующих облегчённые варианты компьютеров.

Из-за невысокого числа переменных построение градуировочных моделей на данных ОМС и последующий прогноз по уравнению (1.4) занимает меньше времени. Это может быть существенным фактором при анализе «больших данных» и онлайн-мониторинге быстрых процессов.

2.4.3 Многоуровневая валидация модели

Важность тщательной валидации моделей в процессе разработки и применения ОМС неоднократно отмечалась в предыдущих разделах. Достоверная проверка на специально разработанном стандартном наборе образцов должна предоставлять объективные критерии для оценки и сравнения предсказательной способности различных моделей. Многочисленные промежуточные модели, создаваемые в процессе разработки ОМС на этапах создания прототипа и оптимизации производства промышленного анализатора, должны сравниваться между собой и с исходной моделью полноспектрального «золотого стандарта».

В настоящем разделе описан подход, названный *многоуровневая валидация*, позволяющий увеличить достоверность проверки модели на ограниченном (а это как правило так) тестовом наборе. Подход предполагает наличие в данных определённой иерархической структуры, возникающей при планировании

эксперимента. Иерархия в данных возникает из-за необходимости систематического учёта различных факторов variability, таких как прецизионность измерения, негомогенность образца, его источник и т. д.

В качестве примера рассмотрим набор данных, измеренных 4-канальной ОМС на светодиодах. Целью исследования была проверка применимости метода для диагностики рака почки у человека [27,78]. Для одной серии измерений (серия «В») были предоставлены клинические образцы 4-х пациентов⁴, полученные в результате нефрэктомии. У каждого пациента были взяты образцы здоровой и поражённой ткани. Поскольку биологическая ткань сильно негомогенна, а область анализа встроенным в ОМС зондом практически точечная, измерения проводились в нескольких позициях каждого образца, обозначенных в соответствии с сеткой разметки (A1, B2 и т. д.). В каждой позиции было сделано 5 повторных измерений, при которых зонд отводился от позиции измерения и позиционировался вновь, имитируя действия врача в время обследования. Детали этого исследования подробно изложены в работе [27] и разделе 5.4.

Полученные таким образом данные имеют иерархическую структуру с 4-мя уровнями вложенности, соответствующими измерению, позиции, образцу ткани и пациенту (рис. 2-5).

Согласно предложенному подходу, для проверки модели (в приведённом примере это была ПЛС-ДА модель для дискриминации измерения по признаку рак/здоровая ткань) следует провести несколько этапов валидации методом ПП сегментами (раздел 1.3.7), по одной проверке на каждый из существующих иерархических уровней. При этом разбивка на сегменты производится по группам измерений соответствующего уровня. В приведённом на рис. 2-5 примере это ППП на уровне «измерение», а также три ПП с сегментами, образованными измерениями: в одной позиции, одного образца или одного пациента на

⁴ Разрешение этической комиссии EA1/134/12 клиники «Charité» (г. Берлин, ФРГ)

соответствующих уровнях. Каждая валидация является тестом устойчивости модели к одной из ситуаций будущего практического анализа, когда модель сталкивается с новыми данными, не участвовавшими в её обучении: новым измерением, новой позицией и т. д. Важно учесть, что для использования в валидации уровень иерархии должен содержать достаточное количество соответствующих групп измерений, то есть он должен быть адекватно представлен. Уже в приведённом случае с 4-мя пациентами уровень «пациент» может быть поставлен под сомнение. При меньшем же числе пациентов он должен быть исключён из ПП как не охваченный экспериментально.

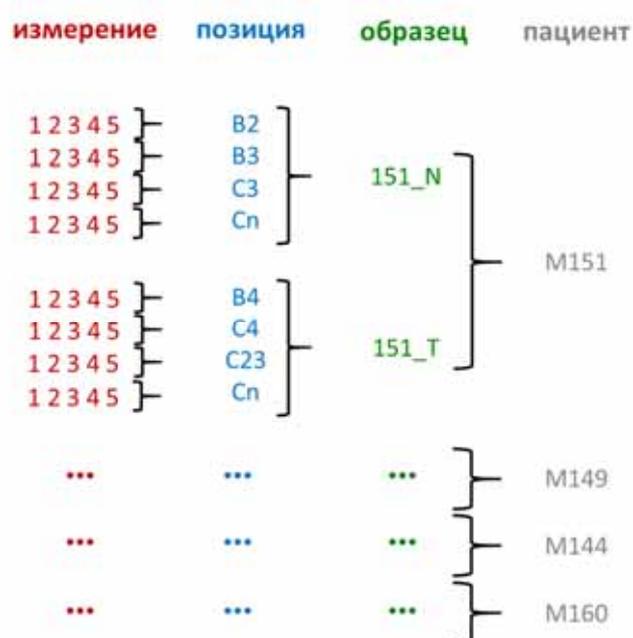


Рисунок 2-5. Структура данных прототипа ОМС для диагностики рака почки.

Для удобства сравнения СКО проведённых ПП сводятся в одну таблицу. В ту же таблицу вносится *СКОГ* и *СКОП* по результатам ПТН (если есть), а также значения коэффициента детерминации R^2 всех проверок. Это особенно важно если сравниваются модели, построенные на различающихся наборах данных. Анализ результатов многоуровневой валидации проводится по нескольким критериям, существенное отклонение которых от нормы может потребовать дополнительных действий. Возможные отклонения и их причины представлены в табл. 2-2. Рекомендуется повторить весь цикл проверок при различных значениях числа ЛП в модели. Результаты многоуровневой валидации важны для оценки качества данных и эффективности аналитического метода, а их сравнение позволяет

выбрать наилучший с учётом структуры данных метод проверки для его использования в качестве основного критерия при сопоставлении анализаторов и моделей. Подход может быть также использован для выбора оптимального числа ЛП в регрессионной модели. В публикациях, как правило, используются лишь наиболее значимые данные [27].

Таблица 2-2. Проверка результатов многоуровневой валидации.

№	Критерий	Отклонение	Возможная причина
1	Расхождение <i>СКОПП</i> между собой	> 1.5 раза	недостаточное качество данных
2	<i>СКОПП</i> «пессимистичной» ПП	выше допустимого	недостаточно данных или неоптимальное ПЭ
3	<i>СКОПП</i> «оптимистичной» ПП	выше допустимого	недостаточно информативный аналитический метод
4	Расхождение <i>СКОПП</i> и <i>СКОГ</i>	> 2 раза	неоптимальное число ЛП
5	Расхождение <i>СКОПП</i> и <i>СКОП</i>	> 1.5 раза	неоптимальный тестовый набор
6	Оптимальное число ЛП в модели разными ПП	разное	недостаточно данных или непредставительные данные

Примечание: приведены ориентировочные значения отклонения для гипотетической задачи построения градуировочной модели для определения процентного содержания компонента смеси при значениях *СКОПП* <10 («пессимистичный» метод ПП).

Значительные расхождения в СКО «пессимистического» и «оптимистического» методов ПП (выдавших наибольшую и наименьшую погрешность соответственно) может свидетельствовать о недостаточном качестве данных, например о присутствии в них выбросов или об их недостаточной представительности, не отражающей всех уровней вариабельности, представленных экспериментальной иерархией, а также о наличии неучтённых при ПЭ факторов. Например, существенно более высокая погрешность на верхнем уровне иерархии может говорить о недостаточном количестве присутствующих на нём групп образцов, а заниженная оценка на первом уровне может свидетельствовать о том, что образующие соответствующие группы повторные измерения слишком похожи друг на друга и, возможно, излишни. Несоответствие лишь «пессимистичной» оценки требованиям по точности прогноза, предъявляемым к разрабатываемой ОМС, может сигнализировать о необходимости дополнить данные новыми измерениями. Если же и «оптимистичная» оценка не

удовлетворяет требованиям, то должна быть поставлена под сомнение сама работоспособность аналитического метода. В хорошо сбалансированных данных, при правильно определённом числе ЛП в модели *СКОГ* и *СКОП* (при наличии тестового набора) не должны сильно отличаться от *СКОПП* любого метода. Чрезмерная «оптимистичность» или, напротив, «пессимистичность» *СКОП* может быть знаком неоптимального тестового набора. Если разные ПП дают отличающиеся значения оптимального числа ЛП, это также может быть следствием дефекта данных.

Приведённые в табл. 2-2 численные критерии расхождения различных СКО весьма условны. Эти границы должны определяться аналитиком с учётом конкретной решаемой задачи. При валидации дискриминационной модели, как в приведённом выше примере ОМС, сравнению подлежат статистические критерии точности, селективности и т. д. [27,78,88], выводимые из числа ложных отнесений, как положительных, так и отрицательных (раздел 1.3.7).

Сравнение методов ПП и ПТН позволяет выбрать наилучший с учётом структуры имеющихся данных метод валидации для его использования в качестве основного критерия сравнения моделей.

2.4.4 «TPT cloud» – комплекс программ для анализа многомерных данных и разработки ОМС «в облаке»

Заложенные в концепцию ОМС миниатюризация и снижение стоимости анализа (табл. 1-1) открывают возможности для широкого тиражирования устройств и их массового использования в распространённых практических приложениях, например для определения питательной ценности молока. Это создаёт необходимость, с одной стороны, обеспечивать множество систем, распределённых по всему миру, актуальной моделью, а с другой – возможность постоянного обновления этой универсальной модели с использованием данных, получаемых на местах отдельными работающими ОМС. В настоящей работе концепция глобального моделирования (рис. 2-6), предполагающая централизованную работу с данными распределённых анализаторов, была реализована на базе Самарского государственного технического университета

(СамГТУ) в комплексе программ «ТРТ cloud» [193-195]. Это ПО не требует установки и работает на удалённом сервере («в облаке») через стандартный интернет-браузер.

«ТРТ cloud» обладает широким набором алгоритмов общего назначения, необходимых для решения практических задач анализа спектральных и других многомерных данных. К ним относятся: загрузка из различных источников и форматирование, проведение предобработки и разведочного анализа, построение и хранение градуировочных и классификационных моделей, а также их использование для прогноза на новых данных, в том числе на удалённом устройстве через протокол ИПП.

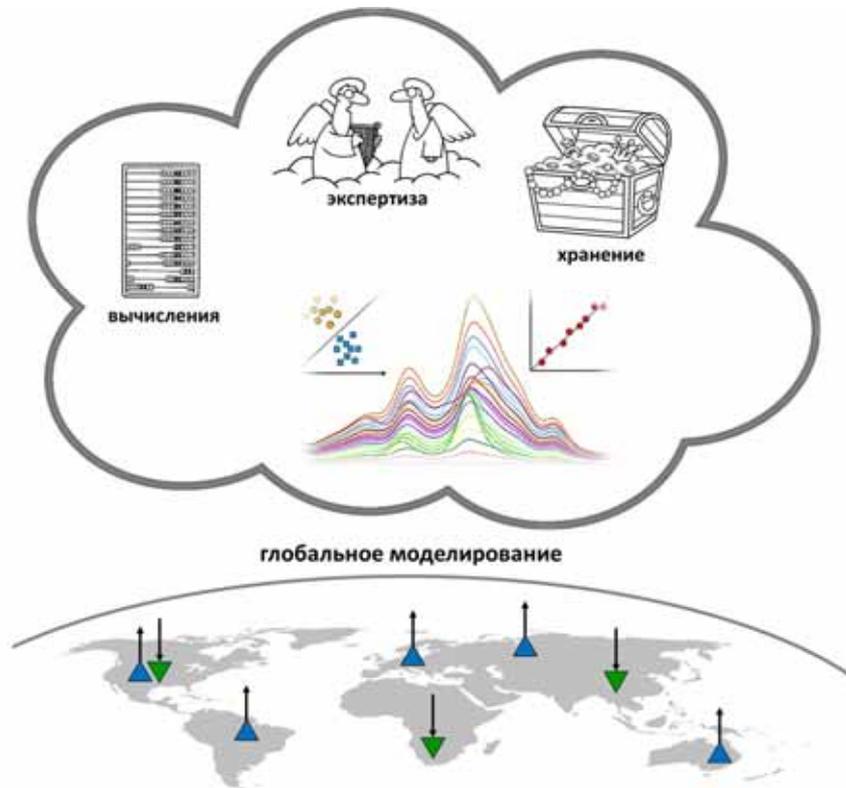


Рисунок 2-6. Схема глобального моделирования для распределённых ОМС.

Программный комплекс оснащён специальным инструментарием для разработки и использования ОМС, в частности оптимизацией каналов на основе полноспектральных данных (раздел 2.2.3), с возможностью включить в расчёт параметры предобработки (раздел 2.2.4), а также опцией сохранения построенных моделей в виде исполняемого кода в форматах, «понимаемых» ПО действующих ОМС. Для сохранения данных и моделей «в облаке» используется специально

разработанная объектная структура, обеспечивающая максимальную связность информации и её пополняемость.

С программой можно ознакомиться по ссылке tptcloud.com.

ЗАКЛЮЧЕНИЕ К ГЛАВЕ 2

Формализован процесс создания ОМС и сформулированы задачи, решаемые на различных этапах разработки. В рамках мультисенсорного подхода предложены отдельные решения, позволяющие облегчить процесс разработки и повысить прогностическую точность результирующих анализаторов.

В отличие от общепринятого априорного метода задания оптической конфигурации разрабатываемой ОМС, когда число и свойства каналов системы определяется экспертом в аналитической спектроскопии, предложено проводить строгую математическую оптимизацию параметров системы для выбранного практического приложения на предварительно полученных полноспектральных данных спланированного эксперимента.

Разработано улучшенное алгоритмическое решение проблемы расчёта оптимальных параметров оптических каналов мультисенсорных систем: центральной длины волны, ширины, допустимой степени перекрытия и других. Оно позволило сократить время вычислений в сотни раз и проводить оптимизацию для большего числа параметров, давая тем самым выигрыш в точности результирующего анализатора. Независимое значение имеет предложенное и практически подтверждённое усовершенствование в области алгоритмов отбора переменных. Согласно ему интервальный отбор переменных по вышеописанному ускоренному алгоритму проводится с одновременной оптимизацией параметров предобработки спектров, давая увеличение точности прогностических моделей, что важно для оптического спектрального анализа в целом.

Даны рекомендации по обеспечению качества получаемых данных ОМС, касающиеся выбора эталонного образца, первичной низкоуровневой обработки и стандартизации измерения. Предложен и апробирован безэталонный метод анализа, использующий один из пространственно-разрешённых каналов

мультисенсорной системы в качестве канала сравнения. Разработаны специальные подходы к предобработке и анализу данных ОМС, учитывающие их особенности. Предложен метод многоуровневой валидации для более достоверного сравнения прогностических моделей с использованием небольшого числа тестовых образцов.

Разработан централизованный подход к математическому обеспечению процесса создания ОМС – от оптимизации каналов до построения рабочих моделей и их практического использования. Подход реализован в разработанном комплексе программ «ТРТ cloud» для анализа многомерных данных «в облаке» – хемометрической платформе для централизованного хранения данных и построенных на них моделей с возможностью их глобального использования распределёнными ОМС.

3. ПЛАНИРОВАНИЕ ГРАДУИРОВОЧНОГО ЭКСПЕРИМЕНТА

Современные спектральные и сенсорные технологии в сочетании с методами анализа многомерных данных всё чаще применяются для количественного анализа сложных смесей. Их точность напрямую зависит от организации данных, взятых для построения и проверки модели. Настоящая глава посвящена решению весьма важной задачи ПЭ в количественном анализе – созданию обучающего и проверочного наборов образцов для построения точных градуировочных моделей по нескольким компонентам одновременно.

3.1 Введение в планирование многокомпонентного градуировочного эксперимента

3.1.1 Понятия и определения

В принятой терминологии ПЭ *факторами* являются изменяемые параметры, границы изменения которых очерчивают *экспериментальное пространство*. В многокомпонентном градуировочном эксперименте это, как правило, концентрации анализируемых компонентов. Однако в экспериментальное пространство могут быть включены и другие компоненты смеси, определение которых не требуется, но учёт их влияния на спектральное измерение необходим для построения устойчивой модели. Границы экспериментального пространства задаются при постановке задачи ПЭ в виде интервалов отдельных факторов, в которых должна быть обеспечена высокая точность и устойчивость градуировочной модели.

Любой метод ПЭ, в том числе градуировочного, определяется неким алгоритмом или подходом к заполнению экспериментального пространства – *дизайном*. Применение дизайна с учётом требований конкретной градуировочной задачи приводит к плану эксперимента, который принято представлять в виде *матрицы дизайна* $C (N \times k)$, где N и k – число образцов и факторов соответственно.

Многие дизайны используют предварительное разбиение факторов по осям экспериментального пространства на дискретные *уровни*, на которых и происходит заполнение образцами. Безуровневые схемы используются, например при

случайном заполнении или в алгоритме Кирсанова и др. [157]. В схемах с предустановленными уровнями элементы матрицы S удобно выражать в условных единицах – целых числах, обозначающих индексы уровней. При этом пересчёт в новый интервал концентраций компонента производится под конкретную задачу. Дизайн обычно предписывает равное число уровней l по каждому фактору, что делает выраженное в индексах уровней экспериментальное пространство квадратом, кубом или гиперкубом.

3.1.2 ПЭ как оптимизационная проблема

Цель ПЭ всегда состоит в решении некоторой аналитической задачи минимальным числом образцов и измерений. Таким образом, речь идёт об оптимизации с точки зрения эффективного использования усилий и средств.

Однако часто ПЭ решает и прямую оптимизационную задачу с использованием математических алгоритмов. В классическом ПЭ оптимизация чаще всего направлена на достижение экстремума внешней целевой функции (например максимального выхода химической реакции) от варьируемых условий эксперимента. В случае же градуировочного эксперимента основной задачей является получение наилучшей модели, то есть снижение ошибки последующего прогноза. Сложность проблемы заключается в том, что оптимальный набор образцов должен быть предложен до эксперимента. Частично с этой задачей справляется *функция размаха* (англ. leverage function), о которой будет речь в разделе 3.1.3. Но она принимает во внимание лишь взаимное расположение образцов и не учитывает множества других факторов, влияющих на точность градуировки. В целом же задачу выбора оптимального с точки зрения точности прогноза градуировочного набора едва ли можно решить чисто алгоритмически. Поэтому путь к построению «хорошей» градуировки лежит через снижение риска внесения различных ошибок. Этого можно добиться, следуя определённым требованиям, изложенным в следующих разделах.

3.1.3 Оценка результата ПЭ

Целью планирования градуировочного эксперимента является получение модели, дающей достаточную (в идеале – максимальную) точность прогноза при её

последующем практическом применении. Сложность оптимизации схемы ПЭ по отношению к ошибке прогноза заключается, во-первых, в необходимости проведения эксперимента для оценки результирующей ошибки любой предложенной схемы и, во-вторых, в отсутствие объективного критерия сравнения различных схем ПЭ. Для оценки ошибки прогноза могут быть использованы валидация независимым (внешним или являющимся частью дизайна) набором или ПП. В первом случае *СКОП* отражает в том числе качество самого проверочного набора. Так, более низкая ошибка при сравнении различных схем построения градуировки может означать (но может и не означать) недостаточную представительность самого набора и наоборот. А при использовании ПП проверочный набор по своему дизайну совпадает с обучающим и тоже подвержен внутренним ошибкам выбранной схемы ПЭ. Неоднозначность результатов делает экспериментальное сравнение схем ПЭ затруднительным, и при выборе надлежащей схемы экспериментаторам часто приходится действовать «вслепую».

Чисто математическим способом априорной (без эксперимента) оценки вклада каждого образца в достоверность последующего прогноза является функция размаха h . Значение h определяется диагональными элементами матрицы проекции:

$$\mathbf{H} = \mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T \quad (3.1)$$

где \mathbf{D} – это матрица со столбцами, включающими все члены уравнения модели аналитического отклика (сами факторы, перекрёстные и квадратичные члены и т. д.), как описано в книге Бреретона [131, с. 47].

Размах можно также выразить уравнениями (3.2) [146]:

$$h = x[\mathbf{D}^T\mathbf{D}]^{-1}x^T; x = [1 \ c_1 \ c_2 \ \dots \ c_i \ c_1c_2 \ \dots \ c_1c_i \ c_2c_i \ c_1^2 \ c_2^2 \ \dots \ c_i^2] \quad (3.2)$$

где c_1, c_2, \dots, c_i – концентрации компонентов 1, 2, ..., i . Подогнав эту функцию к точкам – образцам, значение h можно затем вычислить в любой другой точке экспериментального пространства и за его пределами. В предположении линейного аналитического отклика функция размаха является квадратичным многочленом, выраженным уравнениями (3.3) для двух и (3.4) – для трёх факторов:

$$h = b_0 + b_1c_1 + b_2c_2 + b_{11}c_1^2 + b_{22}c_2^2 + b_{12}c_1c_2 \quad (3.3)$$

$$h = b_0 + b_1c_1 + b_2c_2 + b_3c_3 + b_{11}c_1^2 + b_{22}c_2^2 + b_{33}c_3^2 + b_{12}c_1c_2 + b_{13}c_1c_3 + b_{23}c_2c_3 \quad (3.4)$$

где b_0, b_1, \dots, b_{12} – коэффициенты уравнения.

Чем ниже h , тем выше ожидаемая достоверность прогноза в соответствующей точке экспериментального пространства. Обеспечивая связь между взаимным расположением образцов в пространстве эксперимента и достоверностью прогноза, размах является универсальной мерой баланса схемы ПЭ. В хорошо сбалансированных конструкциях h сохраняет одинаково низкие значения во всей области и описывается простой и симметричной функцией, что указывает на равномерность данных и независимость факторов. В идеальном случае значение размаха приближается к нулю (хотя никогда его не достигает), а $h \geq 1$ указывает на низкую достоверность прогноза [141]. h -функция может быть также весьма полезна для сравнения различных схем ПЭ.

3.2 Требования к набору градуировочных образцов

Настоящий раздел посвящен систематизации различных требований, предъявляемых к набору образцов, используемых для и проверки градуировочной модели. Предложены общие критерии оценки качества дизайна градуировки [196].

3.2.1 Некоррелированность факторов

Попарные корреляции между факторами – столбцами МД – должны быть сведены к минимуму. Это наиболее универсальное требование, относящееся также к классическому ПЭ. Но в случае градуировочного эксперимента оно имеет первостепенное значение: только так можно избежать опосредованной зависимости (*англ.* confounding), когда регрессионная модель для одного из компонентов основана на его статистической корреляции с другим, а не с аналитическим сигналом прибора. Результирующая модель может длительное время успешно предсказывать концентрацию первого компонента, но она даст сбой, как только корреляция исчезнет или ослабнет. Отсюда возникает ещё одно важное требование – коррелирующие факторы должны быть явно включены в экспериментальный план. Следовательно, спектральный количественный анализ смесей должен начинаться с проверки на присутствие коррелирующих факторов.

Степень максимально допустимой корреляции между градуируемыми факторами зависит от приложения. Допуск может быть ослаблен, если корреляция компонентов основана на причинно-следственной связи, то есть имеет физическое, химическое или биологическое объяснение. Так, значения содержания жира и белка в стандартных образцах натурального молока показывают корреляцию выше 0.7 [85], однако её значение может меняться в зависимости от источника образцов [72].

3.2.2 *Равномерность заполнения*

Хорошо сбалансированный градуировочный набор должен быть достаточно представительным и состоять из множества образцов, равномерно распределённых по осям факторов. Следовательно, существует необходимость в достаточно большом количестве уровней планирования. На практике матрица дизайна C ($N \times k$) может включать десятки или сотни измерений, в зависимости от аналитической проблемы, сложности модели и доступности образцов.

С точки зрения математики минимальное число различных образцов (а следовательно и уровней) для построения регрессии должно превышать сложность модели на единицу. Под сложностью или размерностью в полиномиальной регрессионной модели между x и y понимают степень полинома. Аналогичным параметром в многомерной регрессии является число ЛП в факторном пространстве модели. Так, по двум точкам можно построить лишь линейную регрессию (или модель с одной ЛП), по трём – квадратичную функцию (или модель не более, чем с двумя ЛП) и т. д. Очевидно, что такое минимальное число уровней является недостаточным с точки зрения статистики, даже если отдельные уровни заселены многими образцами. Так, в простой линейной градуировке (размерность модели единица) построение прямой даже по многим образцам, расположенным на двух крайних уровнях, например концентрационных, не будет оптимальным. Во-первых, ошибка градуировки в общем случае зависит от значения градуируемого фактора, то есть она неравномерно распределена по его шкале, и её статистическая оценка требует промежуточных экспериментальных точек. Во-вторых, линейность отклика обычно является гипотезой, которая может быть доказана только

экспериментально с использованием достаточного большого числа уровней. Те же причины иметь в градуировке число уровней, значительно превышающее сложность модели, действуют и в случае многомерной регрессии. Кроме того, если многомерная регрессия строится на малом числе уровней, содержащих группы сходных образцов, существует высокий риск переопределения модели (раздел 1.3.7). В этом случае модель начинает описывать спектральные различия между группами образцов как таковыми, а они могут не быть связаны с градуируемым фактором.

Равномерное заполнение экспериментального пространства образцами в общем случае не требует соблюдения абсолютной равномерности в смысле равенства дистанций между образцами вдоль любого фактора. Образцы могут следовать любой схеме при условии, что она учитывает функцию отклика измерения [197], например линейность поглощения по уравнению (1.2).

Необходимое число уровней обучающего набора градуировки зависит и от числа переменных в спектральных данных. При избыточном числе каналов, как в спектроскопии высокого разрешения, корреляция между ними нивелирует влияние шума отдельных каналов на градуировочную модель, то есть на точность прогноза. Поэтому чем меньшим числом оптических каналов представлены данные, тем больше сходных измерений требуется для обеспечения той же точности. Иными словами, недостаток спектрального разрешения компенсируется более высоким «разрешением» по концентрации компонентов в обучающем наборе. Требования к минимальному числу градуировочных образцов для разных типов спектральных данных устанавливаются нормативными документами. Например, ASTM устанавливает требования к минимальному числу градуировочных образцов при различном числе переменных в ИК-спектрах [147]. Из этого следует, что ОМС в целом требуют большего числа измерений для получения градуировочной модели, что должно учитываться в ПЭ.

3.2.3 Покрытие экспериментального пространства

Регрессионная модель предназначена для прогноза на новых образцах, сходных с образцами обучающего набора градуировки. Множество образцов

обучающего набора можно представить в виде выпуклой фигуры, вписанной в квадрат (куб, гиперкуб) экспериментального пространства. Предполагается, что эта фигура занимает значительную область экспериментального пространства, образованную интервалами возможного измерения составляющих его факторов. Если новый образец находится за пределами границ области обучения, прогноз является экстраполяцией, и его достоверность не гарантирована. Поэтому одна из задач ПЭ – максимально полное *покрытие* экспериментального пространства образцами.

Экспериментальное пространство будет заполнено на 100%, если все вершины образуемого им многоугольника (многогранника) содержат образцы. В противном случае процент покрытия следует считать как отношение объёма выпуклого k -мерного многогранника, заданного образцами, к полному объёму экспериментального пространства.

Расчётный пример случайно-равномерного заполнения двумерного пространства эксперимента 25-ю образцами проиллюстрирован на рис. 3-1. Случай плохого покрытия приведён на рис. 3-1 а; корреляция между факторами при этом тоже слишком высока. Судя по гистограмме на рис. 3-1 б, встречаемость значений покрытия менее 50% в случайном методе заполнения все ещё нельзя считать маловероятной.

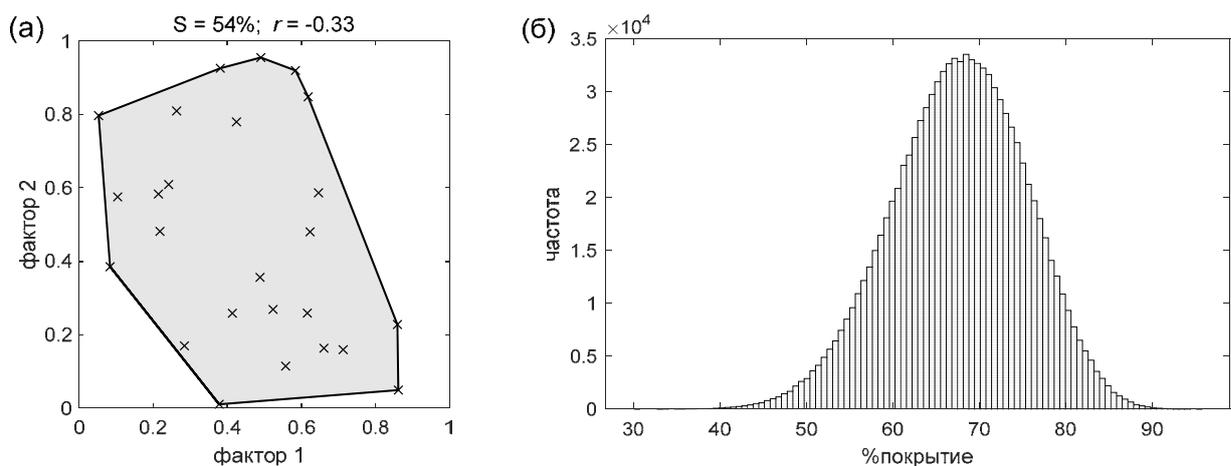


Рисунок 3-1. (а) Пример неоптимального покрытия двухфакторного экспериментального пространства при случайном заполнении 25-ю образцами и (б) гистограмма распределения значений покрытия для того же дизайна (10^6 попыток) [196].

Недостаточное покрытие экспериментального пространства означает, что

при практическом использовании градуировочной модели возможно возникновение нового образца, обладающего необычным с точки зрения модели сочетанием факторов. Необычные образцы могут и должны выявляться на этапе прогноза через оценку расстояния до центра модели и других статистических критериев (раздел 1.3.5), однако их возникновение, например в процессе работы онлайн-анализатора, потребует принятия по каждому случаю отдельного решения, что нежелательно. Недостаточное покрытие ухудшает общий баланс схемы ПЭ, что можно показать, рассчитав h -функцию (раздел 3.1.3). Особенно негативно на качестве прогноза может сказаться незаполненность приграничных областей, примыкающих к вершинам экспериментального пространства, при наличии *взаимодействия* между факторами, например комплексообразования между соответствующими им компонентами.

Едва ли можно выработать универсальный критерий для минимального или достаточного покрытия экспериментального пространства образцами. Интуитивно понятно, что первый должен существенно превышать 50%, а второй – приближаться к 100%. Предложенное численное выражение покрытия удобно использовать как один из критериев сравнения различных схем ПЭ.

3.2.4 Встроенная проверка

Хорошо спланированный градуировочный эксперимент должен включать предопределённый схемой (встроенный) проверочный набор образцов или предусматривать способ его создания в дальнейшем.

Встроенный проверочный набор должен быть представительным по отношению к обучающему и, следовательно, быть в целом подобным ему. В то же время он не должен иметь слишком близких соседей-образцов из обучающего набора (худший случай – совпадение). Кроме того, образцы для проверки не должны занимать позиций на границе многоугольника (многогранника) области покрытия и, тем более, не могут быть его вершинами. Полезно иметь предусмотренную схемой ПЭ возможность построить регрессионную модель как для полного набора образцов, так и только для обучающего набора, исключив проверочные. Исключение проверочных образцов при этом не должно приводить

к дисбалансу и к значимому ухудшению модели.

Несмотря на очевидную важность проблемы, в существующей на сегодня научной литературе не найдено методов ПЭ, сочетающих обучающий и проверочный наборы градуировочных образцов в одной схеме. Этот недостаток присущ даже классическому исчерпывающему полнофакторному дизайну, который часто считают «абсолютным» решением, поскольку он заполняет образцами все уровни всех факторов.

3.2.5 Интерпретируемость

Схема расположения образцов в экспериментальном пространстве отражается в факторном пространстве многомерной градуировочной модели и может быть впоследствии проанализирована, например на графике счетов ПЛС (раздел 1.3.4). Искажения, которые претерпевает схема ПЭ при переносе в пространство ЛП несут информацию о значимости факторов в модели, о наличии межфакторных корреляций и взаимодействий, нелинейностей и выбросов, а также о физическом смысле самих ЛП [65].

3.2.6 Дополнительные характеристики

Для анализа объектов, имеющих некий ожидаемый (например природный, стандартный или целевой) химический состав, хорошо подходят схемы градуировочных ПЭ, построенные вокруг центральной точки. Примерами таких задач являются определение питательных веществ в стандартизированном молоке или проверка состава лекарственного препарата. Образец ожидаемого состава ставится при этом в центр экспериментального плана, а остальные обеспечивают заполнение концентрационных диапазонов анализируемых веществ. В «центробежных» схемах при необходимости легко осуществить достраивание существующей модели новыми образцами для расширения границ экспериментального пространства.

Другим полезным свойством градуировочной схемы, независимо от наличия центральной точки, является возможность её пошагового насыщения новыми образцами, в том числе в существующих границах. Не все дизайны позволяют добавлять образцы в уже существующую схему без нарушения базового принципа

построения. В качестве примера можно привести равномерный дизайн [154] и другие, основанные на ЛГО.

Некоторые методы ПЭ (например, случайный разброс образцов) не обеспечивают уникальной схемы и, следовательно, не могут быть однозначно описаны комбинацией значений N , l , k . Для воспроизведения эксперимента в этом случае необходимо сохранять полную матрицу S .

Немаловажным качеством дизайна является простота построения и использования схем, необходимость привлечения компьютера и специального ПО.

Вышеописанные дополнительные свойства градуировочных ПЭ не являются обязательными требованиями, но они влияют на практичность и распространённость метода. Следует отметить, что встроенная проверка и интерпретируемость – характеристики присущие только дизайну градуировочного эксперимента и не рассматриваемые классической теорией ПЭ [142].

3.3 Диагональный дизайн многокомпонентного градуировочного эксперимента

В работе представлен новый подход к планированию многокомпонентного градуировочного эксперимента, согласно которому образцы систематически размещаются вдоль диагоналей квадрата, куба или гиперкуба экспериментального пространства, изначально минимизируя межфакторные корреляции и обеспечивая равномерность заполнения интервалов факторов [143]. Предложенный диагональный дизайн (ДД) относится к семейству ЛГО, где число образцов и уровней равны ($N = l$) при любом числе факторов k , но он может быть расширен и до более заполненных схем. Предлагаемый ДД алгоритм заполнения экспериментального пространства предусматривает встроенный проверочный набор и может быть реализован без применения компьютера.

3.3.1 Основные допущения и правила конструирования

Анализ смесей инструментальными методами, такими как ОМС, требует построения градуировочной модели. Выбор метода моделирования зависит от функции аналитического сигнала, его зависимости от концентрации аналита.

Линейность и аддитивность по уравнению (1.2) являются распространёнными допущениями спектроскопии, позволяющими применять методы линейного регрессионного анализа данных. Этим допущениям отвечает большинство оптических методов анализа химических систем. Отклонения от линейности могут быть обусловлены самим измерением (нелинейный физический эффект), природой образца (например, рассеивающая среда); или физическим или химическим взаимодействием между компонентами (таким как комплексообразование). Линейная зависимость аналитического сигнала от концентрации является основным допущением диагонального дизайна. Это справедливо как для традиционной спектроскопии, так и для ОМС. Отклонение от линейности не обязательно запрещает применение ДД на практике. Устойчивость схемы к квадратичным членам и взаимодействиям переменных в регрессионном уравнении может быть индивидуально протестировано для любой схемы с использованием функции размаха, как описано в разделе 3.1.3.

Предлагаемый метод ПЭ предполагает, что контролируемые факторы, то есть значения концентраций соответствующих компонентов, можно варьировать в установленных пределах. ДД многофакторного многоуровневого градуировочного эксперимента следует трём простым правилам:

- а) экспериментальное пространство является гиперкубом, то есть интервалы всех факторов разбиваются на равное число уровней;
- б) образцы равномерно распределяются по диагоналям гиперкуба; и
- в) каждый уровень каждого фактора представлен одним и только одним образцом (условие латинского гиперкуба).

Для встроенного тестового набора действуют два дополнительных правила:

- г) проверочные образцы не должны занимать смежные уровни любого фактора экспериментального пространства; и
- д) для проверки не используются крайние образцы на диагоналях.

Некоторые из этих правил построения, сформулированные для базового диагонального ПЭ, могут быть опущены в расширенных дизайнах (раздел 3.3.4).

3.3.2 Диагональный дизайн градуировки для двух компонентов

В настоящем разделе рассмотрен наиболее важный с практической точки зрения случай одновременной градуировки двух компонентов смеси. Детальное рассмотрение принципов построения двухфакторных диагональных схем облегчит дальнейшее обобщение ДД на экспериментальные пространства более высокой мерности.

В двухфакторном планировании возможны схемы с центральной точкой (рис. 3-2) или без неё (рис. 3-3).

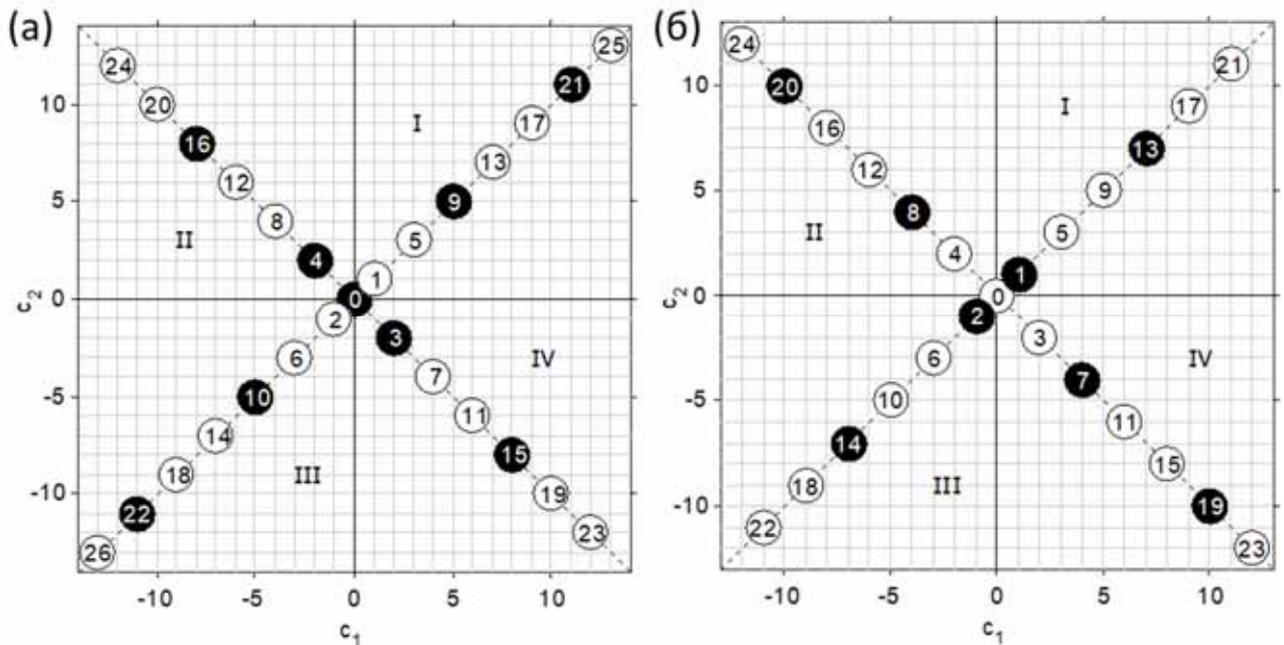


Рисунок 3-2. Двухфакторные диагональные схемы с центральной точкой для: (а) 27-ми экспериментов с проверочным набором (черные круги), начинающимся с нулевого образца ($d2_{0v0}^{27}$)⁵ и (б) 25-ю экспериментов с проверочным набором, начинающимся с образца №1 ($d2_{0v1}^{25}$); римские цифры обозначают квадранты системы координат [143].

Центральный образец, если он существует, всегда помещается в начало координат и получает №0. Идентификационный номер последующих добавляемых образцов инкрементно возрастает. Образцы добавляются от центра к периферии, следуя «правилу креста». Первый образец помещается в ближайшую диагональную позицию первого квадранта. Последующие нечётные элементы

⁵ Введено кодовое обозначение дизайна: dk_{DSVS}^N , где d – обозначение диагонального дизайна, k – число факторов, N – число образцов; DS и VS – номера первого образца обучающего и проверочного (если есть) наборов.

чередуются между *полудиagonалями* в квадрантах I и IV, всегда занимая следующую свободную позицию. После добавления нечётного образца его чётный «двойник» одновременно помещается в симметричное положение на соответствующей смежной полудиagonали в квадранте II или III. Преимуществом предложенной диагональной схемы является практичность – двухкомпонентный эксперимент можно без труда распланировать на клетчатой бумаге.

Центральный образец в ДД не так важен, как в классическом ПЭ, и его можно опустить. В этом случае уровни по осям факторов становятся кратными 0.5 (рис. 3-3). Отсутствие центрального образца не меняет алгоритма заполнения и схемы в целом, но в соответствии с правилом (г) из раздела 3.3.1 его отсутствие не позволяет начать проверочный набор с №1.

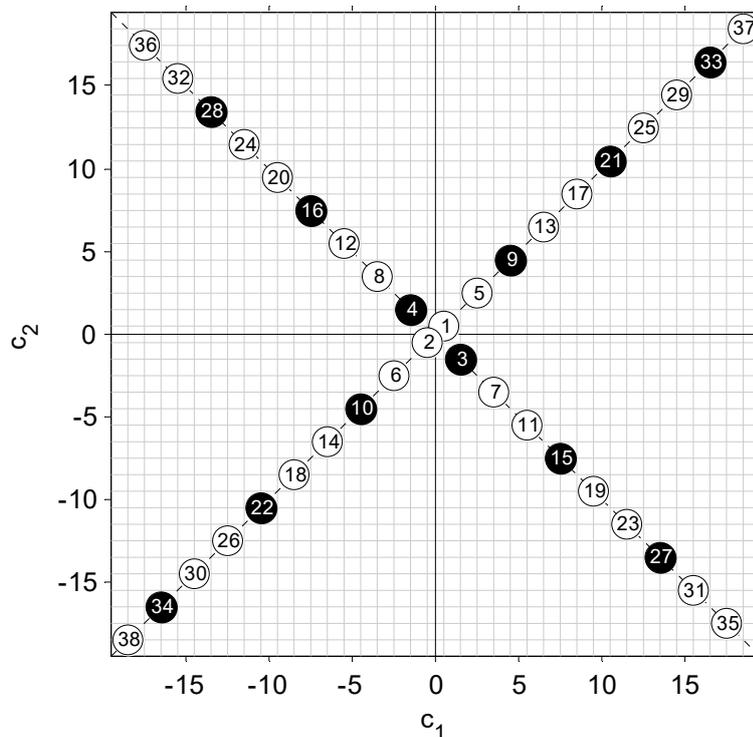


Рисунок 3-3. Двухфакторная диагональная схема без центральной точки с 38-ю образцами; проверочное подмножество (черные круги) начинается с №3 ($d2_{1v3}^{38}$) [143].

Используемые здесь условные нормированные координаты отражают симметрию конструкции. Они более удобны для создания схемы, поскольку они универсальны. При пересчёте в концентрации минимальные и максимальные значения концентрации каждого компонента смеси приписываются границам диапазона соответствующего фактора, а промежуточные рассчитываются

посредством интерполяции.

Хотя число образцов в схемах ДД произвольно, есть несколько «магических» размеров, дающих наиболее оптимальные конструкции. Избранные диагональные схемы и их наиболее распространённые альтернативы из классического ПЭ представлены в табл. 3-1.

Таблица 3-1. Избранные диагональные и классические схемы дизайна для двух факторов.

ПЭ	l	N_c	N_v	r_{cal}	h_{min}	h_{max}	Характеристика
<i>Диагональные схемы без проверочного набора</i>							
$d2_1^9$	9	9	0	0.107	0.14	0.80	минимальный без центральной точки
$d2_0^{10}$	10	10	0	-0.002	0.13	0.66	минимальный с центральной точкой
$d2_0^{25}$	25	25	0	0.004	0.04	0.25	сбалансированный (рис. 3-4 а)
<i>Диагональные схемы с проверочным набором</i>							
$d2_{0v1}^{13}$	13	9	4	-0.081	0.11	0.64	минимальный (проверочный набор 1)
$d2_{0v0}^{15}$	15	10	5	0.063	0.10	0.57	минимальный (проверочный набор 2)
$d2_{0v1}^{25}$	25	17	8	-0.025	0.06	0.36	экономичный (проверочный набор 1)
$d2_{0v0}^{27}$	27	18	9	0.021	0.06	0.34	экономичный (проверочный набор 2)
$d2_{0v1}^{37}$	37	25	12	-0.012	0.04	0.26	полноценный (проверочный набор 1)
$d2_{0v0}^{39}$	39	26	13	0.011	0.04	0.24	полноценный (проверочный набор 2)
$d2_{0v1}^{49}$	49	33	16	-0.007	0.03	0.20	исчерпывающий (проверочный набор 1)
$d2_{1v3}^{50}$	50	34	16	0.007	0.03	0.19	исчерпывающий без центральной точки
<i>Классические схемы (без проверочного набора)</i>							
$ff2^{25}$	5	25	0	0	0.04	0.20	полнофакторный (рис. 3-4 б)
$ccc2^9$	5	9	0	0	0.11	0.61	центрально-композитный
$u2^{10}$	10	10	0	0.006	0.10	0.59	малый равномерный
$u2^{25}$	25	25	0	0.005	0.04	0.26	равномерный (рис. 3-4 в)
$r2^{25}$	-	25	0	0.449	0.04	0.49	равномерно-случайный
$df2^{25}$	13	25	0	0	0	0.04	диагональный факторный (рис. 3-4 г)

Примечания: l – число уровней; N_c , N_v – число образцов в обучающем и проверочном наборах соответственно; r_{cal} – коэффициент линейной корреляции между факторами; h_{min} , h_{max} – минимум и максимум функции размаха соответственно; D – дизайн: d – диагональный, ccc – центрально-композитный [142, с. 150], u – равномерный по Фангу [154], r – равномерно-случайный (значения r_{cal} , h_{min} , h_{max} даны для расчётного примера), df – диагонально-факторный и ff – полнофакторный.

Хотя минимальная схема ДД имеет несколько более высокую корреляцию и максимальное значение h_{max} , чем классический центрально-композитный дизайн с 9-ю образцами (табл. 3-1), она может быть предпочтительнее из-за большего числа уровней и равномерного распределения образцов. В целом же преимуществами

обладают схемы со встроенным проверочным набором. Табл. 3-1 предлагает выбор сбалансированных диагональных схем различных размеров с практическими характеристиками: минимальный, экономичный, полноценный или исчерпывающий.

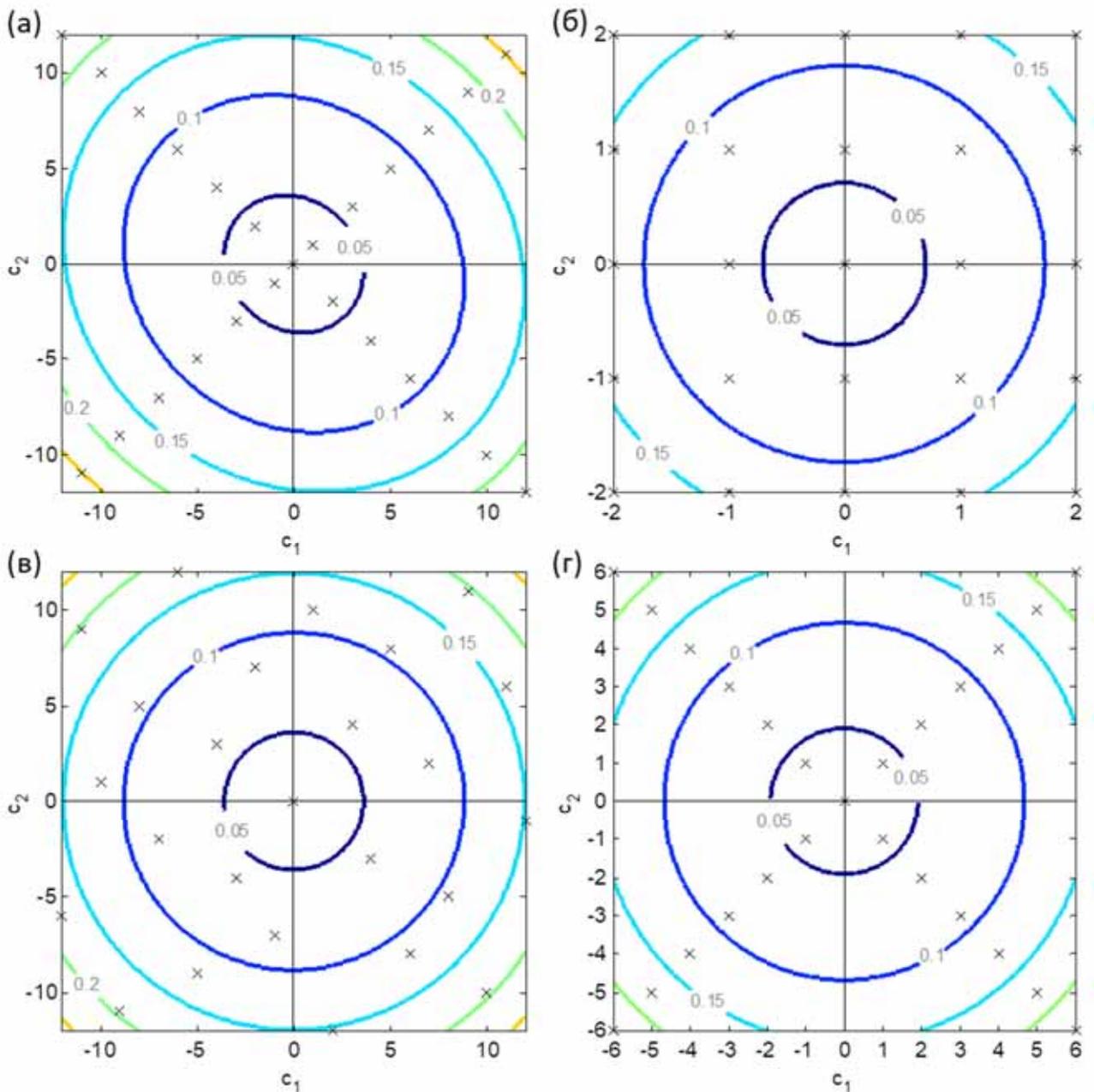


Рисунок 3-4. Контур функции размаха для различных двухфакторных схем с 25-ю образцами: (а) диагональных дизайн, (б) полнофакторный дизайн на пяти уровнях, (в) равномерный дизайн по Фангу [156] и (г) диагонально-факторный дизайн на 13-ти уровнях; образцы обозначены крестами [143].

Хорошо сбалансированный проверочный набор в ДД должен быть максимально симметричным и, таким образом, иметь одинаковое число (не менее одного) образцов на каждой полудиагонали. Чтобы выполнить требование (д)

раздела 3.3.1, за последним образцом проверочного набора должен следовать хотя бы один образец обучающего набора. Это требование делает благоприятными для проверки схемы с определённым количеством образцов, получаемых с периодом 12, а именно с 13-ю, 25-ю, 37-ю и т. д. образцами. Самая компактная ($d2_{0v1}^{13}$) и следующая по размеру ($d2_{0v1}^{25}$) схемы ДД со встроенной проверкой из табл. 3-1 представлены на рис. 3-2 б (для получения первой из второй нужно убрать образцы с номерами выше 12-ти).

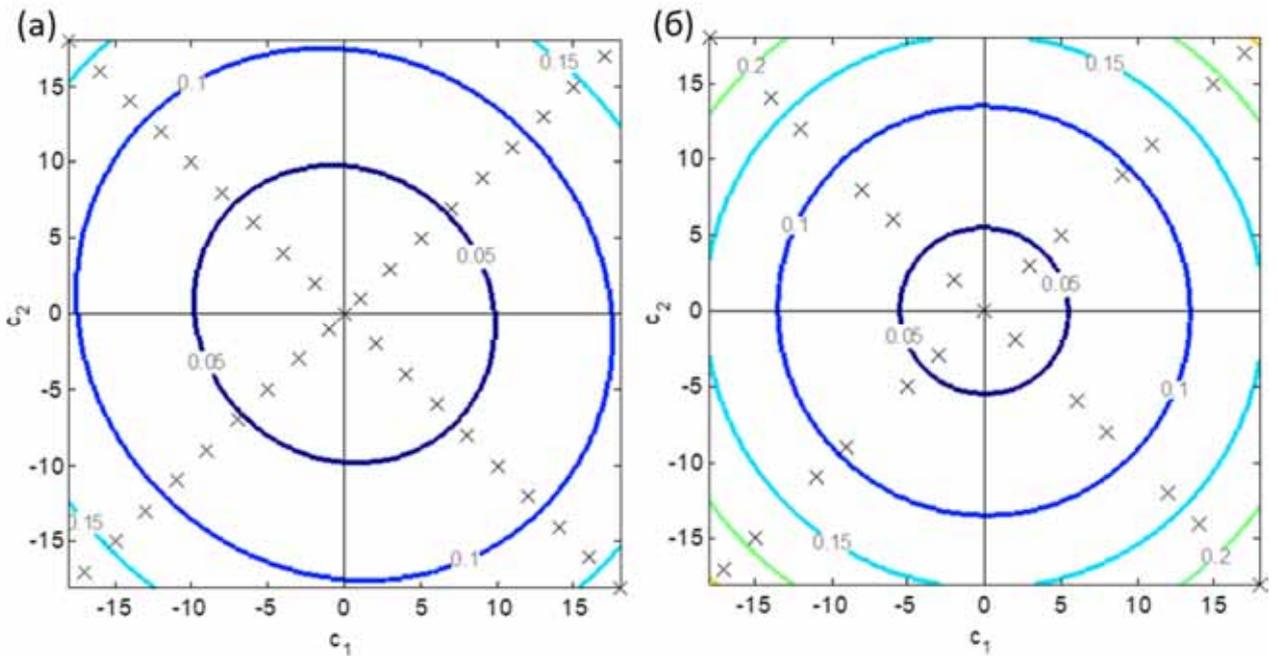


Рисунок 3-5. Контур функции размаха для (а) двухфакторного ДД с 37-ю образцами ($d2_{0v1}^{37}$) и (б) обучающего поднабора; образцы обозначены крестами [143].

Удаление проверочных образцов из полного набора не приводит к «разбалансировке» оставшихся обучающих образцов. Интересно, что в схемах ДД со встроенной проверкой коэффициент корреляции факторов обучающего набора r_{cal} с определённой периодичностью становится ничтожно мал и даже принимает значения меньше (по абсолютной величине) коэффициента корреляции по полному набору. Иными словами, устранение проверочных образцов из модели может иметь положительный эффект. Это иллюстрируют контурные графики на рис. 3-5, где представлена функция размаха для схемы, включающей все 37 образцов или только 25 образцов соответствующего обучающего набора ($d2_{0v1}^{37}$, табл. 3-1). Функции размаха обоих наборов близки по значениям, но функция на рис. 3-5 б показывает заметно более высокую симметрию, чем на рис. 3-5 а.

Хотя с размером градуировочного набора значения h и межфакторной корреляции имеют тенденцию к снижению, на практике редко используют более 50-ти образцов. Исчерпывающие схемы могут быть рекомендованы как абсолютное решение для регрессионного анализа двухкомпонентных смесей лишь в особых случаях: если цена образца относительно низка или точность градуировки имеет решающее значение. Если же независимая проверка не требуется, например при построении промежуточных моделей в процессе разработки ОМС (раздел 2.4), достаточно экономичной схемы с 25-ю образцами (рис. 3-2 б и 3-4 а). По сравнению с равномерным заполнением ЛГО [154,155], имеющим столько же уровней (рис. 3-4 в), ДД может быть предпочтён из-за его простоты, интерпретируемости, пополняемости и «центробежной» организации (раздел 3.2). В целом же оптимальная схема ДД должна индивидуально подбираться для каждого практического случая с учётом различных требований, таких как допуск по корреляции компонентов, ожидаемая точность прогноза, необходимость независимой проверки и наличия центральной точки.

Встроенный проверочный набор является важным новшеством ДД. Проблема организации проверочных образцов в разреженной градуировочной схеме ранее не изучалась систематически. В построении проверочного набора ДД следует общепринятой стратегии валидации градуировочной модели, согласно которой обучающие и проверочные подмножества одинаково представительны по отношению к предполагаемым новым образцам и отвечают одним тем же предположениям о природе данных. Поэтому диагональное расположение проверочных образцов между обучающими является оптимальным, обеспечивая слаженность всей схемы. Встроенная проверка хорошо подходит для определения сложности модели, то есть количества ЛП в регрессии методом, например ПЛС. Поскольку тестовые образцы являются независимыми (в том смысле, что их удаление не приводит к нарушению баланса схемы, как видно из рис. 3-5), такой проверочный набор, как ожидается, позволит достаточно надёжно оценить точность градуировки. При этом по-прежнему предполагается линейная зависимость отклика от концентрации, как в спектральных данных, следующих

уравнению (1.2) Бугера-Лабмерта-Бера.

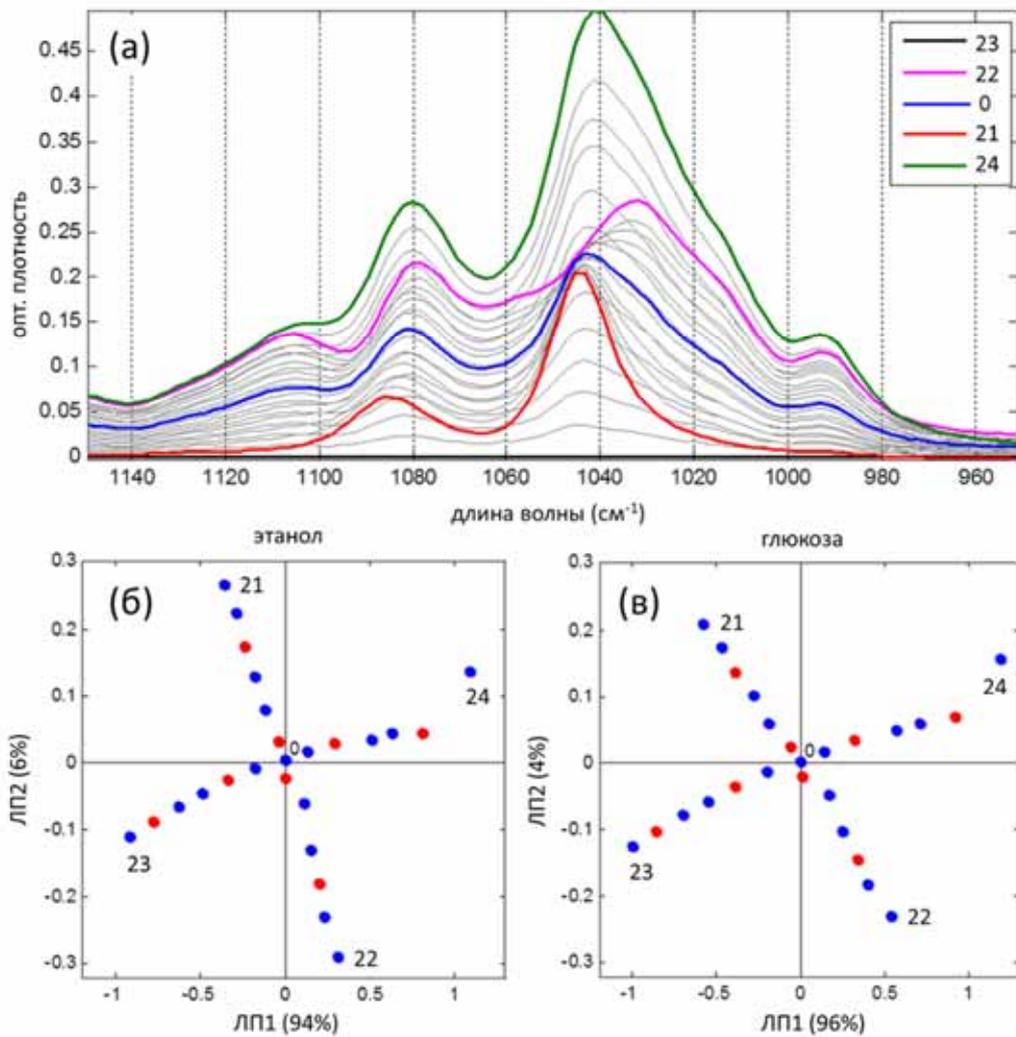


Рисунок 3-6. Применение диагонального дизайна $d2_{0v1}^{25}$ (рис. 3-2 б, где c_1 – убывающая концентрация глюкозы, а c_2 – возрастающая концентрация этанола) для определения этанола и глюкозы в водной смеси по ИК-спектрам: (а) спектры; графики ПЛС-счетов (б) этанола и (в) глюкозы на латентных переменных ЛП1 и ЛП2 [65].

Двухкомпонентные диагональные схемы и их элементы были апробированы в нескольких работах автора и других исследователей по анализу пищевых продуктов [84], контролю окружающей среды [87], мониторингу химических процессов [198] и в биотехнологии [65]. На рис. 3-6 представлены результаты одновременной градуировки этанола и глюкозы из работы [65]. Этот пример иллюстрирует несколько несомненных преимуществ диагональной конструкции. В указанной статье было произведено сравнение разрабатываемых ОМС в области среднего ИК с Фурье-спектрометром высокого разрешения, взятым в данном случае в качестве полноспектрального метода (раздел 2.1.1). Для сравнения

моделей была взята схема ДД из 25-ти образцов с проверочным набором (рис. 3-2 б). Было показано, что наличие встроенного проверочного набора (в дополнение к ПП на полном наборе) прекрасно служит как для определения числа ЛП в ПЛС-регрессии, так и для сравнения инструментальных методов. Благодаря наличию большого числа уровней, крестообразное расположение образцов легко прослеживается на графике ПЛС-счетов (рис. 3-6 б и в), облегчая интерпретацию. В частности, легко увидеть наличие выброса (образец №24), а также заподозрить корреляцию сигналов компонентов (вращение схемы относительно факторного пространства) и их возможную нелинейность (более подробная интерпретация дана в работе [143]).

3.3.3 Обобщение ДД для градуировки трёх и более компонентов

Задаваемое квадратом экспериментальное пространство двухкомпонентной градуировки особенно удобно, так как две перпендикулярные диагонали автоматически обеспечивают некоррелированность факторов. С ростом числа факторов k число диагоналей экспериментального гиперкуба, определяемое по формуле 2^{k-1} , быстро растёт, а их ортогональность в общем случае не соблюдается. Поэтому обобщение ДД для трёх и более компонентов связано с некоторыми компромиссами. Тем не менее, ЛГО с использованием только диагональных элементов может быть реализован для любого числа факторов, а алгоритм должен в первую очередь обеспечивать равномерность заполнения.

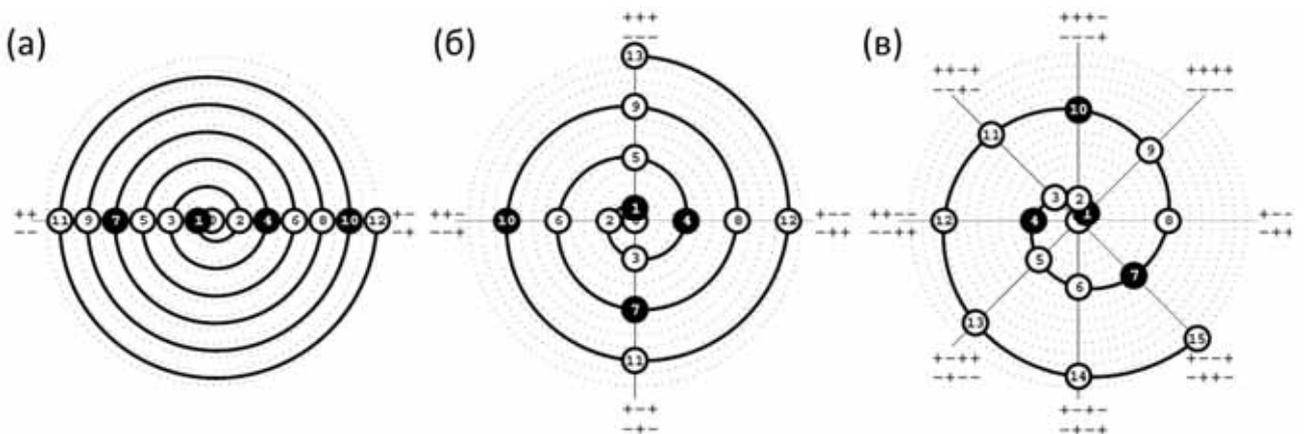


Рисунок 3-7. Схема диагонального заполнения с четырьмя парами проверочных образцов (черные круги) в случае: (а) двух, (б) трёх и (в) 4-х факторов [143].

Вышеописанный алгоритм поочерёдного центрбежного заполнения

диагоналей может быть обобщён для любой мерности k экспериментального пространства. Его наглядной геометрической иллюстрацией является Архимедова спираль в полярных координатах (рис. 3-7), при помощи которой можно без компьютера спланировать градуировочный эксперимент любой сложности, независимо от числа факторов, уровней и образцов.

Алгоритм заполнения оперирует с полудиagonалями k -мерного гиперкуба, однозначно связанными с его вершинами. Декартовы координаты любой вершины единичного гиперкуба, определяемого положительными базисными векторами единичной длины, образуют двоичное число, десятичное представление которого будет в дальнейшем использовано как уникальный идентификатор вершины (от 0 до 2^k-1) и соответствующих полудиagonалей: sd_0 , sd_1 , sd_2 и т. д. Например, вершины единичного куба, заданные координатами $(0,1,0)$ и $(1,1,1)$, соответствуют числам 2 (двоичное «010») и 7 («111»), поэтому прилегающие полудиagonали будут соответственно обозначены как sd_2 и sd_7 .

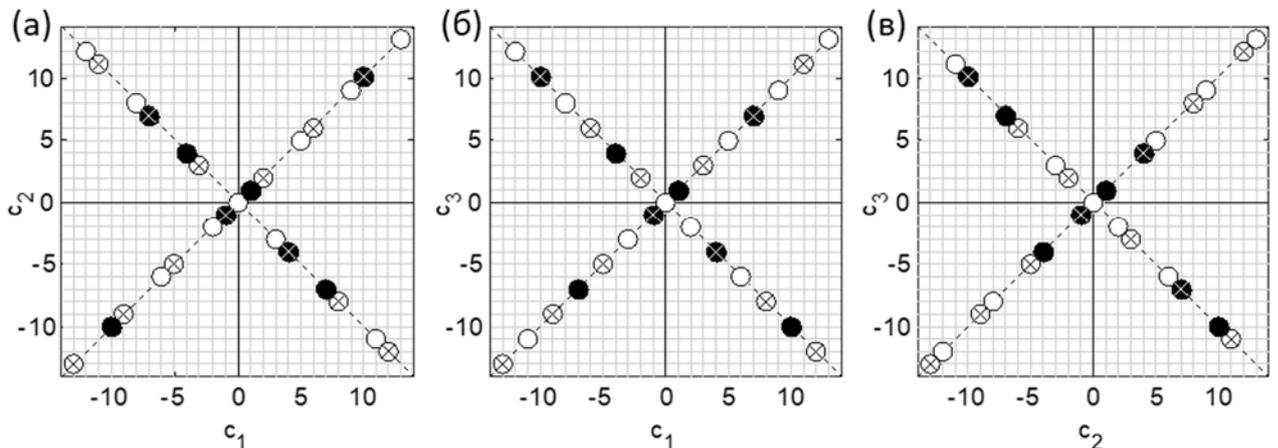


Рисунок 3-8. Диагональная схема для трёх факторов с 27-ю образцами, с центральным образцом и проверочным подмножеством (залитые чёрные круги), начинающимся с образца №1 ($d3_{0v1}^{27}$). В каждой двухфакторной проекции образцы, имеющие отрицательные значения третьего (скрытого) фактора, отмечены крестиком [143].

Для упрощения системы обозначений удобно перенести начало координат в центр гиперкубического экспериментального пространства. Каждая полудиagonаль получает вектор направления от центра к вершине, выраженный в единичных координатах с разными знаками; например, sd_2 выражается как $(-1,1,-1)$, sd_7 – как $(1,1,1)$ или просто «-+-» и «+++» соответственно. По сути, перенос начала координат заменяет нули в двоичных координатах вершин единичного гиперкуба

на -1. Важно отметить, что две половины одной и той же диагонали, например sd2 ('-+-') и sd5 ('++-'), отличаются только знаком вектора направления.

Лежащие на одной и той же диагонали образцы центрально симметричны относительно начала координат. Поэтому их можно добавлять попарно, так что добавление элемента к любой полудиagonали однозначно определяет положение его «двойника». Следовательно, чтобы получить полную схему достаточно заполнить набор полудиagonалей, где каждая из 2^{k-1} диагоналей представлена только одной из двух её половин. Самый простой подход, представленный на рис. 3-7, состоит в том, чтобы взять последовательность идентификаторов полудиagonалей от 0 до $2^{k-1}-1$ в качестве стандартного порядка заполнения. Например, для трёх факторов (рис. 3-7 б) это будет последовательность sd0, sd1, sd2, sd3. Её можно упрощённо записать как sd{0 1 2 3} или sd{0:3}. Такая система обозначений будет использована в дальнейшем изложении.

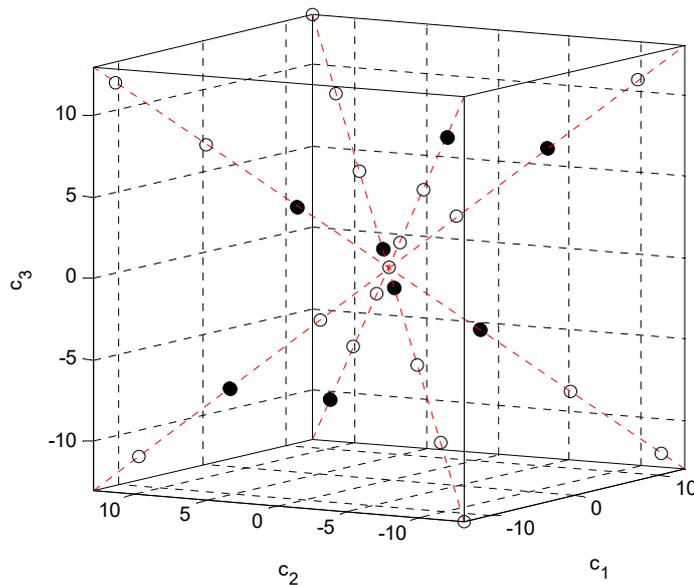


Рисунок 3-9. Трёхмерное представление схемы $d3_{0v1}^{27}$, показанной на рис. 3-8 [143].

Рассмотрим трёхмерную схему (рис. 3-7 б) в качестве иллюстрации алгоритма заполнения. Оси полярного пространства на рис. 3-7 соответствуют диагоналям экспериментального гиперкуба. Образец №0, если он существует (отсутствие центрального образца не влияет на дальнейшее заполнение), ставится в начало координат экспериментального пространства. Каждая последующая пронумерованная позиция вдоль разворачивающейся спирали определяет пару

экспериментов, получаемых путём умножения порядкового номера позиции (в кругах на рис. 3-7) на два противоположных вектора направлений полудиagonалей заполняемой в настоящий момент диагонали. Кодовое обозначение соответствующих векторов направления в виде чередующихся знаков «+» и «-» дано для каждой полярной оси, являющейся образом соответствующей диагонали. Например, переход к позиции 1 добавляет точки (1,1,1) и (-1,-1,-1). Позиция 2 генерирует (-2,2,-2) и (2,-2,2) и т. д. Итоговая схема ДД для 27-ми образцов (последний №26) представлена на рис. 3-8 при помощи проекционных плоскостей с попарными факторами, а трёхмерная версия – на рис. 3-9. На рис. 3-10 показана более насыщенная трёхфакторная схема ДД с 51-м образцом.

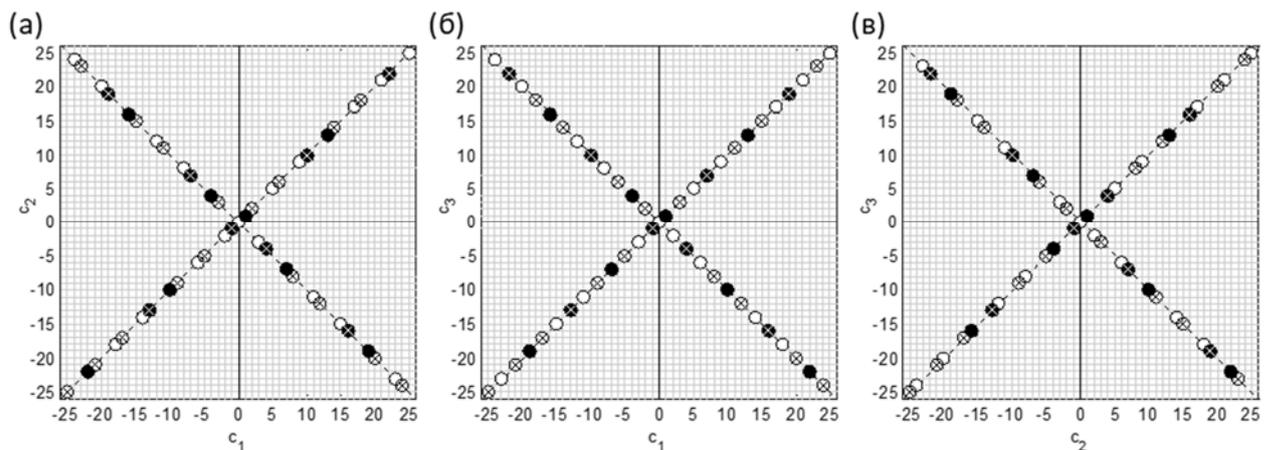


Рисунок 3-10. Диагональная схема для трёх факторов с 51-м образцом ($d3_{0v1}^{51}$) [143].

Проекционная плоскость на рис. 3-8 б напоминает двухфакторную схему, показанную на рис. 3-2 б. На двух других схемах проявляется более сложная периодичность, но принцип чередования продолжает осуществляться, то есть ни один из образцов не имеет соседа на ближайших уровнях той же диагонали. В трёхфакторной схеме последовательность чередования диагоналей при заполнении может отличаться от стандартной $sd\{0:3\}$, как на рис. 3-7 б. Например, она может быть задана комбинациями $sd\{2\ 3\ 0\ 1\}$ или $sd\{0\ 3\ 2\ 1\}$ и т. д. Но независимо от выбранной последовательности, проекции схемы ДД на рис. 3-8 и 3-10 остаются неизменными, хотя и претерпевают при этом симметричные преобразования.

Для выбора подмножества проверочных образцов также существует простое эмпирическое правило с использованием спиральной схемы на рис. 3-7, которое действует для любого числа факторов. Первая пара проверочных образцов

определяется выбранной схемой дизайна, например, на рис. 3-7 это позиция 1 вдоль спирали. Далее каждая третья позиция вдоль развивающейся спирали формирует пару проверочных образцов, удовлетворяющих требованию (г) из раздела 3.3.1.

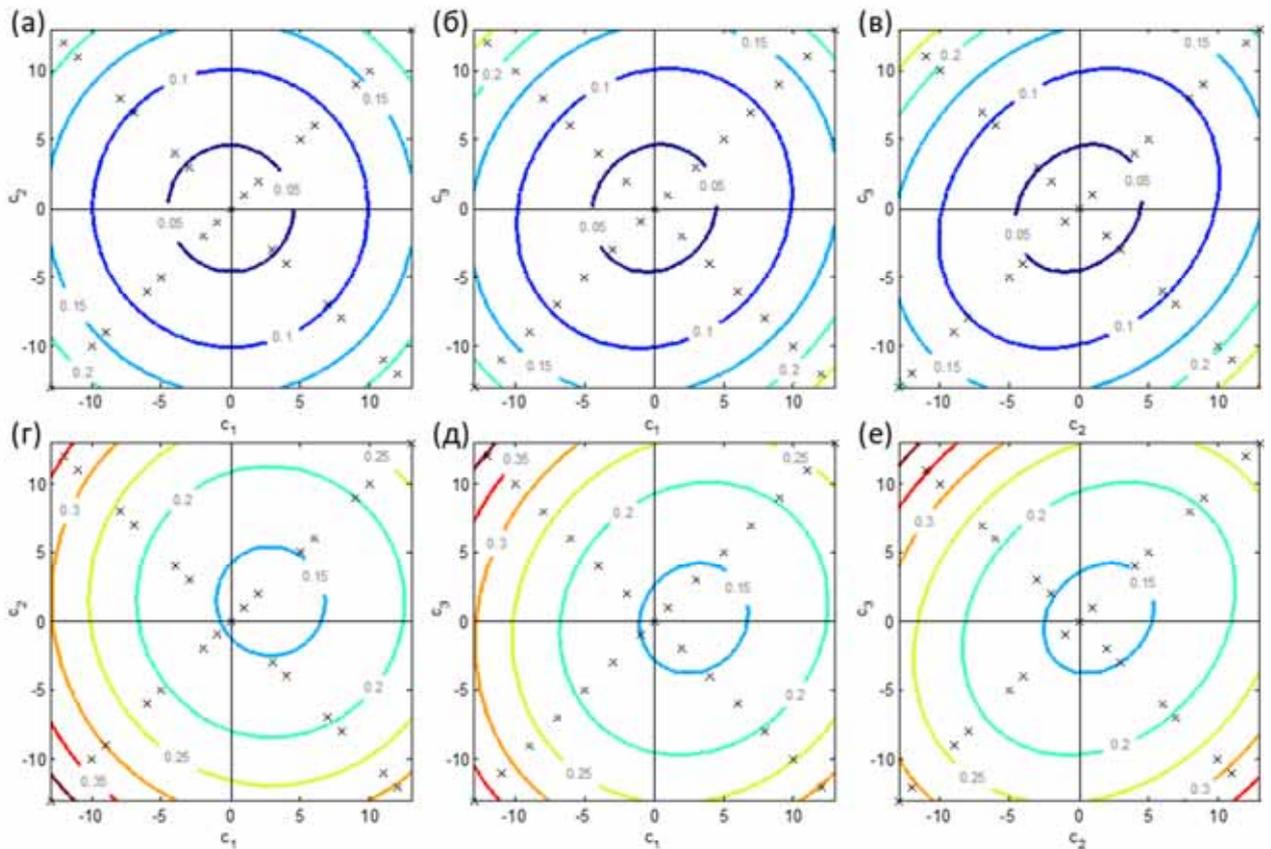


Рисунок 3-11. Контурные графики проекций функции размаха для трёхфакторного ДД на рис. 3-8 ($d3_{0v1}^{27}$) на плоскости: (а) $c_3 = 0$, (б) $c_2 = 0$, (в) $c_1 = 0$, (г) $c_3 = 13$, (д) $c_2 = 13$, (е) $c_1 = 13$ [143].

Правило (д), исключаяющее проверочные образцы на границе экспериментального пространства, требует, чтобы последний полный виток спирали на рис. 3-7 не содержал черных кругов. Однако, как следует из уравнений (3.3) и (3.4), в случае линейного аналитического отклика в функцию размаха входят, помимо линейных, лишь квадратичные и попарные перекрёстные члены концентраций компонентов. Поэтому достаточно отсутствия проверочных образцов в крайних диагональных позициях соответствующих проекционных плоскостей, даже если некоторые из них занимают крайние позиции на диагоналях гиперкуба. Исходя из этого послабления, достаточно включить в запретную для проверки зону только $2k-1+2$ из последних добавляемых образцов (половина от числа вершин гиперкуба плюс два образца – одна спиральная позиция). Таким

образом, в трёхфакторном ДД на рис. 3-8 и 3-9 «запретная зона» формируется шестью последними элементами (три спиральные позиции вместо 4-х). Размер этой зоны растёт с размерностью экспериментального пространства и включает уже 10 образцов для 4-х факторов, 18 – для пяти и т. д.

Функции размаха мерностью три и выше также могут быть проанализированы с использованием плоских срезов. Наиболее информативными являются плоскости, пересекающие центр и грани (гипер)куба. На рис. 3-11 такие срезы используются для представления трёхмерной функции размаха (уравнение (3.3)), соответствующей схеме на рис. 3-8.

На следующем шаге обобщения для 4-х компонентов алгоритм ДД осложняется потерей инвариантности заполнения, то есть различные последовательности перебора диагоналей в алгоритме заполнения приводят к различным результатам, порождая целое семейство схем, удовлетворяющих требованиям (а)–(д) раздела 3.3.1. В 4-хфакторном ДД последовательность $sd\{0:7\}$ (рис. 3-7 в) становится не оптимальной с точки зрения равномерности заполнения. Об этом свидетельствуют группы по 4 образца подряд вдоль некоторых полудиagonalей (рис. 3-12).

Оптимальная последовательность заполнения может быть определена различными критериями. Предложенный нами в работе [143] принцип минимизации суммы скалярных произведений единичных векторов направлений полудиagonalей по ходу циклического заполнения приводит к схеме на рис. 3-13. В этой схеме хорошо сбалансированный проверочный массив достигается при 55-ти образцах, причём полудиagonalей содержат по 4 проверочных образца на каждой. Тема разработки оптимального метода заполнения для ДД мерности 4 и выше заслуживает отдельного исследования и не развивается далее в настоящей работе.

Читатель может самостоятельно разработать диагональную схему градуировки с необходимыми параметрами для любого количества компонентов, используя предоставленную в Приложении А функцию «doe4reg.m» для MATLAB, снабжённую несколькими практическими примерами.

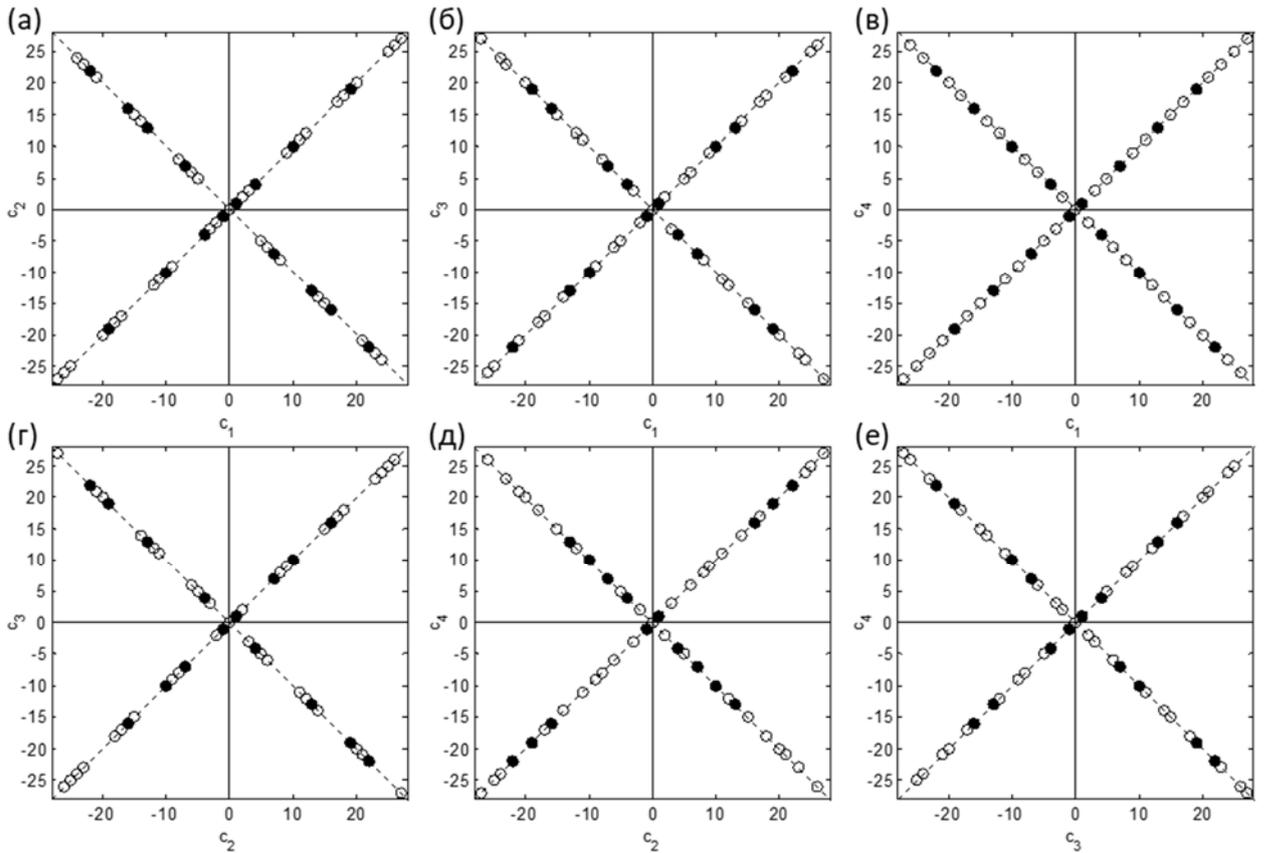


Рисунок 3-12. Схема диагонального дизайна для 4-х факторов с 55-ю образцами с использованием последовательности заполнения $sd\{0:7\}$ [143].

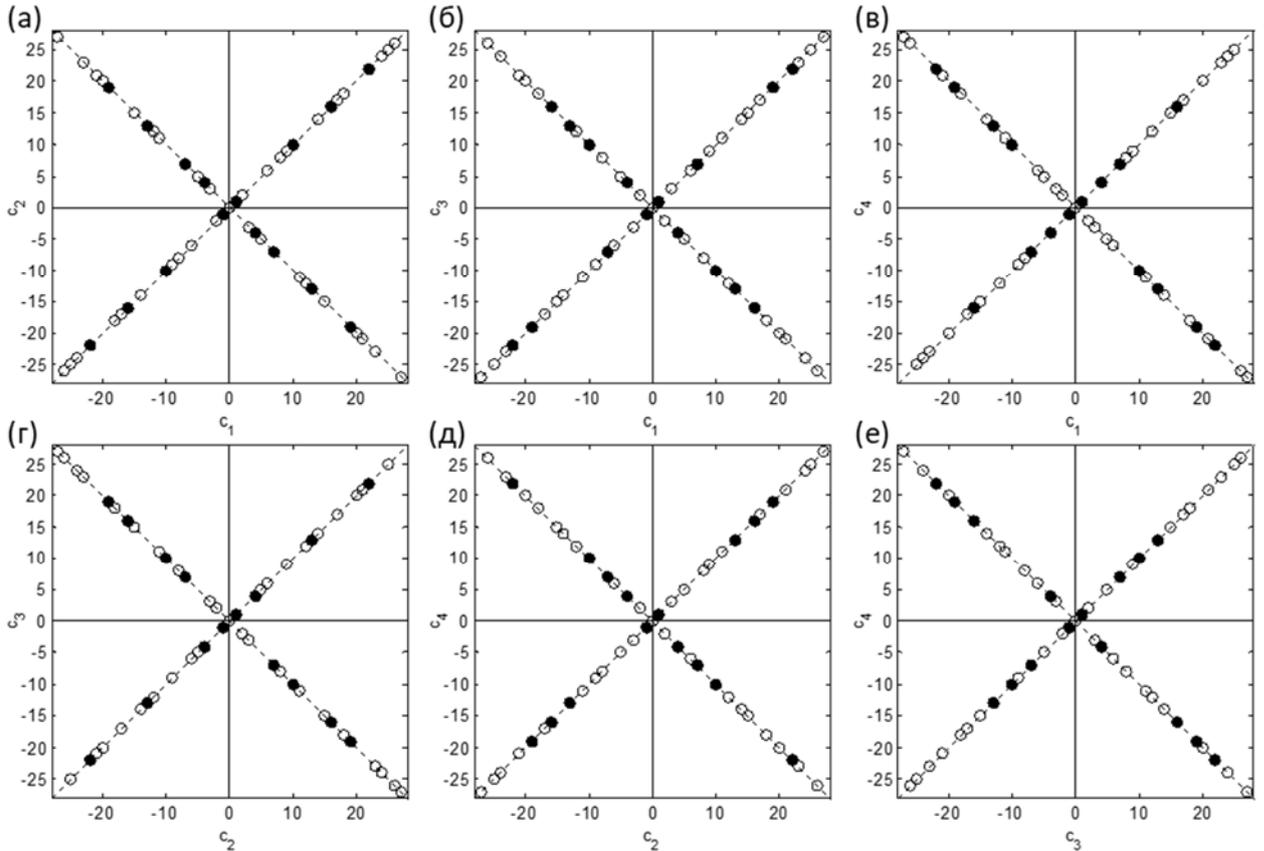


Рисунок 3-13. Диагональный дизайн для 4-х факторов с 55-ю образцами, полученный при последовательности заполнения $sd\{0\ 3\ 5\ 6\ 7\ 1\ 4\ 2\}$ [143].

3.3.4 Расширенные схемы ДД

Рассмотренные выше схемы диагонального дизайна придерживались принципа ЛГО, то есть предполагали, что каждая диагональ присутствует последовательности заполнения один и только один раз. Отклонения от этого принципа как правило не рекомендуются. Однако в схемах высокой размерности существует важный частный случай, который будет проиллюстрирован для 4-факторного ДД. Восемь диагоналей *тессеракта* (4-хмерного гиперкуба) образуют две группы взаимно ортогональных диагоналей: $sd\{0\ 3\ 5\ 6\}$ и $sd\{1\ 2\ 4\ 7\}$; любая пара диагоналей внутри одной группы ортогональна, а любая пара из разных групп – нет. Это свойство тессеракта показано в работе [143]. Использование одной и той же ортогональной последовательности дважды в той же спиральной свёртке, например $sd\{0\ 3\ 5\ 6\ 0\ 3\ 5\ 6\}$ приводит к замечательной схеме, представленной только тремя уникальными плоскими проекциями, каждая из которых повторяется дважды (рис. 3-14). Две тройки проекций (рис. 3-14 а-в или г-е) точно повторяют трёхмерный случай (рис. 3-8). Этот «сдвоенный» ДД переносит преимущества трёхфакторных схем – хорошо сбалансированную градуировку и проверку с меньшим числом образцов – в 4 измерения и добавляет ортогональность опорных диагоналей, которая отсутствует в трёхмерном случае. Таким образом, «сдвоенные» схемы могут быть рекомендованы для уменьшения числа образцов в высокоразмерных экспериментальных пространствах, начиная с 4-х факторов.

Другое важное расширение предлагаемого ДД можно получить, отходя от требования латинского гиперкуба (в) в разделе 3.3.1, позволяя тем самым иметь более, чем один образец на одном и том же уровне фактора. Заполнение пространства в этом случае можно проиллюстрировать двумя одновременно развивающимися спиралями, начинающимися с разных диагоналей. Эти индивидуально сгенерированные наборы данных затем объединяются в одну и ту же расширенную матрицу планирования. Расширенные схемы ДД такого типа (будем называть их дополненными) могут включать в себя до 2^{k-1} генерирующих спиралей, то есть каждая диагональ становится источником отдельного набора образцов. Однако представляется целесообразным использовать только две или три

спирали, равномерно распределённые вдоль круга полярных координат. Пример дополненного ДД с двумя спиралями показан на рис. 3-15.

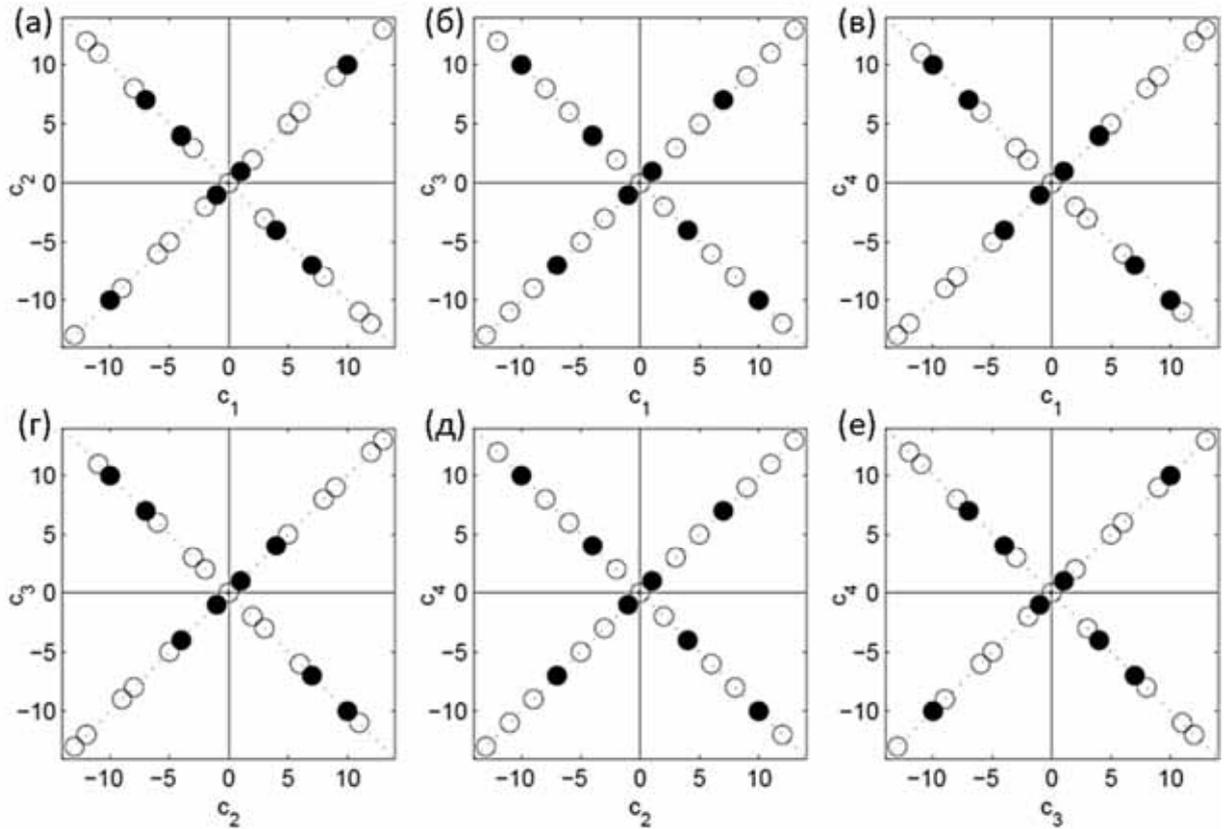


Рисунок 3-14. «Сдвоенная» диагональная конструкция с 27-ю образцами для 4-х факторов (последовательность заполнения $sd\{0\ 3\ 5\ 6\ 0\ 3\ 5\ 6\}$) [143].

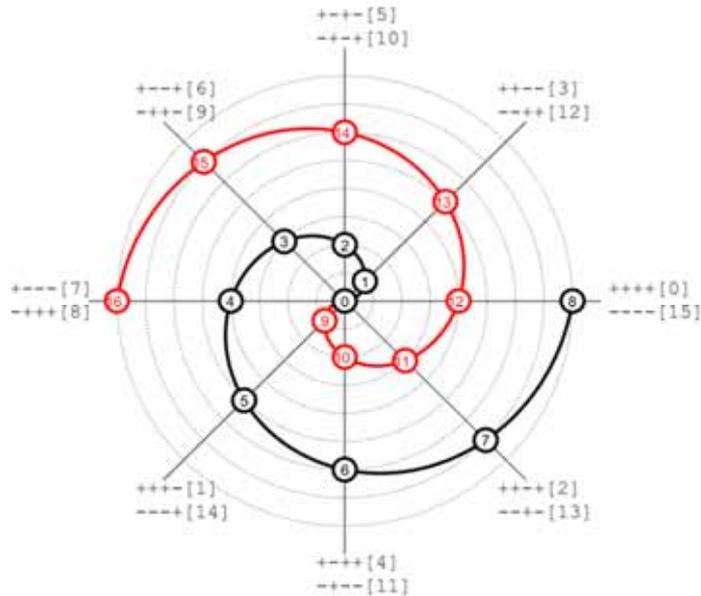


Рисунок 3-15. Схема дополненного диагонального дизайна с 33-мя образцами для 4-х факторов и оптимизированной последовательностью заполнения $sd\{0\ 3\ 5\ 6\ 7\ 1\ 4\ 2\}$. Десятичные идентификаторы полудиагоналей показаны в квадратных скобках рядом с соответствующими двоичными векторами направления [143].

В созданной таким образом схеме на рис. 3-16 пространство 4-х факторов

адекватно покрывается двукратным набором из 16 образцов плюс центральная точка. То есть эта экономичная схема насчитывает всего 33 образца. В алгоритме дополненного ДД отношение числа уровней к числу образцов уменьшается с ростом числа «наложенных» спиралей: от одного (как в ЛГО) до $1/2$ в случае двух частей, $1/3$ для трёх и т. д.

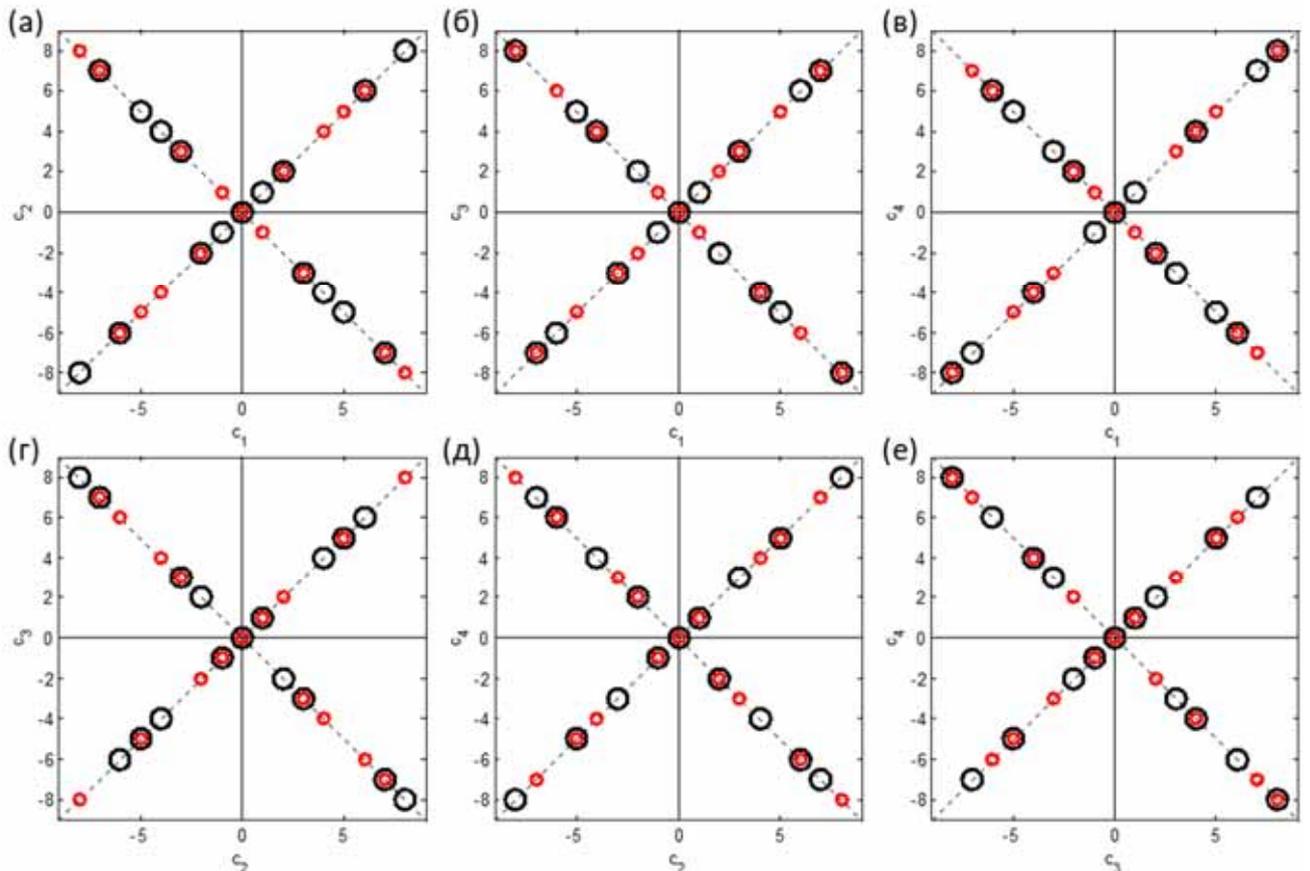


Рисунок 3-16. Дополненный диагональный дизайн с 33-мя образцами для 4-х факторов с центральным образцом. 16 образцов сгенерированы последовательностью заполнения $sd1\{0\ 3\ 5\ 6\ 7\ 1\ 4\ 2\}$ (чёрные круги) и ещё 16 – последовательностью $sd2\{7\ 1\ 4\ 2\ 0\ 3\ 5\ 6\}$ (красные круги) [143].

Общее число диагоналей в гиперкубическом экспериментальном пространстве быстро растёт с числом факторов. Необходимость их заполнения на многих уровнях приводит к стремительному росту числа требуемых образцов в многомерных схемах. Неизбежный компромисс между числом уровней и числом образцов снижает практичность многофакторных схем ДД. С этой точки зрения предложенные расширения являются жизнеспособной альтернативой строгому подходу к диагональному заполнению. Например, 4-факторная схема с 33-мя образцами на рис. 3-16 может конкурировать с классическим ДД, включающим 55

образцов (рис. 3-13), или с более компактной, но менее сбалансированной сдвоенной схемой на рис. 3-14. В пространствах высокой мерности преимущества расширенных подходов становятся ещё более отчётливы. Дополненные ДД могут при этом сочетаться с использованием сдвоенных схем.

Одна или несколько составных частей дополненной схемы могут быть использованы в качестве встроенного проверочного набора, удовлетворяющего требованию (д) конструирования (раздел 3.3.1). Эта новая стратегия проверки, может дополнительно снизить общее число градуировочных образцов в схемах высоких порядков.

Функция в Приложении А генерирует ДД любой мерности, включая расширенные схемы, и снабжена практическими примерами различных подходов.

3.4 Примеры специальных дизайнов градуировочного эксперимента для различных приложений

3.4.1 Планирование градуировочного эксперимента для определения жира и белка в молоке

При разработке методов количественного оптического анализа компонентов молока в области 400–1100 нм использовались принципы планирования многокомпонентного градуировочного эксперимента, рассмотренные в разделе 3.2 настоящей главы. В исследованиях были использованы четыре методики ПЭ (табл. 3-2), разработанные с учётом особенностей молока как аналитического объекта: сложности управляемого варьирования состава образца, ограниченной доступности референтного анализа, малой продолжительности жизни образцов (полный анализ серии должен быть проведён в течение суток) и влияния на спектры дополнительных факторов, таких как размер жировых глобул.

Образцы «Заводской» модели были приготовлены по схеме, сочетающей ДД с классической схемой центрально-композитного дизайна (рис. 3-17) с пятью повторными измерениями в центральной точке. Преимуществом дополненного таким образом ДД является возможность использования и сравнения двух подходов в одном градуировочном наборе с частичным перекрытием образцов.

Важность дополнительных образцов для интерпретации данных при разработке предложенного метода анализа молока показана в разделе 5.1.2.

Таблица 3-2. Наборы градуировочных образцов для анализа молока на содержание жира и общего белка.

Модель	Источник	Число образ.	Референт. анализ ^а	Размер частиц ^б	Ссылки
Заводская	сливки и обрат	21	да	нет	[84]
Лабораторная	стандартные образцы ^в	96	нет	да	[85,90,172]
Товарная	нормализованное молоко из торговых сетей	295	частично ^г	нет	[95,199]
Глобальная	натуральное молоко	>1000	да	да	[182]

Примечания: ^а референтный анализ содержания жира и белка на «MilkoScan™ FT120»; ^б модель устойчива к вариабельности размеров жировых глобул; ^в наборы стандартных образцов (Приложение Б) точно известного состава с преимущественным варьированием жира (набор «F») и белка (набор «P») фирмы «QSE GmbH» (г. Вольнцах, ФРГ); ^г часть образцов проанализирована на приборе «MilkoScan™ FT120», а для остальных информация о жирности взята с упаковок.

Образцы «Лабораторной» модели готовились по представленной на рис. 3-18 схеме путём смешивания стандартных образцов наборов F и P, состав которых в различных экспериментах приведён в табл. Б-1 (Приложение Б). Разница между расчётным содержанием образцов согласно схеме и их фактическим значением не превышала 0.002%.

		%Жир													
		3.02	3.10	3.18	3.26	3.34	3.42	3.50	3.58	3.66	3.74	3.82	3.90	3.98	
%Белок	3.20							<u>19</u>							
	3.15												15		
	3.10			12								<u>20</u>			
	3.05									10					
	3.00					9									
	2.95								7						
	2.90	<u>18</u>						<u>1</u> <u>2-5</u>						<u>21</u>	
	2.85						6								
	2.80									8					
	2.75				11										
	2.70			<u>17</u>								13			
	2.65		14												
	2.60							<u>16</u>							

Рисунок 3-17. Дополненная схема ДД для «Заводской» модели. Проверочные образцы обозначены жирным шрифтом. Дополнительные образцы центрально-композитного дизайна отмечены серым. Подчёркнуты номера образцов, не совпадающих с исходной схемой ДД [84].

Для увеличения числа уровней концентраций анализируемых компонентов стандарты смешивались по схеме, представленной на рис. 3-18. Смешивание производилось внутри каждого из наборов F и P по отдельности. Вершины квадрата схемы представлены 4-мя образцами соответствующего набора, а их попарные смеси расположены на сторонах и диагоналях, то есть использовались все возможные комбинации пар.

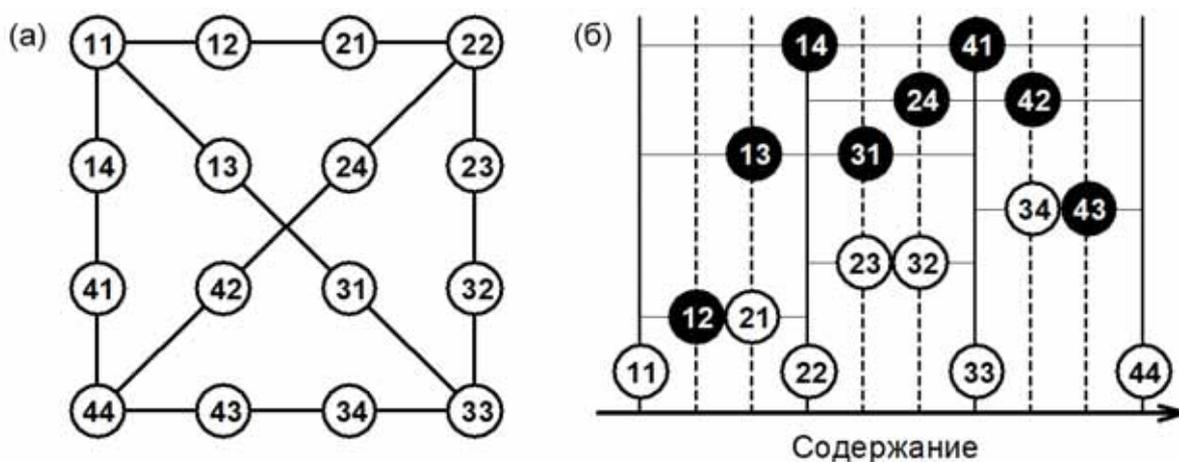


Рисунок 3-18. Дизайн градуировочной серии «Лабораторной» модели: (а) схема смешивания стандартных образцов одной серии (F или P); (б) относительное содержание жира или белка в смешанных образцах одной серии. Горизонтальные уровни обозначают шесть смешиваемых пар. В кругах указаны идентификационные номера образцов; чёрным цветом выделены проверочные образцы [85].

При смешивании каждой пары соблюдается принцип равномерности распределения, то есть равноудалённость образцов друг от друга вдоль линии между исходными стандартами, символизирующей состав образца. Схема на рис. 3-18 представлена серией из 16-ти образцов, получаемых при смешивании каждой пары в соотношении 1:2 и 2:1. Для построения градуировок по жиру и белку серии F и P используются вместе, давая 32 образца. Совмещение позволяет минимизировать корреляции между содержанием жира и белка, присущее отдельным сериям [85]. Изменяя количество смесей, изготавливаемых из каждой пары, то есть число уровней, можно регулировать размер градуировочной серии. Идентификаторы образцов на рис. 3-18 б будут использоваться для маркировки отдельных образцов при обсуждении результатов анализа в разделе 5.1 (см. также Приложение Б). Приставка «f» или «p» перед номером обозначает образцы из серии по жиру или по белку соответственно.

При анализе сырого негомогенизированного молока, в качестве дополнительного – третьего, помимо концентраций жира и белка – фактора выступает природная вариабельность размеров жировых глобул. Для получения устойчивой «Лабораторной» модели эта вариабельность была искусственно создана в градуировочных образцах при помощи пошаговой гомогенизации ультразвуком [85]. Каждый градуировочный образец анализировался три раза: в исходном состоянии и после каждой из двух последовательных 10-тисекундных гомогенизаций. В разделе 5.1.2 образцы с различной степенью (общей продолжительностью) гомогенизации будут называться: «h0» (негомогенизированные), «h1» (10 с) и «h2» (20 с). Таким образом, полный набор данных для построения лабораторных моделей состоял из 96-ти измерений: 32 образца объединённого набора в трёх степенях гомогенизации каждый.

«Товарная» модель для определения жира в нормализованном (гомогенизированном и стандартизированном по содержанию жира) молоке строилась на большом количестве образцов российских и немецких производителей, приобретённых в торговых сетях. Чтобы увеличить количество уровней градуировочного дизайна, образцы одной марки с различным содержанием жира смешивались между собой в разных пропорциях. Пропорции были составлены таким образом, чтобы шаг по массовой доле жира был равным и составлял от 0.1% до 0.3% в зависимости от диапазона концентраций исходных образцов. Массовая доля жира в образцах немецкого молока варьировалась между тремя нормативными уровнями жирности, установленными в Германии – 1.5%, 3.5% и 3.8%. Каждая серия состояла из 11 образцов, включая два исходных образца жирностью 1.5% и 3.5% или 1.5% и 3.8% и девять попарных смесей. Из 25-ти собранных образцов молока из Германии 11 пар были смешаны друг с другом, а три непарные упаковки анализировались отдельно.

Отечественное молоко было представлено семью уровнями жирности в широком диапазоне: 1.5%, 1.8%, 2.5%, 3.2%, 3.5%, 4.0% и 6.0%. По этой причине количество образцов в отдельных сериях было различным. Из 63-х упаковок российского молока были отобраны 24 пары образцов, которые были смешаны друг

с другом. Количество образцов в каждой серии варьировалось от 6-ти до 11-ти в зависимости от содержания жира в исходных образцах, а шесть непарных упаковок были проанализированы отдельно. Остальные три серии одной марки по три упаковки в каждой, имели три уровня жирности: низкий – 1.5%, средний – 3.2% или 3.5% и высокий – 6.0%. При этом образцы с низким и средним содержанием жира, а также образцы средней и высокой жирности попарно смешивались друг с другом.

Таким образом, жирность образцов градуировочных серий немецкого и российского молока варьировалась в пределах 1.5–3.8% и 1.5–6.0% соответственно. Всего для моделирования было приготовлено 124 образца немецкого и 236 образцов российского молока. Суммарное количество образцов после удаления нескольких явных выбросов составило 295.

«Глобальные» градуировочные модели по жиру и белку были построены на образцах свежего натурального молока, предоставленных заводом «Самаралакто». Образцы были собраны в течение одного календарного года из 72-х крупных и мелких молочных хозяйств Приволжского федерального округа (от 1-го до 90-го образцов от каждого хозяйства). Подавляющее большинство образцов отбирались непосредственно из молочных танкеров (автоцистерн), доставлявших молоко на завод – один образец из каждого танкера либо из каждой секции танкера, если они отличались по составу. Кроме того, было использовано некоторое количество смешанных образцов из накопительных резервуаров производственной линии.

Для расширения концентрационного диапазона по жиру 139 образцов молока были смешаны со свежим обратом. Итоговые диапазоны концентраций по жиру и белку составили соответственно 1.55%–4.97% и 2.27%–4.25%, что отражает вариабельность состава натурального молока [200]. После удаления явных выбросов размеры градуировочных серий по жиру и белку составили 1006 и 1000 образцов соответственно. Собранные таким образом обучающие образцы для «Глобальной» модели охватывают изменчивость сезонных, генетических, географических и климатических факторов, а также специфику содержания животных в различных хозяйствах.

3.4.2 Дизайн градуировочного эксперимента в АКТП

Построение качественных градуировочных моделей для аналитического контроля технологических процессов требует специальных методов ПЭ, учитывающих специфику производственных условий.

Градуировочные модели для онлайн-мониторинга одного или нескольких параметров процесса (как концентраций индивидуальных компонентов, так и комплексных химических показателей) строятся на отобранных из среды идущего процесса образцах. Варьирование состава образца обучающего набора при этом возможно лишь за счёт допустимых изменений состава среды и условий проведения процесса. Помимо химического состава среды процесса, на онлайн-спектры неизбежно влияют различные физические факторы, такие как скорость потока, турбулентность, температура, неомогенность и другие. Учёт этих неустраняемых факторов путём их варьирования в обучающих образцах является залогом устойчивости градуировочной модели.

Целенаправленное варьирование факторов может осуществляться путём изменения состава загружаемого в начале процесса и добавляемого в его ходе материала, а также за счёт изменения управляемых параметров – скорости потока, температуры и т. д. При производстве продукта партиями обучающий набор должен состоять из представительного набора загрузок процесса (партий), охватывающих вариативность релевантных факторов. Большое число отдельных процессов в обучающем наборе даёт дополнительную подстраховку от различных, зачастую неизвестных заранее факторов. Многие из таких факторов являются функциями времени, например: изменчивость сырья, износ оборудования, дрейф анализатора и т. д. Поэтому рекомендуется также использовать набор данных отдельных процессов, накопленный в течение достаточно продолжительного периода времени (недель или месяцев). При таком подходе спланированный градуировочный эксперимент дополняется элементами моделирования на исторических данных (раздел 1.3.8). В настоящем разделе рассмотрены примеры экспериментальных дизайнов градуировки, специально разработанных для целей онлайн-мониторинга процессов в фармацевтическом производстве.

Таблица 3-3. Дизайн градуировочного эксперимента для мониторинга остаточной концентрации воды в грануляте в процессе его сушки [86].

№ ^а	Дни ^б	Спек. ^в	Реф. ^г	Лаб. ^д	Наб. ^е	АФИ ^ж , мг	Прод. ^з , с	ОЗ ^и	%H ₂ O		t _{прод} °C ^к	
									мин	макс	мин	макс
B01	1	841	11	0	val	0.1	70.0	0	4.23	21.41	31.6	38.3
B02	2	831	11	0	cal	0.1	69.8	1	3.82	25.92	30.5	38.2
B03	2	1213	20	0	cal	0.1	102.0	2	2.73	22.38	31.2	45.3
B04	3	640	7	0	val	0.1	53.8	1	5.95	21.49	31.6	38.2
B05	3	589	6	0	cal	0.1	50.2	1	5.09	16.29	31.7	38.3
B06	3	537	7	0	cal	0.1	45.2	1	6.34	22.38	30.7	38.1
B07	4	842	10	0	cal	0.1	70.6	1	2.70	13.05	32.7	45.0
B08	4	552	8	0	cal	0.1	45.9	0	5.57	22.38	30.8	37.7
B09	5	498	6	0	cal	1.0	41.4	0	5.82	23.05	32.1	38.2
B10	5	642	12	6	cal	1.0	53.4	0	4.89	22.72	32.5	38.4
B11	6	641	8	3	cal	1.0	55.3	2	4.36	16.25	32.7	38.2
B12	7	835	18	9	val	1.0	71.7	2	3.81	16.88	32.1	38.3
B13	7	638	9	3	cal	1.0	55.7	1	5.12	20.37	32.1	38.2
B14	8	591	14	6	cal	1.0	50.7	1	3.03	14.54	31.1	38.2
B15	9	537	6	2	cal	1.0	44.6	0	5.59	14.97	31.5	38.3
B16	10	548	6	2	val	1.0	45.6	0	5.59	11.57	32.6	38.5
B17	11	403	9	4	cal	10.0	34.7	1	5.76	16.68	30.7	37.1
B18	12	524	5	2	val	10.0	43.6	0	5.76	11.11	33.9	38.3
B19	12	492	10	4	cal	10.0	40.9	0	4.81	17.43	32.2	38.3
B20	13	497	9	4	val	10.0	41.4	0	3.80	9.92	34.1	37.9
B21	13	667	17	9	val	10.0	55.5	0	4.26	19.07	32.6	38.3
B22	13	610	24	15	val	10.0	50.8	0	4.23	20.97	32.6	38.6
B23	14	1098	26	8	cal	10.0	95.9	4	2.44	21.06	32.8	49.7
B24	14	641	24	13	cal	10.0	55.2	1	2.38	15.01	33.7	49.7
B25	15	396	18	11	cal	0	32.9	0	4.25	11.02	31.7	36.9
Всего^л	15	16303	301	101	88^м	0-10	-	19	2.38	25.92	30.7	49.7

Примечания: ^а идентификационный номер процесса (партии); ^б экспериментальный день; ^в число спектров; ^г число отобранных референтных образцов; ^д число образцов в дополнительном лабораторном анализе; ^е принадлежность к обучающему (cal) или проверочному (val) набору; ^ж уровень дозировки активного ингредиента; ^з продолжительность процесса; ^и число циклов очистки зонда; ^к температура продукта; ^л суммарный результат по всему набору; ^м число образцов в проверочном наборе.

В первом примере градуировочный эксперимент проводился для определения воды в грануляте в процессе его сушки в «кипящем» слое. Основная

информация, описывающая дизайн градуировочного эксперимента в данном исследовании, представлена в табл. 3-3. Спектральные и референтные данные 25-ти спланированных процессов сушки партий гранулята были собраны в течение 15-ти экспериментальных дней (полная продолжительность эксперимента составляла около одного календарного месяца). Параметры процесса систематически варьировались для тренировки модели в широком диапазоне производственных условий. Температура воздуха сушки выставлялась между 90°C и 40°C. В начальной, наиболее активной фазе процесса сушки температуру выставляли на 90°, а затем вручную снижали. Скорость потока воздуха варьировалась между 250 и 300 м³/ч. Текущие температуры воздуха сушки, выхлопа и продукта, а также относительная влажность воздуха сушки и выхлопа непрерывно регистрировались в течение всего процесса. Продукт включал четыре различных уровня дозировки активного лекарственного ингредиента. При планировании эксперимента также учитывалась проводимая в ходе процесса автоматическая очистка зонда «LHP» [86], поскольку было необходимо исследовать её влияние на спектры и соответственно на точность определения воды. Планировались также образцы, независимо отправляемые на лабораторный спектроскопический анализ, который проводился с целью сравнения результатов эталонного метода (спектроскопии Фурье-БИК) с использованием фотометрического шара Ульбрихта (табл. 2-1) с результатами тестирования разрабатываемой ОМС, каковой по сути является данный процесс-спектрометр.

Полный градуировочный набор в данном случае был представлен 301-м референтным образцом. Проверочный поднабор был составлен из данных 8-ми отдельных загрузок процесса (88 референтных образцов). При планировании проверочного набора учитывались условия проведения процесса. Так, экстремальные загрузки, имеющие максимальные или минимальные значения диапазонов концентрации воды в продукте и его температуры (выделены жирным в табл. 3-3) близкие к границам общего по всему дизайну диапазона, не использовались для проверки. Это требование следует из условия (д) составления встроенного проверочного набора (раздел 3.3.1). Погрешность определения воды с

использованием разработанных в данном исследовании градуировочных моделей достигала 0.1%. Результаты исследования представлены в разделе 5.2.2 и в работе [86].

В работах [40,185] строились градуировочные модели для онлайн-мониторинга различных параметров процесса покрытия пеллет (раздел 5.2.3), включающего стадии покрытия и последующей сушки. Всего было исследовано 17 загрузок процесса (табл. 3-4). Четыре из них (В1 по В4) были использованы для проверки возможности онлайн мониторинга концентрации воды и суммарного количества нанесенных материалов (часто выражаемого через толщину наносимого покрытия) методами БИК- и КР-спектроскопии [40], а ещё 13 – для построения модели прогноза полной кривой высвобождения АФИ (*профилей растворения*) для будущих готовых пеллет из онлайн-БИК-спектров [185]. Сложность ПЭ в этих исследованиях связана с высокой чувствительностью процесса к условиям его проведения, в особенности к режиму влажности. Выход из рамок оптимальных условий может привести к нарушению нормального хода процесса и даже к отбраковке продукта. Основными варьируемыми параметрами процесса в обоих исследованиях были: материал наносимого покрытия, скорость распыления, массовая доля твёрдой фазы в распыляемой суспензии и температура воздуха сушки. Эти параметры опосредованно влияют на режим влажности процесса и его продолжительность, а также на продолжительность отдельных стадий. Сочетание этих управляемых параметров определяет качество конечного продукта, наиболее важным показателем которого является профиль растворения.

Первое исследование включало процессы нанесения АФИ и двух материалов покрытия на различную основу. Загрузки В2 и В4 были проведены в возможно более влажных условиях, при которых возможен выход процесса из-под контроля – замедление или даже остановка циркуляции пеллет из-за слипания и образования агломератов, что и имело место в обеих загрузках. Пробоотбор проводился примерно каждые 15 мин, а на стадии сушки – каждые 3–5 мин. Отобранные пробы загрузок В1 и В2 в дальнейшем использовались для разработки лабораторного метода определения толщины покрытия путём анализа цифровых изображений

образцов (раздел 5.2.3).

Таблица 3-4. Дизайн градуировочных экспериментов для онлайн-мониторинга различных параметров процесса покрытия пеллет.

№	Основа	Покрытие	Скорость распыл. ^а , г/мин	%Твёрд. фазы ^б	Время покрытия ^в , мин	Продолжит. процесса ^г , мин	t _{суш} °С ^д
<i>Определение воды и толщины покрытия (через общую массу нанесённого материала)^е</i>							
B1	драже	blue ^ж	23.8	14	150	165	-
B2	драже	АФИ ^з	20.6	12	132	143	-
B3	АФИ	red ^и	25.5	15	147	162	-
B4	АФИ	blue	33.3	15	98	143	-
<i>Прогноз профилей растворения</i>							
W1	АФИ	white ^к	24.6	10	152	170	65.7
W2	АФИ	white	24.8	12.5	123	141	65.7
W3	АФИ	white	26.6	15	96	113	65.8
W4	АФИ	white	20	15	124	136	65.8
W5	АФИ	white	28.9	15	88	104	70.8
W6	АФИ	white	10.6	15	240	255	40.9
W7	АФИ	white	20.5	15	124	139	58.7
Y1	АФИ	yellow ^л	16.4	15	155	167	50.8
Y2	АФИ	yellow	23.6	15	108	120	60.8
Y3	АФИ	yellow	14.2	15	179	190	50.7
Y4	АФИ	yellow	12.9	15	198	208	53.9
Y5	АФИ	yellow	26	15	98	109	72.9
Y6	АФИ	yellow	20	15	127	139	59.0

Примечания: ^а средняя скорость распыления покрытия; ^б содержание твёрдой фазы распыляемой суспензии покрытия; ^в продолжительность стадии покрытия; ^г общая продолжительность процесса (включая стадии покрытия и сушки); ^д температура воздуха сушки; ^е целевая толщина покрытия составляла 16 мкм в B2 и 20 мкм в остальных трёх загрузках; ^ж материал покрытия «Acryl-EZE» с голубым красителем; ^з АФИ с поливинилпирролидоном (ПВП); ^и материал покрытия «Opadry red» с ПВП; ^к неокрашенный «Acryl-EZE»; ^л состав жёлтого цвета на основе «Acryl-EZE» с добавками оксидов железа и титана.

В эксперименте по моделированию кинетики растворения использовались два различных состава «Acryl-EZE», обозначенные W (white) и Y (yellow). Целевая толщина покрытия в этом случае составляла 16 мкм. Согласно ПЭ загрузки W1–W7 и Y1–Y6 проводились при широком варьировании температуры входящего воздуха и скорости распыления, покрывающем практическую вариабельность этих факторов. Покрытые АФИ исходные пеллеты были взяты из одной и той же промышленной загрузки большого объёма, чтобы обеспечить однородность

исходного сырья, исключив этот фактор из моделирования в данном исследовании. Распыляемая суспензия содержала стандартную концентрацию «Acryl-EZE» (15%) во всех загрузках за исключением «экстремальных» – W1 (10%) и W2 (12.5%). Из каждой загрузки серий W и Y отбиралось по 9 проб, соответствующих расходу 12.5%, 25%, ..., 100% полной распыляемой массы покрытия. Последняя 9-я проба отбиралась в конце стадии сушки. Всего для моделирования профилей растворения было приготовлено 117 образцов, на которых был проведён тест растворимости.

ЗАКЛЮЧЕНИЕ К ГЛАВЕ 3

Систематизированы базовые требования к наборам обучающих и проверочных образцов для построения точных и надёжных градуировочных моделей по нескольким факторам, таким как концентрации компонентов смеси или суммарные химические показатели. Сформулированные правила построения и сравнения схем многофакторного дизайна дополняют классическую теорию ПЭ рассмотрением и учётом специфических характеристик, присущих только градуировочному эксперименту или имеющих в этом случае повышенное значение. К ним относятся: обеспечение необходимого числа уровней заполнения и достаточной степени покрытия экспериментального пространства образцами, наличие встроенного проверочного набора, интерпретируемость схем, лёгкость построения и другие качества.

С учётом сформулированных требований спланированы эксперименты, впоследствии проведённые при разработке ОМС и аналитических методик (Приложение В), в частности, для определения жира и общего белка в натуральном или нормализованном молоке, а также для онлайн-анализа процессов: определения воды в фармацевтическом грануляте, определения толщины (через общую массу нанесённого материала) защитного покрытия препарата целевой доставки (пеллет) и прогнозирования результирующей кинетики высвобождения лекарства после приёма пациентом.

Разработан новый подход, получивший название диагонального дизайна градуировочного эксперимента, полностью удовлетворяющий сформулированным

требованиям. Предложенный алгоритм заполнения экспериментального пространства образцами позволяет строить экономичные схемы ДД, впервые предлагающие изначально предусмотренный проверочный набор. Подробно рассмотрены практически важные случаи двух- и трёхфакторных экспериментов. Решена задача обобщения ДД на экспериментальные пространства более высоких мерностей. Диагональная схема любой сложности может быть сгенерирована с помощью предложенного геометрического метода или прилагаемой программы.

ДД является также полезным инструментом разработки ОМС, в частности, облегчая процесс многократного тестирования и сравнения прототипов аналитических устройств и методик анализа.

4. ОМС В АНАЛИТИЧЕСКОМ КОНТРОЛЕ ПРОЦЕССОВ

Аналитический контроль процессов, в особенности технологических, является одним из важнейших разделов современной аналитической химии. Исследования этой научной области охватывают широкий диапазон различных процессов – от кинетики лабораторного синтеза до промышленных производств и экологического мониторинга. Сегодня АКТП быстро развивается, давая импульс исследованиям в области спектрального анализа и обработки многомерных данных. Настоящая глава посвящена теории и практике применения ОМС в АКТП.

4.1 Аналитическое пространство и траектория процесса

4.1.1 Основные понятия и определения

Траектория (*англ.* trajectory) является неотъемлемым атрибутом любого процесса, и этот термин можно встретить в литературных источниках по АКТП. Однако его использование для описания хода процесса до сих пор было стихийным, без чёткого определения или разъяснения. Определения «маршрут» (*англ.* route), «профиль» (*англ.* profile) и «сигнатура» (*англ.* signature) процесса в научной периодике используются синонимично. Выбор термина *траектория процесса* [127,201,202] в настоящей работе обусловлен интуитивностью его понимания и свободой от ложных ассоциаций.

Траектория существует в пространстве переменных, выбранных для отображения хода процесса. В этом смысле траектория является проекцией всего процесса как сложного явления на выбранное пространство. Важно понимать, что нет ни «абсолютного» пространства процесса, ни «истинной» траектории. Можно определить бесконечное число возможных пространств, и результирующие траектории одного и того же процесса могут разительно отличаться. Необходимый набор переменных формирует *аналитическое пространство* процесса и определяется в зависимости от поставленной аналитической задачи. Этот факт подчёркивает решающую роль химика-аналитика в успешном достижении цели анализа процесса.

Рис. 4-1 представляет схематическое изображение траектории в

аналитическом пространстве. В дальнейшем изложении схема будет использоваться для иллюстрации и обсуждения основных свойств траектории.

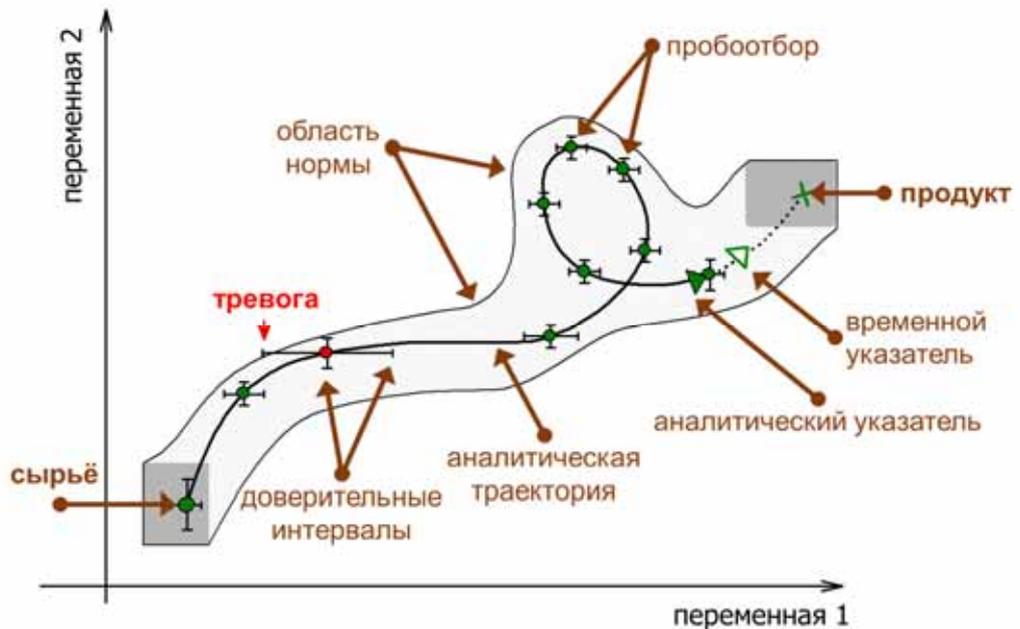


Рисунок 4-1. Схематическое изображение траектории процесса: пройденная (сплошная линия) и прогнозируемая будущая (точечная линия) траектории, точки пробоотбора с оценкой погрешности (зелёные круги), аналитический (зелёный треугольник) и временной (полый треугольник) курсоры, область допустимого хода процесса (заштрихована), области допустимого качества сырья и продукта (серые прямоугольники).

Аналитическое пространство может включать как физические, так и *виртуальные* переменные, являющиеся результатом математических преобразований. Сбор аналитических данных всегда производится в исходном пространстве физических параметров: *изменяемых* (допускающих манипуляции) или *измеряемых* (допускающих только наблюдение). Исходное физическое пространство может определяться как традиционными переменными – характеристиками процесса, такими как давление и температура, так и данными современных детекторов, в том числе спектрометрических. Исходные данные могут быть непосредственно, без какой-либо обработки, использованы для наблюдения за процессом.

Рис. 4-2 представляет пример траектории, полученной при помощи трёх температурных датчиков в ходе процесса сушки фармацевтического гранулята в так называемом псевдооживленном (*англ.* fluidized bed drying) или «кипящем» слое [86]. Три значения температуры, несмотря на взаимные корреляции, отражают

технологические стадии и пертурбации процесса. Традиционные траектории такого типа в настоящее время превалируют в АКТП. Несмотря на то, что они могут быть лишь косвенно связаны с целевым параметром продукта (содержанием воды в вышеприведённом примере), опытные операторы повсеместно используют их для диагностики отклонений хода процесса и определения точки его окончания. Спектральные измерения гораздо более информативны, чем простые физические переменные процесса, однако они менее понятны для визуального восприятия. Сложность интерпретации делает применение новых аналитических приборов и данных эффективным только в комбинации со специальным математическим аппаратом для их представления и моделирования.

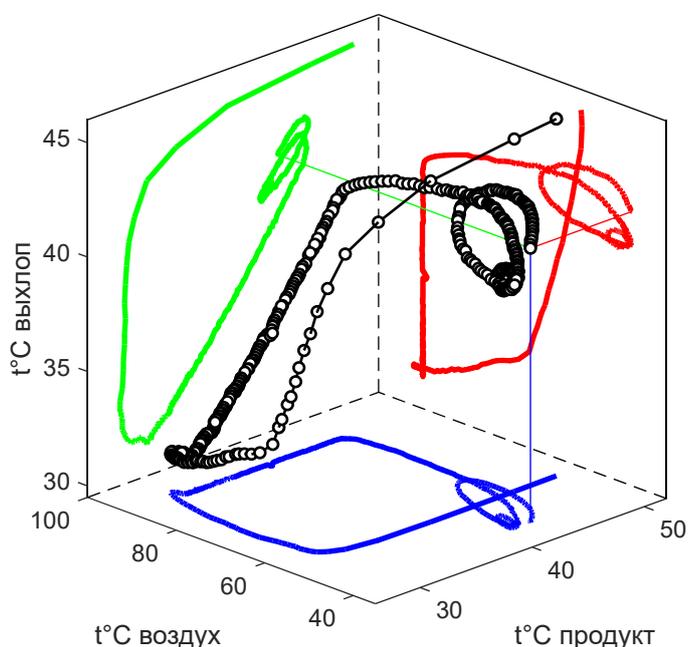


Рисунок 4-2. Траектория процесса сушки гранул в псевдооживленном слое в пространстве трёх физических переменных процесса: температуры воздуха сушки, температуры продукта и температуры выхлопа; попарные зависимости показаны в виде проекций на соответствующие стороны куба. Видео с анимацией траектории доступно в сети Интернет как дополнительный материал к онлайн-версии статьи [9].

Впоследствии исходное аналитическое пространство может быть преобразовано в *рафинированное* пространство, которое упрощает траекторию и лучше подходит для конкретных целей анализа. Эта трансформация состоит в полной или частичной замене исходных физических переменных виртуальными, используя главным образом методы и алгоритмы хемометрики.

В качестве виртуальных переменных могут выступать значения прогноза как

концентраций индивидуальных химических веществ, расходующихся или образующихся в ходе процесса, так и комплексных химических показателей, например содержание биомассы в среде ферментации, общее количество (или толщина слоя) нанесённого защитного покрытия таблетки, или выраженный через ПЛС-счета суммарный спектральный отклик (например при онлайн-анализе гомогенности порошка в процессах смешивания). В приведённых примерах информативность полученной траектории процесса, а значит и результат анализа достигаются без прямого определения индивидуальных химических компонентов анализируемой смеси.

Вводимая в данном разделе методологическая концепция многомерной траектории процесса, будучи естественным атрибутом всех видов процессов, предоставляет собой удобный инструмент для систематического рассмотрения различных аспектов АКТП. Фактически любые задачи аналитического контроля процессов можно рассматривать с точки зрения регистрации, преобразования и последующего использования соответствующих траекторий.

Далее рассмотрим отдельные аспекты хемометрики процессов, имеющие значение при разработке и использовании ОМС.

4.1.2 Дискретизация сбора данных

Траектория процесса может быть расчётно-аналитической, то есть заданной аналитическим уравнением и, следовательно, определённой в каждой точке, непрерывной. Несмотря на важность данного подхода, он выходит за рамки настоящего исследования, которое сфокусировано на экспериментально регистрируемых дискретных траекториях и их хемометрическом анализе.

Траектория задаётся последовательностью измерений, представленных точками в аналитическом пространстве процесса; их координаты задаются значениями выбранных переменных. Используемый здесь термин *измерение* не обязательно предполагает пробоотбор, то есть физическое извлечение образца для последующего анализа. Это может быть и инструментальное онлайн-измерение. Выбранная аналитиком схема последовательности измерений определяет дискретизацию (*англ.* *sampling*) траектории. Проблема оптимальной дискретизации

тесно связана с выбором аналитического инструментария [160,37], а значит, должна быть частью технологической разработки анализатора и планирования процесса.

В недавнем прошлом обычно использовалась максимальная частота измерений, определяемая техническими ограничениями скорости анализа (например при использовании сканирующих оптических спектрометров) или скоростью ручных действий оператора (а позднее – компьютера) по отбору образцов и сохранению данных, то есть использовался минимально возможный шаг дискретизации данных процесса. Сегодняшняя аналитическая и расчётная техника способна обрабатывать десятки и сотни мегабайт данных за доли секунды. К быстрым методам относятся и оптические спектральные сенсоры, например с ДМД (табл. 1-2). Использование анализа изображений или видеозаписи является, вероятно, наиболее ресурсозатратным видом мониторинга в АКТП с точки зрения генерируемого объёма данных [9].

Возможно, в силу этих исторических причин в современных данных существует тенденция к избыточной частоте дискретизации, и количество составляющих траекторию точек может на несколько порядков превышать аналитические нужды. Избыточная дискретизация затрудняет анализ без значимого выигрыша в информативности данных и точности моделирования. В таких случаях прореживание измерений становится обязательным этапом предварительной обработки данных.

Проблема оптимальной дискретизации до сих пор не удостоилась надлежащего рассмотрения в научных трудах по АКТП. Общепринятым является равномерный шаг дискретизации данных процесса. Однако во многих случаях неравномерная дискретизация может быть предпочтительнее. Необходимый шаг измерения следует из крутизны траектории, как видно из рис. 4-2. Следовательно, он должен оптимизироваться исходя из знания закономерностей процесса и динамически подбираться в ходе процесса. В применении к ОМС необходимость оптимального шага дискретизации диктуется также возможными ограничениями объёма памяти устройства, что становится реальной проблемой при

продолжительной автономной работе.

Другая проблема, связанная с дискретизацией сбора данных, заключается в необходимости синхронизировать измерения, поступающие из различных источников, как это бывает в процессе разработки ОМС на основе нескольких спектральных методов. Для того, чтобы синхронизировать сбор данных в ходе процесса, все используемые спектроскопические методы должны управляться одним и тем же ПО, как минимум обеспечивающим единую временную шкалу, а лучше – управляющим триггерами регистрации спектров всех анализаторов. В работе [40] экземпляр коммерчески доступного ПО [203] для управления мультиспектральными системами был использован при совместном онлайн-наблюдении за процессом покрытия пеллет методами БИК- и КР-спектроскопии. При несовпадении шкал разных методов, используемых в дополненных данных, требуется математическая пост-синхронизация (раздел 2.2.1).

4.1.3 Визуализация траекторий

Настоящее исследование преимущественно имеет дело с *периодическими* технологическими процессами, где продукт производится партиями (*англ. batch processes*). Производство партии начинается с загрузки сырья, а заканчивается в момент достижения продуктом заданных свойств. Таким образом, в периодическом процессе траектория – это линия, соединяющая точки, соответствующие исходному сырью и продукту (рис. 4-1).

Траектория идущего процесса прерывается в точке последнего измерения, которую мы будем называть *аналитическим курсором*, по сути «рисующим» траекторию, позволяя производить визуальный мониторинг процесса (рис. 4-1). При наличии задержки, например из-за продолжительности обработки данных, аналитический курсор не совпадает с *курсором временным*, который в этом случае уходит в область гипотетической (прогнозируемой) части траектории, обозначенной пунктиром на рис. 4-1.

Время является особым параметром процесса. Хотя значение переменной времени может быть приписано любой точке траектории, она в общем случае не принадлежит к аналитическому пространству, выполняя описательную функцию

наподобие номера измерения. Особая ситуация, когда переменная времени может присутствовать явно – если она принадлежит к параметрам качества, например время высвобождения активного вещества в процессе растворения лекарственной формы [185]. Таким образом, довольно распространённое понимание траектории процесса как зависимости какого-либо параметра от времени [204] с точки зрения предлагаемой здесь концепции траектории не верно. Этот вывод не противоречит использованию проекций отдельных аналитических переменных на временную ось, называемых *линейными графиками* и часто используемых для изучения различных аспектов процесса. В частности, они служат для визуального исследования и интерпретации сложных многомерных траекторий. Важными примерами временных проекций траектории являются кинетические кривые химических реакций, а также временные зависимости температуры, давления, скорости потока, вязкости и других физических параметров, традиционно используемых для наблюдения за ходом процессов. Примером таких зависимостей для виртуальных переменных являются линейные графики счетов в МГК, ПЛС и других проекционных методах (раздел 1.3.4).

Линейное представление отдельных переменных является простейшим методом изучения траекторий. Альтернативой ему является визуализация траекторий в полном аналитическом пространстве или его подпространстве, образуемом комбинацией двух или трёх аналитических переменных. В отдельных случаях может быть достигнута 4-мерная визуализация, например с использованием цветового градиента в трёхмерном представлении траектории. Наиболее распространёнными являются плоские (двумерные) срезы, такие как график разброса счетов ГК1-ГК2. Визуализация траекторий в пространстве аналитических переменных может выявлять важные закономерности процессов, не передаваемые линейными графиками.

На рис. 4-3 а и б показаны расчётные кинетические кривые реакции с промежуточным продуктом (интермедиатом) [9]. Два представления одних и тех же данных сильно различаются и подчёркивают различные аспекты процесса. Традиционный линейный график с тремя наложенными кинетическими кривыми

концентраций исходного вещества и продуктов реакции (рис. 4-3 а) прекрасно отражает ход реакции, в то время как в трёхмерном аналитическом пространстве концентраций компонентов восприятие времени утрачено (рис. 4-3 б). Зато в последнем случае хорошо видно, что все трёхмерные кинетические кривые лежат в одной плоскости, то есть ранг матрицы данных равен двум (матрица избыточна). При этом разница констант скорости отдельных стадий реакции определяет кривизну соответствующей траектории. Например, при $k_2 \gg k_1$, когда концентрация интермедиата стремится к нулю ($[B] \approx 0$), траектория вырождается в прямую (рис. 4-3 б). Также, на рис. 4-3 б хорошо видна неоптимальность равномерного во времени шага дискретизации сбора данных. Если в начале процесса он недостаточен, то в конце – избыточен. Рассчитав вместо этого равномерно распределённые точки-измерения на двумерных кривых траекторий рис. 4-3 б (или используя другой принцип оптимизации), можно значительно сократить число образцов, необходимых для наблюдения процесса с требуемой степенью детализации.

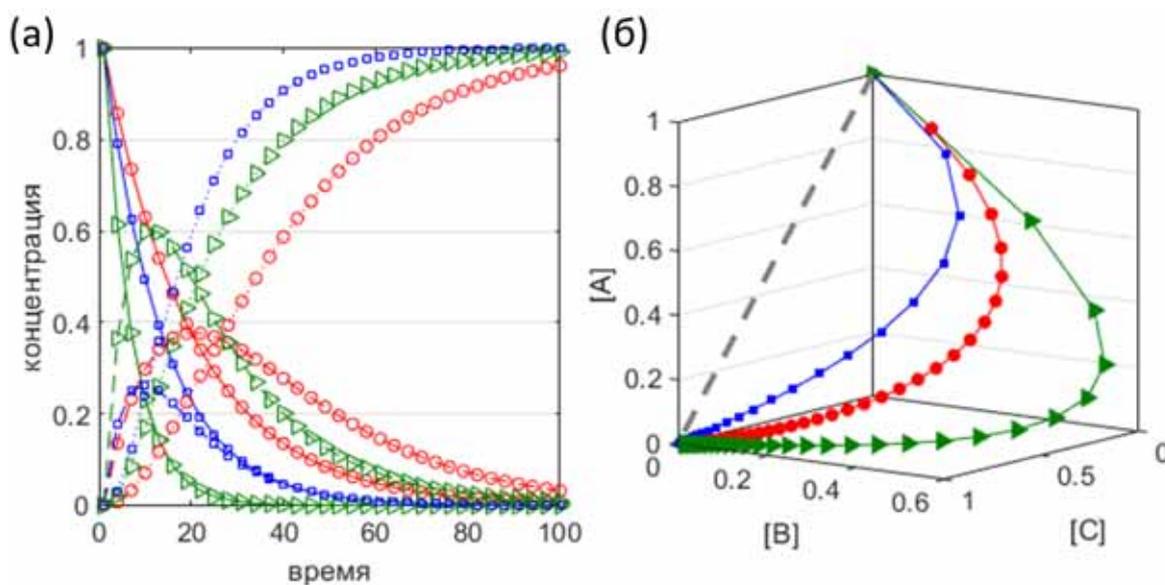


Рисунок 4-3. (а) Модельная кинетика реакции $A \xrightarrow{k_1} B \xrightarrow{k_2} C$: $k_1=0.05, k_2=0.05$ (красные круги); $k_1=0.075, k_2=0.15$ (синие квадраты); и $k_1=0.15, k_2=0.05$ (зелёные треугольники); временные зависимости концентраций компонентов А (сплошная линия), В (пунктирная линия) and С (точечная линия); и (б) трёхмерные концентрационные зависимости; серая пунктирная линия обозначает гипотетическую траекторию при $k_2 \gg k_1$ [9].

На рис. 4-4 даны примеры экспериментальных траекторий процесса. На рис. 4-4 а представлена траектория в пространстве двух ключевых параметров

качества продукта, наблюдаемых онлайн в ходе нанесения защитной оболочки на лекарственный препарат в фармацевтическом производстве пеллет [40]. В исследовании было показано, что используемая ОМС на основе БИК-спектроскопии способна с достаточной точностью определять как содержание воды в продукте, так и среднюю толщину нанесённого покрытия (подробное описание дано в разделе 5.2.3). Влажность среды процесса должна поддерживаться в оптимальном диапазоне, чтобы избежать как слипания (агломерации) в присутствии избыточной воды, так и истирания пеллет вследствие пересушки. При этом точное достижение требуемой средней толщины покрытия определяет точку окончания соответствующей стадии процесса. Для визуализации взаимосвязи концентрации воды и размера пеллет использовали траекторию в пространстве соответствующих параметров. Представление на рис. 4-4 а комбинирует референтные значения, полученные на отобранных образцах для тренировки ПЛС-модели (красные точки), и рассчитанные при помощи этой модели значения прогноза, то есть физические и виртуальные переменные. Освобождённые от эволюционного (временного) аспекта траектории лучше визуализируют локальную динамику процесса, двумерный шум комбинации переменных, эффективность обработки данных (в данном случае – сглаживания), а также точность референтного анализа. Такое графическое представление процесса – его уникальный «автограф» – хорошо подходит для сравнения хода процессов производства различных партий продукта.

В траекторию, представленную в аналитическом пространстве, можно вернуть временное измерение, например при помощи анимации (см. онлайн-версию рис. 4-2 в работе [9]) или цветового кодирования, как показано на рис. 4-4 б. Оси графика представляют значения содержания жира и общего белка в молоке, подлежащие непрерывному мониторингу в ходе производства нормализованного молока для точного достижения целевых характеристик продукта, указанных на упаковке. В этом исследовании [205] содержание жира и белка рассчитывалось из соответствующих ПЛС-регрессионных моделей, построенных на онлайн-спектрах в области 400–1000 нм, которые регистрировались встроенным в

производственную линию процесс-спектрометром⁶. В ходе этого непрерывного производства (общая продолжительность процесса 10 ч 24 мин) производится периодическое переключение между молоком различного состава. Наблюдаемые на рис. 4-4 б «островки» соответствуют временным интервалам стабильного состава продукта. Изменения состава между стандартами жирности производимого молока (3.8%, 3.5% и 1.5%), соответствующее этим продуктам содержание общего белка, а также отклонения от желаемого состава чётко видны на представленной траектории.

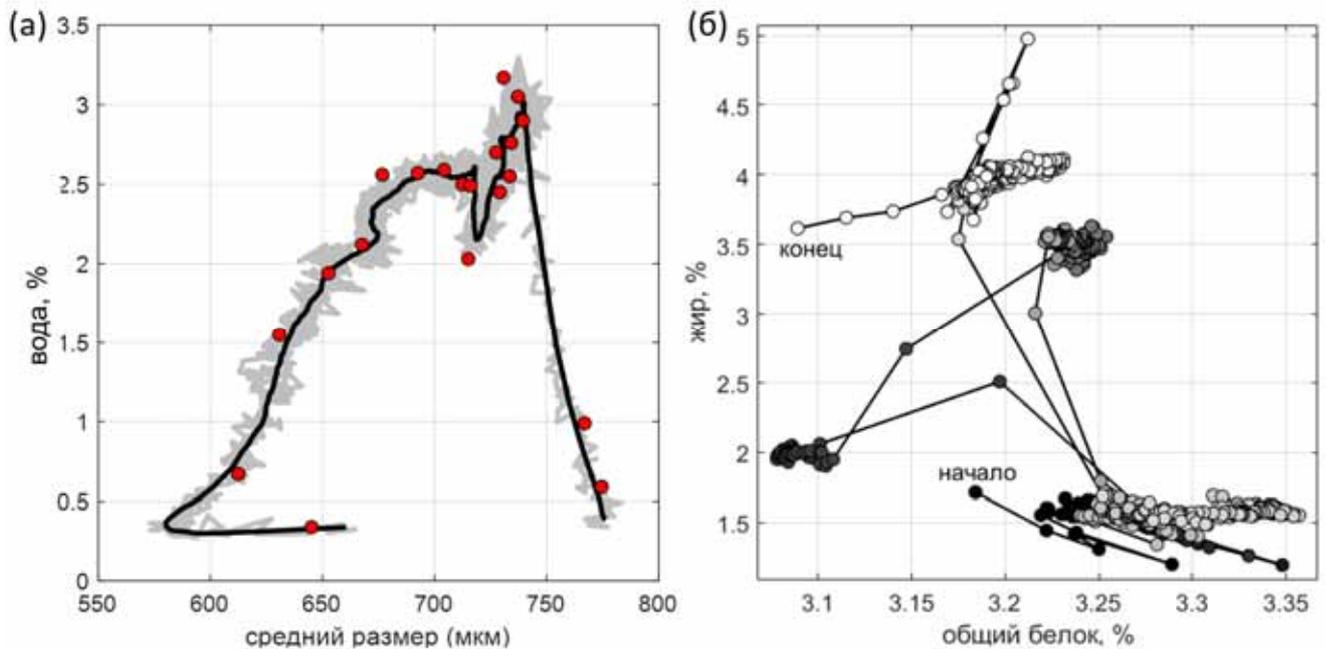


Рисунок 4-4. Примеры двумерных траекторий онлайн-прогноза: (а) процесса покрытия пеллет: сырые (серая линия) и сглаженные (черная линия) траектории, а также референтные измерения (круги); и (б) процесса производства молока в координатах массовой доли жира и общего белка, градиент серого обозначает время процесса [9].

4.1.4 Проблема избыточности аналитического пространства

Мерность немодифицированного аналитического пространства исходных переменных может быть чрезмерно велика. Современные анализаторы генерируют данные, содержащие от сотен и тысяч (в спектроскопии) до миллиардов переменных, например в методах, основанных на получении и анализе цифровых

⁶ Данные получены на заводе «Schwarzwaldmilch GmbH» (г. Фрайбург, ФРГ) и предоставлены фирмой «J&M Analytik AG» (г. Эссинген, ФРГ).

изображений – обычных или спектральных.

Рис. 4-5 даёт пример современных данных видеомониторинга процесса покрытия пеллет (спектральное исследование процесса того же типа обсуждено в разделе 4.1.3) через помещённый в реактор зонд «LHP» со встроенной видеокамерой [9,206]. Видеоданные представляют собой серию кадров – отснятых в ходе процесса изображений. Рис. 4-5 содержит 60 последовательных кадров размером 400×400 точек (края обрезаны), снятых в течение около 1.5 с, что является лишь малой частью обычного в этом случае «образца», включающего более 4000 кадров. Таким образом, исходное пространство содержит более 640 миллионов переменных, упакованных в трёхмерный блок, делая полные данные этого процесса 4-хмодальными. Если бы изображения были цветными, это утроило бы объём данных с добавлением ещё одной размерности.

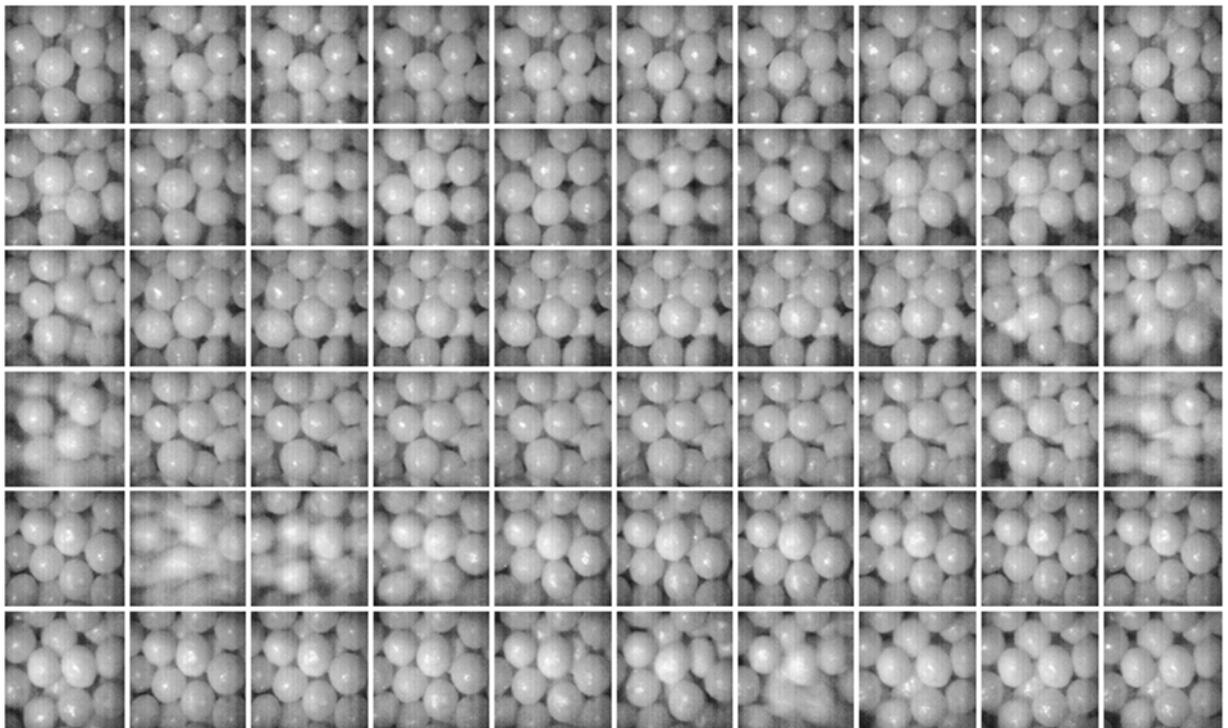


Рисунок 4-5. 60 последовательных кадров видеофильма (слева направо, сверху вниз), снятого в ходе процесса покрытия пеллет [9] (видео доступно в сети Интернет как дополнительный материал к статье).

Интересно, что такие мульти- или мегамерные данные [207], получаемые при использовании изображений, прекрасно подходят для визуального наблюдения процесса, передавая информацию о циркуляции пеллет, образовании мелкой фракции, агломерации и т. д. В то же время любой количественный анализ таких

данных затруднён их аномальным объёмом и избыточностью. Например, основной критерий качества покрытых пеллет – толщина слоя покрытия – напрямую коррелирует с распределением пеллет по размерам или просто с их средним диаметром. Количественный анализ в этом случае может быть основан на предварительном распознавании образов отдельных пеллет в изображениях (методами многомерного или традиционного анализа) и их измерениях [208,209]. После этого на отобранных признаках изображения можно построить градуировочную модель для толщины покрытия. В приведённом гротескном примере размерность аналитического пространства в результате преобразования данных уменьшилась с сотен миллионов (пиксели видеоизображения) до одной переменной (средний размер частицы).

Хотя число переменных в данных онлайн-мониторинга того же самого процесса методами БИК- и КР-спектроскопии (раздел 5.2.3) значительно меньше, оно всё равно является избыточным для решения поставленной аналитической задачи. Предложенная концепция специализации ОМС (раздел 2.1.1) является по сути сокращением избыточного аналитического пространства до оптимального. Причём это сокращение производится ещё на этапе разработки ОМС, а не в ходе анализа уже полученных данных, как при использовании процесс-спектрометра.

4.2 Разрешение и анализ траекторий

4.2.1 Методология анализа процессов

Анализ процесса представлен циклической диаграммой на рис. 4-6. Эта диаграмма разбита на три сектора, объединяющих действия до, в ходе и после окончания процесса (перед следующим циклом). Цикличность – важная концепция, отражающая необходимость эффективного использования уже накопленного (в виде данных и моделей) опыта. Для простоты изложения будем рассматривать задачу наблюдения за процессом посредством ОМС или процесс-спектрометра. Возможное дополнение анализатора неоптическими сенсорами не влияет на суть предложенного подхода. Схема предназначена для помощи в разработке методов АКТП с использованием ОМС. Отдельные шаги схемы

адресуют к ранее рассмотренным и введённым в настоящем исследовании методам и подходам.

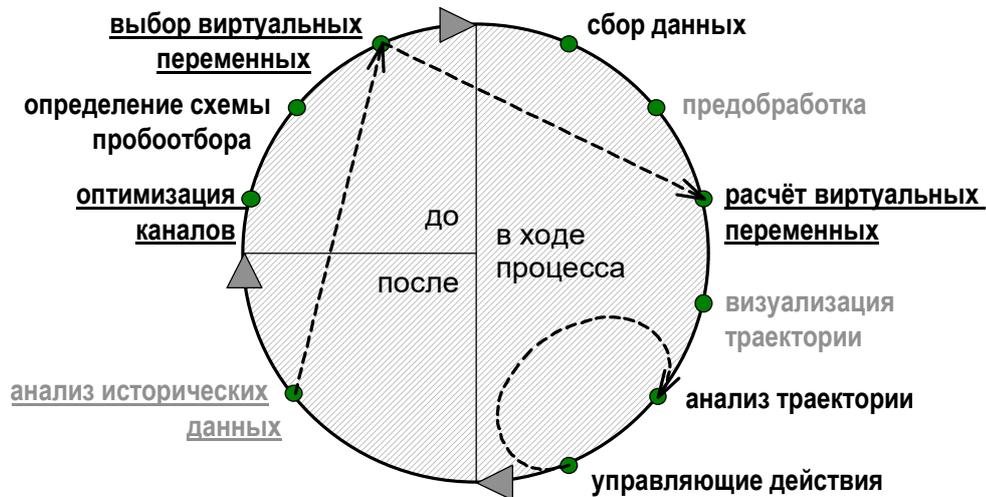


Рисунок 4-6. Обобщённая схема анализа процессов: опциональные действия обозначены серым шрифтом, действия с применением хемометрики – подчеркиванием [9].

Накоплению данных предшествует выбор оптических сенсорных каналов – физических переменных аналитического пространства, в котором будет происходить сбор первичных данных о процессе. Главным требованием к набору каналов является их способность регистрировать данные, необходимые для решения поставленной аналитической задачи. Первоначальный набор сенсорных каналов, как правило, избыточен, как в лабораторном или процесс-спектрометре, представляющем исходный полноспектральный метод (раздел 2.2.1). Соответственно задаче подбираются режим и интерфейс измерения (раздел 2.2.5). Перед началом процесса также необходимо определить различные условия сбора данных, прежде всего выбрать эталонный образец (раздел 2.3.4) и установить шаг (или неравномерную схему) дискретизации сбора данных (раздел 4.1.2). При повторных прохождениях цикла на рис. 4-6 набор сенсоров и параметры сбора данных могут быть оптимизированы на основе полученного опыта. Улучшение может включать переход от спектрометра к ОМС, то есть к оптимальной с точки зрения приложения конфигурации каналов оптической системы (раздел 2.2.2). Кроме того, появляется возможность использования виртуальных переменных – результатов прогноза градуировочной (дискриминационной или иной) модели, построенной на уже собранных данных. На этом этапе, как правило, происходит

сжатие аналитического пространства с применением регрессионных методов, методов отбора переменных, в том числе и признаков изображений.

В ходе процесса собранные исходные данные при необходимости подвергаются предобработке (раздел 1.3.2) и используются для визуализации траектории (раздел 4.1.3), при наличии – с использованием ранее созданной модели. Хотя предобработка и является опциональным шагом, решение о её использовании принимается на стадии моделирования, и она должна рассматриваться как часть имеющейся модели. Для разрешения траектории из многомерных данных могут быть использованы практические любые проекционные методы: МГК, ПЛС-регрессия или МРК (раздел 1.3.5).

Визуализация траектории в одно-, двух- или трёхмерном пространстве нужна, прежде всего, для осуществления основной задачи – мониторинга процесса. Результирующая траектория в её развитии по ходу процесса подлежит непрерывному анализу – визуальному со стороны оператора процесса или автоматизированному при помощи ПО. Целью анализа траектории является своевременное обнаружение ситуации «тревога», означающей выход процесса за рамки нормального протекания (рис. 4-1). В случае возникновения такой ситуации оператором должны быть предприняты управляющие действия для её устранения. Цикл «воздействие на процесс – анализ траектории» повторяется до полного устранения проблемы (петля на рис. 4-6).

После окончания процесса полученные данные становятся историческими и могут быть использованы для создания новой модели или дообучения ранее созданной. Исторические данные могут также использоваться для разведочного анализа (раздел 1.3.4) с целью углублённого понимания данных, а в конечном счёте – для улучшения аналитического метода и даже самого технологического процесса.

На первом цикле схемы на рис. 4-6, например на этапе ввода в эксплуатацию разработанной ОМС (раздел 2.1.2), может быть использована лабораторная модель, построенная на спланированном наборе образцов (глава 3) и адаптированная к условиям процесса при помощи предварительно разработанного метода переноса

модели (раздел 1.3.8). В дальнейшем модель должна обслуживаться наряду с инструментальной частью мультисенсорной системы и при необходимости – дорабатываться (раздел 2.1.2). При использовании стратегии глобального моделирования (раздел 1.3.8) основным материалом для улучшения модели становятся исторические данные всех предыдущих циклов.

Перед началом нового цикла анализа на рис. 4-6 ОМС может быть улучшена как технически – за счёт дальнейшей оптимизации оптической конфигурации каналов (раздел 2.2.2), так и с точки зрения используемого аналитического пространства – выбором новых физических и виртуальных переменных. Действенность улучшения должна быть тщательно протестирована с использованием стандартного тестового набора и соответствующих подходов, например многократной иерархической валидации (раздел 2.4.3).

4.2.2 Качество траектории

Анализ процесса всегда приводит к разрешению траектории. Однако её качество, а значит и польза, могут быть различными. Оценки качества траектории должны учитывать по крайней мере её эффективность и точность.

Эффективность траектории в основном связана с её информативностью, с содержанием релевантной информации о процессе. Однако эта информация должна быть по возможности рафинирована, то есть количество бесполезной информации и шума должно быть минимизировано. Это достигается, во-первых, оптимальным выбором сенсорных каналов, а во-вторых, использованием многомерного анализа данных.

Траектории подвержены многим погрешностям, включая турбулентность процесса, ограничения точности приборов и расчётные ошибки. Таким образом, каждая точка траектории должна характеризоваться доверительными интервалами по каждой аналитической переменной (рис. 4-1), учитывающими различные источники погрешности. Без метрологических оценок траектория практически бесполезна. Точность должна учитываться при всех видах анализа траекторий. Например, выход процесса из-под контроля должен определяться как выход за границы допустимой области не самого аналитического указателя (рис. 4-1), а его

доверительного интервала. Сходным образом, конечная точка процесса должна считаться достигнутой, когда она оказывается в границах доверительного интервала указателя траектории и остаётся там в течение определённого времени для исключения случайных событий. Аналогично, геометрические особенности траектории подлежат физической интерпретации только в том случае, если они значительны по сравнению с погрешностью измерения, с учётом которой траектория является гипертрубкой, а не линией. По мере того, как указатель времени в ходе процесса перемещается, отдельные переменные могут стать более «тревожными» при приближении их доверительных интервалов к границе допустимой области. В то же время другие безопасно находятся внутри неё, несмотря на относительно широкие интервалы доверия. Менее точная траектория увеличивает вероятность ложной тревоги (рис. 4-1).

Неточность другого рода связана с недостаточным шагом дискретизации сбора данных для всей траектории или её критических участков. Здесь необходимо ввести другое важное свойство траектории – разрешение. «Мёртвая петля» на рис. 4-1 иллюстрирует риск потери важных характеристик процесса, если шаг дискретизации слишком велик. При необходимости синхронизировать данные с различными шагами дискретизации (раздел 2.3.1) итоговое разрешение траектории определяется самой разреженной переменной. При использовании равномерной дискретизации для процессов с переменной динамикой, где траектория образует «крутые повороты», как на рис. 4-1 или 4-2, выбранный шаг должен быть достаточно подробным для адекватного описания наиболее динамичных участков. Это может привести к избыточной дискретизации (а значит лишним измерениям и, возможно, ненужным референтным образцам) на остальных участках траектории. Для того, чтобы избежать этого нужна неравномерная дискретизация сбора данных с динамической оптимизацией шага.

Ошибки могут появиться и на этапе применения модели. Даже при идеальной модели недостаточное качество новых данных может привести к неточному прогнозу. Плохие данные могут быть результатом сбоя процесса или аналитического оборудования, а также человеческой ошибки. Оценка соответствия

новых данных модели, а значит, их применимости для прогноза (раздел 1.3.5), имеет решающее значение для достоверности траектории.

Для виртуальных переменных, таких как значения прогноза концентрации или обобщённого химического показателя, важна правильная валидация модели. В ПЛС и других методах многомерной регрессии погрешность обычно выражается через усреднённые оценки, такие как *СКОП*. Однако эта оценка может быть слишком груба для анализа сложных процессов, поскольку она не учитывает зависимости точности от прогнозируемого значения, а значит, оценка точности в отдельных точках траектории может быть завышена или занижена. Исследователи (например [210]) начинают обращать все большее внимание на метрологические проблемы многомерного анализа процессов, но это ещё не стало общепринятой практикой.

4.3 Применение траекторий процесса

Понятие траектории очень полезно при обсуждении анализа процесса в рамках АКТП. Обычно выделяют следующие цели АКТП: мониторинг, контроль, понимание и оптимизация процесса. Каждая из этих целей включает ряд определённых практических задач, решаемых с использованием ОМС. Табл. 4-1 описывает цели, задачи и решения в терминах аналитического пространства и анализа многомерной траектории процесса (см. схему на рис. 4-1).

Таблица 4-1. Цели и задачи, решаемые в АКТП.

Цель	Задача	Решение
Мониторинг	проверка качества сырья/ продукции	<i>сигнализировать о выходе начальной/ конечной точки траектории за пределы допустимой области</i>
	обеспечение нормального хода процесса	<i>обеспечить визуализацию процесса; сигнализировать о выходе траектории за пределы допустимой области</i>
	определение конечной точки процесса	<i>сигнализировать о достижении траекторией предопределённых критериев окончания процесса</i>
	выпуск продукции в реальном времени	<i>анализировать траекторию в аналитическом пространстве, включающем все критические параметры качества, минимизировать отставание анализа (аналитического указателя)</i>
Контроль	реактивный контроль на события (обратная связь)	<i>возвращать отклонившуюся траекторию в допустимую область в ответ на сигнал «тревога»</i>
	проактивный контроль – прогноз конечного состояния (прямая связь)	<i>прогнозировать траекторию с текущего положения аналитического указателя до конечной точки; при ожидаемом выходе за допустимую область – скорректировать ход процесса</i>
Понимание	учёт технических рисков	<i>исследовать траекторию для изучения закономерностей процесса и связи аналитических переменных с событиями процесса</i>
	учёт аналитических рисков	<i>использовать тщательно разработанное аналитическое пространство</i>
Оптимизация	снижение стоимости	<i>проектировать процесс с более короткой траекторией</i>
	повышение качества продукции	<i>оптимизировать аналитическое пространство, сузить допустимую область за счёт повышения точности траектории, повысить управляемость процесса через изменяемые параметры</i>

Примечание: в обзоре [9] даны литературные ссылки с примерами по каждой из задач АКТП.

4.3.1 Анализ процесса в реальном времени

Одной из основных целей АКТП является анализ процесса в реальном времени. Однако аналитическая информация выдаётся с неизбежной задержкой, возникающей при сборе данных, их передаче, обработке и интерпретации, в том числе с помощью математической модели. Следовательно, измерение может считаться произведённым в режиме (почти) реального времени, если отставание аналитического указателя от временного (рис. 4-1) является допустимым. Допуск может варьироваться в очень широком диапазоне в зависимости от общей длины траектории (продолжительности процесса), локальной динамики процесса и аналитических целей. Ключевым критерием при этом является возможность своевременных управляющих действий по предотвращению или ликвидации аварийной ситуации, а также при достижении точки конца процесса. Задержка может увеличиваться в ходе процесса, например, если ПО не успевает обрабатывать данные, или же наоборот сократиться при замедлении процесса. Поэтому соответствие режиму анализа в реальном времени является скорее свойством текущего измерения, а не всего процесса в целом.

Зависимость выдачи результата от скорости обработки и анализа данных делает принятое деление инструментальных методов анализа процессов на «у линии», «на линии» и «в линии» (раздел 1.1.2) не вполне практичным. Легко представить, что онлайн-измерение «в линии» не обеспечивает анализа в реальном времени, если анализатор или алгоритм обработки данных слишком медленный, либо если процесс слишком быстр. С другой стороны, даже лабораторный анализ может считаться анализом в реальном времени, если его продолжительность мала по сравнению с общей длительностью процесса.

4.3.2 Мониторинг процессов

Задачи мониторинга процессов широко распространены в АКТП. Определение точки окончания процесса, раннее выявление аварийных ситуаций, обеспечение качества сырья и конечной продукции – решение этих задач, в конечном счёте, приносит значительное сокращение затрат. С точки зрения траектории и аналитического пространства мониторинг связан с визуальным или

автоматическим отслеживанием аналитического указателя процесса по отношению к границе допустимой области, устанавливающей диапазоны значений аналитических переменных при нормальном ходе процесса с учётом точности аналитического метода (рис. 4-1). Конечной целью мониторинга процесса является выпуск (продукции) в реальном времени (ВРВ), который предполагает, что аналитическое пространство процесса охватывает все критические атрибуты – показатели качества самого процесса и его продукта. В этом случае достижение траекторией конечной точки внутри допустимой области автоматически означает успешное завершение процесса, и дополнительного контроля качества продукта не требуется. Если же дополнительный контроль качества сырья и продукции производится, то контролируемые физические и химические параметры должны быть включены в аналитическое пространство, чтобы оно удовлетворяло требованиям ВРВ.

Так, например, аналитическое пространство на рис. 4-4 а не является полным в смысле ВРВ, поскольку оно не включает важнейшего параметра качества этой лекарственной формы – времени высвобождения активного препарата в желудочно-кишечном тракте пациента. В нынешнем фармацевтическом производстве этот параметр измеряется отдельным лабораторным тестом на специальном оборудовании. Возможность замены теста онлайн-прогнозом этой переменной при помощи БИК-ОМС [185] будет показана в разделе 5.2.4).

Специальный случай представляют траектории процессов, разрешённые при помощи МРК (раздел 1.3.4). Их получение не требует референтного анализа, и может использоваться при тестах ОМС в условиях реального или модельного процесса. Относительные профили МРК, тем не менее, прекрасно подходят для решения различных задач мониторинга в АКТП, в частности для определения конечной точки процесса или отклонения от нормального курса. Они также помогают выявлять промежуточные продукты, механизмы реакций или этапы процесса и поэтому очень полезны для разведочного анализа, который необходим при разработке ОМС для мониторинга процессов.

4.3.3 Контроль процесса

Контроль процесса – ещё одна важнейшая задача АКТП. Используемый здесь термин *контроль* (англ. control) означает активные действия над процессом – управляющие или корректирующие, в отличие от *мониторинга*, который подразумевает только наблюдение. Это разграничение устраняет некоторую терминологическую путаницу, связанную с употреблением слова «контроль», которая особенно часто встречается в англоязычной литературе [159].

Существует два основных подхода к контролю процессов – посредством *обратной связи* (КОС; англ. feed-forward control, FFC) и *прямой связи* (КПС; англ. feedback control, FBC). В обоих случаях контроль осуществляют корректирующие действия при помощи изменяемых переменных. Корректирующие действия совершаются на основании знаний о процессе – опыта оператора или математических моделей. Схема КОС основана на анализе состояния аналитического указателя и реагирует на фактические отклонения, пытаясь сохранить процесс в пределах допустимой области. КПС же сфокусирован на удержании в рамках допустимой области конечной точки траектории процесса, спрогнозированной при помощи модели (пунктирная линия на рис. 4-1). Успех КПС в существенной мере зависит от точности и надёжности модели, используемой для прогнозирования траектории. Такая модель должна по возможности использовать не только данные текущего процесса, но и накопленные исторические данные.

4.3.4 Понимание процесса

Методология АКТП основана на оценке рисков, и улучшение понимания процесса – прямой путь к их предотвращению и минимизации. Волд и др. [159] описывают десять типов рисков, которые можно для простоты разделить на две группы: технические риски, связанные с самим процессом и аналитические, связанные с его анализом. Риски первой группы могут быть сведены к минимуму за счёт углубления знаний о физике, химии, а возможно и биологии процесса, а также его инженерной реализации. Эти знания строятся на аналитическом опыте, полученном при изучении траекторий уже проведённых процессов.

Исследование траекторий позволяет разрешать лабильные интермедиаты и побочные продукты химических реакций, изучать механизмы реакций и строить математические модели процессов. Анализ данных аналитического пространства позволяет выявлять корреляции между переменными, в частности между измеряемыми и изменяемыми параметрами, обеспечивая тем самым эффективность контроля процесса.

Аналитические риски процесса, включая ошибки при регистрации данных, их анализе и интерпретации результатов, а также при их передаче, устраняются (или сводятся к минимуму) с помощью тщательно разработанных, хорошо спланированных аналитических пространств и высококачественных траекторий. Интерпретация траекторий позволяет глубже понять как исследуемый процесс, так и используемый аналитический метод.

4.3.5 Оптимизация процесса

Оптимизация процесса интуитивно ассоциируется с сокращением или упрощением траектории. Этого можно добиться, избегая излишнего потребления сырья, исключая ненужные действия или уменьшая интенсивность этих действий (таких как нагрев, охлаждение, перемешивание, прессование и т. д.). Одной из наиболее распространённых возможностей применения АКТП для экономии в производящей промышленности является точное определение времени готовности продукта. За счёт своевременной остановки процесса исключаются затраты на чрезмерную обработку материала, присутствующие при определении готовности «на глаз» или по результатам лабораторного анализа, приходящим с неизбежной задержкой. Смешивание порошков, стадия сушки при грануляции или покрытии таблеток и пеллет являются другими типичными примерами процессов, которые могут быть оптимизированы с помощью АКТП. С другой стороны, любая оптимизация представляет собой компромисс между издержками производства и качеством продукции. Более высокое качество продукции требует более узкого допустимого диапазона и, следовательно, соответствующего улучшения точности траектории, а значит – бóльших затрат на исследования и разработку.

Производство – это область коммерческой деятельности, и связанные с ним

затраты должны приниматься во внимание при внедрении методов АКТП. Поэтому включение стоимости как одной из переменных аналитического пространства процесса может быть весьма полезным. В конечном счёте жизнеспособность производства зависит от того, способно ли оно обеспечить адекватное и стабильное (а значит – контролируемое) качество продукта в рамках конкурентоспособного бюджета. Поэтому в отраслях с низкой прибылью эффективный аналитический контроль процесса вполне может определить границу между окупаемостью и убытком. В этом смысле развитие ОМС является позитивным явлением не только в рамках АКТП, но и для индустрии в целом.

ЗАКЛЮЧЕНИЕ К ГЛАВЕ 4

Предложена и систематически разработана методологическая концепция траектории процесса в пространстве аналитических переменных. Сформулированы основные свойства траектории и подходы к получению информативных аналитических пространств для решения задач онлайн-мониторинга, контроля и оптимизации технологических процессов как непрерывного цикла улучшений на основе накопленных знаний.

Даны практические рекомендации по использованию траектории и аналитического пространства для разработки аналитических методов, эффективного сбора данных, их непрерывного анализа и управления процессами в АКТП, в том числе с использованием оптических мультисенсорных систем. Траектория является важным инструментом, способным повысить точность ОМС для аналитического контроля процессов и снизить затраты на их разработку.

5. ПРИМЕРЫ РАЗРАБОТКИ И ПРИМЕНЕНИЯ ОМС

В этой главе собраны иллюстративные примеры разработки и практического применения ОМС для решения аналитических задач в различных практических областях: пищевом и фармацевтическом производстве, биотехнологии, медицинской диагностике и экологии.

Представленные системы (табл. 5-1) находятся на разных стадиях разработки (табл. 2-1), следующей принципу упрощения полноспектрального метода путём оптимизации под выбранное приложение. Основой оптимизации при этом служат предварительные измерения на лабораторном или процесс-спектрометре.

5.1 Определение жира и общего белка в молоке

Раздел обобщает серию исследований, в которой экспериментально и теоретически обосновывается возможность количественного определения питательных компонентов молока методом Вид/КВБИК-спектроскопии, осложнённой множественным рассеянием света коллоидными частицами разных размеров. Рассмотрены различные конструкционные решения для создания ОМС-анализатора суммарного жира и суммарного (общего) белка в натуральном и переработанном молоке.

5.1.1 Оптический анализ молока на содержание жира и белка

Массовые доли жира и общего белка – два основных показателя качества молока, определяющие его питательную ценность и подлежащие контролю. Российские производители обязаны соблюдать нормы ГОСТа [211].

Анализ молока в современной промышленности основан на отборе необходимых проб, отправляемых в специализированную лабораторию. Для определения содержания жира и общего белка в молоке широко используются химические методы [212,213], являющиеся наиболее точными, но в то же время – медленными и трудоёмкими. Поэтому всё большее распространение в молочной промышленности получают различные инструментальные физические методы анализа, центральное место среди которых занимает оптическая спектроскопия (табл. 5-2).

Таблица 5-1. Примеры разрабатываемых ОМС.

№	Анализируемые параметры	Объект	Анализ	Спектр. метод, область	Полноспектр. измерения ^а		Анализ данных ^б	Тип ^в ОМС	Публ. ^г
					спектрометр	интерфейс			
1	Жир и общий белок	натуральное молоко	онлайн	Вид/КВБИК	ММС	проп. ячейка	ПЛС	3	[84]
2			экспресс	Вид/КВБИК	ДМД	кювета	ГА, ПЛС	2	[85,182]
3			экспресс	Вид/КВБИК	фотокамера	объектив	КАПВ, СГ, ПЛС	1	[90]
4			полевой	Вид/КВБИК	ПЗС	кювета	ГА, ПЛС	1	[79,89]
5		нормализов. молоко	полевой	Вид/КВБИК	ДМД	кювета	ГА, ПЛС	1	[199]
6	Вода в составе порошков и твёрдых фармацевтич. форм	гранулят пеллеты	онлайн	БИК	ДМД	ДО-зонд	МГК, МСС, ПЛС	2,3	[40,86]
7		пищевая добавка	онлайн	БИК	мини-спектрометр	оптическое окно	МСС, МГК	3	[91]
8	Общая масса покрытия (толщина)	пеллеты	онлайн	БИК+КР	ДМД и КР	ДО-зонд	МРК, КЛК, ПЛС	3	[40]
9			онлайн	Вид	фотокамера	объектив	МУИ, ПЛС	3	[208]
10	Кинетика растворения АФИ		онлайн	БИК	ДМД	ДО-зонд	ПБО, ПЛС	3	[185]
11	Биомасса	ферментац. среда	онлайн	БИК	ДМД	трансфлек. зонд	ПЛС	3	[65]
12			онлайн	флуоресценция	лаз. диод, ПЗС	ДО-зонд	МРК, ПЛС	1	[81]
13			онлайн	2D-флуоресценция	монохроматор, ПЗС	ДО-зонд	МРК, ПЛС	1	[81]
14			Этанол и углеводы	онлайн	ИК	Фурье-ИК	НПВО-зонд	ПЛС	2
15	Онкологические изменения	биологич. ткань (почка)	экспресс	БИК	ДМД	НПВО-зонд	МГК, ПЛС-ДА	1	[27,78]
16			экспресс	ИК+флуоресц.	Фурье-ИК, миниспектр.	НПВО-зонд, ДО-зонд	МГК, ПЛС-ДА	3	[88]
17		пр. кишка	экспресс	БИК+флуоресц.	ДМД, минисп.	ДО-зонды	МГК, ПЛС-ДА	3	[214]
18	Углеводороды нефти	почва	полевой	ИК	Фурье-ИК	НПВО-прист.	ГА, СГ, СНВ, ПЛС	2	[87,215]

Примечания: ^а в соответствии с табл. 1-2; ^б в соответствии с табл. 1-4 с использованием ПО: «PLS_Toolbox» («Eigenvector Research», США), «The Unscrambler» («Camo», Норвегия), «MATLAB» («MathWorks», США) и «TPT cloud»; ^в в соответствии с табл. 1-3; ^г публикации автора.

Анализ производимого в заводских условиях молока на содержание жира и белка осуществляется на разных стадиях технологического процесса. В обязательном порядке контролируется качество поступающего на завод сырья – сырого молока, имеющего различное происхождение. Оперативному контролю подлежит каждая поступающая на завод порция сырья, к примеру, автоцистерна, которая часто содержит смешанное молоко, собранное в нескольких хозяйствах. При несоответствии требованиям, в частности из-за недостатка общего белка, забраковывается весь объём молока, хотя проблема может быть связана с конкретным хозяйством. Чтобы избежать этой ситуации и связанных с ней потерь, анализ должен производиться в момент забора молока в хозяйствах, а для этого необходим полевой экспресс-метод и соответствующий анализатор. С ростом количества молочных хозяйств и наращиванием общего объёма производимой продукции проблема контроля качества становится важнейшей задачей отечественной молочной промышленности.

Другой важный объект анализа – нормализованное гомогенизированное молоко, являющееся самым массовым коммерчески производимым молочным продуктом. Содержание жира в нем может тонко регулироваться в процессе производства, но для этого оно должно постоянно отслеживаться практически в реальном времени. Отсюда возникает задача быстрого и точного определения жира в большом количестве проб, отобранных с производственной линии, то есть из протекающего по трубе молока. Даже современная заводская лаборатория не вполне справляется с этой задачей из-за ограниченной пропускной способности, связанной со значительной продолжительностью и относительно высокой стоимостью анализа. Решением в данном случае было бы внедрение онлайн-метода на основе встроеного в производственную линию анализатора, дающего мгновенный результат без необходимости пробоотбора.

ИК-спектроскопия, благодаря своей высокой чувствительности и селективности, является общепринятым лабораторным методом анализа молока в России и за рубежом (раздел 1.1.1). Допустимые по ГОСТ [216] пределы абсолютной погрешности измерений ИК-анализаторов составляют 0.11% для белка

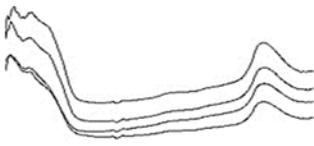
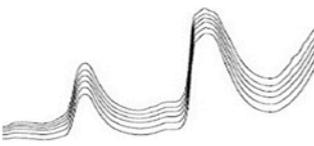
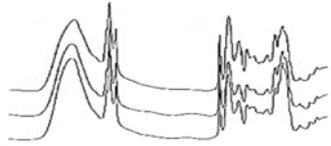
и 0.12% для жира. Те же погрешности на приборе «MilkoScan» по данным производителя [95] значительно ниже (0.05–0.06%), и многие исследователи используют этот прибор для референтного анализа при построении градуировочных моделей [72,95,217,218]. Для точного ИК-анализа молока в режиме пропускания прибор производит интенсивную пробоподготовку, включающую отбор аликвоты, её гомогенизацию и термостатирование. Для обеспечения высокой производительности пробоподготовка, последующий анализ и обработка полученных данных полностью автоматизированы. Высокая техническая сложность существенно повышает стоимость прибора и его обслуживания, что делает «MilkoScan» доступным лишь для крупных предприятий. Будучи устоявшейся лабораторной техникой, ИК-спектроскопия мало пригодна для онлайн-мониторинга процессов из-за высокого поглощения оптических конструкционных материалов в этой спектральной области и технической сложности встраивания метода в производственную линию.

БИК-спектроскопия, основанная на относительно слабых и менее селективных сигналах (табл. 5-2), является конкурентоспособной альтернативой спектроскопии средней ИК-области. БИК-анализ более экономичен, благодаря использованию кювет, светопроводящих волокон и другого оборудования, изготовленного из широко распространённых оптических материалов. БИК-анализаторы отлично подходят для онлайн-мониторинга технологических процессов [8]. Применимость БИК-спектроскопии в практическом определении содержания молочного жира и белка доказана многочисленными исследованиями последних 20 лет [72,217-223]. Существуют и действующие молочные анализаторы, работающие на этом принципе. Однако доминирование полос поглощения воды и множественное рассеяние света коллоидными частицами молока [224-227] осложняют анализ и понижают его точность, что ограничивает практическое применение метода.

Область видимого света обычно не используется для определения питательных веществ натурального молока, поскольку рассеяние коллоидными частицами жира и белка в этой области превосходит поглощение примерно на два

порядка (отсюда белый цвет молока), препятствуя количественному анализу. Существует всего несколько работ, исследующих спектроскопию в этой области для анализа натурального молока [228,229], однако и в них применяется экспериментальная или математическая компенсация эффекта рассеяния в спектрах.

Таблица 5-2. Сравнительная характеристика оптических спектральных методов анализа молока.

Критерий	Вид/КВБИК 360–1100 нм	БИК 1100–2500 нм	ИК 2500–25000 нм
Спектр (схематическое изображение)			
Физический принцип	рассеяние (преимущественно)	поглощение	поглощение
Геометрия измерения	диффузное отражение или пропускание	пропускание	пропускание
Аналитический отклик	суперпозиция спектров рассеяния коллоидными частицами	полосы поглощения обертонов и комбинационных линий	полосы поглощения фундаментальных частот колебаний
Чувствительность	низкая	средняя	высокая
Селективность	очень низкая	низкая	высокая
Онлайн-анализ	легко реализуем	возможен	не применяется
Осложнения	необходимо применение хемометрики	сильное поглощение воды и рассеяние	требуется высокий уровень автоматизации
Распространённость	редко используется	широко распространён	лабораторный стандарт
Стоимость	низкая	средняя	высокая

Предложенный нами в работе [84] подход рассматривает рассеяние света в Вид/КВБИК области не как помеху анализу, а как источник информации о составе молока. Вид/КВБИК-область позволяет разрабатывать технически относительно простые и недорогие анализаторы, а компактность и гибкость в анализе позволяет адаптировать их как для онлайн-мониторинга в АКТП, так и для полевых

измерений. В то же время спектральная область 400–1100 нм, которая полностью покрывается одним классом детекторов, является наиболее благоприятной для разработки оптических мультисенсорных систем. Очень перспективным является использование светодиодов – источников света, излучающих в узком спектральном интервале, то есть близких к монохроматическим. Выбор светодиодов в Вид/КВБИК очень широк, так что светодиод с желаемой (расчётной) интенсивностью и центральной длиной волны может быть подобран достаточно точно. Основные конструкционные элементы ОМС в этой спектральной области, включая источники, световоды и детекторы, доступны от широкого круга производителей по относительно низкой цене.

Спектральный, в том числе мультисенсорный, анализ в Вид/КВБИК-области может производиться в режиме как пропускания, так и диффузного отражения [72,178], предоставляя разработчику ОМС дополнительную гибкость технических решений. Методы пространственно-разрешённого измерения, с использованием двух [178,230] или нескольких [80,179] каналов детектирования, улавливающих рассеянное излучение под различным углом к источнику, способны предоставлять дополнительную (помимо спектральных эффектов рассеяния) информацию о размерах рассеивающих частиц в образце молока, а значит, повышать точность количественного определения жира и белка.

Проведённый анализ выявляет неполное соответствие существующих аналитических методов определения суммарных жира и белка экономическим и нормативным требованиям современной молочной промышленности. Разработка новых эффективных оптических методов онлайн- и полевого экспресс-анализа молока является важной и весьма перспективной.

5.1.2 Обоснование спектрального метода определения молочного жира и белка в области 400–1100 нм

Учитывая востребованность экспресс-методов анализа питательной ценности молока (раздел 5.1.1), в работах [84,85,95] был предложен новый подход к определению жира и общего белка, основанный на рассеянии света. Он был положен в основу разработки соответствующей ОМС для анализа как

гомогенизированного, так и сырого натурального молока [95].

В качестве предварительного исследования была проверена возможность спектрального определения жира в готовом продукте – промышленно нормализованном (гомогенизированном и стандартизированном по составу) молоке, представленном в торговых сетях России и Германии. Анализ проводился с использованием ДМД-спектрометра (табл. 5-1).

Построение градуировочной модели «Товарная» (табл. 3-2) для определения суммарного жира в нормализованном молоке производилось в соответствии со специально разработанным дизайном эксперимента (раздел 3.4.1). Градуировка проводилась методом ПЛС-регрессии без предобработки спектров. Статистические показатели для результирующей модели, построенной на 295-ти образцах молока составили соответственно $SKOPII = 0.131$ и $R^2 = 0.974$ при 6-ти ЛП. Соответствующие $SKOPII$ моделей для молока отдельных стран были: 0.126 для немецкого и 0.125 для российского, обе при пяти ЛП. Поскольку содержание белка в данном эксперименте не варьировалось, градуировочная модель для него не строилась. Более подробно результаты представлены в работах [95,199]. Построение рабочей прогностической модели не было задачей данного эксперимента. Полученные спектры весьма чувствительны к содержанию жира в образцах. На повышение жирности на 0.1% спектр поглощения реагирует заметным положительным сдвигом базовой линии, что соответствует росту рассеяния на всех длинах волн, обусловленному увеличением количества глобул жира на единицу объёма. Относительно высокая сложность моделей свидетельствует о присутствии других, помимо жирности, факторов, влияющих на Вид/КВБИК спектры. Исследование подтвердило принципиальную возможность спектрального определения содержания жира в нормализованном молоке, основываясь на эффекте рассеяния в КВБИК-области. В отличие от известных сходных оптических методов турбидиметрии и нефелометрии, работающих на единичных длинах волн [231-233], предложенный подход работает без предварительной подготовки образцов, а полученные градуировочные модели устойчивы к различиям в происхождении, составе и технологических процессах

изготовления исследуемого молока. По результатам полноспектрального анализа были рассчитаны оптимальные интервалы длин волн, которые могут быть положены в основу создаваемой ОМС, например светодиодной [199].

Следующим шагом исследования была проверка применимости предложенного метода к более сложной задаче количественного определения жира и общего белка в негомогенизированном молоке. В эксперименте использовались искусственные образцы молока близкие к натуральному, содержание жира и белка в которых варьировалось путём смешивания свежих сливок, обрат и раствора лактозы в соответствии с разработанным планом эксперимента (модель «Заводская», раздел 3.4.1). Полученные спектры диффузного пропускания градуировочной серии представлены на рис. 5-1. Наличие в спектре некомпенсированной полосы воды около 980 нм ожидаемо с учётом того, что в качестве эталонного образца для нормирования по уравнению (2.1) использовался оптический фильтр, понижающий оптическую плотность на 4 порядка (раздел 2.3.4).

Использование предложенного дизайна эксперимента (рис. 3-17) позволяет изучить спектральные эффекты жира и белка независимо друг от друга. Так, увеличение содержания жира при равном содержании белка преимущественно проявляется в сдвиге базовой линии (спектры 18 и 21), что, как и в случае гомогенизированного молока, объясняется ростом плотности (количества на единицу объёма) рассеивающих частиц. Спектры образцов 16 и 19 соответственно представляют минимальное и максимальное содержание белка при равенстве содержания в них жира, что проявляется в различии наклона спектров.

Для улучшения точности моделей наиболее подверженные шуму переменные по краям спектров были исключены (рис. 5-1). *СКОП* прогноза по предусмотренному экспериментальным планом тестовому набору составляют 0.045 ($R^2 = 0.964$) для жира и 0.030 ($R^2 = 0.946$) для белка при двух ЛП, дающих оптимальные модели. Как и в эксперименте с нормализованным молоком, полученные статистические характеристики оценочны и могут быть чересчур оптимистичны. Однако целью моделирования на этом этапе было углублённое

изучение данных, а не построение рабочих градуировок.

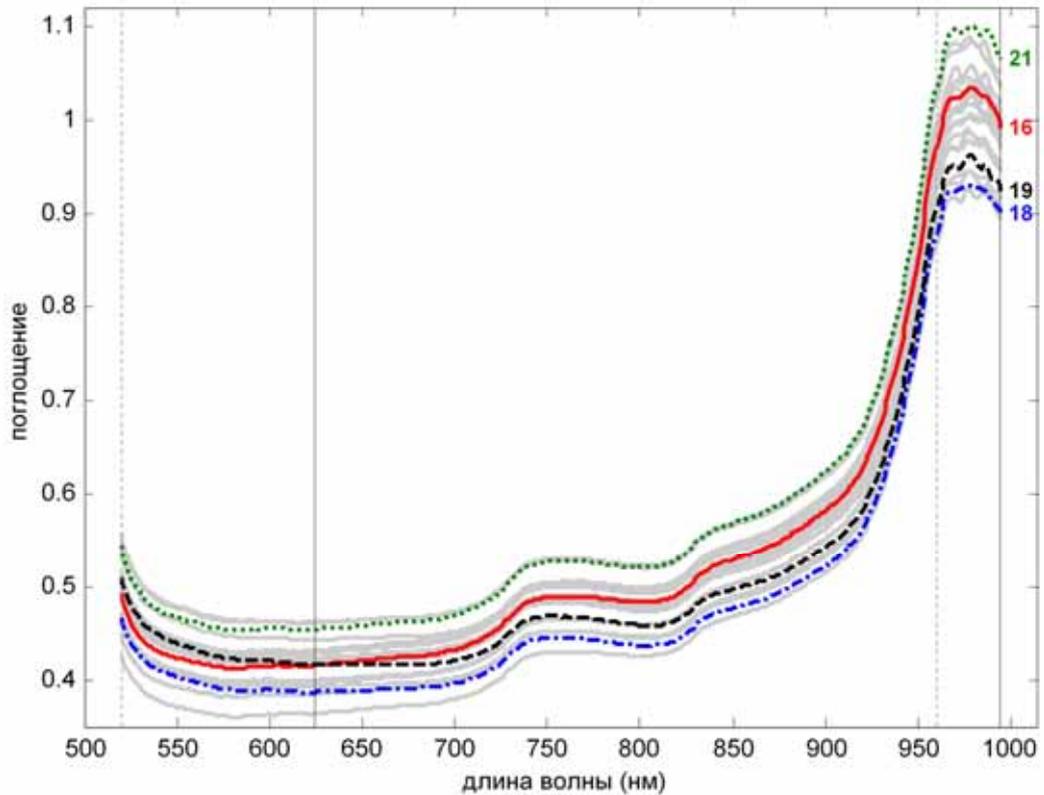


Рисунок 5-1. Вид/КВБИК спектры образцов «Заводской» модели. Выделены четыре образца с максимальным и минимальным содержанием жира (18 и 21) и белка (16 и 19); номера соответствуют схеме на рис. 3-17. Границы спектральных областей, выбранных для построения градуировочных моделей для жира и белка, обозначены сплошной и пунктирной вертикальными линиями соответственно [84].

Сравнение графиков ПЛС-счетов двух первых ЛП (данные для жира и белка совмещены в одном графике на рис. 5-2 а) с дизайном эксперимента на рис. 3-17 выявляет интересный факт. Направление ЛП1 примерно совпадает с направлением роста содержания жира в образцах, а ЛП2 – с содержанием белка. Это позволяет трактовать соответствующие нагрузки по первым двум ЛП (рис. 5-2 б) как разрешённые суммарные спектры жира и белка. Следует отметить, что подобная интерпретируемость абстрактных по своей природе ЛП ПЛС-модели в общем случае не ожидается и является скорее удачным совпадением, чем закономерностью.

Учитывая, что рассеяние значительно преобладает над другими оптическими эффектами в образцах, нагрузки по ЛП1 и ЛП2 на рис. 5-2 б можно условно считать спектрами рассеяния жира и белка соответственно. Спектральный профиль

рассеяния глобулами жира образует почти прямую горизонтальную линию, то есть его интенсивность не зависит от длины волны. Это соответствует теоретическим представлениям так называемого геометрического рассеяния света большими частицами (средний диаметр жировых глобул преимущественно составляет 1–10 мкм [200]). В то же время интенсивность рассеяния значительно меньшими мицеллами белка (до 80–200 нм [229]), отчётливо падает с длиной волны, совпадая с результатами расчёта по теории Ми [84]. Сужение используемых в градуировке данных до очень узкой спектральной области 600–700 нм, где не ожидается никаких оптических явлений, кроме рассеяния, приводит лишь к небольшому повышению *СКОП* (по сравнению с рассмотренными выше оптимальными интервалами длин волн): до 0.057 для жира и 0.044 для белка. Приведённые результаты свидетельствуют о том, что именно рассеяние является основой предложенного спектрального подхода к количественному определению коллоидных компонентов молока.

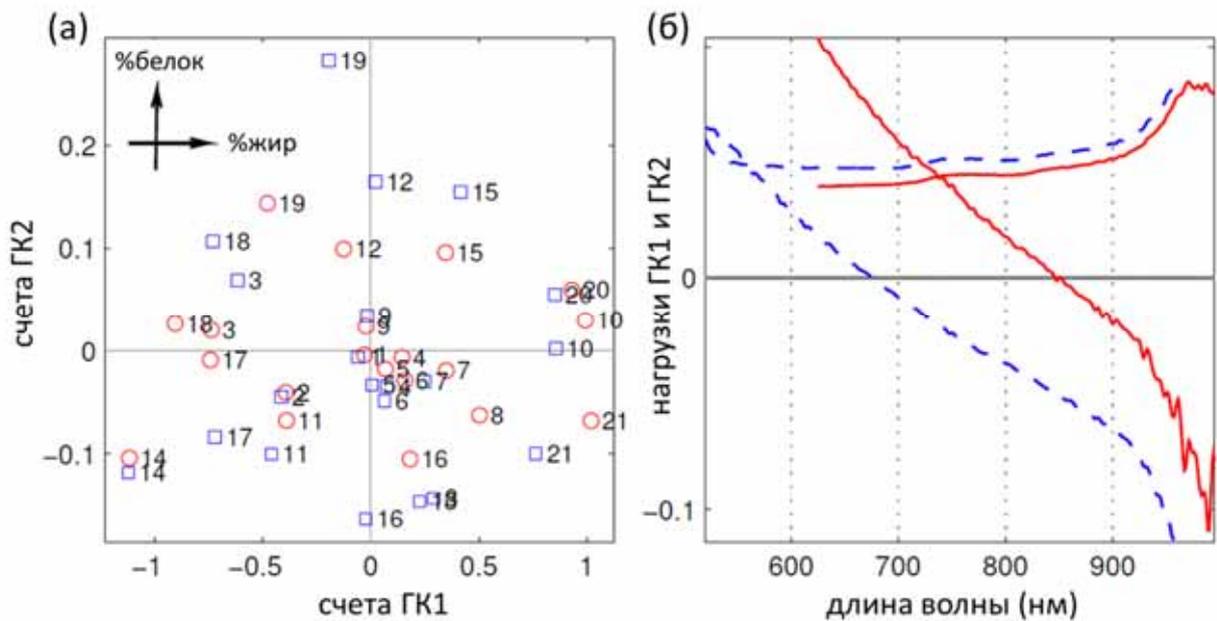


Рисунок 5-2. Две первые ЛП ПЛС-регрессионных моделей содержания жира (красные круги и сплошные линии) и для белка (синие квадраты и пунктирные линии): (а) счета и (б) соответствующие нагрузки [84].

Разделение спектральных эффектов жира и белка в факторном пространстве ПЛС-модели показало, что не только полосы поглощения, но и низкоселективные спектральные отклики, такие как профили рассеяния различных частиц, в частности глобул молочного жира и мицелл белка могут использоваться для

количественного анализа. Ключевым условием при этом является использование методов многомерного анализа данных для компенсации недостатка селективности.

Глобулы молочного жира имеют сложное распределение по размерам [234], которое в природных образцах может варьироваться. Поэтому успех количественного оптического анализа в дисперсной среде натурального молока в значительной мере зависит от способности метода различать рассеяние жировыми глобулами различного размера. Для исследования влияния варибельности размера жировых глобул на устойчивость многомерной модели был спланирован и проведён отдельный лабораторный эксперимент.

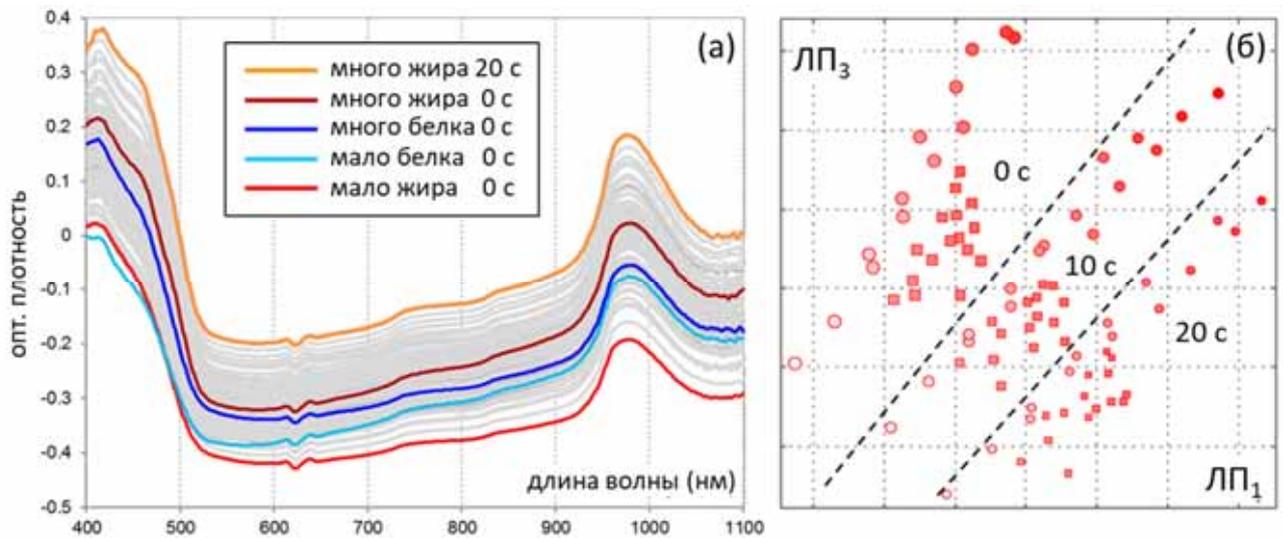


Рисунок 5-3. (а) Спектры 96-ти измерений градуировочных образцов молока «Лабораторной» модели (серые линии) с выделенными цветом характерными образцами при различном времени гомогенизации; и (б) их представление на графике счетов ЛП1–ЛП3 ПЛС-модели для жира. Насыщенность цвета соответствует концентрации жира, а размер символа – степени гомогенизации образца; круги обозначают образцы с преимущественным варьированием жира, а квадраты – белка [235].

Спектры 32-х образцов «Лабораторной» модели в трёх степенях гомогенизации каждый (раздел 3.4.1) показаны на рис. 5-3 а. Как и в случае «Заводской» модели, рост содержания жира характеризуется равномерным увеличением спектральной интенсивности во всей исследуемой области, а различие в содержании белка преимущественно проявляется в наклоне базовой линии. Влияние гомогенизации сходно с эффектом увеличения содержания жира, поскольку также ассоциируется с плотностью рассеивающих частиц в образце, но

имеет свои спектральные особенности (выделенные спектры на рис. 5-3 а).

Табл. 5-3 сравнивает статистические характеристики градуировочных моделей для жира и общего белка, построенных с применением различных методов предобработки данных. В соответствии с принципами многоуровневой валидации, изложенными в разделе 2.4.3, сравнение производится по четырём статистическим критериям, включая три метода проверки.

Таблица 5-3. Статистика ПЛС-градуировок на содержание жира и белка для измерений по дизайну «Лабораторной» модели.

Набор данных	ЛП ^а	Градуировка		ППП ^б		ППС ^в		ПТН ^г	
		СКОГ	R ²	СКОПП	R ²	СКОПП	R ²	СКОП	R ²
<i>содержание жира</i>									
F ^д	5	0.121	0.963	0.134	0.954	0.141	0.949	0.130	0.952
F _{МКР} ^е	4	0.198	0.899	0.215	0.882	0.224	0.872	0.198	0.893
F _{1D2.15} ^ж	4	0.140	0.950	0.159	0.935	0.164	0.931	0.171	0.913
F _{out3} ^з	5	0.104	0.971	0.115	0.965	0.122	0.963	0.110	0.963
F _{out7} ^и	5	0.089	0.977	0.098	0.972	0.102	0.974	0.103	0.969
F _{out7i} ^к	5	0.069	0.986	0.077	0.983	0.077	0.983	0.070	0.984
<i>содержание общего белка</i>									
P	4	0.040	0.965	0.042	0.961	0.043	0.960	0.040	0.959
P _{МКР}	5	0.059	0.923	0.071	0.889	0.073	0.884	0.061	0.911
P _{1D2.15}	4	0.041	0.962	0.048	0.948	0.051	0.943	0.054	0.929
P _{out1} ^л	4	0.040	0.965	0.042	0.961	0.043	0.960	0.040	0.960
P _{out1i} ^м	5	0.037	0.971	0.040	0.965	0.041	0.964	0.038	0.965

Примечания: ^а Число ЛП в ПЛС-модели; ^б полная перекрёстная проверка и ^в перекрёстная проверка сегментами; ^г проверка тестовым набором; ^д полный набор данных в моделях по определению содержания жира (F) или общего белка (P); ^е спектры, предварительно обработанные методом МКР или ^ж первой производной (алгоритм Савицкого-Голея с полиномом второго порядка и шириной фильтра 15 точек); ^и модель F с семью выбросами; ^к модель F_{out7} с выбранными оптимальными спектральными интервалами при помощи ИПЛС: 520–539 и 860–959 нм (в остальных моделях для жира использовался оптимизированный полный интервал 420–1035 нм); ^л модель P с одним выбросом; ^м модель P_{out1} с выбранными при помощи ИПЛС с оптимальными спектральными интервалами: 440–459, 480–519, 560–579, 740–759, 800–819 и 900–959 нм (в остальных моделях для общего белка использовался оптимизированный полный интервал 420–1070 нм).

ППП моделирует ситуацию гипотетического нового образца, который композиционно схож с образцами из набора данных, но отличается распределением глобул жира по размерам, и, следовательно, числом рассеивающих частиц. ППС, где сегментом является один образец в разных степенях

гомогенизации, имитирует новый образец, который отличается от представленных в обучающем наборе содержанием питательных веществ при наличии вариабельности размеров глобул жира. Проверка с помощью тестового набора (по дизайну из раздела 3.4.1) позволяет оценить устойчивость прогноза композиционно близких образцов молока к способу смешивания исходных стандартов (Приложение Б). Таким образом, достигается наиболее полная и достоверная статистическая картина, позволяющая не только оценить общую работоспособность предложенного подхода, но и сравнить различные способы моделирования.

Модели, основанные на полных непредобработанных спектральных данных, уже демонстрируют достаточно высокую точность анализа: СКО перекрёстной проверки сегментами равно 0.14% для жира и 0.04% для белка. После удаления выбросов погрешность снижается до 0.10% для жира, а для белка – незначительно. Использование методов МКР, СНВ или первой производной для корректировки рассеяния (раздел 1.3.2), снижает точность прогноза, что ожидаемо с учётом показанной ранее информативности этого спектрального эффекта в рассматриваемом подходе. Оптимальное число ЛП в моделях на полных исходных данных одинаково для всех методов проверки и составляет в модели по жиру – 5, а по белку – 4, что следует из графика на рис. 5-4. Это оправданная сложность модели, учитывающая факторы изменчивости в данных: содержания питательных веществ (жира и белка), гомогенизации и, возможно, других, неизвестных различий исходных образцов. Обычно ПЛС-модели для определения жира и белка из оптических спектров поглощения получаются сложнее: 8–10 ЛП для БИК-спектроскопии [220-222] и 7–9 ЛП для средней ИК-области [236], где спектры содержат полосы поглощения различных компонентов молока. В отличие от этих методов, Вид/КВБИК-спектроскопия гораздо менее чувствительна к компонентам, отличным от жира и белка, что и даёт наблюдаемое упрощение моделей.

Более глубокое понимание принципов работы модели достигается при углублённом рассмотрении ПЛС-счетов и нагрузок [85,95]. В модели ЛП1 и ЛП3 в большей степени важны для определения содержания жира, а ЛП2 и ЛП4 – белка

(рис. 5-4 б и г). На проекции t_1-t_3 (рис. 5-3 б) чётко прослеживается диагональный рост массовой доли жира в сочетании с чётким разделением образцов по степеням гомогенизации. Этот важный результат доказывает, что модель чувствительна как к различию массовой доли жира, так и к разнице размеров жировых глобул.

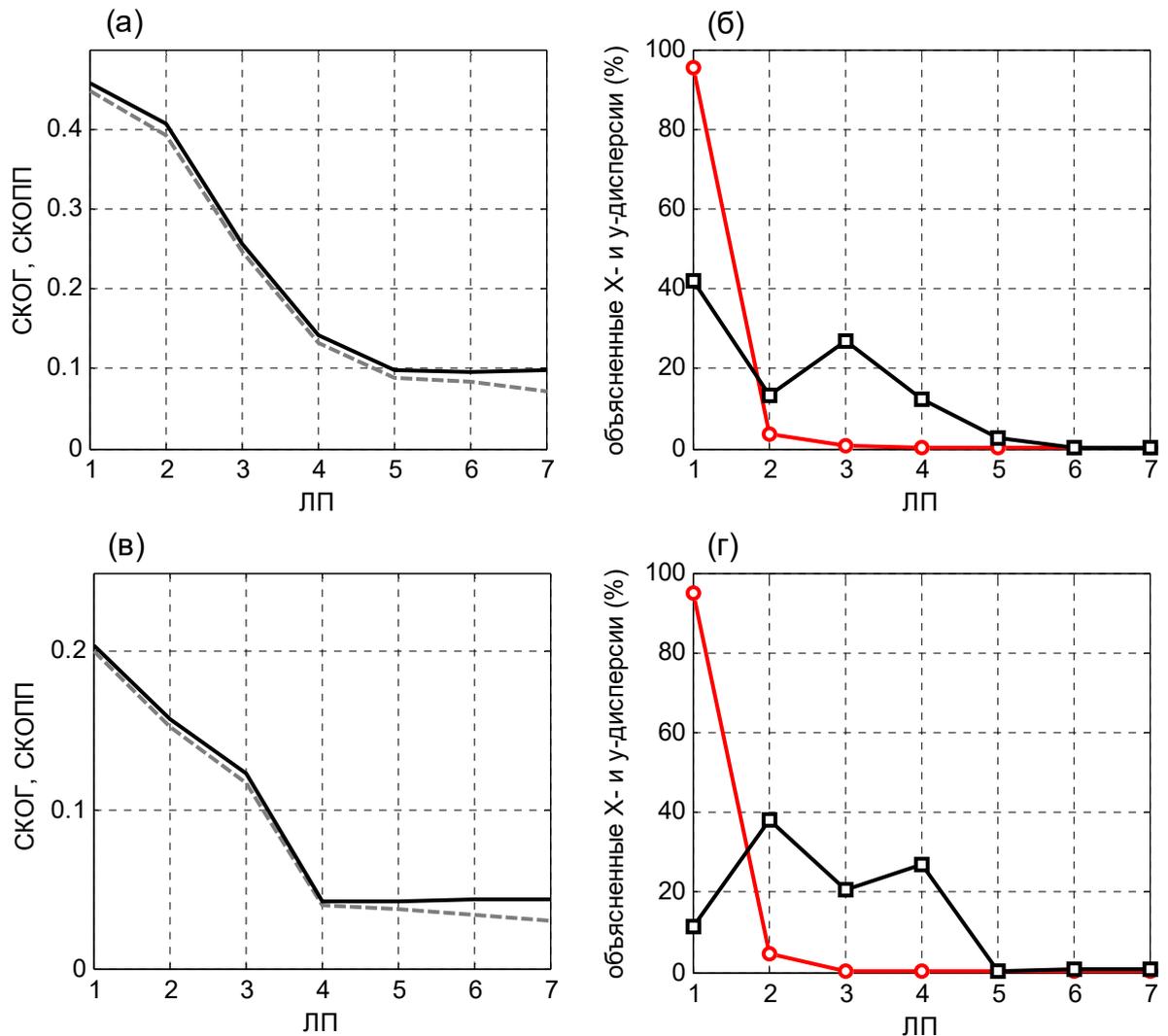


Рисунок 5-4. (а,в) Зависимости СКО градуировки ($SKOG$, серая пунктирная линия) и прогноза ($SKOPP$, черная линия); и (б,г) зависимости объяснённых X- (красная линия, круги) и y-дисперсий (черная линия, квадраты) от числа ЛП в моделях (а, б) F_{out7} и (в, г) P_{out1} [85].

Более простую модель по определению содержания белка, включающую четыре ЛП, можно объяснить тем, что в случае белка не нужно различать размеры частиц с тем же уровнем детализации, что и для жировых глобул. Рассмотрение форм графиков нагрузок (график и его интерпретация даны в работах [85,95]) позволяет предположить, что помимо эффектов рассеяния, поглощение света компонентами молока в Вид/КВБИК области также вносит некоторый вклад в

градуировочную модель.

Использованная в данном исследовании ультразвуковая модификация образцов призвана моделировать природную вариабельность размеров жировых глобул и является относительно слабой по сравнению с промышленной гомогенизацией, как в образцах «Товарной» модели. Исследование спектрального эффекта глубокой пошаговой гомогенизации молока [186] показало, что при снижении размеров основной массы жировых глобул до 1000 нм и менее (то есть до сравнимого с длиной волны диаметра) в Вид/КВБИК-спектрах происходит резкое падение интенсивности длинноволнового конца спектра – на одну единицу оптической плотности и более. Этот эффект связан с изменением закона рассеяния при измельчении жировых глобул, что согласуется с теорией [237].

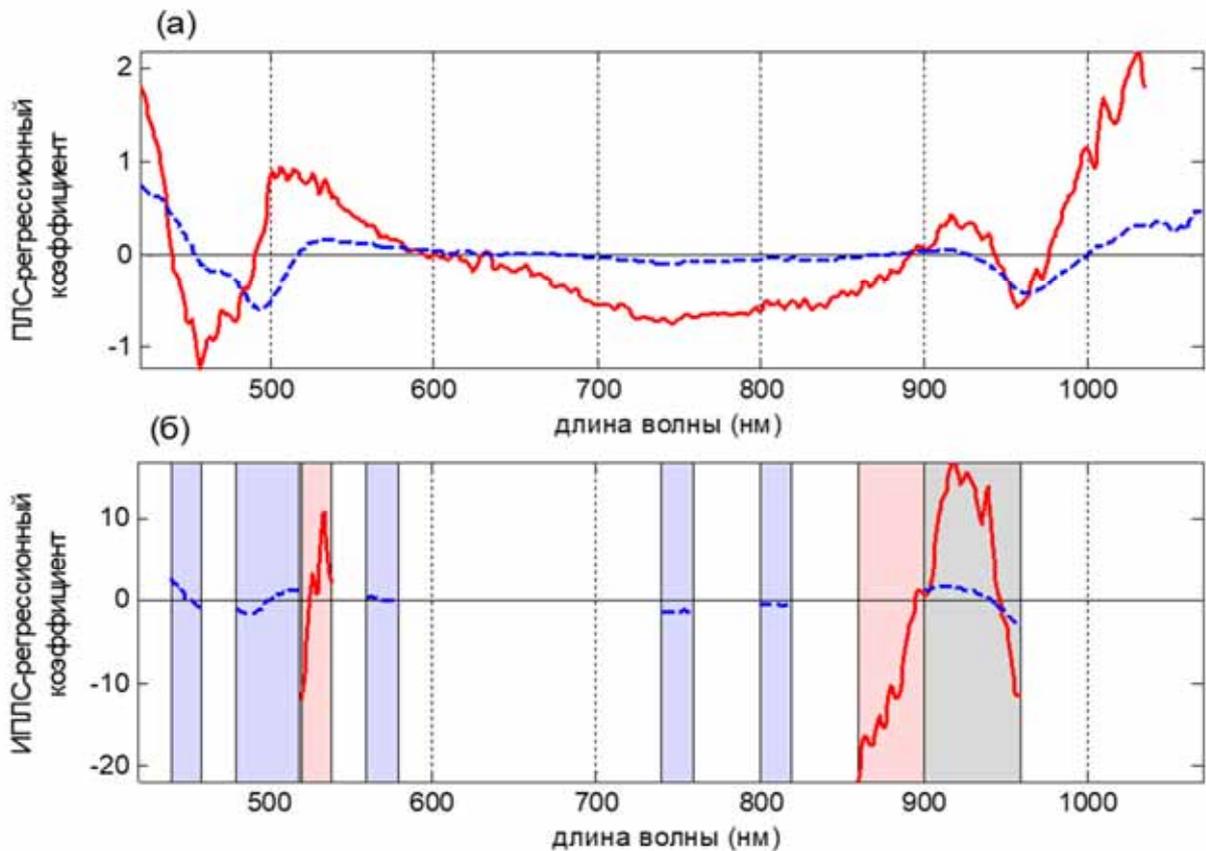


Рисунок 5-5. Векторы регрессионных коэффициентов для градуировочных (а) ПЛС-моделей F_{out7} и P_{out1} и (б) ИПЛС-моделей F_{out7i} и P_{out1i} для определения жира (красная линия) и общего белка (синяя пунктирная линия). Выбранные методом ИПЛС оптимальные интервалы моделей обозначены на рис. (б) цветом заливки: красная – для жира, синяя – для белка, серая – для жира и белка одновременно [85].

Дальнейшее улучшение моделей может быть достигнуто за счёт отбора переменных методом ИПЛС, позволяющего выбрать наиболее информативные для

определения жира и белка спектральные интервалы. На рис. 5-5 представлены векторы регрессионных коэффициентов ПЛС- и ИПЛС-моделей. Выбранные интервалы в целом не совпадают с областями наибольших значений регрессионных коэффициентов для полного спектра, хотя графики ИПЛС-регрессионных коэффициентов сохраняют форму на соответствующих участках (рис. 5-5 а). В то время как для определения содержания жира необходимо лишь два интервала – узкий в видимой области (520–539 нм) и более широкий в БИК области (900–959 нм), ИПЛС-модель для белка использует 6 интервалов в более широком диапазоне длин волн. В обоих случаях общее число используемых для анализа переменных было значительно снижено, причём с выигрышем в точности прогнозирования (табл. 5-3 и рис. 5-6). Особенно ощутимо снижается погрешность при определении жира – до 0.08% по результатам ППС (табл. 5-3). Рассмотрение результатов ИПЛС представляет интерес с точки зрения дальнейшей разработки ОМС на основе полноспектрального метода.

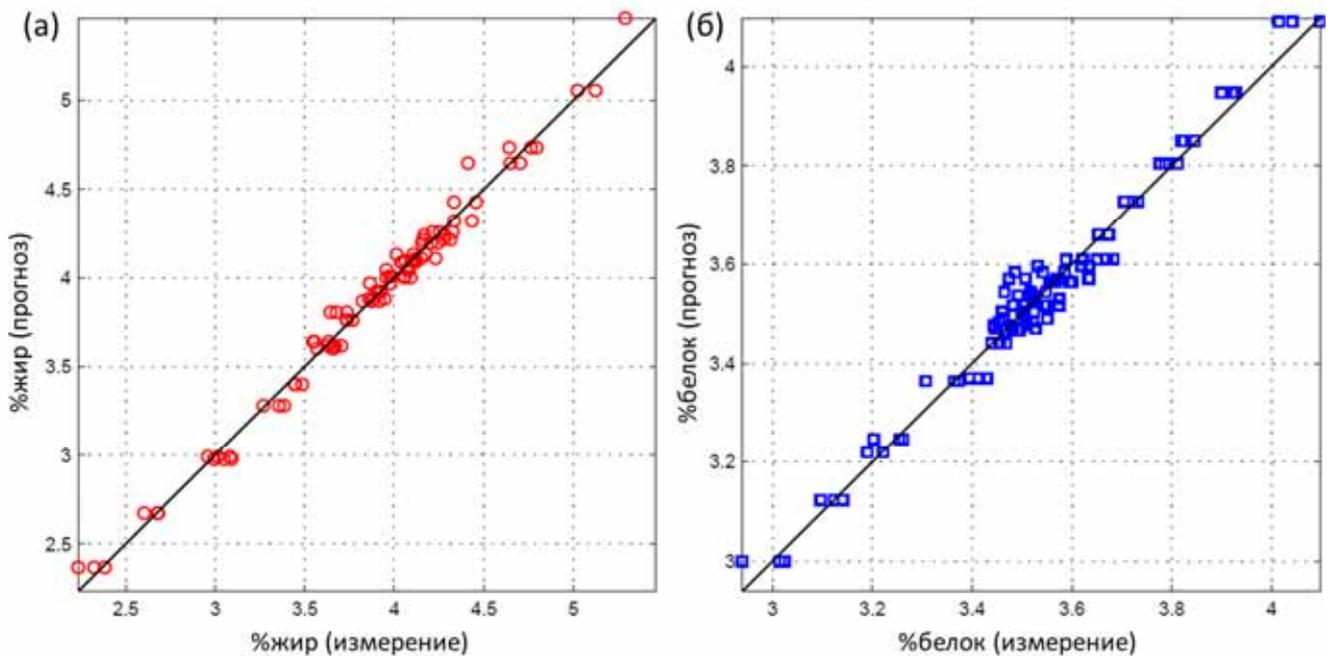


Рисунок 5-6. График «введено-найдено» по результатам ППП градуировочных ПЛС-моделей: (а) F_{out7i} по жиру и (б) P_{out1i} по общему белку (табл. 5-3) [85].

Построение и углублённая интерпретация моделей «Лабораторного» дизайна (раздел 3.4.1) показали, что Вид/КВБИК-спектроскопия в комбинации с методами регрессионного анализа данных позволяет точно определять содержание жира и белка в молоке при значительной, даже превышающей природную, вариабельности

размеров коллоидных частиц. Точность определения жира и белка при этом не уступает методам ИК- и БИК-спектроскопии при значительно меньшей сложности моделей.

5.1.3 Построение рабочих моделей

Для успешного практического применения предложенного подхода к спектральному анализу молока необходимо построить глобальную (раздел 1.3.8) многомерную модель, обученную на представительном наборе образцов, которая будет устойчива к возможным колебаниям его химических и физических свойств, обусловленных сезонностью, генетическими и географическими особенностями, условиями содержания и кормления животных и т. д. В построении таких глобальных моделей для определения массовой доли жира и общего белка были использованы образцы натурального молока, собранные в течение длительного периода времени. Описание дизайна «Глобальной» модели дано в разделе 3.4.1 и в работе [182].

Таблица 5-4. Статистика градуировки и валидации ПЛС-моделей на содержание жира и белка для измерений по дизайну «Глобальной» модели.

Набор данных	Метод	ЛП ^а	Градуировка		ППС ^б		ПТН	
			СКОГ	R ²	СКОПП	R ²	СКОП	R ²
F ₁₀₀₆ ^в	ПЛС ^д	5	0.163	0.882	0.172	0.867	0.165	0.881
	ПЛС+инт ^е		0.093	0.961	0.096	0.959	0.091	0.961
P ₁₀₀₀ ^г	ПЛС	6	0.115	0.691	0.125	0.636	0.114	0.682
	ПЛС+инт ^ж		0.093	0.797	0.102	0.758	0.097	0.805

Примечания: ^а число ЛП; ^б перекрёстная проверка сегментами с десятью смежными блоками; ^в полный набор данных для модели по жиру; ^г полный набор данных для модели по белку; ^д ПЛС-модель на полном наборе спектральных данных; ^е ПЛС-модель по жиру с предобработкой (первая производная по Савицкому-Голею, ширина окна 19, порядок полинома 2) и с отобранными интервалами длин волн (нм): 619–649, 707–737, 720–750, 761–791, 792–822, 837–867, 874–904, 886–916, 901–931, 925–955 с усреднением интенсивностей внутри интервалов; ^ж ПЛС-модель по белку с предобработкой (первая производная по Савицкому-Голею, ширина фильтра 11, порядок полинома 2) и с отобранными интервалами длин волн (нм): 556–582, 604–663, 739–765, 796–829, 860–933, без усреднения.

Результаты градуировки методом ПЛС-регрессии приведены в табл. 5-4. Несмотря на большое разнообразие образцов, полученные погрешности прогноза содержания жира и общего белка в молоке сравнимы с приведёнными ранее

результатами «Лабораторной» модели (табл. 5-3), хотя и несколько уступают им в точности. Учитывая большое количество факторов, потенциально влияющих на состав молока, построенные модели характеризуются относительной простотой: 5 и 6 ЛП для жира и белка соответственно.

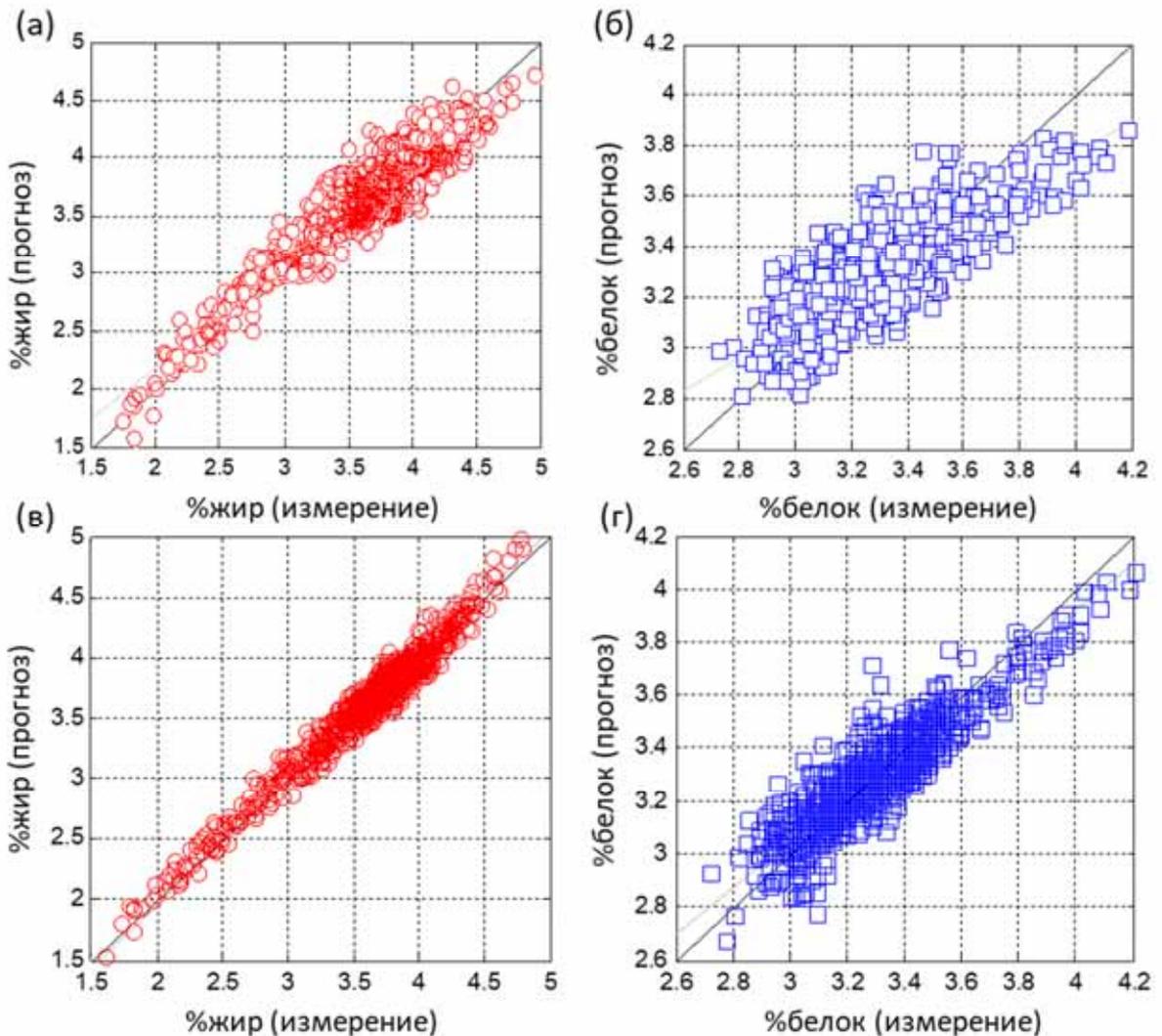


Рисунок 5-7. Графики «введено-найдено» для «Глобальных» градуировочных ПЛС-моделей (а,в) для жира и (б,г) для общего белка (а,б) до и (в,г) после оптимизации: (а) F_{1006} ; (б) P_{1000} ; (в) F_{1006i} ; и (г) P_{1000i} (езде – по результатам ПП, табл. 5-4) [182].

«Глобальная» модель для определения содержания белка имеет на две ЛП больше, чем «Лабораторная» (табл. 5-3). Снижение точности определения белка может быть связано с нелинейностью данных: систематическое занижение прогнозируемых значений при высоком содержании белка видно из рис. 5-7 б. Этот эффект может быть объяснён с точки зрения существующего равновесия между разными формами белка в молоке [200], поскольку доля растворённого белка примерно постоянна и находится в пределах 0.5–0.9%, а чувствительность к нему

Вид/КВБИК-спектрального анализа существенно ниже, чем к рассеивающим мицеллам казеина. Другими словами, метод избирательно обнаруживает мицеллярный казеин, в то время как ИК-спектроскопия, использованная для референтного анализа, определяет суммарное содержание белков (общий белок). Повышение доли растворённого белка может ожидать с увеличением концентрации общего белка в молоке, что и объясняет наблюдаемое нарушение линейности.

Полученные модели были улучшены применением метода интервального отбора переменных с одновременной оптимизацией предобработки данных [172] (раздел 2.2.4). В результате проведённой оптимизации моделей удалось снизить погрешности прогноза (СКОП) «Глобальных» моделей до 0.091% по жиру и 0.097% по общему белку (табл. 5-4 и рис. 5-7 в, г). Найденные оптимальные интервалы (см. примечания к табл. 5-4) существенно перекрываются с интервалами, полученными методом ИПЛС-регрессии (табл. 5-3 и рис. 5-5 б). Общие интервалы: 620–680, 720–780, 840–960 нм для жира и 550–750 и 800–950 нм для белка являются, по всей видимости, наиболее информативными для моделирования.

Показанная данными расчётами возможность перехода от полного спектра к оптимальным интервалам, а значит и к ОМС, без существенных потерь в точности прогноза мотивирует дальнейшую разработку мультисенсорного анализатора молока. Статистически обоснованная представительность обучающего набора образцов «Глобальных» моделей для определения содержания жира и общего белка [95,182] позволяет ожидать высокой устойчивости и надёжности полученных моделей в практическом использовании.

5.1.4 Перенос модели методом коррекции наклона и отсечения

Одной из отличительных черт ОМС является возможное отсутствие аппаратной стандартизации, то есть спектры одного и того же образца, снятые на различных анализаторах, могут заметно отличаться. Поэтому при серийном производстве ОМС необходимо осуществлять перенос модели с одного анализатора на другой. В частности, такое отличие было показано для двух ДМД-

спектрометров серии «TIDAS E» («J&M Analytik AG», г. Эссинген, ФРГ) [182]. В этом случае для переноса модели был успешно применён метод коррекции наклона и отсечения (раздел 1.3.8).

На рис. 5-8 показаны графики зависимостей «введено-найдено» для исходного и скорректированного прогноза содержания жира по «Лабораторной» модели (раздел 3.4.1), построенной на спектрах первичного прибора. Видно, что прогноз на спектрах вторичного прибора сдвинут вверх по отношению к диагонали первого квадранта (идеальному прогнозу). То, что спрогнозированные значения ложатся на прямую (через точки построена регрессионная линия), позволяет использовать метод коррекции наклона и отсечения.

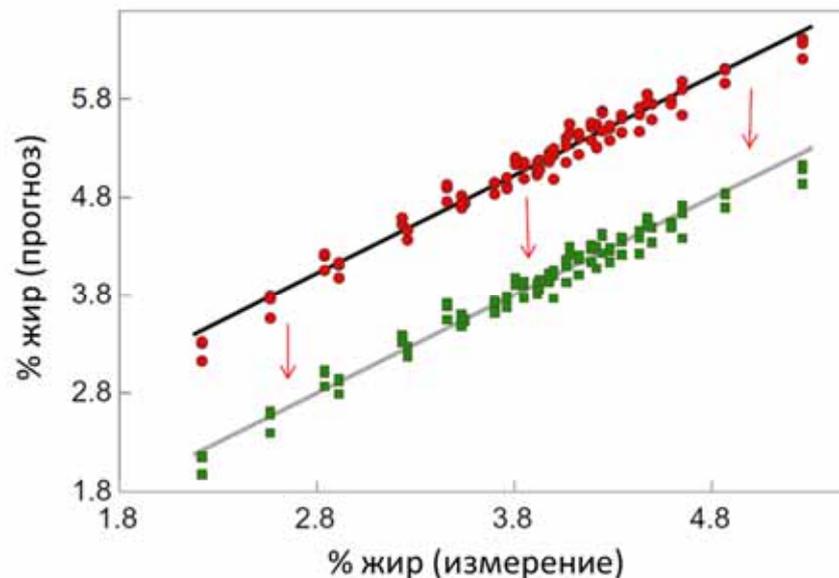


Рисунок 5-8. Перенос модели между молочными анализаторами «TIDAS E» с использованием образцов «Лабораторной» модели: зависимости «введено-найдено» значений прогноза содержания жира на вторичном приборе до (красные круги) и после (зелёные квадраты) применения коррекции наклона и отсечения; чёрная линия – функция регрессии, серая линия – диагональ первого квадранта [182].

Для компенсации ошибки прогноза, связанной с наблюдаемыми спектральными различиями, из линейной регрессии на нескорректированных значениях для вторичного прибора было посчитано её отклонение (в терминах параметров наклона и отсечения) от диагонали первого квадранта по уравнению (1.8), а затем применена коррекция. *СКОГ* прогноза на данных вторичного прибора с применением вышеописанного переноса составило 0.115%, что лишь

незначительно уступает соответствующим результатам первичного прибора ($СКОГ = 0.104\%$). Сходные результаты были получены в случае переноса модели для общего белка между этими же приборами. Это убедительно доказывает применимость коррекции наклона и отсечения для переноса моделей между анализаторами данного типа в задачах количественного определения жира и общего белка в молоке предложенным методом.

Перенос модели различными методами (раздел 1.3.8) можно использовать не только между приборами одной серии, но и между различными анализаторами, например спектрофотометром и светодиодной ОМС [192]. Перенос особенно важен в случае дорогостоящей глобальной модели, которая должна быть адаптирована к новому анализатору или новым условиям анализа.

5.1.5 Разработка светодиодного анализатора

Для исследования принципиальной возможности создания светодиодной ОМС для анализа питательной ценности молока был поставлен следующий эксперимент. Образец объёмом 4 мл помещался в чашку Петри диаметром 30 мм и поочерёдно подсвечивался снизу тремя различными светодиодами через оптоволоконный кабель с волокном кварцевого стекла диаметром 1.2 мм. Изображения диффузных световых пятен на поверхности молока снимались закреплённой цифровой фотокамерой. Светодиоды имели максимумы эмиссии на 465 нм (синий), 526 нм (зелёный) и 630 нм (красный). Измерения проводились в затемнённом помещении. Данные образца для каждого светодиода были представлены серией фотографий при пяти временах экспозиции (брекетинг): 1, 1/2, 1/4, 1/8 и 1/15 с. Таким образом, каждый образец был представлен 15-ю изображениями. Сравнительный анализ данных представлен на рис. 5-9.

Трёхмерные профили (рельефы) световых пятен зависят как от состава образца, так и от времени гомогенизации, причём характер наблюдаемых попарных различий на каждой длине волны светодиода свой. Для зелёного (см. рельефы на рис. 5-9) отчётливо видно, что эффект гомогенизации (рис. 5-9 в) преобладает над эффектами композиционных различий образцов (рис. 5-9 а и б). В целом характер наблюдаемых различий свидетельствует о том, что изображения несут в себе

количественную информацию, необходимую для построения градуировочных моделей для жира и белка. Более подробное рассмотрение наблюдаемых эффектов дано в работе [90].

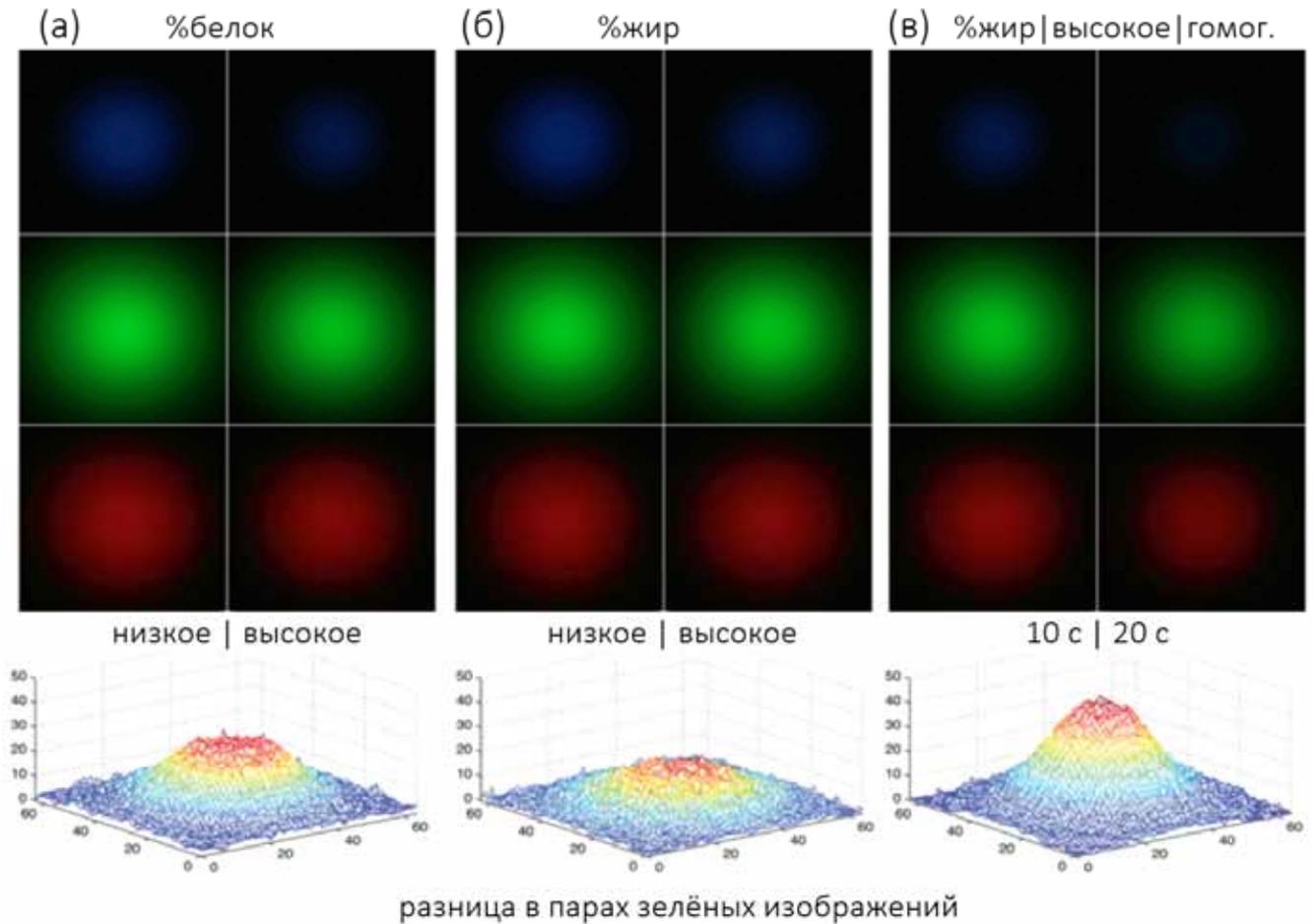


Рисунок 5-9. Сравнение пар изображений (время экспозиции 1/4 с) образцов с различным содержанием жира и белка, а также степенью гомогенизации: (а) *p11h0* и *p16h0* – низкое и высокое содержание белка (нумерация образцов описана в разделе 3.4.1); (б) *f11h0* и *f16h0* – низкое и высокое содержание жира; и (в) *f16h1* и *f16h2* – образец с высоким содержанием жира после 10 и 20 с гомогенизации. В нижней части рис. показаны рельефы разницы в соответствующих парах зелёных изображений [90].

Для построения градуировочных моделей данные каждого измерения должны быть преобразованы в вектор чисел – признаков изображения. Были опробованы различные методы обработки и сбора признаков изображений. Наилучшие результаты в моделировании показал наш собственный метод концентрических колец, по которому каждое изображение описывается интегралами интенсивностей 48 колец, отсчитываемых от центра (максимума яркости). Векторы интенсивностей всех 15 изображений одного образца для анализа объединяются, используя принцип анализа дополненных данных

(раздел 2.2.1). Метод концентрических колец обладает преимуществом перед другими подходами, поскольку наилучшим образом учитывает пространственное распределение яркости рассеянного образцом света. С деталями анализа полученных изображений можно ознакомиться в работе [90].

Градуировочные модели по жиру и белку были получены методом ПЛС-регрессии на признаках изображения, отобранных методом концентрических колец, к которым был затем применён метод отбора переменных КАПВ (раздел 1.3.3). Применение отбора переменных на объединённом векторе признаков даёт значительное улучшение модели, что ожидаемо, учитывая неодинаковую «ценность» различных участков изображений, их подверженность шуму, искажениям, например, насыщением пикселей при высоких яркостях. Оптимизированные градуировочные модели «Лабораторной» серии дают следующие погрешности прогноза (*СКОПП*): 0.215 с 4-мя ЛП для жира и 0.122 с 5-ю ЛП для белка. Для проверки моделей использовалась ППС с 4-мя случайно выбранными сегментами, повторённая 8 раз для сходимости усреднённого результата. Точность определения компонентов молока заметно ниже, чем в полноспектральном методе с использованием ДМД-анализатора (раздел 5.1.2), но этот результат был ожидаем, учитывая значительно меньший объём используемой спектральной информации – трёх не оптимизированных широкополосных оптических каналов на интервале длин волн шириной менее 200 нм вместо сотен переменных в полном Вид/КВБИК-спектре. В то же время некоторый выигрыш был получен за счёт использования пространственного разрешения, присущего изображениям. Полученные погрешности измерений всё ещё приемлемы для практического применения в экспресс-анализе молока.

Основным результатом исследования являлось экспериментальное обоснование возможности определения содержания жира и белка в натуральном молоке при помощи светодиодной ОМС. Использование оригинального метода детектирования при помощи цифровой фотокамеры может иметь практический потенциал. Анализ полученных изображений показал значимость пространственной информации для повышения точности градуировочной модели.

В качестве следующего шага разработки светодиодной ОМС для определения жира и белка в молоке были рассчитаны необходимое число и свойства оптических каналов системы. Расчёт проводился по методике [172], описанной в разделе 2.2.2. При этом использовались светодиоды из собранной ранее базы данных [66]. Была произведена оптимизация трёх возможных конфигураций сенсора: для определения либо жира, либо белка отдельно и для определения обоих компонентов одним измерением (одним набором светодиодов). Наилучшие конфигурации были получены на отдельных моделях (табл. 5-5). Все виды проверок, проведённых в соответствии с принципом многоуровневой валидации (раздел 2.4.3), дали близкие результаты (в обсуждении используются данные ППП). Для модели по жиру *СКОПП* составило 0.068%, а для белка – 0.050%. Однако предпочтительным с точки зрения простоты конструкции и практичности является сенсор, количественно определяющий жир и белок при помощи одного и того же набора светодиодов. Расчёт такой сенсорной системы для семи светодиодов и соответствующие статистики моделирования представлены в табл. 5-5 и на рис. 5-10. При этом результирующие точности расчётной универсальной ОМС всё ещё более чем приемлемы с практической точки зрения.

Существенное снижение погрешности определения жира для всех расчётных конфигураций ОМС по сравнению с полноспектральной моделью может показаться неожиданным, учитывая значительное снижение общей информативности данных – уменьшение числа переменных до 7-ми с существенным закруглением спектрального разрешения. Однако это упрощение имеет и позитивный эффект, заключающийся в отборе наиболее выгодных для градуировки переменных. Примечательно, что в светодиодной ОМС отбор переменных осуществляется на физическом уровне, ещё на этапе сбора данных, в отличие от классического подхода к предобработке спектров (раздел 1.3.3). По-видимому, позитивный эффект отбора переменных в данном случае перевешивает потери от пониженного спектрального разрешения, что вполне логично, учитывая низкую селективность спектральных откликов, на которых строится модель. Погрешность определения белка при этом остаётся практически неизменной.

Таблица 5-5. Статистика градуировки по жиру и общему белку («Лабораторная» модель, раздел 3.4.1) методом ПЛС-регрессии для полноспектральных данных и для расчётной светодиодной ОМС на оптимизированных спектральных интервалах.

Набор данных	СД ^а	ЛП ^б	Градуировка		ППП		ППС ^в		ПТН	
			СКОГ	R ²	СКОПП	R ²	СКОПП	R ²	СКОПП	R ²
<i>полноспектральные модели</i>										
Жир	-	5	0.104	0.975	0.116	0.969	0.125	0.964	0.111	0.975
Белок	-	4	0.052	0.950	0.055	0.944	0.059	0.935	0.058	0.938
<i>модели на оптимальных интервалах^г</i>										
Жир ^д	7	5	0.063	0.991	0.068	0.990	0.068	0.989	0.067	0.987
Белок ^е	7	4	0.048	0.957	0.050	0.953	0.053	0.946	0.055	0.932
<i>модели на общих оптимальных интервалах</i>										
Жир ^ж	7	5	0.088	0.982	0.095	0.979	0.099	0.977	0.090	0.977
Белок	7	4	0.051	0.950	0.054	0.944	0.059	0.934	0.054	0.935

Примечания: ^а число светодиодов; ^б число латентных переменных в модели; ^в в качестве сегментов (три измерения в каждом) использовались отдельные образцы во всех степенях гомогенизации; ^г спектры испускания светодиодов использовались как весовые функции переменных на соответствующих интервалах (ширины спектров светодиодов на полувысоте составляли 19–196 нм; максимумы полос испускания выбранных светодиодов: ^д 681, 761, 770, 853, 850, 935, 950 нм, ^е 400, 450, 463, 530, 625, 761, 820 нм и ^ж 400, 430, 450, 507, 831, 850, 905 нм.

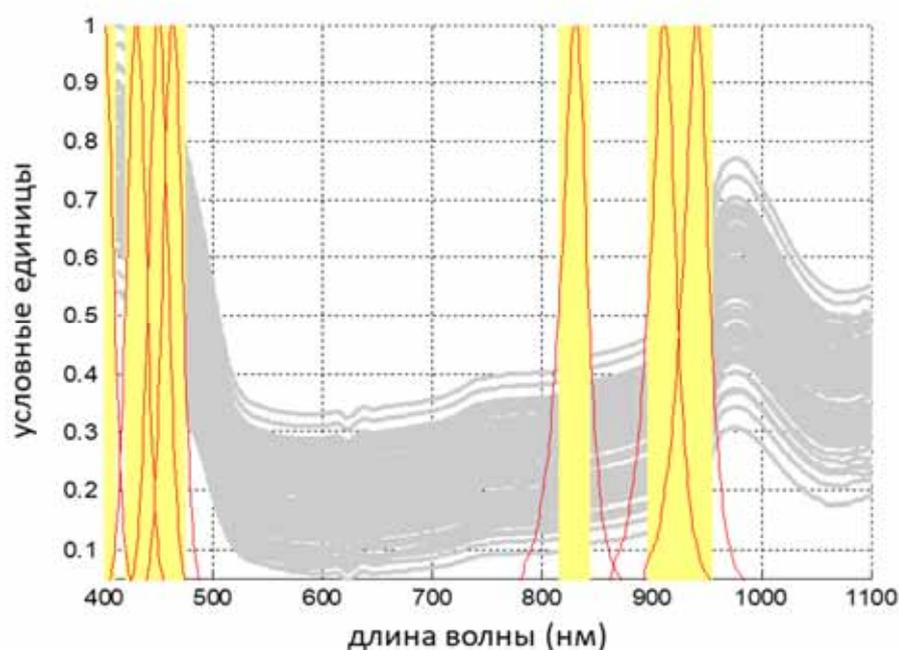


Рисунок 5-10. Оптимальная конфигурация сенсора для одновременного определения жира и общего белка в молоке. Поверх нормированных спектральных данных «Лабораторной» серии (серые линии) изображены спектры испускания светодиодов оптимального набора и соответствующие им интервалы (желтая заливка), определённые как полная ширина на половине высоты [235].

Таким образом, переход к недорогой портативной ОМС для анализа молока

данным методом возможен даже без потерь точности определения жира и белка по сравнению с полноценным спектральным анализатором. В результате проведённых расчётов были выбраны коммерчески доступные светодиоды для использования в прототипе ОМС.

5.1.6 Пространственно-разрешённая спектроскопия и безэталонный метод измерения в анализе молока

Возможность повышения точности определения жира и белка в натуральном молоке за счёт использования пространственного разрешения – дополнительных данных, получаемых при регистрации спектров диффузного отражения в разных позициях по отношению к источнику – была изучена при помощи установки (рис. 2-2), описанной в разделе 2.3.5 и в работе [89]. Один из каналов при этом был использован в качестве внутреннего стандарта (безэталонный анализ, раздел 2.3.5). В эксперименте использовался дизайн «Лабораторной» модели (раздел 3.4.1). Полученные данные представлены на рис. 5-11.

Использование внутреннего эталонного канала в пространственно-разрешённом анализе молока ранее не проводилось. Спектр сравнения, получаемый на том же самом образце, успешно решает две основные задачи эталона для количественного анализа: 1) он служит для нормировки спектров, представляя их в относительных единицах интенсивности, более подходящих для количественного анализа и 2) компенсирует возможную нестабильность источника света (см. раздел 2.3.4). Настоящее исследование является первой экспериментальной апробацией запатентованного нами безэталонного метода и устройства для пространственно-разрешённого анализа рассеивающих сред, в том числе молока [79].

Форма и интенсивность спектра зависит от расстояния до канала освещения. Разница в форме особенно сильна между ближними каналами C_1 , C_2 и C_3 . Разница между минимальной и максимальной интенсивностью серий спектров на отдельных каналах возрастает от C_2 к C_6 (рис. 5-11 б-е). Канал C_1 (рис. 5-11 а), занимающий особое положение между каналом освещения и эталонным (рис. 2-2), показал наибольшие различия между самым высоким и самым низким значениями

оптической плотности на отдельных длинах волн в области от 900 нм.

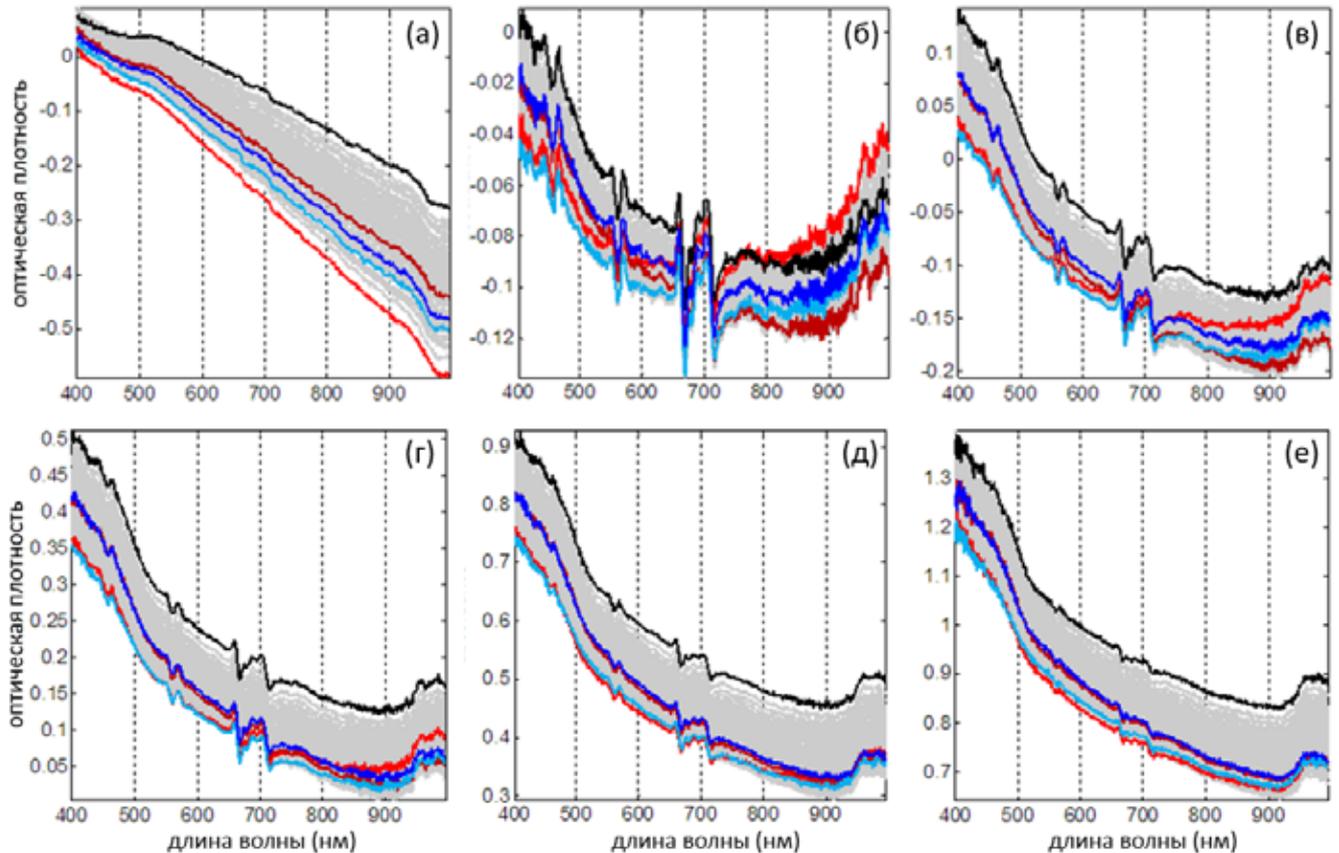


Рисунок 5-11. Спектры образцов молока «Лабораторной» серии на разных пространственных каналах: (а) С₁; (б) С₂; (в) С₃; (г) С₄; (д) С₅; и (е) С₆. Спектры следующих образцов с максимальным/ минимальным содержанием жира и белка и разной степенью гомогенизации выделены цветом: *f11h0* (красный), *f16h0* (коричневый), *f16h2* (черный), *p11h0* (голубой) и *p16h0* (синий); остальные спектры показаны серым цветом [89].

Спектральные формы следует интерпретировать в терминах сложной суперпозиции поглощения и рассеяния света; при этом их вклады в полный спектральный сигнал не являются независимыми. Так, рассеяние света приводит к увеличению пути, пройденному им в образце и, следовательно, к более высокому поглощению. Использование одного из каналов зонда (R на рис. 2-2) в качестве внутреннего эталона выявляет различия между каналами, то есть усиливает вклад пространственной информации в спектрах. Существенная компенсация рассеяния в нормированных таким образом спектрах подчёркивает поглощение, особенно в непосредственной близости от канала сравнения R, где обусловленные рассеянием различия минимальны.

Наиболее примечательным является поведение двух спектров

негомогенизированного молока с наибольшей разницей в содержании жира (образцы *f11h0* и *f16h0*, рис. 3-18 и табл. Б-1 в Приложении Б). Длинноволновый конец спектра образца *f11h0* меняет своё положение от самого низкого (C_1) к самому высокому (C_2) и обратно к самому низкому в канале C_6 . В целом гомогенизация образца, которая резко увеличивает количество рассеивающих частиц в образце, оказывает наиболее сильное влияние на спектры, тем самым определяя диапазон значений поглощения в каждом отдельном канале. Гомогенизированный образец *f16h2* с самым высоким содержанием жира в основном сохраняет своё верхнее положение. Исключением является канал C_2 , где начиная с 800 нм интенсивность спектра становится ниже, чем для того же образца в сыром виде *f16h0*. Это – важное наблюдение, показывающее сильное влияние геометрии измерения на спектры химически идентичных образцов. Влияние содержания белка в спектрах менее выражено и в основном проявляется, как и при измерениях в режиме пропускания (раздел 5.1.2), в наклоне спектра (сравните синие и голубые кривые на рис. 5-11), и пространственное разрешение не кажется столь же эффективным, как в случае с жиром. Более детальная интерпретация спектральных эффектов дана в работе [89]. В целом наблюдаемые спектральные изменения указывают на способность пространственно-разрешённого измерения различать конкурирующие эффекты размеров частиц и их объёмных плотностей при распространении света в среде образца, обеспечивая предпосылки для их количественного определения. В то же время наблюдаемые явления сложны, и анализ данных представляет собой нетривиальную задачу.

Сравнение градуировочных ПЛС-моделей для жира и общего белка, полученных с использованием различных методов предобработки данных, дано в табл. 5-6. Целью углублённого анализа данных было не только получение наилучших моделей прогноза компонентов, но и более глубокое понимание данных и работы самого аналитического метода.

Исходные трёхмодальные данные сначала преобразовывались методом развёртывания (раздел 1.1.3) в матрицу с общими объектами и дополненными спектрами (раздел 2.2.1). Предварительно было показано, что использование более

сложного трёхмодального алгоритма ПЛС-регрессии [238], проводящего прямую градуировку на таких массивах данных, даёт заметно менее точный прогноз как для жира, так и для белка. Поэтому в дальнейшем этот метод не использовался.

Таблица 5-6. Статистика градуировки методом ПЛС-регрессии («Лабораторная» модель) для пространственно-разрешённого анализа молока на содержание жира и белка.

Набор данных	Пер ^а	ЛП ^б	Градуировка		ППП		ППС ^в		ПТН ^г	
			СКОГ	R ²	СКОПП	R ²	СКОПП	R ²	СКОП	R ²
<i>Полные данные^д</i>										
F	3576	6	0.134	0.954	0.171	0.925	0.196	0.902	0.142	0.938
P	3576	6	0.059	0.922	0.077	0.869	0.078	0.865	0.080	0.835
<i>Лучшие отдельные каналы</i>										
F_C ₂	596	5	0.053	0.993	0.177	0.920	0.191	0.907	0.165	0.917
P_C ₅	596	5	0.054	0.935	0.102	0.770	0.107	0.745	0.108	0.701
<i>Лучшие комбинации каналов</i>										
F_C ₂₃₄	1788	5	0.103	0.973	0.134	0.954	0.145	0.946	0.111	0.963
P_C ₁₄₅₆	2384	5	0.073	0.882	0.086	0.837	0.087	0.833	0.085	0.816
<i>Интервальная оптимизация^е</i>										
F_инт ^ж	506	5	0.081	0.983	0.097	0.976	0.099	0.975	0.090	0.975
P_инт ^з	1046	5	0.061	0.917	0.071	0.888	0.072	0.884	0.079	0.840
<i>Моделирование светодиодной ОМС^и</i>										
F_СД ^к	9	5	0.094	0.978	0.102	0.973	0.104	0.973	0.091	0.975
P_СД ^л	7	5	0.065	0.905	0.071	0.887	0.073	0.882	0.072	0.866

Примечания: ^а число переменных в обучающем наборе данных; ^б число латентных переменных в модели; ^в перекрёстная проверка сегментами, образованными отдельными образцами во всех степенях гомогенизации (по три измерения в каждом); ^г спланированный тестовый набор (рис. 3-18); ^д полный набор дополненных данных, полученных методом развёртывания с общими объектами; ^е улучшенные модели по методу совместной оптимизации интервалов и предобработки (раздел 2.2.4) на дополненных данных набора F или P; ^ж оптимизированная модель для жира: сглаживание по Савицкому-Голею (ширина окна 11, степень полинома 2) и выбранные оптимальные интервалы (нм): 640–722 (C₂); 541–623, 898–988 (C₃); 542–624 (C₄); 677–759, 839–921 (C₅), без усреднения интенсивностей внутри интервалов; ^з оптимизированная модель для общего белка: сглаживание по Савицкому-Голею (ширина окна 5, степень полинома 2) и выбранные оптимальные интервалы (нм): 685–892 (C₁), 510–779, 786–995 (C₄), 400–535 (C₅), 435–656 (C₆), без усреднения интенсивностей внутри интервалов; ^и спектры испускания светодиодов моделировались функцией Гаусса с учётом уширения [66] и использовались как весовые функции переменных на соответствующих интервалах, оптимизируемых на дополненных данных набора F или P; ^к оптимизированный светодиодный сенсор для жира (максимум/ширина на полувысоте спектра испускания, нм): 704/32, 973/54 (C₂); 597/24, 616/26, 941/50 (C₃); 516/20, 736/34, 852/42, 899/46 (C₅); ^л оптимизированный светодиодный сенсор для белка (максимум/ширина на полувысоте спектра испускания, нм): 489/18, 876/44 (C₁), 419/14, 697/30 (C₄), 490/18, 588/24 (C₅), 439/16 (C₆).

Сначала была исследована возможность повышения точности прогноза

путём поиска наиболее выгодных комбинаций каналов путём исчерпывающего перебора вариантов. Оказалось, что модели на отдельных каналах сильно различаются по точности прогноза. Наилучшие из них, показанные в табл. 5-6, дают результаты, сходные с полноспектральными моделями. Тот факт, что даже комбинации нескольких каналов могут привести к значительному повышению точности прогноза не только по сравнению с любыми одноканальными измерениями, но и с полноспектральной моделью, как в случае жира (табл. 5-6), убедительно доказывает преимущество пространственного разрешения. В целом эффект достигается сочетанием от двух до 4-х отдельных каналов; при этом наилучшая комбинация не обязательно является сочетанием лучших отдельных каналов. Канал C_2 важен для градуировки содержания жира, что следует ожидать от поведения спектров на рис. 5-11. При определении общего белка ведущая роль принадлежит C_5 .

Аналогичным образом было показано, что пространственное разрешение само по себе (градуировка для данных, включающих сочетания пространственных каналов на единичных длинах волн) не может обеспечить необходимую точность анализа. То есть сочетание спектральной и пространственной информации даёт синергетический эффект в определении компонентов молока данным методом.

Дальнейшие усилия были направлены на улучшение модели с использованием метода совместной оптимизации набора переменных и предобработки данных (раздел 2.2.4). В настоящем исследовании алгоритм применялся вдоль объединённой спектральной оси дополненного набора данных. Преимущество такого подхода заключается в возможности выбора наиболее полезных интервалов длин волн для каждого пространственного канала в отдельности (рис. 5-12).

Интервальная оптимизация моделей (табл. 5-6) даёт наиболее существенный выигрыш в точности определения жира. В случае белка СКО валидации и прогноза ниже, чем при анализе на полном наборе дополненных данных, но незначительно. Обе погрешности несколько выше, чем при измерении на пропускание (табл. 5-3), но находятся при этом на уровне, позволяющем решать все основные практические

задачи. Качество прогноза по оптимизированным моделям можно оценить на рис. 5-13.

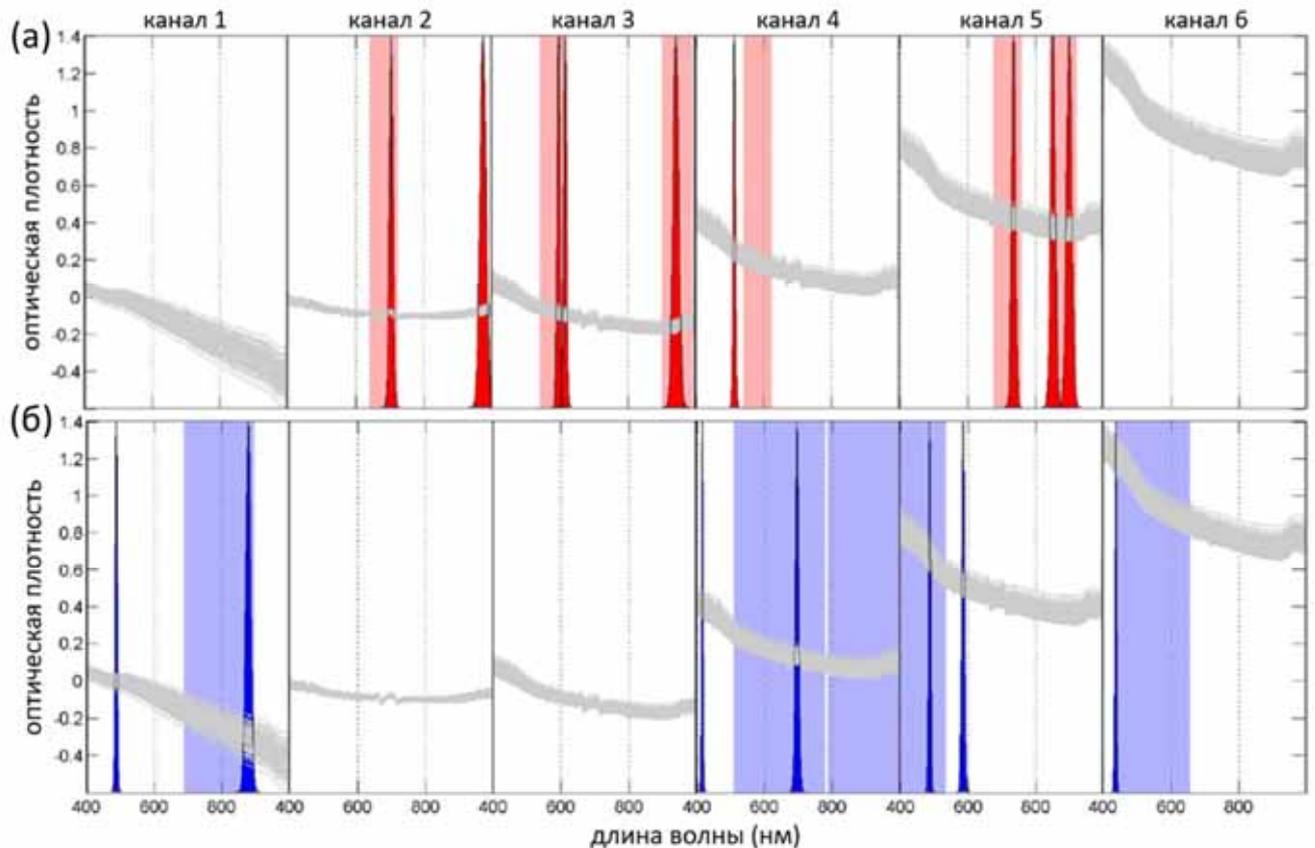


Рисунок 5-12. Интервальная оптимизация градуировочной модели и моделирование светодиодной ОМС на дополненных спектральных данных с пространственным разрешением для определения содержания: (а) жира и (б) общего белка; выделенные вертикальные области указывают выбранные интервалы (раздел «Интервальная оптимизация» табл. 5-6); оптимальные конфигурации модельных ОМС (раздел «Моделирование светодиодной ОМС» табл. 5-6) представлены закрасенными спектрами испускания выбранных оптимальных светодиодов; исходные спектральные данные показаны серыми линиями [89].

Последний раздел в табл. 5-6 («Моделирование светодиодной ОМС») посвящён расчёту светодиодной ОМС (раздел 1.2.6), техническая конструкция которой была запатентована [79]. Оптимальные длины волн определялись на основе анализа данных полного спектра, как описано в [172]. Спектры излучения светодиодов были смоделированы гауссовыми функциями с шириной пика, зависящей от длины волны [66]. Оптимальными были найдены конфигурации ОМС с 9-ю светодиодами для жира и 7-ю – для белка. Как и ожидалось, выбранные светодиоды обычно попадают в ранее определённые оптимальные интервалы длин волн (рис. 5-10). При этом погрешности прогноза для модельных ОМС практически

те же, что и в полностью оптимизированных моделях: 0.10% для жира и 0.07% для содержания белка при наиболее консервативном методе валидации. Отсутствие заметных потерь в точности особенно примечательно, учитывая чрезвычайно сильное загрубление спектрального разрешения в ОМС, где измерение сведено с нескольких тысяч (с учётом объединения) до нескольких переменных. Этот факт свидетельствует о положительном эффекте упрощения данных при переходе к ОМС, что согласуется с результатами раздела 5.1.5. Более подробно с материалами данного исследования можно ознакомиться в работе [89].

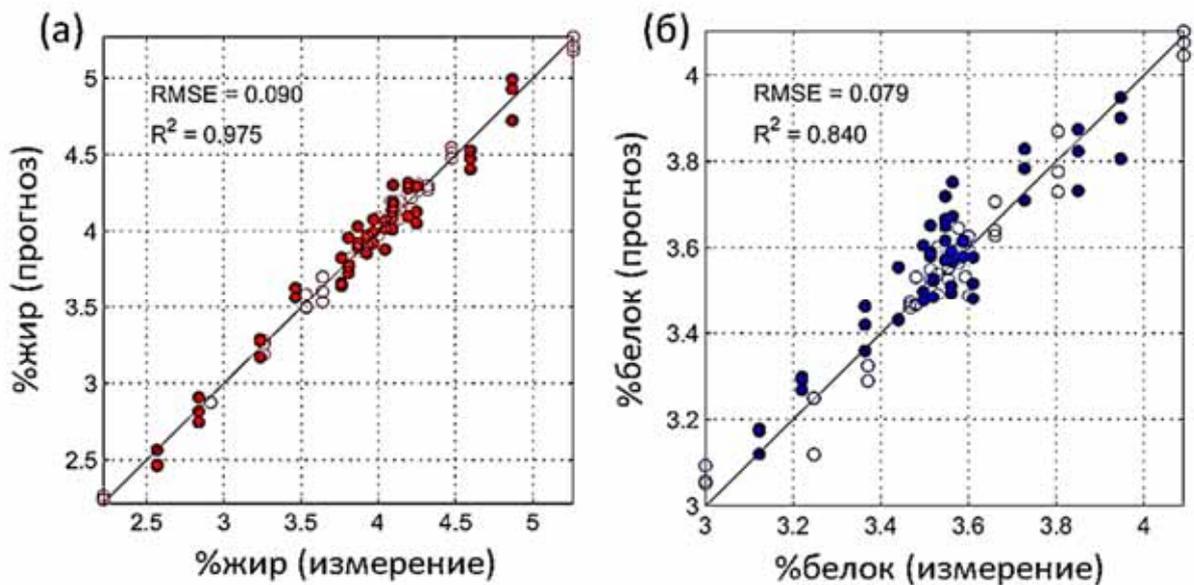


Рисунок 5-13. Графики «введено-найдено» для оптимизированных градуировочных ПЛС-моделей (раздел «Интервальная оптимизация» в табл. 5-6): (а) по жиру и (б) по общему белку; образцы обучающего и проверочного наборов отмечены пустыми и закрашенными символами соответственно [89].

В целом можно констатировать, что светодиодный анализатор жира и общего белка в натуральном молоке в выбранной спектральной области может быть реализован в виде погружного зонда в режиме диффузного отражения. При этом использование пространственно-разрешённого измерения может быть важным фактором достижения требуемой точности определения жира и общего белка. Использование одного из пространственных каналов в качестве внутреннего эталона упрощает метод и даёт дополнительные преимущества при моделировании. В отличие от анализа в режиме пропускания, измерения могут выполняться бесконтактно через стеклянное окно, что важно для онлайн-анализа

динамических сред, например для непрерывного мониторинга процессов. Простота конструкции плоского зонда обеспечивает простоту использования и ухода и идеально подходит для полевых измерений. Дальнейшее усовершенствование анализатора может быть достигнуто путём оптимизации геометрии измерения.

5.2 Контроль качества фармацевтических продуктов

5.2.1 Анализ процессов фармацевтических производств

Точное онлайн-определение воды в составе продукта или среды в ходе различных процессов представляет одну из наиболее актуальных задач современного АКТП. Распространённой технологией, требующей онлайн-контроля, являются процессы изготовления твёрдых лекарственных форм в так называемом «кипящем» или псевдоожиженном слое, когда частицы продукта взвешены в потоке восходящего воздуха. К таким процессам относятся: грануляция, покрытие пеллет и сушка (которая может быть самостоятельным процессом или стадией одного из вышеперечисленных процессов).

Одним из важных показателей качества конечных и промежуточных продуктов фармацевтических производств является массовая доля воды – влажность. Чаще всего, она измеряется в твёрдых формах, имеющих развитую поверхность, таких как порошки, гранулы и пеллеты. Отклонение содержания воды от нормы могут существенно повлиять на потребительские свойства продукта или пригодность сырья. Контроль содержания воды в среде протекающего процесса может быть необходим для своевременного определения точки окончания одной из его стадий (например сушки) или для поддержания оптимального режима протекания процесса, обеспечивающего достаточное качество конечного продукта.

Используемый в современной практике для определения массовой доли воды термогравиметрический анализ «у линии» занимает 20–30 мин и не обеспечивает достаточной оперативности контроля процессов.

Наиболее перспективным и быстро развивающимся оптическим методом онлайн-мониторинга концентрации воды является БИК-спектроскопия в сочетании

с измерением через встроенный в реактор оптоволоконный зонд [239]. Основным недостатком опубликованных по этой теме работ остаётся поверхностный и зачастую формальный подход к моделированию. Большинство представленных в литературе моделей построены на небольших, недостаточно представительных наборах образцов. Докладываемые авторами погрешности определения воды ненадёжны, и их разброс весьма велик.

Серьёзную помеху для оптического онлайн-анализа процессов в кипящем слое представляет турбулентность, приводящая к быстрому случайному изменению плотности окружающего зонд материала. Это, в свою очередь, даёт сильные флуктуации общей спектральной интенсивности (эффект рассеяния), преимущественно проявляющийся в скачкообразном смещении базовой линии, что усложняет построение градуировочных моделей по концентрации воды. Для компенсации этого эффекта принято использовать методы коррекции рассеяния: МКР, СНВ и производные первого или второго порядка (раздел 1.3.2). Такая предобработка данных считается стандартной и используется без исключения во всех известных публикациях. Метод коррекции в большинстве случаев выбирается эмпирически. В тоже время в работах [240-242] была отмечена зависимость спектральной базовой линии от содержания воды в продукте, что было экспериментально доказано Рантаненом и др. [240]. Использование методов коррекции рассеяния, наряду с эффектом турбулентности, попутно устраняет и эту «полезную» составляющую дисперсии данных. С учётом этого, необходимость применения методов коррекции рассеяния требует критического пересмотра.

Другим важным параметром, требующим онлайн-контроля, является толщина покрытия пеллет. Покрытые специальной полимерной оболочкой пеллеты чаще всего используются для замедленного высвобождения активного ингредиента в пищеварительном тракте пациента (так называемая целевая доставка лекарства). Таким образом, свойства защитной оболочки, а также условия процесса изготовления во многом определяют качество продукта.

Процесс изготовления пеллет включает стадии нанесения АФИ, нанесения покрытия и сушки. Общая масса (средняя толщина слоя) нанесённого покрытия,

содержание АФИ и содержание воды – три основных параметра, подлежащих мониторингу в ходе процесса изготовления пеллет в фармацевтической промышленности. Анализ траектории процесса в пространстве этих виртуальных (расчётных) переменных позволяет заблаговременно или на ранней стадии выявлять ситуации выхода процесса за рамки нормального протекания, а также определять точку окончания процесса или его отдельных стадий.

В современной производственной практике толщина нанесённого слоя обычно определяется по количеству распылённого материала, что может давать значительную систематическую ошибку, поскольку не учитывает неизбежные потери материала и истирание пеллет в ходе процесса. Поэтому прямой анализ свойств покрытия в реальном времени процесса представляет собой новую задачу. Проблему в данном случае представляет разработка не только онлайн- (например спектрального) метода измерения толщины покрытия, но и надёжного лабораторного метода или метода «у линии», способного заменить дорогостоящий и статистически недостоверный (из-за невозможности большого количества замеров) прямой метод измерения нанесённого слоя под микроскопом. Использование спектроскопии диффузного отражения для определения толщины нанесённого покрытия основано на достаточно большой глубине проникновения (более 1 мм) БИК-излучения в образец, а значит, способности отражать суммарный состав пеллет. При этом в ходе процесса в спектрах пеллет растёт спектральный сигнал полимерного покрытия и ослабевают сигналы ядра и внутреннего слоя АФИ [40].

Мониторинг отдельных параметров качества пеллет в ходе процесса безусловно важен, но не всегда достаточен. Решение поставленной перед современной фарминдустрией задачи ВРВ [10] требует, чтобы онлайн-мониторингу подлежали все характеристики, определяющие приёмочное качество конечного продукта. В случае покрытых пеллет, выпускаемых для целевой доставки АФИ – это так называемый профиль высвобождения (растворения), измеряемый в стандартном тесте на соответствующем лабораторном приборе. Его заблаговременное прогнозирование онлайн на основе получаемых в ходе процесса

оптических спектров представляется сверхсложной задачей.

Таким образом, несмотря на достигнутый прогресс в разработке оптических онлайн-анализаторов процессов в «кипящем» слое, существует ряд нерешённых задач, требующих дополнительных исследований на обширном экспериментальном материале.

5.2.2 Построение точной градуировочной модели для онлайн-мониторинга содержания воды в грануляте

25 пилотных загрузок процесса сушки фармацевтического гранулята были изучены при помощи ДМД-спектрометра (табл. 5-1), оснащённого зондом «LHP», встроенным в реактор.

Условия процесса в отдельных загрузках варьировались в соответствии с дизайном эксперимента, описанным в разделе 3.4.2. Полный набор данных составлял 16303 онлайн-спектра и референтные значения массовой доли воды для 301-го образца, отобранного из среды процесса и проанализированного термогравиметрическим методом (табл. 3-3). Детали эксперимента описаны в работе [86].

Набор из 1213-ти спектров, снятых в ходе одной из загрузок процесса (рис. 5-14), иллюстрирует основные черты анализируемых данных. Ожидаемое снижение интенсивности основной полосы поглощения воды около 1920–1940 нм сопровождается шумом – скачкообразными смещениями базовой линии, вызванными турбулентностью процесса. В то же время наблюдается постепенное снижение общей спектральной интенсивности, сопутствующее сушке. Эта тенденция отчётливо видна из временных зависимостей интенсивности на двух отдельных длинах волн (рис. 5-15): в максимуме пика воды на 1932 нм и в точке, где поглощение воды не существенно. Обе интенсивности сильно коррелируют с остаточным содержанием воды (рис. 5-15 а), а сглаживание вдоль оси времени (раздел 2.4.1) усиливает корреляцию. Те же тенденции хорошо видны при анимации данных процесса в опубликованных дополнительных материалах к статье [86].

Высокая корреляция спектральных интенсивностей на всех длинах волн в

широком диапазоне объясняется с позиций изменения рефракционных свойств среды в процессе сушки [240], что приводит к изменению маршрутов распространения света в пористой среде гранулята. В ходе сушки жидкие «мостики», соединяющие первичные кристаллы [242], замещаются воздухом. Поскольку показатель преломления на границе кристалл-воздух выше, чем для пары кристалл-вода, сушка приводит к усилению рассеяния, а значит, к росту количества диффузно рассеянного света, достигающего детектора, и соответственно падению наблюдаемой оптической плотности [236].

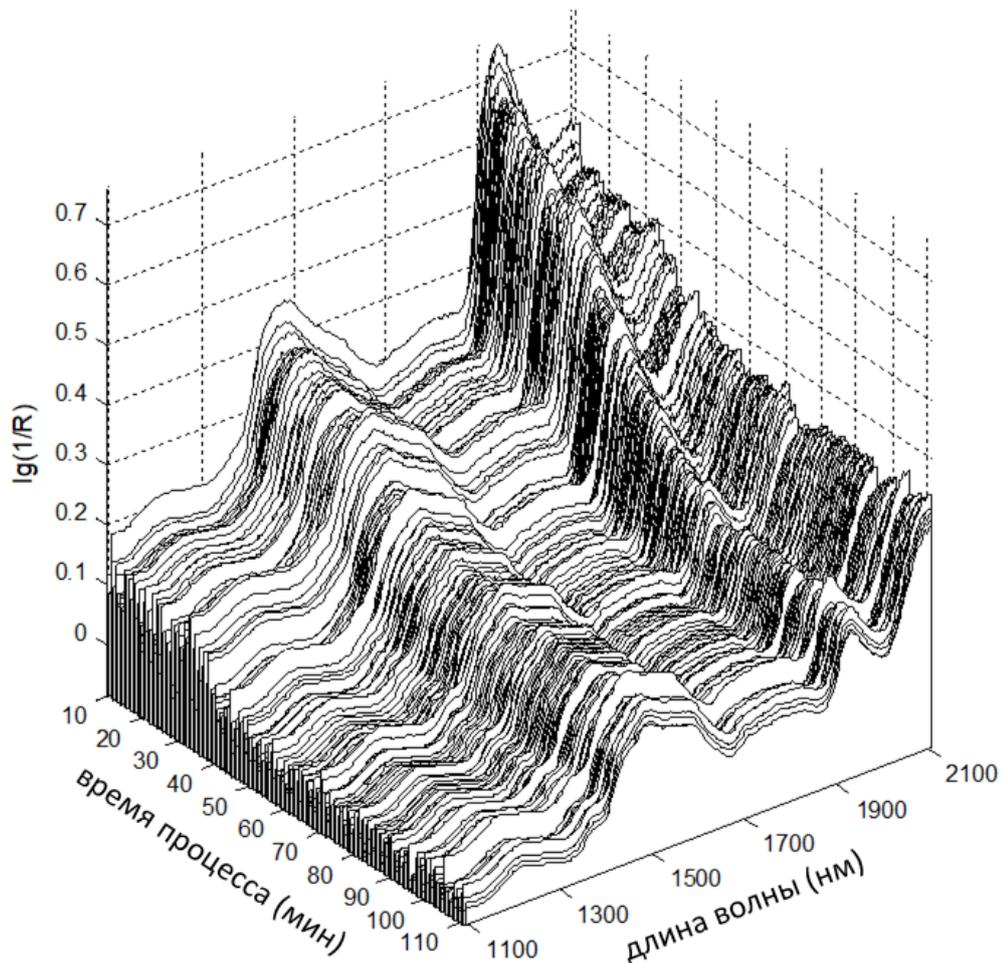


Рисунок 5-14. Онлайн-спектры сушки гранулята (загрузка В03 [86], табл. 3-3).

В этой ситуации необходима предобработка, устраняющая случайный фактор турбулентности, но не затрагивающая полезную – коррелирующую с массовой долей воды – дисперсию данных. Сглаживание данных вдоль оси времени (раздел 2.4.1) отлично справляется с этой задачей.

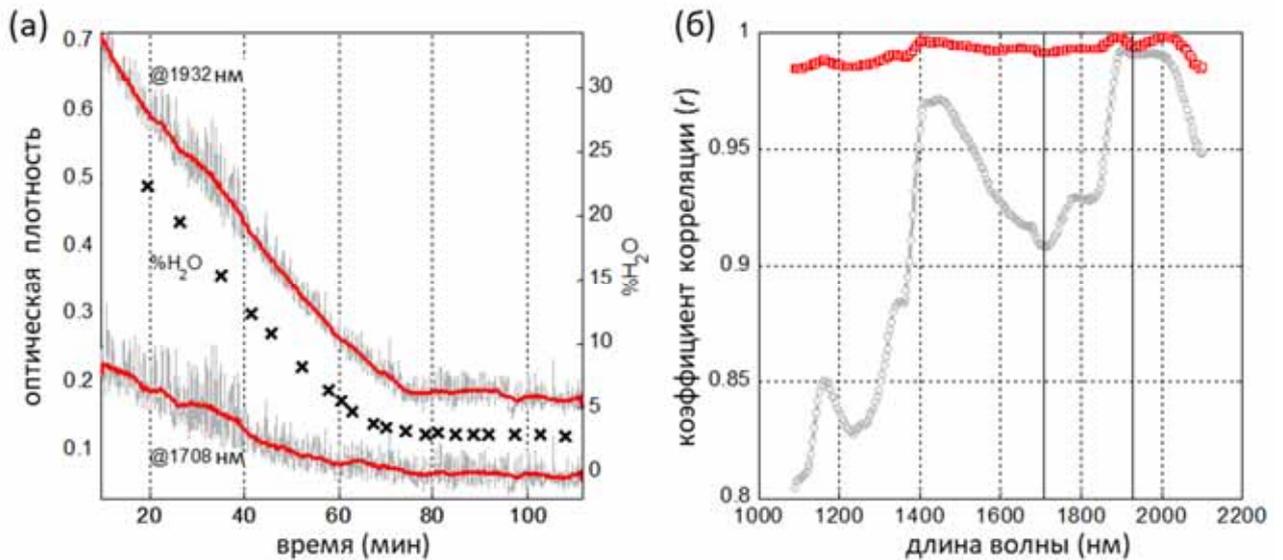


Рисунок 5-15. Анализ корреляций данных для загрузки V03: (а) временные зависимости интенсивностей исходных (серая линия) и сглаженных (красная линия) спектральных данных на выбранных длинах волн и референтные значения массовой доли воды (кресты); и (б) коэффициенты линейной корреляции между содержанием воды и спектральной интенсивностью на отдельных длинах волн для исходных (серые круги) и сглаженных (красные квадраты) данных; вертикальные линии на 1708.1 и 1932.0 нм соответствуют кривым на (а); данные сглажены скользящим средним с окном шириной 47 точек [86].

Выводы визуального анализа спектров подтверждаются результатами МГК на объединённых данных всех 25-ти загрузок с общей спектральной осью. Этот метод разведочного анализа позволяет исследовать траектории индивидуальных загрузок в одном и том же факторном пространстве. Прежде всего было показано, что первые 7 ГК являются значимыми, и соответствующие счета (индивидуальные для каждой загрузки) и общие нагрузки обладают интерпретируемыми чертами. То есть высокая сложность модели согласуется с реальной представительностью данных. Так, счета первого, наиболее значимого ГК всех загрузок, сильно коррелируют с содержанием воды, а счета по ГК2 выглядят как чистый шум, преимущественно описывая турбулентность. При этом нагрузки по ГК1 и ГК2 сильно скоррелированы, например в загрузке V10 (табл. 3-3) $r=0.998$, что свидетельствует о близком спектральном родстве описываемых этими факторами физических явлений. Использование временного сглаживания эффективно устраняет шум из всех семи ГК, выявляя интерпретируемые спектральные черты и элементы траекторий. Более подробно с результатами разведочного анализа можно

ознакомиться в работе [86] и дополнительных материалах к ней.

Построение оптимальной градуировки проводилось путём сравнения ПЛС-моделей в сочетании с различными методами предобработки и модификации исходных данных. Сравнивались как классические методы предобработки БИК-спектров диффузного отражения, а именно, СНВ, МКР и первая производная, так и сглаживание вдоль оси времени с различной шириной окна – 15 или 47 точек. Поскольку исследованный интервал влажности образцов широк, а погрешность прогноза на разных стадиях сушки различна [243], модели были построены как на полном, так и на усечённых наборах данных с верхними границами содержания воды в референтных образцах: 20%, 15% и 10%. Результаты представлены в табл. 5-7.

Относительно высокая сложность модели была обоснована тщательным анализом зависимостей СКО от числа ЛП регрессионных ПЛС-моделей, а также визуальным анализом зашумлённости и интерпретируемости соответствующих счетов и нагрузок. Результат согласуется с полученным ранее для МГК. Количество и разнообразие образцов даже в усечённых моделях было достаточно для построения надёжных моделей такой сложности. При сравнении моделей использовался принцип многоуровневой валидации (раздел 2.4.3).

Снижение верхней границы массовой доли воды в образцах обучающего набора существенно повышает точность градуировки. Так, *СКОПП* сегментативной проверки (в качестве сегментов использовались отдельные загрузки процесса) падает с 0.21 в D до 0.13 в D₁₀. При этом R^2 остаётся на одинаково высоком уровне 0.997–0.998 (табл. 5-7). С учётом такой сильной зависимости погрешности от значения концентрации воды можно рекомендовать ступенчатое переключение с одной модели на другую по мере приближения целевой влажности. Такая автоматическая «фокусировка» модели в ходе процесса может быть реализована в приборном ПО.

Таблица 5-7. Статистика градуировочных ПЛС-моделей для онлайн-мониторинга содержания воды в грануляте методом БИК-спектроскопии: сравнение различных концентрационных диапазонов и методов предобработки данных; все модели построены с 7-ю ЛП.

Набор данн. ^а	Реф. Об. ^б	Предв. Обр. ^в	Градуировка		ППП		ППС ^г		ПТН ^д	
			СКОГ	R ²	СКОПП	R ²	СКОПП	R ²	СКОП	R ²
D	298	нет	0.207	0.998	0.222	0.998	0.236	0.998	0.246	0.998
		С15 ^е	0.181	0.999	0.194	0.998	0.209	0.998	0.198	0.998
		С47 ^ж	0.178	0.999	0.191	0.998	0.210	0.998	0.197	0.998
		МКР	0.264	0.997	0.292	0.996	0.341	0.995	0.290	0.997
		СНВ	0.312	0.996	0.342	0.995	0.395	0.993	0.334	0.996
		1П2.15 ^з	0.203	0.998	0.221	0.998	0.251	0.997	0.256	0.998
D ₂₀	289	нет	0.190	0.998	0.205	0.998	0.216	0.997	0.269	0.997
		С15	0.169	0.998	0.182	0.998	0.195	0.998	0.208	0.998
		С47	0.166	0.998	0.178	0.998	0.190	0.998	0.211	0.998
D ₁₅	268	нет	0.152	0.998	0.161	0.998	0.170	0.997	0.146	0.998
		С15	0.146	0.998	0.155	0.998	0.163	0.998	0.137	0.998
		С47	0.139	0.998	0.147	0.998	0.155	0.998	0.129	0.998
		МКР	0.175	0.997	0.188	0.997	0.210	0.996	0.209	0.996
		СНВ	0.175	0.997	0.191	0.997	0.215	0.996	0.184	0.997
		1П2.15	0.153	0.998	0.164	0.997	0.181	0.997	0.158	0.998
D ₁₀	213	нет	0.116	0.997	0.124	0.996	0.141	0.995	0.129	0.995
		С15	0.109	0.997	0.116	0.997	0.132	0.996	0.122	0.995
		С47	0.109	0.997	0.116	0.997	0.137	0.995	0.121	0.995

Примечания: ^а используемые наборы данных: D – полный (за исключением трёх выбросов), D₂₀, D₁₅ и D₁₀ – данные, ограниченные максимальным значением массовой доли воды в образцах: 20%, 15% и 10% соответственно; ^б число референтных образцов; ^в метод предобработки; ^г перекрёстная проверка сегментами, образованными отдельными загрузками процесса; ^д проверка тестовым набором по дизайну (табл. 3-3); сглаживание вдоль оси времени скользящим средним с шириной окна ^е 15 точек и ^ж 47 точек; ^з первая производная по Савицкому-Голею (ширина окна 15, степень полинома 2).

Предварительные расчёты показали, что увеличение ширины сглаживающего окна в диапазоне от трёх до 47-ми точек ведёт к улучшению точности прогноза, а дальнейшее уширение не даёт заметного эффекта. Исходя из этого, 47 точек (3.9 мин) приняли за максимальную ширину окна. Такое окно ещё является приемлемым с точки зрения продолжительности загрузок, самая короткая из которых составляет около 33 мин. Окно в 15 точек (70 с) примерно соответствует времени циркуляции материала в сушильном реакторе данного типа, в течение

которого бóльшая часть гранулята проходит перед окнами зонда, и можно считать, что усреднённый таким образом спектр представляет весь загруженный материал. Причём с точки зрения общей продолжительности процесса такое усреднённое измерение остаётся практически мгновенным, а задержка результата (отставание аналитического курсора на рис. 4-1) в 35 с (пол-окна) не является значимой. Для окна в 47 точек задержка составляет ровно 2 мин, что может привести к некоторой пересушке. Поэтому с практической точки зрения усреднение по 15-ти точкам может быть предпочтительным, несмотря на несколько худшую точность (табл. 5-7). Окончательное решение принимает аналитик с учётом всех факторов, а для целей сравнения были оставлены оба варианта моделей. Предложенный подход к определению оптимального окна с учётом факторов точности прогноза, продолжительности процесса и времени задержки анализа универсален и рекомендуется к использованию в любых сходных процессах, требующих усреднения данных.

Сравнение различных методов предобработки данных, проведённое на полном наборе D, показывает, что методы коррекции рассеяния (МКР и СНВ) и производная значительно уступают временному сглаживанию скользящим средним, независимо от выбранной ширины окна. Однако этот результат не является совсем неожиданным из-за обнаруженной в работе зависимости общей спектральной интенсивности от массовой доли воды. Классическая коррекция рассеяния уничтожает эту дополнительную информацию, составляющую значительную часть общей дисперсии спектральных данных. Сглаживание во временном домене, напротив, действует в основном на шум, оставляя коррелирующую с водой составляющую спектрального сигнала. Одна из наиболее удачных градуировочных моделей представлена на рис. 5-16. Дальнейшее обоснование этого заключения на основании детального анализа счетов и нагрузок ПЛС-модели, а также дисперсии отдельных ЛП, даны в статье [86].

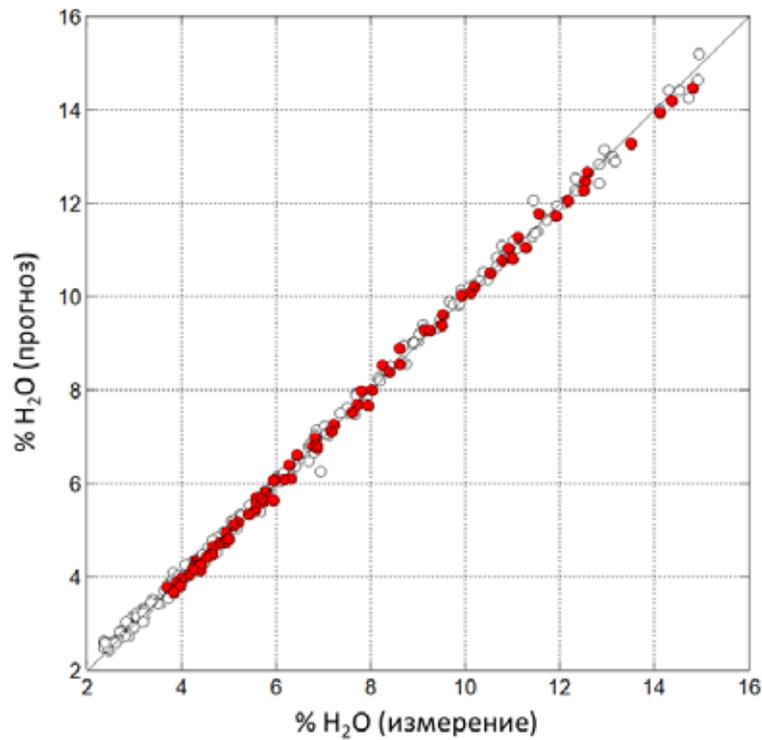


Рисунок 5-16. График «введено-найдено» градуировочной ПЛС-модели для определения массовой доли воды в грануляте онлайн в процессе его сушки (модель D₁₅/C15, табл. 5-7); образцы обучающего и проверочного наборов обозначены полыми и залитыми красными кругами соответственно [86].

Методический подход, основанный на планировании эксперимента, использовании обширного представительного набора обучающих образцов и углублённом анализе данных, позволил изучить основные факторы, влияющие на спектры. Углублённое понимание закономерностей процесса, в свою очередь, позволило разработать оптимальную методику предобработки данных и предложить стратегию «фокусирующегося» моделирования. В результате удалось построить прогностическую модель, обладающую беспрецедентно низкой для онлайн-анализа процессов абсолютной погрешностью – до 0.12% в широком диапазоне концентраций воды от 1% до 10%.

5.2.3 *Онлайн-мониторинг массовой доли воды и толщины покрытия в процессах изготовления пеллет*

Целью настоящего исследования была разработка ОМС на основе БИК- и КР-спектроскопии для определения массовой доли воды в пеллетах и толщины покрытия, спектрально определяемой через общую массу нанесенного материала. КР-мониторинг покрытия пеллет онлайн изучался впервые, поэтому было важно

оценить общую применимость метода к анализу данного процесса и его чувствительность к различным компонентам, а также оценить преимущества совместного использования КР- и БИК-данных для получения адекватной траектории процесса.

Для одновременного сбора данных использовались БИК-спектрометр «J&M TIDAS 1121 SSG» и КР-спектрометр «RxN1» фирмы «Kaiser Optical Systems» (табл. 5-1) с возбуждающим лазером на длине волны 785 нм. Спектрометры были подключены при помощи оптоволоконных кабелей к двум зондам «LHP», установленным в пилотном реакторе. Временные интервалы между последовательно измеряемыми спектрами составляли 10 и 30 с, а время накопления – 75 мс для БИК- и 25 с для КР-спектров соответственно. С методикой эксперимента можно подробно ознакомиться в работе [40]. Спектры загрузки ВЗ представлены на рис. 5-17.

Исчерпывающей характеристикой качества покрытия пеллет является распределение его локальных толщин по размерам. На сегодняшний день нет аналитического оборудования, способного проводить такой анализ на статистически достоверной выборке. В допущении равномерности напыления [40], распределение заменяется средней толщиной покрытия. А при условии равноразмерности ядер пеллет, что также является резонным приближением в силу самой технологии их производства [40], толщину напылённого покрытия можно вычислить из изменения среднего размера пеллет в ходе процесса. Расчёты эффективного размера образца (раздел 2.2.5) с учётом размера пятна, динамики процесса и времени экспозиции показали, что спектры обоих методов отражают среднюю толщину покрытия всех пеллет в реакторе на момент измерения [40].

Предобработка БИК-спектров методом СНВ, а КР-спектров – кусочно-линейной коррекцией базовой линии (раздел 1.3.2) [98,244] благоприятно сказывается на информативности данных – их способности отражать различные стадии процесса производства пеллет (рис. 5-17 б и г).

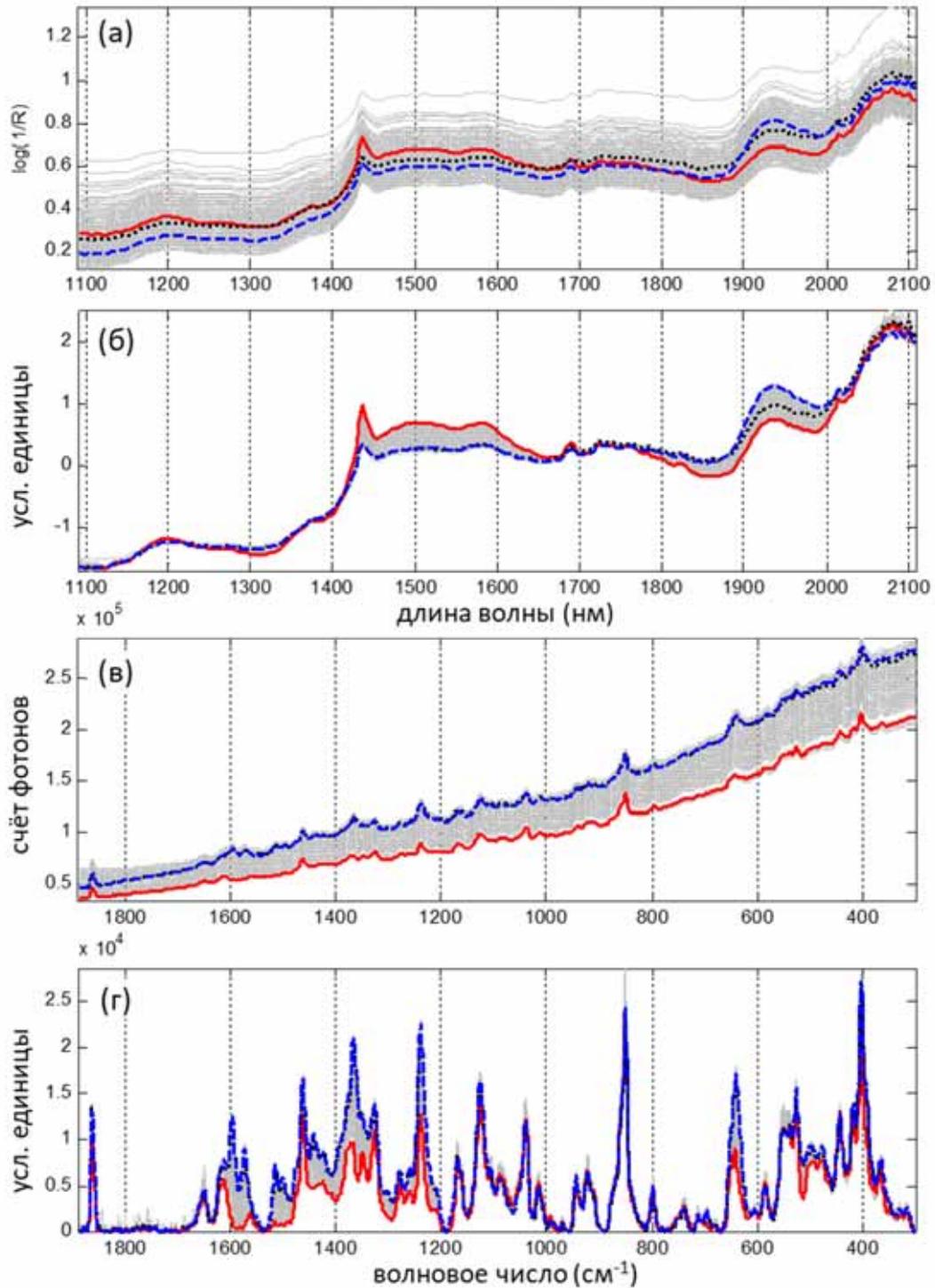


Рисунок 5-17. Спектры загрузки ВЗ: (а) исходные БИК-спектры; (б) БИК-спектры нормированные методом СНВ; (в) исходные КР-спектры; и (г) КР-спектры после КЛК-коррекции базовой линии. Выделены спектры: начало процесса (красная линия), конец стадии покрытия (синяя пунктирная линия) и конец сушки (чёрная точечная линия) [40].

Разведочный анализ БИК- и КР-спектральных данных проводился методом МРК (раздел 1.3.4). Траектории процесса, разрешённые из КР-спектров методом СИМПЛИЗМА (рис. 5-18), были интерпретированы при помощи комплементарных им разрешённых чистых спектров (рис. 5-19). Нисходящие кривые на рис. 5-18 а

и б относятся к сахарозе ядра пеллет. Её спектр сильно доминирует в спектрах всех загрузок (см. например рис. 5-17 г). Этот факт свидетельствует о способности КР-метода доставлять химическую информацию через полимерную оболочку. Концентрационный профиль ацетаминофена (рис. 5-18 б, кривая 2) и соответствующий спектр (рис. 5-19 б) были выделены из данных В2. Разрешение этого компонента основывалось на чистой переменной на длине волны 1611 см^{-1} , совпадающей с полосой валентных колебаний $\text{C}=\text{O}$ вторичных амидов. «Acryl-EZE» – композитный материал, содержащий метакриловую кислоту в качестве основного компонента, формирующего функциональное покрытие пеллет. Его суммарный спектр и концентрационный профиль (рис. 5-18 б, кривая 2) были разрешены из В1. СИМПЛИЗМА может разрешить КР-спектры В3 с двумя или тремя компонентами. Восходящая кривая 1 (рис. 5-18 в) описывает нанесение покрытия, дающего сильный спектральный сигнал.

Однако соответствующие спектры всё ещё содержат черты АФИ и сахарозы, что означает неполное разрешение. Вторая кривая показывает почти зеркальное поведение. Это ожидаемо, учитывая, что рост покрытия сопровождается соответственным уменьшением суммарного содержания других компонентов, а также их маскировкой. Так и в этом случае разрешённые спектральные компоненты показывают смешанные черты сахарозы, АФИ и «Opadry red», сигнализируя о неполном разрешении. Добавление третьего компонента к МРК-модели В3 не помогает разрешить АФИ и ядро; компонент, преимущественно описывающий «Opadry red», также остаётся недоразрешённым. Сходная проблема существует и в данных В4. Неопределённость решения при моделировании смесей со сходными спектрами компонентов или коррелирующими профилями является известным ограничением МРК. В данном случае наблюдаемое недостаточное разрешение объясняется отрицательной корреляцией профилей компонентов при напылении.

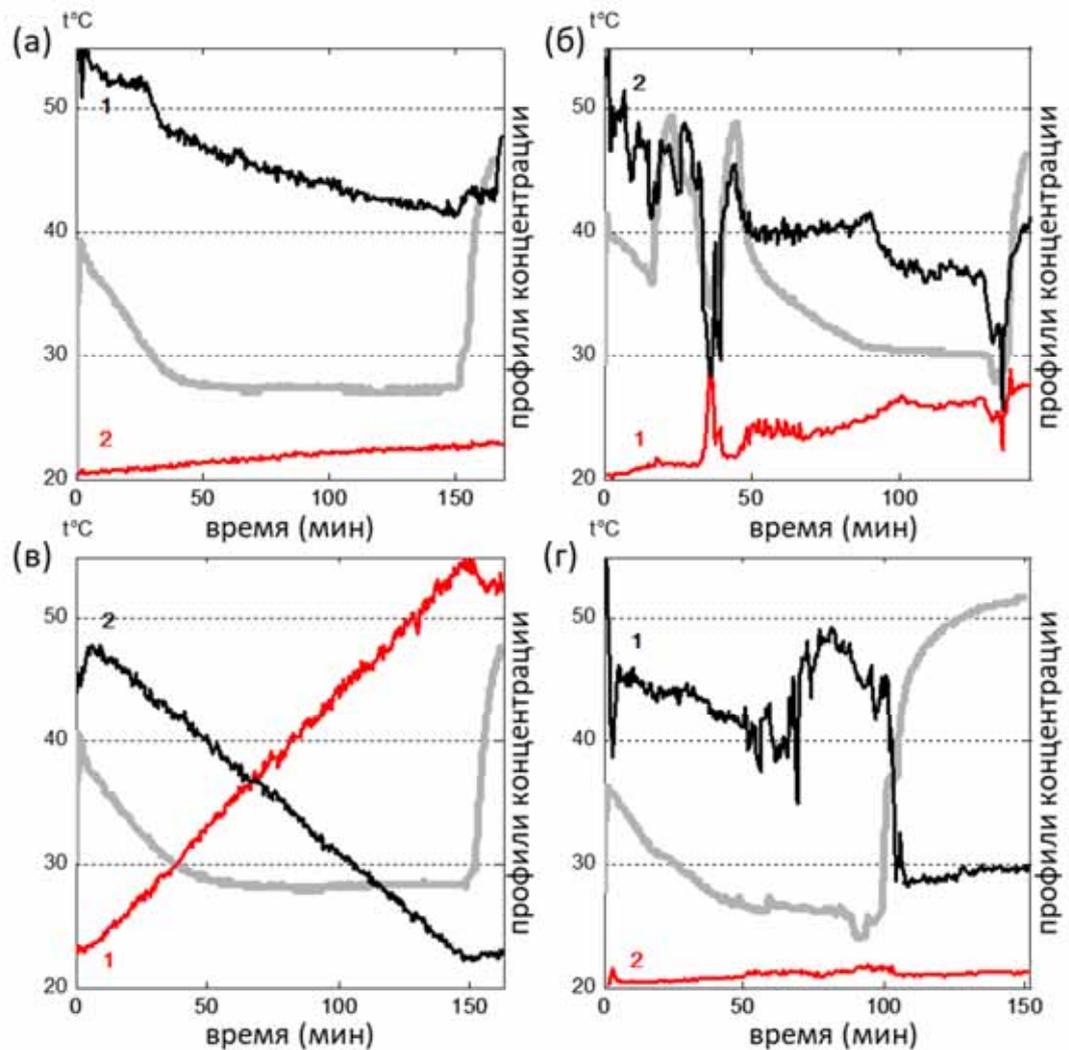


Рисунок 5-18. Измеренная температура (серая линия) и МКР-разрешённые из КР-спектров траектории процесса для отдельных загрузок: (а) В1; (б) В2; (в) В3; и (г) В4. Номер кривой обозначает порядок разрешения соответствующего компонента алгоритмом СИМПЛИЗМА [40].

Артефакты, наблюдаемые в В2, явно связаны с температурой продукта (рис. 5-18 б), косвенно контролируемой через скорость распыления (увеличение скорости ведёт к повышению влажности пеллет и падению температуры). Сильные перепады траектории на рис. 5-18 б совпадают с остановкой «кипящего слоя», когда нормальный ход распыления воды был затруднён агломерацией продукта. Профиль температуры в В4 ближе к нормальному (как в В1 и В3), но температура в целом ниже, что согласуется с более высокой скоростью распыления в этой загрузке. Разрешённые профили на рис. 5-18 г также отражают нестабильность процесса.

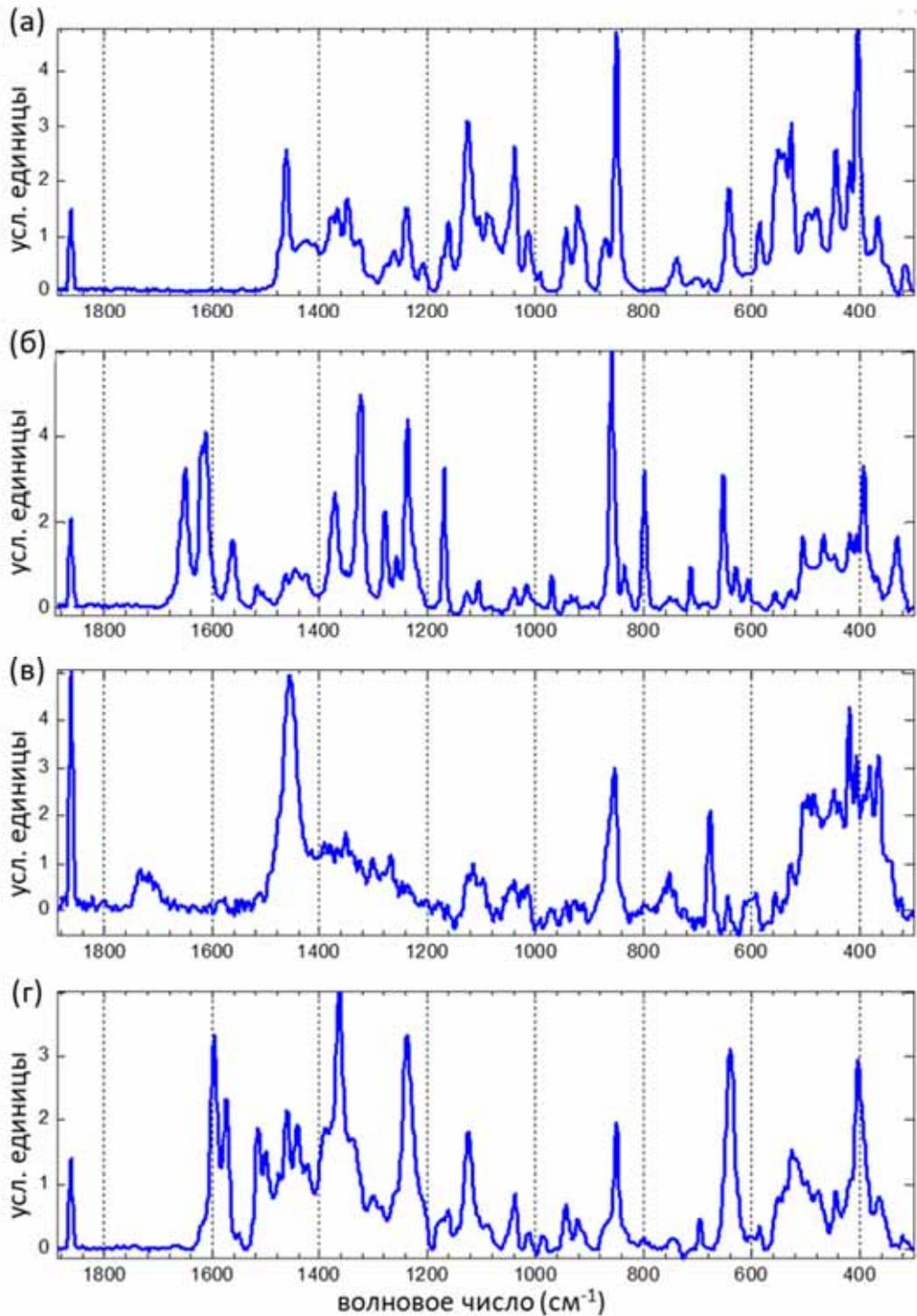


Рисунок 5-19. КР-спектры чистых компонентов, разрешённые из данных отдельных загрузок и их дополненных комбинаций: (а) сахара из В2; (б) АФИ из В2; (в) «Acryl-EZE» (суммарно) из В1; и (г) «Opadry red» (суммарно) из объединённых данных В2+В3 [40].

Недостаточно полное разрешение компонентов в МРК можно преодолеть использованием дополненных данных (раздел 2.2.1), а также объединением данных разных загрузок, связанных друг с другом общими компонентами в один анализируемый блок (X). Такая конкатенация была применена к данным для двух

стадий одного процесса изготовления пеллет: исходного напыления АФИ на основу пеллет (В2) и их дальнейшего покрытия оболочкой (В3). Эти загрузки отлично подходят для совместного анализа, благодаря присутствию в них трёх общих компонентов – сахарозы, АФИ и воды. Отсутствие четвёртого компонента – «Opadry red» в одном из блоков не препятствует совместному анализу методом МРК [120].

Траектория процесса, разрешённая из спектров объединённых загрузок В2 и В3, показана на рис. 5-20 а. Благодаря устранению рассмотренной выше проблемы симметрии профилей в В3 (рис. 5-18 в), «Opadry red» был успешно разрешён (чистый спектр на рис. 5-19 г), что подтверждается отнесением пиков [40]. Спектры чистых компонентов сахарозы и АФИ, разрешённые из В2+В3, также согласуются с литературой [245]. Разрешённые траектории находят однозначную интерпретацию. В комбинированном наборе (В2+В3) спектры образуют естественную последовательность: за напылением АФИ на сахарную основу (В2) следует функциональное покрытие (В3) с фазой сушки в конце каждой загрузки. Хотя кривые стадии В2 отражают нестабильность процесса, явно наблюдается рост концентрации АФИ и уменьшение концентрации сахарозы. Функциональное покрытие в В2 остаётся близким к нулю, что и ожидалось. Близость к нулю профиля «Opadry red» служит дополнительным подтверждением качества МРК-модели. Кривые АФИ и сахарозы на стадии сушки В2 меняют свою форму на конгруэнтную. Это подтверждает наше предположение о существовании косвенной зависимости откликов компонентов от влажности продукта.

В КР-спектрах ($1700\text{--}780\text{ см}^{-1}$) различных стадий процесса (рис. 5-20 б) наблюдаются полосы поглощения основы (начало В2), интенсивность которых постепенно снижается по мере добавления АФИ, а затем «Opadry red», сигналы которых соответственно растут. «Opadry red» имеет область высокой селективности около 1590 см^{-1} , которая и использовалась алгоритмом для его разрешения. Спектр в провале интенсивности кривой 3 (красный пунктир на рис. 5-20 а) напоминает уменьшенную копию спектра основы (рис. 5-19 а) с некоторыми чертами добавленного АФИ выше 1550 см^{-1} , то есть провал связан с

общим падением интенсивности, а не с отдельными спектральными переменными.

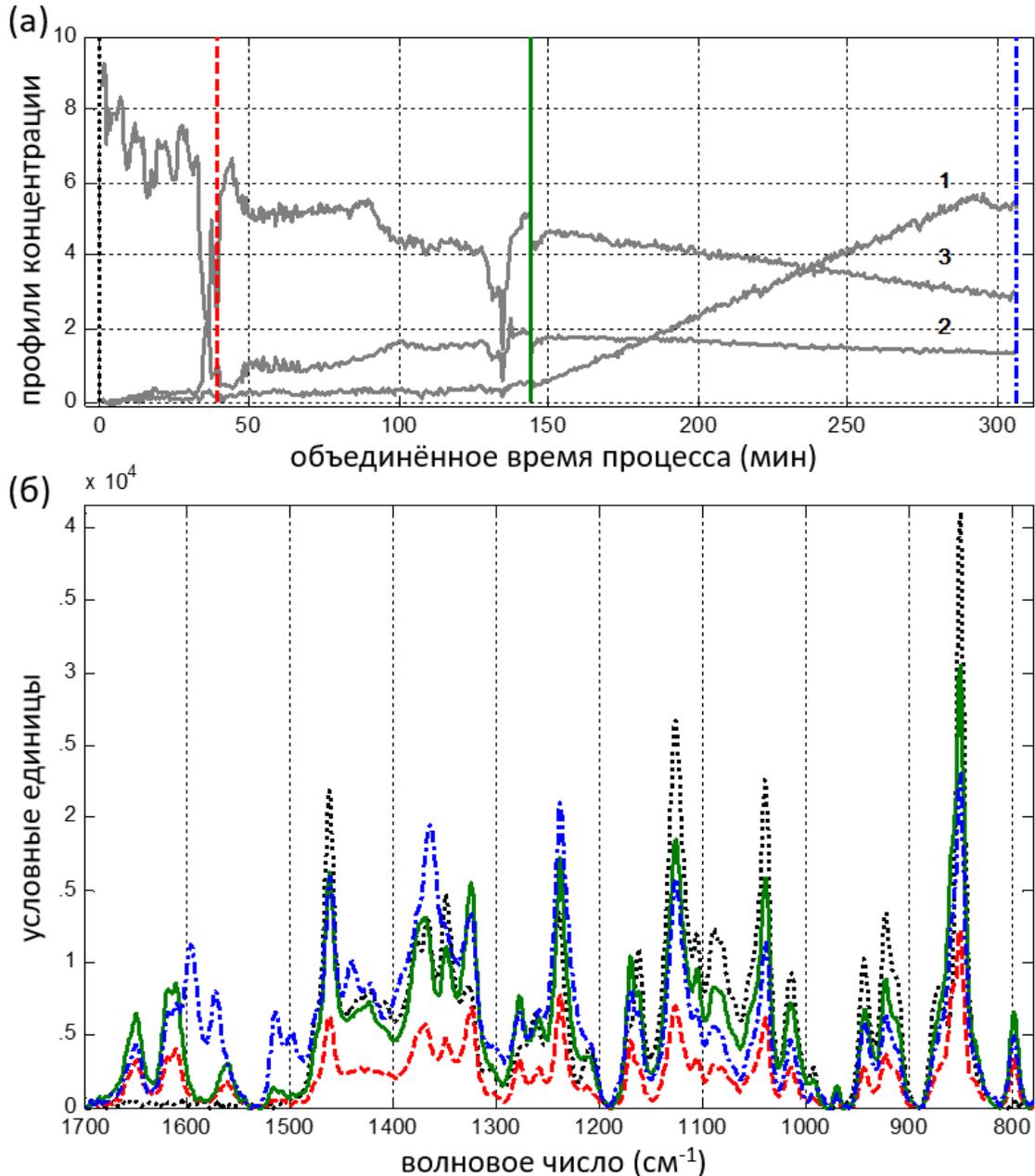


Рисунок 5-20. (а) Концентрационные профили компонентов, разрешённые из КР-спектров объединённых загрузок В2 и В3 и (б) КР-спектры в выбранные моменты времени, обозначенные теми же типами линий на (а).

Концентрационный профиль воды не разрешается из КР-спектров даже объединённых загрузок. Попытка его разрешения из БИК-спектров также не привела к успеху из-за сильного наложения полос. Зато дополненные данные блоков КР- и БИК-спектров объединённых загрузок В2+В3 позволили разрешить профиль воды в модели с тремя компонентами. Разрешённый профиль в точках пробоотбора хорошо коррелирует с данными референтного анализа ($r=0.956$). По сравнению с результатами МРК на КР-спектрах преимущества совместного

анализа спектров очевидны. В целом углублённый разведочный анализ показал достаточную для анализа процесса покрытия пеллет информативность методов БИК- и КР-спектроскопии и целесообразность их совместного использования.

Регрессионный анализ онлайн-данных подтвердил преимущества использования дополненных данных для построения градуировочной модели для определения массовой доли воды в пеллетах. Наилучшая модель, полученная на дополненных (сочетающих КР- и БИК-спектры) данных объединённых загрузок В2+В3, показана на рис. 5-21. *СКОПП* модели в этом случае составило 0.15, что заметно ниже, чем для БИК- (0.18) и КР-данных (0.17) по отдельности. В то же время сравнение моделей для толщины покрытия показало отсутствие выигрыша при совместном использовании спектральных методов. Наилучший результат в этом случае даёт моделирование на БИК-спектрах. Как ранее отмечалось в литературе (раздел 1.1.3), причина синергетического эффекта при объединении разных спектральных методов не вполне изучена, и его отсутствие не является необычным.

В дополнение к изложенным результатам исследования процессов покрытия пеллет в работе [208] был предложен экспресс-метод определения толщины нанесённого слоя «у линии», основанный на анализе цифровых фотографий проб, взятых из среды процесса. Пример такого изображения показан на рис. 5-22. Для разработки метода использовались обезвоженные после референтного гравиметрического анализа образцы пеллет, взятые из загрузок В1 и В2 (раздел 3.4.2). В анализе изображений были опробованы как классические методы сбора признаков, основанные на распознавании и измерении отдельных пеллет, так и многомерные, основанные на математических преобразованиях полного изображения. Наилучшим среди преобразований оказался метод углового измерения (МУИ; *англ.* angle measure technique, АМТ). Вектор признаков затем использовался для построения ПЛС-регрессионной модели. Созданный прототип анализатора показал достаточную точность [208]. Предложенный подход может быть пригоден для дальнейшей разработки экспресс-метода измерения толщины покрытия пеллет «у линии», а при наличии вмонтированной в реактор камеры – и

для создания онлайн-метода. При этом цифровые изображения можно использовать как независимо, так и в комбинации с другим, например спектральным, методом.

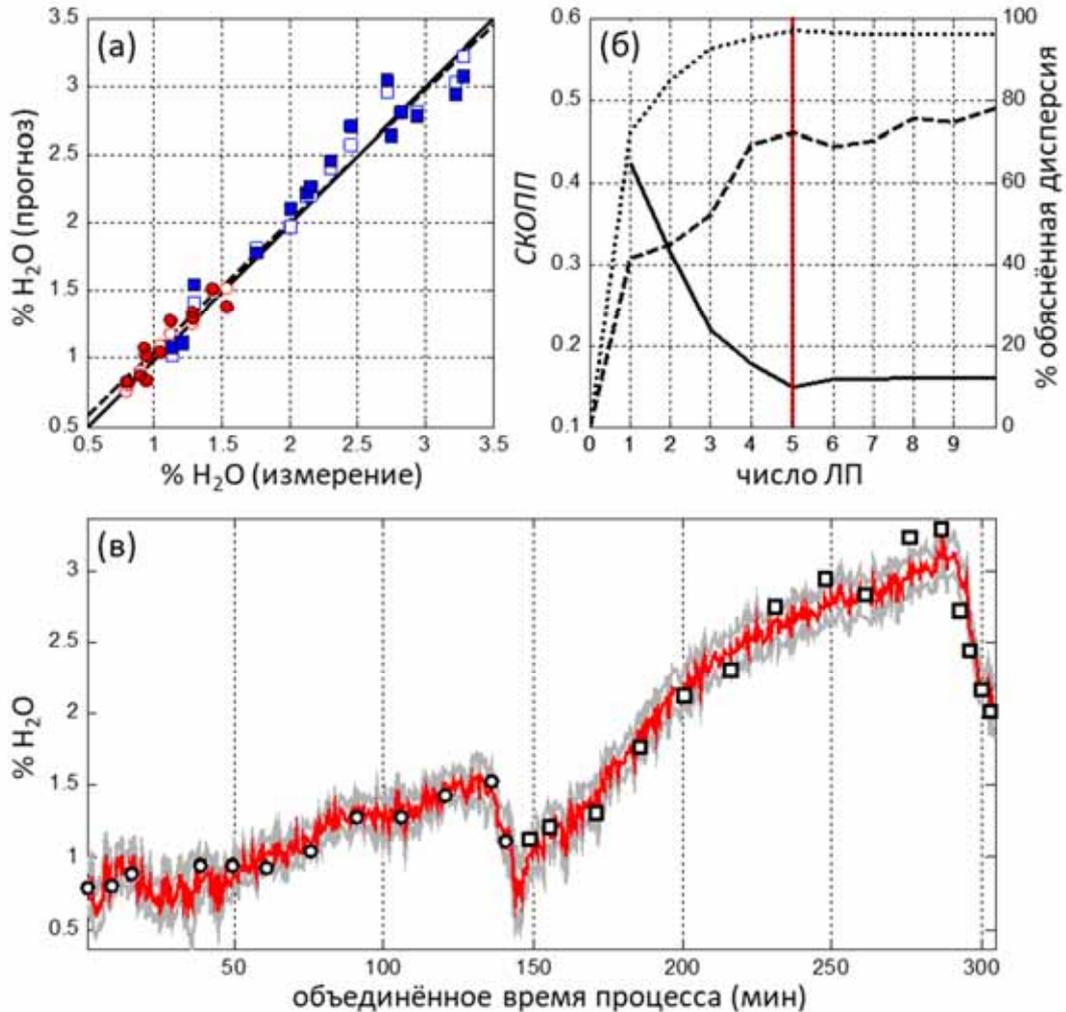


Рисунок 5-21. Градуировочная модель для массовой доли воды в пеллетах, построенная на образцах объединённых данных В2+В3 с использованием комбинации КР- и БИК-спектров: (а) график «введено-найдено»: полые и залитые символы обозначают образцы обучающего и проверочного наборов в В2 (круги) и В3 (квадраты) соответственно; функция линейной регрессии обозначена пунктиром; (б) зависимости *СКОПП* (сплошная линия), объяснённой дисперсии *у* (пунктирная линия) и объяснённой дисперсии *X* (точечная линия) от числа ЛП; и (в) прогноз содержания воды (красная линия) и референтные значения %H₂O в пробах В2 (круги) и В3 (квадраты).

В ходе исследования был независимо разработан алгоритм точного счёта пеллет на снимке пробы. В сочетании с её аналитическим взвешиванием он может использоваться для точного определения среднего веса пеллет, а значит и коррелирующей с ним толщины нанесённого покрытия.



Рисунок 5-22. Пример изображения пеллет для определения толщины покрытия [208].

Приведённый пример интересен как иллюстрация прототипа ОМС (а формально регистрация и анализ цифровых изображений находится в рамках мультисенсорного подхода), не являющегося химическим анализатором в традиционном смысле. В данном случае анализ качества лекарственного препарата, во многом определяемого химическим составом и количеством добавленных компонентов пеллет, происходит путём измерения физического параметра – размера частиц, коррелирующего с количеством (толщиной) нанесённого защитного покрытия и опосредованно определяющим кинетику высвобождения АФИ после приёма лекарства пациентом. В подобных ситуациях следует, видимо, говорить о смежном с химическим анализом применении ОМС (табл. В-1, Приложение В).

5.2.4 Онлайн-прогноз профилей растворения АФИ из БИК-спектров пеллет в процессе их изготовления

Отдельной задачей исследования процессов изготовления пеллет была разработка метода прогнозирования текущего профиля высвобождения АФИ из наблюдаемых в ходе процесса БИК-спектров. Поскольку кинетические характеристики растворения пеллет являются ключевыми параметрами качества продукта, их онлайн-мониторинг необходим для реализации концепции ВРВ (раздел 4.3.2).

Исследование включало 13 экспериментальных загрузок процесса покрытия,

в ходе которых снимались онлайн-БИК-спектры (как описано в разделе 5.2.3), а также отбирались образцы пеллет для проведения тестов растворимости. Дизайн эксперимента описан в разделе 3.4.2. Изучение результатов тестов растворимости показало, что процентная доля высвобожденного АФИ φ от времени (t), независимо от материала и толщины покрытия, хорошо описывается уравнением с двумя константами m и k , сходным с уравнением кинетики автокатализа:

$$\varphi(t, m, k) = 100k \frac{\exp[(m + k)t] - 1}{m + k \exp[(m + k)t]} \quad (5.1)$$

В целом параметр m отвечает за скорость высвобождения (наклон), а k отражает индукционный период (задержку). Дальнейшее исследование показало, что параметр m определяется материалом покрытия, и каждый набор загрузок W и Y процесса (табл. 3-4) характеризуется его общими оценками m^W и m^Y . Их значения были рассчитаны из данных соответствующих загрузок с использованием метода последовательной Байесовой оценки (ПБО; *англ.* successive Bayesian estimation) [246] для многопараметрической оптимизации. Согласно этому подходу данные отдельных загрузок обрабатываются одна за другой, а оценки, полученные нелинейной аппроксимацией предыдущей загрузки, используются как априорная информация для следующей.

Другая константа k тесно связана с толщиной нанесённого покрытия и, следовательно, может быть определена из БИК-спектров процесса (раздел 5.2.3). Оценки \hat{k}_{bj} (b – номер загрузки, j – порядковый номер пробы от 1 до 9 в данной загрузке) получают попутно в ходе ПБО-оптимизации m^W и m^Y . Дальнейшее исследование констант \hat{k}_{bj} показало линейность их обратного логарифма:

$$q = -\lg(k) \quad (5.2)$$

от количества распылённого материала в рамках одной загрузки, хотя отсечения и наклоны соответствующих прямых в разных загрузках отличаются. Последнее свидетельствует о том, что кинетика высвобождения АФИ зависит не только от толщины, но и от качества покрытия, а следовательно – от условий процесса, и БИК-спектры способны эти различия передать.

Для прогноза значений q из онлайн-спектров (обозначим их q_{PLS}) строились

ПЛС-регрессионные модели на СНВ-нормированных БИК-спектрах, ближайших к точкам пробоотбора. Для наборов W и Y были построены отдельные градуировки (рис. 5-23) с тремя и двумя ЛП соответственно. Образцы с порядковым номером 1 каждой загрузки, содержащие практически непокрытые пеллеты, были исключены как явные выбросы. Высокая производительность результирующих моделей, построенных на представительных обучающих наборах образцов спланированного эксперимента, свидетельствует о применимости метода для практического прогноза профилей растворения.

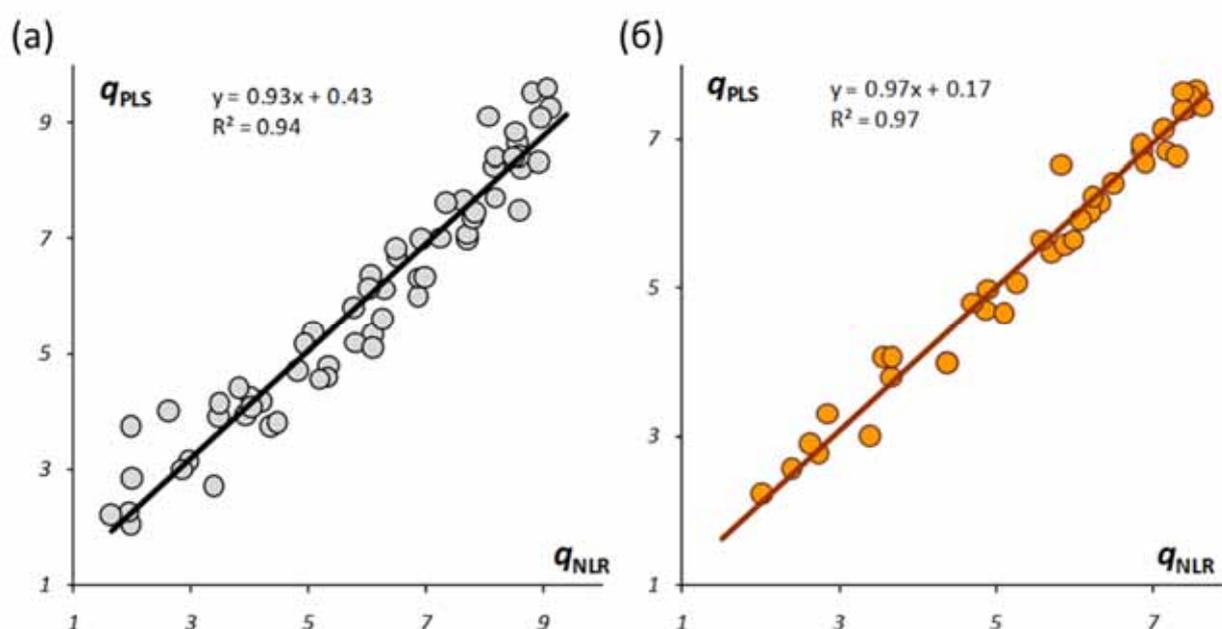


Рисунок 5-23. Графики «введено-найдено» градуировочных ПЛС-моделей для параметра q : (а) для набора W и (б) для набора Y; q_{PLS} – прогноз, q_{NLR} – референтные значения, рассчитанные из данных тестов растворимости. В качестве прогноза даны результаты ППС с размером сегмента в 10% от данных [185].

Для проверки прогностических возможностей построенных моделей и подхода в целом загрузки W2 и Y5 (по 9 образцов в каждой) были исключены из обучающего набора и использованы как новые данные для валидации. Оценки m^W и m^Y , рассчитанные на усечённых данных, лишь незначительно отличаются от оценок, сделанных на полном наборе. Были построены две градуировочные модели для q в наборах W и Y соответственно, которые были затем использованы для прогноза кинетических параметров k в проверочных загрузках W2 и Y5. Спрогнозированные профили высвобождения АФИ для Y5 (рис. 5-24) иллюстрируют работоспособность предложенного метода.

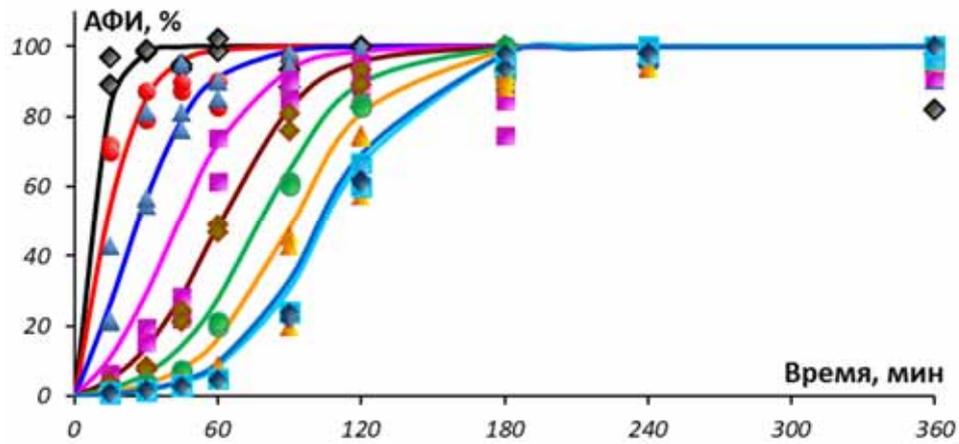


Рисунок 5-24. Прогноз высвобождения АФИ для образцов загрузки У5; полые символы обозначают выбросы [185].

Таким образом, комбинация нелинейной аппроксимации кинетических кривых высвобождения АФИ с ПЛС-регрессионным анализом спектральных данных позволяет прогнозировать будущие профили растворения из регистрируемых онлайн БИК-спектров. Схема проведения моделирования и прогноза представлена на рис. 5-25. Разработка рутинной процедуры для онлайн-прогноза профилей высвобождения АФИ начинается с оценки методом ПБО параметра m из данных теста растворимости. Эта оптимизация параллельно оценивает параметры $q = \{-\ln(k_{bj})\}$, значения которых изменяются в ходе процесса. Затем строится ПЛС-градуировка на БИК-спектральных данных, позволяющая оценивать q_{PLS} , а значит и k в уравнении (5.1) для новых спектров. Как результат, градуировочная модель вместе с уравнением (5.1) может использоваться для рутинного определения полной кривой высвобождения АФИ для каждого спектрального измерения онлайн. В случае смены материала покрытия ПЛС-модель должна быть пересчитана.

В результате исследования разработан математический метод, позволяющий в реальном времени идущего процесса по регистрируемым онлайн БИК-спектрам моделировать будущие профили высвобождения АФИ из производимых таблеток. Полученные результаты представляют ценность в дальнейшей разработке ОМС для мониторинга процесса изготовления таблеток и реализации ВРВ.

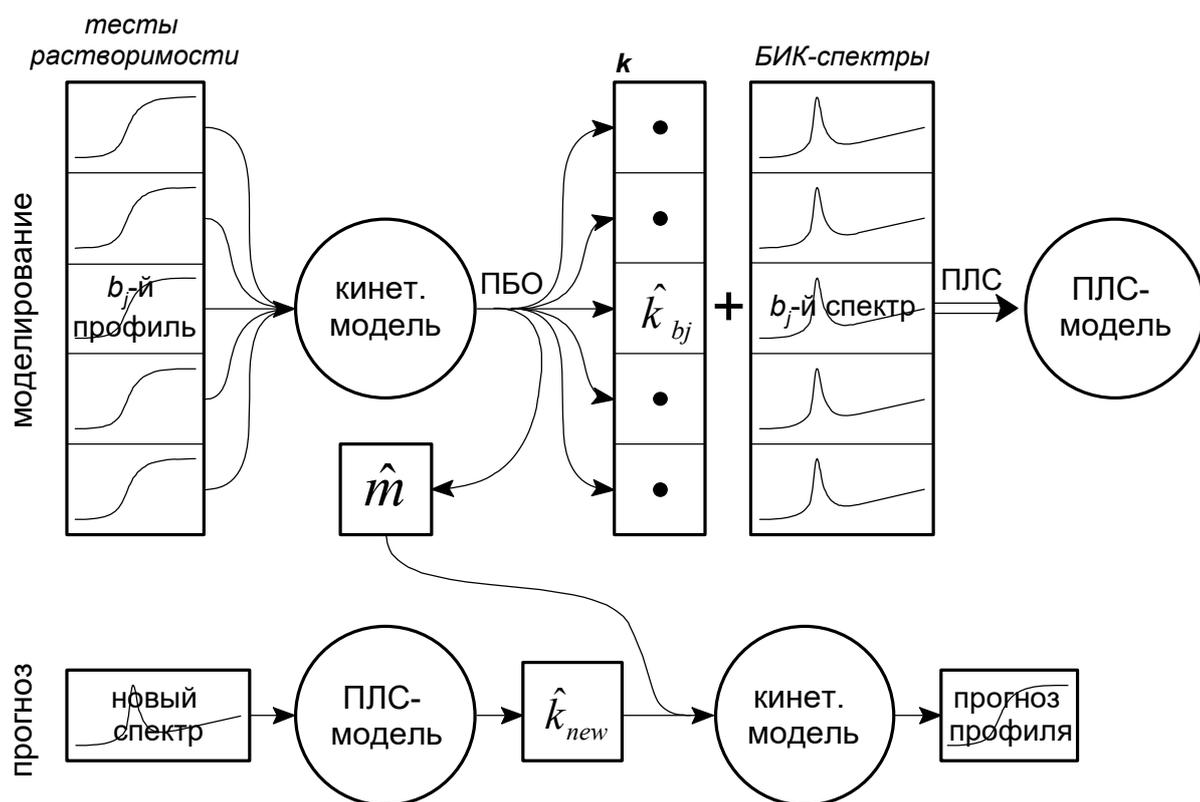


Рисунок 5-25. Схема моделирования и онлайн-прогноза профилей растворения АФИ для процесса покрытия pellets [185].

5.3 Онлайн-мониторинг ферментации *Saccharomyces cerevisiae*

В настоящем исследовании рассмотрены различные ОМС для мониторинга процесса ферментации пекарских дрожжей семейства сахаромициновых (*лат. Saccharomyces cerevisiae, S. Cerevisiae*) на основе флуоресцентной, ИК- и БИК-спектроскопии.

5.3.1 Мониторинг биотехнологических процессов

Большинство промышленных биотехнологических процессов недостаточно контролируются с точки зрения химического анализа. Как правило, наблюдению подлежат только кислотность (pH), окислительно-восстановительная активность, растворенный кислород, биомасса и температура. Мониторинг всех этих параметров производится онлайн с использованием простых сенсоров и зондов. Количественное определение питательных компонентов среды, таких как глюкоза и другие углеводы, а также большинства целевых органических продуктов требует отбора реакционных проб и проведения лабораторного, например хроматографического, анализа. Из-за низкой, по сравнению с ходом процесса,

скорости анализа, традиционный подход часто бывает недостаточным для обеспечения качества продукта и эффективности производства, в том числе для корректирующего вмешательства в случае сбоя [9]. Для решения этих задач необходим онлайн-мониторинг процессов биотехнологии.

Высокоинформативный оптический спектральный анализ ферментационной среды осложняется её многокомпонентным составом, воздействием турбулентности процесса, а также рассеянием света клетками дрожжей и пузырьками выделяющегося газа. Поэтому ПЭ и многомерный факторный и регрессионный анализ данных являются важной наукоёмкой частью разработки методов АКТП в биотехнологии.

Ведущая роль в разработке методов непрерывного мониторинга ферментационных процессов принадлежит БИК-спектроскопии [247,248] и флуориметрии [249]. Эти оптические методы легко адаптируются к анализу процессов онлайн при помощи волоконной оптики и зондов на её основе. ИК-спектроскопия значительно реже используется для онлайн-мониторинга процессов [250-252], что объясняется экспериментальной сложностью использования метода в условиях процесса, требующих специальных интерфейсов измерения. До недавнего времени цена и ограниченная доступность ИК-проводящих материалов были основными препятствиями к разработке и использованию удобных оптоволоконных зондов.

В последние десятилетия было разработано множество различных НПВО-зондов для ИК-анализа на базе ХИК- и ПИК-волокон (раздел 1.1.3), которые были успешно протестированы в различных аналитических приложениях, таких как непрерывный мониторинг процессов и анализ биологических тканей [16-18,21, 22]. Текущие достижения оптических технологий, в частности малобюджетные миниатюрные ИК-детекторы [29-31], создают предпосылки для расширения области применения ИК-спектроскопии и для её возрастающего использования в промышленных и полевых анализаторах.

Флуориметрия – очень эффективный метод анализа биологических систем, в том числе мониторинга биотехнологических процессов [253-254]. Высокая

чувствительность метода позволяет идентифицировать и количественно определять продукты метаболизма и (ко-)энзимы, как вне-, так и внутриклеточные, которые обычно флуоресцируют и могут действовать как маркеры роста культуры и её метаболического состояния [255]. Возможность «вживую» наблюдать внутриклеточную активность превращает флуориметрию в уникальный инструмент для различных биоаналитических приложений. Тенденция последних двух декад – использование спектроскопии двумерной (2D) флуоресценции [253,256-259], генерирующей массив данных, называемый спектром возбуждения-эмиссии (СВЭ). Анализ серии СВЭ, снятых в ходе процесса, требует использования специальных методов анализа данных [260], в том числе трёхмодальных методов, таких как ПАРАФАК [261] или трёхмодальная ПЛС-регрессия [238]. ПАРАФАК применялся для разрешения компонентов по серии 2D-спектров флуоресценции, снятых в ходе процесса ферментации [260,262]. Разрешённые концентрационные профили и индивидуальные СВЭ компонентов позволяют проводить их идентификацию и углублённо изучать закономерности процесса.

Детектируемый сигнал флуоресценции реакционной среды может быть осложнён нестабильной интенсивностью возбуждающего излучения, рассеянием и поглощением различных компонентов, наложением сигналов, эффектом тушения, каскадной флуоресценцией и др. [180]. Дополнительную сложность для анализа данных СВЭ представляет присутствие в них пика рассеянного света возбуждающего излучения. В большинстве случаев волна возбуждения отсекается оптическим фильтром ещё на стадии регистрации данных или позже, при их обработке. Однако при наложении пика рассеяния источника на сигнал флуоресценции такой метод может быть слишком деструктивен и сказываться на результатах моделирования. Чтобы избежать этого, были предложены различные алгоритмы [262-264].

Процесс ферментации *S. cerevisiae* – широко распространённый модельный процесс в биотехнологии. Культивация происходит в закрытом биореакторе в аэробных или анаэробных условиях, в зависимости от желаемого продукта. В обоих случаях метаболизм начинается с глюкозы, приводя к образованию

пирувата, а также небольшого количества полезной энергии для клеток в виде молекул аденозинтрифосфата (АТФ). В отсутствие кислорода пируват конвертируется в этанол и CO_2 . В аэробных условиях пируват трансформируется с образованием CO_2 , H_2O и большого количества АТФ. Для интерпретации данных флуоресценции важно понимать роль носителей энергии для образования АТФ – никотинамидадениндинуклеотида (НАД; *англ.* nicotinamide adenine dinucleotide, NAD) в окисленной (NAD^+) или восстановленной (NADH) формах и флавинадениндинуклеотида (ФАД; *англ.* flavine adenine dinucleotide, FAD) в формах FAD/ FADH. В отсутствие кислорода начинается производство этанола, а концентрации NADH и FADH возрастают, в то время как содержание NAD^+ и FAD падает [258]. NADH является сильным флуорофором, в отличие от NAD^+ , который не флуоресцирует. Таким образом, NADH может использоваться как маркер, указывающий на метаболическое состояние культуры [253,255,265,266]. Флавины, напротив, флуоресцируют только в их окисленной форме [254]. В фазе роста NADH и FAD коррелируют с содержанием биомассы [256].

5.3.2 Разработка мультисенсорной системы в ИК-области

В данном разделе изложены результаты проекта по разработке сенсорного ИК-анализатора для непрерывного определения концентраций этанола, глюкозы и фруктозы с использованием технологии НПВО-зондирования. Были разработаны три прототипа устройств, использующих различные современные детектирующие модули, зонды и оптические волокна. Они были сначала испытаны и охарактеризованы на лабораторном наборе образцов, разработанном с применением диагонального дизайна (раздел 3.3.2), а затем протестированы в условиях процесса ферментации. Параллельно измерения были произведены методом БИК-спектроскопии с использованием зонда на пропускание. Точность методов сравнивается и обсуждается с точки зрения разработки ОМС.

Для тестирования разрабатываемых ИК- и БИК-систем использовались два лабораторных набора образцов, содержащих бинарные водные растворы этанола – глюкозы (EG) и глюкозы – фруктозы (FG), приготовленные в соответствии с диагональной схемой $d2_{0v1}^{25}$ на 25 уровнях (рис. 3-2 б). В тексте образцы

обозначены префиксом соответствующего набора (EG или FG), за которым следует номер образца на рис. 3-2 б. Используемые в наборах диапазоны концентраций составили: для этанола – от 0 до 144 г/л, для глюкозы – от 0 до 288 г/л и для фруктозы – от 0 до 120 г/л.

Эталонные модели для последующего тестирования эффективности разрабатываемых ИК-сенсорных систем (раздел 2.1.1) были построены на полноспектральных измерениях EG- и FG-наборов на Фурье-спектрофотометре «Matrix MF» («Bruker AG», г. Этлинген, ФРГ) через НПВО-зонд. БИК-спектры в области 1100–2100 нм снимались на спектрофотометре «TIDAS S-1000 MS-T50/16» фирмы «J&M Analytik AG» (табл. 5-1).

В измерениях было использовано два типа НПВО-зондов фирмы «art photonics» (г. Берлин, ФРГ) на основе ПИК-волокна (раздел 1.1.3) с использованием различных НПВО-элементов: алмазного (DATR) и сменной ПИК-петли (LATR). Зонд LATR изготовлен без использования дорогостоящих материалов и технологий и значительно дешевле, чем DATR. Онлайн-БИК-спектры снимались с использованием погружного трансфлексионного зонда фирмы «Avantes» (г. Альпендорн, Нидерланды) с регулируемым зазором. Для компенсации высокого поглощения воды оптический путь зонда был установлен на минимальное значение 0.5 мм.

В разработанных сенсорных системах для ИК-области были протестированы две спектрометрические технологии с пироэлектрическими (*англ.* pyroelectric) детекторами. Их технические детали изложены в работе [65] и приложении к ней. Первой системой был дифракционный спектрометр (GrS), построенный на линейном сенсоре PY128LA фирмы «PYREOS» (г. Эдинбург, Шотландия) со 128-ю элементами (рис. 5-26). Участок спектрального сигнала сенсора был настроен таким образом, чтобы покрывать две интенсивных полосы этанола между 9 и 10 мкм (приблизительно 1100–1000 cm^{-1}). Другой сенсор базировался на регулируемом НИС типа Fabry-Pérot (раздел 1.1.3), модель «LFP-80105-337» от «InfraTec GmbH» (г. Дрезден, ФРГ) [33], оптимизированном для измерений в области 8–10 мкм (1250–1000 cm^{-1}).

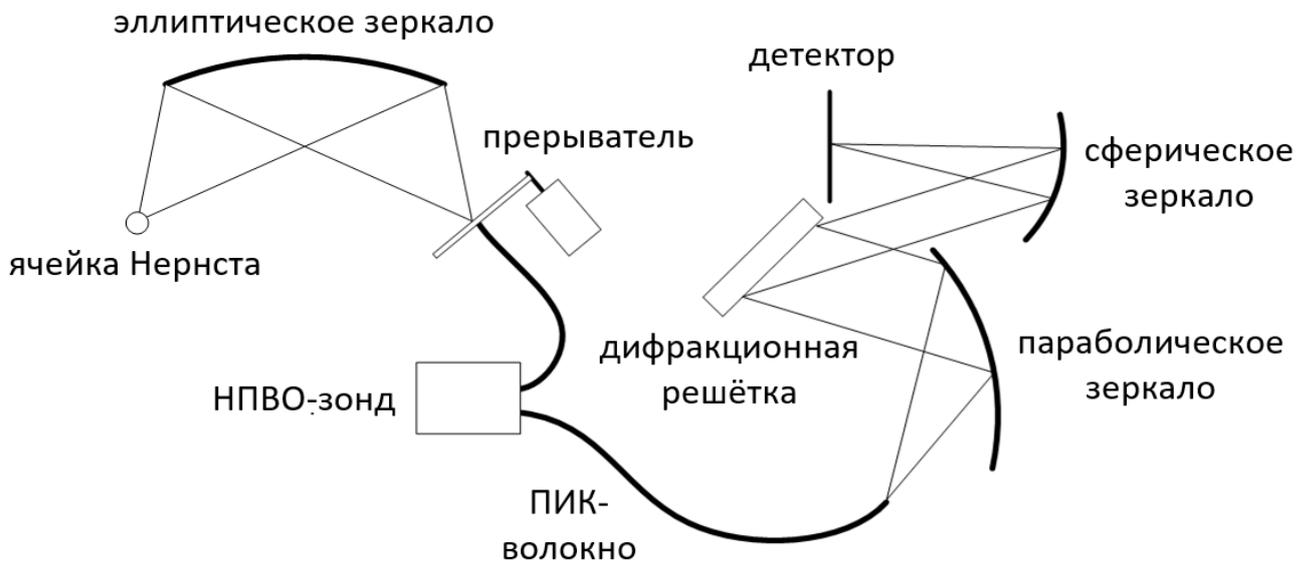


Рисунок 5-26. Схема ИК-ОМС на основе дифракционной решётки (GrS) [65].

Разрабатываемые прототипы анализаторов были испытаны в условиях процесса анаэробной ферментация *S. cerevisiae*. Были проведены 4 загрузки процесса в лабораторном биореакторе ёмкостью 7 л, разработанном Ульмским техническим университетом («Ulm University of Applied Sciences», г. Ульм, ФРГ) [81]. Система управления автоматически поддерживает в реакторе заданные пользователем условия.

Сравнение статистики многоуровневой валидации градуировочных моделей (раздел 2.4.3) для этанола, глюкозы и фруктозы, построенных для различных прототипов ИК-анализаторов на EG- и FG-наборах образцов, приведены в табл. 5-8. Наиболее точный метод ИК-спектроскопии Фурье с алмазным НПВО (FTIR&DATR) зондом использован в качестве полноспектрального «золотого стандарта».

Таблица 5-8. Статистика градуировочных ПЛС-моделей для этанола, глюкозы и фруктозы с использованием спектров EG- и FG-наборов искусственных образцов, снятых различными прототипами ОМС в ИК-области.

Набор данных	Спектральная область	ЛП ^а	Градуировка		ППП		ПТН ^б	
			СКОГ	R ²	СКОПП	R ²	СКОП	R ²
<i>Этанол (EG-набор)</i>								
FTIR ^в -DATR ^г	2999–700 cm ⁻¹	2	5.52	0.984	7.16	0.975	5.41	0.980
FTIR-DATR ^д	1049–951 cm ⁻¹	2	4.55	0.989	5.79	0.983	5.10	0.983
GrS ^е -DATR	1082–1006 cm ⁻¹	2	5.18	0.986	6.67	0.978	5.63	0.979
GrS-LATR ^ж	1093–1015 cm ⁻¹	2	7.76	0.968	10.58	0.945	8.34	0.953
FPI ^з -LATR	1250–990 cm ⁻¹	2	7.48	0.970	9.74	0.953	5.75	0.978
NIR_2D13.2 ^и	1127–2090 nm	3	3.93	0.992	5.09	0.987	3.96	0.990
<i>Глюкоза (EG-набор)</i>								
FTIR-DATR	2999–700 cm ⁻¹	2	3.71	0.998	4.38	0.998	3.58	0.998
FTIR-DATR	1049–951 cm ⁻¹	2	5.47	0.996	6.66	0.995	5.53	0.995
GrS-DATR	1082–1006 cm ⁻¹	2	9.77	0.987	11.99	0.982	10.87	0.980
GrS-LATR	1093–1015 cm ⁻¹	2	9.12	0.989	11.95	0.982	11.62	0.977
FPI-LATR	1250–990 cm ⁻¹	2	8.30	0.991	9.87	0.988	5.96	0.994
NIR_2D13.2	1127–2090 nm	3	6.29	0.995	8.06	0.992	5.79	0.994
<i>Фруктоза (FG-набор)</i>								
FTIR-DATR	2999–700 cm ⁻¹	3	0.79	1.000	1.81	0.998	1.41	0.998
FTIR-DATR	1049–951 cm ⁻¹	3	0.96	0.999	1.14	0.999	1.11	0.999
GrS-DATR	1082–1006 cm ⁻¹	3	2.21	0.996	3.82	0.990	3.71	0.987
GrS-LATR	1093–1015 cm ⁻¹	3	11.07	0.906	17.10	0.793	16.45	0.739
FPI-LATR	1250–990 cm ⁻¹	3	7.23	0.960	13.54	0.870	8.43	0.931
FPI-LATR_o ^к	1127–2090 nm	3	2.92	0.993	6.64	0.968	1.58	0.998
NIR_1D15.2 ^л	1132–2086 nm	4	11.99	0.886	19.47	0.725	14.57	0.797
<i>Глюкоза (FG-набор)</i>								
FTIR-DATR	2999–700 cm ⁻¹	3	2.17	0.999	3.50	0.998	1.78	0.999
FTIR-DATR	1049–951 cm ⁻¹	3	2.35	0.999	2.92	0.999	1.82	0.999
GrS-DATR	1082–1006 cm ⁻¹	3	3.76	0.998	6.76	0.994	4.03	0.997
GrS-LATR	1093–1015 cm ⁻¹	3	7.78	0.992	14.11	0.975	9.30	0.986
FPI-LATR	1250–990 cm ⁻¹	2	26.26	0.908	29.77	0.891	14.43	0.965
FPI-LATR_o	1127–2090 nm	2	8.03	0.991	9.65	0.988	10.45	0.982
NIR_1D15.2	1132–2086 nm	4	12.83	0.977	20.80	0.946	15.12	0.962

Примечания: ^а число латентных переменных в ПЛС-модели; ^б проверочный набор по дизайну d2_{0v1}²⁵ (рис. 3-2 б); ^в Фурье-ИК спектрометр «Matrix MF»; ^г НПВО зонд с алмазным рабочим элементом; ^д эталонный метод; ^е прототип спектрометра на основе дифракционной решётки (рис. 5-26) ^ж НПВО зонд с рабочим элементом в виде сменной ПИК-петли; ^з НИС Fabry-Pérot; ^и вторая производная БИК-спектров по Савицкому-Голею (ширина окна 13 точек, порядок полинома 2); ^к данные FPI-LATR после удаления выбросов – образцов FG02 и FG21; ^л первая производная БИК-спектров по Савицкому-Голею (ширина окна 15 точек, порядок полинома 2).

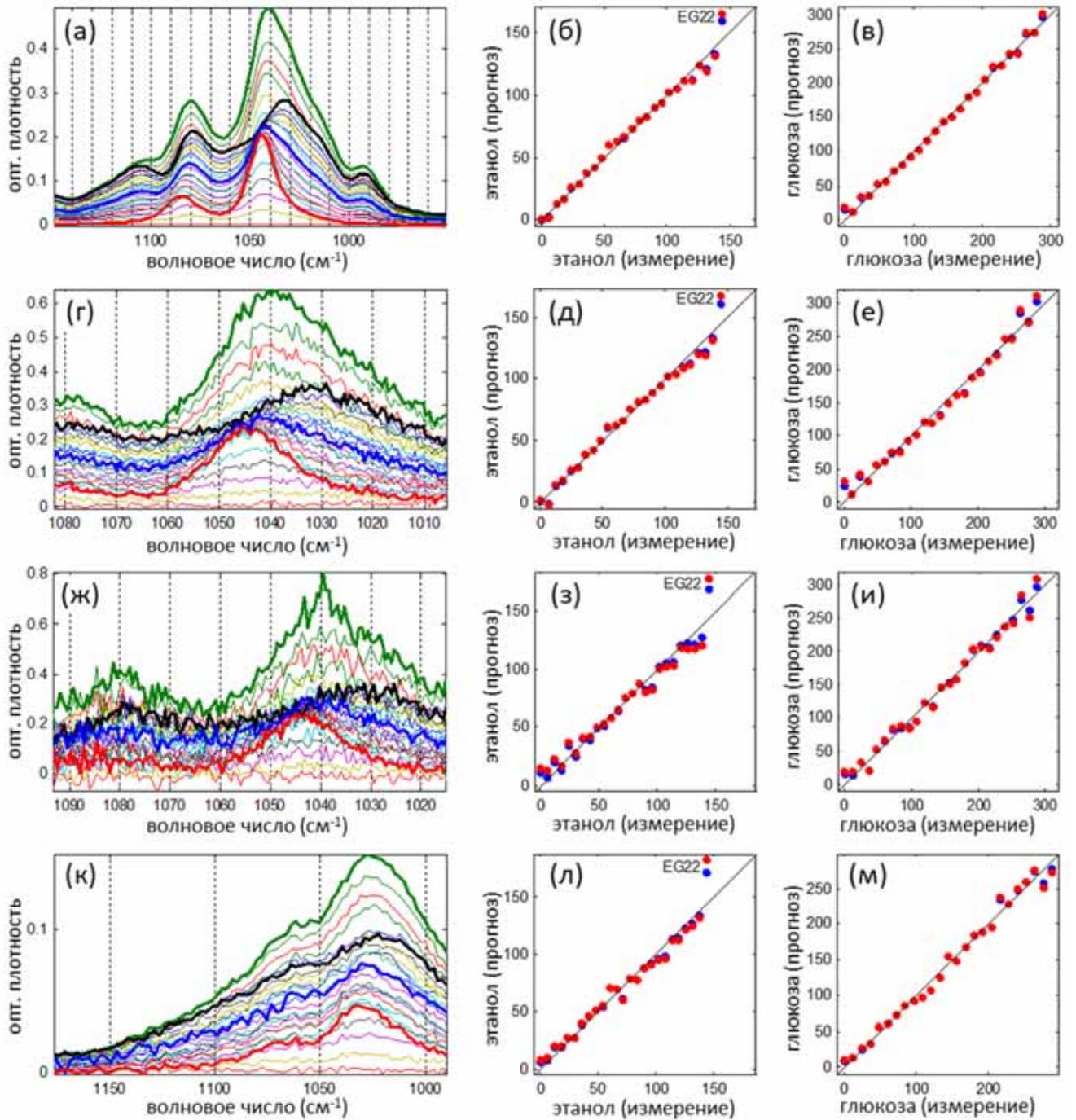


Рисунок 5-27. Спектры EG-набора образцов (левая колонка; характерные образцы выделены жирной цветной линией: зелёный – 24; черный – 22; синий – 0 и красный – 21; нумерация по рис. 3-6 и 3-2 б), полученные с помощью следующих методов: (а) FTIR-DATR, (г) GrS-DATR, (ж) GrS-LATR и (к) FPI-LATR (обозначения согласно табл. 5-8); средняя (б, д, з, л) и правая (в, е, и, м) колонки содержат соответствующие графики «введено-найдено» для концентраций (г/л) этанола и глюкозы (голубые и красные круги для SKOG и SKOPP соответственно) [65].

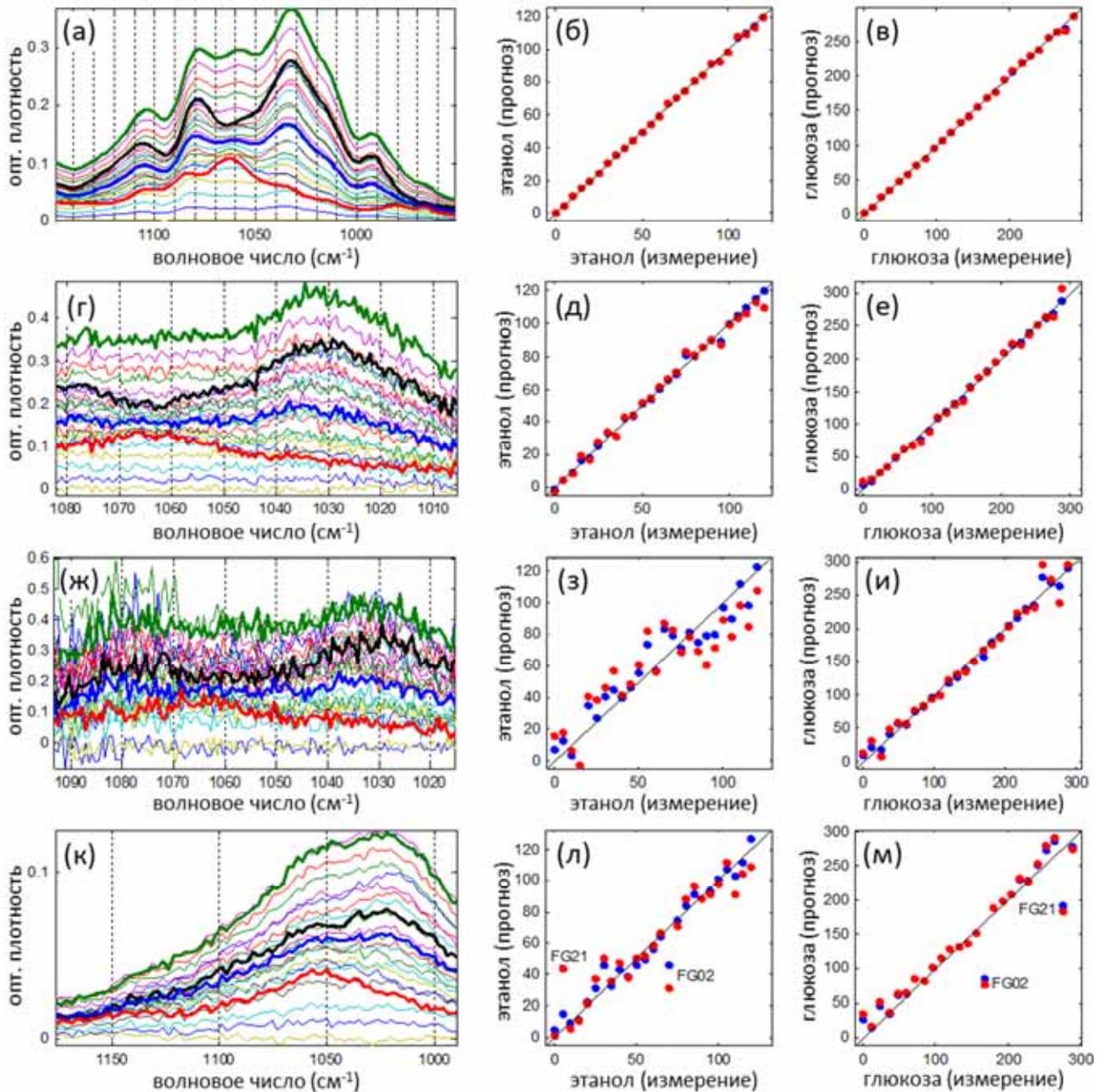


Рисунок 5-28. Спектры FG-набора образцов (левая колонка; характерные образцы выделены жирной цветной линией: зелёный – 24; черный – 22; синий – 0 и красный – 21; нумерация по рис. 3-6 и 3-2 б), полученные с помощью следующих методов: (а) FTIR-DATR, (г) GrS-DATR, (ж) GrS-LATR и (к) FPI-LATR (обозначения согласно табл. 5-8); средняя (б, д, з, л) и правая (в, е, и, м) колонки содержат соответствующие графики «введено-найдено» для концентраций (г/л) фруктозы и глюкозы (голубые и красные круги для *СКОГ* и *СКОПП* соответственно) [65].

Предварительный анализ полноспектральных данных [65] показал, что для анализа наиболее благоприятна область «отпечатков пальцев». Этанол имеет здесь отчётливые пики полос поглощения в районе 1090 cm^{-1} и 1050 cm^{-1} , относящихся к валентным колебаниям С–О и С–С, а также менее интенсивной полосой С–Н при

880 cm^{-1} . Глюкоза и фруктоза в этой области представлены сложной комбинацией плохо разрешённых (даже в Фурье-спектрах!) и трудно интерпретируемых полос. Наиболее интенсивные отличительные пики, наблюдаемые около 1060 и 1030 cm^{-1} (1057 и 1025 cm^{-1} согласно литературе), предположительно принадлежат валентному колебанию C–O D-фруктозы и D-глюкозы соответственно [267,268]. Для количественного анализа важно, что, несмотря на общее сходство, этанол, глюкоза и фруктоза имеют отчётливые спектральные черты (рис. 5-27 а и 5-28 а). Область 1149–951 cm^{-1} кажется наиболее подходящей для количественного анализа, проводимого в настоящем исследовании.

Действительно, ограничение модели выбранным интервалом не приводит к значимым потерям точности, а в некоторых случаях наблюдается снижение погрешности валидации (табл. 5-8). Модели полноспектрального «золотого стандарта» для этой области важны для оценки точности разработанных сенсорных систем, оперирующих практически в том же интервале волновых чисел. Достигнутую в этом аналитическом методе точность прогноза можно считать абсолютной. *СКОП* равные 4–6 г/л и 1–2 г/л для EG- и FG-наборов, соответственно, являются объективно низкими и достаточными для широкого спектра практических нужд. Несколько более высокая погрешность для EG, особенно для этанола (*СКОПП* = 5.79), объясняется присутствием выброса – образца EG22, наблюдаемого на рис. 5-27 б. В моделях на данных ОМС (рис. 5-27 д, з, л) прогнозируемое содержание этанола в EG22 также завышено, что свидетельствует скорее о проблеме приготовления образца, чем о спектральном выбросе. Все ПЛС-модели на выбранном интервале требуют лишь двух ЛП, что является для двухкомпонентной системы признаком надёжности модели. Однако модели серии FG используют три ЛП. Дополнительный фактор в этом случае может быть нужен для компенсации нелинейности, присутствующей в данных из-за насыщения спектральных откликов двух углеводов при их высоком содержании в растворе.

Замена Фурье-спектрометра разработанным в настоящем исследовании дифракционным сенсором (GrS) связана с потерей спектрального разрешения (сравните графики (а) и (г) на рис. 5-27 и 5-28). Другая отличительная особенность

данных сенсорной системы GrS-DATR – их шумный характер. Однако при ближайшем рассмотрении спектров на рис. 5-27 г и 5-28 г видно, что эффект лишь частично связан с шумом измерения. Некоторые острые пики имеют тенденцию к повторению от спектра к спектру и, следовательно, отражают естественные перепады чувствительности пикселей (ячеек) детектора, что не так значимо для многомерного анализа данных. Наблюдаемое снижение качества спектров и их селективности неизбежно отражается на точности многомерных регрессионных моделей. Так, *СКОПП* определения глюкозы вырастает до 12.0 г/л и 6.8 г/л в EG- и FG-наборах соответственно, в то время как модель для этанола страдает в минимальной степени (табл. 5-8). Тем не менее, значения R^2 показывают общую работоспособность моделей, что также подтверждается представленными на рис. 5-27 д, е и 5-28 д, е зависимостями «введено-найдено».

Отказ от алмазного элемента НПВО и его замена на петлю из ПИК-волокна ведёт к весьма существенному удешевлению ОМС. Но это упрощение отражается на качестве спектров (рис. 5-27 ж и 5-28 ж) и точности моделей (рис. 5-27 з и 5-28 з). Случайный спектральный шум усиливается и начинает доминировать над пиксельной структурой. Несмотря на двукратное отражение как в ПИК-волоконной петле, так и в алмазном элементе НПВО-зондов, значительное ослабление сигнала в ПИК-волокне объясняет наблюдаемое снижение качества спектров. Наибольший ущерб при этом наносится определению фруктозы: её *СКОПП* возрастает до 17.1 г/л (сравните с 1.1 г/л в эталонном случае FTIR-DATR). По всей видимости, более выраженное спектральное сходство фруктозы с глюкозой, чем с этанолом, делает её градуировку в смеси более чувствительной к качеству спектров. Так что даже при постоянстве спектрального разрешения в системах GrS-DATR и GrS-LATR переход к петельному ПИК-элементу определяет дальнейшее ухудшение моделей. Но несмотря на существенные потери по сравнению с «золотым стандартом», точность прогноза в наборах EG и FG, достигаемая сенсором GrS-LATR (*СКОПП* = 10–15 г/л для всех анализируемых компонентов за исключением фруктозы), остаётся практически приемлемой для ряда приложений мониторинга процессов, например в пищевой промышленности. Разработанная ОМС способна

определять различные углеводы, присутствующие в смеси, независимо друг от друга, что даёт несомненное преимущество по сравнению с существующими малобюджетными, к примеру, рефрактометрическими анализаторами. Можно сделать вывод, что многомерный анализ данных способен в значительной степени компенсировать понижение спектральной селективности, связанное с упрощением конструктивных составляющих анализатора. Основным источником ошибки ОМС на основе дифракционного спектрометра является скорее эксперимент и точность измерения, чем недостаточная селективность спектров.

Наконец, было важно сравнить две различные ИК-спектрометрические технологии, активно развивающиеся в наши дни – НИС-интерферометр Fabry-Pérot (FPI) и основанную на детекторе «PYREOS» дифракционную сенсорную систему GrS (раздел 1.1.3). Спектры, полученные с помощью зонда LATR в паре с FPI (FPI-LATR) представлены на рис. 5-27 к и 5-28 к. Хотя зашумлённость спектров значительно ниже, чем в дифракционных сенсорах GrS, разрешающая способность FPI очевидно является наиболее низкой среди всех рассмотренных здесь систем. Этот эффект частично объясняется практически двукратным падением номинального спектрального разрешения. Пики компонентов в спектрах разрешаются гораздо хуже, а отношения интенсивностей пиков поглощения характерных образцов, выделенных жирными линиями на рис. 5-27 к и 5-28 к, отличаются от других спектральных методов: см. графики (а), (г), (ж) на рис. 5-27 и 5-28. Некоторым преимуществом спектров FPI является более широкий интервал волновых чисел, покрываемый детектором, по сравнению с GrS. В данном приложении дополнительная спектральная область не приносит заметной выгоды, поскольку спектры не имеют информативных полос выше 1170 см^{-1} , поэтому она была исключена из анализа. Точности прогноза, достигаемые системой FPI-LATR в целом сравнимы с результатами GrS-LATR и даже несколько выше, если не считать глюкозу в наборе FG (СКОПП около 30 г/л), модель для которой страдает от двух сильных выбросов (рис. 5-28 м). В отличие от выброса по этанолу на рис. 5-28 б, д, з, л, вызванному ошибкой приготовления образца, выбросы в модели для глюкозы проявляются только в случае FPI и, следовательно, связаны с

отклонениями спектральной природы. Эти отклонения находят объяснение с точки зрения повышенной чувствительности FPI к внешним механическим воздействиям, являющейся известной технической проблемой этой технологии. После удаления вышеуказанных выбросов модель FPI-LATR оказывается даже несколько лучше, чем GrS-LATR (табл. 5-8). Следует отметить, что модель для глюкозы в наборе FG построена с двумя ЛП, а следовательно, она проще и надёжнее, чем в случае GrS.

Результаты анализа наборов искусственных образцов в ИК-области были также подвергнуты сравнению с результатами измерения методом БИК-спектроскопии. Хотя БИК-спектроскопия и не принадлежит к наиболее прецизионным спектральным методам анализа водных сред, её широкое применение в анализе биотехнологических объектов и процессов [8,247-249] делает такое сравнение интересным и полезным методологически. На БИК-спектрах EG- и FG-наборов искусственных образцов (см. дополнительные материалы к [65]) можно увидеть, что подавляюще высокое поглощение воды сильно осложняет определение других составляющих. Основная дисперсия в БИК-спектрах вызвана изменением сигнала воды. Такая непрямая корреляция с концентрациями растворённых компонентов смесей является нежелательной для их количественного определения, поскольку их характеристические пики сильно маскируются полосами воды, становясь едва отличимыми друг от друга и, следовательно, подверженными негативному воздействию ограниченного динамического диапазона спектрального измерения. Использование спектральных производных второго порядка в данных EG и первого порядка для FG-набора помогает подчеркнуть химические различия и, следовательно, увеличить точность моделирования. После такой интенсивной предобработки спектров, погрешности градуировок для набора EG становятся сравнимыми с эталонным методом FTIR-DATR, но за счёт одной дополнительной ЛП, указывающей на возможную нелинейность и сопряжённые с ней проблемы надёжности прогноза. Градуировочные модели для FG-набора демонстрируют практически неприемлемое качество прогноза (*СКОПП* около 20 г/л при 4-х ЛП), показывая неспособность БИК-системы различить глюкозу и фруктозу по их вялым

спектральным чертам в присутствии доминирующего поглощения воды.

Рассмотренные результаты убедительно показывают принципиальную возможность создания оптических мультисенсорных систем в ИК-области для анализа среды биотехнологического процесса на содержание этанола, глюкозы и фруктозы в типичных для ферментации концентрациях (табл. В-1, Приложение В). Сравнительный анализ данных трёх прототипов анализатора и эталонного метода, проведённый в соответствии с концепцией разработки ОМС (раздел 2.1.1) выявил преимущества и недостатки рассмотренных систем и основные причины потери точности при упрощении конструкции, что должно учитываться в дальнейшей разработке. Исследования также показали, что для поставленной цели ИК-спектроскопия в режиме НПВО в целом предпочтительнее БИК-спектроскопии на пропускание.

5.3.3 Мониторинг процесса ферментации посредством ИК- и БИК-спектроскопии

Одна из разработанных (раздел 5.3.2) ИК-сенсорных систем – дифракционный спектрометр с ПИК-петлёй в качестве рабочего элемента НПВО-зонда – была протестирована в условиях процесса ферментации *S. cerevisiae*. В ходе процесса параллельно регистрировались БИК-спектры среды методом, описанным в разделе 5.3.1. Были проанализированы две загрузки процесса, в ходе которых из биореактора периодически отбирались образцы. После окончания каждого процесса ферментации оставшийся в биореакторе питательный раствор модифицировался путём добавления расчётных количеств раствора глюкозы и этанола с целью расширения их концентрационных диапазонов для более надёжной градуировки. Добавленные количества глюкозы и этанола были подобраны таким образом, чтобы минимизировать корреляцию компонентов в соответствии с требованиями к планированию градуировочного эксперимента (раздел 3.2.1). Результирующие наборы, состоящие из 41-го образца, отобранных как в ходе процесса, так и по его окончании, были использованы для последующей многомерной градуировки. Дальнейшая информация по процессу ферментации приведена в статье [81] и дополнительных материалах к ней. Статистика

построенных методом ПЛС-регрессии градуировок для одной из загрузок приведена в табл. 5-9.

Таблица 5-9. Статистика градуировочных моделей, построенных методом ПЛС-регрессии для этанола, глюкозы и биомассы на ИК- и БИК-спектральных данных процесса ферментации.

Аналит	Пред. обраб. ^a	Спектр. область, нм/см ⁻¹⁶	ЛП ^b	Градуировка		ППП		ПТН ^г	
				СКОГ	R ²	СКОПП	R ²	СКОП	R ²
<i>БИК-спектрометр с трансфлексионным зондом^d</i>									
Биомасса	нет	1100-1860	2	0.43	0.952	0.47	0.946	0.48	0.928
Этанол ^e	2D2.15 ^ж	1100-1900	4	1.65	0.992	2.05	0.988	2.10	0.989
Глюкоза	2D2.15	1100-1900	4	3.08	0.995	3.71	0.993	3.96	0.993
<i>Дифракционный ИК-спектрометр с НПВО-зондом с ПИК-нетлём (GrS-LATR)</i>									
Этанол	нет	1093-1015	3	3.86	0.956	4.87	0.934	4.74	0.945
Глюкоза	нет	1093-1015	4	12.86	0.918	17.25	0.859	13.33	0.918

Примечания: ^a предобработка спектров; ^b нм для БИК- и см⁻¹ для ИК-спектров; ^b число латентных переменных в ПЛС-модели; ^г проверка тестовым набором (см. доп. материал к [65]); ^д к данным применялась предобработка, устраняющая влияние пузырьков газа на спектры [65]; ^e этанол и глюкоза определялись по результатам ферментативного метода [269,270] с УФ-фотометрией на 340 нм; ^ж вторая производная методом Савицкого-Голея (ширина окна 15, степень полинома 2)

Как БИК-анализатор, так и разработанный прототип ИК-ОМС показали удовлетворительную статистику моделей. В целом результирующая точность несколько хуже точности анализа искусственных образцов EG-набора (раздел 5.3.2). ПЛС-модели были проверены как методом ППП, так и при помощи проверочного набора из 15 проб процесса (см. дополнительный материал к [65]), выбранных в соответствии с правилами планирования градуировочного эксперимента (раздел 3.2), чтобы максимально полно покрыть концентрационные диапазоны компонентов, в то же время, избегая их минимальных и максимальных значений. Сравнению подлежат значения R² по ПП, поскольку различие концентрационных диапазонов этанола и глюкозы делает сравнение значений СКО соответствующих моделей некорректным. Анализ показал, что ПЛС-регрессия успешно описывает изменения как в ходе каждого отдельного процесса, так и разницу между двумя проведёнными загрузками. Модели для этанола и глюкозы включали большее число ЛП по сравнению с моделями на искусственных образцах EG-набора, а также по сравнению с моделями, построенными для каждой загрузки

отдельно, что объясняется большей сложностью и разнообразием реальных систем. Общее постоянство градуировочной и проверочной статистик при сравнении моделей (табл. 5-9) между собой, а также при их сравнении с ранее доложенными результатами для искусственных EG-смесей (табл. 5-8) подкрепляет наш вывод о возможности использования разработанного ИК-анализатора типа ОМС с НПВО-зондом на основе ПИК-петли для мониторинга процессов в биотехнологии в качестве разумной альтернативы более устоявшемуся методу БИК-спектроскопии (рис. 5-29).

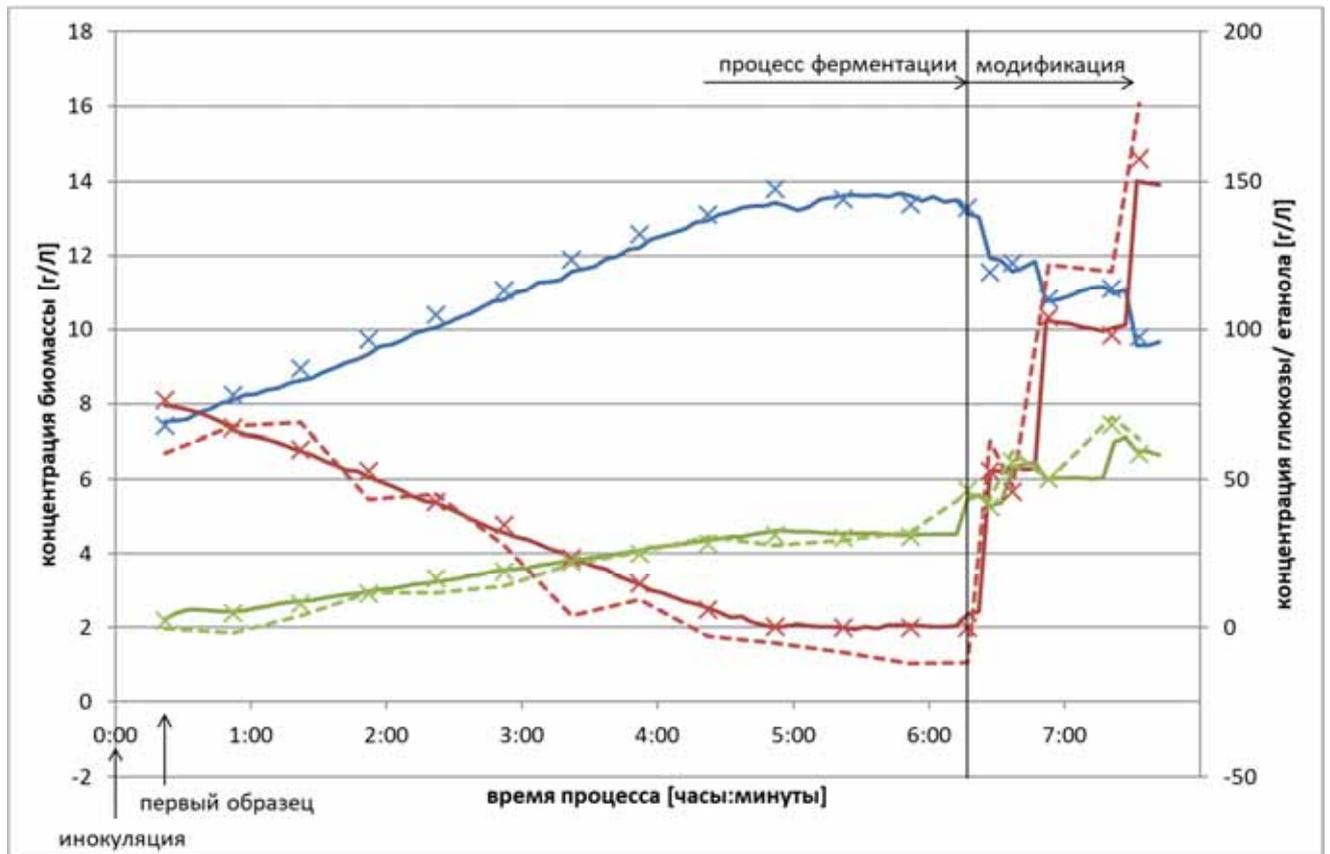


Рисунок 5-29. Траектории процесса ферментации: концентрационные профили биомассы (синий), глюкозы (коричневый) и этанола (зелёный) в ходе процесса и последующей модификации реакционного раствора. Кресты: концентрации определённые референтным анализом; сплошные линии: прогноз по БИК-спектрам, снятым каждые 5 мин; пунктир: прогноз по спектрам среднего ИК для градуировочных образцов [65].

Наблюдаемый в БИК-спектрах дрейф базовой линии в существенной мере связан с рассеянием света клетками дрожжей и, следовательно, включает информацию о содержании биомассы. Поэтому спектры для построения градуировочной модели для биомассы использовались без предобработки. При

использовании второй производной, как в моделях для глюкозы и этанола, градуировка по биомассе не строится. Модель для определения биомассы проста (две ЛП), а её точность ($СКОПП = 0.47$ и соответствующий $R^2 = 0.946$) доказывает практическую состоятельность метода, в частности, как дополнительной возможности БИК-спектрального анализа процесса ферментации. ИК-спектры НПВО не чувствительны к биомассе из-за малой глубины проникновения ИК-излучения в образец.

Малобюджетная комбинация сенсора «PYREOS» на основе дифракционной решётки с НПВО-зондом, включающим измерительный элемент в виде петли ПИК-волокна, пока проигрывает в точности более традиционной БИК-спектроскопии, особенно при определении глюкозы. Однако гораздо более низкая цена мультисенсорного ИК-анализатора делает оптические методы конкурентоспособными в практическом смысле, в особенности принимая во внимание тот факт, что технический прорыв в ИК-сенсорике только начинается. Способность ИК-ОМС с зондами НПВО различать углеводы в типичной ферментационной смеси является ещё одним значительным преимуществом метода по отношению к традиционной БИК-спектроскопии в целом.

5.3.4 Флуориметрический мониторинг процесса ферментации

Разработаны и опробованы два прототипа ОМС на основе флуоресцентной спектроскопии с использованием зонда «LHP» (раздел 1.1.3) для онлайн-мониторинга процесса ферментации дрожжей *S. Cerevisiae*. Акцент исследования был сделан на анализе данных флуориметрии в присутствии интенсивных помех и на эффективное использование полной спектральной информации, в том числе нефлуоресцентных сигналов. В частности, методами ПЛС-регрессии и МРК исследован эффект присутствия в спектре флуоресценции сигнала внешнего источника света. Предложенный нами [81] четырёхступенчатый алгоритм разрешения из серии СВЭ откликов индивидуальных флуорофоров и их суммарных концентрационных профилей был применён к анализу данных мониторинга процесса ферментации. Процессы проводились в лабораторном реакторе, описанном в разделе 5.3.3. Спектры регистрировались

спектрофотометром «MCS-CCD» от «J&M Analytik AG» (табл. 5-1) с шагом 1 нм.

В первом исследовании онлайн-спектры флуоресценции, возбуждаемой лазерным светодиодом на волне 365 нм, регистрировались каждые 45 мин. В ходе процесса каждые 5 мин регистрировались данные сенсоров оксигенации (pO_2), кислотности, температуры и рассеяния (по оптической плотности на 860 нм). Входящий в спектрометр сигнал проходил через оптический фильтр, отсекающий область ниже 450 нм для удаления пика полосы возбуждения. Результирующие данные показаны на рис. 5-30. Из среды процесса периодически отбирались образцы (50 мл) для определения глюкозы и биомассы референтным методом.

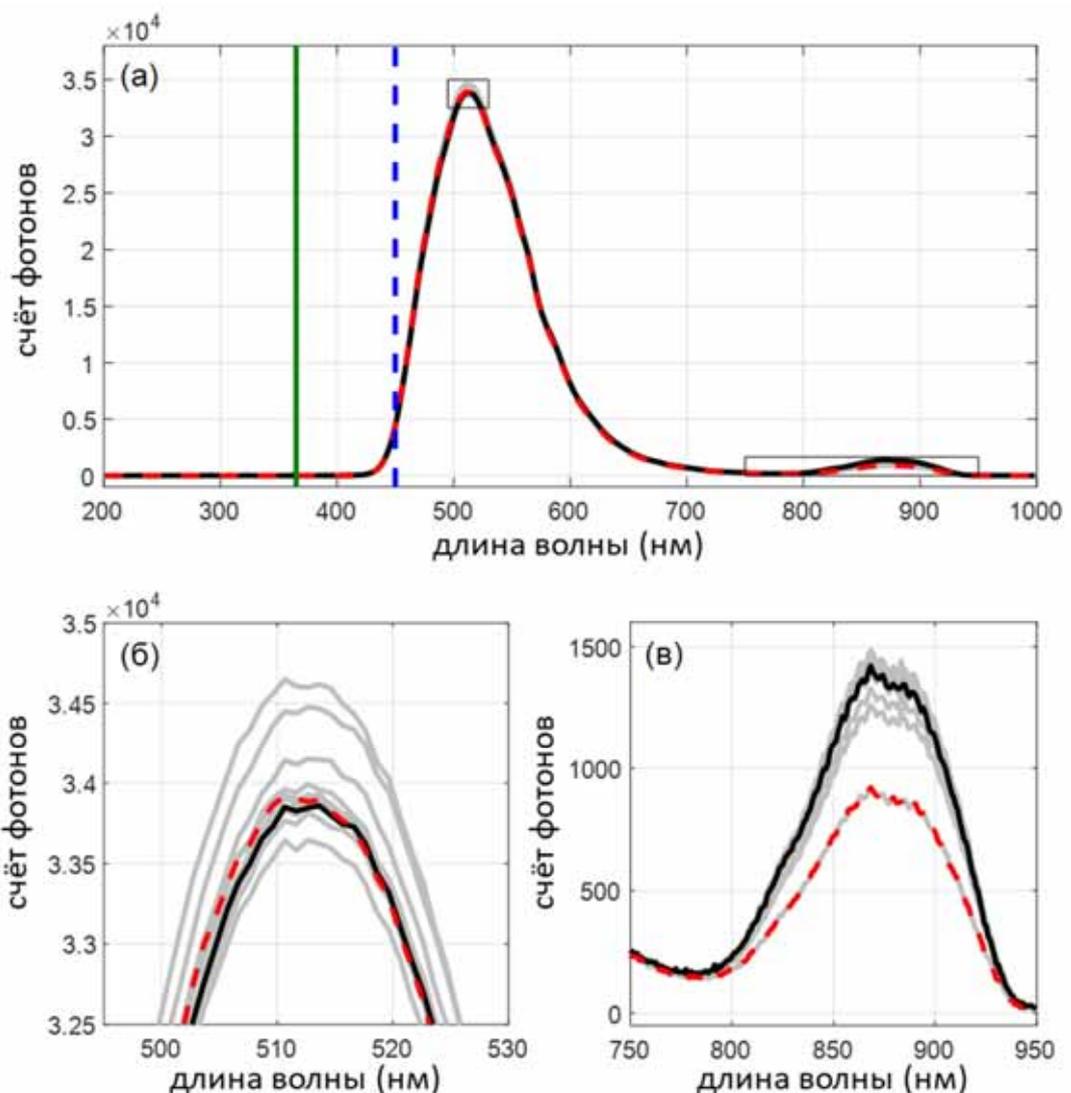


Рисунок 5-30. Онлайн-спектры флуоресценции процесса ферментации *S. cerevisiae*: (а) полные спектры; увеличенные полосы (б) флуоресценции и (в) излучения сенсора рассеяния. Зелёная линия обозначает длину волны возбуждения, пунктирная синяя – длину волны отсечения оптического фильтра. Выделены спектры: в начале (красная пунктирная линия) и в конце процесса (чёрная линия); остальные спектры показаны серыми линиями [81].

Основной детектируемый сигнал относится к флавинам (рибофлавин, ФАД), имеющим близкие максимумы поглощения на 360–370 нм и эмиссии на 520–530 нм [180,254,258]. Также ожидается некоторый вклад флуоресценции НАД в форме NADH на 450 нм. Все эти флуорофоры могут содержаться как в ферментационной среде, так и в клетках дрожжей. Флуоресценция среды доминирует из-за присутствия экстракта дрожжей в исходном инокуляте, поэтому в ходе процесса наблюдаются лишь небольшие изменения спектров испускания.

В дополнение к флуоресценции в спектрах наблюдался значительно менее интенсивный широкий пик при 870 нм, возникший из-за работающего зонда – сенсора биомассы по рассеянию БИК-излучения на единичной длине волны. Этого можно было бы избежать при другом расположении устройства. Однако, как будет показано ниже, эта случайность имела позитивное воздействие.

ПНК-МРК с граничными условиями неотрицательности (раздел 1.3.4) на полных спектральных данных дал 2 концентрационных профиля (рис. 5-31 а), интерпретация которых была сделана на основе соответствующих спектров (рис. 5-36 б). Присутствие сходных пиков на 530 нм в обоих разрешённых спектрах показывает, что оба компонента описывают изменение флуоресценции в ходе процесса. Единственное отличие спектров на рис. 5-31 б определяется присутствием в одном из них полосы сенсора биомассы. Из-за высокой дисперсности среды и значительного оптического пути (несколько сантиметров между зондами) интенсивность пика на 870 нм должна сильно зависеть от рассеяния, которое в свою очередь определяется количеством биомассы, как это и происходит в самом сенсоре биомассы. Следовательно, восходящий профиль на рис. 5-31 а, чей спектр содержит пики как флуоресценции, так и рассеяния, отражает внутриклеточное содержание биофлуорофоров в растущей культуре. Напротив, другой разрешённый спектр, содержащий только флуоресцентный сигнал, относится к компоненту, не зависящему от биомассы. Таким образом, нисходящий профиль на рис. 5-31 а объясняет изменения исключительно связанные с составом среды (например, соотношение флуорофоров), а также другие возможные химические и физические изменения, такие как вариации

ИНТЕНСИВНОСТИ ИСТОЧНИКА СВЕТА.

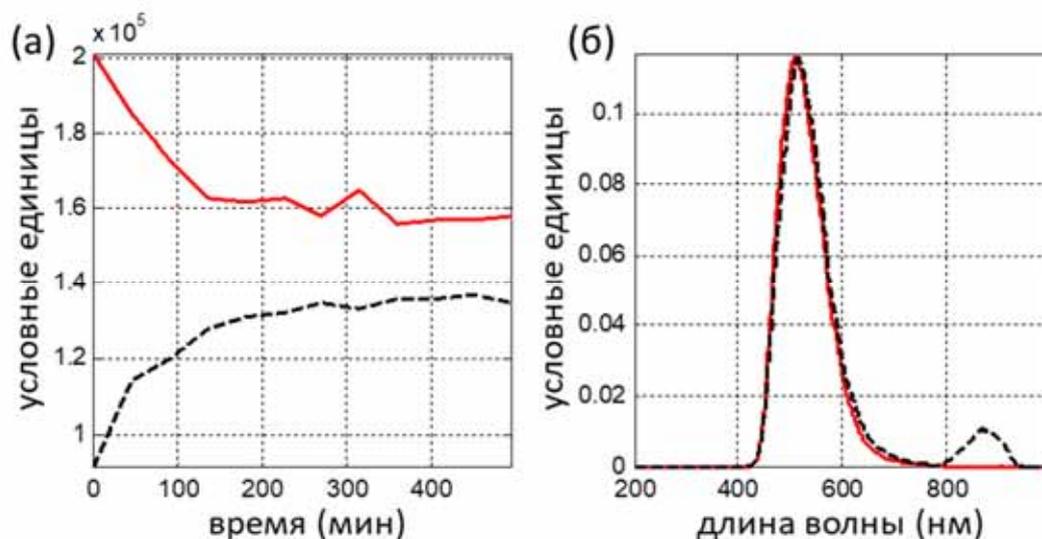


Рисунок 5-31. Результаты МРК на онлайн-спектрах флуоресценции: (а) разрешённые концентрационные профили и (б) чистые спектры двух компонентов, соответствующих чёрной пунктирной и красной сплошной линиям [81].

Без сигнала сенсора биомассы спектры компонентов (рис. 5-31 б) практически идентичны и не могут быть разрешены только на области флуоресценции 420–775 нм (так называемая проблема недостаточного ранга; *англ.* rank deficiency problem). Преимущества совместного анализа сигналов флуоресценции и рассеяния для определения биомассы подтверждаются статистикой градуировочных моделей (табл. 5-10).

ПЛС-регрессионные модели для градуировки биомассы были построены с использованием двух различных референтных наборов данных – счёта клеток (лабораторный метод) и онлайн сенсора биомассы по оптической плотности (онлайн-метод). В обоих случаях полноспектральные модели работали лучше, чем модели, построенные на областях только флуоресценции или только излучения сенсора. Это говорит о синергетическом эффекте совместного анализа данных флуоресценции и рассеяния.

Концентрация кислорода может также быть определена из данных флуоресценции. Это интересное наблюдение было сделано ещё в работе [258] и объяснено респираторным откликом клеток на степень аэрации, что отражается в интенсивности флуоресценции. Как и ожидалось, никакой значимой корреляции $\%pO_2$ с сигналом сенсора биомассы по оптической плотности не найдено.

Включение области сенсора 776–950 нм в анализ только ухудшает прогноз (табл. 5-10). Следует заметить, что проведённый здесь регрессионный анализ является скорее разведочным, чем количественным. Ограниченное число образцов и их принадлежность к одной загрузке процесса делает оценку ошибки ненадёжной. Однако этот подход отлично работает для сравнения моделей.

Таблица 5-10. Статистика градуировок различных референтных значений методом ПЛС-регрессии для различных спектральных областей.

Градуируемое значение	Параметр модели	Полный спектр 420–950 нм	Область флуоресценции 420–775 нм	Область сигнала сенсора 776–950 нм
СС ^а	$SKOPII/10^7$	1.029	1.252	1.396
	R^2	0.936	0.906	0.882
	Число ЛП	3	4	3
OD ^б	$RMSECV$	0.046	0.146	0.103
	R^2	0.988	0.706	0.854
	Число ЛП	3	3	1
%pO ₂	$RMSECV$	7.482	5.358	-
	R^2	0.956	0.977	-
	Число ЛП	4	3	-

Примечания: ^а счёт клеток биомассы по данным лабораторного метода (первый образец удалён как выброс); ^б данные оптического онлайн-сенсора; ^в данные сенсора на содержание растворённого кислорода.

5.3.5 Разрешение траектории процесса ферментации из данных 2D-флуориметрии

Мониторинг другого процесса ферментации дрожжей *S. cerevisiae* производился посредством 2D-флуориметрии. Двумерные спектры (СВЭ) были получены при 39-ти временах процесса (каждые 10 мин) и при следующих 24-х длинах волн возбуждения (максимумы пиков эмиссии, нм): 316, 323, 331, 339, 347, 354, 362, 371, 381, 389, 399, 408, 417, 424, 433, 441, 451, 460, 469, 475, 485, 494, 502, 512. Полная ширина полос эмиссии на половине высоты составляла около 20 нм. Для возбуждения люминесценции использовался решёточный монохроматор «FL3095» («J&M Analytik AG») с ксеноновой лампой. Лабораторный анализ образцов не проводился. Детальное описание эксперимента дано в работе [81].

Оптическая фильтрация сигнала в этом исследовании не использовалась, и в

спектрах присутствует сильная полоса рассеянного света от источника возбуждающего излучения. Флуоресценция видна в виде слабого и зашумлённого сигнала между 390 и 650 нм, наложенного на преобладающий (минимум в 400 раз) сигнал рассеянного света источника (рис. 5-32).

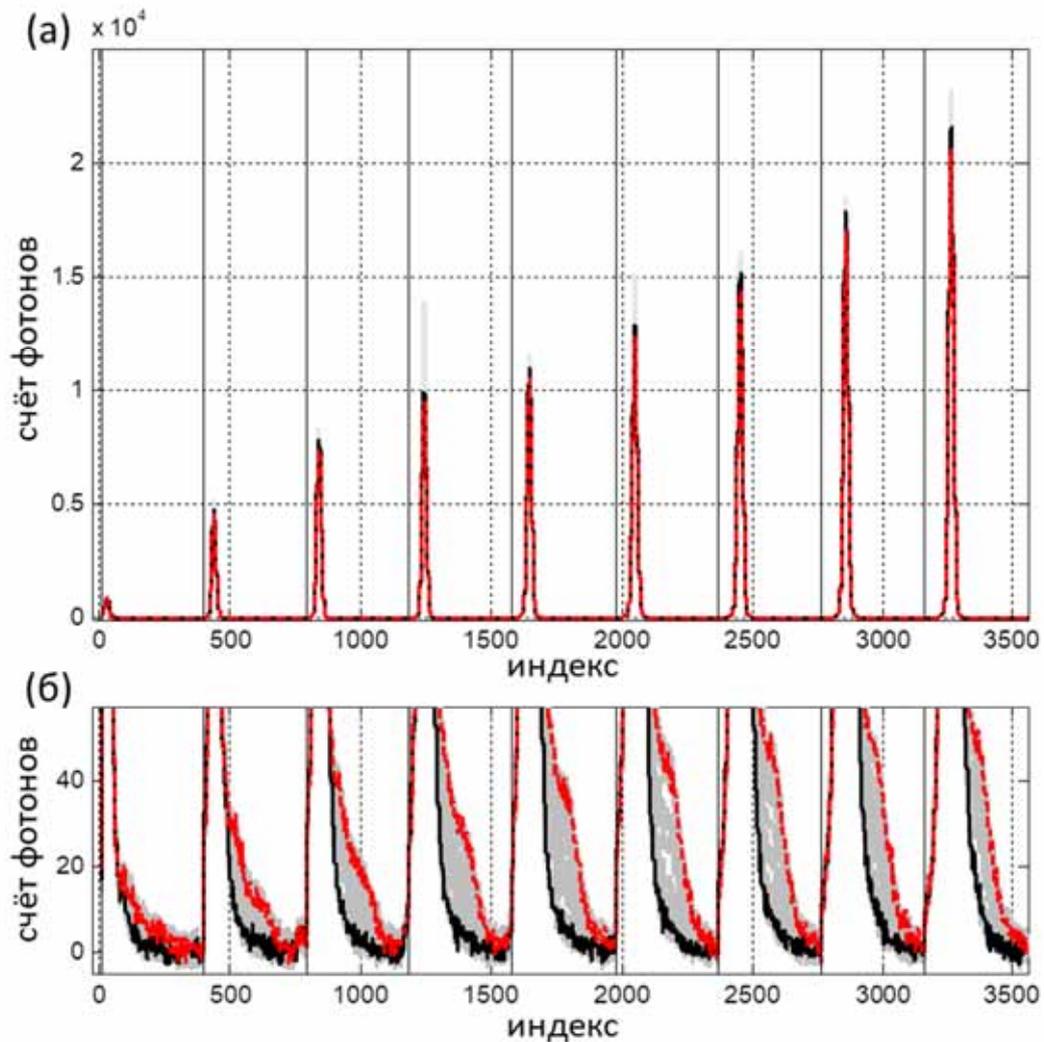


Рисунок 5-32. (а) Дополненные онлайн-спектральные данные при 9-ти длинах волн возбуждения и (б) увеличенная область сигнала флуоресценции; выделены первый (красная пунктирная линия) и последний (чёрная линия) спектры процесса [81].

Из-за сильного доминирования полосы возбуждающего излучения в спектрах исходные трёхмодальные данные процесса плохо пригодны для факторного анализа, а предобработка деструктивна для слабого сигнала флуоресценции. Для разрешения СВЭ-спектров отдельных флуорофоров был разработан специальный алгоритм [81].

Для совместного анализа использовались дополненные данные с общими объектами (осью времени) для 9-ти наиболее значимых длин волн возбуждения,

полученные методом развёртывания (раздел 1.1.3). Подготовленные таким образом данные показаны на рис. 5-32. МРК методом СИМПЛИЗМА показало наличие единственного фактора, связанного с изменением интенсивности флуоресценции, который и был разрешён в виде комплементарной пары «концентрационный профиль – объединённый 2D-спектр». Дальнейшая интерпретация разрешённой траектории показала сбой процесса из-за вероятной экспериментальной ошибки [81].

СВЭ, полученный обратной свёрткой разрешённых объединённых спектров флуоресценции на шаге 3 предложенного алгоритма, представлен на рис. 5-33 а. Очевидно, что, несмотря на общую траекторию, данные СВЭ включают отклики более, чем одного флуорофора, возбуждаемые на разных длинах волн. Второе применение МРК позволило из СВЭ разрешить чистые спектры возбуждения-эмиссии двух флуорофоров (рис. 5-33 б и в), максимумы которых указывают на присутствие NADH и группы флавинов [180,254,259].

Проведённые испытания различных оптических спектральных методов в 4-х процессах ферментации дрожжей *S. cerevisiae* показали перспективность развития ОМС на базе ИК-спектроскопии и флуориметрии (табл. В-1, Приложение В), в том числе двумерной. ИК-системы на основе современных систем детектирования – «PYREOS» и Fabry-Pérot способны определять этанол и углеводы с достаточной для практических целей (и с более высокой, чем у традиционной БИК-спектроскопии) точностью, причём различные углеводы, в частности глюкоза и фруктоза, могут определяться отдельно. Флуориметрия дала хорошие результаты в определении биомассы и оксигенации среды. То есть эти аналитические методы дополняют друг друга и могут быть использованы совместно для построения аналитического пространства траектории процесса.

Максимальное использование спектральной информации играет важную роль в разработке флуориметрических ОМС для онлайн-анализа процессов. Применение МРК позволяет достичь глубокого понимания самого процесса и информативности получаемых онлайн данных, что можно использовать для оптимизации оптических каналов разрабатываемых мультисенсорных систем.

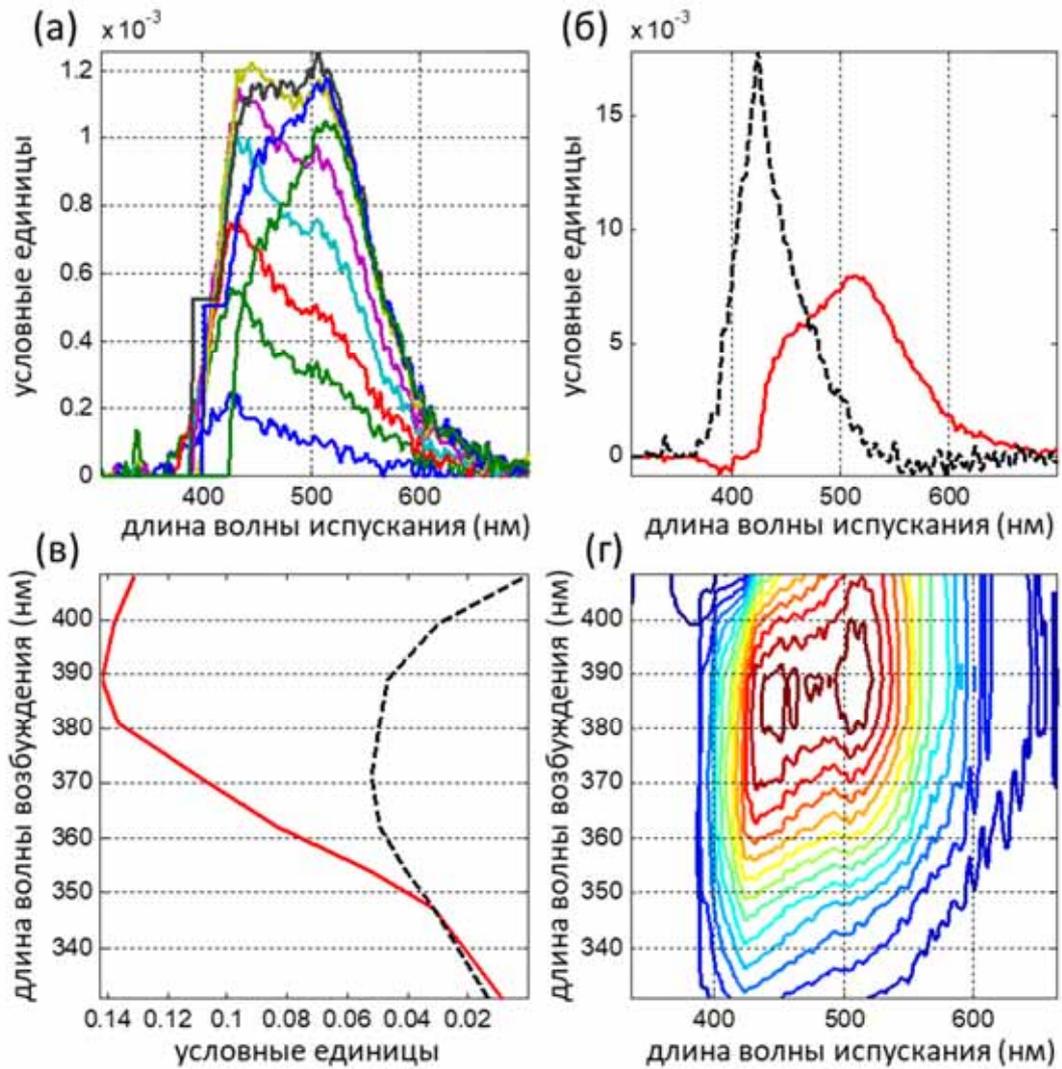


Рисунок 5-33. Разрешение спектров отдельных флуорофоров из двумерных спектров флуоресценции процесса ферментации дрожжей: (а) исходные спектры флуоресценции на 9-ти длинах волн возбуждения, разрешённые из трёхмодальных онлайн-данных; разрешённые из них методом МРК чистые спектры (б) эмиссии и (в) возбуждения отдельных флуорофоров; и (г) СВЭ, реконструированный из разрешённых чистых спектров [81].

5.4 Оптические методы диагностики рака почки

Исследование, описанное в этом разделе, посвящено разработке ОМС для определения границы опухоли при диагностике рака почки и прямой кишки у человека различными методами и поиску их эффективных комбинаций. В первом примере были разработаны и испытаны два прототипа светодиодной ОМС в спектральной области 900–1500 нм и экспериментально проверена возможность распознавания патологически изменённой ткани на четырёх выбранных длинах волн. Во втором примере исследованы диагностические возможности комбинации флуоресцентной и ИК-спектроскопии и её эффективность по сравнению с

индивидуальными методами. Обе разработки протестированы с использованием клинических образцов нормальной и поражённой ткани пациентов.

5.4.1 Методы, проблемы и задачи онкологической диагностики

Онкологические заболевания остаются основной причиной смертности во всём мире. Для улучшения ситуации необходимо проводить большое количество диагностических анализов на различных этапах профилактики и лечения заболевания: от превентивного скрининга до контроля эффективности лечения. Воплощение этой стратегии требует разработки информативных, лёгких в обращении и недорогих диагностических устройств для решения различных задач.

Оптическая спектроскопия биотканей, называемая также спектральной гистопатологией [271,272], является относительно новым подходом неинвазивной диагностики, активно развивающимся с 1990-х. Быстрое и недорогое спектральное измерение подходит как для профилактических проверок, так и для определения границы опухоли, например в ходе хирургической операции, представляя собой реальную альтернативу традиционной оперативной гистопатологии.

Неопластические процессы в тканях неизбежно ведут к химическим и физическим модификациям, которые могут быть обнаружены спектральными методами. Недавние исследования подтвердили возможность различить поражённую и здоровую ткань с использованием флуориметрии, КР-, ИК- и БИК-спектроскопии поглощения и диффузного отражения [272-274]. Развитие диагностической спектроскопии позволило лучше понять биохимию патологии на клеточном уровне, что в свою очередь стимулировало разработку новых методов диагностики.

БИК-спектроскопия диффузного отражения является одним из наиболее перспективных методов оптического анализа биотканей. Неся в себе смешанную информацию о поглощении и рассеянии электромагнитного излучения, она позволяет детектировать как молекулярные, так и морфологические изменения, связанные с трансформацией раковых клеток. Успешное применение БИК-анализа для обнаружения злокачественных опухолей показано для рака печени [275], груди [276,277], шейки матки [278], языка [279], простаты [280] и других органов.

Диапазон длин волн 900–1700 нм, покрываемый большинством современных БИК-спектрометров, содержит интенсивные полосы поглощения жира и воды – известных биомаркеров некоторых видов рака [277]. Важно, что БИК-излучение проникает в рассеивающий образец на глубину до 3 мм [281], чего не могут обеспечить другие спектральные методы. Другим важным преимуществом БИК-диагностики тканей является гибкость измерения, связанная с возможностью использования оптоволоконных кабелей и зондов, что важно во время операции. Литературное исследование не выявило примеров использования оптоволоконной БИК-спектроскопии в диагностике рака почки. Возможность замены БИК-спектроскопии оптической мультисенсорной системой, показанная для других приложений [26,95,171,282], представляет безусловный интерес и для онкодиагностики.

Флуориметрия – один из наиболее распространённых диагностических методов спектроскопии. В частности, она используется для выявления границы между нормальной тканью и опухолью [283,284]. Высокая чувствительность ко многим флуорофорам – биомаркерам рака [285] является основным преимуществом метода.

Менее чувствительная, но более селективная ИК-спектроскопия способна распознавать разнообразные органические молекулы и количественно определять компоненты сложных биологических образцов. Исследования показали большой диагностический потенциал метода в различных видах рака у человека [286-288].

Комбинация различных спектральных методов в одном измерении, способна значительно увеличить эффективность онкологической диагностики [289-291]. Большинство опубликованных работ посвящено созданию мультимодальных систем на основе флуориметрии в сочетании с методами спектроскопии диффузного отражения [272,291-294]. Методы БИК-, ИК- и флуоресцентной спектроскопии по отдельности или в различных комбинациях могут быть реализованы в виде оптоволоконного зонда; в случае ИК-спектроскопии – это зонд НПВО. Соединение двух и более типов спектральных данных далеко не всегда даёт синергетический эффект (раздел 1.1.3 и обзор [9]), поэтому эффективность такого

объединения всегда требует тщательной экспериментальной проверки.

5.4.2 Образцы

В исследовании использованы замороженные образцы неокрашенной биопсии семи пациентов, по образцу нормальной ткани и опухоли от каждого, полученные в результате нефрэктомии и предоставленные Институтом урологии университетской клиники «Шаритэ» («Charité», г. Берлин, ФРГ)⁷. Образцы размораживались непосредственно перед анализом в течение 5 мин при комнатной температуре. Классификация образцов по степени и стадии заболевания приведена в работе [27].

В первой серии измерений с первым прототипом светодиодного сенсора (серия А) исследовались 3 пары образцов от пациентов 191, 194 и 198. В следующей серии измерений (Б) с улучшенным прототипом сенсора исследовались 4 пары фрагментов почки – номера 144, 149, 151 и 160. Все исследованные образцы относились к типу светлоклеточной почечной карциномы. Образцы были разложены в чашки Петри (35 мм) и зафиксированы небольшим количеством клея «TRUGLUE®». На внешней поверхности дна чашки Петри была нанесена клетчатая разметка для обозначения позиции измерения. Примеры подготовленных образцов показаны на рис. 5-34. В тексте измерения обозначены кодом, состоящим из типа ткани («N» и «T» для здоровой и поражённой соответственно), номера пациента, позиции зонда (в соответствии с нанесённой сеткой: B2, C3, Cn и т. д.) и номера повторного измерения в той же позиции. Например, N144_Cn_5 означает измерение №5 в позиции Cn (дополнительная позиция в центре фрагмента) нормальной ткани пациента 144. В исследовании методами ИК-спектроскопии и флуоресценции [88] использовались только образцы серии Б.

Выбор подходящей комбинации спектральных методов и последующая разработка анализатора и аналитической методики включает несколько этапов измерений с клиническими образцами. С учётом природы образцов, их

⁷ Разрешение EA1/134/12 этической комиссии «Шаритэ».

ограниченной доступности и этических аспектов использования, промежуточные исследования, например сравнение методов, которые необходимы для принятия обоснованных разработческих решений, должны проводиться на минимально возможном наборе пациентов. Получение необходимых образцов биопсии в случае рака почки осложняется требованием, чтобы пары «здоровая ткань – опухоль» принадлежали одной и той же почке, то есть они могут быть получены только при полной нефрэктомии (при частичном удалении здоровая часть органа должна быть по возможности сохранена). Такие операции составляют только около 30% производимых в «Шарите» хирургических вмешательств при почечной карциноме⁸. Для увеличения информационного наполнения данных эксперимент был спланирован таким образом, чтобы охватить максимум реальной variability ткани.

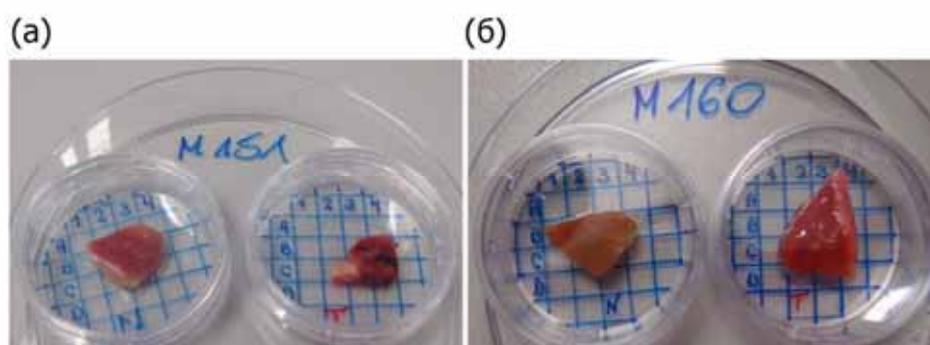


Рисунок 5-34. Образцы фрагментов почки пациентов (а) 151 и (б) 160; в каждом случае слева располагается здоровая ткань, а справа – опухоль [27].

5.4.3 Разработка и тестирование светодиодного сенсора для диагностики рака почки

Работа светодиодного сенсора (табл. 1-3 из раздела 1.2.6) основана на быстром сканировании нескольких (как правило, от двух до восьми) выбранных длин волн. Светодиоды на выбранных волнах попеременно облучают образец, а диффузно отражённый или пропущенный (в зависимости от геометрии измерения) свет улавливается и измеряется фотодиодным детектором.

В настоящем исследовании были выбраны 4 длины волны: 0.94 мкм

⁸ Личное общение с практикующими врачами клиники «Шаритэ» (г. Берлин, ФРГ).

(канал 4), 1.17 мкм (канал 3), 1.30 мкм (канал 1) и 1.44 мкм (канал 2). Длины волн первого и 4-го каналов были выбраны вблизи максимумов полос поглощения воды, а третьего – липидов [295]. Возможность распознавания раковой опухоли по соотношению концентраций вода-липиды (суммарно) показана в [277]. Четвёртый канал в относительно «пустой» спектральной области предназначался для коррекции рассеяния и служил своеобразным внутренним эталоном в соответствии с концепцией безэталонного измерения (раздел 2.3.5).

Измерения проводились при помощи вертикально установленного зонда в металлическом корпусе (рис. 5-35 а). В экспериментальной серии А зонд был в контакте с тонким (0.5 мм) кварцевым стеклом, защищающим образец. В серии измерений Б с улучшенным прототипом ОМС зонд был оснащён сапфировым окном, установленным под углом Брюстера ($\sim 60.5^\circ$) (рис. 5-35 б), и только острый конец зонда касался образца при измерении, то есть измерение проводилось на некотором расстоянии от непокрытой поверхности образца.

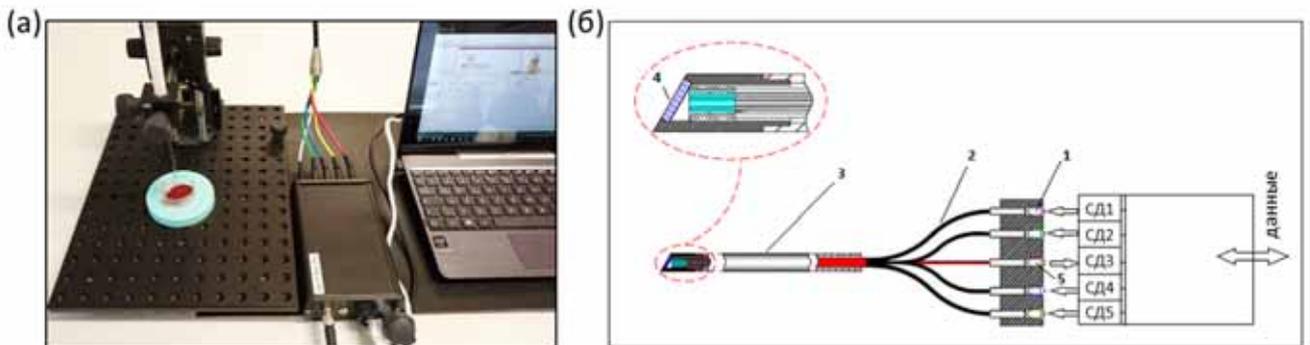


Рисунок 5-35. Экспериментальная установка (слева) и схема конструкции улучшенного прототипа светодиодной ОМС (справа): 1 – светодиоды, 2 – оптоволоконные кабели, 3 – трубка из нержавеющей стали, 4 – сапфировое окно (в конструкции первого прототипа для серии А зонд имел прямоугольную форму и не имел защитного окна) [27].

Данные двух прототипов ОМС (серии А и Б) сходны (рис. 5-36). Усреднённые сигналы имеют одинаковую форму с отчётливым минимумом на 1.44 мкм из-за поглощения воды. Канал 1 (1.30 мкм), напротив, даёт максимальную интенсивность. Вследствие бесконтактного измерения, общая интенсивность отклика в серии Б ниже.

Усреднённый сигнал образцов опухоли (рис. 5-36 в, г) имеет существенно более высокие интенсивности детектируемого сигнала на всех каналах, что

свидетельствует о более высоком рассеянии света поражённой тканью. Это объясняется тем, что раковые клетки имеют несимметричные увеличенные ядра с большим содержанием ДНК. Диаметр клеток опухоли может быть более 20 мкм, в отличие от 5-10 мкм для здоровых клеток. Предполагается, что в плотной клеточной структуре именно ядра ответственны за рассеяние света [296]. В целом биохимические и морфологические изменения поражённой ткани приводят к усилению её светорассеивающих свойств [297].

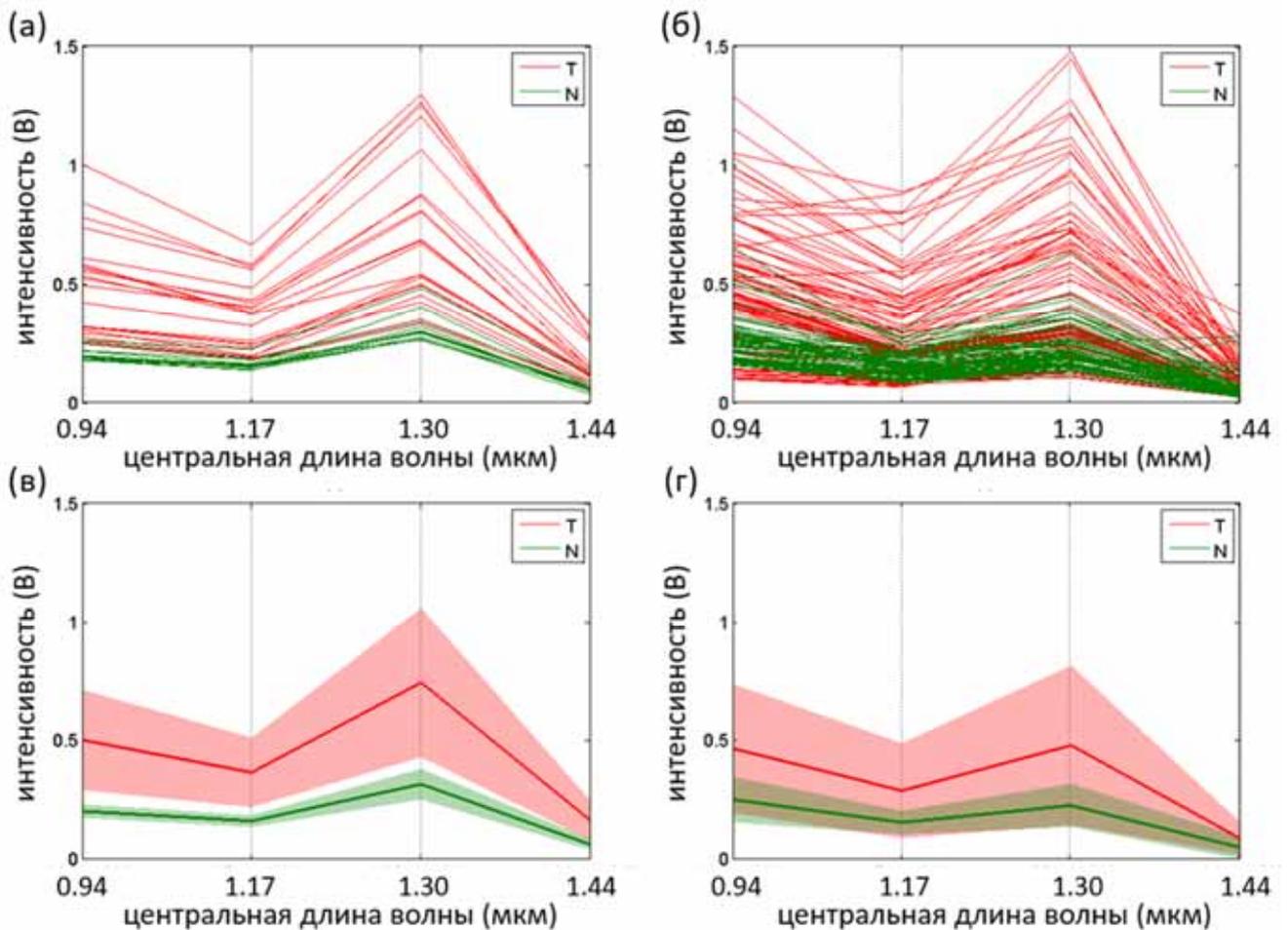


Рисунок 5-36. Данные светодиодной ОМС для диагностики рака почки: (а) серия А; (б) серия Б; и (в, г) средние значения (линия) и стандартные отклонения (залитая область) в сериях А и Б соответственно. Образцы опухоли и нормальной ткани показаны красным и зелёным цветом соответственно [27].

Другая очевидная особенность – гораздо более низкая вариабельность измерений на здоровых образцах. Простейшее объяснение этого факта – более упорядоченная структура здоровой ткани. Светлоклеточная карцинома имеет характерный жёлтый оттенок благодаря высокому содержанию липидов. На срез ткань выглядит сильно неомогенной, что часто видно невооружённым глазом.

Таким образом, интенсивность диффузно рассеянного БИК-излучения является важным для диагностики рака признаком. Это подтверждается тем, что применение методов коррекции рассеяния, например СНВ, или использование канала 1 (1.3 мкм) в качестве внутреннего эталона значительно ухудшало дискриминационную силу модели. Поэтому данные обеих серий А и Б анализировались без предобработки.

Разведочный анализ данных проводился методом МГК. Графики МГК-счетов (рис. 5-37) удобны для оценки общей дискриминационной способности аналитического метода. В серии А (рис. 5-37 а) разделение классов на плоскости первых двух ГК практически полное, за исключением пациента 194. Спектральное сходство образцов опухоли и нормальной ткани можно объяснить визуально наблюдаемым высоким содержанием в них жира. Несмотря на сильный разброс данных для образцов опухоли в МГК-пространстве, разделение классов убедительное и иллюстрирует преимущество мультисенсорного подхода по сравнению с одноканальным измерением. Ни один из отдельных каналов не даёт той же степени разделения классов. В серии Б (рис. 5-37 б) разделение классов хуже, что также частично связано с необычными оптическими свойствами отдельных образцов. Так, образец *T151* образует очень компактную группу измерений слева по направлению ГК1, хотя оно указывает на рак. Аномальное поведение образца может быть объяснено сильной неоднородностью опухоли с большим количеством некротических областей, кистозных изменений, сгустков крови и жира (рис. 5-34 а). Другие несоответствия на графике счетов на рис. 5-37 б связаны с отдельными позициями измерения. Так, результаты измерения *N144_Cn* являются типичными ложно положительными. Удаление вышеуказанных образца и позиции (около 18% всех измерений) значительно улучшает результат ПЛС-ДА, как описано ниже.

Дискриминация образцов проводилась методом ПЛС-ДА (раздел 1.3.6). Рак был принят за «позитивный», а здоровая ткань – за «негативный» результат анализа, численно закодированные как 1 и 0 соответственно. Факторное пространство ПЛС-ДА строится с учётом гистологического отнесения образцов, то

есть оно, в отличие от МГК, оптимизировано под диагностическую задачу. Статистика градуировки и ПП для серий А и Б представлены в табл. 5-11. Хотя качество дискриминации в серии Б хуже, полученные значения чувствительности, селективности и точности (раздел 1.3.7) указывают на общую пригодность светодиодных ОМС в БИК-области как вспомогательных анализаторов для распознавания опухоли. Учитывая большой размер и разнообразие серии Б, она лучше соответствует реальной вариабельности ткани почки у онкологических пациентов, так что данные не вполне подходят для прямого сравнения двух моделей ОМС.

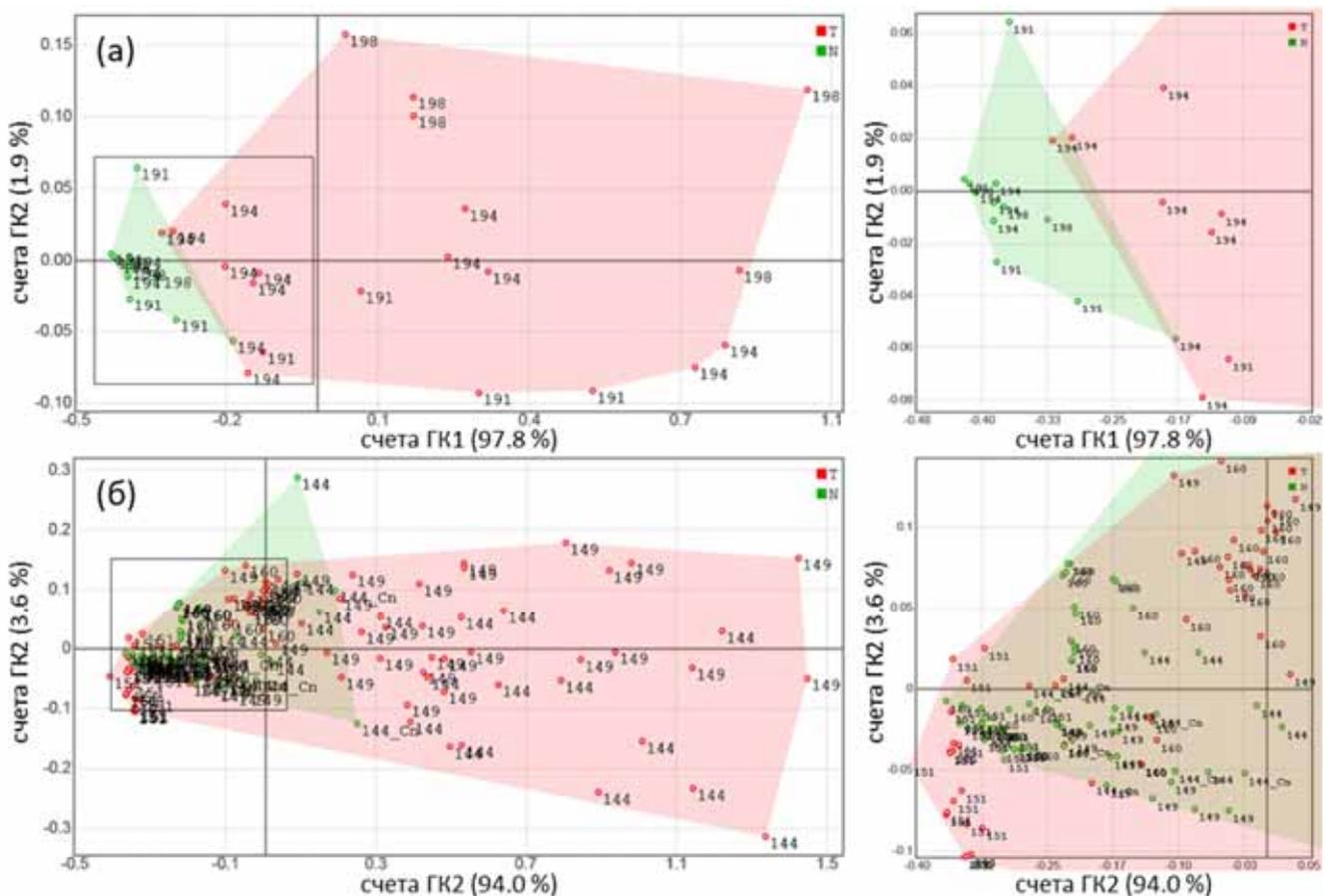


Рисунок 5-37. Графики МГК-счетов для моделей (а) серии А и (б) серии Б. Области, обозначенные прямоугольниками, показаны увеличенными справа. Красный и зелёный цвет обозначают опухоль и здоровую ткань соответственно. Залитые многоугольники обозначают области соответствующих классов [27].

Довольно высокая прецизионность измерений для отдельных образцов и позиций (даже для выбросов) свидетельствует о достаточном техническом качестве анализатора. Недостаток разделения классов в наборе Б вызван сложностью данных, которая не в полной мере передаётся прототипами 4-хканальных ОМС.

Основной проблемой остаётся недостаточная чувствительность (%Ч, табл. 5-11), вызванная большим разбросом измерений раковых образцов. Светодиодный сенсор для распознавания границы раковой опухоли – относительно новое приложение, находящееся в ранней стадии своей разработки. Успешная проверка концепции, проведённая в настоящем исследовании, даёт мотивацию для дальнейшей разработки оптической диагностической системы. На следующих её этапах необходима оптимизация оптической конфигурации – количества и рабочих длин волн светодиодов. Качество данных должно быть улучшено за счёт оптимизации конструкции зонда и геометрии измерения. Благодаря простоте и миниатюрности, светодиодная ОМС может составить серьёзную конкуренцию БИК-спектроскопии, особенно в медицинских приложениях. Возможность реализации в виде автономного устройства делает данный подход ещё более перспективным для хирургической и гистопатологической практики.

Таблица 5-11. Статистика диагностики рака почки методом ПЛС-ДА. Все модели построены для двух ЛП.

Набор данных	Градуировка							ППС ^а						
	ИП ^б	ЛП	ИО	ЛО	%Ч	%С	%Т	ИП	ЛП	ИО	ЛО	%Ч	%С	%Т
А33 ^в	19	1	11	2	91	92	91	18	1	11	3	86	92	88
В170 ^г	62	5	70	33	65	93	78	59	7	68	36	64	92	75
В140 ^д	64	3	67	6	91	96	94	63	3	67	7	90	96	93

Примечания: ^а перекрёстная проверка сегментами, образованными отдельными позициями измерения; ^б истинно положительные результаты отнесения и т. д. см. раздел 1.3.6; ^в полные данные серии А; ^г полные данные серии В; ^д данные серии В без образца T151, а также без позиций N144_Cn и T144_Cn.

5.4.4 Эффект комбинации флуоресцентной и ИК-спектроскопии

Для экспериментальной проверки эффективности мультиспектрального анализа ИК-спектры и спектры флуоресценции снимались в одних и тех же позициях образца (рис. 5-38).

ИК-спектры измерялись Фурье-спектрометром «Matrix MF», оснащённым НПВО-зондом на основе ПИК-волокна фирмы «art photonics» (табл. 5-1). Для снятия спектра сравнения использовался стерильный физиологический раствор (0.9% NaCl). Спектры были сняты с разрешением 8 см^{-1} при 64-х накоплениях.

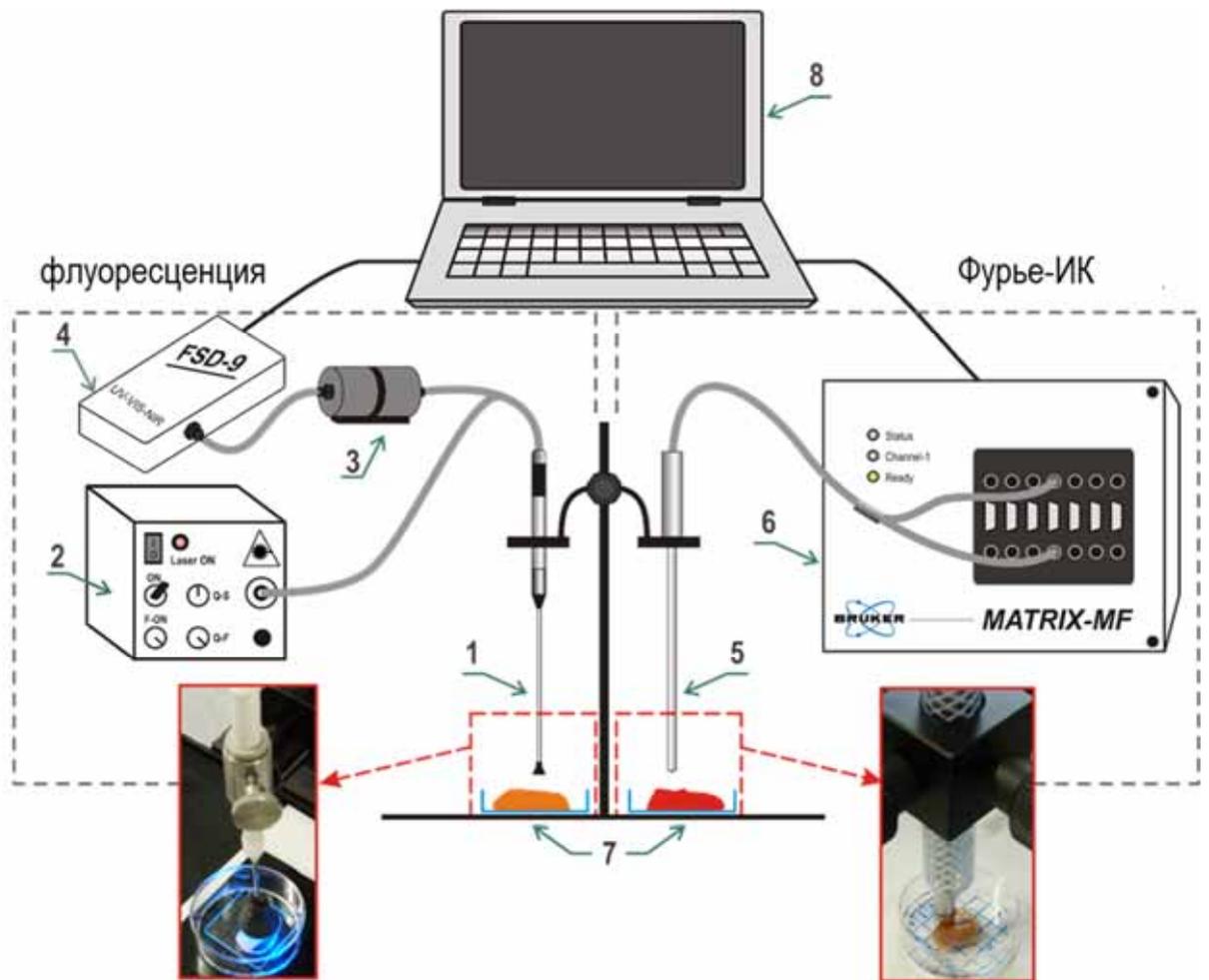


Рисунок 5-38. Схема комбинированного измерения методами флуориметрии (слева) и ИК-спектроскопии (справа): 1 – флуоресцентный зонд; 2 – лазер; 3 – оптический фильтр; 4 – Вид/КВБИК-спектрометр; 5 – НПВО-зонд; 6 – ИК-спектрометр; 7 – образцы; 8 – компьютер [88].

Спектры флуоресценции возбуждались на 473 нм лазером мощностью 25 мВт и регистрировались спектрофотометром «FSD-9» через иглообразный зонд, содержащий покрытое алюминием детектирующее оптоволокно, окружённое 13-ю облучающими волокнами. Время регистрации спектра индивидуально подбиралось для каждого образца в интервале 125–2000 мс, чтобы обеспечить интенсивность в оптимальном для спектрометра диапазоне. Повторные измерения в одной позиции проводились с тем же временем накопления. Подробное описание эксперимента опубликовано в работе [88].

Спектральные измерения 8-ми доступных образцов были проведены в каждой из 31-й (от трёх до пяти на образец) позиций, закодированных как на рис. 5-34. Таким образом, результирующие спектральные данные отражали изменчивость ткани как в пределах каждого образца, так и между ними. Результирующий

объединённый набор данных включал 92 спектра на 82 (ИК) + 736 (флуориметрия) переменных. Исходные спектральные данные показаны на рис. 5-39 а, б.

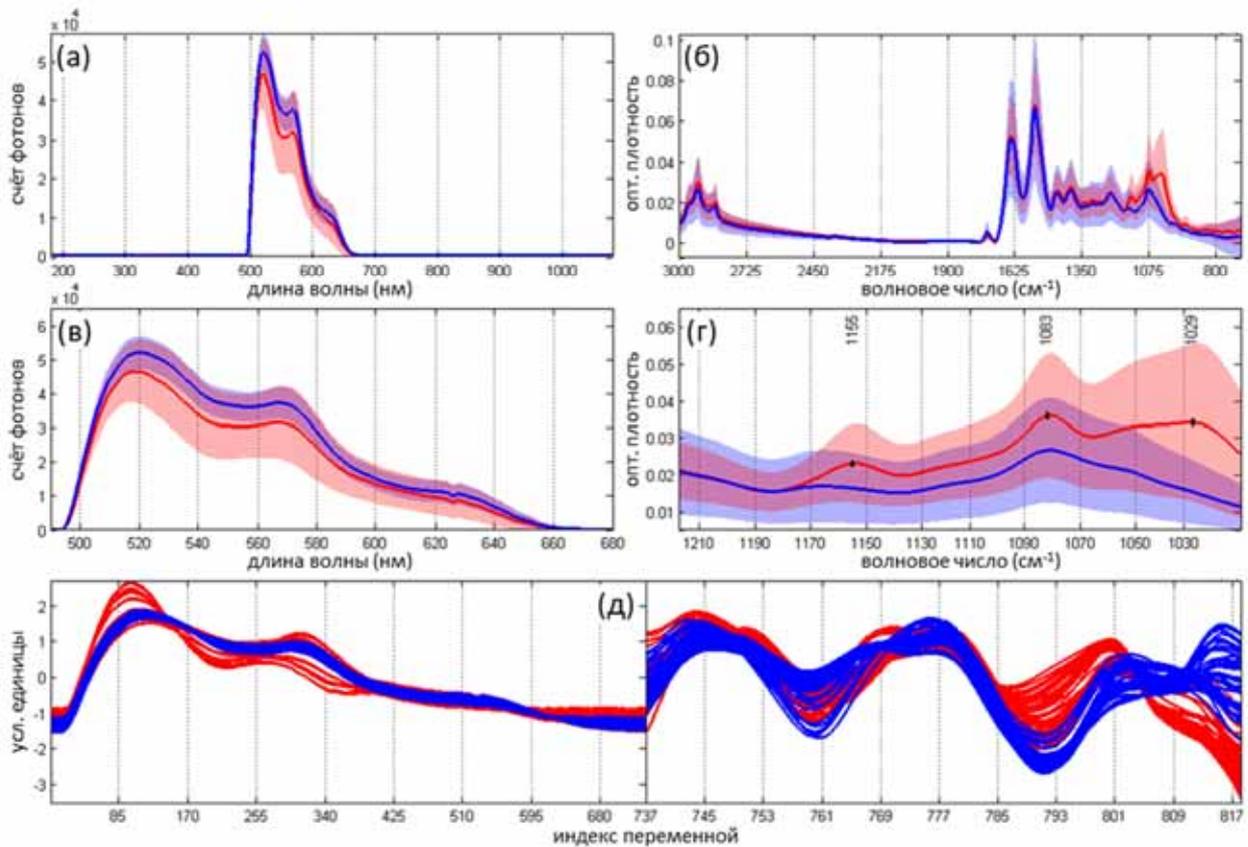


Рисунок 5-39. Спектральные данные мультиспектрального измерения: (а) исходные спектры флуоресценции; (б) исходные ИК-спектры; (в) флуоресцентные спектры в области 490–680 нм; (г) ИК-спектры в области 1220–1010 см^{-1} ; и (д) объединённый набор предобработанных спектров флуоресценции (слева) и ИК (справа). Применены следующие предобработки: СНВ для флуоресценции и 2-я производная по Савицкому-Голею с последующим СНВ для ИК-спектров; красный и синий цвет соответствует образцам опухоли и здоровой ткани соответственно [88].

Вторая спектральная производная по методу Савицкого-Голея использовалась со следующими предварительно подобранными параметрами: второй порядок полинома и сглаживающее окно в 25 точек. Перед конкатенацией предобработанные флуоресцентные и ИК-спектры были нормированы одним из двух методов: СНВ или автошкалированием (раздел 1.3.2).

Для анализа полные ИК-спектры (рис. 5-39 б) были сужены до области 1220–1010 см^{-1} , содержащей три пика характерные для поражённой ткани (рис. 5-39 г). Их интерпретация в терминах биохимии заболевания дана в работе [88]. Спектры флуоресценции (рис. 5-39 в) также были сужены до области наблюдаемого сигнала:

490–680 нм. Наиболее значимые флуорофоры в образцах биопсии почки – это ФАД (520 нм), коллаген (560 нм) и порфирины (630 нм). Внутритканевое выделение коллагена типов I и III при почечной карциноме отмечалось в работе [298]. Разница флуоресцентных сигналов здоровой и раковой ткани, по всей видимости, связана с изменением соотношения этих компонентов. Как следует из отнесения спектральных сигналов, методы флуоресцентной и ИК-спектроскопии способны нести комплементарную химическую информацию и, следовательно, подходят для совместного анализа.

Разведочный МГК-анализ на исходных спектральных данных показал значительно лучшее разделение классов «рак-норма» в случае комбинированных данных (рис. 5-40 а, в, д). Для сравнения дискриминирующей способности отдельных спектральных методов и их комбинации был использован следующий подход. Данные отдельных методов были проанализированы ПЛС-ДА с различными техниками предобработки: для флуоресценции – без предобработки или после СНВ; для ИК-данных – взятием второй производной (2П) или 2П+СНВ (2П с последующим СНВ). При объединении спектров обе части данных должны быть нормированы, чтобы уравнивать шкалы интенсивностей. Поэтому в случае объединения метод «без предобработки» не рассматривался. Обе части данных были нормированы СНВ или автошкалированием. Выбор лучшего метода основывался на точности (%T), а также статистиках градуировки и валидации. Результаты, представленные в табл. 5-12, позволяют сравнить диагностическую эффективность отдельных спектральных методов и их комбинации после индивидуально подобранной техники предобработки.

Статистики градуировки, ПП и проверка случайным набором (раздел 1.3.7) показывают сходные значения и порядок ранжирования методов, что свидетельствует об отсутствии переопределённости моделей. Риск переопределения всегда должен быть рассмотрен и исключён при моделировании на ограниченных наборах данных. Представленный набор данных достаточен для целей сравнения, но рабочая модель для диагностики должна строиться на гораздо большем числе пациентов, покрывающем естественную вариабельность образцов.

Примечательно, что модели как для отдельных, так и для комбинированных данных потребовали лишь две ЛП. Сохранение простоты модели при дополнении данных – признак комплементарности объединяемых блоков. Статистика табл. 5-12 показывает, что ПП сегментами – позициями измерения – наиболее консервативна, а значит, это наилучшая стратегия валидации. Проверка случайным набором (раздел 1.3.7), сочетающая в себе черты тестового набора и ПП, также хорошо подходит для малых объёмов данных.

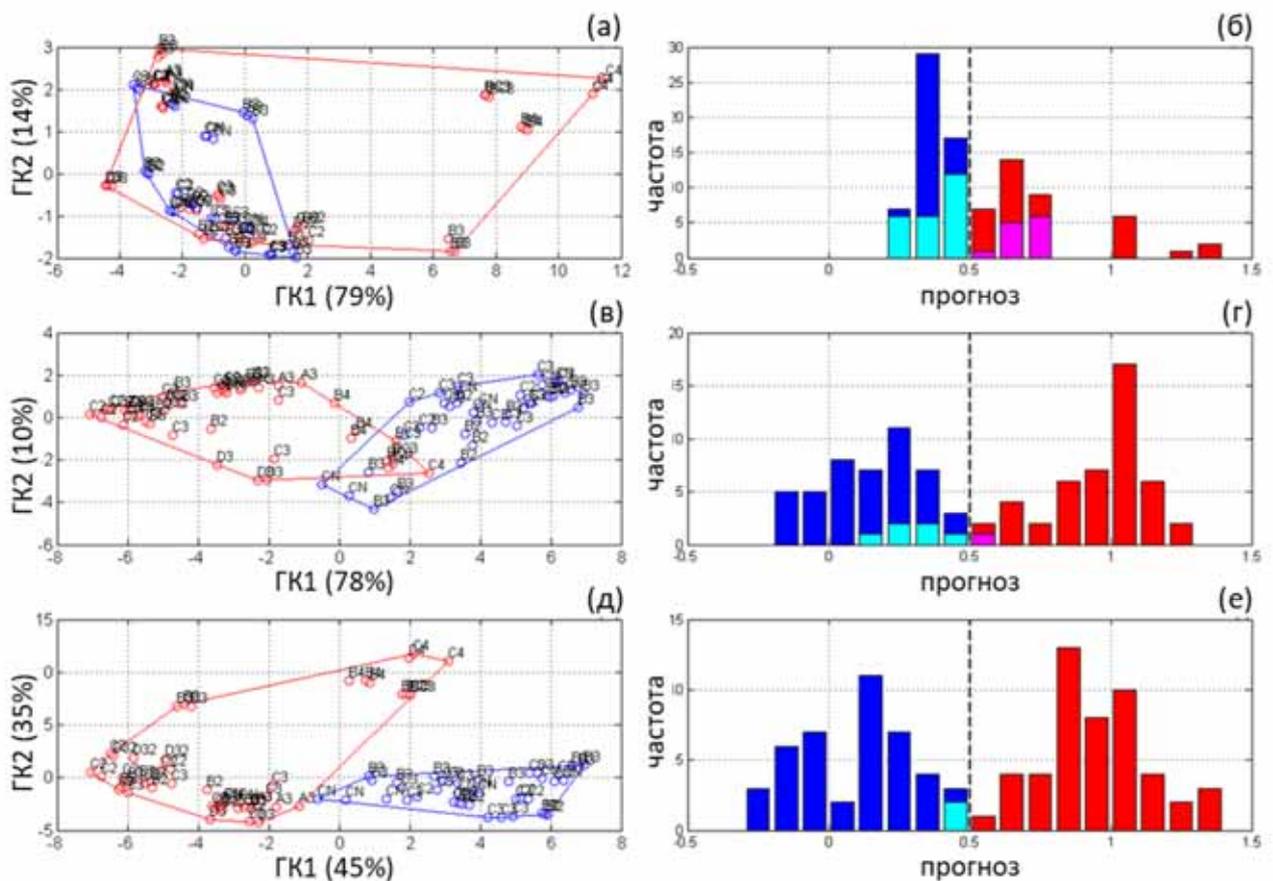


Рисунок 5-40. МГК- и ПЛС-ДА модели: (а, в, д) графики счетов; и (б, г, е) частотные гистограммы прогнозов ПЛС-ДА моделей (по данным ППС) для следующих наборов данных: (а, б) СНВ-нормированные спектры флуоресценции, (в, г) СНВ-нормированные вторые производные ИК-спектров и (д, е) объединённые данные. На (а, в, д) красный и синий цвета обозначают опухоль и здоровую ткань соответственно; метки обозначают позиции измерения; процентная доля объяснённой отдельными ГК дисперсии показана в скобках на подписях к осям. На (б, г, е): синий, голубой, красный и пурпурный цвета обозначают *ИО*, *ЛО*, *ИП* и *ЛП* результаты отнесения соответственно [88].

Данные флуориметрии сами по себе показывают более низкую дискриминационную способность, чем ИК-спектроскопия. Соответствующие им лучшие значения точности $\%T$ по результатам ПП составили 61% и 92% (жирный

шрифт в табл. 5-12). СНВ даёт некоторое улучшение моделей на спектрах флуоресценции, и напротив, приводит к наихудшим результатам для ИК. В последнем случае лучшим методом оказался 2П+СНВ (табл. 5-12). Неверная классификация представлена преимущественно нежелательными ЛО, то есть нераспознанным раком (рис. 5-40 г).

Конкатенация спектров флуоресценции и ИК-спектров (вторая производная) с последующей СНВ-нормировкой ведёт к снижению ложных классификаций до двух (из 92-х измерений), что соответствует точности $\%T = 92$ (табл. 5-12 и рис. 5-40 е). Это существенное улучшение по сравнению с результатами отдельных спектральных методов. Предобработка играет важную роль в достижении оптимального результата. Полученный синергетический эффект позволяет предположить, что методы флуориметрии и ИК-спектроскопии отвечают за различные биомаркеры. Из-за малой глубины проникновения ИК-спектроскопия преимущественно «работает» на клеточном уровне, в то время как флуоресцентный сигнал может приходиться с глубины нескольких миллиметров. Это показывает дополнительную комплементарность методов. Результат данного исследования обосновывает целесообразность их сочетания в одном аналитическом устройстве. При разработке мультимодальной ОМС ИК-спектрометр Фурье должен быть заменён мультисенсорным (например, НИС Fabry-Pérot, раздел 1.1.3). Регистрация спектров флуоресценции также может быть существенно упрощена.

В более поздней работе [199] представлено сходное исследование, показавшее синергетический выигрыш от комбинации флуориметрии и БИК-спектроскопии для диагностики границы опухоли при раке прямой кишки. В отличие от ИК-спектроскопии, БИК-область предоставляет бóльшую гибкость для разработки различных ОМС, включая светодиодные анализаторы (раздел 5.4.3).

Таблица 5-12. Сравнение моделей ПЛС-ДА для диагностики рака почки методами флуориметрии и ИК-спектроскопии; все модели построены с двумя ЛП.

Данные	Предобработка	ИП	ЛП	ИО	ЛО	%Т	%Ч	%С
<i>Градуировка^а</i>								
Флуоресценция	нет	37	20	21	14	63	73	51
	СНВ	32	12	29	19	66	63	71
ИК	нет	38	1	40	13	85	75	98
	СНВ	42	2	39	9	88	82	95
	2П ^б	45	4	37	6	89	88	90
	2П+СНВ	49	0	41	2	98	96	100
Флуоресценция ИК	АШ ^в АШ	39	13	28	12	73	76	68
	АШ 2П+АШ	44	8	33	7	84	86	80
	СНВ СНВ	48	0	41	3	97	94	100
	СНВ 2П+СНВ	51	0	41	0	100	100	100
<i>Перекрёстная проверка сегментами^г</i>								
Флуоресценция	нет	32	22	19	19	55	63	46
	СНВ	27	12	29	24	61	53	71
ИК	нет	38	3	38	13	83	75	93
	СНВ	42	4	37	9	86	82	90
	2П	45	5	36	6	88	88	88
	2П+СНВ	45	1	40	6	92	88	98
Флуоресценция MIR	АШ АШ	35	15	26	16	66	69	63
	АШ 2П+АШ	37	9	32	14	75	73	78
	СНВ СНВ	47	0	41	4	96	92	100
	СНВ 2П+СНВ	49	0	41	2	98	96	100
<i>Проверка случайным набором^д</i>								
Флуоресценция	нет					61	70	50
	СНВ					65	61	70
ИК	нет					84	75	95
	СНВ					88	83	94
	2П					89	87	91
	2П+СНВ					95	92	99
Флуоресценция ИК	АШ АШ					71	75	66
	АШ 2П+АШ					81	83	80
	СНВ СНВ					96	94	100
	СНВ 2П+СНВ					99	98	100

Примечания: ^а обучение модели и прогноз с использованием полного набора данных; ^б вторая производная по Савицкому-Голею (ширина окна 25, степень полинома 2); ^в автошкалирование; ^г 31 сегмент, образованный позициями измерения; ^д 1000 итераций со случайно выбранными измерениями (15% полных данных) с усреднением результата.

5.5 Экологический мониторинг почв

Аварийные разливы нефтепродуктов случаются на всех этапах добычи и переработки нефти. Анализ почвы на содержание нефтепродуктов необходим для своевременного обнаружения наличия и степени загрязнений, а также для оценки эффективности мер по их ликвидации. Периодическому экологическому контролю подлежат огромные территории. Остаточное содержание углеводородов нефти должно также контролироваться в ходе процессов ремедиации загрязнённых почв и при переработке нефтешламов. Существующие лабораторные методики определения суммарного содержания нефтепродуктов в почве (их обзор дан в работе [215]) продолжительны, зачастую требуют интенсивной пробоподготовки и не всегда обладают требуемой точностью. Работы по спектральному анализу углеводородов в почве немногочисленны и используют преимущественно режим диффузного отражения в БИК-области [299,300].

В работах [87,215] нами исследована возможность создания ОМС для лабораторного экспресс-анализа почв на содержание суммарных нефтепродуктов и воды методом ИК-спектроскопии с использованием НПВО-зонда.

На этапе разведочного анализа была исследована серия 57-ми незагрязнённых искусственных образцов почвогунта, приготовленных смешиванием в различных пропорциях 4-х компонентов: почвенного субстрата, глины, песка и доломитовой муки в соответствии с разработанным планом эксперимента, показанным на рис. 5-41 [215]. Для измерений образцы были стандартным образом спрессованы в таблетки. Измерения проводились на лабораторном Фурье-ИК-спектрофотометре «Avatar 360», оснащённом приставкой однократного НПВО «Smart Performer» фирмы «Thermo Nicolet» (табл. 5-1).

Интерпретация полученных спектров позволила изучить спектральные характеристики почв различного состава и сравнить их со спектрами нефти. На основе этих данных для количественного анализа общих углеводородов нефти (ОУН) была выбрана спектральная область $4000\text{--}1700\text{ см}^{-1}$ и использован оптимальный для неё НПВО-зонд фирмы «art photonics» на основе ХИК-волокна с пирамидальным кристаллом ZrO_2 , дающим два внутренних отражения. Для

проведения анализа зонд был подсоединён к тому же ИК-спектрофотометру через адаптер. При измерении зонд приводился в плотный контакт с таблеткой образца с полным погружением в неё кристалла.

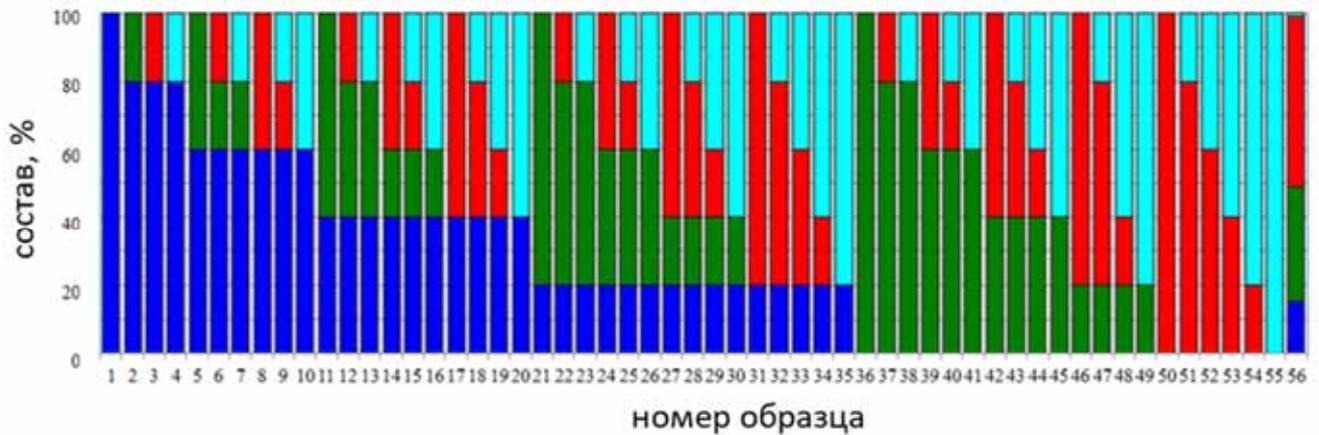


Рисунок 5-41. Состав образцов незагрязнённой почвы: синий – гумус, зелёный – песок, красный – глина, голубой – доломит [215].

Градуировочная серия загрязнённых образцов была приготовлена путём добавления нефти месторождения «Майорское» (Оренбургская обл., РФ) и воды к 100 г искусственного образца почвогрунта, по составу близкому к распространённому в Самарской обл. (15% почвенного субстрата, 50% глины, 34% песка и 1% доломитовой муки). Введение в дизайн эксперимента дополнительного фактора – содержания воды – было необходимо для проверки применимости метода в полевых условиях, когда влажность почвы может широко варьироваться из-за осадков и других природных условий. Обучающий набор из 25-ти образцов следовал диагональной схеме d_{0v1}^{25} (рис. 3-2 б), в которой оба компонента варьировались в диапазоне 1–13%.

Помимо градуировочной модели для ОУН на этом же наборе была построена модель для содержания воды, которое также может быть определяемым параметром, например при ремедиации нефтешлама. Наименьшие погрешности определения были достигнуты при использовании метода совместной оптимизации предобработки данных и отбора переменных (раздел 2.2.4): с использованием пяти трёхточечных интервалов с усреднением без предобработки для ОУН (рис. 5-42) и четырёх индивидуальных переменных после сглаживания данных по Савицкому-Голею с последующим применением СНВ для воды. Значения *СКОП* для

встроенного проверочного набора соответственно составили 1.1 и 0.6 при трёх ЛП (табл. 5-13).

Таблица 5-13. Статистика проверки градуировочных ПЛС-моделей для различных методов предобработки и отобранных наборов переменных.

Предобработка	ЛП	Градуировка		ППП		ПТН ^а	
		СКОГ	R ²	СКОПП	R ²	СКОП	R ²
<i>ОУН (%)</i>							
Исходные данные	3	1.66	0.787	1.99	0.699	2.09	0.656
СГО.2.15+МКР ^б	2	1.51	0.824	1.71	0.778	1.41	0.819
Отбор инт. с усреднением ^в	3	1.04	0.916	1.26	0.879	1.14	0.874
<i>Вода (%)</i>							
Исходные данные	2	2.67	0.453	3.25	0.275	3.00	0.290
СГО.2.15+СНВ ^г	4	0.97	0.927	1.44	0.846	1.21	0.864
СГО.0.5+СНВ+инт ^д	3	0.79	0.952	0.91	0.936	0.62	0.963

Примечания: ^а обучающий и проверочный наборы в соответствии со схемой ДД d_{0v1}^{25} (рис. 3-2 б); ^б сглаживание по Савицкому-Голею (степень полинома 2, ширина окна 15 точек) с последующим применением МКР; ^в отобранные интервалы (см⁻¹): 1740–1738, 2839–2837, 3329–3327, 3391–3389, 3449–3447 (оптимальная ширина интервала 3 точки с усреднением); ^г сглаживание по Савицкому-Голею (вторая степень полинома, ширина окна 15 точек) с последующим применением метода стандартизации нормировкой вариации; ^д сглаживание по Савицкому-Голею (нулевая степень полинома, ширина окна 5 точек) с последующим применением метода стандартизации нормировкой вариации; отобранные переменные (см⁻¹): 1662, 2796, 3182, 3209, 3240 (оптимальная ширина интервала – одна точка).

Основной экспериментальной проблемой метода является низкая общая интенсивность спектров более сухих образцов и плохая прецизионность, что вообще характерно для НПВО-анализа твёрдых сред. Предложенные оригинальные техники предобработки и анализа данных в значительной степени преодолевают это негативное влияние, и достигаемые погрешности определения являются приемлемыми для ряда практических задач. Повышение точности метода потребует новых экспериментальных решений, что и является целью дальнейшей разработки [301].

Полученные результаты позволяют рекомендовать дальнейшую разработку ОМС для полевого определения общих углеводов нефти на основе ИК-спектроскопии в области 4000–1700 см⁻¹ через оптоволоконный НПВО-зонд (табл. 1-В, Приложение В). Предложенный подход получил дальнейшее развитие в диссертационной работе Гурьяновой [301].

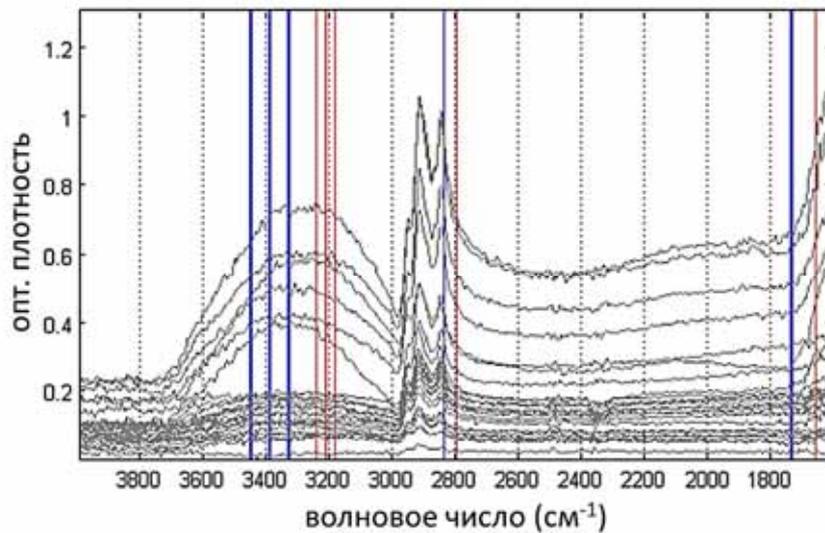


Рисунок 5-42. Отобранные оптимальные интервалы: красные – в модели для определения воды, синие – в модели для определения ОУН [87].

ЗАКЛЮЧЕНИЕ ГЛАВЕ 5

На большом экспериментальном материале показана принципиальная возможность перехода от полноспектрального измерения к ОМС с незначительной потерей точности анализа или без неё. Важную роль в достижении оптимального результата играет использование принципов разработки оптических мультисенсорных систем, изложенных в главах 2–4 настоящей диссертационной работы, а именно: правильный выбор спектрального метода или комбинации методов оптического анализа, математическая оптимизация оптических каналов будущего сенсора, в том числе геометрии измерения, планирование эксперимента, разведочный анализ данных для углублённого понимания объектов и методов анализа, а также максимально полное использование содержащейся в данных информации на этапе моделирования. Дополнительный выигрыш в точности анализа может быть получен при использовании в измерениях пространственного разрешения наряду со спектральным.

Предложены оптические мультисенсорные системы, обладающие небольшим размером, портативностью, лёгкой встраиваемостью в производственную линию при многократном снижении стоимости анализа по сравнению с лабораторной спектроскопией. Предложенные прототипы ОМС и методики расширяют круг сложных объектов анализа, что показано на примере определения индивидуальных компонентов и обобщённых показателей

химического состава различных объектов: натурального молока, твёрдых лекарственных форм, среды процесса ферментации, образцов биологической ткани и загрязнённых почв.

При помощи разработанных ОМС и соответствующих методик определения компонентов сложных смесей решён ряд аналитических задач полевого, экспресс- и онлайн-анализа. Низкоселективные светодиодные ОМС в видимой и КВБИК-области, способные определять содержание жира и общего белка в натуральном или нормализованном молоке с погрешностями менее 0.1%, создают альтернативу стандартным лабораторным методам ИК- и БИК-спектроскопии. Погрешность онлайн-определения воды в ходе сушки порошкообразных фармацевтических и пищевых продуктов в диапазоне 0-10% снижена до 0.1% (в три раза по сравнению с существующими подходами). Построена модель, которая из БИК-спектров, снятых онлайн в ходе процесса изготовления препарата целевой доставки (пеллет), прогнозирует будущую кинетику высвобождения ацетаминофена после приёма лекарства пациентом. Показано улучшение аналитических характеристик соответствующих методик при дополнении флуориметрии данными измерений в ИК- и БИК-области, что даёт более надёжную диагностику рака почки (снижение ложных отнесений – ошибок первого и второго рода) и повышение точности определения биомассы в ходе ферментации (снижение погрешности на четверть и более). Разработана улучшенная методика зондирования почв на содержание суммарных (общих) углеводов нефти методом ИК-спектроскопии НПВО с учётом существенно различающейся влажности объекта.

Практические разработки оптических мультисенсорных систем для различных практических приложений обобщены в табл. В-1 (Приложение В).

ЗАКЛЮЧЕНИЕ

Диссертационная работа закладывает научные основы создания специализированных оптических анализаторов низкой селективности – оптических мультисенсорных систем, представляющих собой новое направление развития аналитической спектроскопии. Благодаря своим потребительским качествам, таким как миниатюрность, портативность, автономность, возможность использования онлайн и ценовая доступность, ОМС способны существенно расширить аналитические возможности оптической спектроскопии по сравнению с традиционным лабораторным анализом. Их намечающееся широкое использование для решения различных задач качественного и количественного анализа в промышленности, в медицине и других практических областях повышает уровень аналитического контроля многих важных аспектов человеческой деятельности.

Успешная разработка оптических мультисенсорных систем требует решения целого ряда научных и технических задач, рассмотренных в настоящей работе: оптимизации спектрального анализатора под приложение, обучения и проверки математических моделей, аналитического контроля процессов и других. Предложена общая методология разработки ОМС и практические решения для её отдельных стадий, подкреплённые большим числом реальных примеров.

Отдельную ценность с научной точки зрения представляют положения, которые могут применяться и в других областях химии. К ним относятся: предложенные принципы планирования многокомпонентного градуировочного эксперимента, концепция траектории в аналитическом контроле процессов, а также новые подходы многомерного анализа низкоселективных спектральных данных.

Результаты работы представляют большую практическую важность в решении разнообразных задач разработки, совершенствования и эффективного использования ОМС для полевого или экспресс-анализа и онлайн-мониторинга различных объектов и сред. С повышением востребованности химического анализа роль научного подхода к разработке мультисенсорных систем будет возрастать.

ВЫВОДЫ

- 1) Формализован подход к созданию ОМС как специализированных анализаторов низкой селективности, сформулированы основные этапы разработки и решаемые задачи, предложены отдельные решения, облегчающие разработку и повышающие точность анализа. Как альтернатива общепринятому экспертному подходу к конфигурации ОМС предложена математическая оптимизация каналов системы для конкретного практического приложения на предварительно полученных полноспектральных данных спланированного эксперимента. Разработан ускоренный оптимизационный алгоритм, позволяющий повысить точность результирующего анализатора путём включения в расчёт большего числа параметров и имеющих независимую ценность как обобщённый интервальный отбор переменных в хемометрике. Разработан метод многоуровневой валидации для более достоверного сравнения анализаторов и их прогностических моделей с использованием небольшого числа тестовых образцов.
- 2) Разработан централизованный подход к математическому обеспечению полного цикла создания ОМС – от оптимизации каналов до построения рабочих моделей и их практического использования. Подход реализован «в облаке» в комплексе программ «ТРТ cloud» для хранения и анализа многомерных данных и хемометрических моделей с возможностью их глобального использования распределёнными ОМС, а также перманентного обучения на новых данных, поставляемых работающими анализаторами.
- 3) Систематизированы базовые требования к набору образцов для построения с минимальными затратами точных градуировочных моделей для определения нескольких концентраций (или обобщённых показателей состава) из одного аналитического измерения, спектрального или ОМС. С учётом сформулированных требований разработано новое семейство схем, получившее название диагонального дизайна, имеющего преимущества перед известными в теории планирования эксперимента подходами, такие как наличие встроенных проверочных образцов, простота и интерпретируемость

результатов. Эффективность диагонального дизайна доказана на примерах разработки ОМС для определения питательной ценности молока, для мониторинга процесса ферментации, а также ИК-ОМС и методики определения общих углеводов нефти в почвах с различной влажностью.

- 4) Предложена и систематически разработана концепция траектории в пространстве аналитических переменных – методологический инструмент, упрощающий решение задач аналитического контроля и оптимизации технологических процессов. Сформулированы основные свойства траектории процесса и даны практические рекомендации по её использованию в АКТП для непрерывного сбора спектральных данных, их анализа, а также для разработки онлайн-методов определения индивидуальных компонентов и обобщённых показателей качества производимой продукции, в том числе с использованием мультисенсорного подхода.
- 5) Разработан ряд оптических мультисенсорных систем с улучшенными, по сравнению с известными аналогами, метрологическими характеристиками для количественного и качественного полевого, экспресс- и онлайн-анализа, в том числе при определении различных обобщённых показателей химического состава. Разработана ОМС для экспресс-анализа молока на содержание жира и общего белка, обеспечивающая сравнимую с существующими полноспектральными ИК- и БИК-методами точность. Разработаны подходы к созданию и отдельные прототипы ИК-ОМС для мониторинга состава среды (глюкоза, фруктоза, этанол) в процессах ферментации в биотехнологическом производстве, а также определения суммарного содержания углеводов нефти и нефтепродуктов в целях экомониторинга почв. Разработаны прототипы для смежных с химическим анализом сфер биомедицинской диагностики (спектрального распознавания границы опухоли при оперативном лечении рака почки) и онлайн-контроля качества материалов (спектральный анализ процессов производства твёрдых фармацевтических форм для мониторинга содержания воды, толщины защитного покрытия и кинетики высвобождения лекарства).

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

АКТП	аналитический контроль технологических процессов (<i>англ.</i> process analytical technology, PAT)
АТФ	аденозинтрифосфат
АФИ	активный фармацевтический ингредиент (<i>англ.</i> active pharmaceutical ingredient, API)
БИК	ближний инфракрасный
Вид	видимый
ВРВ	выпуск (продукции) в реальном времени (<i>англ.</i> real-time release, RTR)
ГА	генетический алгоритм (<i>англ.</i> genetic algorithm)
ГОСТ	Межгосударственный стандарт (в СССР – Государственный общесоюзный стандарт)
ДД	диагональный дизайн
ДМД	диодно-матричный детектор (<i>англ.</i> diode-array detector, DAD)
ДНК	дезоксирибонуклеиновая кислота
ДО	диффузное отражение
ИК	инфракрасный
ИПЛС	интервальный ПЛС (<i>англ.</i> interval PLS, iPLS)
ИПП	интерфейс прикладного программирования (<i>англ.</i> application program interface, API)
ИЮПАК	Международный союз теоретической и прикладной химии (МСТПХ; <i>англ.</i> International Union of Pure and Applied Chemistry, IUPAC)
КАПВ	конкурентная адаптивная перевзвешенная выборка (<i>англ.</i> competitive adaptive reweighted sampling, CARS)
КВБИК	коротковолновый БИК
КЛК	кусочно-линейная коррекция (базовой линии)
КОС	контроль посредством обратной связи
КПС	контроль посредством прямой связи
КР	комбинационное рассеяние
ЛГО	латинский гиперкуб образцов (<i>англ.</i> latin hypercube of samples,

	LHS)
МГК	метод главных компонент (<i>англ.</i> principal component analysis, PCA)
МД	матрица дизайна (в планировании эксперимента)
МКР	множественная коррекция рассеяния или коррекция множественного рассеяния (<i>англ.</i> multiplicative scatter correction, MSC)
МЛР	множественная линейная регрессия (<i>англ.</i> multiple linear regression, MLR)
ММС	монолитный миниатюрный спектрометр (<i>нем.</i> monolithisches Miniarurspektrometer, MMS)
МРК	метод разделения кривых (<i>англ.</i> multivariate curve resolution)
МСС	метод скользящего среднего
МУИ	метод углового измерения (<i>англ.</i> angle measure technique, AMT)
МЭ	матрица эксперимента
МЭМС	микроэлектромеханические системы (<i>англ.</i> microelectromechanical systems, MEMS)
НАД	никотинамидадениндинуклеотид (<i>англ.</i> nicotinamide adenine dinucleotide, NAD)
НИС	настраиваемый интерферометрический спектрометр
ОМС	оптическая мультисенсорная система
ПАПАФАК	параллельный факторный анализ (<i>англ.</i> parallel factor analysis, PARAFAC)
ПБО	последовательная Байесова оценка (<i>англ.</i> successive Bayesian estimation)
ПЗС	прибор с зарядовой связью (<i>англ.</i> charge-coupled device, CCD)
ПИК	поликристаллическое ИК (оптоволокно)
ПЛС	проекция на латентные структуры (<i>англ.</i> projection to latent structures или partial least-squares, PLS)
ПЛС-ДА	ПЛС-дискриминантный анализ (<i>англ.</i> PLS discriminant analysis, PLS-DA)
ПНК-МРК	МРК методом перемежающихся наименьших квадратов (<i>англ.</i> alternating least squares MCR, ALS-MCR)

ПО	программное обеспечение
ПП	перекрёстная проверка (<i>англ.</i> cross-validation)
ППК	перекрёстная проверка на классах
ПТН	проверка тестовым набором
ППП	полная перекрёстная проверка
ППС	перекрёстная проверка сегментами
ПЭ	планирование эксперимента
СВЭ	спектр возбуждения-эмиссии (<i>англ.</i> excitation-emission matrix)
СГ	метод сглаживания и дифференцирования по Савицкому-Голею (<i>англ.</i> Savitzky-Golay);
СИМПЛИЗМА	простой интерактивный алгоритм автоматического МРК для анализа смесей (<i>англ.</i> simple-to-use interactive self-modeling mixture analysis, SIMPLISMA)
СКО	среднеквадратичное отклонение
СКОГ	среднеквадратичное отклонение градуировки
СКОП	среднеквадратичное отклонение прогноза
СКОПП	среднеквадратичное отклонение перекрёстной проверки
СНВ	стандартизация нормировкой вариации (<i>англ.</i> standard normal variate, SNV)
УФ	ультрафиолетовый
ФАД	флавинадениндинуклеотида (<i>англ.</i> flavine adenine dinucleotide, FAD)
ХИК	халькогенидное ИК (оптоволокно)
ЭВМ	электронно-вычислительная машина

СПИСОК ЛИТЕРАТУРЫ

1. Золотов Ю.А. Российский вклад в аналитическую химию. Москва: ИП Лысенко А.Д. PRESS-BOOK.RU, 2017. 190 с.
2. Reif-Acherman S. Anders Jonas Ångström and the foundation of spectroscopy — Commemorative article on the second centenary of his birth. // *Spectrochim. Acta, Part B*. 2014. V. 102. P. 12–23.
3. ГОСТ 32255-2013. Инструментальный экспресс-метод определения физико-химических показателей идентификации с применением инфракрасного анализатора. // Межгосударственный стандарт. М.: Стандартинформ, 2014. 14 с.
4. Bakeev K.A. *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. Ames, IA: Blackwell Publishing Professional, 2005. 451 p.
5. Juran J.M. *Juran on quality by design: the new steps for planning quality into goods and services*. New York, NY, USA: The Free Press, 1992. 538 p.
6. Kessler R.W. Perspectives in process analysis. // *J. Chemom.* 2013. V. 27, № 11. P. 369–378.
7. Pomerantsev A.L., Rodionova O.Ye. Process analytical technology: a critical view of the chemometricians. // *J. Chemom.* 2012. V. 26, № 6. P. 299–310.
8. PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. FDA Guidance for Industry. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070305.pdf> [03.03.2019].
9. Bogomolov A. Multivariate process trajectories: Capture, resolution and analysis. // *Chemom. Intell. Lab. Syst.* 2011. V. 108, № 1. P. 49–63.
10. Кучерявский С.В. Технология анализа процессов. // *Методы менеджмента качества*. 2006, № 5. С. 12–17.
11. Reich G. *Mid and Near Infrared Spectroscopy, Analytical Techniques in the Pharmaceutical Sciences*. New York: Springer-Verlag, 2016. P. 61–138.
12. Nagy B., Farkas A., Borbás E., Vass P., Nagy Z.K., Marosi G. Raman spectroscopy

- for process analytical technologies of pharmaceutical secondary manufacturing. // *Am. Assoc. Pharm. Sci.* 2019. V. 20, № 1. P. 1–16.
13. Engler M., Bogomolov A., Mannhardt J. Die Lighthouse-Probe, eine neuartige Sonde für die Prozessanalytik. // *Chem. Ing. Tech.* 2009. V. 81, № 8. P. 1114–1115.
 14. Shiryaev V.S., Churbanov M.F., Dianov E.M., Plotnichenko V.G., Adam J.-L., Lucas J. Recent progress in preparation of chalcogenide As-Se-Te glasses with low impurity content. // *J. Optoelectron. Adv. Mater.* 2005. V. 7, № 4. P. 1773–1779.
 15. Artjushenko V.G., Baskov P.B., Kuzrnicheva G.M., Musina M.D., Sakharov V.V., Sakharova T.V. Structure and properties of $\text{AgCl}_{1-x}\text{Br}_x$ ($x=0.5\text{--}0.8$) optical fibers. // *Inorg. Mater.* 2005. V. 41. P. 178–181.
 16. Artyushenko V., Bocharnikov A., Colquhoun G., Leach C., Lobachev V., Sakharova T., Savitsky D. Mid-IR fibre optics spectroscopy in the $3300\text{--}600\text{ cm}^{-1}$ range. // *Vib. Spectrosc.* 2008. V. 48. P. 168–171.
 17. Lumpi D., Wagner Ch., Schöpf M., Horkel E., Ramer G., Lendl B., Fröhlich J. Fibre optic ATR-IR spectroscopy at cryogenic temperatures: in-line reaction monitoring on organolithium compounds electronic supplementary material (esi) for chemical communications. // *Chem. Commun.* 2012. V. 48. P. 2451–2453.
 18. Khaustova S., Shkurnikov M., Tonevitsky E., Artyushenko V., Tonevitsky A. Noninvasive biochemical monitoring of physiological stress by Fourier transform infrared saliva spectroscopy. // *Analyst.* 2010. V. 135. P. 3183–3192.
 19. Damien F., Boncheva M. The extent of orthorhombic lipid phases in the stratum corneum determines the barrier efficiency of human skin in vivo. // *J. Investig. Dermatol.* 2010. V. 130. P. 611–614.
 20. Bruch R.F., Sukuta S., Afanasyeva N.I., Kolyakov S.F., Butvina L.N. Fourier transform infrared evanescent wave (FTIR-FEW) spectroscopy of tissue. // *Proc. SPIE 2970, Lasers in Surgery: Advanced Characterization, Therapeutics, and Systems VII.* 1997. V. 2970.
 21. Wang L., Mizaikoff B. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. // *Anal. Bioanal. Chem.* 2008. V. 391. P. 1641–1654.

22. Mackanos M.A., Contag C.H. Fiber-optic probes enable cancer detection with FTIR spectroscopy. // *Trends Biotechnol.* 2010. V. 28, № 6. P. 317–323.
23. Bui D.A., Hauser P.C. Analytical devices based on light-emitting diodes – a review of the state-of-the-art. // *Anal. Chim. Acta.* 2015. V. 853. P. 46–58.
24. Золотов Ю.А. Основы аналитической химии, 5-е изд., Т.2. М.: Издательский центр «Академия», 2012. 408 с.
25. Greer J., Petrov G.I., Yakovlev V.V. Raman spectroscopy with LED excitation source. // *J. Raman Spectrosc.* 2013. V. 44, № 7 P. 1058–1059.
26. de Lima K.M.G. A portable photometer based on LED for the determination of aromatic hydrocarbons in water. // *Microchem. J.* 2012. V. 103. P. 62–67.
27. Bogomolov A., Ageev V., Zabarylo U., Usenov I., Schulte F., Kirsanov D., Belikova V., Minet O., Feliksberger E., Meshkovsky I., Artyushenko V. LED-based near infrared sensor for cancer diagnostics. // *Proc. SPIE.* 2016. V. 9715. art. no. 971510.
28. Giovenzana V., Civelli R., Beghi R., Oberti R., Guidetti R. Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera* L.) for Franciacorta wine. // *Talanta.* 2015. V. 144. P. 584–591.
29. Malinen J., Rissanen A., Saari H., Karioja P., Karppinen M., Aalto T., Tukkiniemi K. Advances in miniature spectrometer and sensor development. // *Proc. SPIE.* 2014. V. 9101. P. 1–15.
30. Kim D.S., Lee T.-R., Yoon G. Development of an ultra-compact mid-infrared attenuated total reflectance spectrophotometer. // *Opt. Eng.* 2015. V. 53, № 7. art. no. 074108.
31. Chew W., Sharratt P. Trends in process analytical technology. // *Anal. Methods.* 2010. V. 2. P. 1412–1438.
32. Demtröder W. *Laser Spectroscopy: Basic Concepts and Instrumentation* (3rd ed.). Berlin: Springer, 2003. 997 p.
33. Kurth S., Hiller K., Neumann N., Heinze M., Doetzel W., Gessner T. Tunable Fabry-Perot-Interferometer for 3-5 μm wavelength with bulk micromachined reflector carrier. // *Proc. SPIE.* 2003. V. 4983. P. 215–226.
34. «Qmini» User Manual. <https://www.rgb->

- photonics.com/fileadmin/user_upload/downloads/manuals/Qmini_2_Spectrometer.pdf [15.04.2019].
35. «Spectral Engines». <https://www.spectralengines.com/> [15.04.2019].
 36. «MicroNIR PAT-U»: A fit-for-purpose near-infrared (NIR) analyzer for quality-by-design (QbD) manufacturing. <https://www.viavisolutions.com/en-us/osp/products/micronir-pat-u> [15.04.2019].
 37. De Beer T.R.M., Wiggenhorn M., Veillon R., Debaq C., Mayeresse Y., Moreau B., Burggraeve A., Quinten T., Friess W., Winter G., Vervae C., Remon J.P., Baeyens W.R.G. Importance of using complementary process analyzers for the process monitoring, analysis, and understanding of freeze drying. // *Anal. Chem.* 2009. V. 81, № 18. P. 7639–7649.
 38. De Beer T.R.M., Vercruyse P., Burggraeve A., Quinten T., Ouyang J., Zhang X., Vervae C., Remon J.P., Baeyens W.R.G. In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. // *J. Pharm. Sci.* 2009. V. 98, № 9. P. 3430–3446.
 39. De Groot P.G., Swierenga H., Postma G.J., Melssen W.J., Buydens L.M.C. Effect on the partial least-squares prediction of yarn properties combining Raman and infrared measurements and applying wavelength selection. // *Appl. Spectrosc.* 2003. V. 57. P. 642–648.
 40. Bogomolov A., Engler M., Melichar M., Wigmore A. In-line analysis of a fluid bed pellet coating process using a combination of near infrared and Raman spectroscopy. // *J. Chemom.* 2010. V. 24. P. 544–557.
 41. Smilde A.G., Westerhuis J.A., Boqué R. Multiway multiblock component and covariates regression models. // *J. Chemom.* 2000. V. 14. P. 301–331.
 42. Lopes J.A., Menezes J.C., Westerhuis J.A., Smilde A.K. Multiblock PLS analysis of an industrial pharmaceutical process. // *Biotechnol. Bioeng.* 2002. V. 80. P. 419–427.
 43. Choi S.W., Lee I.B. Multiblock PLS-based localized process diagnosis. // *J. Process Control.* 2005. V. 15. P. 295–306.

44. Höskuldsson A., Rodionova O.Ye., Pomerantsev A.L. Path modeling and process control. // *Chemom. Intell. Lab. Syst.* 2007. V. 88. P. 84–99.
45. Pages J., Tenenhaus M. Multiple factor analysis combined with PLS path modeling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. // *Chemom. Intell. Lab. Syst.* 2001. V. 58. P. 261–273.
46. Родионова О.Е. Интервальный метод обработки результатов многоканальных экспериментов: дис. д-ра физ.-мат. наук: 01.04.01. Москва, 2008. 273 с.
47. Blum L.J., Coulet P.R. (Eds.). *Biosensor Principles and Applications*. New York: Marcel Dekker, 1991. 376 p.
48. Sarker S.D., Nahar L. Hyphenated techniques and their applications in natural products analysis. // *Methods Mol Biol.* 2012. V. 864. P. 301–340.
49. Kucheryavski S. Extracting useful information from images. // *Chemom. Intell. Lab. Syst.* 2011. V. 108. P. 2–12.
50. Kiers H. Some procedures for displaying results from three-way methods. // *J. Chemom.* 2000. V. 14. P. 151–170.
51. Wold S., Martens H., Wold H. The multivariate calibration problem in chemistry solved by the PLS method. Eds. Ruhe A., Kågström B. In: *Conference proceedings, Matrix Pencils. Lecture Notes in Mathematics*. Pite Havsbad, Sweden, March 22–24. 1983. V. 973. P. 286–293.
52. Wold S., Sjöström M., Eriksson L. PLS-regression: a basic tool of chemometrics. // *Chemom. Intell. Lab. Syst.* 2001. V. 58. P. 109–130.
53. Esbensen K.H. *Multivariate data analysis – in practice: An introduction to multivariate analysis and experimental design (5th ed.)*. Oslo: CAMO Process AS, 2001. 598 p.
54. Родионова О.Е., Померанцев А.Л. Хемометрика: достижения и перспективы. // *Усп. хим.* 2006. Т. 74, № 4. С. 302–317.
55. Severinghaus J.W., Honda Y. History of blood gas analysis. VII. Pulse oximetry. // *J. Clin. Monit.* 1987. V. 3, № 2. P. 135–138.
56. Vashist S.K. Non-invasive glucose monitoring technology in diabetes management :

- A review. // *Anal. Chim. Acta*. 2012. V. 750. P. 16–27.
57. Debus B., Kirsanov D., Yaroshenko I., Sidorova A., Piven A., Legin A. Two low-cost digital camera-based platforms for quantitative creatinine analysis in urine. // *Anal. Chim. Acta*. 2015. V. 895. P. 71–79.
58. Roda A., Michelini E., Zangheri M., Di Fusco M., Calabria D., Simoni P. Smartphone-based biosensors: A critical review and perspectives. // *Trends Analyt. Chem.* 2016. V. 79. P. 317–325.
59. Hulanicki A., Glab S., Ingman F. Chemical sensors: definitions and classification. // *Pure Appl. Chem.* 1991. V.63. P. 1247–1250.
60. Persaud K., Dodd G.H. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. // *Nature*. 1982. V. 299. P. 352–355.
61. Vlasov Yu., Legin A., Rudnitskaya A., DI Natale C., D'Amico A. Nonspecific sensor arrays ("electronic tongue") for chemical analysis of liquids (IUPAC Technical Report). // *Pure Appl. Chem.* 2005. V. 77, № 11. P. 1965–1983.
62. Власов Ю.Г., Легин А.В., Рудницкая А.М. Электронный язык — системы химических сенсоров для анализа водных сред. // *Рос. хим. ж. (Ж. Рос. хим. об-ва им. Д.И. Менделеева)*. 2008. Т. LII, № 2. С. 101–112.
63. Smith T., Guild J. The C.I.E. colorimetric standards and their use. // *Trans. Opt. Soc.* 1931. V. 33, № 3. P. 73–134.
64. Osorio D., Vorobyev M. A review of the evolution of animal colour vision and visual communication signals. // *Vision Res.* 2008. V. 48, № 20. P. 2042–2051.
65. Bogomolov A., Heßling M., Wenzel U., Princz S., Hellmuth T., Barraza Bernal M.J., Sakharova T., Usenov I., Artyushenko V., Meyer H. Development and testing of mid-infrared sensors for in-line process monitoring in biotechnology. // *Sens. Actuator B-Chem.* 2015. V. 221. P. 1601–1610.
66. Galyanin V., Belikov V., Belikova V., Bogomolov A. Emission band width approximation of light-emitting diodes in the region 350–2100 nm. // *Sens. Actuator B-Chem.* 2017. V. 252. P. 773–776.
67. Medvedovici A., David V. Spectrophotometry: Diode Array. Eds. Worsfold P.J., Townshend A., Poole C.F. In: *Encyclopedia of Analytical Science* (2nd ed.). Oxford:

- Elsevier, 2005. P. 321–335.
68. Kapustin D., Zhilin S. Sorting for special cereal-based products needs new quality analysis standards. // 11th Winter Symposium on Chemometrics (WSC11). St. Petersburg, Russia, February 26 – March 2, 2018. P. 57–58.
 69. Сулима Е.Л. Алгоритмы переноса градуировочных моделей внутри серии ИКФ-спектрометров: дис. канд. тех. наук: 05.11.13. Санкт-Петербург, 2005. 123 с.
 70. Bouveresse E., Massart D.L. Standardization of near-infrared spectrometric instruments: a review. // *Vib. Spectrosc.* 1996. V. 11, № 1. P. 3–15.
 71. Wang Y.D., Veltkamp D.J., Kowalski B.R. Multivariate instrument standardization. // *Anal. Chem.* 1991. V. 63, № 23. P. 2750–2756.
 72. Aernouts B., Polshin E., Lammertyn J., Saeys W. Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: reflectance or transmittance? // *J. Dairy Sci.* 2011. V. 94, № 11. P. 5315–5329.
 73. Lito P.F., Magalhães A.L., Gomes J.R.B., Silva C.M. Universal model for accurate calculation of tracer diffusion coefficients in gas, liquid and supercritical systems. // *J. Chromatogr. A.* 2013. V. 1290. P. 1–26.
 74. Zhang X.B., Feng Y.C., Hu C.Q. Feasibility and extension of universal quantitative models for moisture content determination in beta-lactam powder injections by near-infrared spectroscopy. // *Anal. Chim. Acta.* 2008. V. 630, № 2. P. 131–140.
 75. MS-80P: LED-based sensor for in-line moisture monitoring. <https://www.pharma-test.de/en/jm/products/tidas-process/p-ms-80p/> [15.04.2019].
 76. Davies A.M.C., Fearn T. Back to basics: the “final” calibration. // *Spectrosc. Eur.* 2007. V.19, № 6. P. 15–18.
 77. Шмидт В. Оптическая спектроскопия для химиков и биологов. М.: Техносфера, 2007. 368 с.
 78. Bogomolov A., Zabarylo U., Kirsanov D., Belikova V., Ageev V., Usenov I., Galyanin V., Minet O., Sakharova T., Danielyan G., Feliksberger E., Artyushenko V. Development and testing of an LED-based near-infrared sensor for human kidney tumor diagnostics. // *Sensors.* 2017. V. 17, № 8. art. no. 1914.

79. Meyer H., Bogomolov A. Optische Messvorrichtung und Messverfahren. Patent DE 10 2013 219 932 B4.
80. Schmilovitch Z., Katz G., Maltz E., Kutscher M.L. Spectroscopic fluid analyzer. Patent EP 1 444 501 B1.
81. Bogomolov A., Grasser T., Heßling M. In-line monitoring of *Saccharomyces cerevisiae* fermentation with a fluorescence probe: new approaches to data collection and analysis. // J. Chemom. 2011. V. 25. P. 389–399.
82. NDC Technologies. <https://www.ndc.com/en/Products/On-Line-NIR-Gauges.aspx> [15.04.2019].
83. Schügerl K., Zeng A.-P. Tools and applications of biochemical engineering science Ed. Scheper T. In: Advances in Biochemical Engineering/Biotechnology. Berlin Heidelberg: Springer, 2002. V. 74. 273 p.
84. Bogomolov A., Dietrich S., Boldrini B., Kessler R.W. Quantitative determination of fat and total protein in milk based on visible light scatter. // Food Chem. 2012. V. 134. P. 412–418.
85. Bogomolov A., Melenteva A. Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability. // Chemom. Intell. Lab. Syst. 2013. V. 126. P. 129–139.
86. Bogomolov A., Mannhardt J., Heinzerling O. Accuracy improvement of in-line near-infrared spectroscopic moisture monitoring in a fluidized bed drying process. // Front. Chem. 2018. V. 6 P. 1–10.
87. Guryanova A., Ermakov V., Galyanin V., Artyushenko V., Sakharova T., Usenov I., Bykov D.E., Bogomolov A. Quantitative analysis of total hydrocarbons and water in oil-contaminated soils with attenuated total reflection (ATR) infrared spectroscopy. // J. Chemom. 2017. V. 31. P. 1–10.
88. Bogomolov A., Belikova V., Zabarylo U.J., Bibikova O., Usenov I., Sakharova T., Krause H., Minet O., Feliksberger E., Artyushenko V. Synergy effect of combining fluorescence and mid infrared fiber spectroscopy for kidney tumor diagnostics. // Sensors. 2017. V. 17, № 11. art. no. 2548.
89. Bogomolov A., Belikova V., Galyanin V., Melenteva A., Meyer H. Reference-free

- spectroscopic determination of fat and protein in milk in the visible and near infrared region below 1000 nm using spatially resolved diffuse reflectance fiber probe. // *Talanta*. 2017. V. 167. P. 563–572.
90. Kucheryavskiy S.V., Melenteva A., Bogomolov A. Determination of fat and total protein content in milk using conventional digital imaging. // *Talanta*. 2014. V. 121. P. 144–152.
91. Bogomolov A. Optical multisensor systems. // 11th Winter Symposium on Chemometrics (WSC11). St. Petersburg, Russia, February 26 – March 2, 2018. P. 35.
92. Application programming interface. https://en.wikipedia.org/wiki/Application_programming_interface [15.04.2019]
93. Massart D.L. Vandeginste B.G.M., Deming S.N., Michotte Y., Kaufman L. *Chemometrics: a textbook*. Amsterdam: Elsevier, 1988. 488 p.
94. Померанцев А.Л. Химия и жизнь. Четвёртая парадигма. // *Химия и жизнь* – XXI век. 2006, № 6. С. 22–26.
95. Мелентьева А.А. Анализ молока на содержание жира и общего белка методами оптической спектроскопии рассеяния: дис. канд. хим. наук: 02.00.02. Саратов, 2015. 134 с.
96. Naes T., Isaksson T., Fearn T., Davies T. *A user friendly guide to multivariate calibration and classification*. Chichester UK: NIR publications, 2002. 344 p.
97. Geladi P., MacDougall D., Martens H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. // *Appl. Spectrosc.* 1985. V. 39, № 3. P. 491–500.
98. Bogomolov A., Windig W., Geer S.M., Blondell D.B., Robbins M.J. Interactive series baseline correction algorithm. // 4th Winter Symposium on Chemometrics (WSC4), Chernogolovka, Russia, February 12–18, 2005. P. 20.
99. Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. // *Anal. Chem.* 1964. V. 36. P. 1627–1639.
100. Niazi A., Leardi R. Genetic algorithms in chemometrics. // *J. Chemom.* 2012. V. 26, № 3. P. 345–351.

101. Nørgaard L., Saudland A., Wagner J., Nielsen J.P., Munck L., Engelsen S.B. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. // *Appl. Spectrosc.* 2000. V. 54. P. 413–419.
102. Li H., Liang Y., Xu Q., Cao D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. // *Anal. Chim. Acta.* 2009. V. 648, № 1. P. 77–84.
103. Centner V., Massart D.L., de Nord O.E., de Jong S., Vandeginste B., Sterna C. Elimination of uninformative variables for multivariate calibration. // *Anal. Chem.* 1996. V. 68. P. 3851–3858.
104. Jouan-Rimbaud D., Walczak B., Poppi R.J., de Nord O.E., Massart D.L. The development of calibration models for spectroscopic data using principal component regression. // *Anal. Chem.* 1997. V. 69. P. 4317–4323.
105. Leardi R. Application of genetic algorithm – PLS for feature selection in spectral data sets. // *J. Chemom.* 2000. V. 14. P. 643–655.
106. Hagerman J.A., Streppel M., Wehrens R., Buydens L.M.C. Wavelength selection with Tabu Search. // *J. Chemom.* 2004. V. 17. P. 427–437.
107. Nording T.E.M., Koljonen J., Alander J.T., Geladi P. Genetic algorithms as a tool for wavelength selection. // Eds. Alander J.T., AlaSiuru P., Hyötyniemi H. In: *STeP-2004, Proceeding of the 11th Finnish Artificial Intelligence Conferences.* Vantaa, Finland, September 1–3, 2005. V. 3. P. 99–113.
108. Lucasius C.B., Becker M.L.M., Kateman G. Genetic algorithms in wavelength selection: a comparative study. // *Anal. Chim. Acta.* 1994. V. 286. P. 135–153.
109. Hörchner U., Kalivas J.H. Further investigation on a comparative study of simulated annealing and genetic algorithm for wavelength selection. // *Anal. Chim. Acta.* 1995. V. 311. P. 1–13.
110. Navea S., Tauler R., de Juan A. Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure. // *Anal. Biochem.* 2005. V. 336. P. 231–241.
111. Navea S., Tauler R., Goormaghtigh E., de Juan A. Chemometric tools for

- classification and elucidation of protein secondary structure from infrared and circular dichroism spectroscopic measurements. // *Proteins*. 2006. V. 63. P. 527–541.
112. Galvão R.K.H., Araújo M.C.U. Variable selection. Eds. Brown S.D., Tauler R., Walczak B. In: *Comprehensive chemometrics: chemical and biochemical data analysis*. Oxford: Elsevier, 2009. Ch. 3.05. P. 233–283.
113. Höskuldsson A. Variable and subset selection in PLS regression. // *Chemom. Intell. Lab. Syst.* 2001. V. 55. P. 23–38.
114. Bogomolov A., Hachey M.R.J. Application of SIMPLISMA purity function for variable selection in multivariate regression analysis: a case study of protein secondary structure determination from infrared spectra. // *Chemom. Intell. Lab. Syst.* 2007. V. 88. P. 132–142.
115. Hachey M., Bogomolov A., Boruta M. ANYL 71-Application of a new variable selection method to the determination of protein secondary structure determination using infrared spectroscopy. // *Abstracts of papers of the American chemical society*. 2007. V. 234.
116. Holland, J. H. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Oxford, England: University of Michigan Press, 1975. 183 p.
117. Koljonen J., Nordling T.E.M., Alander J.T. A review of genetic algorithms in nearinfrared spectroscopy and chemometrics: past and future. // *J. Near Infrared Spectrosc.* 2008. V. 16. P. 189–197.
118. Wold S., Esbensen K., Geladi P. Principal component analysis. // *Chemom. Intell. Lab. Syst.* 1987. V. 2. P. 37–52.
119. Karpushkin E., Bogomolov A., Boruta M. New system for computer-aided infrared and Raman spectrum interpretation. // *Chemom. Intell. Lab. Syst.* 2007. V. 88. P. 107–117.
120. Монахова Ю.Б. Методология хемометрического моделирования спектрометрических сигналов в анализе объектов сложного состава: дис. д-ра хим. наук: 02.00.02. Саратов, 2016. 318 с.

121. de Juan A., Rutan S.C., Tauler R. Two-way data analysis: multivariate curve resolution—iterative resolution methods. Eds. Brown S.D., Tauler R., Walczak B. In: *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Oxford: Elsevier, 2009. Ch. 2.19. P. 325–344.
122. Windig W., Bogomolov A., Kucheryavskiy S. Two-Way Data Analysis: Detection of Purest Variables. In: *Comprehensive Chemometrics (2nd ed.)*, Chapter 2.08; Brown S., Tauler R., Walczak B., Eds. Amsterdam: Elsevier B.V., 2020, P. 107–136.
123. Bogomolov A., Hachey M., Williams A. Software for interactive curve resolution using SIMPLISMA. Ed. Pomerantsev A.L. In: *Progress in Chemometrics Research*. New York: Nova Science Publishers, 2005. Ch. 10. P. 119–135.
124. Kucheryavskiy S., Windig W., Bogomolov A. Spectral unmixing using the concept of pure variables. Ed. Ruckebusch C. In: *Data Handling in Science and Technology: V. 30: Resolving Spectral Mixtures*. Amsterdam: Elsevier, 2016. Ch. 3. P. 53–99.
125. Debus B., Kirsanov D.O., Panchuk V.V., Semenov V.G., Legin A. Three-point multivariate calibration models by correlation constrained MCR-ALS: A feasibility study for quantitative analysis of complex mixtures. // *Talanta*. 2017. V. 163. P. 39–47.
126. Sjöström M., Wold S., Lindberg W., Persson J.-Å., Martens H. A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. // *Anal. Chim. Acta*. 1983. V. 150. P. 61–70.
127. Wise B.M., Gallagher N.B. The process chemometrics approach to process monitoring and fault detection. // *J. Proc. Cont.* 1996. V. 6, № 6. P. 329–348.
128. The Unscrambler: Methods.
<https://www.camo.com/downloads/U9.6%20pdf%20manual/The%20Unscrambler%20Methods.pdf> [15.04.2019].
129. Belikova V., Bogomolov A. Distance estimation between objects in spectral data analysis. // 11th Winter Symposium on Chemometrics (WSC11). St. Petersburg, Russia, February 26 – March 2, 2018. P. 43-44.
130. Brereton R.G. *Chemometrics for pattern recognition*. Bristol, UK: John Wiley &

- Sons, Ltd., 2009. 522 p.
131. Brereton R.G. *Chemometrics: Data analysis for the laboratory and chemical plant*. Chichester, UK: John Wiley & Sons, 2003. 489 p.
 132. Solovieva S., Karnaukh M., Panchuk V., Andreev E., Kartsova L., Bessonova E., Legin A., Wang P., Wan H., Jahatspanian I., Kirsanov D. Potentiometric multisensor system as a possible simple tool for non-invasive prostate cancer diagnostics through urine analysis // *Sens. Actuator B-Chem.* 2019. V. 289. P. 42–47.
 133. Liu Y.J., Tran T., Postma G., Buydens L.M.C., Jansen J. Estimating the number of components and detecting outliers using Angle Distribution of Loading Subspaces (ADLS) in PCA analysis. // *Anal. Chim. Acta.* 2018. V. 1020. P. 17–29.
 134. Цикин А.М. Хемометрика в анализе многокомпонентных объектов на основе данных ИК, ЯМР и рентгенофлуоресцентной спектроскопии, отличающихся большим значением взаимной информации: дис. канд. хим. наук: 02.00.02. Саратов, 2014. 181 с.
 135. Westerhuis J.A., Velzen E.J.J., Hoefsloot H.C.J., Smilde A.K. Discriminant Q^2 (DQ^2) for improved discrimination in PLS-DA models. // *Metabolomics.* 2008. V. 4. P. 293–296.
 136. Coppa M., Revello-Chion A., Giaccone D., Ferlay A., Tabacco E., Borreani G. Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. // *Food Chem.* 2014. V. 150. P. 49–57.
 137. Sola-Larrañaga C., Navarro-Blasco I. Chemometric analysis of minerals and trace elements in raw cow milk from the community of Navarra, Spain. // *Food Chem.* 2009. V. 112, № 1. P. 189–196.
 138. Gemperline P.J., Cho J.H., Aldridge P.K., Sekulic S.S. Appearance of discontinuities in spectra transformed by the piecewise direct instrument standardization procedure. // *Anal. Chem.* 1996. V. 68, № 17. P. 2913–2915.
 139. Bouveresse E., Hartmann C., Massart D.L., Last I.R., Prebble K.A. Standardization of near-infrared spectrometric instruments. // *Anal. Chem.* 1996. V. 68. P. 982–990.
 140. Lind J. *A treatise of the scurvy in three parts. Containing an inquiry into the nature, causes, and cure, of that disease; Together with a critical and chronological view of*

- what has been published on the subject (1st ed.). Cambridge: Cambridge University Press, 1753. 456 p.
141. Leardi R. Experimental design in chemistry: A tutorial. // *Anal. Chim. Acta*. 2009. V. 652. P.161–172.
 142. Eriksson L., Johansson E., Kettaneh-Wold N., Wikström C., Wold S. Design of experiments: Principles and applications (3rd ed.). Umeå: Umetrics AB, 2008. 459 p.
 143. Bogomolov A. Diagonal designs for a multi-component calibration experiment. // *Anal. Chim. Acta*. 2017. V. 951. P. 46–57.
 144. Eriksson L., Johansson E., Wikström C. Mixture design—design generation, PLS analysis, and model usage. // *Chemom. Intell. Lab. Syst.* 1998. V. 43. P. 1–24.
 145. Bondi R.W. Jr., Igne B., Drennen J.K. 3rd, Anderson C.A. Effect of experimental design on the prediction performance of calibration models based on near-infrared spectroscopy for pharmaceutical applications. // *Appl. Spectrosc.* 2012. V. 66, № 12, P. 1442–1453.
 146. Araujo P.W., Brereton R.G. Visualisation of confidence in two-factor designs where model, replication and star points are varied. // *Analyst*. 1997. V. 122. P. 621–630.
 147. ASTM E1655-05 (Reapproved 2012). Standard practices for infrared multivariate quantitative analysis.
 148. Brereton R.G. Multilevel multifactor designs for multivariate calibration. // *Analyst*. 1997. V. 122. P. 1521–1529.
 149. Muñoz J.A., Brereton R.G. Partial factorial designs for multivariate calibration: extension to seven levels and comparison of strategy. // *Chemom. Intell. Lab. Syst.* 1998. V. 43. P. 89–105.
 150. Şahin S., Demir C., Güçer Ş. Simultaneous UV–Vis spectrophotometric determination of disperse dyes in textile wastewater by partial least squares and principal component regression. // *Dyes Pigm.* 2007. V. 73. P. 368–376.
 151. Devos O., Fanget B., Saber A.-I., Paturel L., Naffrechoux E., Jarosz J. Use of a plackett-burman design with multivariate calibration for the analysis of polycyclic aromatic hydrocarbons in micellar media by synchronous fluorescence. // *Anal.*

- Chem. 2002. V. 74. P. 678–683.
152. Cirovic D.A. Influence of mixture design on multivariate prediction of PAHs in mixture spectra. // *Talanta*. 1998. V. 45. P. 989–1000.
153. McKay M.D., Beckman R.J., Conover W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. // *Technometrics*. 2000. V. 42. P. 55–61.
154. Fang K.T., Lin D.K.J. Uniform experimental design and its applications in industry. Eds. Rao C.R., Khattree R. In: *Handbook of Statistics in Industry*. New York: North Holland, 2003. Ch. 4. V. 22. P. 131–170.
155. Liang Y.Z., Fang K.T., Xu Q.S. Uniform design and its applications in chemistry and chemical engineering. // *Chemom. Intell. Lab. Syst.* 2001. V. 58. P. 43–57.
156. Uniform design tables. <http://sites.stat.psu.edu/~rli/DMCE/UniformDesign/> [16.04.2019].
157. Kirsanov D., Panchuk V., Agafonova-Moroz M., Khaydukova M., Lumpov A., Semenova V., Legin A. A sample-effective calibration design for multiple components. // *Analyst*. 2014. V. 139. P. 4303–4309.
158. Workman J.Jr., Koch M., Lavine B., Chrisman R. Process Analytical Chemistry. // *Anal. Chem.* 2009. V. 81. P. 4623–4643.
159. Wold S., Cheney J., Kettaneh N., McCready C. The chemometric analysis of point and dynamic data in pharmaceutical and biotech production (PAT) – some objectives and approaches. // *Chemom. Intell. Lab. Syst.* 2006. V. 84. P. 159–163.
160. van den Berg F.W.J., Hoefsloot H.C.J., Smilde A.K. Selection of optimal process analyzers for plant-wide monitoring. // *Anal. Chem.* 2002. V. 74. P. 3105–3111.
161. Gy P. Sampling of discrete materials — a new introduction to the theory of sampling: I. Qualitative approach. // *Chemom. Intell. Lab. Syst.* 2004. V. 74. P. 7–24.
162. Esbensen K.H., Friis-Petersen H.H., Petersen L., Holm-Nielsen J.B., Mortensen P.P. Representative process sampling—in practice: Variographic analysis and estimation of total sampling errors (TSE). // *Chemom. Intell. Lab. Syst.* 2007. V. 88. P. 41–59.
163. Skibsted E.T.S. PAT and beyond. Ph.D. Thesis. University of Amsterdam, 2005.

232 p.

164. Morris J., Chen Z. From data to knowledge through smart process analytical technologies (PAT) and process systems engineering. // *European Pharm. Review*. 2008. V. 6. P. 82–91.
165. Chen Z.P., Lovett D., Morris J. Process Analytical Technologies (PAT) — The impact for process systems engineering. // *Comput. Aided Chem. Eng.* Eds. Braunschweig B., Xavier J. In: 18th European Symposium on Computer Aided Process Engineering. Lyon, France, 2008. V. 25. P. 967–972.
166. Davis B., Lundsberg L., Cook G. PQLI control strategy model and concepts. // *J. Pharm. Innov.* 2008. V. 3. P. 95–104.
167. Грибов Л.А., Баранов В.И., Эляшберг М.Е. Безэталонный молекулярный спектральный анализ. Теоретические основы. Москва: Изд. Эдиториал УРСС, 2002. 317 с.
168. Fiege M., Harkai S., (...), Bogomolov A., Fähler B., Schurmann K. Automated detection of counterfeit drugs using multimodal spectroscopy and advanced web-based software platforms. // *Spectroscopy Europe*. 2015. V. 27(4) P. 19–22.
169. Artyushenko V., Zabarylo U., Bogomolov A., (...), Schulte F., Eichler H.J. Fiber spectroscopy for tumor margin detection – Selection of the best methods. // *Optics InfoBase Conference Papers*. 2014.
170. Artyushenko V., Zabarylo U.J., Bogomolov A., Minet O., Belikova V., Sakharova T., Usenov I. Using fiber spectroscopy methods for rapid cancer diagnosis. // *SPEC 2016 Conference*. Montreal, Canada, June 26–30, 2016.
171. Artyushenko V., Schulte F., Zabarylo U., (...), Krause H., Minet O. Spectral fiber sensors for cancer diagnostics in vitro. // *Optics InfoBase Conference Papers*. 2014.
172. Galyanin V., Melenteva A., Bogomolov A. Selecting optimal wavelength intervals for an optical sensor: A case study of milk fat and total protein analysis in the region 400–1100 nm. // *Sens. Actuator B-Chem*. 2015. V. 218. P. 97–104.
173. Fischer A.G. Injection electroluminescence. // *Solid State Electron*. 1961. V. 2, № 4. P. 232–246.
174. Hachey M.R.J., Bogomolov A., Gordon K.C., Rades T. How using Raman

- spectroscopy and SIMPLISMA can accelerate the study of polymorphs: a case study using Carbamazepine. // *Raman Technology for Today's Spectroscopists*. 2004 (June). P. 17–21.
175. Bogomolov A., McBrien M. Mutual peak matching in a series of HPLC/DAD mixture analyses. // *Anal. Chim. Acta*. 2003. V. 490. P. 41–58.
176. Igne B., Bogomolov A., Bu D., Dardenne P., Galyanin V., Tillmann P. Summary of the 2014 IDRC software shoot-out. // *NIR News*. 2015. V. 20. P. 8–14.
177. Gottfries J., Depui H., Fransson M., Jongeneelen M., Josefson M., Langkilde F.W., Witte D.T. Vibrational spectrometry for the assessment of active substance in metoprolol tablets: a comparison between transmission and diffuse reflectance near-infrared spectrometry. // *J. Pharm. Biomed. Anal.* 1996. V. 14. P. 1495–1503.
178. Kalinin A.V., Krashennnikov V.N., Krivtsun V.M. Short wave near infrared spectrometry of back scattering and transmission of light by milk for multi-component analysis. // *J. Near Infrared Spectrosc.* 2013. V. 21. P. 35–41.
179. Schmilovitch Z., Shmulevich I., Notea A., Maltz E. Near infrared spectrometry of milk in its heterogeneous state. // *Comput. Electron. Agric.* 2000. V. 29, № 3. P. 195–207.
180. Li J., Humphrey A.E. Use of fluorometry for monitoring and control of a bioreactor. // *Biotechnol. Bioeng.* 1991. V. 37. P. 1043–1049.
181. Богомолов А. Проекционные методы в линейном регрессионном анализе: РГК/ ПЛС (лекция) // Школа «Введение в анализ многомерных данных», Самарский государственный технический университет, г. Самара, РФ, 16 февраля 2005 г.
182. Melenteva A., Galyanin V., Savenkova E., Bogomolov A. Building global models for fat and total protein content in raw milk based on historical spectroscopic data in the visible and short-wave near infrared range. // *Food Chem.* 2016. V. 203. P. 190–198.
183. ASTM E3029-15. Standard Practice for Determining Relative Spectral Correction Factors for Emission Signal of Fluorescence Spectrometers,
184. ASTM E2911-13. Standard Guide for Relative Intensity Correction of Raman

Spectrometers.

185. Pomerantsev A.L., Rodionova O.Ye., Melichar M., Wigmore A.J., Bogomolov A. In-line prediction of drug release profiles for pH-sensitive coated pellets. // *Analyst*. 2011. V. 136, № 22. P. 4830–4838.
186. Bogomolov A., Melenteva A., Dahm D.J. Fat globule size effect on visible and short-wave near infrared spectra of milk. // *J. Near Infrared Spectrosc.* 2013. V. 21. P. 435–440.
187. Moberg L., Robertson G., Karlberg B. Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis. // *Talanta*. 2001. V. 54. P. 161–170.
188. Tan Y.X., Jiang J.-H., Wu H.-L., Cui H., Yu R-Q. Resolution of kinetic system of simultaneous degradations of chlorophyll a and b by PARAFAC. // *Anal. Chim. Acta*. 2000. V. 412. P. 195–202.
189. Ermakov V.V., Bogomolov A., Bykov D.E. Oil sludge depository assessment using multivariate data analysis. // *J. Environ. Manage.* 2012. V. 105. P. 144–151.
190. Karpushkin E., Bogomolov A. Morphology Assessment of Poly(2-Hydroxyethyl Methacrylate) Hydrogels Using Multivariate Analysis of Viscoelastic and Swelling properties. // *Polymer*. 2015. V. 58. P. 222–229.
191. Orman B.A., Schumann R.A., Jr. Comparison of Near-Infrared Spectroscopy Calibration Methods for the Prediction of Protein, Oil, and Starch in Maize Grain // *J. Agric. Food Chem.* 1991. V. 39, № 5. P. 883–886.
192. Surkova A., Bogomolov A., Legin A., Kirsanov D. Calibration transfer for LED-based optical multisensor systems. // *ACS Sensors*. 2020. V. 5. P. 2587–2595.
193. Свидетельство о государственной регистрации программы для ЭВМ «ТРТ Cloud» № 2016616371, Дата выдачи 09.06.2016. Авторы: Галянин В.В., Богомолов А.Ю., Беликов В.А., Беликова В.А., Мелентьева А.А.
194. Свидетельство о государственной регистрации программы для ЭВМ "Interval Selection toolbox" № 2015660526, Дата выдачи 01.10.2015. Авторы: Галянин В.В., Богомолов А.Ю., Беликов В.А., Беликова В.А., Мелентьева А.А.
195. Bogomolov A., Melenteva A. Tenth Winter Symposium on Chemometrics

- (WSC10). // *J. Chemom.* 2017. V. 31 (8). art. no. e2906.
196. Bogomolov A. Designing a Multi-Component Calibration Experiment: Basic Principles and Diagonal Approach. In: *Comprehensive Chemometrics* (2nd ed.), Chapter 1.15; Brown S., Tauler R., Walczak B., Eds. Amsterdam: Elsevier B.V., 2020, P. 411–430.
197. Brereton R.G. Multilevel multifactor designs for multivariate calibration. // *Analyst.* 1997. V. 122. P. 1521–1529.
198. Kirsanov D., Babain V., Agafonova-Moroz M., Lumpov A., Legin A. Combination of optical spectroscopy and chemometric techniques – a possible way for on-line monitoring of spent nuclear fuel (SNF) reprocessing. // *Radiochim. Acta.* 2012. V. 100. P. 185–188.
199. Surkova A., Belikova V., Kirsanov D., Legin A., Bogomolov A. Towards an optical multisensor system for dairy: Global calibration for fat analysis in homogenized milk. // *Microchem. J.* 2019. V. 149. art. no. 104012.
200. Walstra P., Geurts T.J., Noomen A., Jellema A., van Boekel M.A.J.S. *Dairy technology: principles of milk properties and processes.* New York: Marcel Dekker Inc., 1999. 752 p.
201. Cimander C., Mandenius C.-F. Bioprocess control from a multivariate process trajectory. // *Bioprocess Biosyst. Eng.* 2004. V. 26. P. 401–411.
202. MacGregor J.F. Data-based latent variable methods for process analysis, monitoring and control. // *Comput. Aided Chem. Eng.* Eds. Barbosa A., Matos H. In: *European Symposium on Computer-Aided Process Engineering-14.* 2004. V. 18. P. 87–98.
203. synTQ for PAT software package. <http://www.optimal-ltd.co.uk/syntq-for-pat> [23.08.2019].
204. Kourti T. The process analytical technology initiative and multivariate process analysis, monitoring and control. // *Anal. Bioanal. Chem.* 2006. V. 384. P. 1043–1048.
205. Bogomolov A. An in-line method for testing of fat and protein in milk. // *DIAA Victoria State Conference.* Melbourne, Australia, May 8–9, 2013.
206. Bogomolov A., Kucheryavski S., Melichar M., Wigmore A. Multi-Technique

- process monitoring: method development and application to pellet coating. // 11th Scandinavian Symposium on Chemometrics (SSC11). Loen, Norway, June 8–11, 2009.
207. Eriksson L., Johansson E., Kettaneh-Wold N., Trigg J., Wirkström C., Wold S. Multi- and megavariable data analysis. Umeå, Sweden: Umetrics Academy, 2006. 425 p.
208. Kucheryavski S., Esbensen K.H., Bogomolov A. Monitoring of pellet coating process with image analysis – a feasibility study. // J. Chemom. 2010. V. 24. P. 472–480.
209. Kucheryavski S., Bogomolov A., Esbensen K.H. Monitoring of pellet coating process with multivariate image regression. // 11th Scandinavian Symposium on Chemometrics (SSC11). Loen, Norway, June 8–11, 2009.
210. Andersson M., Folestad S., Gottfries J., Johansson M.O., Josefson M., Wahlund K.-G. Quantitative analysis of film coating in a fluidized bed process by in-line NIR spectrometry and multivariate batch calibration. // Anal. Chem. 2000. V. 72. P. 2099–2108.
211. ГОСТ Р 52054–2003. Молоко коровье сырое. Технические условия. М.: Стандартинформ, 2008. 12 с.
212. ГОСТ 5867–90. Молоко и молочные продукты. Методы определения жира. М.: Стандартинформ, 2009. 12 с.
213. ГОСТ 25179–90. Молоко. Методы определения белка. М.: Стандартинформ, 2009. 6 с.
214. Ehlen L., Zabarylo U.J., Speichinger F., Bogomolov A., Belikova V., Bibikova O., Artyushenko V., Minet O., Beyer K., Kreis M.E., Kamphues C. Synergy of fluorescence and near-infrared spectroscopy in detection of colorectal cancer. // J. Surg. Res. 2019. 2019. V. 242. P. 349–356.
215. Гурьянова А.О., Ермаков В.В., Максина Е.В., Раменская Е.В., Сахарова Т.В., Быков Д.Е., Артюшенко В.Г., Богомолов А.Ю. Определение массовой доли нефтепродуктов в почвогрунте по спектрам нарушенного полного внутреннего отражения. // Экол. и пром-сть России. 2013. Т. 12. С. 24–28.

216. ГОСТ 32255–2013 Молоко и молочные продукты. Инструментальный экспресс-метод определения физико-химических показателей идентификации с применением инфракрасного анализатора. М.: Стандартинформ, 2014. 16 с.
217. Tsenkova R., Atanassova S., Toyoda K., Ozaki Y., Itoh K., Fearn T. Near-infrared spectroscopy for dairy management: measurement of unhomogenized milk composition. // *J. Dairy Sci.* 1999. V. 82, № 11. P. 2344–2351.
218. Tsenkova R., Atanassova S., Itoh K., Ozaki Y., Toyoda K. Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers. // *J. Animal Sci.* 2000. V. 78. P. 515–522.
219. Wu D., He Y., Feng S. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. // *Anal. Chim. Acta.* 2008. V. 610, № 2. P. 232–242.
220. Kalinin A.V., Krasheninnikov V.N. Portable milk product quality analyzer based on spectrometry in the near IR range. // *Appl. Spectrosc.* 2008. V. 75. P. 274–279.
221. Kalinin A., Krivtsun V., Krasheninnikov V., Sadovskiy S., Denisovich H., Yurova H. Calibration models for multi-component quantitative analyses of dairy with the use of two different types of portable near infrared spectrometer. // *J. Near Infrared Spectrosc.* 2008. V. 16. P. 343–348.
222. Šašić S., Ozaki Y. Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein and lactose in raw milk by partial least-squares regression and band assignment. // *Anal. Chem.* 2001. V. 73, № 1. P. 64–71.
223. Chen J.Y. Effect of multiplicative scatter correction on wavelength selection for near infrared calibration to determine fat content in raw milk. // *J. Near Infrared Spectrosc.* 2002. V. 10, № 4. P. 301–307.
224. Rémillard N., Robin O., Martel R., Paquin P. Influence of homogenization efficiency on milk fat content determination by infrared analysis. // *Int. Dairy J.* 1993. V. 3, № 3. P. 197–208.
225. Robin O., Paquin P. Evaluation of the particle size of fat globules in a milk model emulsion by photon correlation spectroscopy. // *J. Dairy Sci.* 1991. V. 74, № 8.

- P. 2440–2447.
226. Urban C., Schurtenberger P. Application of a new light scattering technique to avoid the influence of dilution in light scattering experiments with milk. // *Phys. Chem. Chem. Phys.* 1999. V. 17, № 1. P. 3911–3915.
227. Attaie R., Richter R.L. Size distribution of fat globules in goat milk. // *J. Dairy Sci.* 2000. V. 83, № 5. P. 940–944.
228. Muñiz R., Pérez M.A., De la Torre C., Carleos C.E., Corral N., Baro J.A. Comparison of principal component regression (PCR) and partial least square (PLS) methods in prediction of raw milk composition by Vis–NIR spectrometry, application to development of on-line sensors for fat, protein and lactose content. // XIX IMEKO World Congress Fundamental and Applied Metrology. Lisbon, Portugal, September 6–11, 2009. Conference materials, ISBN: 978-963-88410-0-1. 2009. P. 2564–2568.
229. Kalinin A.V., Krasheninnikov V.N., Potapov A.V. Effect of dispersion structure variation on chemometrical calibration of near-infrared spectrometer: protein fractions in milk and reversed micelles solutions. // *Chemom. Intell. Lab. Syst.* 2009. V. 97. P. 33–38.
230. Xin Q., Ling H.Z., Long T.J., Zhu Y. The rapid determination of fat and protein content in fresh raw milk using the laser light scattering technology. // *Opt. Lasers Eng.* 2006. V. 44. P. 858–869.
231. Walstra P. Turbidimetric method for milk fat determination. // *J. Dairy Sci.* 1967. V. 50, № 11. P. 1839–1840.
232. Ashworth U.S. Turbidimetric methods for measuring fat content of homogenized milk. // *J. Dairy Sci.* 1969. V. 52, № 2. P. 262–263.
233. Abebe E. Determination of fat and protein content of milk using He-Ne laser light scattering. M.Sc. Thesis. Addis Ababa University, 2007.
234. Kalinin A.V. Near infrared spectroscopy method for simultaneously determining the components of raw milk. Patent EP 2 418 474 A1.
235. Bogomolov A., Melenteva, A. Galyanin V. Hedgehog in the fog: Scatter-based quantitative analysis of milk below 1100 nm. 17th International Diffuse Reflectance

- Conference (IDRC). // Chambersburg, PA, USA, August 2 – 8. https://www.cnirs.org/content.aspx?page_id=22&club_id=409746&module_id=160542 [23.08.2019]
236. Luginbühl W. Evaluation of designed calibration samples for casein calibration in Fourier transform infrared analysis of milk. // *Lebensmittel-Wissenschaft und -Technologie*. 2002. V. 35, № 6. P. 554–558.
237. Dahm D.J. Explaining some light scattering properties of milk using representative layer theory. // *J. Near Infrared Spectrosc.* 2013. V. 21, № 5. P. 322.
238. Bro R. Multiway calibration. Multilinear PLS. // *J. Chemom.* 1996. V. 10. P. 47–61.
239. Roggo Y., Chalus P., Maurer L., Lema-Martinez C., Edmond A., Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. // *J. Pharm. Biomed. Anal.* 2007. V. 44. P. 683–700.
240. Rantanen J., Räsänen E., Tenhunen J., Käsäkoski M., Mannermaa J.-P., Yliruusi J. In-line moisture measurement during granulation with a four-wavelength near infrared sensor: an evaluation of particle size and binder effects. // *Eur. J. Pharm. Biopharm.* 2000. V. 50. P. 271–276.
241. Luukkonen P., Fransson M., Björn I. N., Hautala J., Lagerholm B., Folestad S. Real-time assessment of granule and tablet properties using inline data from a high-shear granulation process. // *J. Pharm. Sci.* 2008. V. 97. P. 950–959.
242. Burggraeve A., Silva A.F., van den Kerkhof T., Hellings M., Vervaet C., Remon J. P., Heyden Y.V., De Beer T. Development of a fluid bed granulation process control strategy based on real-time process and product measurements. // *Talanta*. 2012. V. 100. P. 293–302.
243. Mantanus J., Ziémons E., Lebrun P., Rozet E., Klinkenberg R., Streel B., Evrard B., Hubert Ph. Moisture content determination of pharmaceutical pellets by near infrared spectroscopy: method development and validation. // *Anal. Chim. Acta*. 2009. V. 642. P. 186–192.
244. Windig W., Gallagher N., Boruta M., Hachey M., Bogomolov A. Self-modeling mixture analysis by pure variable methods. // Federation of Analytical Chemistry and Spectroscopy Societies (FACSS). Quebec, Canada, October 9–13, 2005.

245. Socrates G. Infrared characteristic group frequencies: tables and charts (2nd ed.). Toronto: Wiley, 1994.
246. Bystritskaya E.V., Pomerantsev A.L., Rodionova O.Ye. Non-linear regression analysis: new approach to traditional implementations. // *J. Chemom.* 2000. V. 14. P. 667–692.
247. Jamrógiewicz M. Application of the near-infrared spectroscopy in the pharmaceutical technology. // *J. Pharm. Biomed. Anal.* 2012. V. 66. P. 1–10.
248. Tamburini E., Marchetti M.G., Pedrini P. Monitoring key parameters in bioprocesses using near-infrared technology. // *Sensors.* 2014. V. 14. P. 18941–18959.
249. Lourenço N.D., Lopes J.A., Almeida C.F., Sarraguça M.C., Pinheiro H.M. Bioreactor monitoring with spectroscopy and chemometrics: a review. // *Anal. Bioanal. Chem.* 2012. V. 404. P. 1211–1237.
250. Landgrebe D., Haake C., Hopfner T., Beutel S., Hitzmann B., Scheper T., Rhiel M., Reardon K.F. On-line infrared spectroscopy for bioprocess monitoring. // *Appl. Microbiol. Biotechnol.* 2010. V. 88. P. 11–22.
251. Fayolle Ph., Picque D., Corrieu G. On-line monitoring of fermentation processes by a new remote dispersive middle-infrared spectrometer. // *Food Control.* 2000. V. 11. P. 291–296.
252. Kornmann H., Rhiel M., Cannizzaro C., Marison I., von Stockar U. Methodology for real-time, multianalyte monitoring of fermentations using an in-situ mid-infrared sensor. // *Biotechnol. Bioeng.* 2003. V. 82. P. 702–709.
253. Lidén G. In situ fluorescence measurements – clarifying or blurring the picture? // *Pure & Appl. Chem.* 1993. V. 65. P. 1927–1932.
254. Lindemann C., Marose S., Nielsen H.O., Scheper T. 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring. // *Sens. Actuator B-Chem.* 1998. V. 51. P. 273–277.
255. Volesky B., Srivastava A.K. Culture fluorescence as a marker of physiological state in fermentation process optimization. // *Anal. Chim. Acta.* 1991. V. 249. P. 279–290.

256. Hisiger S., Jolicoeur M. A multiwavelength fluorescence probe: is one probe capable for on-line monitoring of recombinant protein production and biomass activity? // *J. Biotechnol.* 2005. V. 117. P. 325–336.
257. Haak M.B., Eliasson A., Olsson L. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. // *J. Biotechnol.* 2004. V. 114. P. 199–208.
258. Hantelmann K., Kollecker M., Hüll D., Hitzmann B., Scheper T. Two-dimensional fluorescence spectroscopy: a novel approach for controlling fed-batch cultivations. // *J. Biotechnol.* 2006. V. 121. P. 410–417.
259. Skibsted E., Lindemann C., Roca C., Olsson L. On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. // *J. Biotechnol.* 2001. V. 88. P. 47–57.
260. Mortensen P.P., Bro R. Real-time monitoring and chemical profiling of a cultivation process. // *Chemom. Intell. Lab. Syst.* 2006. V. 84. P. 106–113.
261. Bro R. PARAFAC. Tutorial and applications. // *Chemom. Intell. Lab. Syst.* 1997. V. 38. P. 149–171.
262. Thygesen L.G., Rinnan Å., Barsberg S., Møller E.K.S. Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. // *Chemom. Intell. Lab. Syst.* 2004. V. 71. P. 97–106.
263. Engelen S., Møller S.F., Hubert M. Automatically identifying scatter in fluorescence data using robust techniques. // *Chemom. Intell. Lab. Syst.* 2007. V. 86. P. 35–51.
264. Rinnan Å., Booksh K.S., Bro R. First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. // *Anal. Chim. Acta.* 2005. V. 537. P. 349–358.
265. Scheper Th., Schügerl K. Culture fluorescence studies on aerobic continuous cultures of *Saccharomyces cerevisiae*. // *Appl. Microbiol. Biotechnol.* 1986. V. 23. P. 440–444.
266. Siano S.A., Mutharasan N. NADH and flavin fluorescence responses of starved yeast cultures to substrate additions. // *Biotechnol. Bioeng.* 1989. V. 34. P. 660–670.
267. Ibrahim M., Alaam M., El-Haes H., Jalbout A. F., de Leon A. Analysis of the

- structure and vibrational spectra of glucose and fructose. // *Ecl. Quim.* 2006. V. 31, № 3. P. 15–21.
268. Tipson R.S. Infrared spectroscopy of carbohydrates. National Bureau of Standards Monograph 110, Washington D.C. 1968.
269. Enzymatic ethanol test by UV-method. https://food.r-biopharm.com/wp-content/uploads/sites/2/2012/06/roche_ifu_ethanol_en_10176290035_2017-08.pdf [01.06.2019]
270. Enzymatic D-glucose/ D-fructose test by UV-method. https://food.r-biopharm.com/wp-content/uploads/sites/2/2012/06/roche_ifu_glucose-fructose_de_10139106035_2017-08.pdf [01.06.2019]
271. Fullwood L.M., Clemens G., Griffiths D., Ashton K., Dawson T.P., Lea R.W., Davis C., Bonnier F., Byrned H.J., Baker M.J. Investigating the use of Raman and immersion Raman spectroscopy for spectral histopathology of metastatic brain cancer and primary sites of origin. // *Anal. Methods.* 2014. V. 6. P. 3948–3961.
272. Artyushenko V., Schulte F., Zabarylo U., Berlien H.-P., Usenov I., Saeb Gilani T., Eichler H., Pieszczek Ł., Bogomolov A., Krause H., Minet O. Spectral fiber sensors for cancer diagnostics *in vitro*. // *Proceedings of the European Conference on Biomedical Optics.* Munich, Germany, June 21–25, 2015.
273. Kallenbach-Thieltges A., Großerüschkamp F., Mosig A., Diem M., Tannapfel A., Gerwert K. Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections. // *J. Biophotonics.* 2013. V. 6. P. 88–100.
274. Dukor R.K. Vibrational spectroscopy in the detection of cancer. In: *Handbook of Vibrational Spectroscopy.* Hoboken, NJ, USA: John Wiley & Sons, Inc. 2006.
275. Nachabé R., Evers D.J., Hendriks B.H., Lucassen G.W., Van Der Voort M., Wesseling J., Ruers T.J. Effect of bile absorption coefficients on the estimation of liver tissue optical properties and related implications in discriminating healthy and tumorous samples. // *Biomed. Opt. Express.* 2011. V. 2. P. 600–614.
276. Zhang Y., Chen Y., Yu Y., Xue X., Tuchin V.V., Zhu D. Visible and near-infrared spectroscopy for distinguishing malignant tumor tissue from benign tumor and

- normal breast tissues in vitro. // *J. Biomed. Opt.* 2013. V. 18, № 7. art. no. 077003.
277. De Boer L.L., Molenkamp B.G., Bydlon T.M., Hendriks B.H., Wesseling J., Sterenberg H.J., Ruers T.J.M. Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries. // *Breast Cancer Res. Treat.* 2015. V. 152. P. 509–518.
278. Chang V.T.C., Bean S.M., Cartwright P.S., Ramanujam N. Visible light optical spectroscopy is sensitive to neovascularization in the dysplastic cervix. // *J. Biomed. Opt.* 2010. V. 15. art. no. 057006.
279. Mallia R.J., Narayanan S., Madhavan J., Sebastian P., Kumar R., Mathews A., Thomas G., Radhakrishnan J. Diffuse reflection spectroscopy: An alternative to autofluorescence spectroscopy in tongue cancer detection. // *Appl. Spectrosc.* 2010. V. 64. P. 409–418.
280. Kim S.B., Temiyasathit C., Bensalah K., Tuncel A., Cadeddu J., Kabbani W., Liu H. An effective classification procedure for diagnosis of prostate cancer in near infrared spectra. // *Expert Syst. Appl.* 2010. V. 37. P. 3863–3869.
281. Svaasand L.O. On the physical rationale of photodynamic therapy. // *Proc. SPIE* 10306. Future directions and application in photodynamic therapy, 103060I. San Diego, CA, USA, January 19–21, 1990.
282. Qi S., Ouyang Q., Chen Q., Zhao J. Real-time monitoring of total polyphenols content in tea using a developed optical sensors system. // *J. Pharm. Biomed. Anal.* 2014. V. 97. P. 116–122.
283. Yang H., He B., Dai X., Satpathy M., Yang L., Jiang H. FMTPen: a miniaturized handheld fluorescence molecular tomography probe for image-guided cancer surgery. // *Photonics.* 2015. V. 2. P. 279–287.
284. Ozaki M., Kagawa K., Arimoto H., Kominami Y., Sanomura Y., Yoshida S., Seo M.W., Kawahito S., Tanaka S. Spectral Study of metabolism-based autofluorescence and white-light reflectance for endoscopic tumor imaging. // *Proceedings of the 2015 37th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC)*. Milan, Italy, August 25–29, 2015. P. 5879–5883.

285. Valdés P.A., Kim A., Leblond F., Conde O.M., Harris B.T., Paulsen K.D., Wilson B.C., Roberts D.W. Combined fluorescence and reflectance spectroscopy for in vivo quantification of cancer biomarkers in low- and high-grade glioma surgery. // *J. Biomed. Opt.* 2011. V. 16. art. no. 116007.
286. Baker M.J., Trevisan J., Bassan P., Bhargava R., Butler H.J., Dorling K.M., Fielden P.R., Fogarty S.W., Fullwood N., Heys K.A., Hughes C., Lasch P., Martin-Hirsch P.L., Obinaju B., Sockalingum G.D., Sulé-Suso J., Strong R.J., Walsh M.J., Wood B.R., Gardner P., Martin F.L. Using Fourier transform IR spectroscopy to analyze biological materials. // *Nat. Protoc.* 2014. V. 9, № 8. P. 1771–1791.
287. Šablinskis V., Urboniene V., Ceponkus J., Laurinavicius A., Dasevicius D., Jankevičius F., Hendrixson V., Koch E., Steiner G. Infrared spectroscopic imaging of renal tumour tissue. // *J. Biomed. Opt.* 2011. V. 16. art. no. 096006.
288. Lima C., Goulart V., Côrrea L., Pereira T., Zzell D. ATR-FTIR Spectroscopy for the assessment of biochemical changes in skin due to cutaneous squamous cell carcinoma. // *Int. J. Mol. Sci.* 2015. V. 16. P. 6621–6630.
289. Kim A., Khurana M., Moriyama Y., Wilson B.C. Quantification of in vivo fluorescence decoupled from the effects of tissue optical properties using fiber-optic spectroscopy measurements. // *J. Biomed. Opt.* 2010. V. 15. art. no. 067006.
290. Kang J. W., Lue N., Kong C.-R., Barman I., Dingari N.C., Goldfless S.J., Niles J.C., Dasari R. R., Feld M.S. Combined confocal Raman and quantitative phase microscopy system for biomedical diagnosis. // *Biomed. Opt. Express.* 2011. V. 2. P. 2484–2492.
291. Artyushenko V., Zabarylo U., Bogomolov A., Minet O., Krause H., Sakharova T., Usenov I., Schulte F., Eichler H.J. Fiber spectroscopy for tumor margin detection – selection of the best methods. // *Biomed. Opt. – Optical Society of America Technical Digest (online)*. 2016. art. no. JTu3A.
292. Tunnell J.W., Desjardins A.E., Galindo L., Georgakoudi I., McGee S.A., Mirkovic J., Mueller M.G., Nazemi J., Nguyen F.T., Wax A., Zhang Q., Dasari R.R., Feld M.S. Instrumentation for multi-modal spectroscopic diagnosis of epithelial dysplasia. // *Technol. Cancer Res. Treat.* 2003. V. 2, № 6. P. 505–514.

293. Volynskaya Z., Haka A.S., Bechtel K.L., Fitzmaurice M., Shenk R., Wang N., Nazemi J., Dasari R.R., Feld M.S. Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy. // *J. Biomed. Opt.* 2008. V. 13, № 2. art. no. 024012.
294. Chang S.K., Mirabal Y.N., Atkinson E.N., Cox D., Malpica A., Follen M., Richards-Kortum R.J. Combined reflectance and fluorescence spectroscopy for in vivo detection of cervical pre-cancer. // *J. Biomed. Opt.* 2005. V. 10. art. no. 024031.
295. Nachabé R., Evers D.J., Hendriks B.H.W., Lucassen G.W., van der Voort M., Rutgers E.J., Vrancken Peeters M.J., Van der Hage J.A., Oldenburg H.S., Wesseling J., Ruers T.J.M. Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods. // *J. Biomed. Opt.* 2011. V. 16, № 8. art. no. 087010.
296. Drezek R., Dunn A., Richards-Kortum R. A pulsed finite-difference time-domain (FDTD) method for calculating light scattering from biological cells over broad wavelength ranges. // *Opt. Express.* 2000. V. 6. P. 147–156.
297. Roth W., Macher-Goeppinger S., Schirmacher P. Pathology and molecular pathogenesis of renal cell carcinoma. // *Nephrologe.* 2011. V. 6. P. 315–322.
298. Sokolov K., Galvan J., Myakov A., Lacy A., Lotan R., Richards-Kortum R.J. Realistic three-dimensional epithelial tissue phantoms for biomedical optics. // *J. Biomed. Opt.* 2002. V. 7. P. 148–156.
299. Панкратова К.Г., Щелоков В.И., Ступакова Г.А., Стрепетова А.В. Определение содержания нефтепродуктов в почве методом БИК-спектроскопии. 1. Основные предпосылки использования БИК-спектроскопии для оценки загрязнения почв. // *Плодородие.* 2012. Т. 2, № 65. С. 49–50.
300. Chakraborty S., Weindorf D.C., Zhu Y., Li B., Morgan C.L.S., Ge Y., Galbraith J. Spectral reflectance variability from soil physicochemical properties in oil contaminated soils. // *Geoderma.* 2012. V. 177–178. P. 80–89.
301. Гурьянова А.О. Совершенствование аналитического определения загрязняющих веществ в системе экологического мониторинга нефтезагрязненных почв: дис. канд. хим. наук: 03.02.08. Казань, 2019. 105 с.

ПРИЛОЖЕНИЕ А. ГЕНЕРАТОР СХЕМ ДИАГОНАЛЬНОГО ДИЗАЙНА НА ЯЗЫКЕ «МАТЛАВ» С ПРИМЕРАМИ

```

function [C,val] = doe4reg(k,first,last,firstval,sequence)

% DOE4REG Creates diagonal design matrix C for k factors:

% Syntax:
% [C,Nv] = doe4reg(k,first,last,firstval,sequence)

% Input:
% k - the number of factors;
% first - first ID (0 or 1 for designs with or without central point,
% respectively;
% last - terminal ID;
% firstval - first validation sample ID (optional: 1 by default);
% sequence - diagonal sequence (optional, ordinal numbers by default).

% Output:
% C - (N x k) design matrix, where N = last - first + 1 is the number of experiments
% val - validation set IDs

% Tested for Matlab version R2018b (9.5.0.944444)
% Author: Dr. Andrey Bogomolov
% Version 1.01 of 21 March, 2020

nd = 2^(k-1);
if nargin<5, sequence = 0:nd-1; end
if nargin<4, firstval = 1; end
if sequence(1:nd/2)-sequence(nd/2+1:end)==0, fold = 1; else fold = 0; end
is0 = first==0;

% Define k-dimensional hypercube
Us = dec2bin(sequence,k);
for i = 1:k, U(:,i) = str2num(Us(:,i)); end
D = ~U-U;

% Generate design matrix
C = [];
for i = 1:(last+1)/2
    Ci = i*D(mod(i-1,nd)+1,:);
    C = [C;Ci;-Ci];
end
C = C(1:last,:);
if is0, C = [zeros(1,k);C]; else C = sign(C).*(abs(C)-0.5); end

% Generate validation set
val = [];
if firstval==0, val = 0; firstval = 3; end
valfree = 2^(k-1-fold)+2;
val = sort([val firstval:6:last-valfree (firstval:6:last-1-valfree)+1]);
end

```

```
% Examples:
% 1) "Risky" extra short (9 diagonal samples plus central one) two-factor
% diagonal design without validation set:
% C = doe4reg(2,0,9)

% 2) Diagonal design with 26 samples (no central point) for two factors;
% validation subset starts from ID = 3:
% [C,val] = doe4reg(2,1,26,3)

% 3) Well-balanced diagonal design with 51 samples for three factors
% including central sample and validation set starting from ID = 1:
% [C,val] = doe4reg(3,0,50,1)

% 4) Four-factor diagonal design with 55 samples using an optimized sequence:
% [C,val] = doe4reg(4,0,54,1,[0 3 5 6 7 1 4 2])

% 5) Folded four-factor diagonal design based on example 4) above:
% [C,val] = doe4reg(4,0,54,1,[0 3 5 6 0 3 5 6])

% 6) Example of building an augmented design for four factors with 16 + 16 + 1 = 33
% samples
% C1 = doe4reg(4,0,16,1,[0 3 5 6 7 1 4 2])
% C2 = doe4reg(4,0,16,1,[7 1 4 2 0 3 5 6])
% C = [C1;C2(2:end,:)]
```

ПРИЛОЖЕНИЕ Б. СТАНДАРТНЫЕ ОБРАЗЦЫ

Таблица Б-1. Состав (г/100 г) наборов стандартных образцов для приготовления градуировочных серий «Лабораторной» модели.

Эксперимент	Набор	Ид. ^а	%Жир	%Белок	%Лактоза
Построение лабораторной градуировочной модели (раздел 5.1.3)	F	F1	2.369±0.008 ^б	3.47	4.82
		F2	3.280±0.004	3.49	4.76
		F3	4.238±0.011	3.573±0.013	4.790±0.016
		F4	5.459±0.015	3.61	4.75
	P	P1	3.64	2.997±0.008	4.234±0.018
		P2	4.00	3.373±0.012	4.70
		P3	4.32	3.657±0.011	4.78
		P4	4.13	4.089±0.013	5.373±0.018
Перенос модели (раздел 5.1.4) и расчёт оптимальной конфигурации светодиодной ОМС (раздел 5.1.5);	F	F1	2.22	3.58	4.82
		F2	3.26	3.60	4.75
		F3	4.08	3.53	4.79
		F4	5.26	3.48	4.80
	P	P1	3.55	2.98	4.24
		P2	4.00	3.36	4.68
		P3	4.65	3.76	4.81
		P4	4.19	4.14	5.06
Построение градуировочной модели на цифровых изображениях (раздел 5.1.5) и пространственно-разрешённый безэталонный анализ (раздел 5.1.6)	F	F1	1.99±0.003	3.45	4.90
		F2	3.23±0.005	3.47	4.82
		F3	4.22±0.009	3.61±0.007	4.73±0.005
		F4	5.47±0.010	3.21	4.80
	P	P1	3.63	2.99	4.24
		P2	4.27	3.30	4.65
		P3	4.03	3.71	4.85
		P4	4.33	4.05	5.32
Исследование влияния размеров жировых глобул на спектральные свойства образца (раздел 5.1.2)	F	F _{low}	1.99	3.45	-
		F _{high}	5.47	3.21	-
	P	P _{low}	4.00	3.36	-
		P _{high}	4.19	4.14	-

Примечание: ^а идентификационный номер стандарта (оригинальная маркировка производителя); ^б Среднее значение и 95%-й доверительный интервал с учётом однородности материала и результата арбитражных измерений от нескольких независимых аккредитованных лабораторий. Были использованы следующие арбитражные методы: метод Кьельдаля для определения массовой доли белка (ASU L 06.00-10), метод Розе-Готтлиба для определения массовой доли жира (ASU L 01.00-9), ферментативный метод (ASU L 01.00-17) и высокоэффективная жидкостная хроматография для определения массовой доли лактозы; ^в значения без доверительного интервала получены в «QSE GmbH» с помощью ИК-спектроскопии на приборе «MilkoscanTM FT6000» («FOSS», Дания).

**ПРИЛОЖЕНИЕ В. ПРАКТИЧЕСКИЕ РАЗРАБОТКИ АНАЛИТИЧЕСКИХ УСТРОЙСТВ С
ИСПОЛЬЗОВАНИЕМ МУЛЬТИСЕНСОРНОГО ПОДХОДА**

Таблица В-1. Разработанные методики и устройства с использованием мультисенсорного подхода для химического анализа в различных прикладных областях.

Аналиты, обобщенные показатели	Спектральный метод	Аналитические устройства	Прикладная задача	Достигнутые показатели	Раздел диссер.
<i>Определение индивидуальных химических веществ и обобщенных показателей химического состава</i>					
Жиры (суммарно), белок (суммарно)	Вид/КВБИК-спектроскопия диффузного пропускания или отражения	компактный ДМД-спектрометр, светодиодная ОМС, зонд ДО с пространственным разрешением	новая лабораторная, полевая или онлайн-методика определения питательной ценности натурального или нормализованного молока	погрешности ^а определения жира и белка <0.1% ^б	5.1
Вода	БИК- и КР-спектроскопия ДО, в т.ч. комбинация методов	оптоволоконный онлайн-анализатор (процесс-спектрометр) и зонд ДО, ОМС на основе микроспектрометра	улучшенная онлайн-методика мониторинга технологического процесса сушки фармацевтических препаратов и добавок	погрешность определения воды снижена до <0.1% ^в	5.2.2, 5.2.3
	ИК-спектроскопия НПВО	оптоволоконный анализатор с НПВО-зондом	новая полевая методика определения воды в почве совместно с углеводородами нефти	погрешность определения воды ~0.6% ^г	5.5
Сахароза, ацетаминофен, полимерный состав защитной оболочки (суммарно)	БИК- и КР-спектроскопия ДО, в т.ч. комбинация методов	зондовые онлайн-анализаторы для БИК- и КР-спектроскопии (в т.ч. совместный анализ)	новые и улучшенные онлайн-методики контроля качества в производстве фармацевтических препаратов (пеллет)	экономия ресурсов за счёт более точного определения времён окончания стадий процесса	5.2.3

Аналиты, обобщенные показатели	Спектральный метод	Аналитические устройства	Прикладная задача	Достигнутые показатели	Раздел диссер.
Глюкоза, фруктоза, этанол, биомасса (суммарно)	флуориметрия, 2D-флуориметрия, БИК-спектроскопия, ИК-спектроскопия	оптоволоконные процесс-спектрометры и зонды ДО, НПВО	новые и улучшенные онлайн-методики контроля процесса ферментации в биотехнологическом производстве	совместное определение компонентов с погрешностями <math><10 \text{ г/л}^{\text{д}}</math>	5.3
Углеводороды нефти (суммарно)	ИК-спектроскопия НПВО	оптоволоконный анализатор и НПВО-зонд	новая полевая методика экологического мониторинга загрязнения почв нефтепродуктами при различной влажности почвы	определение углеводородов нефти с погрешностью $\sim 1\%^{\text{е}}$	5.5
<i>Смежные с химическим анализом сферы: биомедицинская диагностика, онлайн-контроль качества материалов в ходе технологического процесса (анализ без индивидуального определения химических компонентов)</i>					
Химические (липиды, вода, гликоген, PO_2^- , ФАД, коллаген, порфирины, аминокислотные остатки и другие) и морфологические характеристики биотканей	БИК-спектроскопия	светодиодная ОМС с зондом ДО	новая методика оптической диагностики биотканей для определения границы опухоли (рак почки)	показан диагностический потенциал светодиодной ОМС	5.4.3
	ИК-спектроскопия НПВО и флуориметрия (в т.ч. совместно)	оптоволоконные спектрометры и зонды НПВО, ДО	новая методика оптической диагностики биотканей для определения границы опухоли (рак почки)	снижена вероятность ошибочных отнесений I-го и II-го рода ^ж	5.4.4
	БИК-спектроскопия и флуориметрия (в т.ч. совместно)	оптоволоконные спектрометры и зонды ДО	новая методика оптической диагностики биотканей для определения границы опухоли (рак прямой кишки)	снижена вероятность ошибочных отнесений I-го и II-го рода	5.4.4

Аналиты, обобщенные показатели	Спектральный метод	Аналитические устройства	Прикладная задача	Достигнутые показатели	Раздел диссер.
Химические (вода, сахара, ацетаминофен, суммарно определяемые компоненты защитного покрытия), морфологические и динамические характеристики среды процесса	БИК-спектроскопия	оптоволоконный процесс-спектрометр и зонд ДО	новая онлайн-методика прогнозирования в ходе технологического процесса будущей кинетики высвобождения ацетаминофена после приёма лекарства пациентом	показана возможность повышения качества препаратов целевой доставки (пеллет) за счёт проактивного (КПС) контроля процесса производства	5.2.4
Толщина покрытия пеллет (размер)	Анализ изображений в Вид-области	Цифровая фотокамера	Новая методика определения конечной точки процесса покрытия пеллет «у линии»	экономия ресурсов за счёт более точного определения времени окончания процесса	5.2.3

Примечания: ^a здесь и далее в таблице: погрешность (абсолютная) – среднеквадратичное отклонение прогноза на проверочном наборе (СКОП); ^b в практически используемом диапазоне концентраций жира и общего белка (табл. Б-1, приложение Б); ^b в диапазоне концентраций воды 1–10%; ^r в диапазоне концентраций воды 1–13%; ^д в практически используемых диапазонах концентраций (этанола: 0–144 г/л, глюкозы: 0–288 г/л, фруктозы: 0–120 г/л); ^e в диапазоне концентраций ОУН 1–13%; ^ж число ложно положительных (ошибка I-го рода) и ложно отрицательных (ошибка II-го рода) отнесений по результатам проверки модели.