

# ИНТЕРАКТИВНОЕ ОПРЕДЕЛЕНИЕ СВОЙСТВ НАФТЫ (ЛИГРОИНА) КАК ЭТИЛЕНОВОГО СЫРЬЯ С ИСПОЛЬЗОВАНИЕМ СПЕКТРОСКОПИИ В БЛИЖНЕМ ИНФРАКРАСНОМ ДИАПАЗОНЕ

© 2023 г. Chen Fan<sup>1</sup>, Tianbo Liu<sup>2</sup>, Guihua Hu<sup>1</sup>, Minglei Yang<sup>1</sup>, Jian Long<sup>1,3,\*</sup>

<sup>1</sup> Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, 200237 China

<sup>2</sup> Sinopec Jinan company, Jinan, 250102 China

<sup>3</sup> Qingyuan Innovation Laboratory, Quanzhou, 362801 China

\*E-mail: longjian@ecust.edu.cn

Поступила в редакцию 28 июля 2021 г.

После доработки 8 ноября 2022 г.

Принята к публикации 27 ноября 2023 г.

Предоставление актуальной информации о свойствах нефти как этиленового сырья в режиме реального времени имеет большое значение для улучшения моделирования, управления и оптимизации процессов. Интерактивная (в режиме онлайн) система измерения спектра полного пропускания в ближнем ИК-диапазоне (NIR), а также метод регрессии основных компонентов и метод частичных наименьших квадратов (PLS) использованы для разработки моделей, прогнозирующих свойства нефти при различных методах предварительной обработки. Прогнозируемые относительные ошибки моделей NIR, разработанных методом PLS, особенно для точек температуры фракционирования тестового набора, составляют в основном 1–5%, т.е. их можно использовать для создания приемлемых моделей прогнозирования в ближнем ИК-диапазоне при интерактивном определении свойств нефти. Сглаживание Савицкого–Голея в сочетании с предварительной обработкой первой производной обеспечивает наилучший эффект устранения шумов по сравнению с другими методами.

**Ключевые слова:** спектроскопия в ближнем ИК-диапазоне, свойства нефти, предварительная обработка, частичные наименьшие квадраты, интерактивное определение

**DOI:** 10.31857/S0028242123050076, **EDN:** RZPYBP

## Терминология

$X$  – необработанные данные;;

$X_{\text{norm}}$  – нормализованные данные;;

$X_{\text{max}}$  – максимальные значения в необработанных данных;;

$X_{\text{min}}$  – минимальные значения в необработанных данных;

$y$  – известное свойство;

$T$  – оценка;

$T^+$  – транспонирование  $T$ ;

$\hat{r}$  – вектор-столбец;

$R$  – набор  $r$ ;

$y_{(\text{PCR})}$  – PCR-аппроксимация вектора свойств;

$y_{\text{pre}}$  – свойство прогнозирования;

$y_{\text{real}}$  – свойство измерения;

$x_i$  – оптическая плотность точки спектра с  $i$ -м волновым числом;

$\beta_i$  – коэффициент.

## Сокращения по статье

ASTM – Американское общество испытаний и материалов;

BC – метод коррекции базовой линии;

FD – метод первой производной;

FBP – конечная температура кипения;

GC – газовый хроматограф;

IBP – начальная температура кипения;

NIR – ближний ИК-диапазон;

LVs – скрытые переменные;

PCA – анализ основных компонентов;

PCR – метод регрессии основных компонентов;

PCs – основные компоненты;

PIONA – парафины, изопарафины, олефины, нафтены и ароматические соединения;

PLS – метод регрессии с частичными наименьшими квадратами;

$R^2$  – коэффициент детерминации;

RMSE – среднеквадратическая ошибка оценки;

SG – метод Савицкого–Голея;

SG–FD – метод Савицкого–Голея в сочетании с методом первой производной;

SG–SNV – метод Савицкого–Голея в сочетании с методом анализа среднеквадратического отклонения случайной величины с нормальным распределением;

SG–BC – метод Савицкого–Голея в сочетании с методом коррекции базовой линии;

SNV – метод анализа среднеквадратического отклонения случайной величины с нормальным распределением.

Этилен – важное органическое соединение нефтехимической промышленности, характеризующееся крупнейшим массовым производством, значительно превосходящим производство других продуктов нефтехимии [1]. Затраты на сырье составляют более 90% общих производственных затрат [2]. Нафта – разновидность смешанной легкой нефти, состоящей из алканов, циклоалканов, ароматических соединений, олефинов и других веществ. Она является сырьем для производства этилена, синтетического аммиака, полиэстера и других материалов. В современных рыночных условиях в Китае способ получения олефинов крекингом нафты будет оставаться преобладающим в течение длительного времени.

Свойства нафты имеют решающее значение для реализации процесса производства этилена, включая моделирование, управление и оптимизацию процесса, особенно мониторинг и оптимизацию в

реальном времени [3]. Эти свойства контролируются с помощью хорошо зарекомендовавших себя стандартных процедур, таких как газовая хроматография (ГХ), но они отнимают много времени, являются дорогостоящими и трудоемкими. Важно предоставлять информацию о химических свойствах используемой нафты в режиме реального времени, минимизируя при этом время анализа [4].

Использование спектроскопии позволяет получать информацию о различных типах внутримолекулярных химических связей (таких как C–H, O–H или N–H) в ближнем ИК-диапазоне (NIR) для прогнозирования качества нафты и имеет такие преимущества, как быстрая, недорогая и минимально необходимая подготовка проб [5]. По мере развития технологии анализа методом спектроскопии в ближнем ИК-диапазоне появилось множество исследований, основанных на сочетании этого метода с хемометрическим калибровочным моделированием для быстрого определения некоторых важных химических и физических параметров нефти и ее дистиллятов [6], таких как неочищенная нефть [7, 8], бензин [9, 10], дизельное топливо [11, 12] и нафта [13]. Многие критические параметры нефти и ее продуктов, включая дистилляционные свойства, октановое число, химический состав и т. д., были точно спрогнозированы в лабораторных условиях с использованием новых хемометрических подходов, таких как метод машинного обучения [14–17]. Технологию NIR все чаще используют для интерактивного анализа, мониторинга и помощи при управлении такими процессами нефтепереработки, как крекинг нафты и смешивание бензинов [18]. Кроме того, для мониторинга промышленных потоков и процессов в ближнем инфракрасном диапазоне было предложено дополнительное применение NIR-спектроскопии через интегрированные волоконно-оптические сети, микроспектроскопию комбинационного рассеяния с Фурье-преобразованием [19] и релаксометрию протонного ЯМР [20]. Ученые выполнили ряд исследований по быстрому прогнозированию свойств нафты и предложили множество методов моделирования для автономного определения (в режиме онлайн) свойств нафты в лабораторных условиях.

В то же время нафта представляет собой сложную гидрогенизированную нефть, характеризующуюся большим разнообразием внутреннего строения, включая распределение числа атомов

углерода, молекулярную структуру, температуру кипения, содержание парафинов, изопарафинов, олефинов, нафтенов и ароматических соединений (PIONA). Свойства нефти и ее компонентов меняются в зависимости от свойств нефти, технологии переработки и условий эксплуатации оборудования. Более того, производственная среда процесса нефтепереработки динамична и изменчива. Температура и скорость нефти, подаваемой по трубопроводу, варьируются; кроме того, нефть может содержать некоторые примеси. Можно предсказать, что спектральный пик нефти в ближнем инфракрасном диапазоне будет большим, а это может привести к серьезному перекрытию и избыточности данных спектрального шума.

Режим пропускания NIR является весьма перспективным для комплексной оценки качества образца, если может быть получена и хорошо обработана спектральная информация. Одним из основных требований для интерактивной системы обнаружения является сбор спектров пропускания с высоким соотношением сигнал/шум. Для корреляции спектральных изменений с изменениями состава или свойств анализируемого соединения, обычно следует предварительно обработать спектральные данные образцов, чтобы улучшить возможности прогнозирования и повысить надежность построенных моделей. До настоящего времени опубликовано лишь несколько исследований по применению NIR-анализа для быстрого интерактивного измерения физических и химических параметров нефти в реальных заводских условиях. Данное исследование направлено на изучение использования NIR-спектроскопии с датчиком пропускания и разработки NIR-моделей в сочетании с хемометрикой с целью кратковременного и экономически эффективного прогнозирования свойств нефти для обеспечения быстрого анализа в режиме онлайн. Для анализа эффекта интерактивного прогнозирования свойств нефти были объединены различные методы предварительной обработки NIR-спектра и различные методы моделирования для разработки прогнозных моделей.

## ЭКСПЕРИМЕНТЫ И МЕТОДЫ

**Образцы нефти.** Образцы нефти в течение последних нескольких лет отбирали из подающего

трубопровода печи крекинга этилена на нефтехимическом предприятии, расположенном на востоке Китая. Они представляли собой смесь прямогонной нефти, риформинг-нафты и гидрогенизированной нефти, полученных из нефти после ее переработки методами атмосферной и вакуумной перегонки, каталитического риформинга и гидрокрекинга соответственно. Пробу объемом около 100 мл непрерывно отбирали в виде капель нефти из нефтепровода с помощью автоматического пробоотборника; время отбора проб составляло 30 мин. Частота пробоотбора зависела от изменения свойств нефти, перерабатываемой перегонной установкой, и от условий эксплуатации указанного оборудования. Ожидалось, что это приведет к явным различиям в свойствах различных образцов нефти, отбираемых каждые 2–3 дня. Плотность нефти составляла 620–650 г·см<sup>-3</sup>, нефть содержала большое количество углеводородных соединений, таких как C<sub>4</sub>–C<sub>10</sub>, и малое количество соединений серы и азота.

Характеристика жидкого сырья по диапазону дистилляции и содержанию PIONA имеет высокую исследовательскую и прикладную значимость для управления процессом крекинга этилена и его оптимизации. Диапазон дистилляции определяли стандартными методами испытаний Американского общества по испытаниям и материалам (ASTM) – D86. Анализ PIONA широко используют в области анализа нефти благодаря его качественной и количественной точности и большому объему информации. Газовая хроматография является важным способом определения состава и функциональных групп углеводородных мономеров нефти. Сотни компонентов нефти были разделены в зависимости от их температуры кипения на неполярной хроматографической капиллярной колонке высокого разрешения, при этом определенное количество соединений ассоциировалось с обнаруженными пиками. Содержание PIONA в нефти определяли методом лабораторного анализа по стандартной методике испытаний SH/T 0714-2002 с использованием газового хроматографа Agilent с размером колонки 50 м × 200 мкм × 0.5 мкм. Эксперименты проводились в следующих условиях: температура на входе для отбора проб составляла 250°C, коэффициент разделения – 280 : 1, нагрев осуществляли по следующему протоколу: начальная температура – 35°C, продолжительность нагрева – 10 мин, ско-

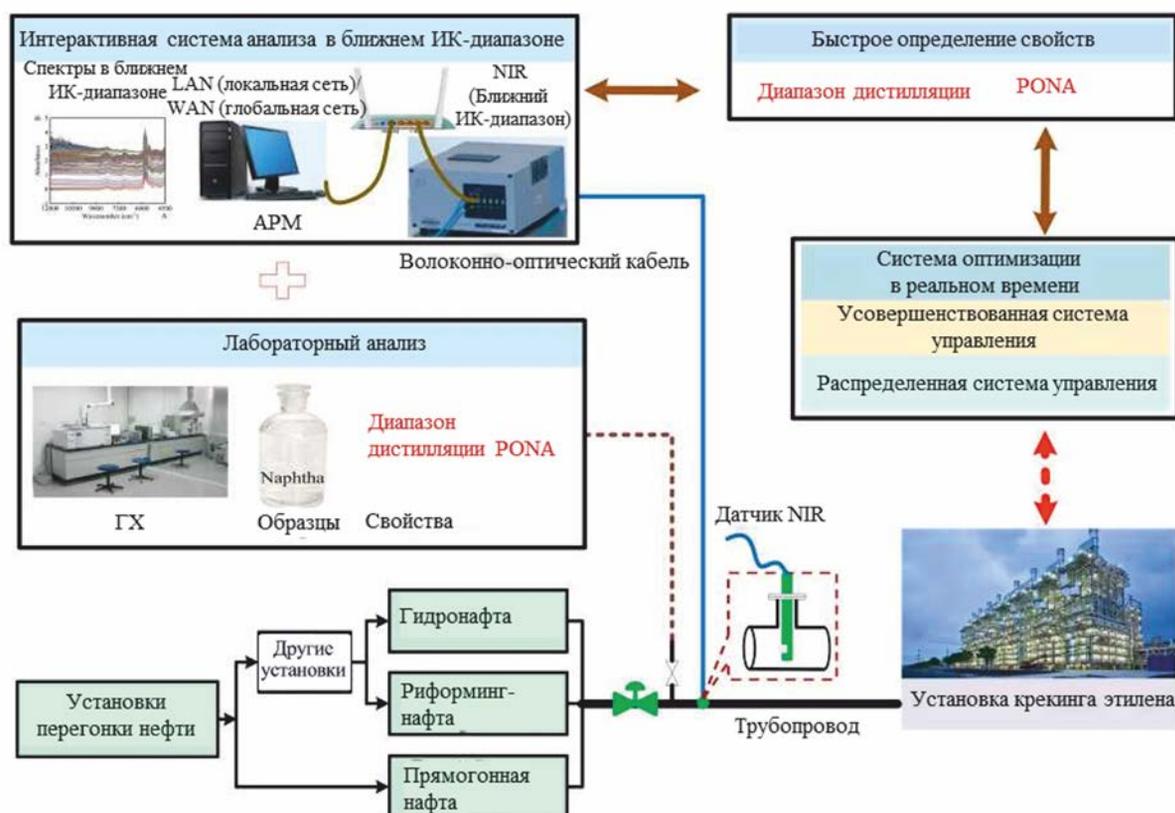


Рис. 1. Схема отбора проб нефти и NIR-измерений в режиме онлайн.

рость нагрева –  $0.5^{\circ}\text{C}\cdot\text{мин}^{-1}$  до  $60^{\circ}\text{C}$  и  $2^{\circ}\text{C}\cdot\text{мин}^{-1}$  до  $180^{\circ}\text{C}$ , температура пламенно-ионизационного детектора –  $300^{\circ}\text{C}$ , скорость потока  $\text{H}_2$  –  $35\text{ мл}\cdot\text{мин}^{-1}$ , а скорость потока воздуха –  $400\text{ мл}\cdot\text{мин}^{-1}$ .

**Интерактивное устройство обнаружения в ближнем инфракрасном диапазоне спектра.** Измерения пропускания (оптический путь  $1.0\text{ мм}$ ) нефти, текущей по трубопроводу, осуществляли погружным волоконно-оптическим датчиком, вставленным в верхнюю часть трубопровода. Для регистрации NIR-спектра образца с помощью программного обеспечения OPUS-NT, полученный через датчик сигнал направляли в интерактивный NIR-спектрометр, установленный на расстоянии  $30\text{ м}$  и соединенный с компьютером. Сканированные значения длины волны находились в диапазоне от  $4500$  до  $12\ 000\text{ см}^{-1}$ . Образцы нефти, транспортируемые по трубопроводу в период отбора проб, хранились при температуре примерно  $10\text{--}25^{\circ}\text{C}$ . В течение 30-минутного периода отбора проб получали

$15\text{ NIR}$ -спектров нефти с  $32$  сканами накопления для каждого спектра с интервалами  $20\text{ мин}$ . NIR-спектр образца, использованного для моделирования, строили в среднем на основе  $15$ -ти спектров для обеспечения спектральной воспроизводимости. Схематическое изображение системы отбора проб нефти и интерактивной системы NIR-измерений представлено на рис. 1. Данные, обработанные компьютером, и модели на их основе могут быть отправлены в распределенную систему управления, расширенную систему управления, систему оптимизации в реальном времени и т.д.

**Метод предварительной обработки спектральных данных.** В большинстве случаев фоновая информация (или шум) необработанных спектральных данных, отклонения и систематическая погрешность, вызванные несовершенством спектрометров, могут влиять на исходные спектры. Следовательно, для уменьшения воздействия на характеристики разработанных моделей прогнози-

вания любых помех необходима предварительная обработка исходных спектральных данных [19]. В настоящем исследовании изучалась пригодность этих методов с использованием некоторых методов предварительной спектральной обработки, таких как сглаживание Савицкого–Голея (SG), среднеквадратическое отклонение случайной величины с нормальным распределением (SNV), коррекция базовой линии (BC) и первая производная (FD).

Сглаживающая фильтрация является широко используемым методом предварительной обработки для спектрального анализа. Этот метод может улучшить спектральную гладкость, уменьшить шумовые помехи и выбрать различную ширину окна и порядок полиномиального аппроксимирования. Метод SNV также используют для приведения данных разных представлений к одному и тому же масштабу путем повторной настройки значения каждого измерения таким образом, чтобы конечный вектор данных попадал в интервал  $[0, 1]$  или  $[-1, 1]$  (1):

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

где  $X_{\text{norm}}$  – нормализованные данные;  $X$  – необработанные исходные данные;  $X_{\min}$  и  $X_{\max}$  – минимальное и максимальное значения необработанных данных.

Для метода BC спектр разбивают на несколько равных частей по длинам волн и используют ширину каждой части в качестве ширины окна для анализа. Значения, отображаемые в каждом окне, обрабатывают в соответствии с типом измеряемого спектра. Точки собственных значений в каждом окне соединяют прямой линией, а соединяющую кривую аппроксимируют приближенной линейной зависимостью. Точки, не соответствующие аппроксимированной кривой, считают недействительными и игнорируют. Кроме того, получают базовую функцию, а разницу между исходными спектральными данными и базовыми данными представляют как скорректированные спектральные данные.

Дифференцирование, которое также обычно используют для предварительной обработки спектра, может увеличить спектральную разницу, устранить дрейф базовой линии и улучшить спектральное

разрешение. Для обработки спектральных данных чаще всего применяют производные первого или второго порядка. При применении производной первого порядка можно удалить постоянную базовую линию, а использование производной второго порядка может удалить базовую линию в виде линейной функции; производные более высокого порядка используют редко. Существует ряд алгоритмов дифференцирования. В качестве примера рассмотрим метод дифференцирования Савицкого–Голея. Суть этого метода, также как и метода сглаживания SG, заключается в получении нескольких локальных сегментов всего спектра посредством перемещения окна, а затем – полиномиальных параметров путем аппроксимации каждого сегмента к полиномиальной функции. Вычисляют производную подобранной функции в точке центральной длины волны. Стратегия перемещения окна позволяет получить производные центральной длины волны всех малых участков спектра.

Для исследования использовали сглаживающий фильтр метода SG обработки NIR-спектров, объединенный с другими методами обработки сглаженных данных и, таким образом были сформированы различные методы предварительной обработки.

**Метод моделирования.** Для построения регрессионной модели на базе основных компонентов скрининга использовали метод регрессии основных компонентов (PCR). Анализ основных компонентов – это классический метод многомерного статистического анализа, который для уменьшения размерности данных может преобразовать несколько перекрестно коррелированных переменных в несколько некоррелированных переменных, а именно в основные компоненты (PC). Метод PCR реализуется в четыре этапа. Целью калибровки является нахождение зависимости между оценками и истинными концентрациями каждого соединения в смеси, которая описывается следующим уравнением (2) [22]:

$$y = T\hat{f}, \quad (2)$$

где  $y$  – известное свойство спектра смеси;  $T$  – оценка;  $\hat{f}$  – вектор-столбец, представляющий собой оцененный коэффициент регрессии, длина которого равна количеству основных компонентов.

Значение  $\hat{\mathbf{r}}$  можно оценить по известным значениям  $\mathbf{y}$  и  $\mathbf{T}$  (3):

$$\hat{\mathbf{r}} = \mathbf{T}^+ \mathbf{y}, \quad (3)$$

где  $\mathbf{T}^+$  – транспонирование оценки  $\mathbf{T}$ .

PCR-аппроксимацию вектора свойств  $\mathbf{y}_{(\text{PCR})}$  вычисляют по уравнению (4):

$$\mathbf{y}_{(\text{PCR})} = \mathbf{T}\mathbf{T}^+ \mathbf{y}. \quad (4)$$

Если имеется множество представляющих интерес соединений, вектор  $\mathbf{y}$  можно расширить до матрицы, в которой каждый столбец соответствует соединению (5):

$$\mathbf{Y} = \mathbf{T}\mathbf{R}, \quad (5)$$

где  $\mathbf{R}$  – набор векторов-столбцов  $\hat{\mathbf{r}}$ .

**Метод частичных наименьших квадратов (PLS).** Метод PLS является одним из самых перспективных методов многомерной калибровки и широко используется благодаря его нечувствительности к коллинеарным переменным и к большому количеству переменных. Основная идея алгоритма PLS заключается в том, что система управляется несколькими скрытыми переменными, и этими переменными может быть описана модель PLS. Метод PLS позволяет объяснить взаимосвязь между спектральными данными и представляющими интерес свойствами извлечением небольшого количества скрытых переменных (LV), отражающих исходную независимую информацию [23]. PLS-анализ может быть выполнен с использованием всех калибровочных спектров в представляющей интерес области. Скрытые переменные (LV) фиксируют максимальную ковариантность между эталонными данными и записанным спектром. Модели PLS были разработаны для создания математической модели корреляции между свойствами нефти и спектральными данными ближнего инфракрасного диапазона. Математическая корреляционная модель этого метода показана в уравнении (6):

$$y_{\text{pre}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i, \quad (6)$$

где  $y_{\text{pre}}$  – свойство прогнозирования;  $x_i$  – параметр модели;  $\beta_i$  – оптическая плотность точки спектра с  $i$ -м волновым числом.

**Оценка точности модели.** Во время построения модели прогнозирования для разработки калибровочной модели образцы соответствующего набора данных выбирают случайным образом, а оставшиеся образцы используют для тестирования характеристик модели. Калибровочные модели, основанные на подмножестве калибровок, оценивают посредством перекрестной калибровки. Для оценки моделей NIR обычно используют среднеквадратические ошибки (RMSE) и коэффициент детерминации ( $R^2$ ) наборов калибровки или прогнозирования. Чем меньше значения RMSE, тем выше точность. Чем ближе значение  $R^2$  к 1, тем лучше характеристики модели. Точность прогнозирования композиционных данных с использованием такой модели оценивают по вышеупомянутым показателям эффективности. Тем не менее, для большинства приложений в нефтеперерабатывающей промышленности обычно требуется, чтобы абсолютная и/или относительная погрешность прогнозирования свойств нефти попадала в определенный диапазон.

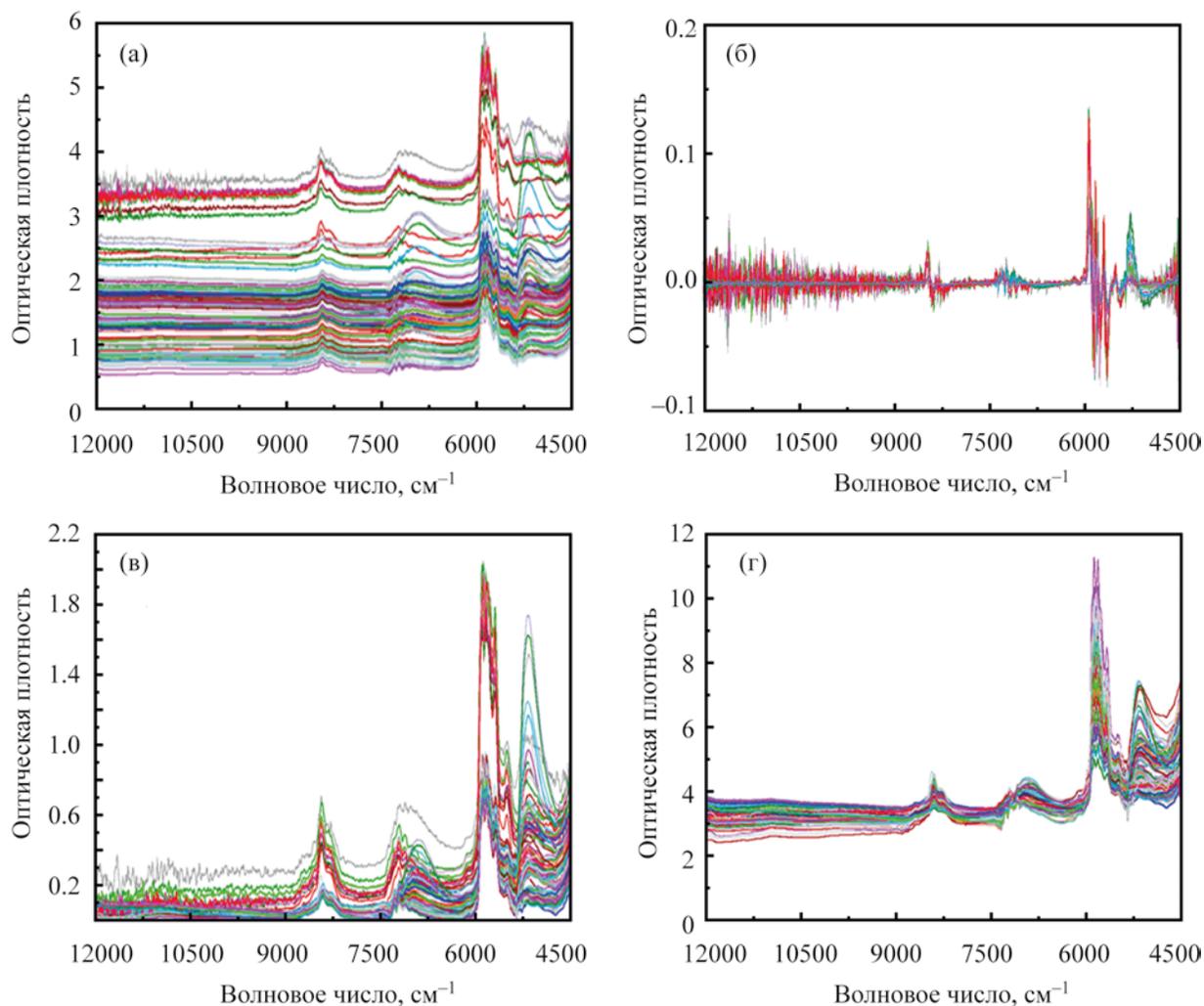
В данном исследовании для оценки точности модели прогнозирования использовались относительные ошибки предсказанных и измеренных значений NIR. Точность была рассчитана для наборов прогнозов по уравнению (7):

$$\delta = \left| \frac{y_{\text{pre}} - y_{\text{real}}}{y_{\text{real}}} \right| \times 100, \quad (7)$$

где  $\delta$  – относительная ошибка, характеризующая качество результатов при использовании модели прогнозирования;  $y_{\text{pre}}$  – прогнозируемое значение;  $y_{\text{real}}$  – измеренное значение.

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

**Предварительная обработка образца.** После удаления аномальных результатов измерений было получено 160 наборов данных (100 случайных групп использовались в качестве калибровочного набора, а остальные данные – в качестве тестового набора). Диапазон свойств и их средние значения



**Рис. 2.** Предварительная обработка NIR-спектра различными методами: а – необработанный спектр; б – спектр, предварительно обработанный методами SG-FD; в – спектр, предварительно обработанный методами SG-BC; г – спектр, предварительно обработанный методами SG-SNV.

были рассчитаны для полного набора данных, подмножества калибровки и подмножества прогнозирования, чтобы получить пригодные диапазон и распределение образцов как в подмножестве калибровки, так и в подмножестве прогнозирования.

**Сравнение методов предварительной обработки NIR-спектра.** Некоторая случайная информация, содержащаяся в полученных исходных спектрах, оказалась непригодной для построения надежных моделей, поэтому было изучено влияние предварительной обработки данных методом SG в сочетании с методами FD, BC и SNV на

спектральные данные анализируемых образцов. Спектры, полученные до и после такой предварительной обработки, представлены на рис. 2. Видно, что необработанные спектры имеют значительный случайный шум, поэтому их необходимо предварительно обработать для исключения нежелательных помех. При использовании методов SG-FD и SG-SNV можно сделать исходную спектральную интенсивность более плотной по сравнению с методом SG-BC, при этом также улучшаются пиковые свойства предварительно обработанных спектров, что позволяет предположить, что эти методы более эффективны в отношении снижения шума,

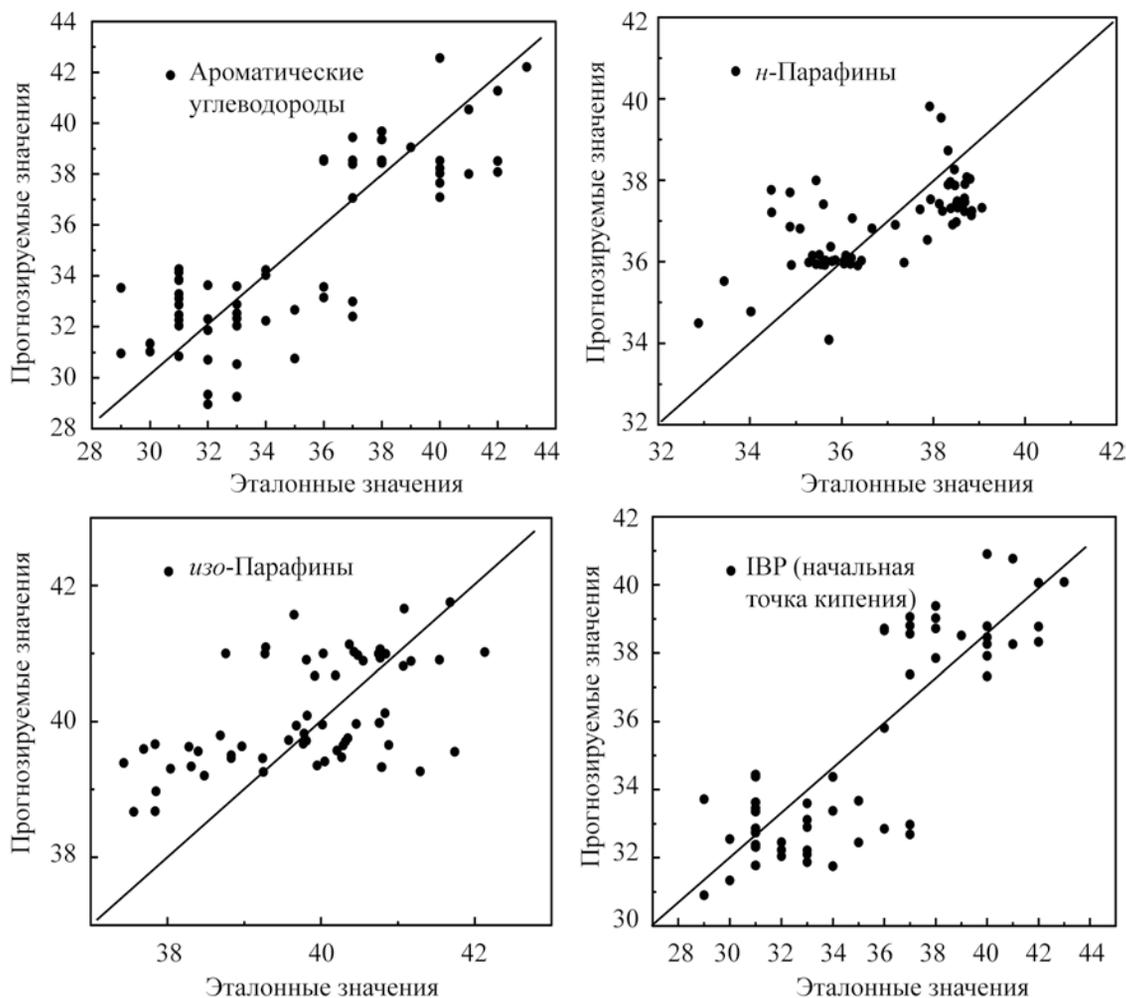


Рис. 3. Прогностический эффект моделей BG–BC и PCR в отношении начальной температуры кипения (IBP) и содержания *изо*-парафинов, *n*-парафинов и ароматических соединений.

сдвига сигнала и рассеяния света в исходных спектрах нефти.

**Выбор длины волны.** Учитывая, что в данных спектра образцов нефти слишком много выборок для сравнения переменных (длин волн), предварительно обработанные спектральные данные были сегментированы и смоделированы (сегмент за сегментом) с последующей оценкой качества модели. Для моделирования методом ступенчатой регрессии был выбран диапазон с высокой точностью модели, затем были выбраны длины волн в качестве переменных для повышения робастной устойчивости модели и надежности ее прогнозов. После стандартизации спектральных данных по ста образцам был проанализирован результат мо-

делирования с помощью метода PLS путем выбора различных волновых чисел. Для дальнейшей оптимизации переменных (длин волн) использовали метод ступенчатой регрессии до тех пор, пока не добивались соответствия значений RMSE и  $R^2$  моделей установленным требованиям. Наконец, для разработки моделей прогнозирования были выбраны обработанные спектральные данные в диапазоне  $4800\text{--}6800\text{ см}^{-1}$ , который охватывает типичные длины волн, относящиеся к критическим химическим связям, и содержит основные характеристические пики.

**Разработка модели NIR с использованием метода регрессии основных компонент.** Для построения модели после предварительной обра-

**Таблица 1.** Средняя относительная ошибка модели прогнозирования методом PCR, построенной для тестового набора данных различными методами предварительной обработки

Свойство		Метод		
		SG-FD	SG-BC	SG-SNV
Диапазон дистилляции	IBP (начальная температура кипения)	5.04	5.50	6.39
	10% Температура фракционирования	4.42	4.74	5.51
	30% Температура фракционирования	3.63	3.92	4.10
	50% Температура фракционирования	2.87	2.94	3.27
	70% Температура фракционирования	2.73	2.97	2.84
	90% Температура фракционирования	2.87	2.90	2.59
	95% Температура фракционирования	1.91	1.95	1.78
	IBP (конечная температура кипения)	1.13	1.14	1.04
Состав группы	Алканы	2.13	2.20	2.31
	<i>n</i> -Парафины	2.75	9.13	3.30
	<i>изо</i> -Парафины	2.18	2.03	2.05
	Ароматические углеводороды	7.47	7.65	7.72
	Нафтены	7.20	7.53	7.58

ботки методами SG-FD, SG-BC и SG-SNV был исследован прогностический эффект от использования метода PCR для таких параметров, как температура кипения и содержание PIONA. Эффективность моделей прогнозирования, разработанных в сочетании с методом SG-BC для калибровочного набора, оцененных в отношении начальной температуры кипения (IBP) и содержания *изо*- и *n*-парафинов и ароматических соединений, показана на рис. 3. Горизонтальные и вертикальные оси графиков представляют фактические и прогнозируемые значения свойств нефти соответственно. Диагональная линия на графике является целевой линией. Чем ближе точки к целевой линии, тем лучше прогнозирующий эффект модели. Модели, полученные методом PCR в сочетании со сглаживанием SG и коррекцией базовой линии в качестве предварительной обработки, продемонстрировали хороший прогностический эффект для калибровочного набора.

Кроме того, модели прогнозирования свойств нефти оценивались с использованием тестового набора для изучения влияния различных сочетаний методов PCR-моделирования и вышеупомянутых методов предварительной обработки. Средние относительные ошибки между прогнозируемым значением каждой модели и фактическим значением рассчитывались в соответствии с уравнением (7).

Результаты представлены в табл. 1. Метод предварительной обработки, сочетающий подходы SG и FD, продемонстрировал значительное преимущество в сочетании с методом PCR-моделирования; при этом результаты прогнозирования большей части свойств показали меньшие относительные ошибки. Метод предварительной обработки SG-SNV продемонстрировал некоторые преимущества в отношении определенных свойств, тогда как метод предварительной обработки SG-BC оказался менее эффективным по сравнению с двумя другими методами.

**Разработка NIR-модели методом частичных наименьших квадратов.** После предварительной обработки исходных данных NIR-спектров методами SG-FD, SG-BC и SG-SNV были построены модели прогнозирования NIR с использованием метода PLS. Точность NIR-моделей проверяли с помощью как калибровочных, так и тестовых наборов. Примеры прогнозирования с использованием NIR-моделей в отношении таких параметров, как IBP, 10%-ная температура фракционирования, содержание *n*-парафинов и *изо*-парафинов, показаны на рис. 4. Метод PLS оказался приемлемым для моделирования после предварительной обработки спектра методами сглаживания Савицкого-Голея и коррекции базовой линии (BC). Эти модели продемонстрировали хорошие прогностические эф-

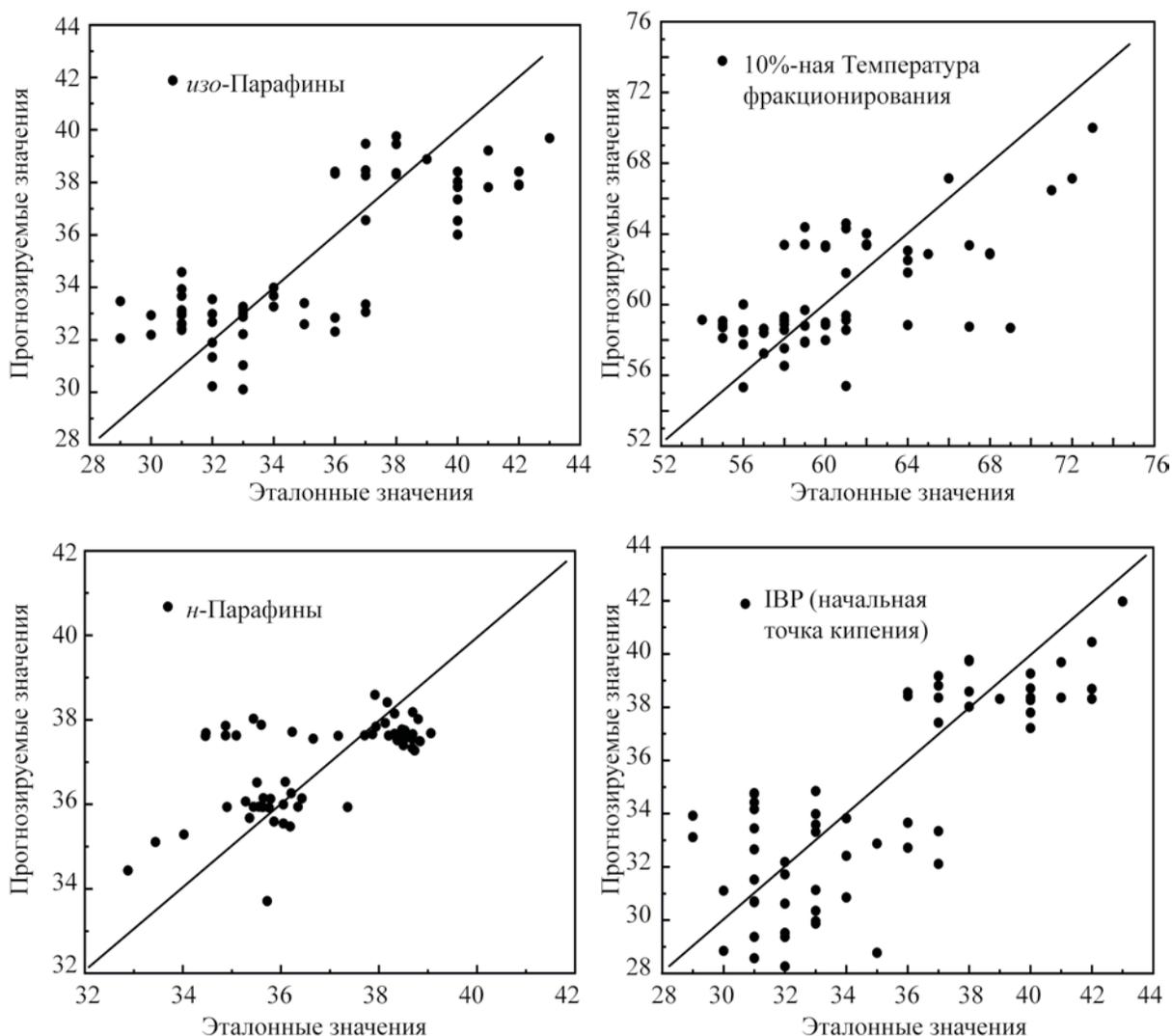


Рис. 4. Прогнозирующий эффект моделей, построенных с помощью подходов PLS и SG-BC для калибровочного набора.

факты на ста наборах калибровочных данных, что было подтверждено близостью прогнозируемых значений к эталонным (значениям, полученным в результате лабораторного анализа). Результаты изучения других методов предварительной обработки позволили сделать аналогичные выводы.

На следующем этапе исследования сравнили средние относительные ошибки прогноза 14 параметров нефти, полученных методом PLS в сочетании с методами предварительной обработки SG-FD, SG-BC или SG-SNV для тестового набора, которые приведены в табл. 2. Сочетание методов PLS и SG-FD обеспечило наилучший прогнозирующий эффект, за ним следовало PLS-SG-SNV, а

наименее эффективным оказалось сочетание методов PLS-SG-BC. В случае использования моделей PLS-SG-BC допустимые средние относительные ошибки прогнозируемых свойств нефти (особенно температуры фракционирования) для тестового набора составляли 1–5%. Возможные объяснения сравнительно высоких средних относительных погрешностей некоторых свойств, например, содержания нафтенов сводились к следующему:

- 1) неправильный выбор длины волны приводит к низкой точности разработанной модели;
- 2) данные NIR, собранные на заводе, были неточными, что привело к смешиванию необработанных спектральных данных с обильными помехами;

**Таблица 2.** Средняя относительная ошибка модели прогнозирования методом PLS в ближнем инфракрасном диапазоне, построенной с использованием различных методов предварительной обработки тестового набора параметров

Свойство		Метод		
		SG-FD	SG-BC	SG-SNV
Диапазон дистилляции	IBP (начальная температура кипения)	5.04	4.98	6.35
	10% Температура фракционирования	4.43	4.81	5.50
	30% Температура фракционирования	3.76	4.13	4.48
	50% Температура фракционирования	2.86	2.94	3.41
	70% Температура фракционирования	2.73	3.03	2.97
	90% Температура фракционирования	2.89	2.97	2.33
	95% Температура фракционирования	1.89	1.69	1.57
	IBP (конечная температура кипения)	1.13	1.20	1.09
Состав группы	Алканы	2.11	2.18	2.31
	<i>n</i> -Парафины	2.73	2.42	2.52
	<i>изо</i> -Парафины	2.17	2.02	1.90
	Ароматические углеводороды	7.43	7.62	7.74
	Нафтены	7.05	7.41	7.57

3) выбранные методы предварительной обработки или моделирования не были лучшим вариантом для этих параметров свойств, что привело к неудовлетворительным результатам.

### ЗАКЛЮЧЕНИЕ

Для удовлетворения спроса на высококачественную нефть, используемую для производства этилена, были одновременно разработаны оптимальные модели прогнозирования диапазона дистилляции и группового состава нефти на основе NIR-спектров, собранных интерактивной системой измерения спектра полного пропускания. Необработанные усредненные спектры пропускания, полученные от более чем 160 различных образцов нефти, были схожими и содержали большой объем исходной информации. Для исследования влияния интерактивного прогнозирования с использованием NIR-моделей на свойства нефти различные методы предварительной обработки (SG-FD, SG-BC, SG-SNV) были объединены с методами моделирования (PCR и PLS). Различные методы предварительной обработки оказывали несколько отличное влияние на точность модели. Метод SG-FD обеспечил наилучшие результаты прогнозирования, хотя и без существенных преимуществ по сравнению с

остальными методами. В случае использования одного и того же метода предварительной обработки относительные ошибки моделей, разработанных методом PLS, были немного ниже, чем у моделей, разработанных методом PCR, следовательно, эта комбинация может обеспечить приемлемые модели прогнозирования свойств нефти. Средние относительные погрешности прогнозируемых свойств нефти, особенно точек температуры фракционирования для тестового набора, в основном составляли 1–5%, т.е. были приемлемыми.

### ФИНАНСИРОВАНИЕ РАБОТЫ

Работа была проведена при поддержке Национальной программы ключевых исследований и разработок Китая (2022YFB3305900), Национального фонда естественных наук Китая (61973124, 62373155), Основной программы Инновационной лаборатории Цинъюань (грант № 00122002) и Программы внедрения научных талантов в университеты (проект 111) в рамках гранта B17017.

### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов, требующего раскрытия в данной статье.

## ИНФОРМАЦИЯ ОБ АВТОРАХ

Jian Long, ORCID: <https://orcid.org/0000-0002-3221-309X>

Chen Fan, ORCID: <https://orcid.org/0000-0003-2588-2516>

## СПИСОК ЛИТЕРАТУРЫ

1. Zhang S., Wang S., Xu Q. Emission constrained dynamic scheduling for ethylene cracking furnace system // *Ind. Eng. Chem. Res.* 2017. V. 56. № 5. P. 1327–1340. <https://doi.org/10.1021/acs.iecr.6b02822>
2. Gu H., Li J., Mu P., Zhu Q. Improving the operational efficiency of ethylene cracking integrated with refining by molecular management // *Ind. Eng. Chem. Res.* 2020. V. 59. № 29, P. 13160–13174. <https://doi.org/10.1021/acs.iecr.0c01705>
3. Zhao C., Liu C., Xu Q. Dynamic scheduling for ethylene cracking furnace system // *Ind. Eng. Chem. Res.* 2011. V. 50. № 21. P. 12026–12040. <https://doi.org/10.1021/ie200318p>
4. Trichard F., Forquet V., Gilon N., Lienemann C.P., Baco-Antoniaali F. Detection and quantification of sulfur in oil products by laser-induced breakdown spectroscopy for on-line analysis // *Spectrochim. Acta, Part B.* 2016. V. 118. P. 72–80. <https://doi.org/10.1016/j.sab.2016.02.001>
5. Wang K., Du W., Long J. Near-infrared wavelength-selection method based on joint mutual information and weighted bootstrap sampling // *IEEE Trans. Ind. Inf.* 2020. V. 16. № 9. P. 5884–5894. <https://doi.org/10.1109/TII.2020.2972351>
6. Santos F.D., Santos L.P., Cunha P.H.P., Borghi F.T., Romão W., de Castro E.V.R., de Oliveira E.C., Filgueiras P.R. Discrimination of oils and fuels using a portable NIR spectrometer // *Fuel.* 2021. V. 283. P. 118854. <https://doi.org/10.1016/j.fuel.2020.118854>
7. Joshi N.B., Mullins O.C., Jamaluddin A., Creek J., McFadden J. Asphaltene precipitation from live crude oil // *Energy Fuels.* 2001. V. 15. № 4. P. 979–986. <https://doi.org/10.1021/ef010047i>
8. Barreira F.R., Reis L.G., Nunes R.d.C.P., Filipakis S.D., Lucas E.F. Asphaltenes precipitation onset: Influence of the addition of a second crude oil or its asphaltene fractions (C3I and C5I) // *Energy Fuels.* 2018. V. 32. № 10. P. 10391–10397. <https://doi.org/10.1021/acs.energyfuels.8b01749>
9. Wang K., He K., Du W., Long J. Novel adaptive sample space expansion approach of NIR model for in-situ measurement of gasoline octane number in online gasoline blending processes // *Chem. Eng. Sci.* 2021. V. 242, p. 116672. <https://doi.org/10.1016/j.ces.2021.116672>
10. Long J., Jiang S. Y., Liu T. B., Wang K., He R. C., Zhao L. Modified hybrid strategy integrating online adjustable oil property characterization and data-driven robust optimization under uncertainty: Application in gasoline blending // *Energy Fuels.* 2022. V. 36. № 12. P. 6581–6596. <https://doi.org/10.1021/acs.energyfuels.2c00195>
11. Long J., Jiang S. Y., He R. C., Zhao L. Diesel blending under property uncertainty: A data-driven robust optimization approach // *Fuel.* 2021. V. 306. P. 121647. <https://doi.org/10.1016/j.fuel.2021.121647>
12. Chen M., Khare S., Huang B., Zhang H., Lau E., Feng E. Recursive wavelength-selection strategy to update near-infrared spectroscopy model with an industrial application // *Ind. Eng. Chem. Res.* 2013. V. 52. № 23. P. 7886–7895. <https://doi.org/10.1021/ie4008248>
13. Yu H., Du W., Lang Z. Q., Wang K., Long J. A Novel integrated approach to characterization of petroleum naphtha properties from near-infrared spectroscopy // *IEEE Trans. Instrum. Meas.* 2021. V. 70. P. 1–13. Art. № 2510413. <https://doi.org/10.1109/TIM.2021.3077659>
14. Wang S., Liu S., Zhang J., Che X., Wang Z., Kong D. Feasibility study on prediction of gasoline octane number using NIR spectroscopy combined with manifold learning and neural network // *Spectrochim. Acta, Part A.* 2020. V. 228. P. 117836. <https://doi.org/10.1016/j.saa.2019.117836>
15. Wang X., Yang K., Kalivas J.H. Comparison of extreme learning machine models for gasoline octane number forecasting by near-infrared spectra analysis // *Optik.* 2020. V. 200. P. 163325. <https://doi.org/10.1016/j.ijleo.2019.163325>
16. Correa Gonzalez S., Kroyan Y., Sarjovaara T., Kiiski U., Karvo A., Toldy A.I., Larmi M., Santasalo-Aarnio A. Prediction of gasoline blend ignition characteristics using machine learning models // *Energy Fuels.* 2021. V. 35. № 11. P. 9332–9340. <https://doi.org/10.1021/acs.energyfuels.1c00749>
17. He K., Zhong M., Fang J., Li Y. Biased minimax probability machine-based adaptive regression for online analysis of gasoline property // *IEEE Trans. Ind. Inf.* 2020. V. 16. № 4. P. 2799–2808. <https://doi.org/10.1109/TII.2019.2945001>
18. Li J., Chu X. Rapid determination of physical and chemical parameters of reformed gasoline by near-infrared (NIR) spectroscopy combined with the Monte Carlo virtual spectrum identification method // *Energy Fuels.* 2018. V. 32. № 12. P. 12013–12020. <https://doi.org/10.1021/acs.energyfuels.8b00854>
19. Кунцов А.Х., Арбузова Т.В. Исследование тяжелых фракций нефти методом Фурье-спектроскопии КР ближнего ИК-диапазона // *Нефтехимия.* 2011. Т. 51. № 3. С. 214–222. EDN NWCMOB

- [Kuptsov A.K., Arbuzova T.V. A study of heavy oil fractions by Fourier-transform near-infrared Raman spectroscopy // *Petrol. Chemistry*. 2011. V. 51. № 3. P. 203–211. <https://doi.org/10.1134/S0965544111020083>].
20. Kashaev R.S. Structural–dynamic phase transitions in petroleum disperse systems // *Petrol. Chemistry*. 2019. V. 59. № 2. P. 143–150. <https://doi.org/10.1134/S0965544119020087>
  21. Huang Y., Dong W., Chen Y., Wang X., Luo W., Zhan B., Liu X., Zhang H. Online detection of soluble solids content and maturity of tomatoes using Vis/NIR full transmittance spectra // *Chemometrics and Intelligent Laboratory Systems*. 2021. V. 210. P. 104243. <https://doi.org/10.1016/j.chemolab.2021.104243>
  22. Hosani E.A., Meribout M., Al-Durra A., Al-Wahedi K., Teniou S. A new optical-based device for online black powder detection in gas pipelines // *IEEE Trans. Instrum. Meas.* 2014. V. 63. № 9. P. 2238–2252. <https://doi.org/10.1109/TIM.2014.2308985>
  23. Haghi R.K., Yang J., Tohidi B. Fourier transform near-infrared (FTNIR) spectroscopy and partial least-squares (PLS) algorithm for monitoring compositional changes in hydrocarbon gases under *in situ* pressure // *Energy Fuels*. 2017. V. 31. № 9. P. 10245–10259. <https://doi.org/10.1021/acs.energyfuels.7b01677>