

Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа теоретической механики и математической физики

Работа допущена к защите
Директор ВШТМиМФ,
д.ф.-м.н., член-корр. РАН
_____ А.М. Кривцов
« _____ » _____ 2023 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
магистерская диссертация
ИСПОЛЬЗОВАНИЕ ГИБРИДНОГО ПОДХОДА ДЛЯ РЕШЕНИЯ ЗАДАЧИ
ПРОГНОЗИРОВАНИЯ ДЕБИТА НА СКВАЖИНАХ, ОБОРУДОВАННЫХ
УЭЦН
по направлению подготовки
01.04.03 Механика и математическое моделирование
Направленность (профиль)
01.04.03_04 Математическое моделирование процессов нефтегазодобычи

Выполнил
студент гр. 5040103/10401

Г.А. Мосягин

Руководитель
профессор ВШТМиМФ, д.ф.-м.н.,

В.М. Иванов

Консультант
по нормоконтролю

Е.А. Хайбулова

Санкт-Петербург
2023

**САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО**

Физико-механический институт

Высшая школа теоретической механики и математической физики

УТВЕРЖДАЮ

Директор ВШТМиМФ

_____ А.М. Кривцов

« _____ » _____ 2023г.

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

студенту Мосягину Григорию Александровичу гр. 5040103/10401

1. Тема работы: Использование гибридного подхода для решения задачи прогнозирования дебита на скважинах, оборудованных УЭЦН.
2. Срок сдачи студентом законченной работы: 26.05.2023.
3. Исходные данные по работе: справочная литература, актуальные публикации по теме исследования, данные о работе скважин, типе погружного оборудования, конструкции скважин, физико-химические свойства флюидов
4. Содержание работы (перечень подлежащих разработке вопросов): обзор литературы по теме ВКР, создание модели машинного обучения для прогнозирования калибровочных коэффициентов, создание гибридного алгоритма восстановления динамики работы скважины, проверка гибридного алгоритма на реальных данных и сравнение с физической моделью
5. Перечень графического материала (с указанием обязательных чертежей): не предусмотрены
6. Дата выдачи задания: 27.02.2023.

Руководитель ВКР _____ В.М. Иванов, профессор ВШТМиМФ, д.ф.-м.н.

Задание принял к исполнению 27.02.2023

Студент _____ Г.А. Мосягин

РЕФЕРАТ

На 61 с., 35 рисунков, 6 таблиц, 0 приложений

ВИРТУАЛЬНАЯ РАСХОДОМЕТРИЯ, VFM, МАШИННОЕ ОБУЧЕНИЕ, ГРАДИЕНТНЫЙ БУСТИНГ, ГИБРИДНАЯ МОДЕЛЬ, ПРОГНОЗИРОВАНИЕ ДЕБИТА, ЗАДАЧА РЕГРЕССИИ.

В данной работе разработан гибридный подход к решению задачи виртуальной расходомерии для скважин, оборудованных электроприводным центробежным насосом (ЭЦН). Разработанная гибридная модель сочетает в себе использование физических принципов моделирования работы УЭЦН и использование методов машинного обучения. Проанализировано поведение предложенного алгоритма при различных эксплуатационных условиях. Проведено сравнение работы алгоритма с физической моделью виртуального расходомера. Разработан программный код на языке Python.

ABSTRACT

61 pages, 35 figures, 6 tables, 0 appendices

VIRTUAL FLOW MEASUREMENT, VFM, MACHINE LEARNING, GRADIENT BOOSTING, HYBRID MODEL, FLOW RATE FORECAST, REGRESSION PROBLEM.

This paper develops a hybrid approach to solving the virtual flow-metering problem for wells equipped with an electrically driven centrifugal pump (ESP). The developed hybrid model combines the use of physical principles of ESP operation modeling and the use of machine learning methods. The behavior of the proposed algorithm under different operating conditions is analyzed. The performance of the algorithm is compared with the physical model of the virtual flow meter. The program code in Python language has been developed.

СОДЕРЖАНИЕ

Введение	5
Глава 1. ОБЗОР ПОДХОДОВ К РЕШЕНИЮ ЗАДАЧИ ВИРТУАЛЬНОЙ РАСХОДОМЕТРИИ	7
Глава 2. ОБЩАЯ ТЕОРИЯ ПОСТРОЕНИЯ АЛГОРИТМОВ ВИРТУАЛЬНОГО РАСХОДОМЕРА ДЛЯ СКВАЖИН, ОБОРУДОВАННЫХ УЭЦН	14
2.1. Физическая модель	14
2.1.1. Модель PVT	14
2.1.2. Модель течения жидкости в НКТ	15
2.1.3. Модель УЭЦН	20
2.2. Статистическая модель	25
2.2.1. Задачи машинного обучения	26
2.2.2. Модели машинного обучения	27
2.2.3. Обучение статистических моделей	29
2.2.4. Методы оценки модели машинного обучения	31
Глава 3. РЕАЛИЗАЦИЯ АЛГОРИТМА ГИБРИДНОЙ МОДЕЛИ ВИРТУАЛЬНОЙ РАСХОДОМЕТРИИ	33
3.1. Схема работы алгоритма	33
3.2. Физическая модель	35
3.2.1. Обработка данных	36
3.2.2. Адаптация модели	37
3.2.3. Прогнозирование калибровочных коэффициентов	39
3.2.4. Восстановление дебита	41
3.3. Статистическая модель	42
3.3.1. Подготовка данных	44
3.3.2. Обучение модели	47
Глава 4. Результаты применения алгоритма	50
4.1. Смена способа эксплуатации	50
4.2. Ввод новой скважины	52
4.3. Смена ЭЦН	52
4.4. Поломка АГЗУ	53
4.5. Сравнение гибридной модели с физической моделью	54
4.6. Выводы	57
Заключение	58
Список использованных источников	60

ВВЕДЕНИЕ

Сегодня около 70% процентов добычи нефти в Российской Федерации осуществляется с помощью погружных электроцентробежных насосов (ЭЦН). Доля данного способа добычи нефти высока, именно поэтому постоянный мониторинг работы электроцентробежных насосов крайне необходим для обеспечения оптимальных условий эксплуатации и избежания задержек в добычи нефти.

Одним из самых важных показателей работы скважины является дебит нефти. Информация о дебите нефти позволяет операторам принимать важные решения по оптимизации и управлению добычи, прогнозировать будущую производительность месторождения. Поэтому точное измерение дебита является чрезвычайно ценным в производственной деятельности.

В Российской Федерации для возможности контроля дебита скважины нефтегазовые компании обычно проводят испытания скважин с помощью автоматических групповых замерных установок (АГЗУ), которые по очереди переключаются между кустовыми скважинами для определения их текущего дебита. У замера скважины с помощью АГЗУ существует специфика – точность замера возрастает с увеличением времени замера [9]. Данная специфика является особенно актуальной на малодобитных скважинах, работающих в периодическом (ПКВ) режиме. Такой способ измерения скважин накладывает ограничение на количество испытаний на скважину, и, следовательно, дебиты скважин не будут постоянно доступны. Кроме того, на результаты испытаний влияют изменения условий работы скважины. Это может привести к недостаточному пониманию дебита для скважин, в которых добыча быстро меняется.

Одним из альтернативных способов измерения дебита является создание математической модели скважины. Данный подход известен как задача виртуальной расходомерии. Виртуальный расходомер позволяет по регистрируемым данным во время эксплуатации скважины определять ее добычу с более высокой частотой по сравнению АГЗУ. Применение виртуального расходомера особо актуально для месторождений с большим фондом скважин, так как он способен оценивать дебит в режиме реального времени, что позволяет своевременно получать информацию, выявлять аномалии в работе скважин и значительно снижать нагрузку на операторов добычи. Виртуальный расходомер не требует установки дополнительного оборудования, поэтому он может снизить капитальные и эксплуатационные затраты на разработку месторождения [10].

Целью данной работы является рассмотрение и сравнение возможных подходов к решению задачи виртуальной расходомерии, а также создание и проверка возможности использования объединенной физической и статистической (гибридной) модели на реальных промышленных данных.

ГЛАВА 1. ОБЗОР ПОДХОДОВ К РЕШЕНИЮ ЗАДАЧИ ВИРТУАЛЬНОЙ РАСХОДОМЕТРИИ

Для оптимального управления нефтяным активом и максимального извлечения углеводородов необходимо понимание поведения системы добычи. На сегодняшний день большинство скважин оборудовано специальными датчиками телеметрии, с помощью данных из которых специалисты сектора оперативного управления в режиме реального времени могут контролировать и анализировать работу скважин. Анализ таких данных позволяет быстро найти отклонения в работе скважин, однако напрямую не способен сказать о потерях в добыче углеводородов по скважине.

Важнейшим параметром для контроля работы фонда, его оптимизации, определения состояния разработки месторождения и прогнозирования будущей производительности является дебит нефти. На данный момент для измерения дебита часто используются автоматические групповые замерные установки (АГЗУ), флюид к которым должен подводиться по отдельному трубопроводу. Как правило, дебит скважины измеряется путем периодических испытаний (например, ежемесячно или еженедельно). Такой подход позволяет понять режим работы скважины во время замера, однако не может сказать, как работала скважина между замерами. Использование АГЗУ может накладывать свои ограничения на измерения дебита механизированного фонда [14]:

- увеличенное время замера при испытании ПКВ скважин и скважин из малодебитного фонда;
- невозможность применения АГЗУ из-за высокого линейного давления;
- невозможность подключения скважины к АГЗУ из-за проблем с инфраструктурой;
- увеличение погрешности замера из-за поломки или неисправности, наличия засорений и отложений механизированных примесей, деградации с течением времени замерной установки;
- наличие нескольких скважин, работающих на один трубопровод, подключенный к АГЗУ.

Использование передвижных замерных установок, способных измерять дебит скважины в режиме реального времени, позволяет решить часть описанных проблем. Однако такое решение требует высоких затрат на проведение каждого замера.

Кроме того, такие установки могут иметь определенный рабочий диапазон, за пределами которого точность определения дебита может значительно снижаться.

В качестве альтернативы дебит скважины может быть измерен косвенно с помощью виртуального расходомера. Цель виртуальной расходомерии – без прямого измерения дебита с помощью различных математических моделей по косвенным параметрам оценить дебит скважины. В данном случае, под косвенными параметрами понимаются параметры, напрямую не говорящие о том, с каким дебитом работает скважина (например, давление и температура). За последнее время вышло немало работ по теме виртуальной расходомерии, предложено множество подходов к решению данной задачи, рассмотрим кратко некоторые из этих работ.

Авторы [10] подготовили обзор и анализ существующих подходов к решению задачи виртуальной расходомерии. В работе выделено два основных подхода к решению задачи:

- создание виртуального расходомера на базе физической модели (First principles VFM);
- создание виртуального расходомера на базе статистической модели (Data-driven VFM).

Системы «First principles VFM» основаны на аналитическом моделировании процессов подъема нефти и газа на поверхность. Моделирование систем с помощью «First principles VFM» является надежным способом описания поведения производственных систем, в том числе виртуальные расходомеры на базе физической модели являются наиболее широко распространенными в отрасли. В целом, данный алгоритм включает в себя следующие основные компоненты: модель флюида, модель добывающей системы (модель притока, гидродинамическая модель лифта, модель штуцера и модель ЭЦН) и алгоритмы оптимизации, используемые для поиска решений системы (определения согласованного значения расхода) [9].

В свою очередь, системы «Data-driven VFM» основаны на поиске взаимосвязи между входными (данными телеметрии) и выходными (дебитом скважины) параметрами с помощью моделей машинного обучения. Основным преимуществом данного метода является то, что он позволяет отказаться от детального физического моделирования систем, для которых поиск точного решения может являться трудновыполнимой задачей. В целом, данный алгоритм включает в себя следующие основные стадии: обработка и фильтрация данных, выбор и разработка алгоритма машинного обучения, а также использование обученной модели

для прогнозирования дебита на новых данных. Подробнее данные стадии будут рассмотрены во 2 главе.

Авторы [10] приходят к выводу, что, несмотря на большое количество работ и исследования по теме виртуальной расходомерии, все еще существует множество способов улучшить и сделать систему виртуального расходомера более надежной. В работе отмечается, что наиболее перспективной системой виртуальной расходомерии является гибридная модель, которая сочетает в себе подходы физического и статистического моделирования.

В работах [11; 12] рассматривается создание гибридной модели штуцера, а также различные способы построения гибридной модели в целом. Авторы рассматривают физическую модель как модель «белого ящика», а статистическую модель как модель «черного ящика», оценивая модели по степени предварительных знаний о системе (по физической интерпретации параметров или взаимосвязи между переменными, рис.1.1). Авторы определяют гибридную модель, как модель «серого ящика». Цель создания гибридной модели состоит в том, чтобы получить модель, обладающую высокой гибкостью, интерпретируемостью и высокой научной обоснованностью.



Рис.1.1. Диапазон моделей виртуального расходомера от «белого ящика» до «черного ящика» и некоторые их характеристики [12]

В работе выделяется два способа построения гибридной модели рис.1.2:

- последовательное соединение моделей;
- параллельное соединение моделей.

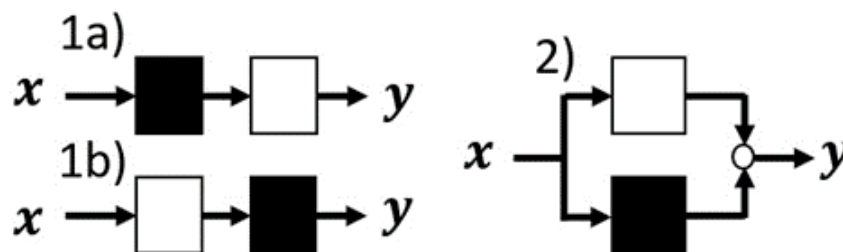


Рис.1.2. Схемы построения гибридной модели [11]

К последовательному соединению моделей можно отнести случаи, при которых статистическая модель прогнозирует физические параметры (например обводненность или оценку плотности газа), которые идут на вход физической модели (1a). Также возможно построение гибридной модели, когда физическая модель создает параметры, которые идут на вход в статистическую модель, такой прием известен как *feature engineering* (1b). К параллельному соединению моделей можно отнести ситуации, при которых для каждой скважины строится отдельно физическая и статистическая модель, а итоговый дебит получается выбором лучшей модели или путем объединения результатов моделей.

В работе [11] рассматривается физическая и статистическая модель, а также несколько вариаций гибридной модели виртуального расходомера для 10 скважин Норвежского континентального шельфа. Авторы приходят к выводу, что смещение модели от «белого ящика» к «серому ящику» повышает качество модели. В целом, для большинства скважин гибридная модель показывает себя лучше физической и статистической модели. Стоит отметить, что для одной скважины авторы проводили исследование реагирования моделей виртуального расходомера на изменение режима работы скважины для проверки экстраполяционной способности алгоритмов. От моделей ожидалось, что при увеличении диаметра штуцера они предскажут увеличение добываемого дебита. Все модели, кроме статистической, корректно отреагировали на изменение параметров. Авторы связывают данный факт с тем, что статистическая модель требует большое количество данных для обучения, которые должны охватывать различные режимы работы скважины. Для некоторых гибридных моделей и статистической модели на определенных скважинах наблюдались не естественные значения дебита. Авторы объясняют данное наблюдение возможными ошибочными и недостаточно обработанными входными данными, а также недостаточным размером обучающей выборки. В работе отмечается, что в будущих исследованиях стоит обратить особое внимание на формирование обучающей выборки, фильтрацию данных и возможность

создания общей статистической модели, обучение которой должно происходить на данных, содержащих большое количество скважин. Таким образом, авторы предлагают вместо построения отдельной статистической модели для каждой скважины, сформировать общую статистическую модель, способную прогнозировать значения целевой переменной для множества скважин. Также отмечается, что, независимо от типа модели «серого ящика», необходимо соблюдать баланс между физической и статистической частью гибридной модели. Например, если процесс демонстрирует неизвестное для модели поведение, то наибольшее влияние должна иметь статистическая часть модели.

В работе [9] авторы проводят обзор существующих подходов к решению задач виртуальной расходомерии, а также проводят обзор и анализ замерных установок. Авторы приходят к выводу, что виртуальная расходомерия является востребованным подходом как для измерения дебита скважин, так и для выявления осложнений в работе фонда.

Авторы [7] рассматривают алгоритм виртуальной расходомерии для скважин, оборудованных УЭЦН, основанный на совмещении модели УЭЦН, модели РVT свойств и модели течения многофазного флюида в скважине. Для реализации физической модели авторы использовали энергетическую схему построения модели, основывающуюся на потребляемой мощности насоса. Для сопоставления фактического и модельного режима работы скважины авторы использовали коэффициент деградации по мощностной характеристике, который интерполировался между фактическими замерами дебита. Предложенный алгоритм апробирован на ряде скважин компании ПАО «Газпром нефть» и хорошо зарекомендовал себя на постоянно действующем фонде скважин. Применение данного алгоритма позволяет оперативно выявлять нестабильно работающие скважины и планировать на них мероприятия по управлению их работой [9].

Автор в работе [8] рассматривает построение гибридной модели для фонтанирующих скважин на основе расчёта движения потока через штуцер. Для реализации физической модели автор использовал подход Перкинса, для определения калибровочных коэффициентов между замерами дебита автор использовал машинное обучение. Описанный алгоритм был протестирован на реальных промышленных данных и хорошо зарекомендовал себя для ряда скважин. Автор отметил, что в дальнейших исследованиях необходимо уделить особое внимание обработке и фильтрации данных, а также рассмотреть новые подходы к прогнозированию калибровочных коэффициентов.

В работе [13] рассматривается параллельное соединение физической и статистической модели для построения гибридной модели на примере 2 нефтяных скважин. Для каждой скважины строилась физическая и статистическая модель, после чего прогнозы дебита с каждой модели направлялись в оценщик, который по исторической сходимости и степени доверия к расчету объединял оценки каждой модели для обеспечения большей точности модели. В результате 6-месячного тестирования авторы отметили, что применение гибридной модели привело к 50% улучшению точности прогноза по сравнению с отдельными моделями. Авторы отметили, что системы виртуального расходомера могут быть использованы для мониторинга работы скважины в режиме реального времени, а также в качестве резервного источника информации в случае поломки замерной установки.

В работах [10—12] был проведен подробный анализ и сравнение двух подходов к решению задачи виртуальной расходомерии. Сравнение подходов представлено в табл.1.1.

После проведенного анализа литературы можно сделать вывод о том, что внедрение виртуального расходомера позволит получать данные о работе фонда в режиме реального времени, что открывает широкий спектр возможностей для оперативного реагирования на отклонения при работе фонда. В отличие от замерных установок система виртуального расходомера не требует установки дополнительного оборудования, что может снизить капитальные и эксплуатационные затраты на разработку месторождения [10].

Система виртуального расходомера может использоваться как самостоятельное решение или в комбинации с замерными установками в качестве резервной системы, которая может описывать работу скважины между замерами, а также валидировать значения, полученные от замерных установок [10]. Стоит отметить, что при хорошей исторической сходимости виртуального расходомера на конкретной скважине, возможен вариант постепенного отказа от использования физического расходомера, лишь иногда используя его для калибровки модели. Такое применение виртуального расходомера позволит снизить нагрузку на замерные установки, что позволит чаще замерять наиболее проблемные скважины. Необходимо отметить, что, потенциально, виртуальный расходомер может позволить проводить обратное распределение добычи с конечного узла сбора.

Таблица 1.1

Сравнение подходов к решению задачи виртуальной расходомерии

Модель виртуального расходомера	Достоинства	Недостатки
Модель виртуального расходомера в общем	<ul style="list-style-type: none"> – Мониторинг работы скважины в режиме реального времени – Не требует больших капитальных и эксплуатационных затрат – Не требует физического вмешательства в случае наличия проблем – Возможность интегрирования с другим программным обеспечением для оптимизации добычи 	<ul style="list-style-type: none"> – Зависит от точности датчиков – Требует периодической настройки и перекалибровки модели
Физическая модель	<ul style="list-style-type: none"> – Используются проверенные и известные методы моделирования – Хорошо подходит для моделирования устойчивых состояний – На рынке доступно множество коммерческих решений – Имеет хорошую интерпретируемость и предсказательную способность 	<ul style="list-style-type: none"> – Требует глубоких знаний о физике рассматриваемого процесса – Сильно зависит от точности данных PVT – Высокие вычислительные затраты – Процесс настройки модели не является простым
Статистическая модель	<ul style="list-style-type: none"> – Не требует глубоких знаний о физике рассматриваемого процесса – Обученная модель требует низкие вычислительные затраты – Легко объединить различные параметры системы без построения сложной физической модели – Модель легко обновляется с учетом новых полученных данных – Модель способна адаптироваться к неизвестным или не моделируемым явлениям (имеет высокую гибкость) 	<ul style="list-style-type: none"> – Требует наличие большого количества исторических данных для обучения – Имеет малую экстраполяционную способность на данных, которых не было в обучающей выборке – Требовательные к качеству данных и их обработке – Имеет малую интерпретируемость и научную обоснованность

В целом, можно сделать вывод о том, что создание и внедрение гибридной модели виртуального расходомера является актуальной задачей и может позволить значительно увеличить точность и доверие к системам виртуальной расходомерии.

ГЛАВА 2. ОБЩАЯ ТЕОРИЯ ПОСТРОЕНИЯ АЛГОРИТМОВ ВИРТУАЛЬНОГО РАСХОДОМЕРА ДЛЯ СКВАЖИН, ОБОРУДОВАННЫХ УЭЦН

В главе рассматривается общая концепция построения физической и статистической модели виртуального расходомера. В параграфе 2.1 рассматривается физическая модель виртуального расходомера, необходимые компоненты для построения модели. Параграф 2.2 затрагивает этапы построения статистической модели виртуального расходомера, а именно типы задач машинного обучения, методы разбиения данных для обучения, методы поиска оптимальной комбинации гиперпараметров и метрики оценки качества моделей.

2.1. Физическая модель

Физические модели или модели «белого ящика» основаны на предварительных знаниях о процессе, для описания взаимосвязи между входными и выходными параметрами используют физические законы. Преимуществом данных моделей является интерпретируемость результатов, хорошая экстраполяционная способность и научная согласованность. Именно поэтому физическая модель является наиболее распространенным подходом к решению задачи виртуальной расходомерии [10].

Стоит отметить, что рассматриваемая модель «белого ящика» не будет учитывать модель пласта и модель штуцера. Таким образом, для определения физической модели необходимо объединить в систему модель PVT, модель течения жидкости в НКТ (насосно-компрессорных трубах), а также модель ЭЦН. Далее подобрать параметры для сопоставления модели с реальным режимом работы скважины, после чего необходимо решить задачу узлового анализа для определения дебита жидкости. Реализация физической модели будет рассмотрена в главе 3.

2.1.1. Модель PVT

Углеводородные смеси представляют собой вещества, свойства которых меняются в зависимости от условий давления и температуры в системе добычи. Свойства флюида оказывают большое влияние на прогнозы системы виртуального расходомера [10]. Неправильное указание плотности фаз, например, может приве-

сти к отклонению в прогнозе виртуального расходомера, что вызовет проблемы в поиске решения и адаптации модели.

Обычно модель PVT можно определить с помощью двух подходов:

- модель «черной нефти» (Black Oil);
- композиционная модель.

Модель «Black Oil» является упрощенной моделью, которая предполагает, что флюид представляет собой смесь трех углеводородных компонентов: нефти, газа и воды, свойства которых рассчитываются с помощью корреляций. В данной модели предполагается, что вода и нефть не смешиваются между собой и не обмениваются массами, газ растворим в нефти и нерастворим в воде. В модели учитываются объемные коэффициенты воды и нефти, газосодержание, а также коэффициенты, отвечающие за соотношение нефть-газ-вода в смеси. Композиционная модель более точно описывает сложное фазовое состояние смеси и процессы массообмена, а также позволяет проводить оценку количества компонентов, состав каждой фазы и учитывать более сложные процессы, такие как фазовые переходы. Согласно [10], современные системы виртуального расходомера поддерживают композиционный подход и модель «черной нефти», однако упрощенные модели виртуального расходомера, как правило, используют модель «черной нефти» из-за ее простоты. В данной работе выбрана модель «черной нефти», так как она требует относительно небольших вычислительных ресурсов.

2.1.2. Модель течения жидкости в НКТ

При движении флюида в скважине его локальные свойства изменяются, в том числе давление и температура, что влияет на свойства флюида. После настройки PVT модели, с помощью выбранных корреляций система необходимо вычислить плотность и вязкость фаз по локальным температурам и давлениям.

Для установившегося течения однофазного потока существует уравнение градиента давления, которое было выведено из закона сохранения импульса и закона сохранения масс [1]:

$$\begin{aligned}
\left(\frac{dp}{dL}\right)_{\text{общ}} &= \left(\frac{dp}{dL}\right)_{\text{трения}} + \left(\frac{dp}{dL}\right)_{\text{гравит}} + \left(\frac{dp}{dL}\right)_{\text{уск.}} \\
\left(\frac{dp}{dL}\right)_{\text{трения}} &= \frac{f\rho\vartheta^2}{2d} \\
\left(\frac{dp}{dL}\right)_{\text{гравит}} &= \rho g \sin(\theta) \\
\left(\frac{dp}{dL}\right)_{\text{уск.}} &= \rho\vartheta \frac{d\vartheta}{dL}
\end{aligned} \tag{2.1}$$

Однако, задача расчета характеристик потока, состоящего из двух и более фаз намного сложнее, чем для однофазного потока. Поэтому задача прогнозирования характеристик многофазного потока сопряжена с применением эмпирических зависимостей [1]. Вследствие разности плотностей и вязкостей, поведение фаз в потоке может отличаться, газ и жидкость могут перемещаться по трубе с разными скоростями, может наблюдаться явление проскальзывания. В одной скважине может наблюдаться несколько режимов потока, поскольку в различных точках скважины значения давления и температуры могут быть не одинаковыми. Особенно сильно от режима потока зависит изменение градиента давления. Поэтому важно правильно предсказывать режим потока для корректного моделирования движения флюида через трубу.

В данной работе для описания течения потока в НКТ была выбрана эмпирическая корреляция Беггза и Брилла [1]. Основными преимуществами данного метода являются учет эффекта проскальзывания, режима трения и угла наклона скважины. Градиент давления в наклонной трубе Беггз и Брилл предложили рассчитывать по формуле:

$$\begin{aligned}
\frac{dp}{dL} &= \frac{\frac{f\rho_n\vartheta_m^2}{2d} + \rho_s g \sin \theta}{1 - E_k} \\
\rho_s &= \rho_L H_{L(\theta)} + \rho_g (1 - H_{L(\theta)}) \\
E_k &= \frac{\vartheta_m \vartheta_{S_g} \rho_n}{p}
\end{aligned} \tag{2.2}$$

где f – коэффициент трения; ρ_n – плотность смеси без учета проскальзывания; ρ_s – плотность смеси с учетом проскальзывания; $H_{L(\theta)}$ – объемное содержание жидкости с учетом проскальзывания; ρ_L – плотность жидкости; ρ_g – плотность газа; ϑ_m – скорость смеси; θ – угол наклона трубы от горизонтали; d – диаметр

трубы; E_k – безразмерная кинетическая энергия; ϑ_{S_g} – скорость газа, приведенная на единицу сечения трубы.

Первым этапом необходимо определить режим потока смеси. Как говорилось выше, чрезвычайно важно определить режим потока в каждой точке скважины для корректного определения градиента давления. Большинство исследователей выделяют четыре режима восходящего многофазного потока газа и жидкости [1]:

- пузырьковый режим, который характеризуется равномерным распределением газовой фазы в виде отдельных пузырьков в непрерывной жидкой фазе;
- пробковый режим, в котором присутствует несколько блоков пробки, каждый из которых состоит из газового пузыря, пробки жидкости (снаряда) и пленки жидкости вокруг газового пузыря;
- эмульсионный режим, который характеризуется хаотичным движением газа и жидкости. Для данного режима характерны колебательные движения или движения с переменным направлением;
- кольцевой режим, в котором газовая фаза в центральном ядре является непрерывной, а фаза жидкости, представляемая тонкой пленкой на стенках трубы и рассеянными в газовом ядре каплями жидкости, движется в восходящем направлении.

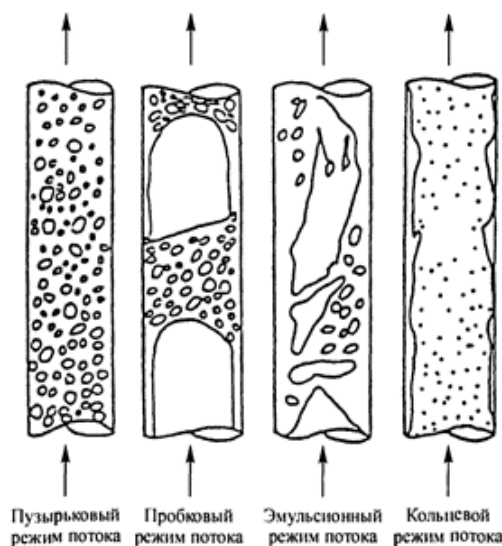


Рис.2.1. Режимы потока течения многофазной смеси [1]

Для расчета границ переходных режимов потока и объемного содержания жидкости Беггз и Брилл использовали число Фруда для смеси:

$$N_{Fr} = \frac{\vartheta_m^2}{gd} \quad (2.3)$$

Переходные границы режимов потока вычисляются по следующим формулам:

$$\begin{aligned} L_1 &= 316\lambda_L^{0.302} \\ L_2 &= 0.000925\lambda_L^{-2.468} \\ L_3 &= 0.1\lambda_L^{-1.452} \\ L_4 &= 0.5\lambda_L^{-6.738} \end{aligned} \quad (2.4)$$

где $\lambda_L = \frac{q_L}{q_L + q_g}$ – объемное содержание жидкости через дебиты по фазам в пластовых условиях.

Далее по предложенным неравенствам определяется режим потока в горизонтальной трубе:

– Раздельный режим:

$$\begin{aligned} \lambda_L < 0.01 \text{ и } N_{Fr} < L_1 \\ \text{или} \\ \lambda_L \geq 0.01 \text{ и } N_{Fr} < L_2 \end{aligned} \quad (2.5)$$

– Переходный режим:

$$\lambda_L \geq 0.01 \text{ и } L_2 \leq N_{Fr} < L_3 \quad (2.6)$$

– Прерывистый режим:

$$\begin{aligned} 0.01 \leq \lambda_L < 0.4 \text{ и } L_3 < N_{Fr} \leq L_1 \\ \text{или} \\ \lambda_L \geq 0.4 \text{ и } L_3 < N_{Fr} \leq L_4 \end{aligned} \quad (2.7)$$

– Распределенный режим:

$$\begin{aligned} \lambda_L < 0.4 \text{ и } N_{Fr} \geq L_1 \\ \text{или} \\ \lambda_L \geq 0.4 \text{ и } N_{Fr} > L_4 \end{aligned} \quad (2.8)$$

Следующим этапом необходимо спрогнозировать объемное содержание жидкости. Объемное содержание жидкости для всех режимов потока рассчитывается по одним уравнениям, однако для каждого режима необходимо использовать свои эмпирические коэффициенты табл.2.1.

Таблица 2.1

Эмпирические коэффициенты для объемного содержания жидкости в горизонтальных режимах потока по Беггзу и Бриллиу

Режим потока	a	b	c
Расслоенный	0.980	0.4846	0.0868
Прерывистый	0.855	0.5351	0.0173
Распределенный	1.065	0.8524	0.0609

Сначала рассчитывается объемное содержание жидкости в горизонтальной трубе при ограничении $H_{L(0)} \geq \lambda_L$:

$$H_{L(0)} = \frac{a\lambda_L^b}{N_{Fr}^c} \quad (2.9)$$

Объемное содержание жидкости с поправкой на угол наклона определяется следующим образом:

$$H_{L(\theta)} = H_{L(0)} \Psi$$

$$\Psi = 1 + C [\sin(1.8\theta) - 0.333 \sin^3(1.8\theta)] \quad (2.10)$$

$$C = (1 - \lambda_L) \ln(e\lambda_L^f N_{L\theta}^g N_{Fr}^h)$$

где θ – фактический угол наклона трубы; Ψ – поправочный коэффициент на угол наклона трубы; коэффициенты e, f, g, h определяются по табл.2.2.

Таблица 2.2

Эмпирические коэффициенты для значения C по Беггзу и Бриллиу

Режим потока	e	f	g	h
Расслоенный в напр. снизу вверх	0.011	-3.7680	3.5390	-1.6140
Прерывистый в напр. снизу вверх	2.960	0.3050	-0.4473	0.0978
Распределенный в напр. снизу вверх	Поправки нет, $C = 0, \Psi = 1$			
Все режимы в напр. сверху вниз	6.700	-0.3692	0.1244	-0.5056

Если предполагаемый режим потока соответствует переходной зоне, то необходимо интерполировать значения объемного содержания жидкости для расслоенного и прерывистого режимов потока:

$$H_{L(\theta)_{перех.}} = AH_{L(\theta)_{раздел.}} + (1 - A)H_{L(\theta)_{прерывист.}}$$

$$A = \frac{L_3 - N_{Fr}}{L_3 - L_2} \quad (2.11)$$

Следующим этапом необходимо спрогнозировать коэффициент трения, который предлагается вычислять по формуле

$$\frac{f}{f_n} = e^s$$

$$s = \frac{\ln y}{-0.0523 + 3.182 \ln^2 y + 0.01853 \ln^4 y} \quad (2.12)$$

$$y = \frac{\lambda_L}{H_{L(\theta)}^2}$$

где f – коэффициент трения; f_n – нормирующий коэффициент трения, который определяется по диаграмме Му迪 или из соотношений:

$$f_n = \frac{64}{N_{Re}} \text{ для ламинарного потока } (N_{Re} < 3000)$$

$$f_n = 0.0056 + 0.5 N_{Re}^{0.32} \text{ для турбулентного } (3000 < N_{Re} < 3 * 10^6) \quad (2.13)$$

$$N_{Re} = \frac{\rho_n \vartheta_m d}{\mu_n}$$

В зависимости от значения y функция s может иметь разрывы, поэтому для значения $y = 1$ необходимо брать $s = 0$ чтобы свести корреляцию к масштабу однофазного потока. При условии, когда $1 < y < 1.2$ Беггз и Брилл предложили следующее соотношение для s :

$$s = \ln(2.2y - 1.2) \quad (2.14)$$

Таким образом, для определения давления в каждой точке скважины необходимо последовательно проделать следующие шаги:

1. Определить режим потока смеси;
2. Спрогнозировать объемное содержание жидкости;
3. Спрогнозировать коэффициент трения;
4. Подставить все значения в формулу (2.4) и вычислить значение давления в заданной точке.

2.1.3. Модель УЭЦН

В России установки электроцентробежных насосов (УЭЦН) являются самым распространенным способом добычи нефти. При этом, на протяжении последних лет в Российской Федерации доля скважин с УЭЦН только растет [9]. УЭЦН состоит из рис.2.2:

- погружной части (станция управления, трансформатор, устьевое оборудование);

- наземной части (многосекционный центробежный насос, погружной электродвигатель с гидрозащитой, кабельная линия).

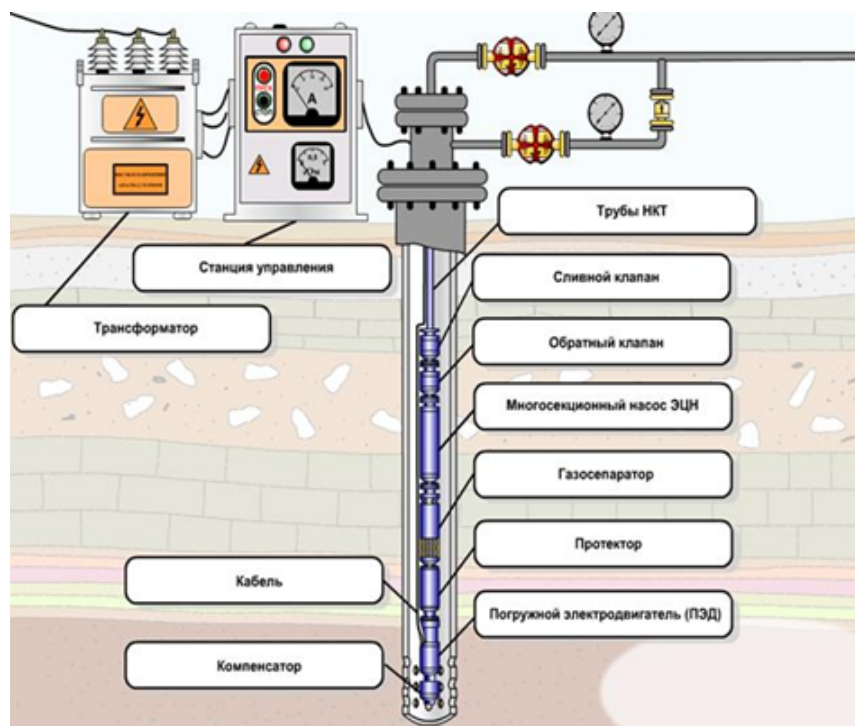


Рис.2.2. Схема УЭЦН

Погружной центробежный насос состоит из нескольких секций, которые, в свою очередь, состоят из множества ступеней. Жидкость, которая поступает в насос через приемный модуль, направляется в ступени насоса, где ее кинетическая энергия увеличивается благодаря крутящему моменту, передаваемому погружным электродвигателем. Под действием центробежных сил жидкость последовательно проходит через каждую ступень. На выходе из каждой ступени жидкость имеет большее давление, чем на входе ступени. Таким образом, последовательно после каждой ступени флюид накапливает необходимый перепад давлений необходимый для его поднятия на поверхность.

Центробежный насос преобразует механическую энергию (поступающую от двигателя, приводящего его в движение) в кинетическую энергию перемещаемой жидкости [6]. При постоянной скорости нагнетания в зависимости от плотности перекачиваемой жидкости насос передает флюиду разное количество энергии. После прохождения ступени флюиды имеют различный перепад давления в зависимости от их плотности, поэтому при работе с центробежными насосами принято использовать напор вместо давления, так как напор не зависит от плотности жидкости [6].

$$H = \frac{P}{\rho g} \quad (2.15)$$

где p – давление флюида; ρ – плотность флюида.

Основной характеристикой центробежного насоса является его напорно-расходная характеристика (зависимость КПД, напора, потребляемой мощности от подачи центробежного насоса) [2].

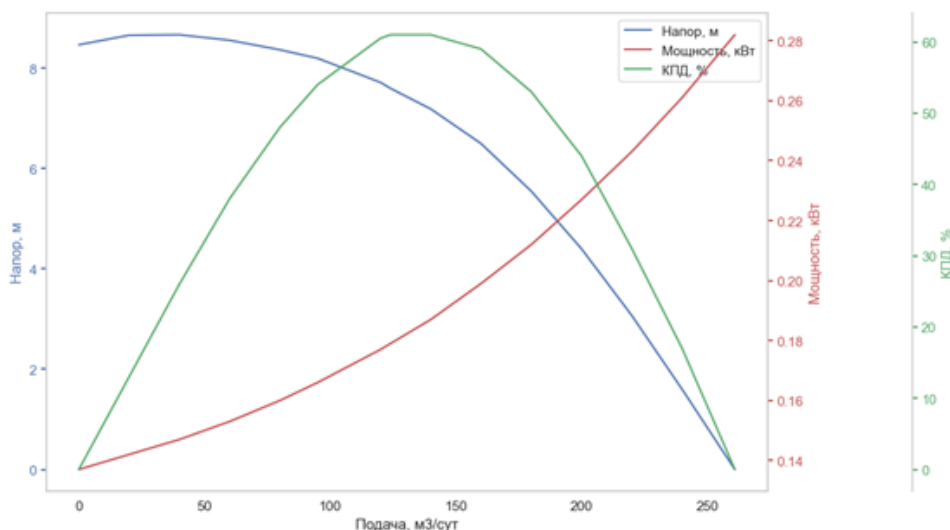


Рис.2.3. Напорно-расходная характеристика для 1 ступени

Как правило, напорно-расходная характеристика снимается на технической воде при номинальной частоте вращения 50 Гц [2]. Однако, в условиях нефтяных скважин насос перекачивает углеводородные смеси, состоящие из воды, нефти и газа. Вязкость углеводородной смеси значительно отличается от вязкости технической воды, что вызывает повышенные потери на трение в ступенях ЭЦН и приводит к снижению развиваемого напора, снижению КПД и повышению мощности.

Использование характеристик, полученных в результате испытаний на технической воде, не может гарантировать достаточную точность при моделировании скважины, поэтому необходимо корректировать паспортную характеристику ЭЦН для обеспечения достаточной точности модели. В качестве решения описанной проблемы институт гидравлики [6] предложил умножать паспортные характеристики, полученные на воде, на поправочные коэффициенты:

$$\begin{aligned} Q_{visc} &= C_Q Q_w \\ H_{visc} &= C_H H_w \\ \eta_{visc} &= C_\eta \eta_w \end{aligned} \quad (2.16)$$

где $Q_{visc}, H_{visc}, \eta_{visc}$ – скорректированная подача, напор и КПД; Q_w, H_w, η_w – подача, напор и КПД, полученные в результате испытаний на воде; C_Q, C_H, C_η – поправочные коэффициенты по подаче, напору и КПД.

Поправочные коэффициенты являются функциями скорректированного расхода жидкости, Q^* , который зависит от вязкости жидкости, напора и номинальной подачи:

$$Q^* = \exp\left(\frac{39.5276 + 26.5605 \ln(\nu) - y}{51.6565}\right) \quad (2.17)$$

$$y = -7.5946 + 6.6504 \ln(H_{wBEP} + 12.8429 \ln(Q_{wBEP}))$$

где Q_{wBEP}, H_{wBEP} – номинальная подача и напор для воды в галлонах в день и футах, ν – кинематическая вязкость (сСт).

Тогда поправочные коэффициенты определяются следующим образом:

$$C_Q = 1 - 4.0327 \cdot 10^{-3} Q^* - 1.724 \cdot 10^{-4} (Q^*)^2$$

$$C_\eta = 1 - 3.3075 \cdot 10^{-2} Q^* + 2.8875 \cdot 10^{-4} (Q^*)^2 \quad (2.18)$$

Определение поправочного коэффициента по напору отличается от определения поправочных коэффициентов по подаче и КПД. Для напора вводятся 4 поправочных коэффициента, которые соответствуют 60%, 80%, 100% и 120% подаче от номинальной подачи по воде (Q_{wBEP}):

$$C_{H0.6} = 1 - 3.68 \cdot 10^{-3} Q^* - 4.36 \cdot 10^{-5} (Q^*)^2$$

$$C_{H0.8} = 1 - 4.4723 \cdot 10^{-3} Q^* - 4.18 \cdot 10^{-5} (Q^*)^2$$

$$C_{H1.0} = 1 - 7.00763 \cdot 10^{-3} Q^* - 1.41 \cdot 10^{-5} (Q^*)^2$$

$$C_{H1.2} = 1 - 9.01 \cdot 10^{-3} Q^* + 1.31 \cdot 10^{-5} (Q^*)^2 \quad (2.19)$$

После определения описанных точек восстановить полную зависимость напора от расхода можно, например, с помощью методов интерполяции.

Как было сказано выше, все паспортные характеристики снимаются при частоте вращения 50 Гц. Однако, стоит отметить, что на данный момент работу центробежных насосов можно корректировать за счет увеличения или уменьшения частоты вращения, тем самым контролируя добычу на скважине. Работа ЭЦН при различных частотах давно изучена [6], существуют законы подобия, которые описывают зависимость между текущей частотой вращения насоса и основных его параметров:

$$\begin{aligned}
 Q_2 &= Q_1 \left(\frac{F_2}{F_1} \right) \\
 H_2 &= H_1 \left(\frac{F_2}{F_1} \right)^2 \\
 N_2 &= N_1 \left(\frac{F_2}{F_1} \right)^3
 \end{aligned}
 \tag{2.20}$$

где F_1, F_2 — номинальная и текущая частота вращения; Q_1, Q_2 — подача при номинальной и текущей частоте вращения; H_1, H_2 — напор при номинальной и текущей частоте вращения; N_1, N_2 — мощность при номинальной и текущей частоте вращения.

Из законов подобия можно сделать следующие выводы:

- подача насоса прямо пропорционально зависит от частоты вращения;
- напор насоса от частоты вращения имеет квадратичную зависимость;
- потребляемая мощность имеет кубическую зависимость от частоты вращения;
- КПД насоса не зависит от частоты вращения.

На рис.2.4 изображена зависимость напора от подачи при различных частотах вращения

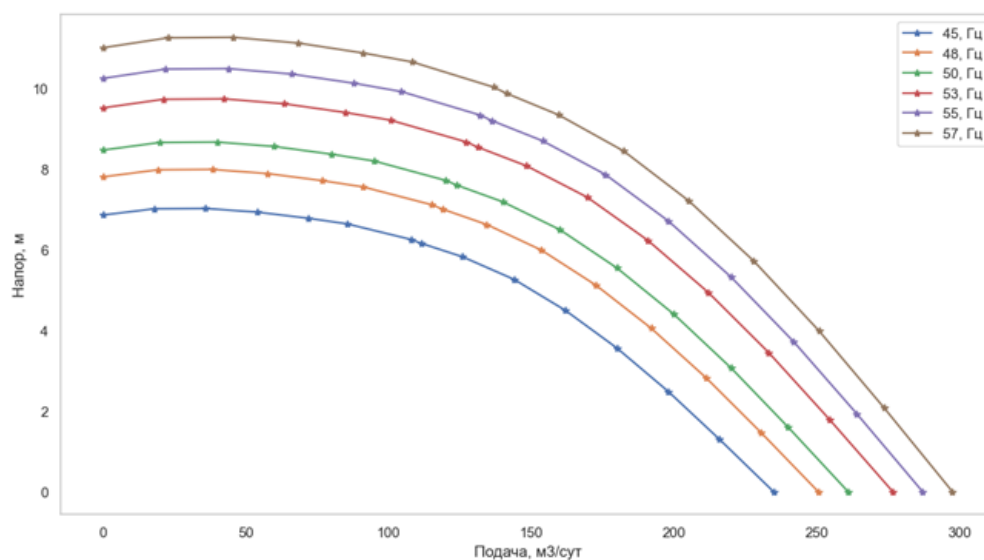


Рис.2.4. Зависимость напорной характеристики от частоты вращения ЭЦН для одной ступени

Таким образом, полученная модель ЭЦН позволяет моделировать работу ЭЦН при частотах вращения, отличающихся от номинальной, а также учитывать вязкость углеводородной смеси, отличающейся от вязкости технической воды.

2.2. Статистическая модель

У каждой физической модели есть свои ограничения, в том числе для оптимизации модели с вычислительной точки зрения приходится вносить в модель упрощения. В сложных процессах часто можно наблюдать неизвестные физические явления, о которых физическая модель не имеет представления. По данным причинам различие между физической моделью и реальностью неизбежно [10]. Физическим моделям часто не хватает гибкости, то есть способности адаптироваться к неизвестным и не моделируемым физическим явлениям.

В качестве альтернативы моделям на физических принципах, для оценки дебита можно использовать статистические модели. Статистические модели основаны на нахождении взаимосвязи между входными и выходными параметрами системы без точного физического описания рассматриваемых процессов [10].

Основным достоинством статистических моделей является то, что такие модели позволяют отказаться от детального физического моделирования систем или процессов, для которых точное решение может быть трудно найти численно. Статистические методы полагаются на то, что экспериментальные или промышленные данные хорошо представляют систему, и пытаются узнать взаимосвязи, которые описывают систему, непосредственно из данных [10].

Статистические модели обладают высокой гибкостью и способны адаптироваться к неизвестным и не моделируемым физическим явлениям. Преимуществом использования статистических моделей также является их низкая вычислительная стоимость по сравнению с физическими моделями, которые обычно решают сложные системы уравнений состояния. Это преимущество позволяет статистическим моделям прогнозировать дебит скважины в режиме реального времени [11].

Процесс создания и эксплуатации статистической модели представлен на рис.2.5.

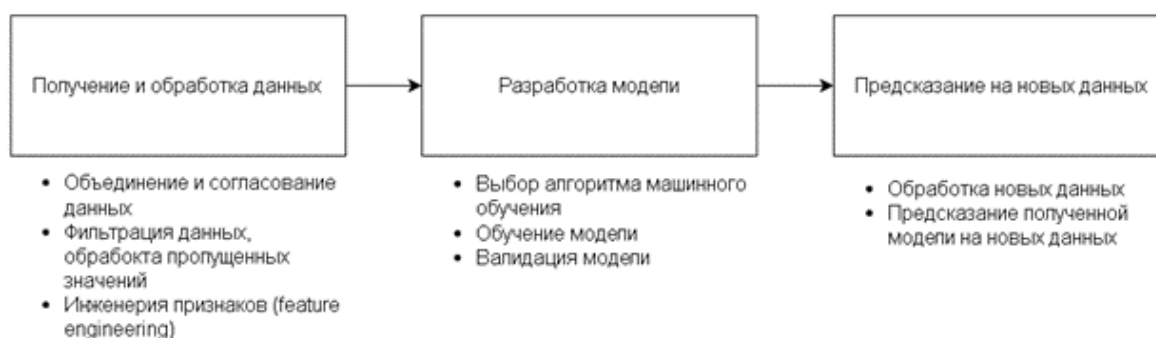


Рис.2.5. Процесс создания и эксплуатации статистической модели [10]

Для разработки модели первым этапом необходимо собрать исторические данные для обучения. Вторым этапом необходимо объединить данные из разных источников и согласовать их. Убедиться, что обработанные данные не имеют шумов, выбросов. В зависимости от модели машинного обучения убедиться, что отсутствуют пропущенные значения в данных. Третьим этапом необходимо провести feature engineering. Feature engineering – процесс выделения из данных новых признаков, которые могут помочь в обучении модели, подробнее эта техника будет рассмотрена в 2.2.3. Четвертым этапом необходимо выбрать модель машинного обучения и обучить ее на тренировочных данных. Полученная модель должна быть протестирована на новом наборе данных для проверки обобщающей способности алгоритма. После успешной валидации модель может предсказывать выходную переменную по новым входным данным.

2.2.1. Задачи машинного обучения

Глобально, алгоритмы машинного обучения можно разделить на 2 вида:

- обучение с учителем (supervised learning);
- обучение без учителя (unsupervised learning).

При обучении с учителем на вход модели подается размеченные данные, которые включают входные данные X и целевую переменную Y (выходные данные). В процессе обучения модель находит взаимосвязь между входными и выходными данными. Результатом обучения модели является функция, которая способна по новым входным данным X предсказывать целевую переменную Y . Типичной задачей обучения с учителем является классификация. Алгоритмы классификации используются, когда выходные данные ограничены набором значений (метками или классами). Хорошим примером классификации является спам-фильтр: он обучается на множестве примеров писем с указанием их класса (спам или нет), и ему необходимо научиться классифицировать новые письма [4]. Другой типичной задачей обучения с учителем является предсказание целевой числовой переменной, например цены на автомобиль или его пробег. Такая задача называется регрессией [4].

При обучении без учителя на вход алгоритма подаются только входные данные. Такие данные не имеют разметки, в процессе обучения алгоритм находит в данных структуру, общие черты и реагирует на изменение найденных черт в каждой новой порции данных. Типичными задачами обучения без учителя

являются задачи кластеризации, визуализации, снижения размерности и задачи по выявлению аномалий [4].

Задача предсказания дебита по входным параметрам является задачей регрессии, так как во время замера скважины известен ее дебит. В качестве целевого признака выступает дебит с замерной установки, а в качестве входных данных могут быть использованы регистрируемые данные о работе погружного оборудования.

2.2.2. Модели машинного обучения

Существует большое множество алгоритмов машинного обучения для решения задач регрессии, основанных на различных математических принципах. У каждой модели машинного обучения есть свои ограничения, производительность различных алгоритмов сильно зависит от структуры и размера данных, поэтому каждый алгоритм машинного обучения может в лучшей или худшей степени находить и описывать имеющиеся в данных зависимости.

Одними из самых популярных и простых моделей машинного обучения являются решающие деревья. Деревья решений реализуют логику иерархически организованной системы вопросов. Цель дерева состоит в том, чтобы, основываясь на обучающей выборке, подобрать серии решающих правил, результатом которых является конечный ответ из тренировочной выборки. Условия, разделяющие данные на две различные группы называются узлами, направления дальнейшего разбиения данных являются ветвями, конечные решения листьями (рис.2.6).

Одним из преимуществ дерева решений является его интерпретируемость за счет возможности визуализации решающих правил. Решающие деревья могут применяться в задачах регрессии и классификации. В задаче классификации в качестве предсказательного значения выступает метка класса в листке, а в случае задачи регрессии предсказательное значение выступает в роли среднего арифметического листьев последнего узла.

Однако, одиночные деревья имеют недостатки – они легко поддаются переобучению [5]. Переобучение – явление, при котором модель чрезмерно подгоняется к зависимостям в обучающей выборке, не выявляя при этом общих зависимостей. Переобученные модели имеют малую ошибку на обучающих данных, но имеют слабую предсказательную способность на новых данных. При отсутствии ограничений древовидная структура будет адаптироваться к обучающим данным очень

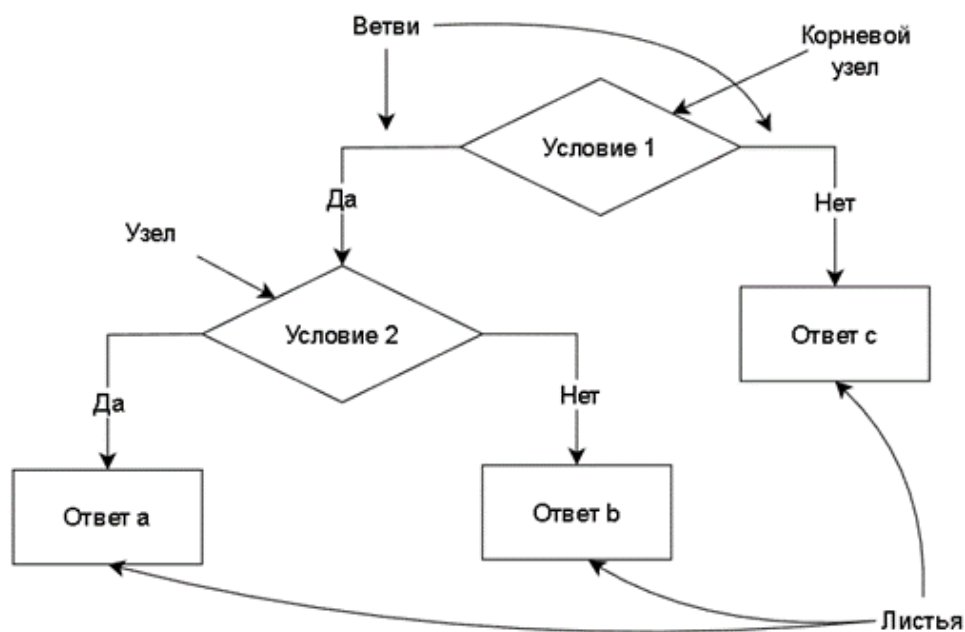


Рис.2.6. Схема деревьев решений

точно подгоняя их. Чтобы избежать переобучения, структуру дерева ограничивают во время обучения. Такая процедура называется регуляризацией [5]. Например, структуру дерева можно ограничивать с помощью следующих параметров:

- максимальная глубина дерева;
- максимальное количество листьев;
- минимальное количество объектов, которые должны находиться в листе дерева;
- минимальное количество объектов, необходимых для разделения узла.

Другим способом борьбы с переобучением является ансамблирование или композиция моделей. В отличие от одиночных деревьев решений, которые могут быть склонны к переобучению, ансамблирование деревьев позволяет снизить ошибку прогнозирования и повысить общую точность модели. На сегодняшний день градиентный бустинг является одним из наиболее популярных и эффективных алгоритмов машинного обучения, который используется для решения задач классификации и регрессии. Он основан на идее последовательного построения слабых моделей (как правило деревьев решений) и их комбинирования в одну сильную модель. Главная идея градиентного бустинга заключается в том, чтобы каждая последующая модель исправляла ошибки предыдущей модели. Это позволяет достигать высокой точности прогнозирования даже в сложных задачах, таких как распознавание образов и анализ текстовых данных.

Основными преимуществами градиентного бустинга являются:

- высокая точность предсказаний;
- способность эффективно находить нелинейные зависимости в данных различной природы;
- способность отлично работать на выборках с «табличными», неоднородными данными;
- нечувствительность к масштабированию признаков;
- способность обрабатывать данные большой размерности, сохраняя при этом высокую скорость обучения;
- устойчивость к явлению переобучения;
- устойчивость к выбросам и шуму в данных.

В том числе, благодаря перечисленным достоинствам он широко применяется во многих конкурсах по машинному обучению [4].

2.2.3. Обучение статистических моделей

При проверке эффективности модели могут возникнуть две ситуации:

- ошибки на тренировочных и тестовых данных велики;
- ошибка на тренировочных данных мала, но велика на тестовых.

В первой ситуации модель не может найти взаимосвязь между входными и выходными параметрами, такое явление называют недообучением модели. Во второй ситуации модель слишком хорошо подстроилась под зависимости тренировочных данных, не найдя при этом общего правила в данных, такое явление называется переобучением. Целью обучения алгоритма является найти середину между этими ситуациями, в таком случае алгоритм будет обладать хорошей обобщающей способностью и может быть использован для прогнозирования выходных параметров на новых данных.

Одним из самых простых способов оценки обобщающей способности алгоритма является разделение данных на обучающую и тестовую выборку. Весь набор данных разделяется на две части в примерном отношении 80% на 20% или 70% на 30%. Модель обучается на тренировочной выборке, чтобы впоследствии применить найденные зависимости на новые данные. Тестовая выборка используется, чтобы проверить обобщающую способность обученной модели. Такой подход обладает существенными недостатками:

- уменьшается размер тренировочной выборки. Особенно актуальной проблемой является при небольших наборах данных;

- результат тестирования алгоритма сильно зависит от способа разбиения данных.

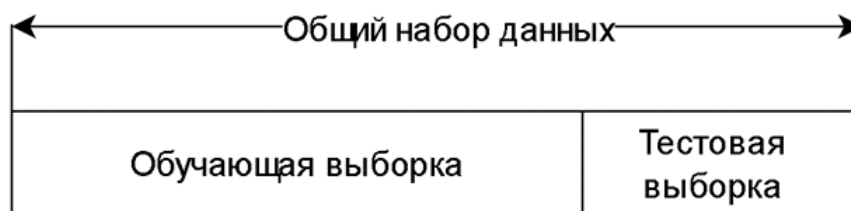


Рис.2.7. Схема разбиения общего набора данных на обучающую и тестовую выборку.

Альтернативным и более устойчивым к способу разбиения данных является метод кросс-валидации. Кросс-валидация позволяет более точно оценить производительность модели и уменьшить вероятность переобучения на конкретных выборках данных. Данный подход разделяет набор данных на N подмножеств. Далее происходит N итераций, во время каждой из которых $N-1$ подмножеств выступают в роли обучающей выборки, а оставшееся подмножество выступает в роли тестовой выборки. Итоговой оценкой модели является среднее значение N получившихся оценок модели или вычисляется оценка модели на отложенном множестве, которое не участвовало в кросс-валидации (рис.2.8).

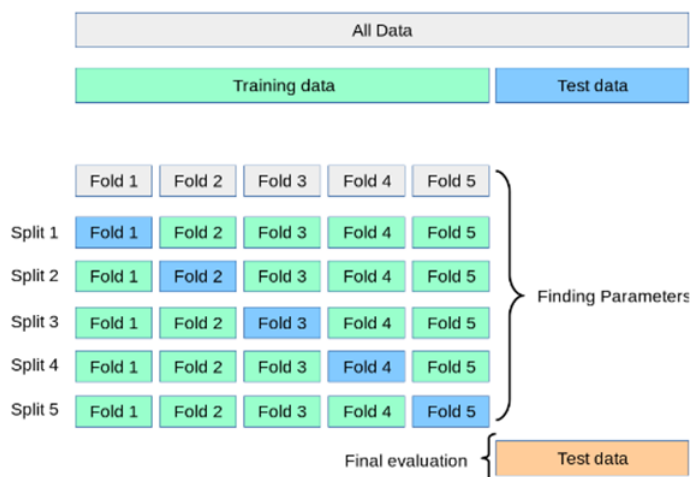


Рис.2.8. Схема кросс-валидации.

Еще одним способом борьбы с переобучением является настройка гиперпараметров модели. Гиперпараметры — это параметры модели, которые не определяются в процессе обучения модели, а устанавливаются до начала ее обучения [10]. Гиперпараметры напрямую влияют на предсказательную способность алгоритма и позволяют сделать модель более подходящей для конкретной задачи. Они позволяют уменьшить влияние шума и выбросов на конечные предсказания алгоритма. Наиболее популярными видами выбора оптимальных гиперпараметров

являются метод сеточного поиска (GridSearchCV) и метод случайного поиска (RandomSearchCV) с использованием перекрестной проверки.

Сеточный поиск позволяет перебрать все возможные комбинации значений гиперпараметров из заданного диапазона и выбрать наилучшую комбинацию параметров для модели. При использовании GridSearchCV из заданного диапазона гиперпараметров создается сетка, затем для каждой комбинации параметров модель машинного обучения тренируется и оценивается с помощью кросс-валидации. Затем выбирается лучший набор гиперпараметров на основе наивысшей точности оценки модели кросс-валидации.

Одними из недостатков сеточного поиска является его скорость и ограниченность в выборе гиперпараметров. Для поиска лучшей комбинации гиперпараметров необходимо протестировать каждую комбинацию. Альтернативным методом для поиска лучшей комбинации является случайный поиск (RandomSearchCV). RandomSearchCV вместо перебора всех возможных комбинаций, оценивает заданное количество случайных комбинаций, выбирая произвольное значение для каждого гиперпараметра на каждой итерации [4].

2.2.4. Методы оценки модели машинного обучения

Для понимания обобщающей способности модели машинного обучения ее необходимо оценить. Для задач регрессии существует большое множество метрик оценки, которые имеют свои преимущества и недостатки. Для оценки производительности модели рекомендуется использовать несколько метрик одновременно, чтобы получить более полное представление о ее работе. Кроме того, при выборе метрик следует учитывать особенности конкретной задачи.

Одной из самых популярных метрик является среднеквадратическая ошибка (MSE), которая измеряет среднеквадратичную разницу между прогнозируемыми и фактическими значениями.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^{pred})^2 \quad (2.21)$$

где y_i — реальное значение целевой переменной; y_i^{pred} — предсказанное значение целевой переменной.

Одним из недостатков MSE является чувствительность к выбросам и плохая интерпретируемость. Например, при вычислении MSE для дебита скважин мы

получим размерность ошибки [м3/сут]². В таких случаях прибегают к метрике MAE (mean absolute error), которая измеряет среднюю абсолютную разницу между прогнозируемыми и фактическими значениями:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^{pred}| \quad (2.22)$$

Описанные метрики позволяют оценить фактическое отклонение от значений целевого признака, но не дают представления об эффективности применения алгоритма. В таких случаях используют коэффициент детерминации (R^2):

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^{pred})^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.23)$$

Где \hat{y}_i — среднее значение по выборке.

Коэффициент детерминации принимает значения от нуля до единицы и отражает эффективность применения модели машинного обучения по сравнению с интуитивно понятным решением в виде среднего по выборке [3].

ГЛАВА 3. РЕАЛИЗАЦИЯ АЛГОРИТМА ГИБРИДНОЙ МОДЕЛИ ВИРТУАЛЬНОЙ РАСХОДОМЕТРИИ

В главе рассматривается предлагаемый подход к гибридизации моделей виртуального расходомера. В параграфе 3.2 описываются основные этапы, необходимые для восстановления значений дебита по косвенным параметрам с помощью физической модели: обработка данных, адаптация модели, восстановление дебитов. В параграфе 3.3 описывается создание статистической модели, выбор алгоритма для машинного обучения, подходы к обработке данных, фильтрации, создание новых признаков (feature engineering), подходы к улучшению модели.

3.1. Схема работы алгоритма

Предложенный алгоритм реализует последовательное соединение физической и статистической модели. А именно последовательность «физическая модель – статистическая модель – физическая модель» (рис.3.1).



Рис.3.1. Схема построения гибридной модели.

Классический физический виртуальный расходомер хорошо себя показывает при наличии стационарного режима течения флюида, при наличии фактических замеров дебита с помощью замерной установки для адаптации модели под реальный режим работы скважины [7; 10]. Физическая модель сопоставляет модельный режим работы скважины с ее фактическим режимом работы с помощью калибровочных коэффициентов.

Однако, есть случаи, при которых калибровочные коэффициенты не всегда могут быть определены или могут быть определены некорректно. К этим случаям можно отнести:

1. Незамеряемый фонд скважин. К этому фонду скважин можно отнести скважины, которые не замерялись больше 1 месяца. При таких ситуациях с момента последнего замера режим скважины мог измениться, например сменится режим работы (с ПДФ на ПКВ) или проведены геолого-технические мероприятия. При таких ситуациях калибровочный коэффициент будет отражать прошлый режим работы скважины, что может не соответствовать действительности;
2. Запуск новых скважин. При запуске новых скважин виртуальному расходомеру не будет хватать информации и системе, так как у него не будет знаний о дебите скважины, которые необходимы для адаптации модели. В таких ситуациях калибровочных коэффициент не определен;
3. Смена ЭЦН на скважинах. При смене ЭЦН без наличия нового замера калибровочный коэффициент также считается неопределенным. При наличии по скважине калибровочного коэффициента по прошлому ЭЦН его брать некорректно, так как у нового насоса может быть другой производитель и характеристики;
4. При неисправности замерной установки. При неисправности замерной установки виртуальный расходомер может показывать некорректные значения восстанавливаемого дебита, так как некорректные замеры дебита с помощью замерной установки напрямую влияют на корректность калибровочных коэффициентов.

Для решения описанных проблем и дополнения физического виртуального расходомера предлагается использовать статистическую модель для прогнозирования калибровочных коэффициентов. Предлагаемый алгоритм можно разделить на три этапа:

1. Расчет калибровочных коэффициентов с помощью физической модели за длительный промежуток времени на большом количестве скважин. Большое количество скважин за длительный промежуток времени позволит собрать репрезентативную выборку, которая будет содержать в себе эксплуатационные условия конкретного месторождения для наилучшей обобщающей способности статистического алгоритма;
2. Создание и обучение статистической модели с последующим предсказанием калибровочных коэффициентов на новых данных;
3. Расчет физического виртуального расходомера с предсказанными калибровочными коэффициентами с помощью статистической модели.

3.2. Физическая модель

Основная цель виртуальной расходомерии для скважин, оснащенных ЭЦН - восстановление дебита скважины по косвенным параметрам, замеры которых обладают большей дискретностью, чем замеры АГЗУ. Системы виртуальной расходомерии на базе физических моделей являются наиболее широко используемыми виртуальными расходомерами в отрасли [10].

Общая концепция решения задачи виртуальной расходомерии представлена на рис.3.2



Рис.3.2. Архитектура модуля виртуальной расходомерии для скважин с ЭЦН на базе физической модели.

Для восстановления значений дебита необходимо выполнить последовательно следующие шаги:

1. Этап обработки данных. На данном этапе статические и динамические данные согласуются между собой, после чего происходит фильтрация и обработка пропущенных значений;
2. Этап адаптации модели. На этапе адаптации по высокочастотным замерам телеметрии восстанавливаются мгновенные значения замерного дебита АГЗУ, после чего модель скважины соотносится с реальным режимом работы с помощью определения поправочных коэффициентов (далее “коэффициенты калибровки”). Далее происходит анализ полученных коэффициентов калибровки, с последующим их определением в моменты времени, когда дебит отсутствовал;

3. Этап восстановления дебитов. На этапе восстановления дебитов решается обратная оптимизационная задача – по высокочастотным данным телеметрии и калибровочным коэффициентам подбирается дебит скважины. После восстановления дебита происходит анализ полученных результатов.

Таким образом, реализованный алгоритм позволяет с помощью физического моделирования получать полезную информацию из сырых необработанных данных. Стоит отметить, что рассматриваемый модуль был разработан на языке Python с использованием библиотеки Unifloc и позволяет масштабировать, улучшать отдельные расчетные подмодули с общим увеличением качества работы алгоритма.

3.2.1. Обработка данных

Данные, необходимые для расчета виртуального расходомера, могут быть получены из различных источников. Данные могут приходиться с различной частотой, например, давление на приеме ЭЦН является высокочастотным параметром и может приходиться раз в 15 минут, а замерной дебит жидкости АГЗУ, в свою очередь, может замеряться на скважине 1 раз в 7 дней. Поэтому полученные данные необходимо объединить и привести к единой временной шкале.

Объединённые данные можно разделить на два типа:

- статические данные – параметры, не меняющиеся в течение работы скважины с ЭЦН: конструкция скважины, инклинометрия, спущенное оборудование (ЭЦН, НКТ), PVT свойства флюидов и т.п.;
- динамические данные – параметры, меняющиеся во время работы скважины: частота питающего тока, напряжения, токи по фазам, дебит скважины, активные мощности, давления и т.п.

Набор входных данных, необходимых для обучения восстановления дебита с помощью физической модели виртуального расходомера представлен в табл.3.1

После объединения данных важным этапом является фильтрация. В полученных данных необходимо устранить артефакты сохранения. Могут существовать ошибки в размерностях данных, например, атмосферы и мега паскалы могут соседствовать в одном временном ряде линейного давления. Также вышедшие из строя датчики, особенно давления на приеме, могут передавать «нереальные» значения. Подобные артефакты в данных устраняются с помощью задания физических границ для каждого из параметров, в пределах которых значения можно использовать в алгоритме виртуальной расходомерии.

Таблица 3.1

Описание данных, необходимых для расчета виртуального расходомера.

Группа параметров	Описание	Тип данных
Давления	Линейное давление, давление на приеме ЭЦН	Динамические данные
Данные о работе погружного оборудования	Активная мощность, токи, напряжение, коэффициент мощности, температура ЭЦН и ПЭД	Динамические данные
Конструкция скважины	Диаметр НКТ и ЭК, инклинометрия	Статические данные
PVT свойства	Плотности фаз, газосодержание, вязкость нефти, объемный коэффициент нефти, давление насыщения	Статические данные
Информация о погружном оборудовании	Модель ЭЦН, количество секций, количество ступеней, коэффициент сепарации газа, глубина спуска ЭЦН, напорно-расходная характеристика ЭЦН	Статические данные

Необходимо понимать, что некоторых данных, необходимых для расчета виртуального расходомера, может не быть. При таких ситуациях пропущенные значения могут быть заменены на значения по умолчанию или могут быть вычислены по другим параметрам. Например, одним из самых критичных параметров для виртуального расходомера является давление на приеме. В случае поломки датчика, для обеспечения стабильности данные о давлении на приеме берутся из технологического режима.

3.2.2. Адаптация модели

После объединения, фильтрации, обработки пропущенных значений и удаления шумов необходимо подготовить данные для адаптации модели и нахождения калибровочных коэффициентов.

Для сопоставления модели с реальным режимом работы скважины необходимо согласовать высокочастотные параметры (электротелеметрии) и низкочастотные (замеры дебита).

В работах [7; 8] для согласования параметров использовалось осреднение всех параметров по времени замера дебита (обычно 1 час). Однако, при осреднении параметров может теряться потенциально важная информация о работе скважины, например, для скважин в ПКВ режиме при применении осреднения пропадает информация о мгновенном расходе через насос. Информация о мгновенном расходе в насосе может помочь понять в какой зоне и насколько эффективно работает погружное оборудование.

Другим способом для согласования параметров является восстановление мгновенного расхода через насос во время замера. Стоит отметить, что по своей

природе замер на АГЗУ является интегральным [9], итоговое значение замеры показывает какое количество продукции “прошло” за определенный промежуток времени, приведенное к размерности в м³/сут. При этом в случае работы ПКВ с периодом работы 30 минут и 30 минут накопления часовой замер на АГЗУ выдаст результат в 100 м³/сут, но при этом мгновенный дебит через насос будет 200 м³/сут в течение 30 минут работы, и 0 м³/сут во время накопления. Поэтому при восстановлении мгновенного расхода необходимо разграничивать циклы работы и накопления, например, ориентируясь на нулевые значения по току). Также при ПКВ режиме критически важна частота сохранения данных – при недостаточной дискретности циклы работы и накопления абсолютно не видны.

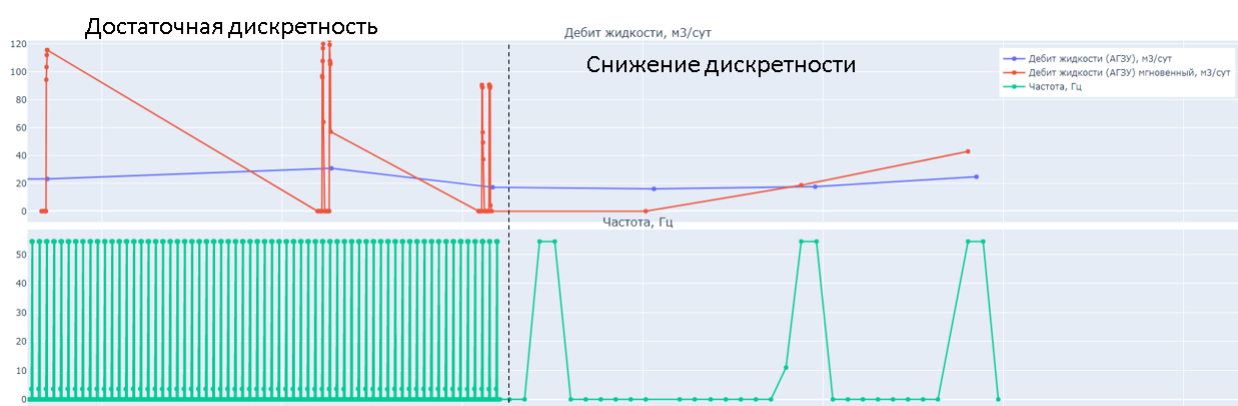


Рис.3.3. Восстановление мгновенного расхода через насос с достаточной и недостаточной дискретностью.

На рис.3.3 видно, что при достаточной дискретности алгоритм при восстановлении мгновенного расхода во время замера корректно разделяет время работы и накопления. При снижении дискретности циклы работы пропадают, что влечет за собой ошибку в определении мгновенного расхода через насос и калибровочных коэффициентов.

Для построения гидравлической модели необходимы данные по давлению на приеме и выкиде ЭЦН. В случае отсутствия датчика давления на выкиде ЭЦН, давление на выкиде может быть рассчитано от устья скважины по корреляции Бэггза и Брилла, описанной в подпараграфе 2.1.2.

После определения перепада давлений в насосе, в паспортные характеристики напорно-расходной характеристики вносятся поправки на вязкость нефти и текущую частоту вращения. По скорректированной напорно-расходной характеристике для каждого момента времени, где есть мгновенный расход, определяется модельный перепад давления через насос.

В итоге в один момент времени существует давление на выкиде ЭЦН, рассчитанное через модельный перепад давления через насос и давление на выкиде ЭЦН, рассчитанное по корреляции Беггза и Брилла от линейного давления (при отсутствии датчиков давления на выкиде). В силу износа оборудования, влияния свободного газа, механических примесей, отложений солей или других осложненных условий эксплуатации модельное и фактическое значения давления на выкиде ЭЦН могут отличаться. Для сопоставления модели скважины и ее фактического режима работы вводятся калибровочные коэффициенты, которые позволяют установить связь между дебитом скважины и косвенными параметрами:

$$K_{degr} = \frac{P_{dis}^{fact}}{P_{dis}^{model}} \quad (3.1)$$

$$P_{dis}^{model} = P_{intake} + dP_{model}$$

где K_{degr} – коэффициент калибровки (деградации); P_{dis}^{fact} – давление на приеме ЭЦН, рассчитанное по корреляциям от линейного давления (в случае отсутствия датчика давления на выкиде); P_{dis}^{model} – модельное давление на выкиде ЭЦН; P_{intake} – давление на приеме ЭЦН; dP_{model} – перепад давления через насос.

3.2.3. Прогнозирование калибровочных коэффициентов

После адаптации модели калибровочные коэффициенты определены только тогда, когда существует мгновенный расход дебита (рис.3.4). После этапа адаптации модели скважины и получения калибровочных коэффициентов напорно-расходной характеристики ЭЦН, необходимо восстановить их значения в тех моментах времени, где нет данных по дебиту.

Существует несколько подходов для решения задачи восстановления калибровочных коэффициентов. Например, в работе [7] для восстановления калибровочных коэффициентов используется линейная интерполяция для модели УЭЦН. В работе [8] используется машинное обучение, а именно гребневая регрессия для восстановления калибровочных коэффициентов по косвенным параметрам для модели штуцера.

Однако, для некоторых скважин замеры АГЗУ могут быть некорректными и иметь характерный шум. В таких случаях применение линейной интерполяции может слишком сильно подгонять модель под “шумные” замеры АГЗУ и не отражать реальной работы скважины (рис.3.5). Применение методов машинного обучения

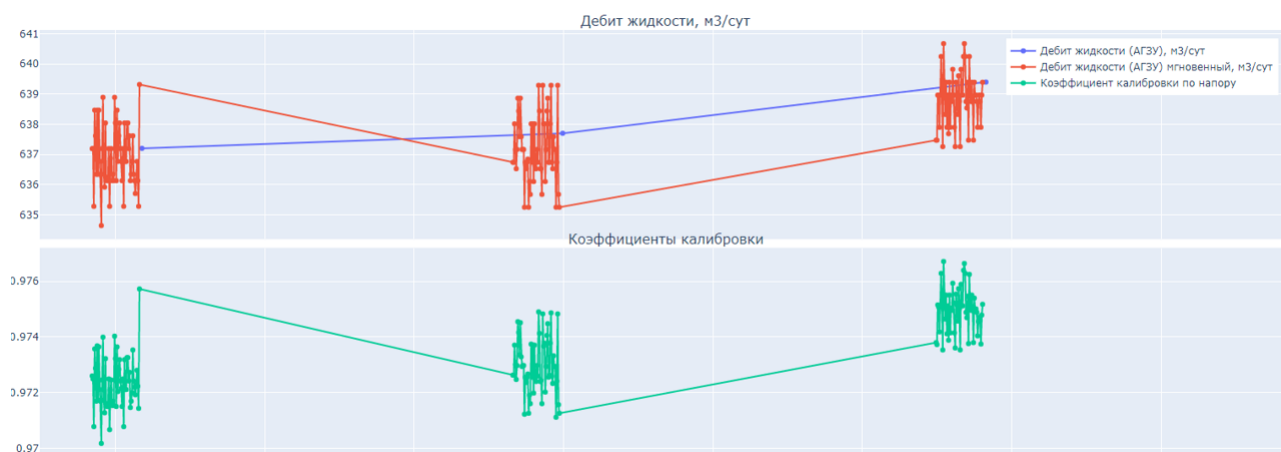


Рис.3.4. Восстановление мгновенного расхода через насос с достаточной и недостаточной дискретностью.

может быть слишком требовательным и в целом увеличить время расчета модели, а также излишним при отсутствии смены режима работы скважины. В данной работе для восстановления калибровочных коэффициентов в каждый момент времени предлагается брать медианное значение. Предложенный подход позволит увеличить обобщающую способность алгоритма и снизить влияние выбросов и некорректных значений дебитов с АГЗУ (рис.3.7).

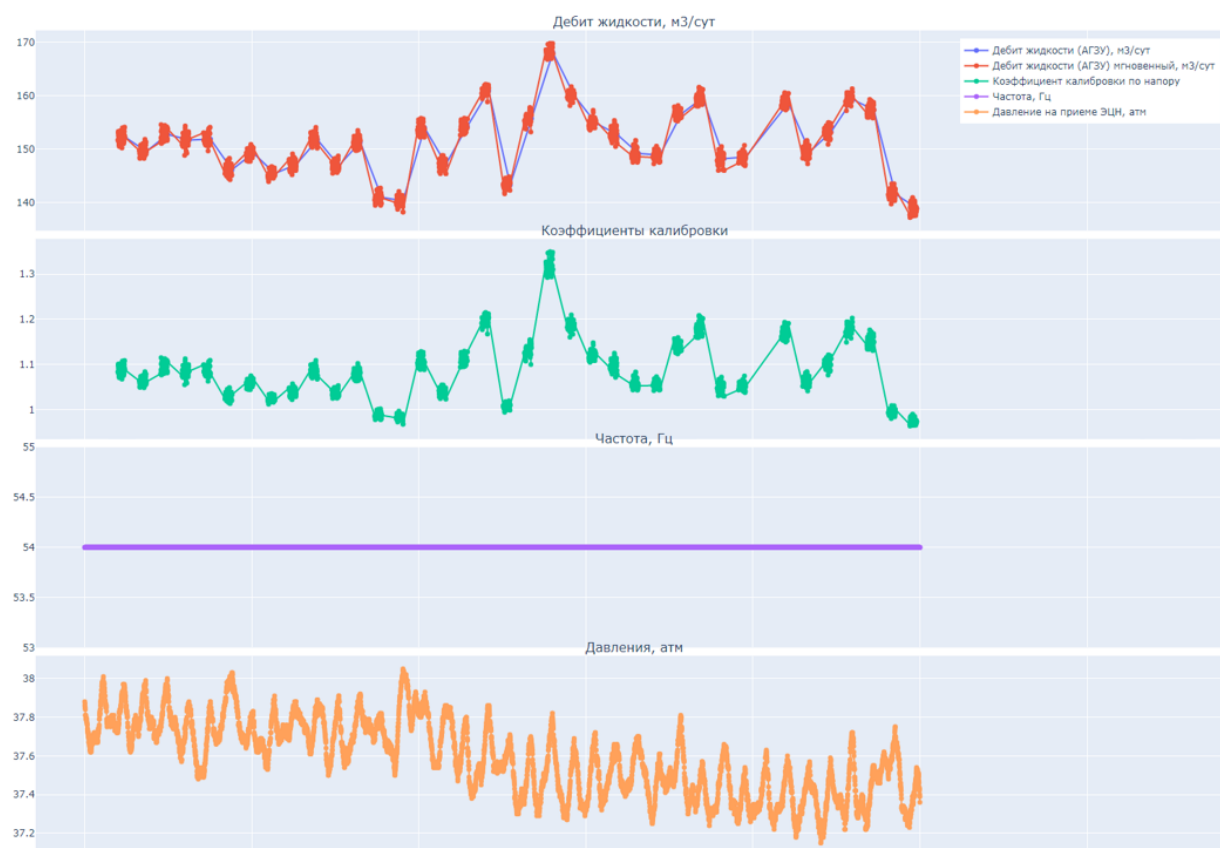


Рис.3.5. Коэффициенты калибровки для скважины 11 с «шумными» замерами АГЗУ.

3.2.4. Восстановление дебита

Следующим этапом после адаптации модели и восстановления калибровочных коэффициентов является решение обратной оптимизационной задачи – восстановление дебита по высокочастотным данным телеметрии и калибровочным коэффициентам. Для этого необходимо:

1. В каждый момент времени скорректировать модельные значения давления на выкиде ЭЦН с помощью калибровочных коэффициентов;
2. В каждый момент времени скорректировать напорно-расходную характеристику на текущую частоту вращения;
3. С помощью узлового анализа определить дебит в условиях насоса. Метод узлового анализа заключается в разбиении системы на отдельные элементы с определенным узлом по пути движения потока, для каждой из которых строится характеристика (зависимость давления в точке от дебита). Пересечения кривых и будет являться решением для данных условий (рис.3.6).
4. Перевести дебит скважины в поверхностные условия.

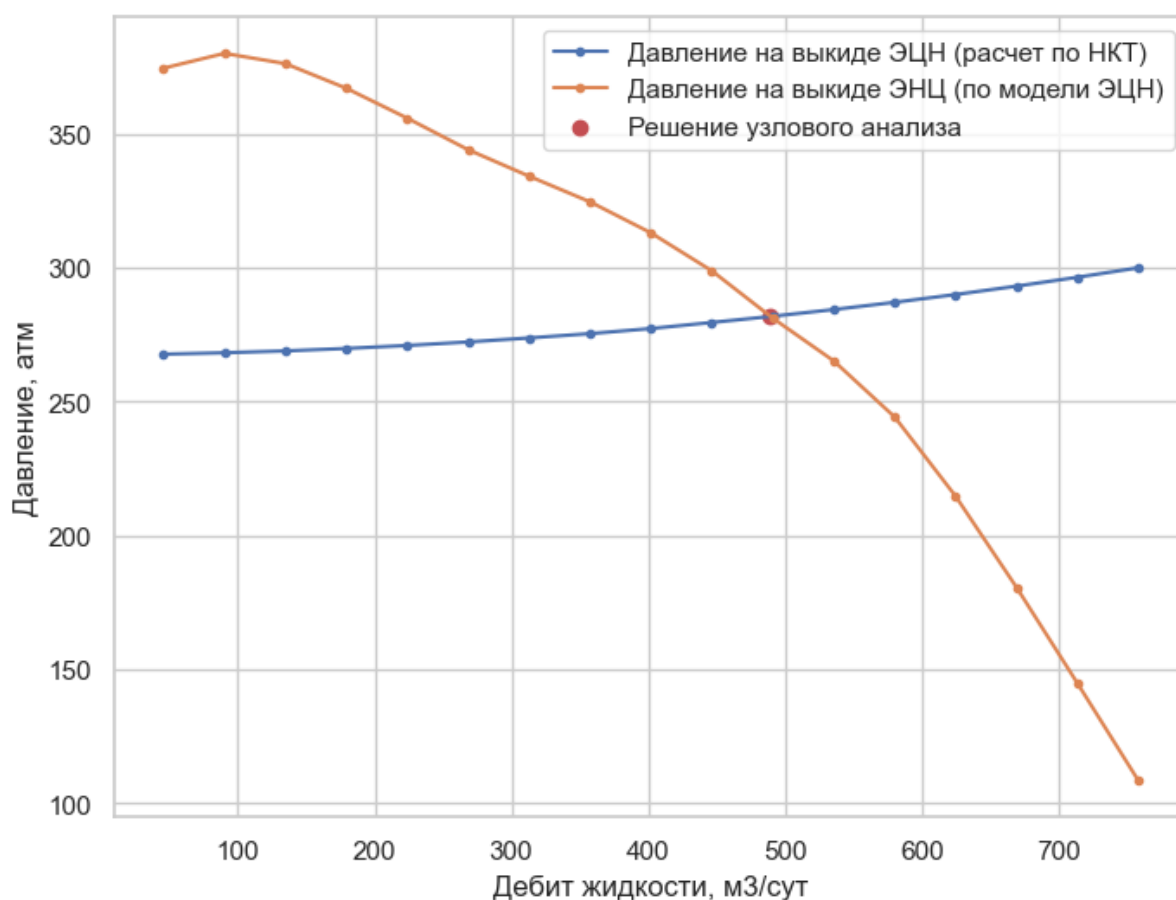


Рис.3.6. Пример узлового анализа скважины с узлом на выходе ЭЦН.

Проделав все описанные шаги, можно получить дебит для скважин, оборудованных ЭЦН. В качестве примера рассмотрим скважину 11 с сравнением двух методов восстановления калибровочных коэффициентов (рис.3.7), описанных в подпараграфе 3.2.3.

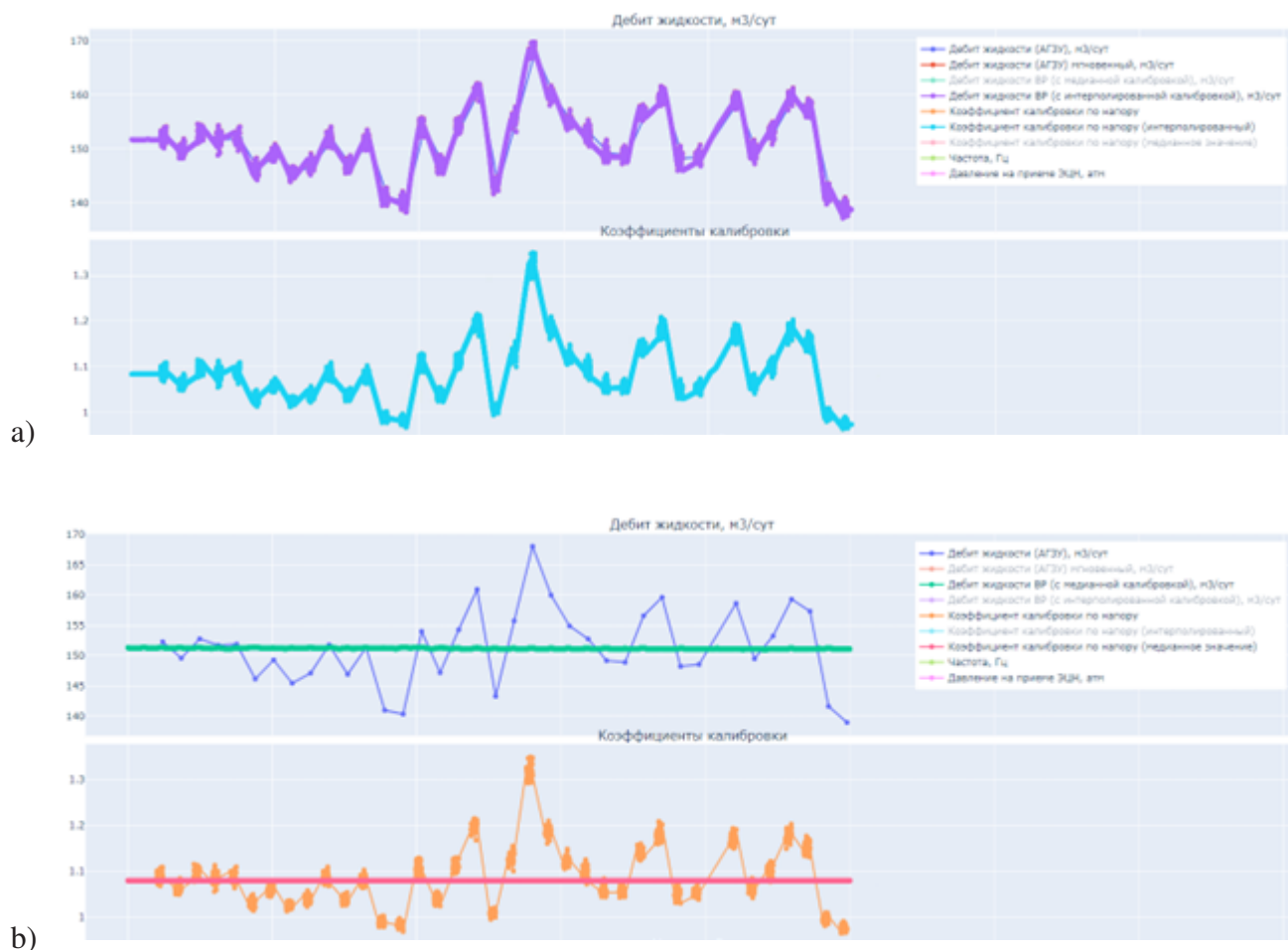


Рис.3.7. Пример восстановления дебита с помощью физической модели виртуального расходомера для скважины 11 (*a* — интерполированные значения калибровочных коэффициентов; *b* — медианное значение калибровочных коэффициентов.)

3.3. Статистическая модель

Основная цель статистической составляющей гибридной модели – найти взаимосвязь между целевой переменной (калибровочными коэффициентами) и входными параметрами (параметрами телеметрии, данных PVT и т. д.) для возможности прогнозирования калибровочных коэффициентов на новых данных.

Входные данные, необходимые для обучения модели совпадают с данными, необходимыми для расчета физической модели виртуального расходомера, которые описаны в подпараграфе 3.2.1.

В качестве целевой переменной будут выступать калибровочные коэффициенты, которые предварительно должны быть рассчитаны с помощью физической модели виртуального расходомера. Поэтому для формирования обучающей выборки предварительно с помощью физической модели виртуального расходомера был посчитан фонд скважин месторождения X за временной интервал 6 месяцев. Стоит отметить, что для месторождения X распределение калибровочных коэффициентов имеет медианное значение 0.9. Данное значение говорит о том, что большинство насосов работают с деградацией порядка 10% (рис.3.8).

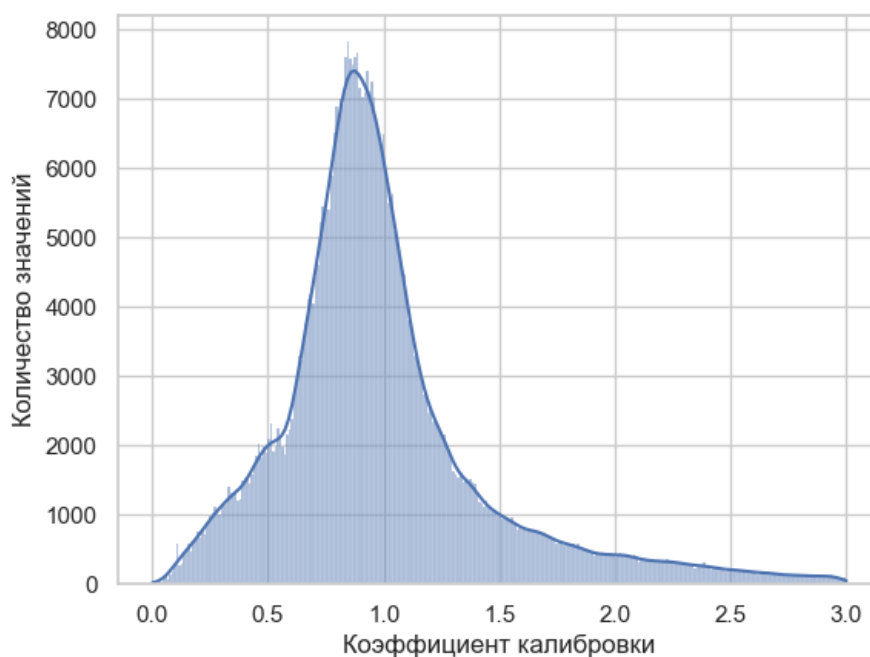


Рис.3.8. Распределение калибровочных коэффициентов.

Схема работы статистической модели изображена на рис.3.9.

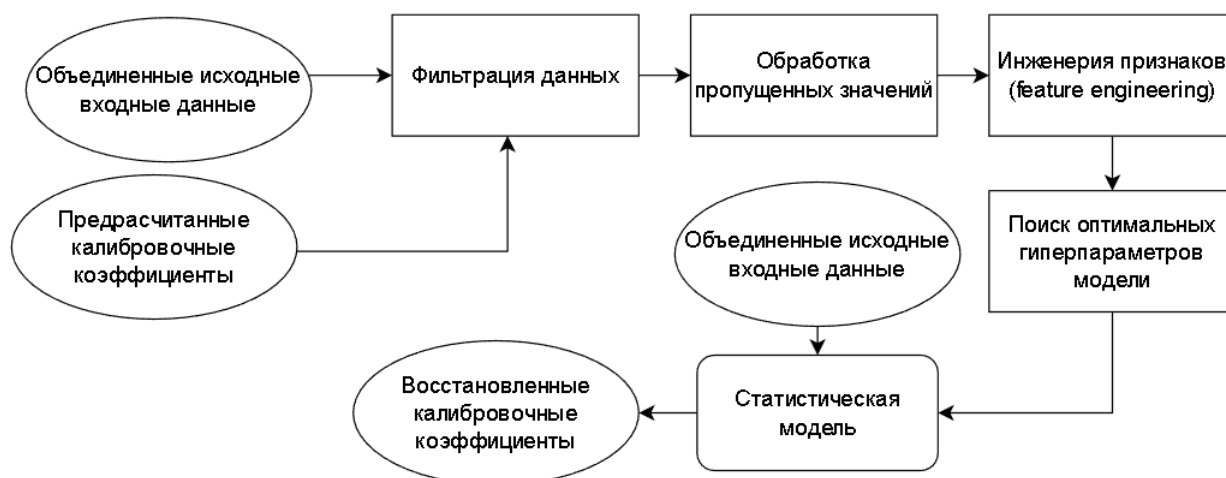


Рис.3.9. Схема работы статистической модели.

В описанной архитектуре последовательно производятся следующие действия:

1. Получение входных и выходных данных;
2. Фильтрация данных, которая будет описана в подпараграфе 3.3.1;
3. Обработка пропущенных значений, необходимых для конструирования новых признаков;
4. Создание новых признаков с помощью элементов расчета физической модели виртуального расходомера;
5. Поиск оптимальных гиперпараметров модели;
6. По выбранным оптимальной комбинации гиперпараметров обучается выбранная модель машинного обучения;
7. Для прогнозирования калибровочных коэффициентов обученная статистическая модель получает новые выходные данные.

3.3.1. Подготовка данных

Реальные промышленные данные могут иметь выбросы или показывать некорректные значения из-за неисправности в работе датчика, иметь отсутствие данных по некоторым параметрам, необходимым для расчета модели. Модели машинного обучения очень чувствительны к качеству данных, корректная подготовка данных имеет ключевое значение для достижения хорошей обобщающей способности алгоритма. Согласно [10], прежде чем начинать обучение модели необходимо убедиться, что собранные данные подходят для решения задачи путем удаления выбросов, обработки недостающих значений и удаления шумов.

Необходимо понимать, что алгоритм виртуального расходомера может не всегда корректно восстанавливать дебит скважины. Например, некорректный расчет может происходить из-за следующих факторов:

- не оптимальная зона работы насоса или наличие неопределенности в напорно–расходной характеристике [15];
- некорректная работа замерной установки;
- некорректно внесенные данные о работе скважины. Например, некорректные PVT свойства или паспортные характеристики ЭЦН;
- ограничение использования выбранной математической модели.

При формировании обучающей выборки для удаления шумов и наибольшей обобщающей способности алгоритма из всех скважин месторождения X были

выбраны скважины с хорошей исторической сходимостью дебита по виртуальному расходомеру и наибольшим доверием к модели. При этом, как было описано выше (подпараграф 3.2.2), дискретность данных при наличии интегрального замера дебита имеет огромную роль. Поэтому из списка скважин исключались скважины с недостаточной дискретностью данных.

Промысловые параметры телеметрии имеют выбросы и некорректные значения. Для фильтрации всех типов данных применялись физические уставки. В случае если параметр вышел за уставки, значение параметра удалялось.

Также одним важным этапом подготовки данных является обработка пропущенных или недостающих значений.

Стоит отметить, что калибровочные коэффициенты содержат в себе информацию о подготовке данных физической модели. Например, одним из критически важных параметров, необходимых для расчета физической модели, является давление на приеме. В случае поломки датчика давления на приеме показания можно брать давление на приеме из технологического режима. Поэтому необходимо синхронизировать обработку пропущенных значений в расчете физического виртуального расходомера с обработкой данных в статистической модели для прогнозирования калибровочных коэффициентов.

Как было сказано ранее, эффективность моделей сильно зависит от конструирования признаков [10]. Инженерия признаков получила свое название от того, что в сообществе машинного обучения входные данные часто называют признаками, поэтому инженерия признаков – это процесс манипулирования входными данными для выявления полезной информации, которая может помочь в процессе обучения модели. Feature engineering может помочь алгоритму найти сложные взаимосвязи между исходными данными и выходной переменной. Создание хороших признаков на основе входных данных, которые описывают процесс течения многофазного потока, может помочь получить более точные прогнозы [10].

Для конкретной скважины за один день может быть произведено несколько замеров. При большом количестве замеров АГЗУ калибровочные коэффициенты могут быть достаточно зашумленными (рис.3.5). Для обеспечения устойчивости алгоритм физического виртуального расходомера берет медианное значение в качестве итогового калибровочного коэффициента, поэтому предложенный алгоритм будет предсказывать не мгновенные значения калибровочных коэффициентов, а итоговое медианное значение, которое можно напрямую направлять в физическую

модель. Такой подход позволит уменьшить шум в данных и позволит повысить производительность алгоритма.

Параметры телемеханики крайне важны в определении калибровочных коэффициентов, однако, как и калибровочные коэффициенты, их необходимо агрегировать. Для сохранения динамики изменения параметров в качестве признаков брались максимальное и минимальное, среднее и медианное значение, а также стандартное отклонение за день

Напорно-расходная характеристика играет ключевую роль в восстановлении дебита с помощью алгоритма виртуального расходомера. В том числе от вида НРХ зависит точность виртуального расходомера и определение калибровочных коэффициентов. Например, при работе насоса в крайне левой зоне напорно-расходная характеристика имеет плоский тренд, что затрудняет восстановление дебита и определение калибровочного коэффициента. В данной зоне работы лучше себя ведет модель по мощности, так как в левой зоне мощностная характеристика не имеет плоской тенденции, однако модель по мощности проигрывает модели по напору в крайне правой зоне работы из-за плоской тенденции мощностной характеристики [15]. Для возможности учета вида НРХ и формирования новых признаков каждая характеристика нормируется и интерполируется. После чего в фиксированных срезах каждое значение выступает в качестве нового признака (рис.3.10).

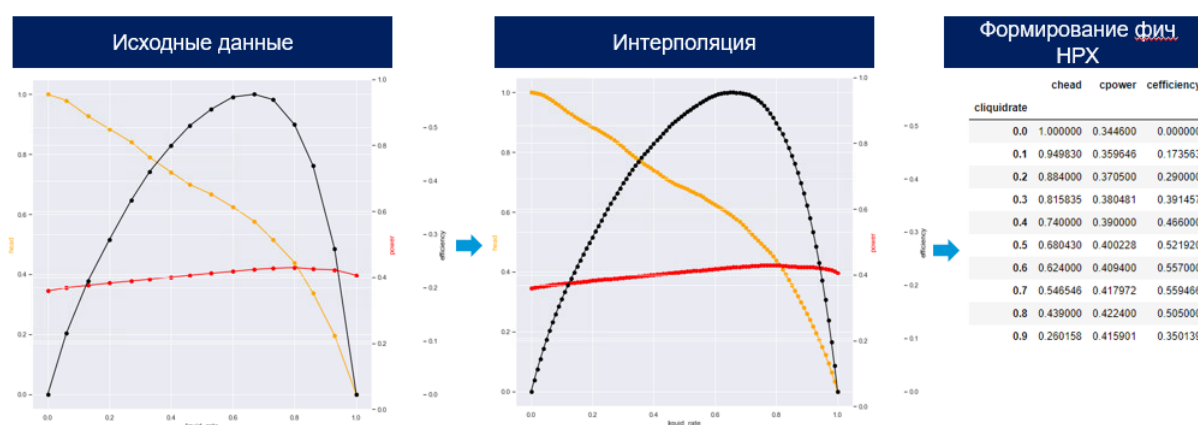


Рис.3.10. Конструирование новых признаков из НРХ.

Во время построения физической модели виртуального расходомера по напорно-расходной характеристике для расчета перепада давления в насосе требуются данные о давлении на выкиде ЭЦН. При отсутствии данных датчиков по скважине, что является распространенной практикой из-за дороговизны, как

и в модели виртуального расходомера, данные значения восстанавливались по косвенным данным и корреляциям, описанным в подпараграфе 2.1.2.

После фильтрации, обработки пропущенных значений и инженерии признаков, динамические и статистические входные данные необходимо объединить и согласовать. Итоговый набор входных признаков можно разбить на следующие группы параметров:

1. Параметры телемеханики. Для каждого из параметров за каждый день берутся минимальные, максимальные, средние, медианные значения и стандартное отклонение. А также давление на выкиде ЭЦН;
2. Конструкция скважины. Измеренная глубина скважины (MD), истинная глубина скважины (TVD), диаметр НКТ;
3. Информация о погружном оборудовании. Модель ЭЦН. Номинальная подача, максимальная подача, номинальный и максимальный напор, количество секций, количество ступеней, глубина спуска ЭЦН, номинальная частота, обработанные данные напорно-расходной характеристики;
4. PVT свойства. Плотности фаз, газосодержание, вязкость нефти, объемный коэффициент, давление насыщения.

3.3.2. Обучение модели

В качестве модели машинного обучения был выбран алгоритм LightGBM. LightGBM – библиотека с открытым исходным кодом, разработанная компанией Microsoft, которая предоставляет эффективную и действенную реализацию алгоритма градиентного бустинга.

LightGBM способен решать задачи классификации и регрессии. Основными преимуществами алгоритма являются:

- высокая скорость обучения;
- меньшее использование памяти;
- высокая точность;
- поддержка параллельного, распределенного обучения.

Стоит отметить, что LightGBM строит не симметричные деревья, как например алгоритм XGBoost. Таким образом, поддеревья, построенные с помощью LightGBM, могут иметь разную длину (рис.3.11). Такой подход к построению позволяет быстро найти зависимости в обучающей выборке, однако стоит понимать, что неконтролируемый рост дерева в глубину неизбежно может привести к переобучению

модели [16]. Поэтому, для обеспечения наилучшей обобщающей способности алгоритма необходимо настроить гиперпараметры модели.

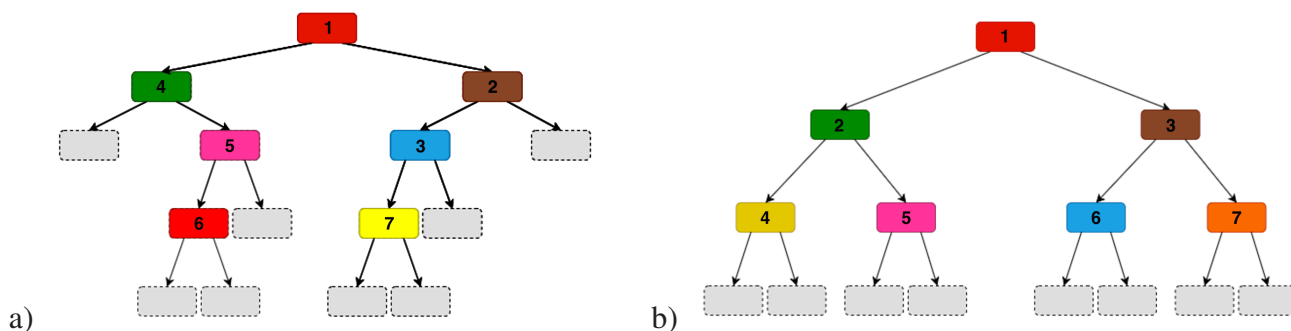


Рис.3.11. Принцип построения деревьев различных алгоритмов градиентного бустинга (*a* — принцип построения деревьев по алгоритму LightGBM; *b* — принцип построения деревьев по алгоритму XGBoost).

Наилучшие параметры искались с помощью библиотеки Sklearn и метода GridSearchCV. Для поиска оптимальных гиперпараметров GridSearchCV тестирует каждую комбинацию из сетки заданных параметров используя кросс-валидацию (параграф 2.2.3). Для каждой комбинации параметров считается метрика MAE. Модель с наивысшими метриками выбирается в качестве лучшей. Для настройки модели в качестве гиперпараметров брались следующие параметры:

- скорость обучения (learning rate);
- глубина дерева;
- количество деревьев;
- минимальное количество объектов в листе.

После выбора наилучший параметров для определения степени обученности модели и достаточности объема обучающей выборки построим кривую обучения (learning curve). Кривая обучения строится путем обучения модели на разных подмножествах данных с последующим измерением метрик на тренировочной и тестовой выборке, а также она показывает, как изменяется точность модели в зависимости от размера обучающей выборки. Learning curve позволяет оценить, достаточно ли данных для обучения модели, выбрать оптимальное их количество и определить, является ли модель переобученной или недообученной. Например, в случае недообучения модели на кривой обучения ошибка на тренировочных и тестовых данных будет высока, что говорит о том, что модель не смогла найти взаимосвязи, описывающие данные (рис.3.12а). Малая ошибка на тренировочных и большая ошибка на тестовых данных, наоборот, скажет о переобучении модели (рис.3.12с). При «идеальном» варианте кривая, характеризующая ошибку

на тестовой выборке должна сходиться к кривой, характеризующей ошибку на тренировочной выборке (рис.3.12b). Такая кривая говорит о том, что данных для обучения модели достаточно и выбранная модель хорошо настроена и подходит для решения поставленной задачи.

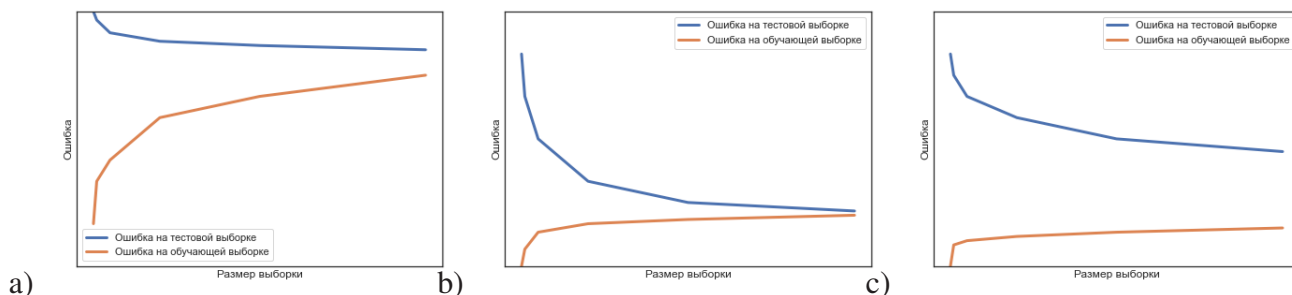


Рис.3.12. Кривые обучения (*a* — при недообучении модели; *b* — в «идеальном» случае; *c* — при переобучении модели).

На (рис.3.13) приведены кривые обучения для предложенной статистической модели. Как можно увидеть, модели достаточно данных для нахождения зависимостей в данных, при этом $R^2 = 0.7$ и $MAE = 0.16$, что является неплохим результатом и говорит о хорошей обобщающей способности алгоритма.

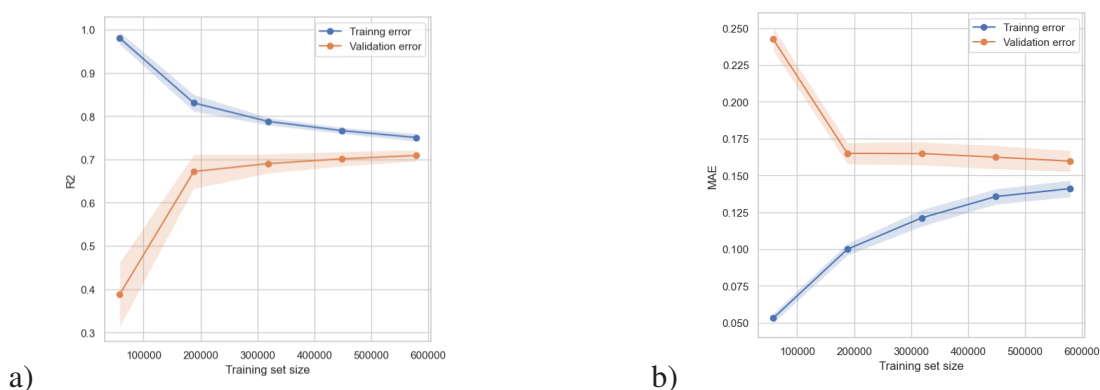


Рис.3.13. Кривые обучения предложенной статистической модели (*a* — метрика R^2 ; *b* — метрика MAE).

ГЛАВА 4. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ АЛГОРИТМА

В данной главе рассматриваются результаты применения предложенного алгоритма при различных эксплуатационных случаях в условиях месторождения X. Также в данной главе сравнивается физическая модель виртуального расходомера с предложенной гибридной моделью.

4.1. Смена способа эксплуатации

Рассмотрим случай перевода скважины с фонтанного способа эксплуатации на ЭЦН. Стоит отметить, что во время фонтанирования скважины возможно применение виртуальной расходомерии с помощью модели штуцера, описание которой можно найти в работах [8; 10—12]. Как и в виртуальной расходомерии для скважин, оборудованных УЭЦН, для расчета дебита по модели штуцера необходимо определить калибровочные коэффициенты для сопоставления модели с реальным режимом работы скважины.

При поломке замерной установки или другим причинам, по которым скважину невозможно замерить, особо актуальным является оценка дебита с помощью виртуальной расходомерии. Однако для возможности прогнозирования дебита алгоритму необходимы калибровочные коэффициенты, которые без замеров дебита являются неопределенными. Стоит отметить, что после смены способа эксплуатации на ЭЦН применение калибровочных коэффициентов модели штуцера к модели ЭЦН недопустимо. В такого рода ситуациях для оценки дебита скважины можно применить предложенный гибридный подход, который способен определять калибровочные коэффициенты без наличия замеров АГЗУ.

На примере скважины 12 рассмотрим применение гибридного алгоритма. Как можно видеть на (рис.4.1) скважина 12 сначала запускалась в режиме ПКВ, после чего была переведена в режим ПДФ. Во время рассматриваемого интервала скважина хорошо замерялась, по ней хорошо известны фактические калибровочные коэффициенты. Для тестирования алгоритма данная скважина не участвовала в обучающей выборке, алгоритм машинного обучения не знал фактические калибровочные коэффициенты для рассматриваемой скважины.

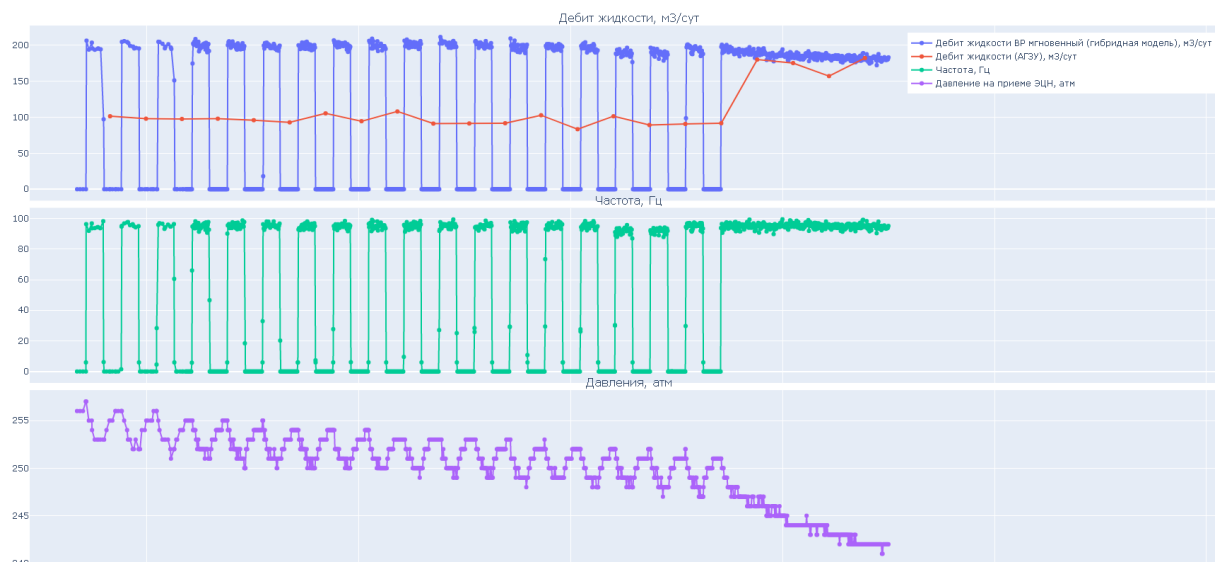


Рис.4.1. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 12.

На рис.4.1 видно, что предложенный алгоритм восстанавливает мгновенный расход через насос. Для корректного сравнения с АГЗУ, время замера которого составляет для 12 скважины 60 минут, значения виртуального расходомера интегрировались скользящим окном в 60 минут (рис.4.2). Полученный дебит имеет среднюю ошибку 7.37 м³/сут и относительную ошибку 6.09%. На рис.4.3 представлено сравнение фактические калибровочных коэффициентов с предсказанным значением (0.775).



Рис.4.2. Интегральный дебит по гибричному алгоритму виртуального расходомера для 12 скважины.



Рис.4.3. Калибровочные коэффициенты для скважины 12.

4.2. Ввод новой скважины

Рассмотрим случай ввода новой скважины. В отличие от случая со сменой способа эксплуатации, в рассматриваемом случае нет информации с каким дебитом раньше работала скважина. Известны лишь данные телеметрии, инклинометрии, данные о погружном оборудовании. По скважине 13 также известны замеры дебита и калибровочные коэффициенты, которые для чистоты эксперимента, не вносились в обучающую выборку алгоритма. Результаты восстановления дебита и полученного калибровочного коэффициента можно рассмотреть на рис.4.4, рис.4.5 (средняя ошибка гибридной модели составляет 47.53 м³/сут, а относительная ошибка составляет 9.23%).



Рис.4.4. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 13.



Рис.4.5. Калибровочные коэффициенты для 13 скважины.

4.3. Смена ЭЦН

Рассмотрим случай смены ЭЦН. Обычно смена погружного оборудования происходит после отказа скважины. После смены ЭЦН при невозможности замерить скважину с помощью замерной установки, актуальной является оценка

дебита скважины с помощью альтернативных способов, например, с помощью виртуальной расходомерии.

При невозможности замерить скважину с новым ЭЦН, не получится оценить дебит скважины с помощью виртуального расходомера. Насосы могут отличаться по напорно-расходным характеристикам, конструкции, так же стоит учитывать разные осложняющие факты, поэтому использование калибровочных коэффициентов от прошлого ЭЦН недопустимо. Для скважины 14 после смены ЭЦН известны замеры дебита и калибровочные коэффициенты. Для тестирования гибридного алгоритма эти данные не были внесены в обучающую выборку. На рис.4.6, рис.4.7 продемонстрированы результаты расчета (средняя ошибка гибридной модели составляет 6.43 м³/сут, а относительная ошибка составляет 2.43%).

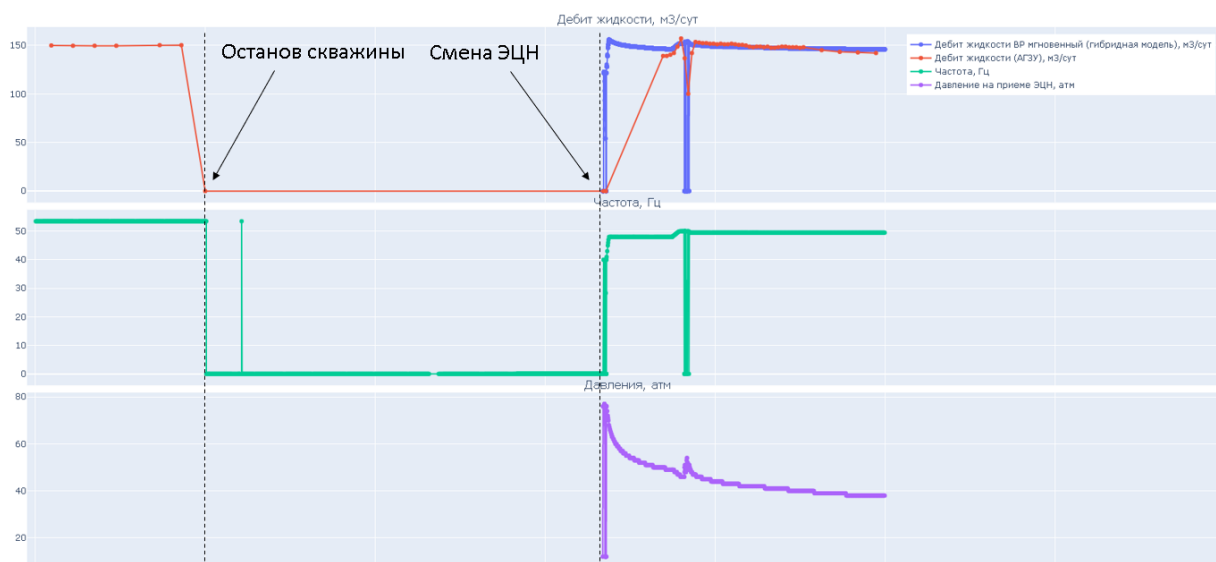


Рис.4.6. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 14.



Рис.4.7. Калибровочные коэффициенты для скважины 14.

4.4. Поломка АГЗУ

Рассмотрим поведение предложенного алгоритма при наличии некорректных замеров АГЗУ. Некорректные замеры могут быть из-за выхода из строя оборудования, засорения, влияния газа, недостаточного времени замера. Специалисту

геологу необходимо время от времени, утверждать режим работы по скважине. Хорошо помогают в этом специалистам замерные установки, однако в случае их поломки, процесс утверждения дебита становится сложной задачей. В таких ситуациях может помочь оценка дебита по виртуальному расходомеру. Однако, при наличии множества некорректных замеров дебита, алгоритм виртуального расходомера может подстроиться под эти замеры, и также показывать некорректные дебиты. В таких ситуациях может быть полезен предложенный алгоритм, так как для его расчета замеры АГЗУ не являются обязательными параметрами.



Рис.4.8. Калибровочные коэффициенты для скважины 15.



Рис.4.9. Калибровочные коэффициенты для скважины 15.

4.5. Сравнение гибридной модели с физической моделью

Предложенный гибридный подход к решению задачи виртуальной расходомерии обладает рядом преимуществ, перед классической физической моделью, для расчета которой необходимы знания о прошлом режиме работы скважины (замеры АГЗУ). Однако, для большего доверия к гибридной модели и внедрения ее в рабочие процессы необходимо протестировать и сравнить предложенный подход с физической моделью. Для проведения тестирования было выбрано 10 скважин

месторождения X с известным дебитом и калибровочными коэффициентами. Физическая модель, описанная в 3, адаптировалась на известные замеры дебита, а для предложенной гибридной модели данная группа скважин была исключена из обучающей выборки. Результаты расчета можно увидеть в табл.4.1, табл.4.2.

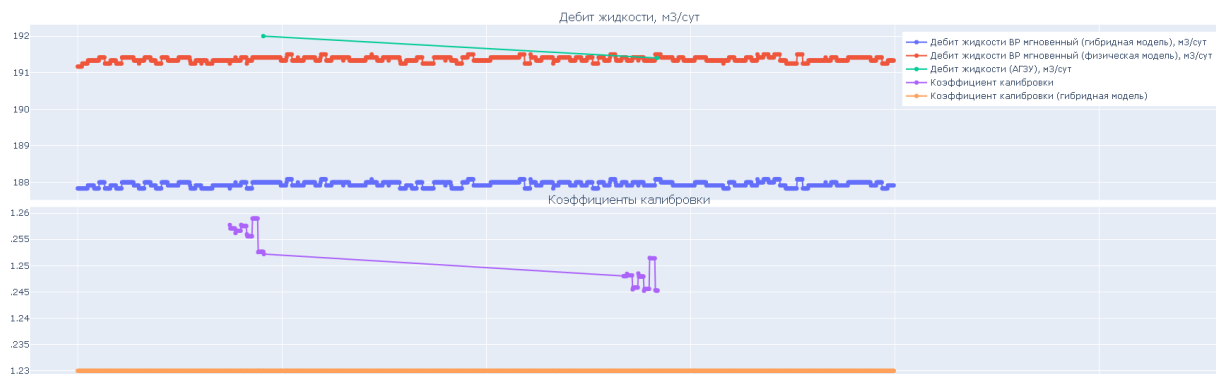


Рис.4.10. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 8.

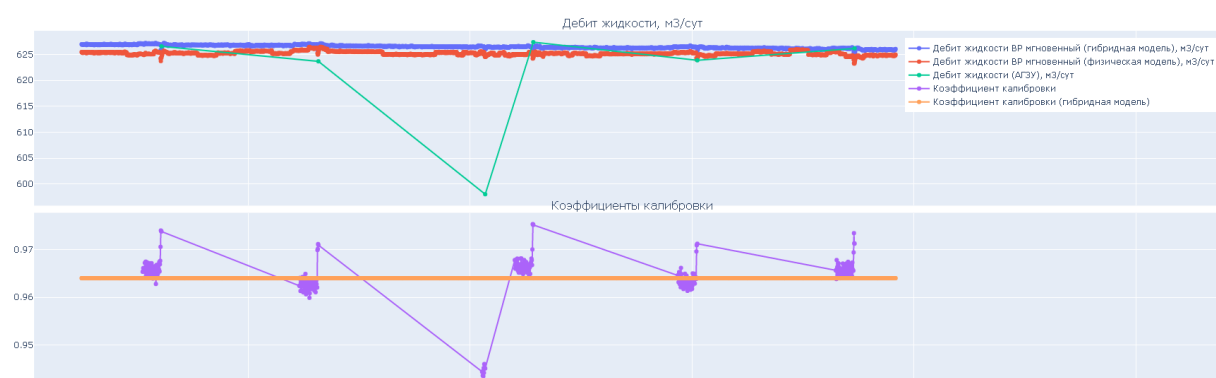


Рис.4.11. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 9.

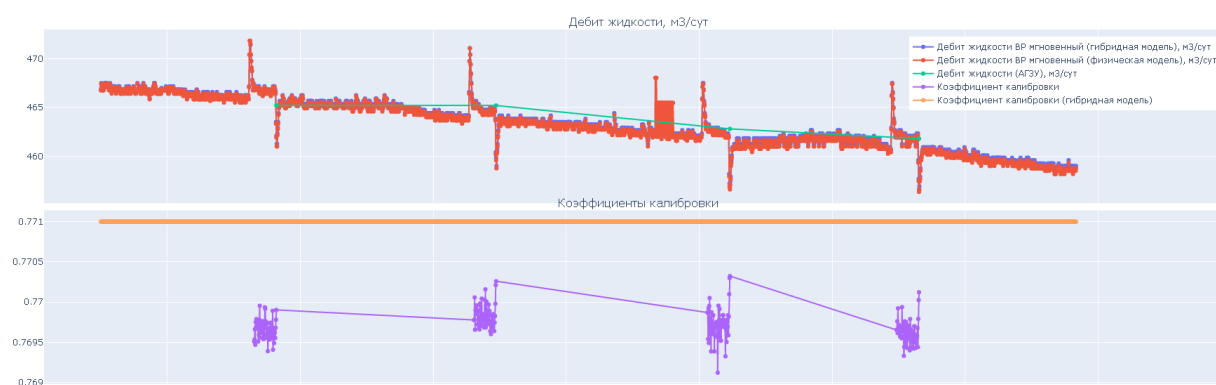


Рис.4.12. Пример восстановления дебита с помощью гибридного алгоритма виртуального расходомера для скважины 10.

Таблица 4.1

Результаты решения задачи виртуальной расходомерии с помощью физической и гибридной модели на 10 скважинах.

Номер скважины	Режимный дебит, м3/сут жидкости	Дебит ж-ти (физ. модель), м3/сут	Дебит ж-ти (гибрид), м3/сут	Средняя ошибка (гибрид), м3/сут	Относительная ошибка (гибрид), %
1	136	136	129	7	5.14
2	61.8	61.8	60.6	1.2	1.94
3	58.5	60.35	63.49	4.99	8.52
4	434	434	417.17	16.83	3.87
5	414	415.91	390.83	23.17	5.59
6	30.7	30.95	37.25	6.55	21.33
7	230	230.1	218.3	11.7	5.08
8	191.4	191.5	188.4	3	1.56
9	625.1	624.9	626.3	1.2	0.19
10	462.4	461.9	462	0.4	0.08

Таблица 4.2

Сравнение предсказанных калибровочных коэффициентов с фактическими на 10 скважинах.

Номер скважины	Калибр-й коэф-т	Предсказанный калибр-й коэф-т	Абсолютная ошибка	Относительная ошибка, %	Относительная ошибка по дебиту ж-ти (гибрид), %
1	1.29	1.223	0.067	5.19	5.14
2	1.1	1.05	0.05	4.54	1.94
3	0.93	1.04	0.11	11.82	8.52
4	0.97	0.92	0.05	5.15	3.87
5	1.13	1.06	0.07	6.19	5.59
6	0.92	1.09	0.17	18.47	21.33
7	1.18	1.1	0.08	6.77	5.08
8	1.25	1.23	0.02	1.6	1.56
9	0.965	0.964	0.001	0.1	0.19
10	0.77	0.771	0.001	0.12	0.08

4.6. Выводы

Таким образом, предложенная гибридная модель продемонстрировала хорошие результаты и показала, что она может дополнять физическую модель виртуального расходомера в случаях, когда ее расчет не определен или является некорректным (при смене способа эксплуатации рис.4.1, при вводе новой скважины рис.4.5, при смене ЭЦН рис.4.7, при неисправности АГЗУ рис.4.9). Виртуальный расходомер в описанных случаях может помочь понять с каким дебитом работает скважина, в какой зоне НРХ работает ЭЦН, что позволит вовремя скорректировать режим работы скважины для обеспечения максимизации добычи нефти. Также стоит отметить, что гибридный подход может позволить валидировать работу замерных установок, что позволит специалистам по добыче нефти корректно утверждать режим работы скважины и в целом максимально эффективно управлять нефтяным фондом.

Было проведено сравнение гибридного и физического подхода на 10 скважинах месторождения X, которые хорошо замерялись и имели стабильный режим работы. Сравнение показало (табл.4.1, табл.4.2), что предложенная гибридная модель может использоваться совместно с физической моделью с последующей заменой из-за большего охвата различных эксплуатационных ситуаций. В среднем гибридный подход ведет себя хуже физического. Данный факт связан с тем, что физический подход напрямую адаптирует модель по фактическим замерам, а гибридный подход на фактические замеры не ориентируется и использует для определения калибровочных коэффициентов модель машинного обучения. Поэтому, такое отклонение гибридной модели можно связать с проблемами в обработке данных, фильтрации данных и обучении модели машинного обучения, на которые необходимо обратить внимание в дальнейшей работах по улучшению алгоритма.

ЗАКЛЮЧЕНИЕ

В данной работе был рассмотрен гибридный подход к решению задачи виртуальной расходомерии для скважин, оборудованных ЭЦН. В качестве физической составляющей гибридной модели был рассмотрен подход, который совмещает модель PVT, модель ЭЦН и модель течения жидкости в НКТ, для восстановления дебита используя метод узлового анализа. В качестве статистической составляющей была рассмотрена модель машинного обучения, которая позволяет по косвенным параметрам предсказывать калибровочные коэффициенты, необходимые для расчета виртуального расходомера.

Предложенная гибридная модель дополняет физическую модель в тех случаях, когда расчет физической модели не определен или является некорректным:

- при некорректных замерах АГЗУ;
- при вводе новой скважины;
- при смене ЭЦН;
- при смене способа эксплуатации;
- при отсутствии возможности замерить скважину в связи с поломкой АГЗУ.

Рассмотренный гибридный алгоритм был протестирован на реальных промысловых данных в случаях, при которых расчет по физической модели был не определен. Также было проведено сравнение гибридного подхода и физической модели (4.5). Анализ результатов показал, что алгоритм способен восстанавливать дебит по косвенным параметрам с приемлемой погрешностью и потенциально способен помогать специалистам с нефтепромысла надежно и своевременно контролировать процесс добычи, проводить обратное распределение добычи с конечного узла сбора, а также снизить нагрузку с замерных установок.

В качестве дальнейших путей развития и улучшения физической составляющей алгоритма можно отметить улучшение фильтрации данных, а также настройка одновременного расчета физической модели через перепад давлений через насос и физической модели на основе потребляемой мощности, рассмотренной в работе [7], с последующим выбором лучшей модели на основе текущей и исторической сходимости к замерам АГЗУ.

В качестве дальнейшего улучшения физической составляющей алгоритма стоит отметить улучшение фильтрации данных, работы с пропущенными значениями, создание новых признаков для улучшения обобщающей способности алгоритма. Также в качестве развития модели можно рассмотреть создание моде-

ли не только под условия конкретного добывающего общества с определенными эксплуатационными условиями, а попробовать использовать данные для обучения с разных добывающих обществ. Такой подход, потенциально, может увеличить сходимость для скважин из новых добывающих обществ, данных у которых может быть недостаточно для обучения и корректной работы модели.

Стоит отметить чувствительность алгоритма к дискретности данных, например для корректного расчета ПКВ скважин необходимо разграничить циклы работы и накопления, найти рабочую долю времени, что требует хорошей дискретности данных (рис.3.3). В будущих работах стоит обратить внимание на то, какие группы параметров телеметрии оказывают наибольшее влияние на улучшение сходимости виртуального расходомера. Например, вкладываться в хорошие многофазные замерные установки, в датчики давления на выкиде для определения перепада давлений через насос, или в улучшение датчиков потребляемой мощности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Брилл Д., Мукерджи Х.* Многофазный поток в скважинах. — 2006.
2. *Мищенко И. Т.* Скважинная добыча нефти. — Нефть и газ, 2003.
3. *Чернова Н. И.* Математическая статистика. — Новосибирский гос. ун-т, 2007.
4. *Géron A.* Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. — " O'Reilly Media, Inc.", 2022.
5. *Sjardin B., Massaron L., Boschetti A.* Large scale machine learning with Python. — Packt Publishing Ltd, 2016.
6. *Takacs G.* Electrical submersible pumps manual: design, operations, and maintenance. — Gulf professional publishing, 2017.
7. *Андрианова А., Логинов А., Хабибуллин Р., Кобзарь О.* Виртуальная расходомерия как инструмент мониторинга эффективности работы скважины с УЭЦН // ПРОНЕФТЬ. Профессионально о нефти. — 2022. — № 4. — С. 75—81.
8. *Полинов М. А.* Прогнозирование дебита фонтанирующих скважин на основе расчёта движения потока через штуцер: выпускная квалификационная работа магистра: направление 01.04. 03 «Механика и математическое моделирование»; образовательная программа 01.04. 03_04 «Математическое моделирование процессов нефтегазодобычи». — 2022.
9. *Шабонас А., Горидько К.* Обзор подходов к реализации алгоритма виртуального расходомера на скважинах, оборудованных установками электроцентробежного насоса // Нефтепромышленное дело. — 2022. — Т. 637, № 1. — С. 33—41.
10. *Bikmukhametov T., Jäschke J.* First principles and machine learning virtual flow metering: a literature review // Journal of Petroleum Science and Engineering. — 2020. — Vol. 184. — P. 106487.
11. *Hotvedt M., Grimstad B., Imsland L.* Developing a hybrid data-driven, mechanistic virtual flow meter—a case study // IFAC-PapersOnLine. — 2020. — Vol. 53, no. 2. — P. 11692–11697.
12. *Hotvedt M., Grimstad B., Ljungquist D., Imsland L.* On gray-box modeling for virtual flow metering // Control Engineering Practice. — 2022. — Vol. 118. — P. 104974.
13. *AL-Qutami T. A., Ishak M. A., Wollebaek L., W Ahmad W. A.* Combining Physics and Machine Learning for Multimodal Virtual Flow Metering with Confidence. — 2022.

14. *Vinogradov D., Vorobev D.* Virtual flowmetering novuport field examples. — 2020.
15. *Wang G., Wang Z., Song L.* Uncertainty analysis for different virtual pump water flow meters // *Science and Technology for the Built Environment*. — 2019. — Vol. 25, no. 3. — P. 297–308.
16. Учебник по машинному обучению. — URL: <https://academy.yandex.ru/handbook/ml>.