

Использование алгоритмов оптимизации и машинного обучения в задачах добычи нефти и газа

Кобзарь Олег
Газпромнефть Цифровые решения



ПЛАН ПРЕЗЕНТАЦИИ

- Введение
- Физическое моделирование. Решение оптимизационных задач для обеспечения эффективности процесса нефтедобычи
- Машинное обучение. Использование статистического подхода.
- Гибридное моделирование. Способы объединения физического моделирования и методов машинного обучения. Развитие методов.
- Заключение





Прирост запасов

- Определение нефтенасыщенных пропластков по РИГИС
- Открытие нового месторождения



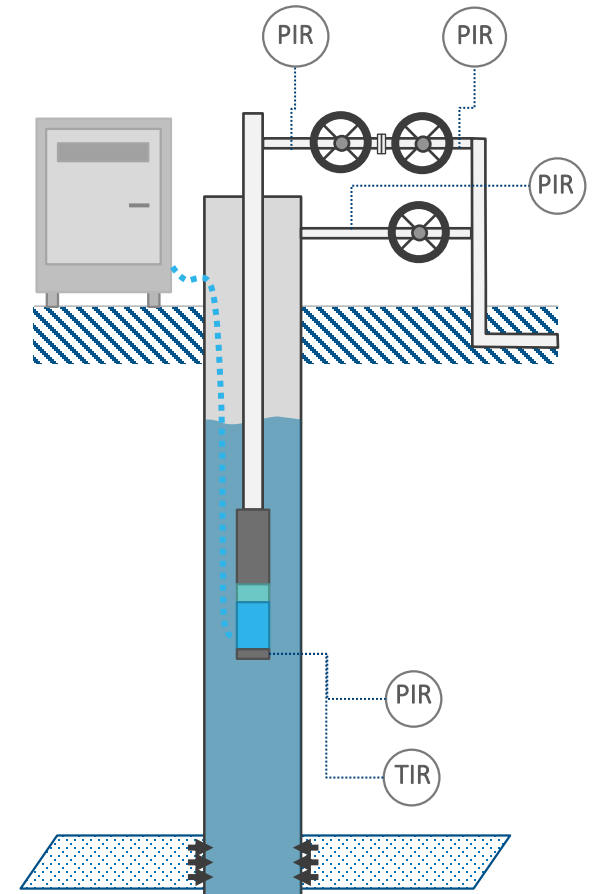
Увеличение добычи

- Бурение скважин
- Перевод на механизированные способы эксплуатации с увеличением добычи
- Максимизация добычи на текущих мощностях



Снижения затрат

- Оптимизация работы скважины
- Групповая оптимизация работы скважин





Физические модели

Численное моделирование физических процессов

Плюсы: физически обоснованное моделирование многофазного процесса

Минусы: модели сложные с точки зрения разработки и требовательные к качеству входных данных

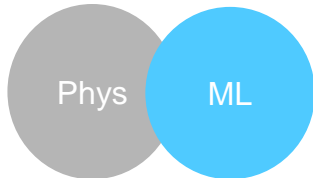


Методы машинного обучения

Использование моделей машинного обучения

Плюсы: статистические модели могут выявлять корреляции между измеренными параметрами, отсутствие физических ограничений на модель

Минусы: модель не описывает реальное физическое поведение системы



Гибридные модели

Использует физические модели с параметрами, откалиброванными по историческим данным на основе методов машинного обучения

Плюсы: модели обеспечивают компромисс между точностью результатов и сложностью моделирования

Минусы: отсутствие унифицированного подхода, модель для каждого объект (месторождения) должна быть построена с учетом его особенностей

Задачи

- Оптимизация расхода электроэнергии
- Максимизация добычи нефти
- Групповая оптимизация скважин

Физическое моделирование.

Физическое моделирование.

Решение оптимизационных задач для обеспечения эффективности процесса нефтедобычи



Основные темы

- Схема подхода по составлению решения
- Определения однозначности решения
- Методы оптимизации
- Узловой анализ

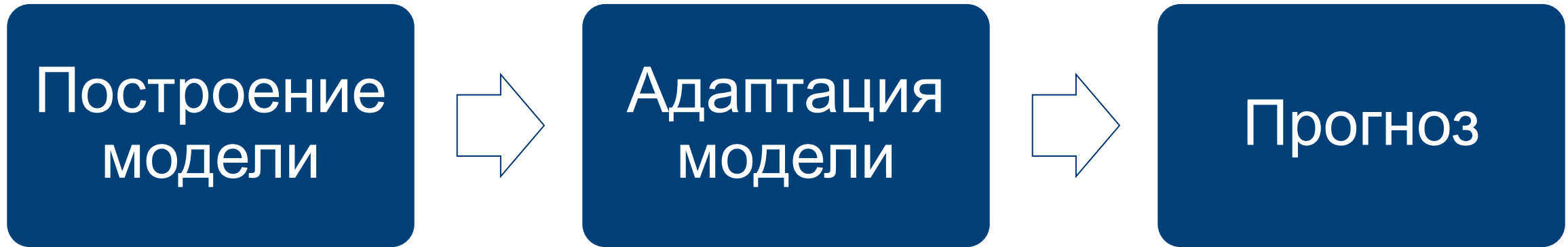


Результат

- Увеличение добычи
- Снижение затрат
- Прочие косвенные эффекты

Классический подход

«Все модели неверны, но некоторые из них полезны»



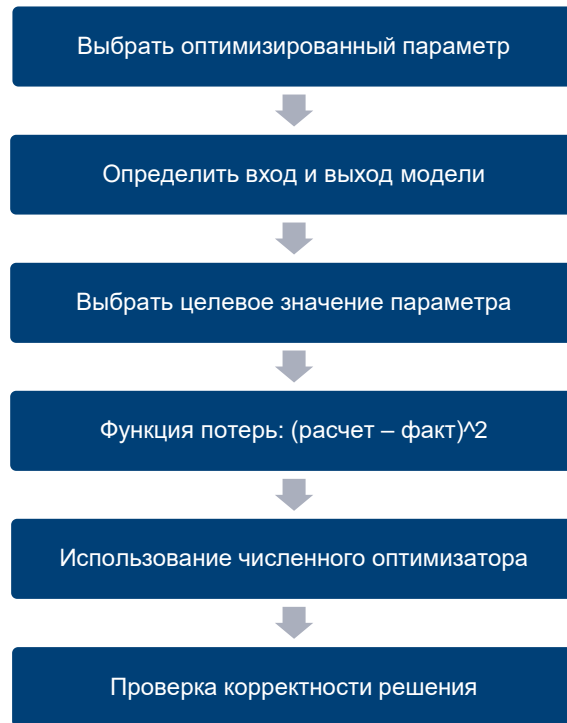
- У каждой физической модели есть свои пределы применимости
- Как правило, чем проще модель, тем она лучше
- Задача, данные и ресурсы определяют выбор модели

Точность прогноза зависит от качества модели, точности ее адаптации, метода прогноза

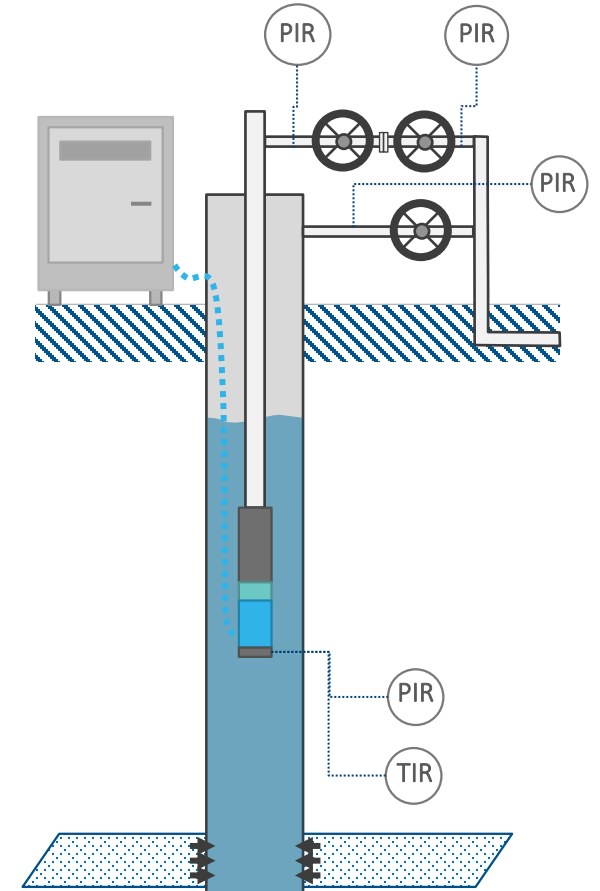
Постановка задачи

Основная задача, составление функции потерь

Пример: адаптация модели скважины на реальные данные



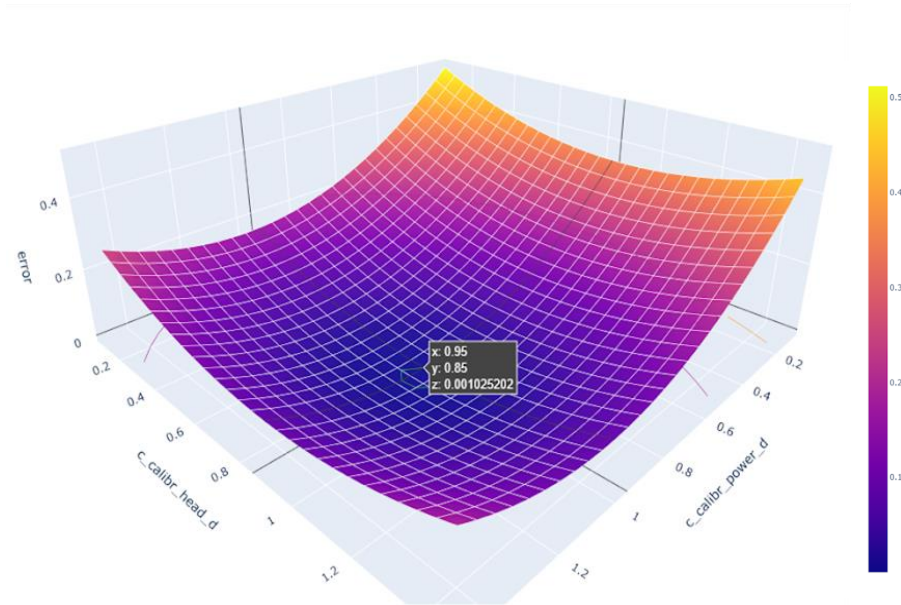
- Найти коэффициенты деградации ЭЦН
- Фактические данные (вход модели): замер дебита, давления в основных узлах системы
- Выход модели: буферное давление
- Целевое значение: фактический замер
- Полученные значения должны принадлежать диапазону $[0.5, 1.5]$



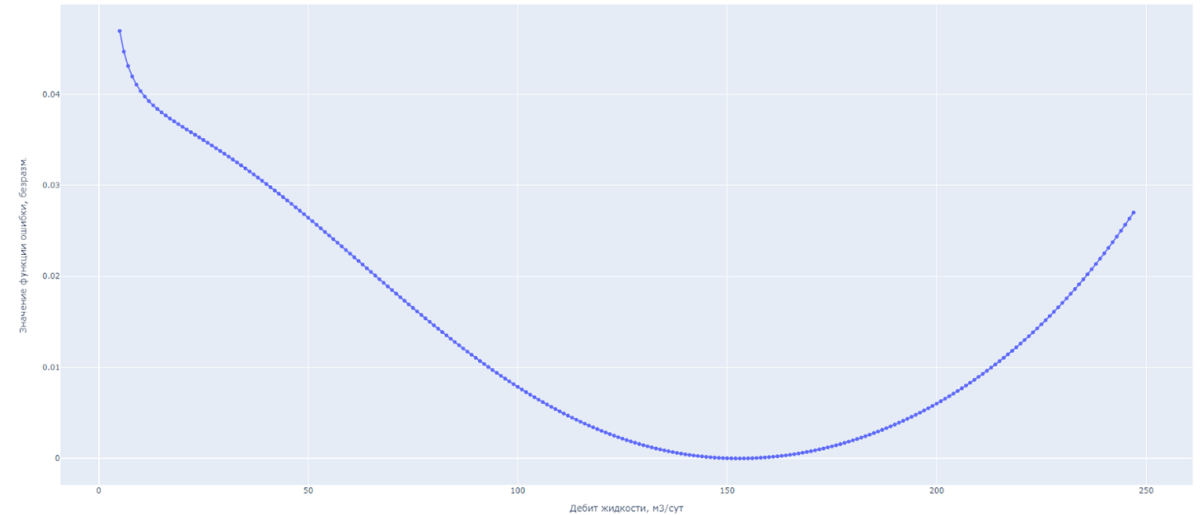
Необходима проверка на однозначность функции

Решение оптимизационной задачи

Адаптация модели скважины и определение дебитов



Калибровочные коэффициенты для ЭЦН по напору и мощности однозначно описывают систему



Единственное решение в задаче определения дебита скважины

$$Error = w_h \left(\frac{P^{calc} - P^{fact}}{P_{max}} \right)^2 + (1 - w_h) \left(\frac{N^{calc} - N^{fact}}{N_{max}} \right)^2 \Rightarrow \min$$

ТР 2.0: Оптимизация режимов работы фонда

ТР 2.0: Блок моделирования



Блок массовой работы с моделями скважин



Блок инд. работы с моделями скважин



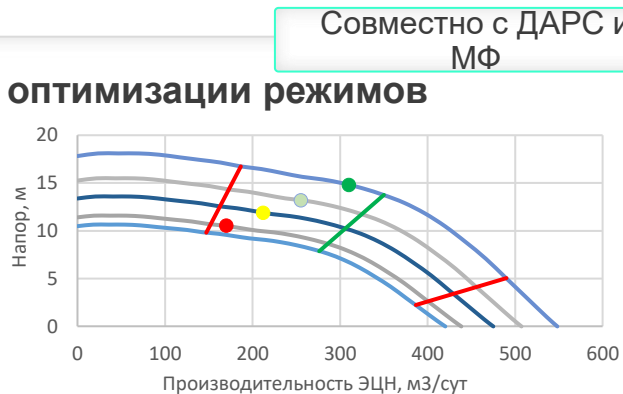
Модель скважины ТР 2.0

ТР2.0: Блок оптимизации режимов работы скважин



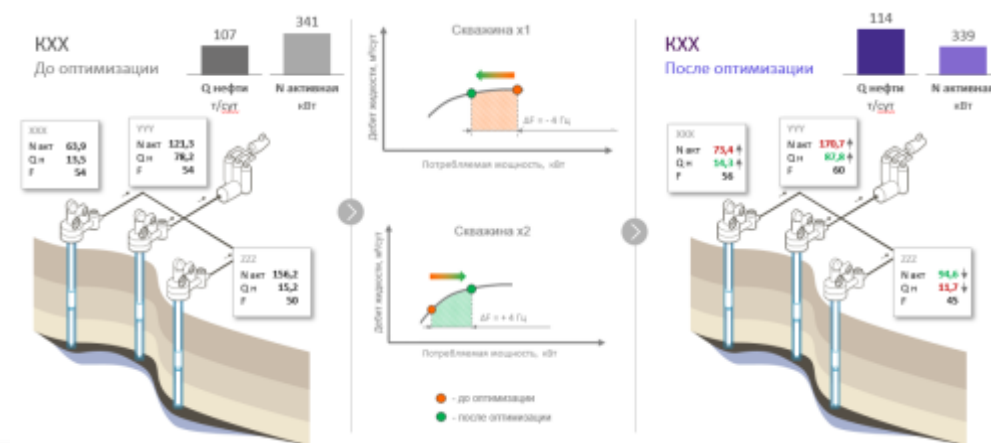
Индивидуальные оптимизации режимов

- Оптимизация:
 - F
 - dшт
 - циклов ПКВ/АПВ/ЧЧ
 - уставок ПИД
 - Qг/г
- Учёт тех. и гео. ограничений, истории работы скважины
- Экспертиза и мониторинг



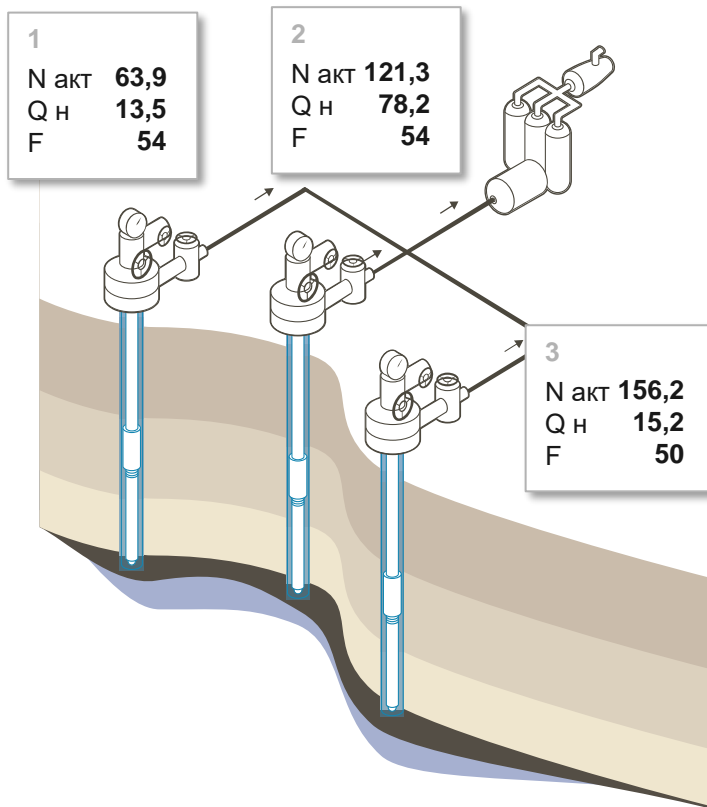

Групповые оптимизации режимов

- Оптимальное управление фондом в целом
- Экспертиза, мониторинг и ежесуточная актуализация программ групповой оптимизации

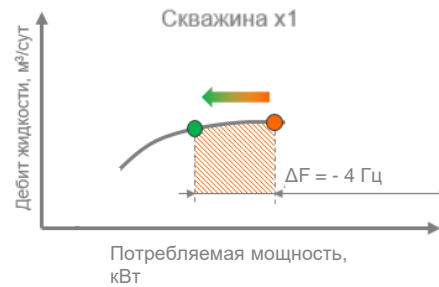


ТР 2.0: Оптимизация режимов работы фонда УЭЦН

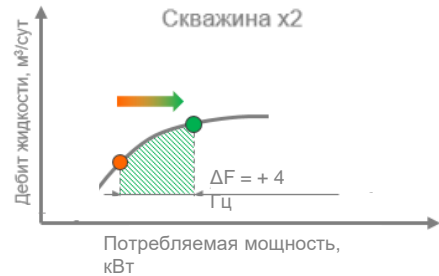
До оптимизации



Групповая оптимизация технологического режима скважин с УЭЦН



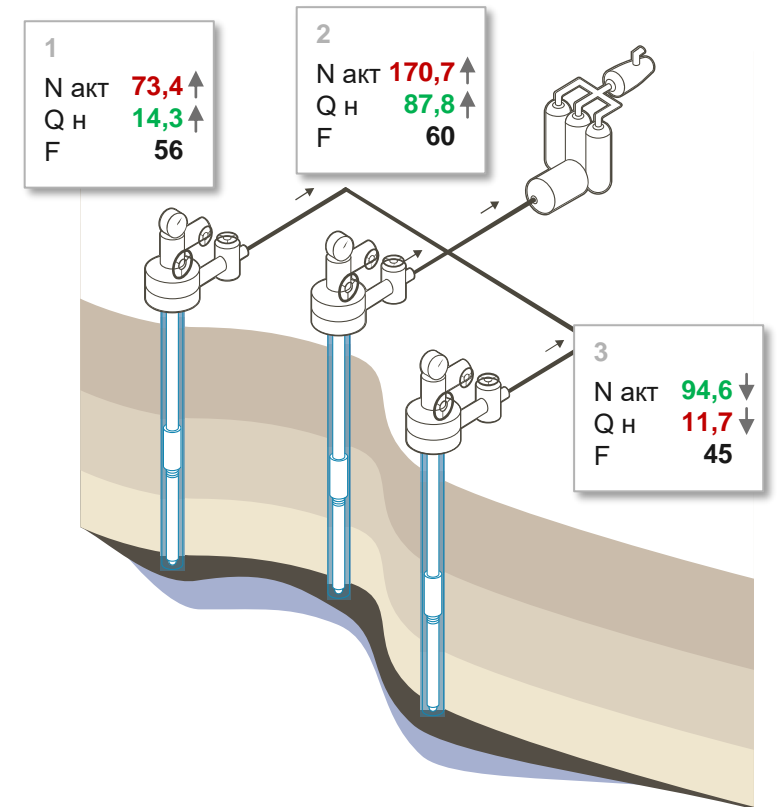
Скв №1:
Уменьшаем частоту – $Q_{\text{н}}$ меняется слабо



Скв №2:
Увеличиваем частоту – $Q_{\text{н}}$ существенно возрастает

- - до оптимизации
- - после оптимизации

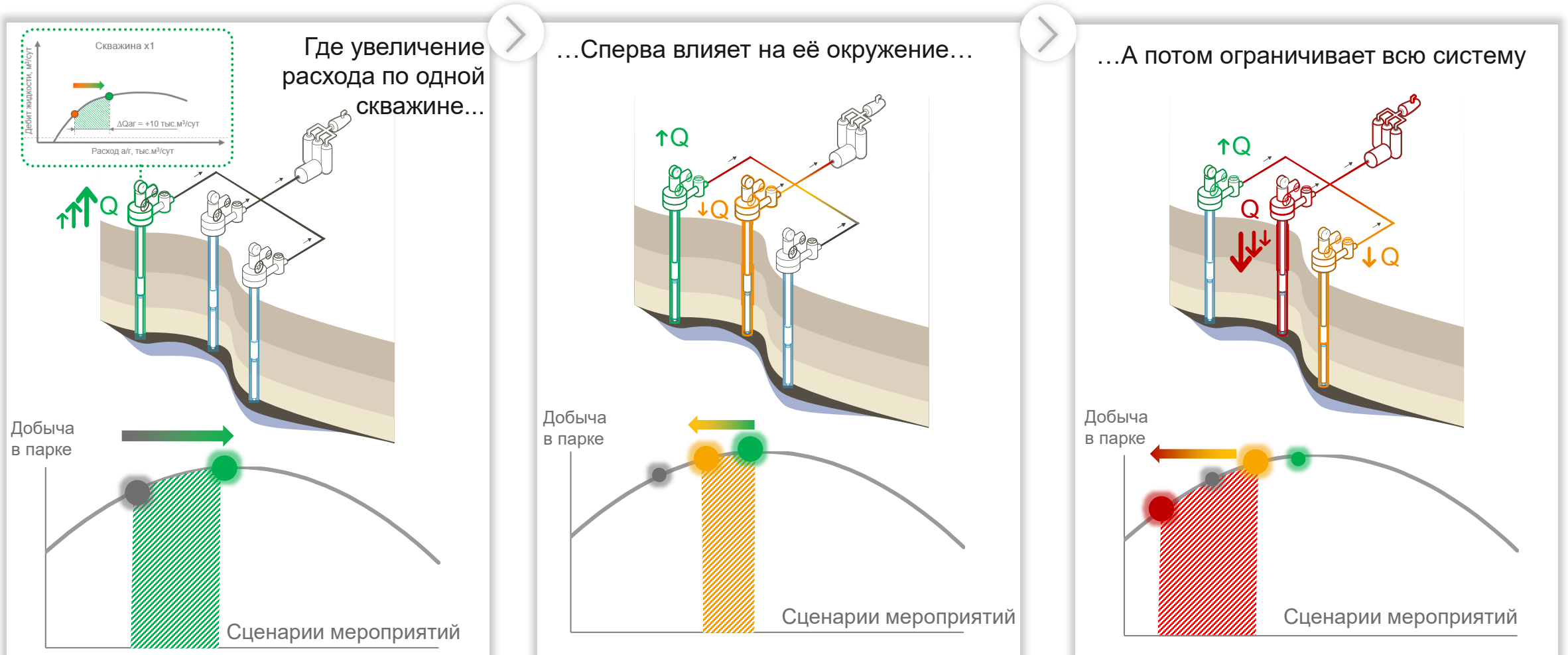
После оптимизации



$N_{\text{акт}}$ – активная потребляемая мощность ЭЦН, кВт $Q_{\text{н}}$ – дебит нефти, т/сут F – частота ЭЦН, Гц

TR2.0: кейсы групповой оптимизации режимов работы газлифта

Разработка функционала групповой оптимизации берёт своё начало с газлифтного фонда



● Исходное значение

● Значение непосредственно после реализации мероприятия

● Значение после отклика ближайшего окружения

● Значение после отклика всей системы

Задачи

- Верификация промысловых данных
- Определение аномальных циклов ПКВ

Машинное обучение.

Использование статистического подхода.

Преимущества и недостатки



Основные темы

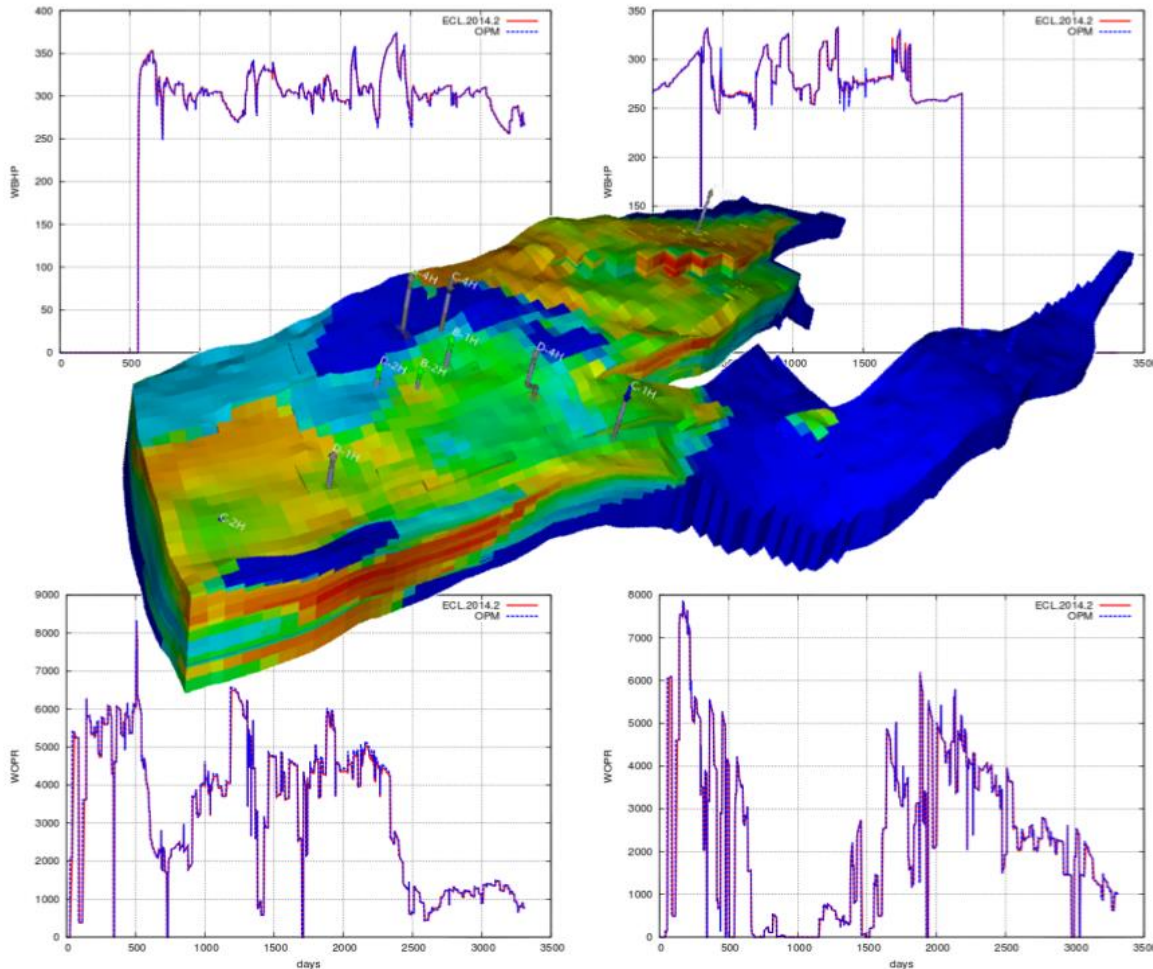
- Корректность постановки задачи
- Используемые модели
- Способы проверки решений
- Подготовка данных



Результат

- Снижение затрат
- Прочие косвенные эффекты

Косвенные эффекты



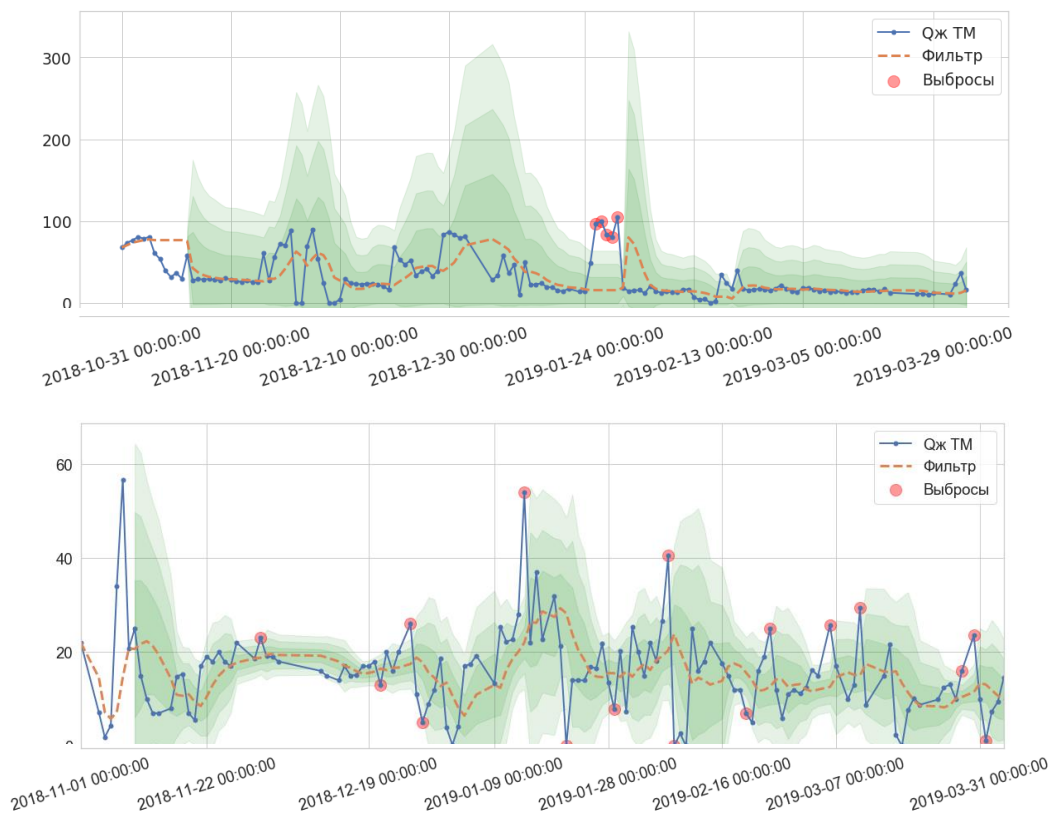
- Гидродинамические модели
 - Секторные (часть месторождения)
 - Цифровые двойники (все месторождения)
- Требуют большого количества данных и времени расчета
- Подготовленные и верифицированные данные позволят сделать прогноз более точным



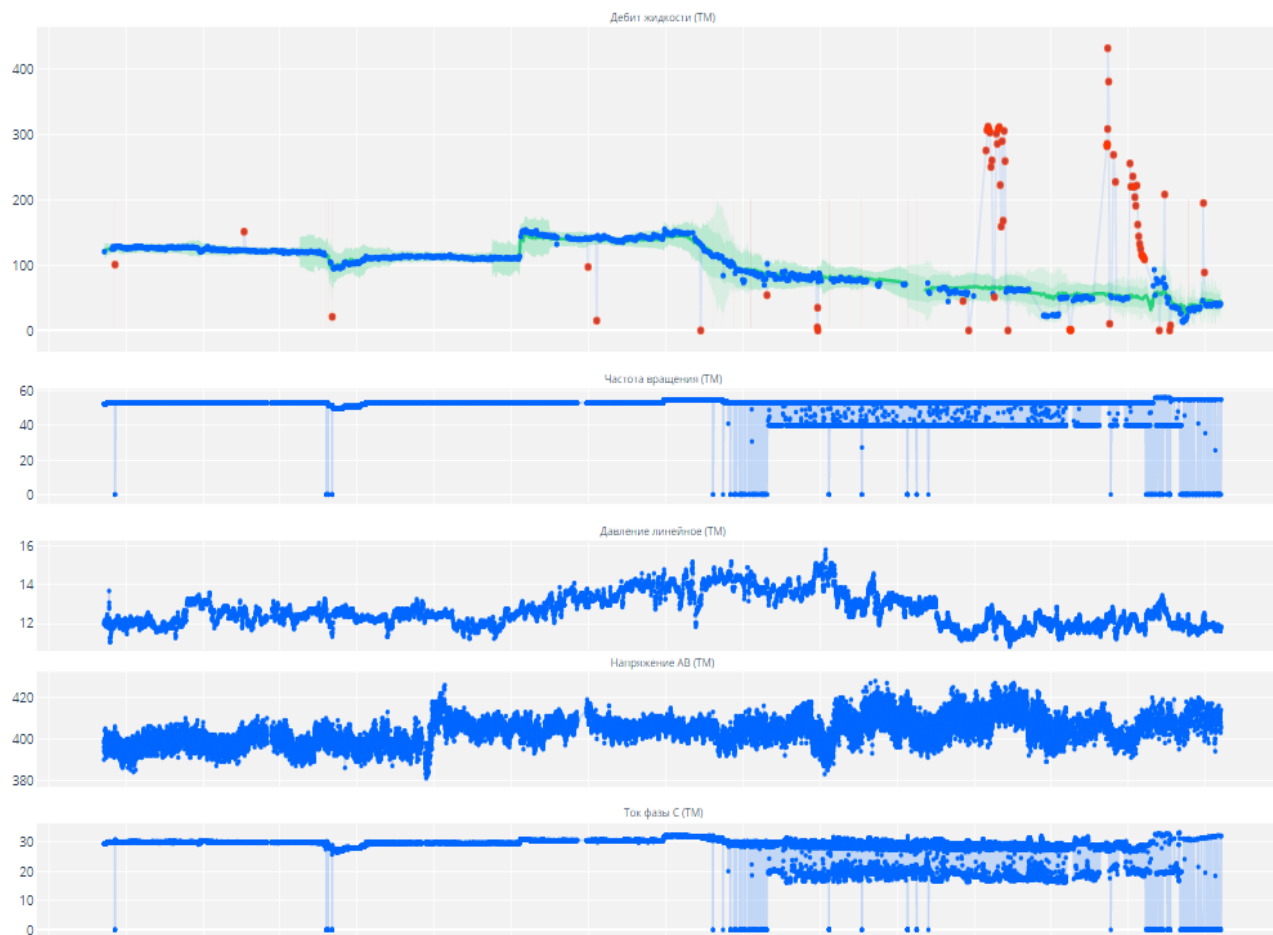
Эффекты от валидации данных сложно измерить, но качество данных – основа успешного прогноза для всех типов моделей

Преимущества реализованного подхода

Фильтрация тривиальным фильтром

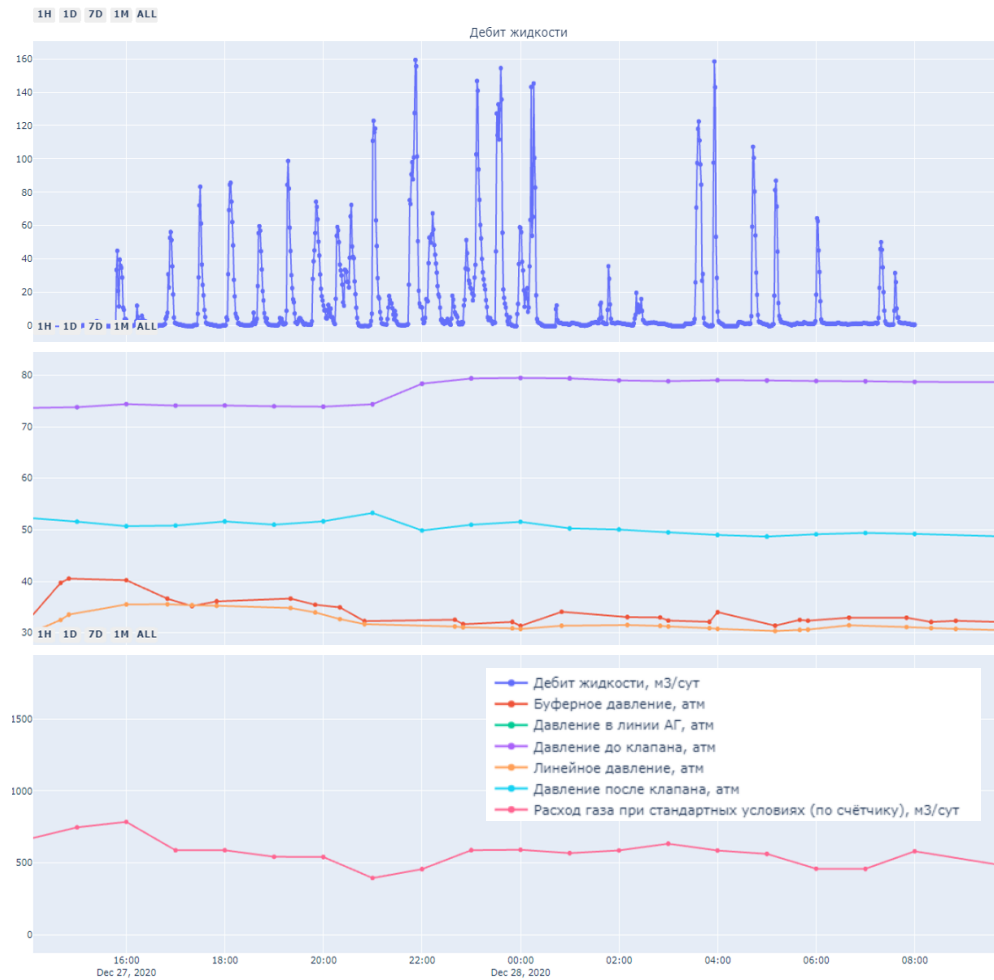


Фильтрация на основе автоэнкодера



Подготовка данных

Низкая дискретность данных



Переменная дискретность



Алгоритмы расчета и источники данных для Виртуального расходомера

Замер и регулировка ключевых параметров СУ:

- Определение потребляемой УЭЦН мощности
- Обновление функции ошибки по мощности

Кабель: передача электроэнергии от СУ до ПЭД

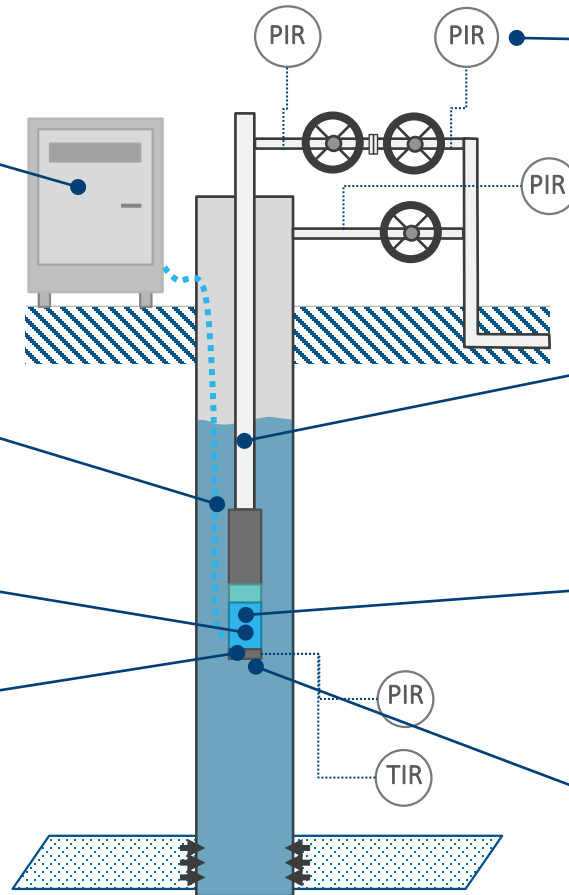
- Расчет энергетических потерь при передаче электроэнергии

ПЭД: передача мощности валу ЭЦН

- Учёт КПД погружного двигателя

Датчики телеметрии ЭЦН

- Корректировка PVT-свойств в условиях приема насоса



Датчики давления и температуры

- Расчет перепада давления в штуцере по корреляции Перкинса
- Обновление функции ошибки по буферному давлению

НКТ: транспорт НГЖС на поверхность

- Расчет распределения давления и температуры в стволе скважины с использованием корреляций (Ансари, Бегз-Брилл)
- Определение буферного давления ЭЦН: передача энергии потоку НГЖС

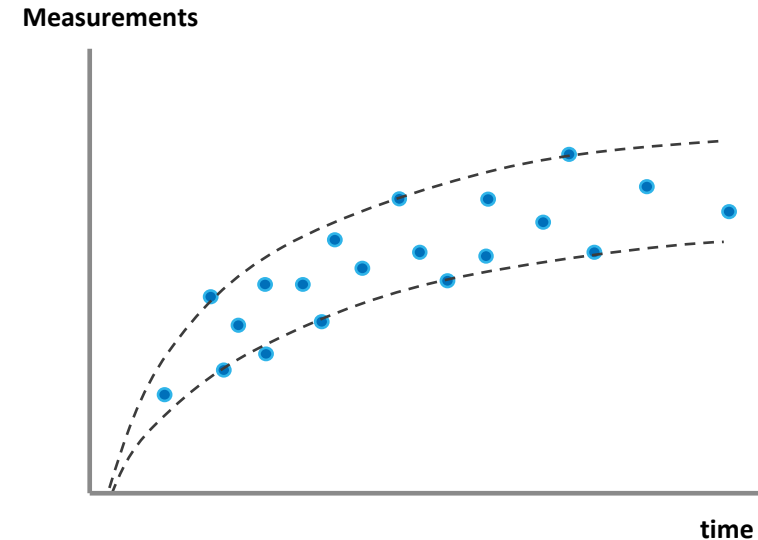
- Расчет ЭЦН по ступеням с учетом паспортных характеристик
- Поправки для расчетов: частота вращения, вязкость, плотность жидкости

ГС: частичное удаление свободного газа на приеме ЭЦН

- Расчет полной сепарации с учетом естественной сепарации и коэффициента сепарации ГС (корреляция Маркеза)

Обзор применяемых моделей машинного обучения

- Для задачи регрессии (виртуальной расходомерии)
 - Линейные модели
 - Приемлемая точность и скорость расчета
 - Деревья решений и градиентный бустинг на решающих деревьях
 - Слабая способность к экстраполяции
 - Авторегрессионные модели
 - Не учитывают влияние управляющих параметров (нет контроля)
 - Неглубокие нейронные сети
 - Превращаются в линейные модели
 - Глубокие полносвязные нейронные сети
 - Требование к большому количеству данных
 - Рекуррентные нейронные сети
 - Высокие требования к данным и ресурсам
 - Трансформеры (автокодировщик)
 - Высокие требования к ресурсам
 - Линейная интерполяция
 - Хорошая базовая модель



Выбор сделать непросто

Фильтрация данных

Предобработка данных

Цель: подготовить данные для VFM

Фильтрация + Сглаживание

Взвешенный медианный фильтр

- ✗ Не учитывает многомерность задачи
- ✗ Не учитывает физические закономерности
- ✗ Не учитывает неоднородность измерений по времени
- ✗ Не позволяет заполнять пропуски
- ✗ Неоднозначные краевые эффекты

Заполнение пропусков

Линейная интерполяция

- ✗ Не учитывает многомерность задачи
- ✗ Не учитывает физические закономерности
- ✓ Учитывает неоднородность измерений по времени
- ✗ Неоднозначные краевые эффекты

Виртуальная расходомерия

Цель: создание априорной информации для байесовской модели

Физическая модель

- ✓ Учитывает многомерность задачи
- ✓ Учитывает физические закономерности
- ✓ Учитывает неоднородность измерений по времени
- ✗ Не учитывает временные зависимости
- ✓ Позволяет заполнять пропуски в дебите
- ✗ Не позволяет заполнять пропуски в других признаках
- ✗ Не позволяет вычислить степень уверенности в предсказании
- ✗ Не работает с выбросами и шумными данными
- ✗ Неоднозначные краевые эффекты
- ✗ «Неидеальные» физические законы

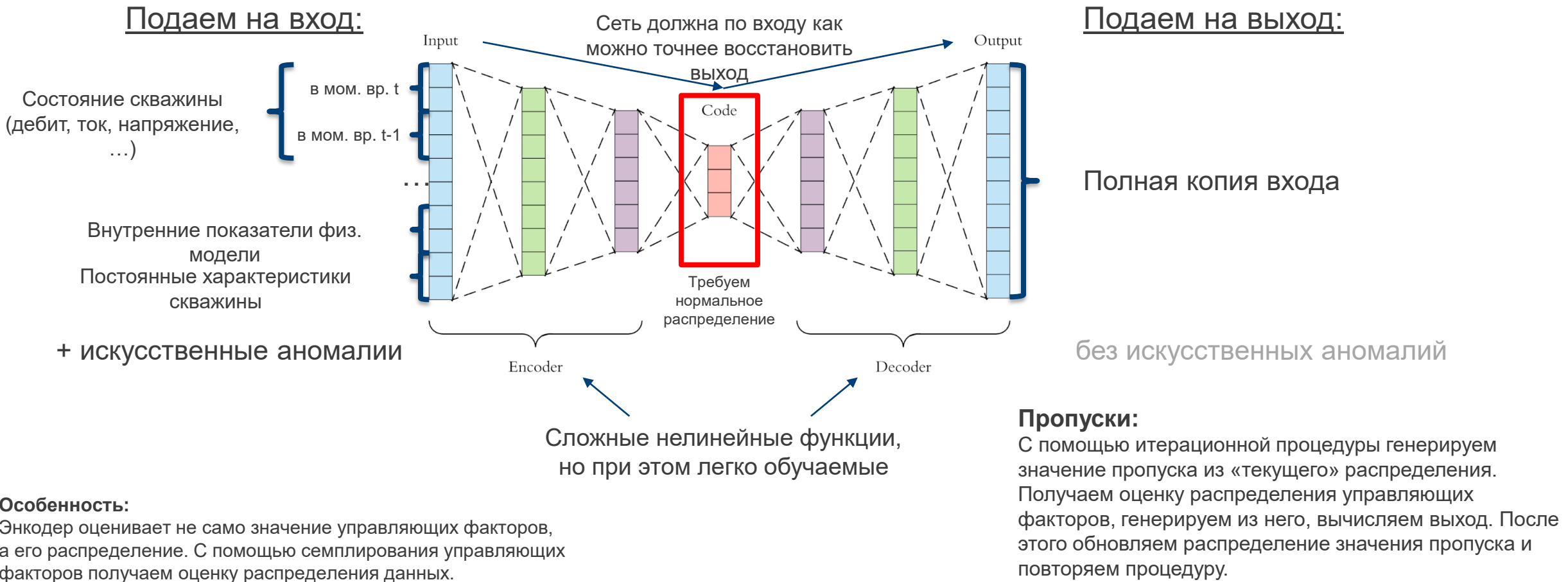
Байес & Нейронные сети

Результат:

- Уточнение прогнозов физической модели
 - Решение всего комплекса задач
-
- ✓ Учитывает многомерность задачи
 - ✓ Учитывает физические закономерности, поправленные на «неидеальность»
 - ✓ Учитывает неоднородность измерений по времени
 - ✓ Учитывает временные зависимости
 - ✓ Позволяет заполнять пропуски в любых признаках
 - ✓ Позволяет вычислить степень уверенности в предсказании
 - ✓ Работает с выбросами и шумными данными
 - ✓ Оффлайн и онлайн (на 1 шаг)
 - ? Онлайн работа на несколько шагов
 - ? Краевые эффекты
 - ? Возможность согласованной фильтрации

Модель фильтрации на базе вариационного автоэнкодера

Идея: все процессы на скважине управляются небольшим количеством **управляющих факторов**. Эти факторы могут быть полностью абстрактными и не иметь человеческой или физической интерпретации.



Особенность:

Энкодер оценивает не само значение управляющих факторов, а его распределение. С помощью семплирования управляющих факторов получаем оценку распределения данных.

Верификация данных методом вариационного автоэнкодера на основе значений нескольких признаков одновременно

В каждый целевой момент времени на вход в автоэнкодер подается:

- Средние значения замеров дебита и остальных признаков за несколько интервалов времени.
- Специальная метка, если в интервале не было замеров.

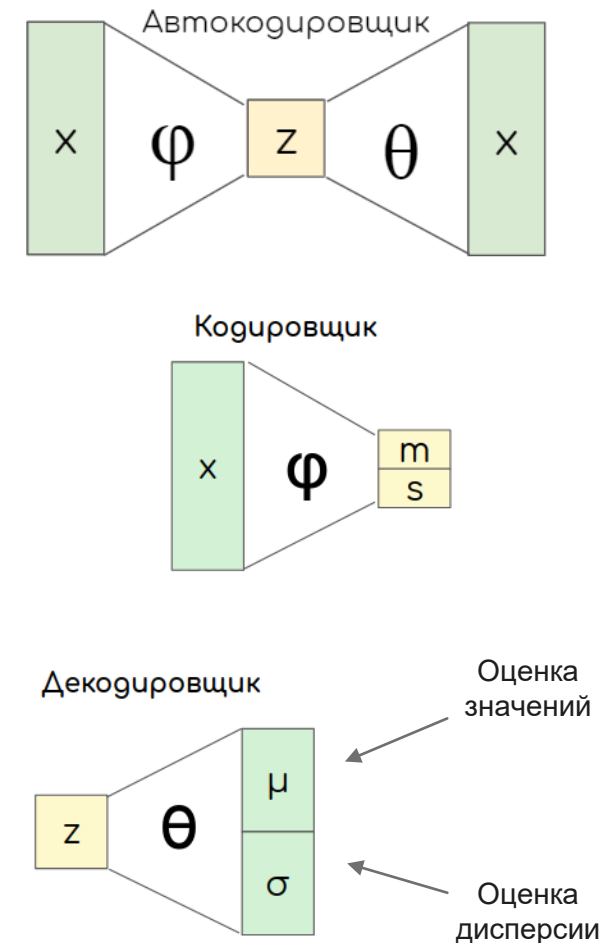
Автоэнкодер – композиция двух нейронных сетей:

- Кодировщик сжимает данные, пытаясь выделить в них основные зависимости.
- Декодировщик по сжатым данным пытается как можно точнее восстановить исходные данные.

Вариационный автоэнкодер кроме предсказания получает также оценку дисперсии, из которой вычисляется доверительный интервал.

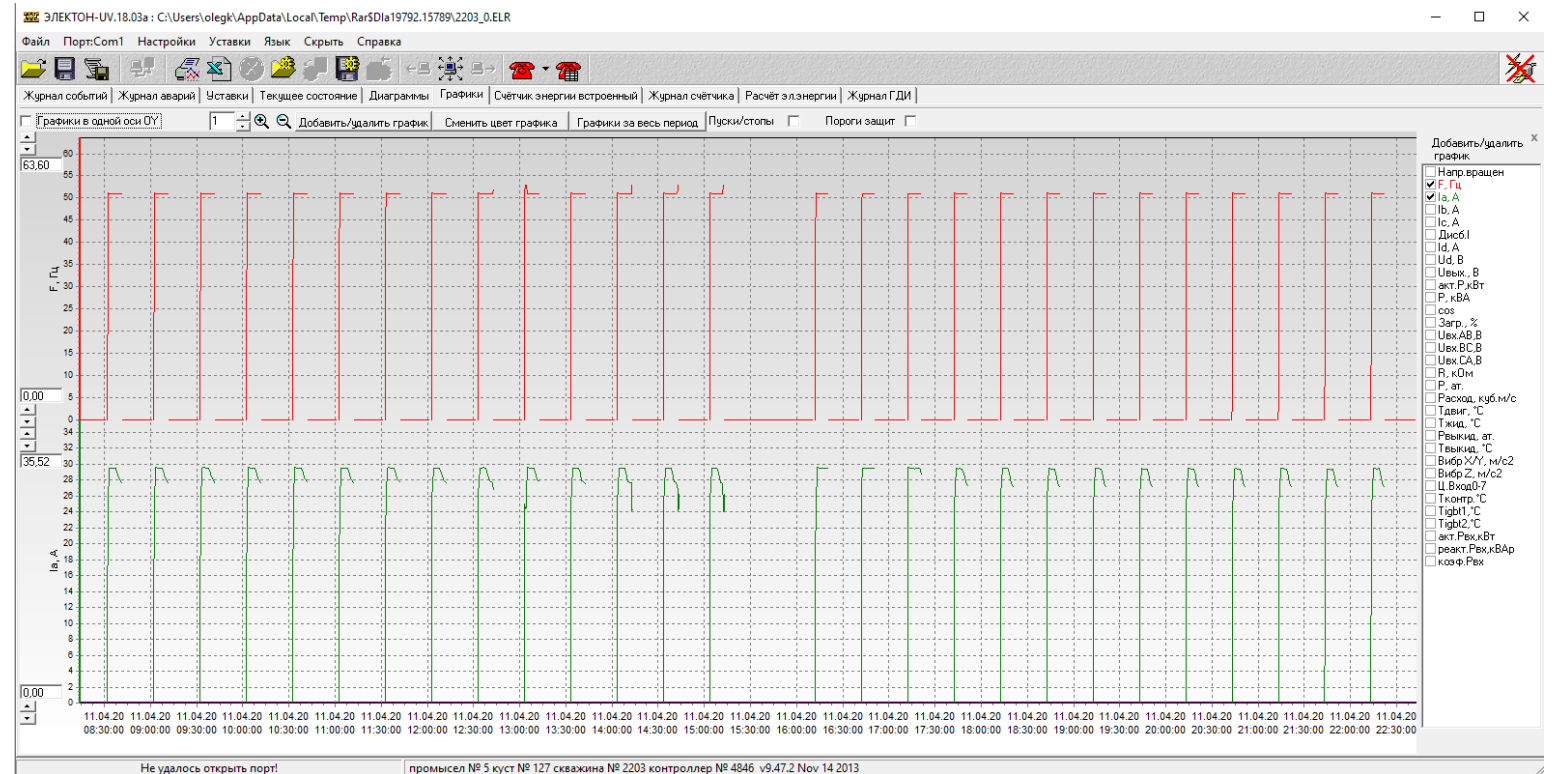
Особенности использования:

- Взвешенное усреднение (учет разной дискретности на временных интервалах)
- Добавление искусственных пропусков и выбросов к обучающим данным.



Аномальные циклы ПКВ

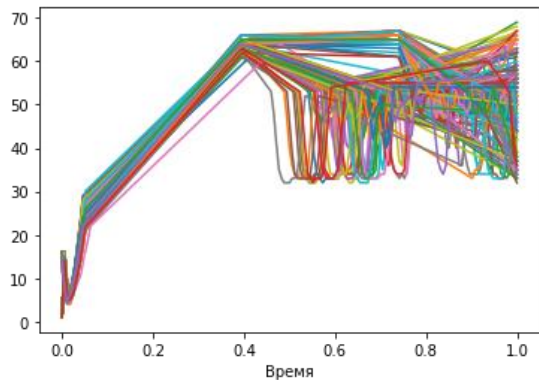
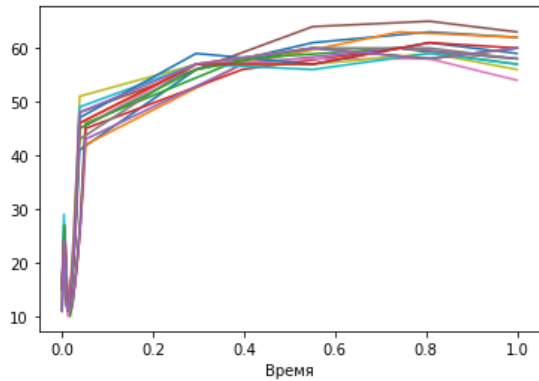
- Скважины, работающими в ПКВ режиме (цикл работы и накопления)
- В норме цикл похож на букву «П»
- Изначально нет потоковых данных, только выгрузки в специальном формате
- Технологию нужно вручную просматривать 10 параметров (временных рядов) по каждой из 100 скважин раз в неделю
- Есть аномалии работы (форма «зубца»), которые связаны с определенным блоком проблем
- Отсутствие разметки



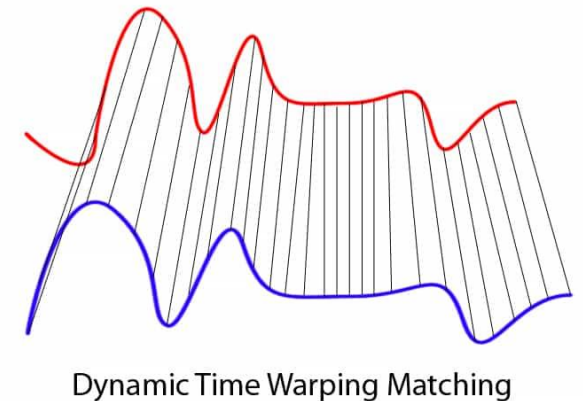
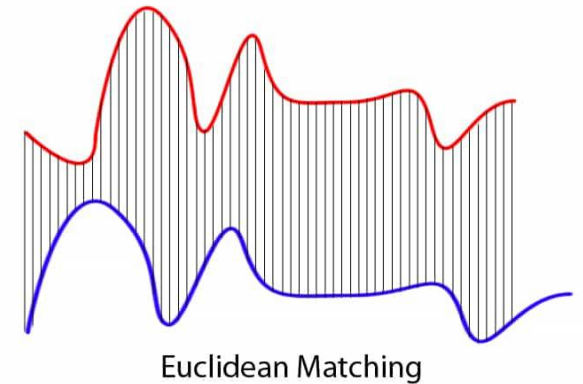
Проблема: много данных, которые анализируются вручную

Аномальные циклы ПКВ

Технологии



- Pandas и терпение для парсинга исходных данных
- If-else на разнице в качестве базовой модели
- DTW для определение аномалий по близости примера из библиотеки и рассматриваемого цикла
- Декомпозиция и метод накопленных сумм для анализа давления на приеме
 - Удаление цикличности и определение тренда
- Иерархическая кластеризация для формирования библиотеки аномалий

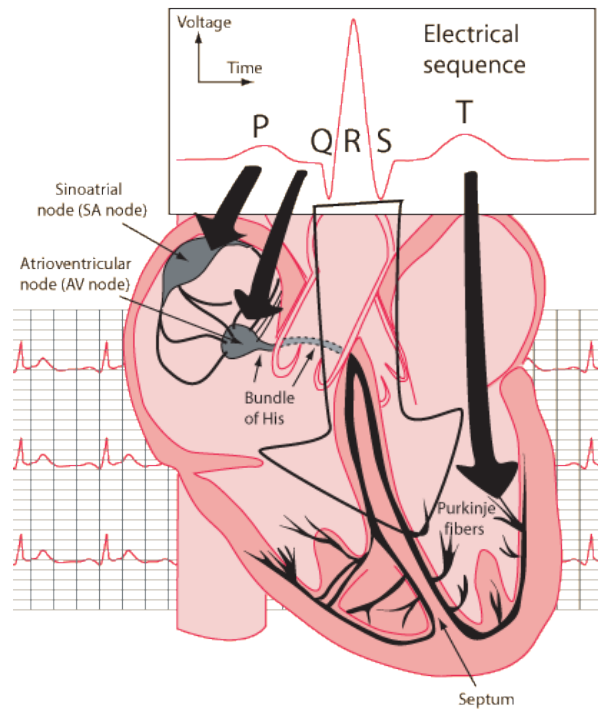


Грамотный процесс решения задачи с легким выходом на продуктив

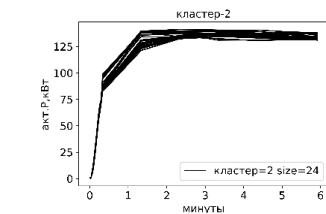
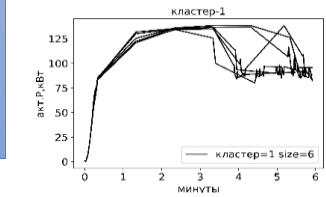
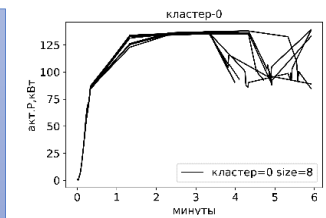
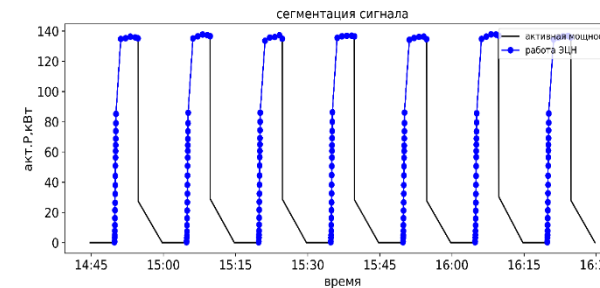
Перенос лучших практик из других отраслей

Стандартное решение в одной отрасли может быть уникальным в другой

HEALTCHARE – анализ данных ЭКГ



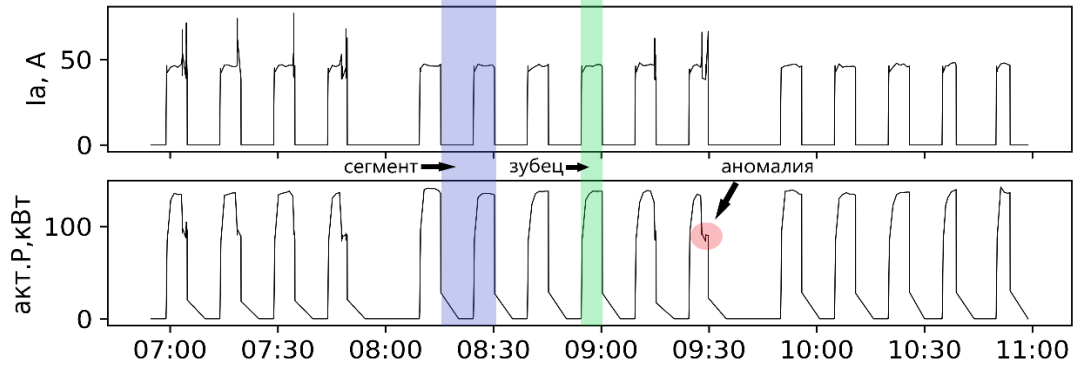
PETROLEUM ENGINEERING – анализ данных ПКВ



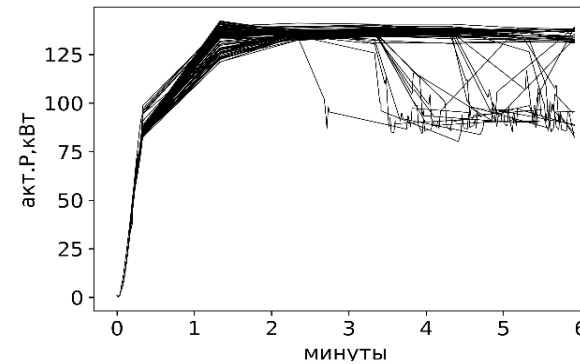
Успешная интеграция подходов из области обработки медицинских данных в задачу по выявлению аномальных циклов периодических ЭЦН-овских скважин

Пример анализа исходных данных по скважине

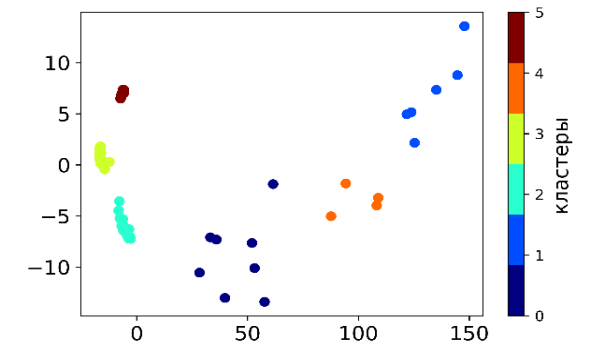
1. Исходный сигнал



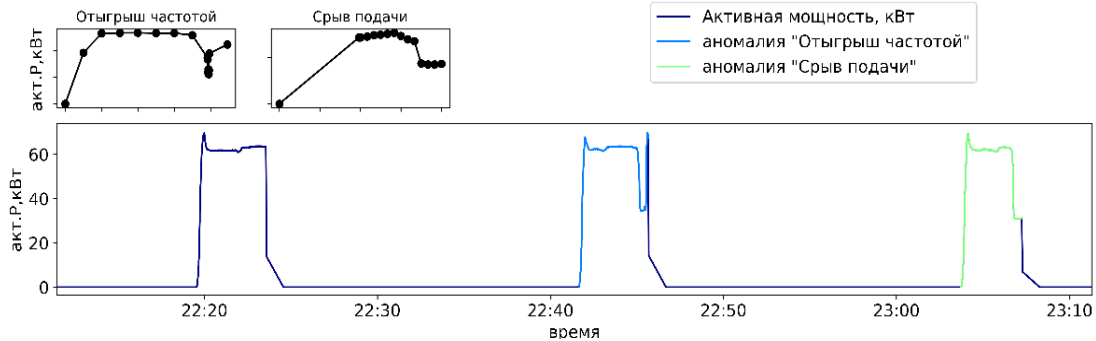
2. Сегментация сигнала



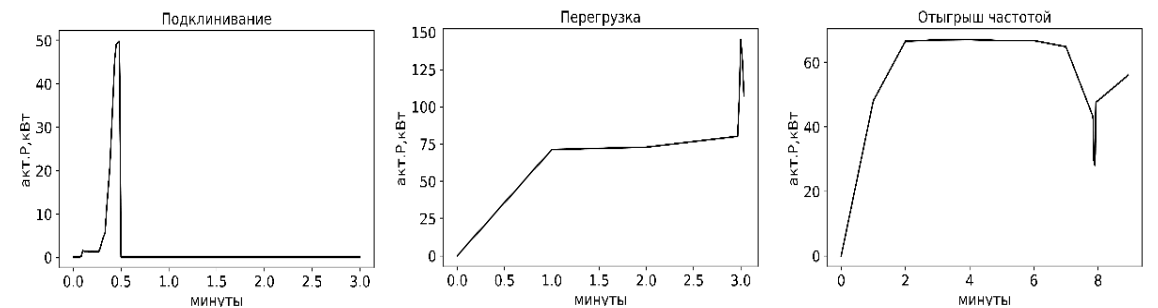
3. Кластеризация



5. Поиск аномальных циклов для новой скважины

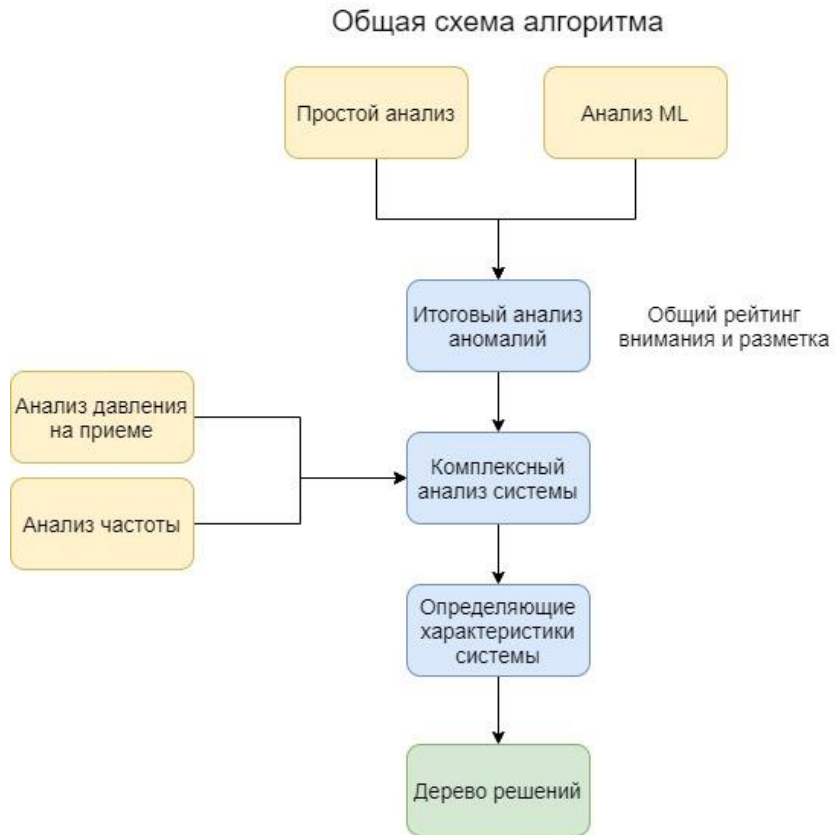


4. Сформирована библиотека аномалий

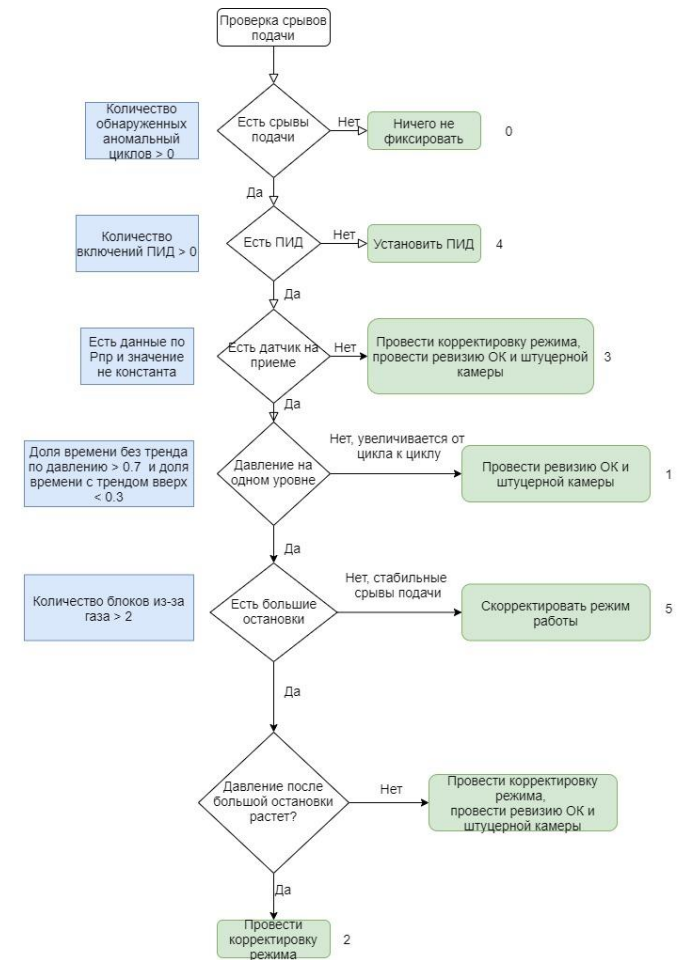


Система рекомендаций

Анализ работы скважины



Рекомендация по оптимизации



Задачи

- Виртуальная расходомерия для:
 - Фонтанных скважин
 - ЭЦН
 - Газлифт

Гибридное моделирование

Способы объединения физических и статистических методов



Основные темы

- Схема подхода по составлению решения
- Определения однозначности решения
- Методы оптимизации
- Узловой анализ



Результат

- Увеличение добычи
- Снижение затрат
- Косвенные эффекты

Физические расходомеры

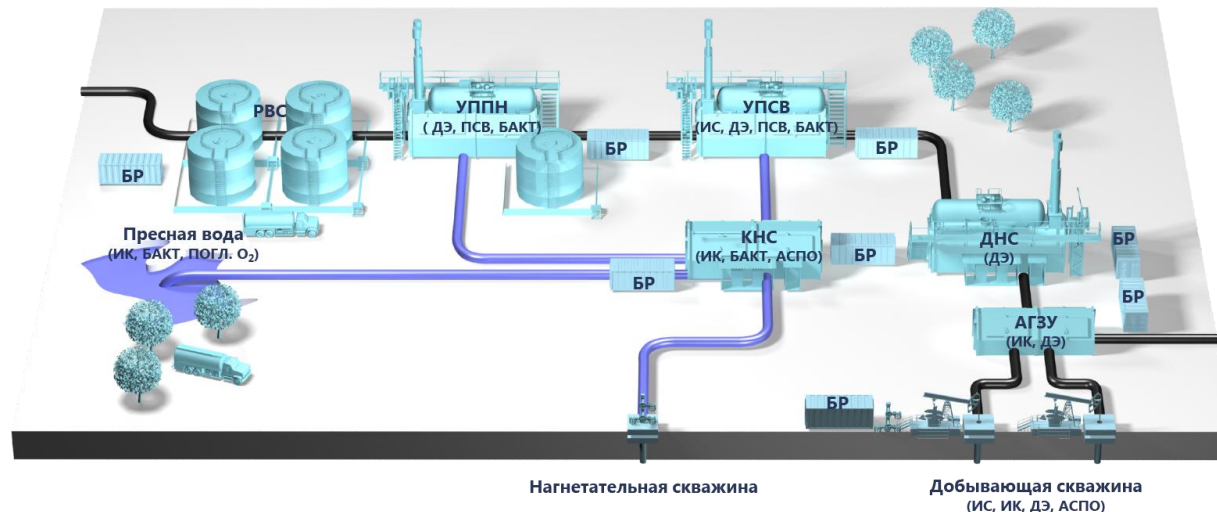
- **Дебит скважины** – сколько продукции добывает скважин за определенный промежуток времени (м³/сут) – основной показатель работы скважины
- **Автоматизированная групповая замерная установка (АГЗУ)** предназначена для автоматического периодического определения продукции нефтяных скважин и контроля за их технологическими режимами.



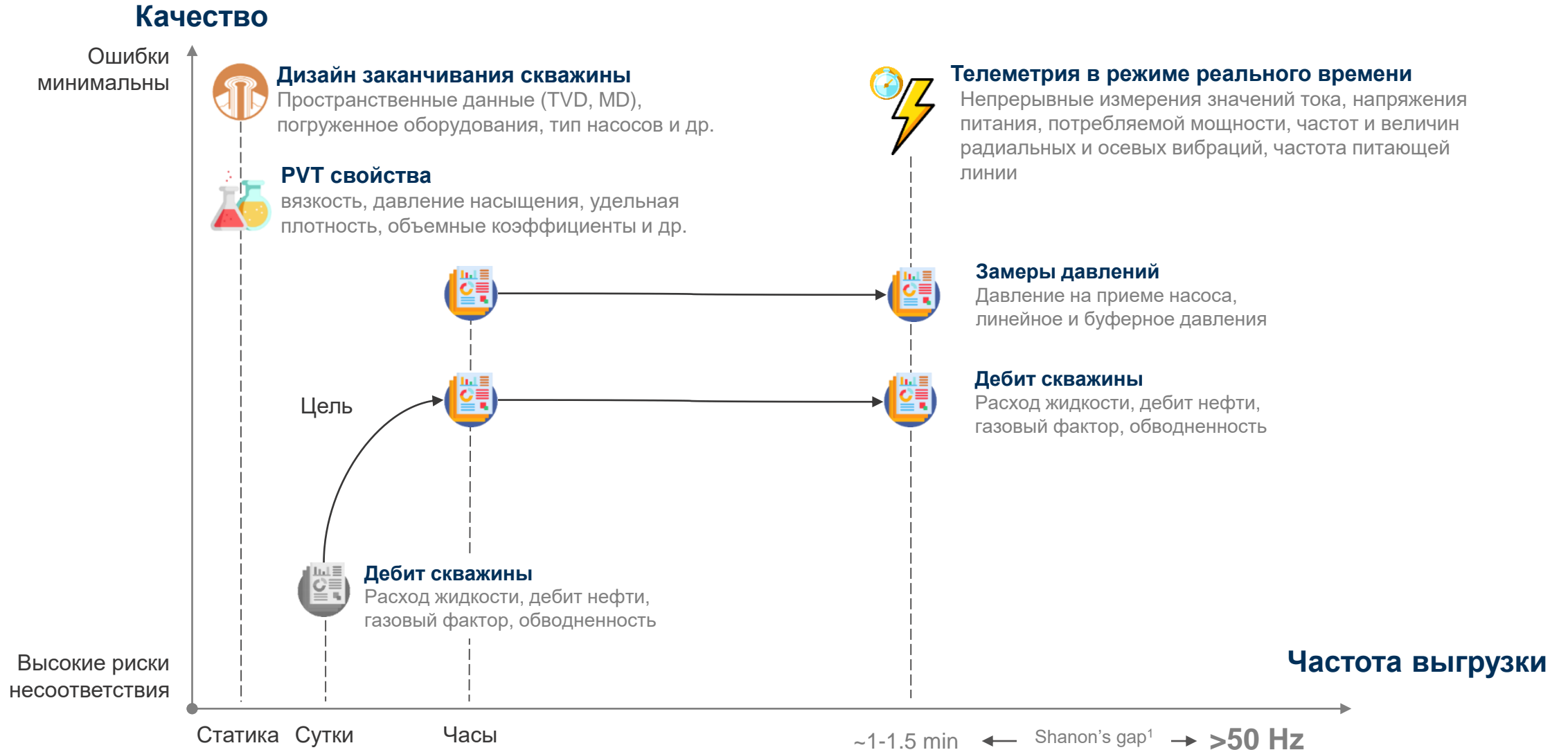
Особенности использования АГЗУ

- Дорого, т.к. устройство сложное.
- Низкая частота замеров.
- Низкий охват.
- Нехватка АГЗУ.

Проблема контроля работы скважин (игра «Поиск добычи»)



Алгоритмы машинного обучения нацелены в первую очередь на улучшение качества данных



Пример расчета Виртуального расходомера

Увеличение частоты данных по дебиту скважины с помощью виртуальных замеров позволяет явно отслеживать работу добывающей системы



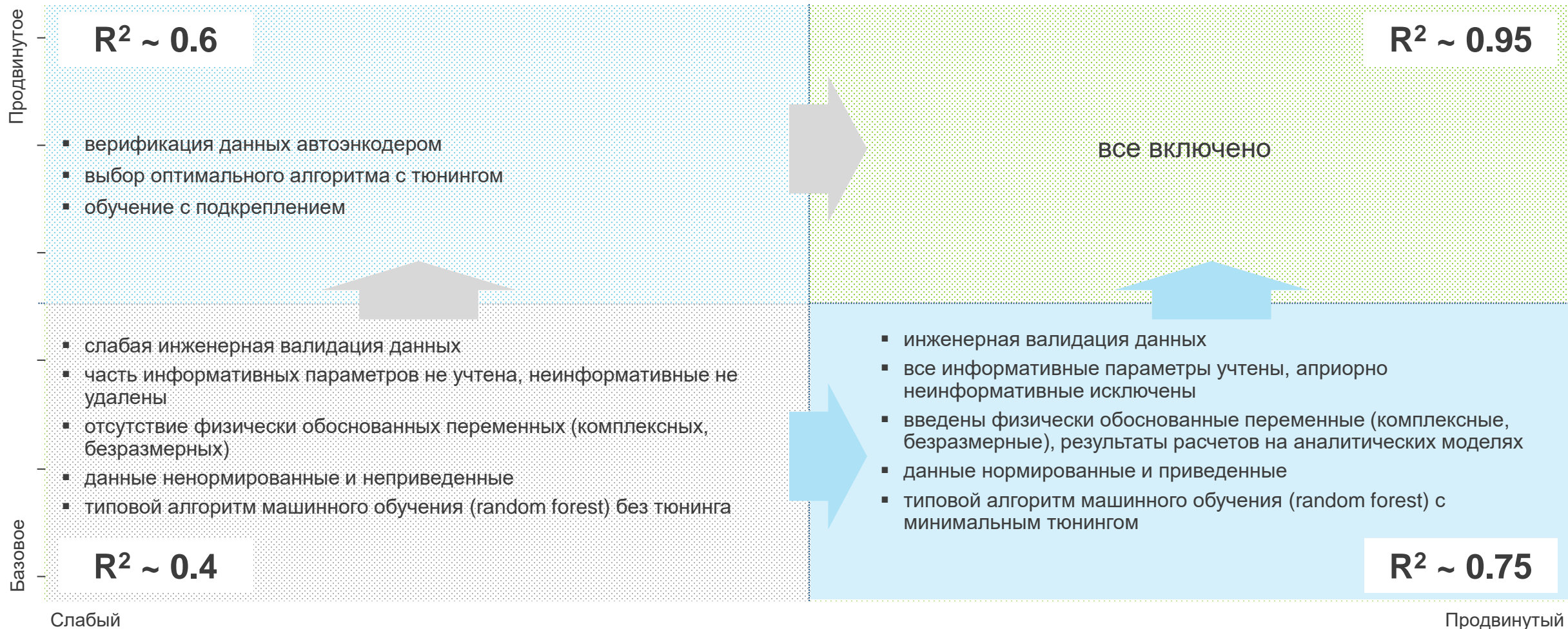
Виртуальная расходомерия позволяет наблюдать изменение в режиме работы скважины



Гибридные подходы 2/2

Факторы повышения прогнозной способности апостериорных моделей

Машинное обучение



Нефтяной инжиниринг

Реализация ряда проектов в нефтяном секторе позволила сделать следующие выводы

Выводы:

- Качество и количество данных, **необходимое**, но не достаточное условие для решения задачи в нефтяной промышленности
- Понимание процессов нефтедобычи **обязательно** для любого типа моделирования
- Машинное обучение в одиночку не способно бороться с технологическими вызовами отрасли, актуальность **гибридных подходов** растет
- Решение **не может считаться успешным** без внедрения и промышленной эксплуатации
- **~80% сложности** возникает в части data engineering, влечет к большим временным затратам при подготовке «почвы» для аналитиков данных