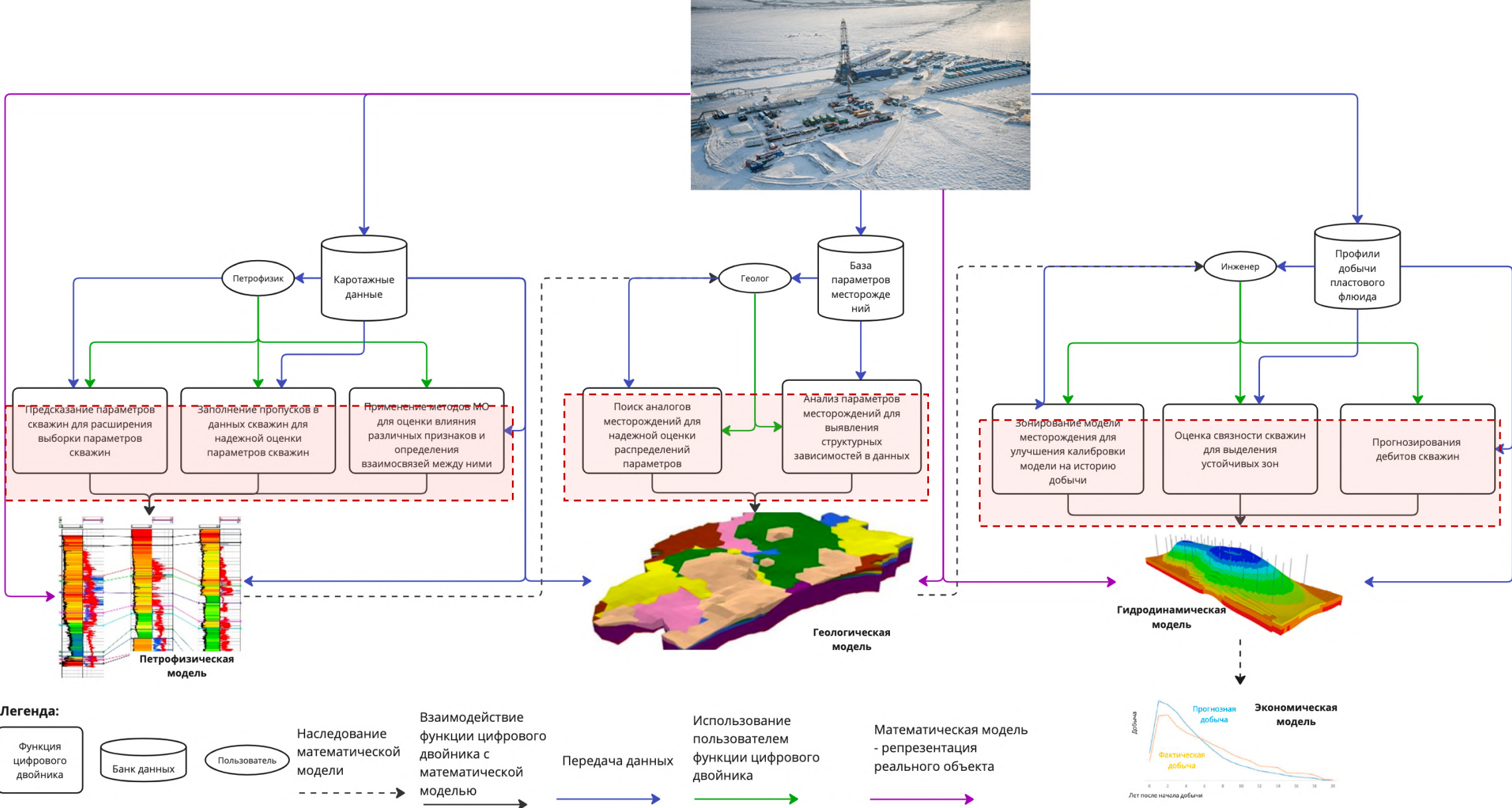


Методы интерпретации математических моделей нефтегазовой отрасли

Антон Воскресенский – PhD, CEO Voskresenskii Consulting

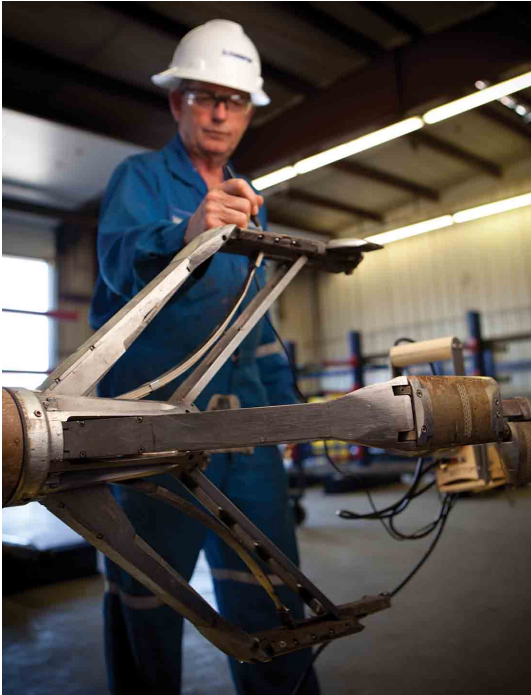
5 марта 2025

Цифровой двойник резервуара месторождения [4]

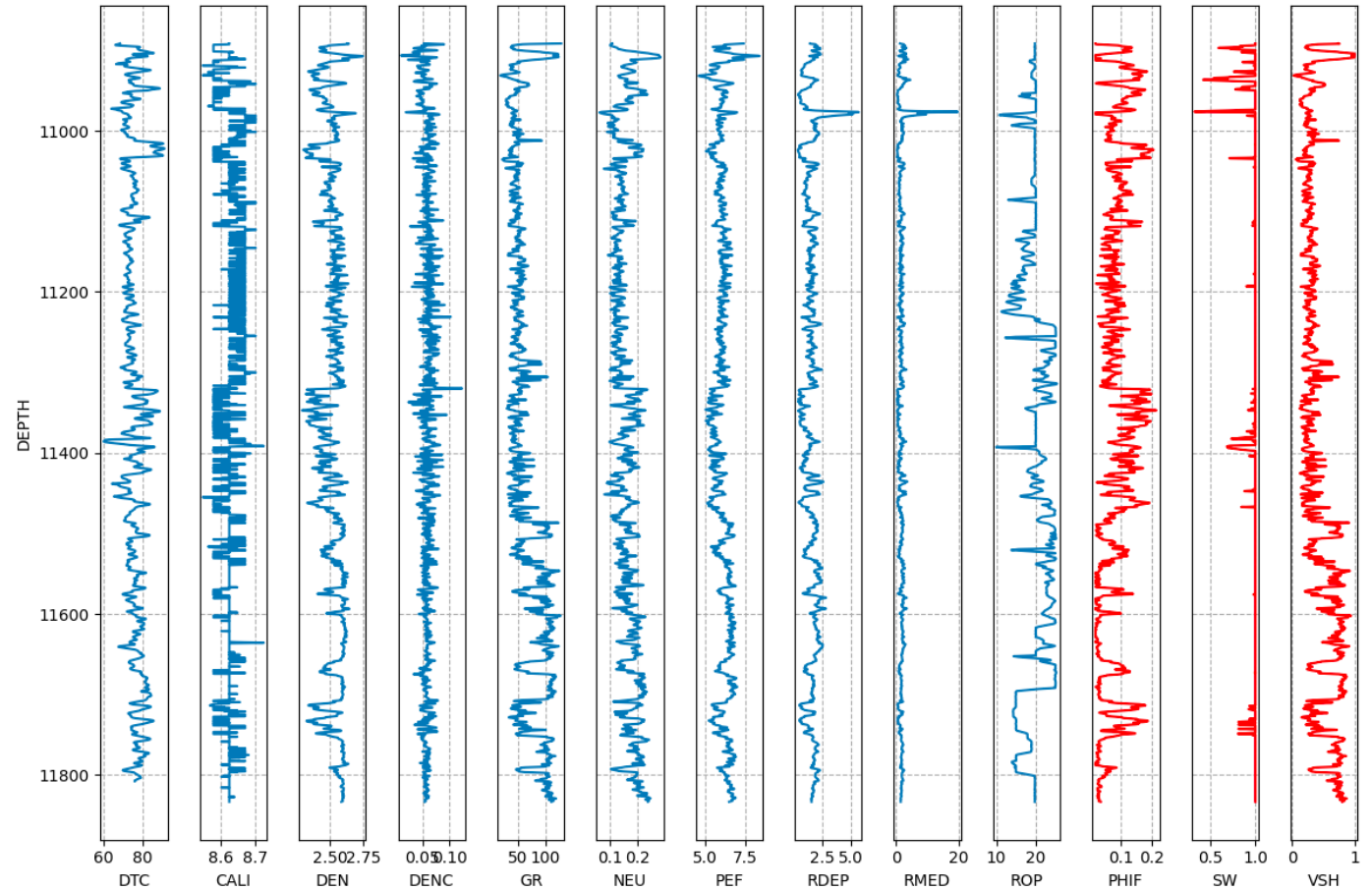


Что такое каротажные данные и как их получают?

а)



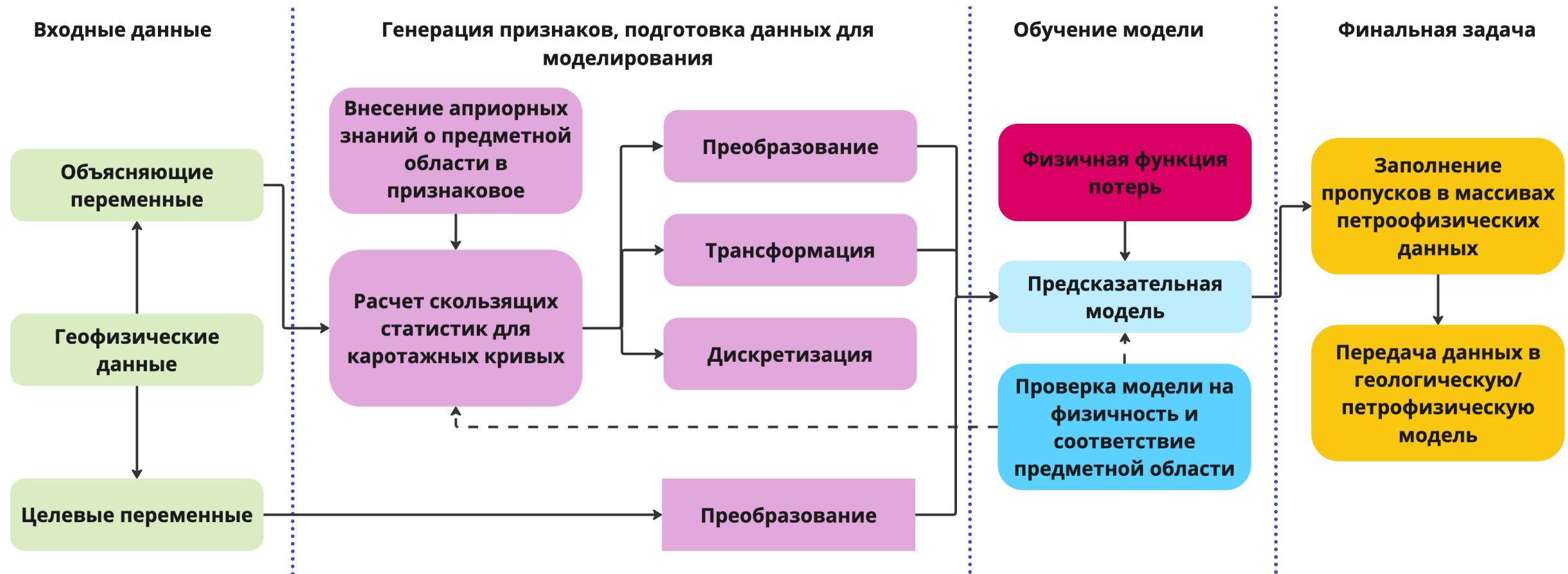
б)



Пример визуализации каротажных данных

Прибор для записи каротажных данных а) рядом с человеком (для масштаба) и б) в стволе скважины [5]

Метод моделирования геофизических данных скважин



Общая схема метода заполнения пропусков на основе моделирования и предсказанию геофизических параметров скважин [4]

Алгоритм заполнения пропусков в геофизических данных скважин [4]

Шаг 1. Подготовка данных: собрать X и y (с учётом пропусков).

Шаг 2. Первичный инжиниринг признаков.

Шаг 3. Добавление априорных знаний (на примере направления тренда):

- Задать углы $\varphi_k \in \{0^\circ, 1^\circ, \dots, 180^\circ\}$; направления $u_k = (\cos \varphi_k, \sin \varphi_k)$.
- Для каждого k : $d_i^{(u_k)} = x_i \cos \varphi_k + y_i \sin \varphi_k$.
- Найти $\rho_k = \text{corr}(d_i^{(u_k)}, y_i)$.
- Определить $u^* = \arg \max_k |\rho_k|$.
- Новый признак: $x_i^{\text{spatial_trend}} = x_i \cos \varphi^* + y_i \sin \varphi^*$.

Шаг 4. Преобразование целевого признака:

$$y_t^{\text{rel}} = \frac{y_t - y_{t-1}}{y_{t-1}}.$$

Шаг 5. Обучение модели:

- $\hat{y}_t^{\text{rel}} = f(X_t; \theta)$, где $\theta^* = \arg \min_{\theta} L(y^{\text{rel}}, \hat{y}^{\text{rel}})$.
- Восстановить \hat{y}_t (пропуски) по формуле $\hat{y}_t = \hat{y}_{t-1}(1 + \hat{y}_t^{\text{rel}})$.

Шаг 6. Проверка физической обоснованности:

- Анализ с помощью значений Шепли.
- $\frac{\partial \rho}{\partial z} \geq 0$ (увеличение плотности с глубиной), $\frac{\partial \phi}{\partial z} \leq 0$ (уменьшение пористости с глубиной).
- $\hat{y}_t \in [y_t^{\min}, y_t^{\max}]$.
- $\Delta \hat{y}_i = \hat{y}_{i,t} - \hat{y}_{i,t-1}, \quad |\Delta \hat{y}_i| < \varepsilon$.

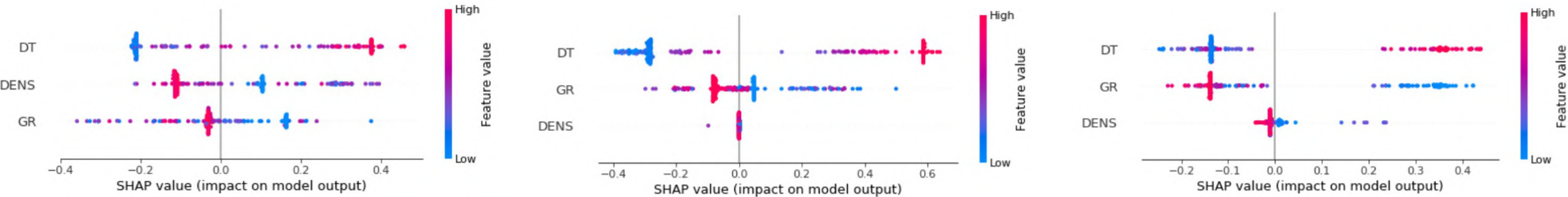
Data: X (матрица объясняющих переменных, включая координаты (x_i, y_i)), y (целевой признак с пропусками)

Result: \hat{y} (заполненные значения),
расширенное
признаковое
пространство

Кейс 1 и 2. Результаты применения метода заполнения пропусков в геофизических данных скважин

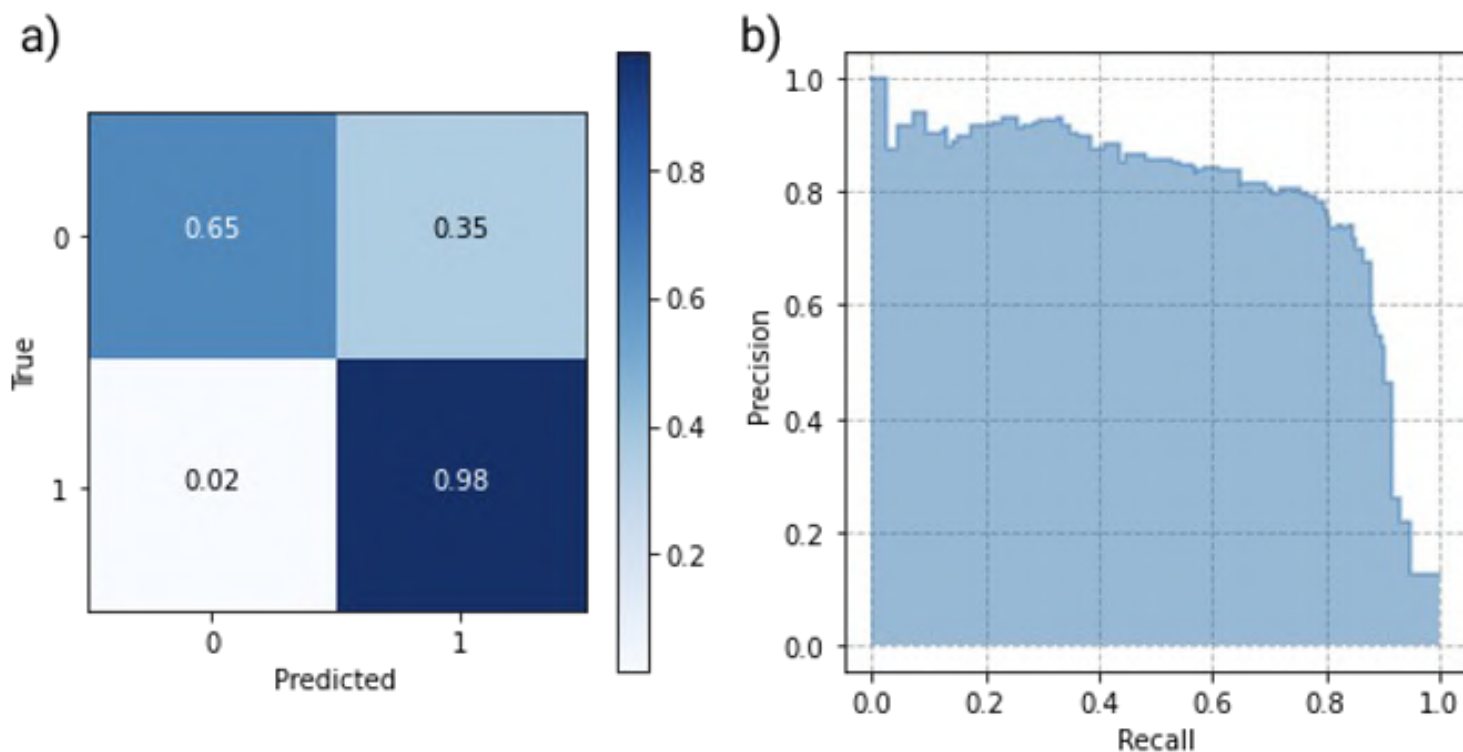
Класс	Метрики до применения метода			Метрики после применения метода			Приращение (%)
	Точность	Полнота	F1-мера	Точность	Полнота	F1-мера	
0	0.70	0.50	0.58	0.84	0.65	0.73	+20.0, +30.0, +25.9
1	0.90	0.85	0.87	0.95	0.98	0.97	+5.6, +15.3, +11.5

Сравнение метрик модели бинарной классификации до и после применения алгоритма внесения априорных данных в признаковое пространство модели. Внесение априорных данных увеличивает метрики классификации геофизических параметров [4]



Сводные графики чисел Шепли признаков моделей МО, построенных с использованием алгоритма внесения априорных данных. Несмотря на использование различных моделей, самый важный (первый) признак остается неизменным; характер влияния признаков одинаков у всех представленных моделей [4]

Кейс 3. Метрики точности модели бинарной классификации

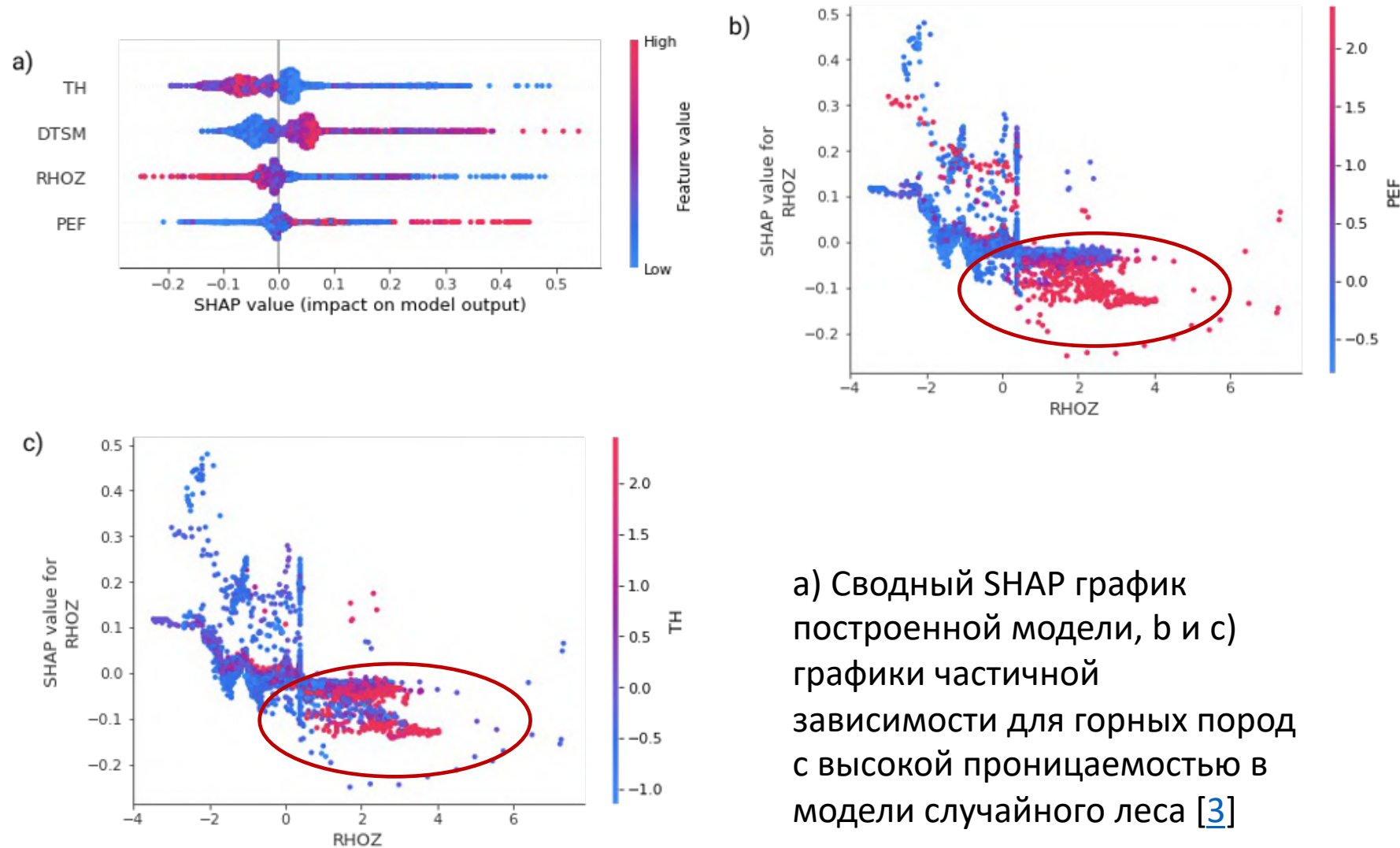


Rock type	Precision	Recall	F1-score
0	0.84	0.65	0.73
1	0.95	0.98	0.97
Accuracy	0.94	0.94	0.94
Macro avg.	0.89	0.82	0.85

Статистические показатели классификационной модели, учитывающие перекрестную проверку с удалением одной скважины и инжиниринга признаков

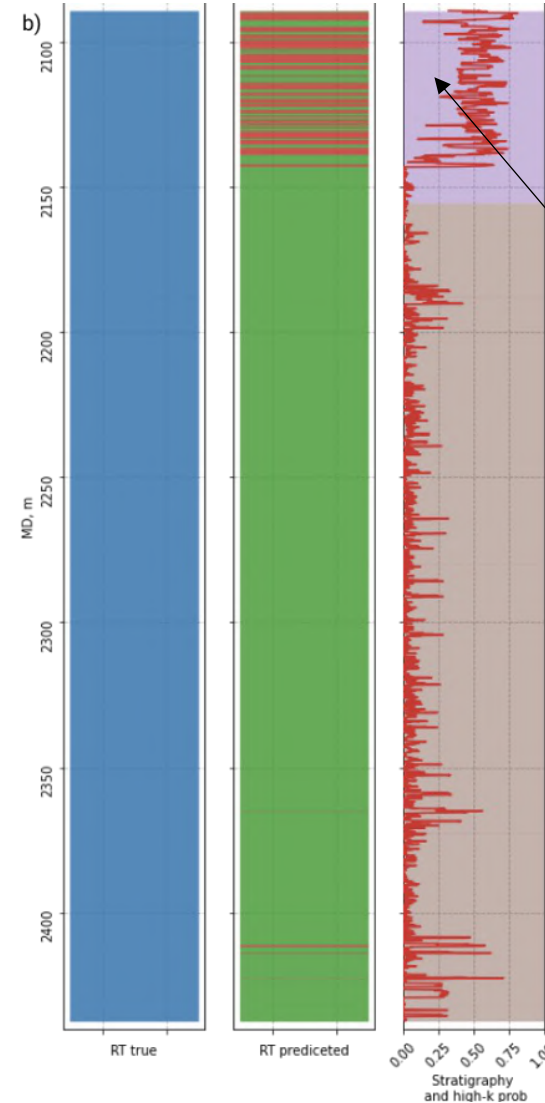
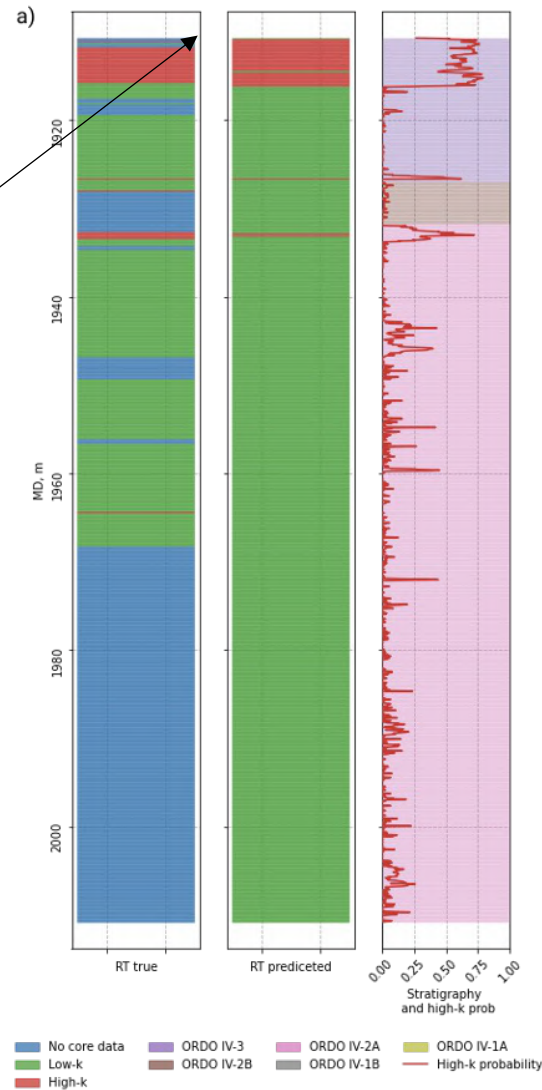
а) матрица несоответствий для тестовых данных классифицирует типы горных пород как "0" для высокой и "1" для низкой проницаемость. 2) кривая точности-полноты показывает баланс между истинными положительными показателями и точностью предсказания модели при различных пороговых значениях псевдовероятности [3]

Кейс 3. Интерпретация предсказаний модели



Кейс 3. Примеры предсказаний на скважинах

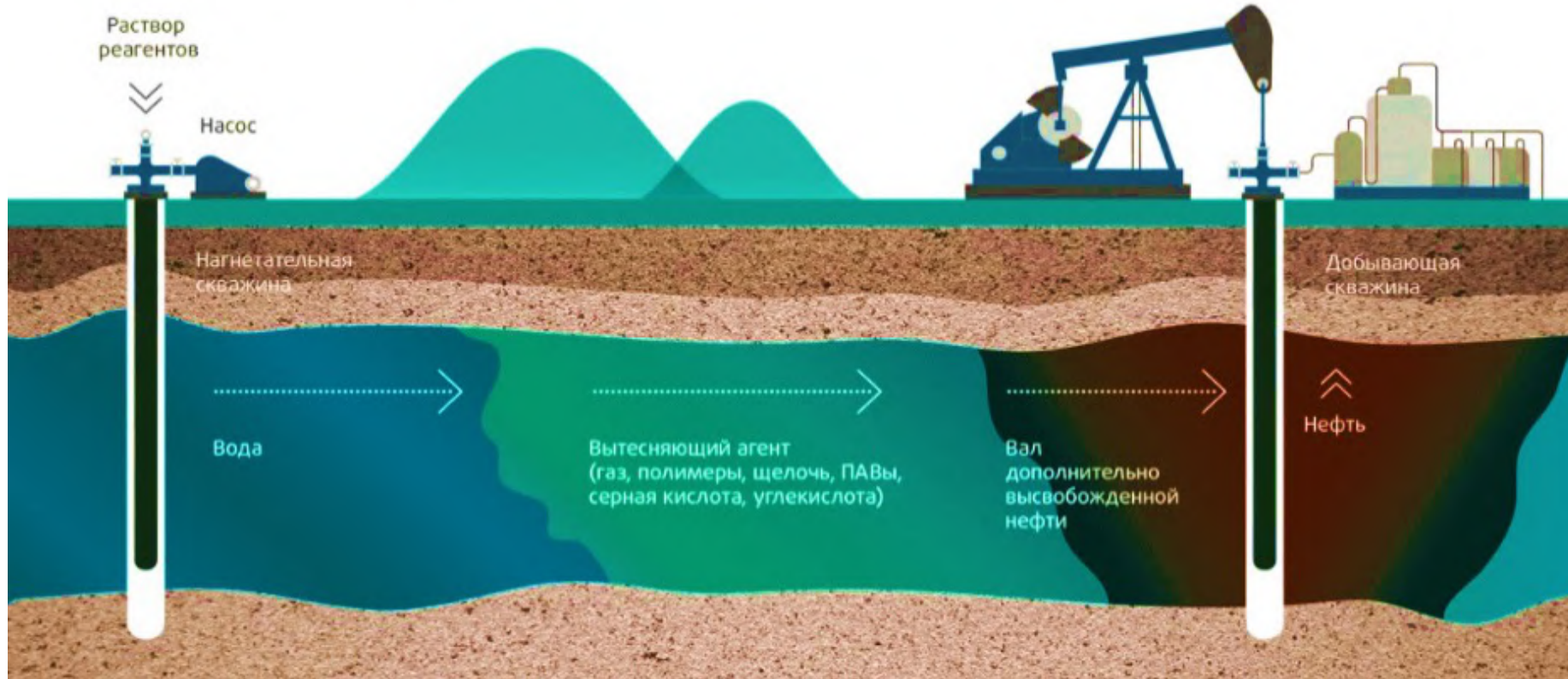
Модель была протестирована на тестовых данных и успешно предсказала интервалы с высокими значениями проницаемости в скважинах с керном



Мощный интервал с высокой проницаемостью на скважине не осященной керном.

Предсказания для: а) тестовой скважины с данными истинных меток, б) скважины без истинных меток [3]

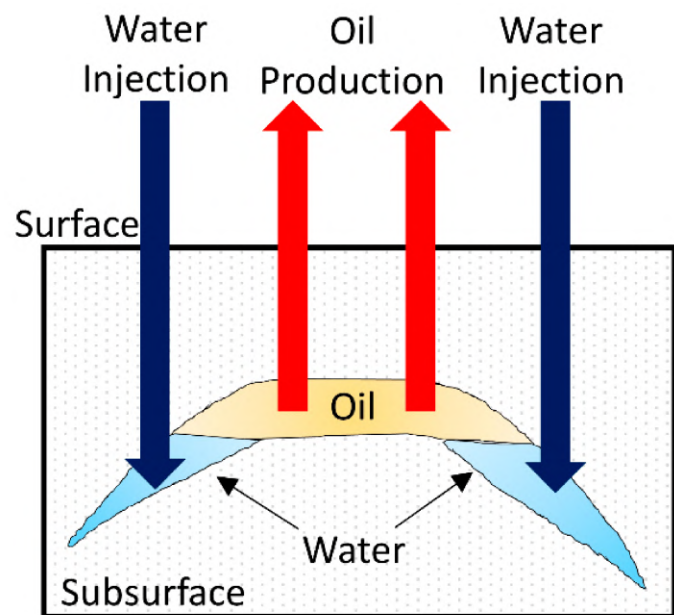
Использование ИИ для прогнозирования добычи и управления заводнением



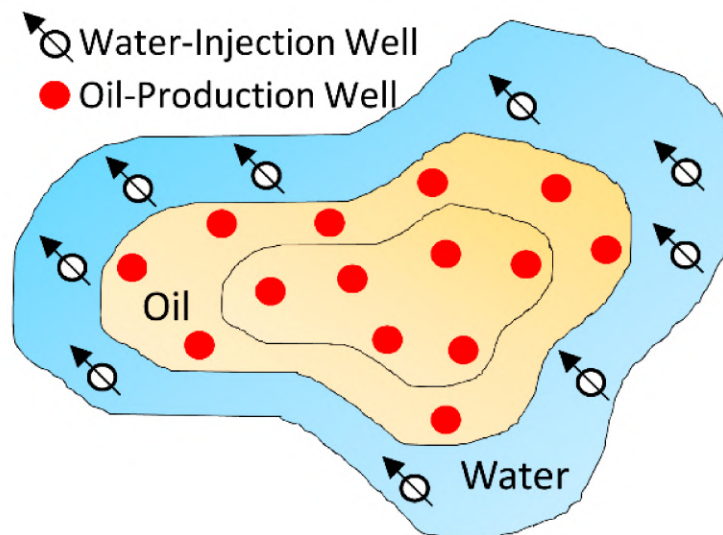
Заводнение месторождения с целью поддержания пластового давления

- Разработка ML решения для обеспечения прогнозирования дебитов скважин.
- Экономический эффект достигается за счет снижения операционных затрат на закачку воды и добычу скважин за счет снижения обводненности добывающих скважин.

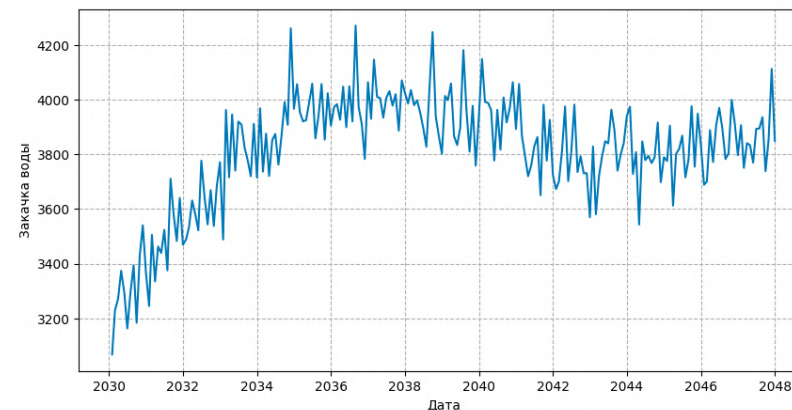
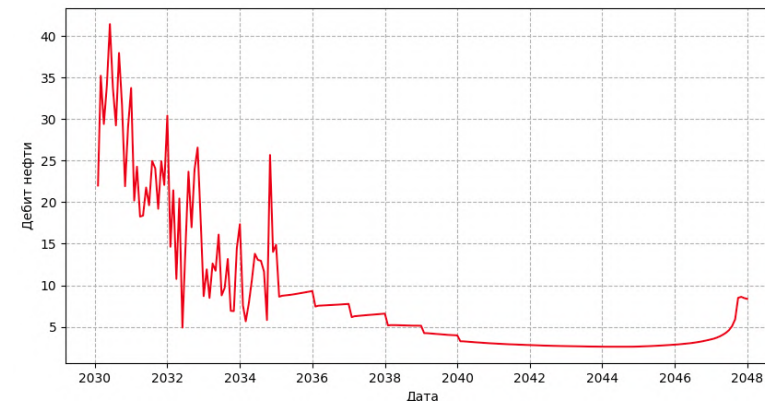
Данные: геоданные и временные ряды



Местоположения скважин (в профиль) закачки и добычи флюида из резервуара месторождения [7]



Местоположения скважин (вид сверху) различных типов на площади месторождения [7]



Профили добычи нефти и приемистости воды

Метод моделирования дебитов и связности скважин



Схема моделирования дебитов скважин и их связности [4]

Схема моделирования связности скважин

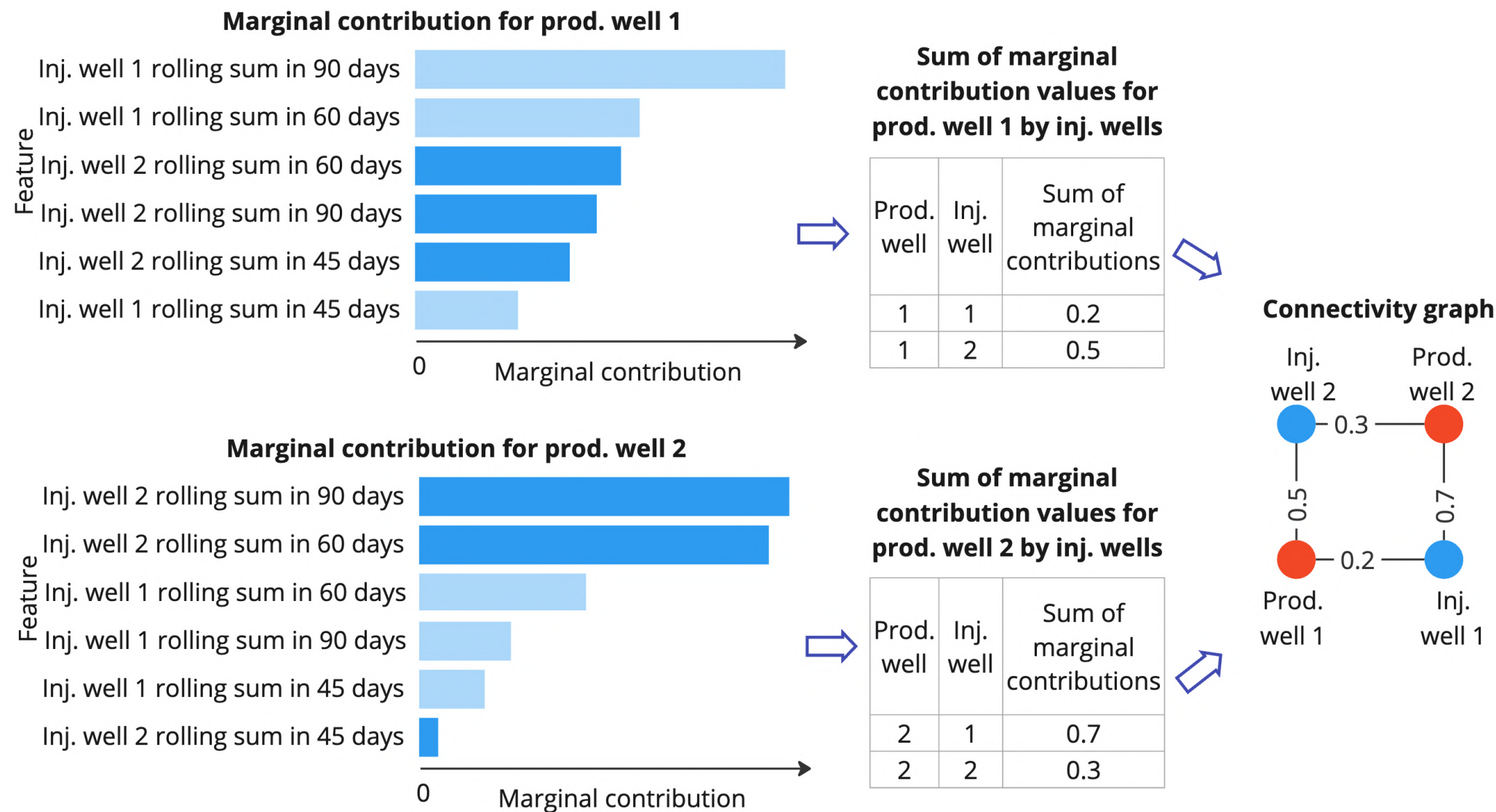


Схема моделирования связности скважин [1]

Схема моделирования связности скважин в динамике

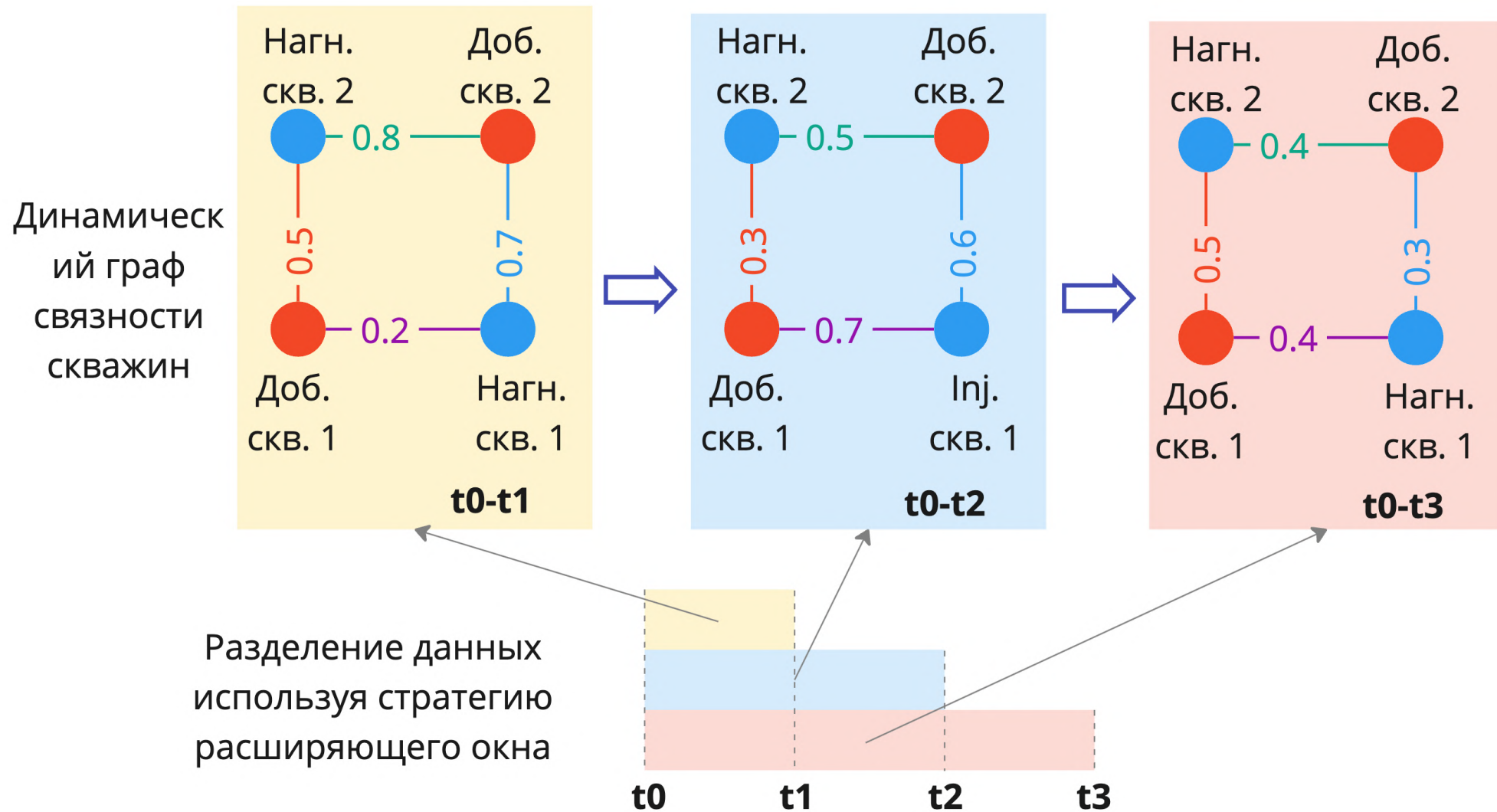


Схема моделирования связности скважин в динамике [1]

Алгоритм суррогатного моделирования связности скважин [4]

1. *Представление данных.* Разделяем исходный тензор \mathbf{X} по времени. Для каждого шага $t \in \{1, \dots, T\}$ имеем срез $\mathbf{X}_t \in \mathbb{R}^{N \times F}$.
2. *Обучение модели прогнозирования.* Тренируем $\hat{y}_t = f(\mathbf{X}_t)$ (любая подходящая ML-модель) на интервале времени.
3. *Интерпретация модели через значения Шепли.* Пусть $\mathcal{N} = \{1, 2, \dots, F\}$ — множество всех признаков. Тогда значение Шепли для i -го признака:

$$\phi_i(\mathbf{X}_t) = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \frac{|S|! (|\mathcal{N}| - |S| - 1)!}{|\mathcal{N}|!} [f(S \cup \{i\}) - f(S)].$$

4. *Агрегация значений Шепли.* Для каждой нагнетательной скважины W_i определяем набор индексов $K_i \subseteq \{1, \dots, F\}$ (признаки, относящиеся к W_i). Тогда совокупный вклад:

$$\Phi(W_i) = \sum_{k \in K_i} \phi_k(\mathbf{X}_t).$$

5. *Создание матрицы смежности \mathbf{A} .* Пусть элемент a_{ij} отражает степень связности между добывающей скважиной i и нагнетательной скважиной j . Формируем $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{N_{\text{prod}} \times N_{\text{inj}}}$.
6. *Построение графа связности на интервале.* Определяем $G = (V, E)$, где V — множество скважин, а $E = \{(i, j) \mid a_{ij} > 0\}$.
7. *Формирование графа связности для жизненного цикла.* Для каждого шага t имеем $G_t = (V_t, E_t)$. Совокупность $\{G_t\}_{t=1}^T$ описывает динамику связности во времени.

Require:

$\mathbf{X} \in \mathbb{R}^{T \times N \times F}$
где:

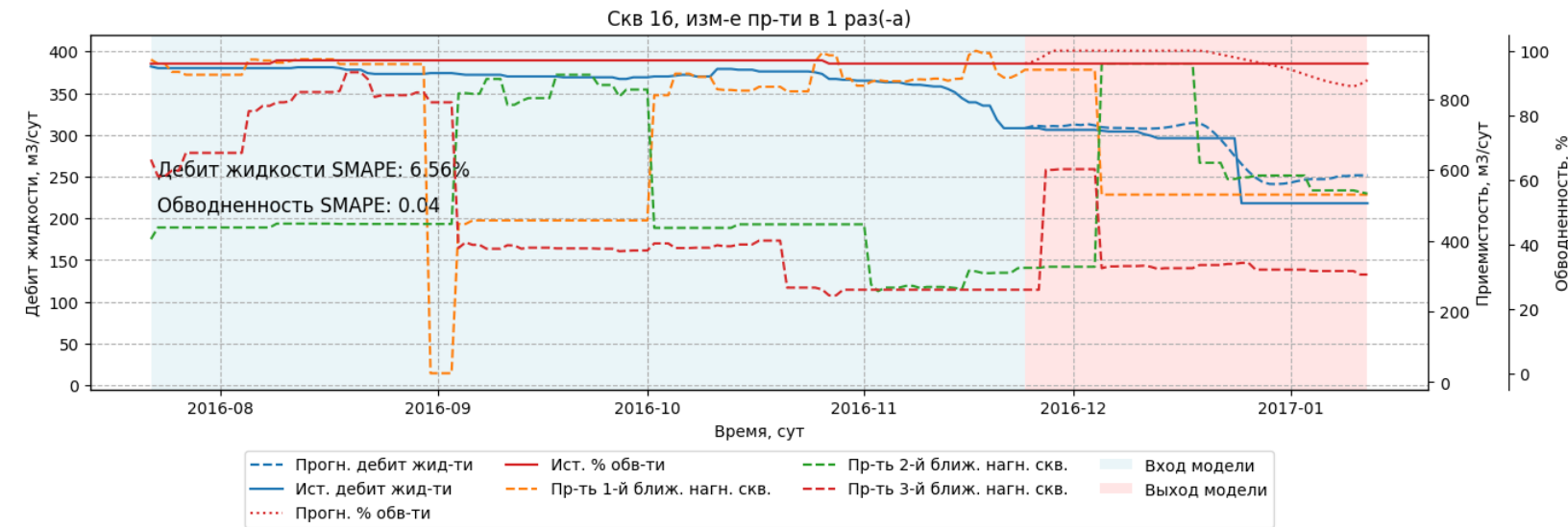
- T — количество временных шагов,
- N — количество скважин,
- F — число признаков (например, давление, дебит и т. д.).

Ensure: Граф(ы) связности скважин

Данные: группы рассматриваемых признаков модели

Общие	Технология	Окружение	Флюид	Пласт
Идентификаторы и временные отметки, такие как дата измерения и идентификатор скважины, обеспечивающие структурную организацию данных	Связаны с операционной деятельностью и параметрами скважин, включая дебит жидкости, обводненность, и другие показатели, влияющие на производительность скважин, такие как затрубное давление и параметры оборудования (например, диаметр обсадной колонны, диаметр дросселя)	Отражают влияние соседних скважин и общие условия добычи, например, приемистость соседних скважин и дебит жидкости соседних добывающих скважин	Характеристики добываемых флюидов, такие как вязкость нефти и жидкости, что важно для понимания поведения флюида в пласте и скважине	Геологические и физические характеристики пласта, куда входят координаты скважины, абсолютная глубина, мощность пласта, литология, пористость, проницаемость и другие показатели, определяющие потенциал пласта для добычи

Примеры физических прогнозов дебитов скважин

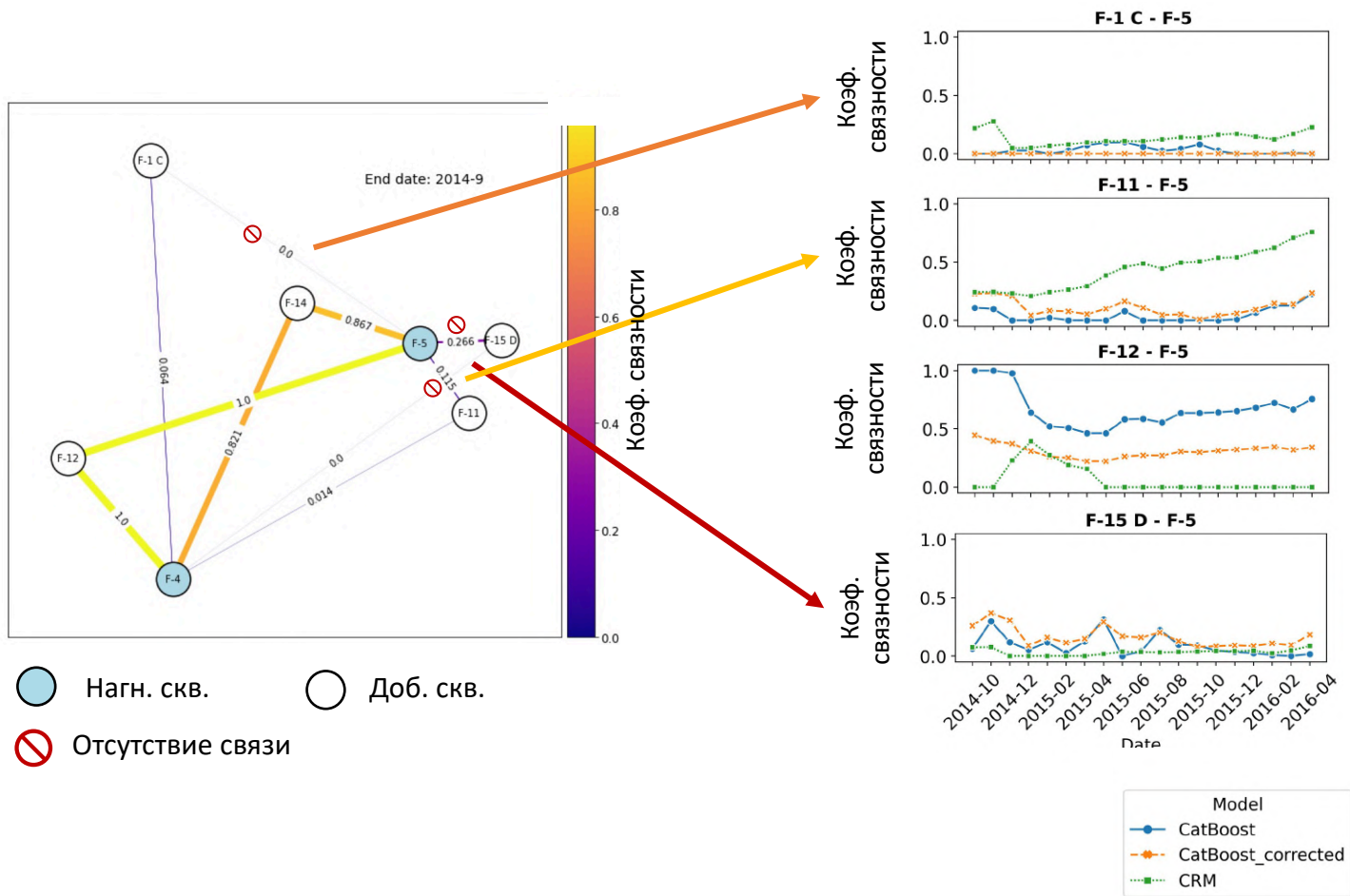


Модель физично реагирует на изменение приемистости в ближайших нагнетательных скважинах



Примеры прогнозов на двух скважинах [4]

Примеры оценки связности скважин



Метод суррогатного моделирования связности скважин определяет отсутствие гидродинамической связи в 100% случаев отсутствия связи, что подтверждается ретроспективными исследованиями. Однако, традиционная модель CRM демонстрирует наличие связей которые невозможны с предметной точки зрения [1]

Схема связности скважин [1] построенная на датасете Volve [2]

Выводы

- Внесение априорных знаний о предметной области и физических функций потерь улучшают не только точность, но и улучшают физичность моделей
- Модели построенные с учетом физики процесса демонстрируют физические предсказания
- Интерпретация моделей позволяет находить инсайты в данных пропущенные предметными специалистами

Физический подход

- Долго
- Сложно
- Нужны предметные специалисты
- Сложно поддерживать
- Понятно предметному специалисту
- Не нужно много данных
- Работает

Data-driven подход

- Быстро
- Просто
- Не нужны узкие специалисты
- Легко поддерживать
- Непонятно предметному специалисту
- Нужно достаточно данных
- Не факт, что заработает

Спасибо за внимание!