

ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ С ИСПОЛЬЗОВАНИЕМ ОБОРУДОВАНИЯ ИВЦ НГУ

1. Аннотация

С ростом внедрения методов горизонтально-направленного бурения точное определение параметров окружающей среды стало особенно важным: вся поверхность горизонтального окончания является пятном контакта, поэтому бурение должно осуществляться в наиболее нефтенасыщенной части коллектора.

Устройство ВИКПБ, использующее высокочастотное электромагнитное зондирование, позволяет получать данные (сигналы), необходимые для определения характеристик среды около скважины.

Однако интерпретация этих сигналов часто проводится качественно, так как существующие программные системы количественной интерпретации данных либо имеют ограниченное применение, либо затруднены в реализации в производственном процессе. В то же время, из-за нелинейности влияния характеристик среды на сигналы ВИКПБ, качественная интерпретация подвержена ошибкам.

Поскольку разрабатывается всё больше месторождений с тонкослоистыми коллекторами, потребность в точной и своевременной интерпретации данных ВИКПБ возрастает. Это подчёркивает необходимость создания инструментов, позволяющих решать как прямые, так и обратные задачи обработки сигналов ВИКПБ в процессе бурения. В данном исследовании предлагается разработка и применение моделей машинного обучения для анализа сигналов устройства ВИКПБ. Эти модели направлены на преодоление ограничений существующих методов количественной интерпретации, повышая точность анализа данных.

Интеграция подходов машинного обучения в обработку данных высокочастотного электромагнитного зондирования позволяет создать надёжное решение, способствующее более глубокому пониманию подповерхностных условий.

Полученные модели и реализованный программный модуль на Python продемонстрировали высокую точность при прогнозировании как прямого поведения сигналов, так и решении обратной задачи с использованием синтетических данных, что делает их надёжным и практичным инструментом для количественной интерпретации данных ВИКПБ в условиях реального бурения.

2. Тема работы

Использование средств машинного обучения для решения прямых и обратных задач при определении параметров среды вблизи скважины по сигналам прибора ВИКПБ.

3. Состав коллектива

1. Масуд Рофикул Ал, студент группы 21214, Направление подготовки: 09.03.01 Информатика и вычислительная техника, Направленность (профиль): Компьютерные науки и системотехника, Факультет информационных технологий, Новосибирский национальный исследовательский государственный университет
2. Власов Александр Александрович, к.т.н., доцент, Кафедра систем информатики, Факультет информационных технологий, Новосибирский национальный исследовательский государственный университет, руководитель ВКР

3. Быкова Галина Павловна, ассистент, Кафедра систем информатики, Факультет информационных технологий, Новосибирский национальный исследовательский государственный университет, соруководитель ВКР

4. Научное содержание работы

4.1 Постановка задачи

Целью данного исследования является изучение применимости методов машинного обучения для количественного анализа данных, получаемых с прибора ВИКПБ (Высокочастотный Индукционный Каротаж в Процессе Бурения), при бурении горизонтальных окончаний нефтегазовых скважин.

Современные методы интерпретации сигналов ВИКПБ в основном носят качественный характер, что ограничивает их точность и применимость в условиях бурения тонкослоистых и сложнопостроенных коллекторов. Существующие программные решения для количественной интерпретации либо недостаточно точны, либо трудны в реализации на практике.

В связи с этим была поставлена задача: разработать программный модуль, использующий методы машинного обучения, для быстрой и точной количественной интерпретации сигналов ВИКПБ, поступающих в процессе бурения горизонтального ствола скважины.

Основные требования к разрабатываемому модулю включают:

- обеспечение высокой точности предсказания параметров среды как в прямой, так и в обратной задачах;
- устойчивость модели к шуму и особенностям синтетических и потенциально реальных данных;
- вычислительную эффективность, позволяющую использовать решение в режиме, близком к реальному времени.

4.2 Современное состояние проблемы

Современные технологии горизонтально-направленного бурения в нефтегазовой отрасли значительно повысили требования к точной и оперативной интерпретации параметров окружающей среды вблизи скважины. Это обусловлено необходимостью оптимизации траектории бурения для максимизации добычи, особенно в сложных геологических условиях, таких как тонкослоистые коллекторы [8]. Прибор высокочастотного индукционного каротажа (ВИКПБ), разработанный НПП ГА «Луч» и интегрированный в телеметрическую систему «Луч», играет ключевую роль в решении этой задачи [3, 7]. ВИКПБ использует высокочастотное электромагнитное зондирование для регистрации сигналов, отражающих такие параметры среды, как удельное электрическое сопротивление вмещающего пласта (R_{sfV}) и пласта, в котором находится прибор (R_{fv}), расстояние до границы слоя (H) и зенитный угол ($Zenith$). Эти параметры позволяют определять структуру и состав пород вблизи скважины, что критически важно для бурения [3].

Традиционные методы интерпретации данных ВИКПБ основаны на аналитических и численных подходах к решению прямых и обратных задач геофизики. Прямая задача заключается в моделировании сигналов прибора на основе известных параметров среды, тогда как обратная задача направлена на реконструкцию этих параметров по измеренным сигналам [4]. Например, изопараметрическое каротажное зондирование, описанное Антоновым, иллюстрирует ранние попытки аналитического моделирования сигналов, подобных ВИКПБ, но такие подходы часто ограничены упрощенными геологическими моделями [2]. Современные численные методы также имеют недостатки: многие из них применимы только в узком диапазоне геологических условий, что снижает их универсальность [9]. Кроме того, они требуют значительных вычислительных ресурсов, что затрудняет их использование в реальном времени во время бурения, где простой буровой установки связан с высокими финансовыми затратами [10]. Нелинейная зависимость сигналов ВИКПБ от параметров среды дополнительно усложняет интерпретацию, повышая риск ошибок и снижая оперативность принятия решений. Как отмечается в литературе, максимальное время предсказания для одной модели не должно превышать 10 мс, чтобы соответствовать требованиям производственных процессов [1].

Методы машинного обучения (ML) открывают новые перспективы для преодоления указанных ограничений. Современные исследования демонстрируют успешное применение нейронных сетей и алгоритмов градиентного бустинга, таких как LightGBM, XGBoost и CatBoost, для моделирования сложных нелинейных зависимостей в геофизических данных. Например, исследование Бобровского показало возможность быстрой симуляции гальванических и электромагнитных сигналов с использованием нейронных сетей в одномерных моделях околоскважинного пространства при боковом каротажном зондировании, что применимо и к задачам ВИКПБ [5]. В рамках текущих исследований для интерпретации сигналов ВИКПБ были обучены несколько моделей нейронных сетей на вычислительных кластерах Института вычислительного центра Новосибирского государственного университета (ИВЦ НГУ), что позволило обрабатывать большие объемы синтетических данных, таких как наборы *palette.csv* и *random.csv*, и достигать высокой точности предсказаний. Алгоритмы ML способны учитывать сложные взаимосвязи между параметрами и обеспечивать точность, сопоставимую с погрешностью прибора, что делает их перспективными для количественной интерпретации сигналов ВИКПБ.

Исследование Власова и Быковой подчеркивает возможность создания быстрых симуляторов сигналов ВИКПБ на основе эквивалентных моделей, которые могут быть дополнены ML-подходами для повышения точности и скорости [9]. Ранее Быкова разработала программный модуль для быстрого определения расстояния до границы пласта по данным ВИКПБ, что демонстрирует потенциал автоматизированных ML-решений для оперативного бурения [6]. Численное моделирование и инверсия данных электромагнитного каротажа, описанные в работе Эпова и др., показывают, что ML может сократить время вычислений и улучшить качество интерпретации по сравнению с традиционными методами [10]. Обучение нейронных сетей на кластерах ИВЦ НГУ, оснащенных высокопроизводительными процессорами и графическими ускорителями, позволило оптимизировать процесс тренировки моделей, обеспечив высокую вычислительную

эффективность. Однако успешное применение ML требует решения ряда технических задач, включая предобработку данных (например, обработку выбросов и нормализацию), выбор оптимальных алгоритмов и гиперпараметров, а также обеспечение вычислительной эффективности для работы в условиях ограниченных ресурсов.

Ключевые вызовы в применении методов ML для интерпретации данных ВИКПБ включают:

1. Точность предсказаний: Погрешность моделей должна быть сопоставима с погрешностью прибора (0.3 для фазовых измерений и 0.001 для амплитудных отношений), чтобы обеспечить надежность результатов [3, 7].
2. Оперативность: Время предсказания одной модели должно быть минимальным (не более 10 мс), чтобы соответствовать требованиям бурения в реальном времени [1].
3. Качество данных: Геофизические данные часто содержат выбросы и шум, что требует тщательной предобработки, например, логарифмического преобразования или удаления выбросов методом межквартильного размаха (IQR) [10].
4. Интеграция в производственные процессы: Разработанные модели должны быть адаптированы для работы в составе телеметрических систем, таких как «Луч», что требует оптимизации их размера и вычислительных ресурсов [3, 7].
5. Тестирование на реальных данных: Большинство исследований, включая упомянутые, проводилось на синтетических данных. Проверка моделей на реальных данных остается важной задачей для подтверждения их практической применимости [9, 10].

Перспективы дальнейших исследований связаны с разработкой более устойчивых и универсальных моделей ML, способных работать в сложных геологических условиях, включая тонкослоистые коллекторы. Ансамблевые методы, такие как Stacking Ensemble, и оптимизированные алгоритмы градиентного бустинга (например, LightGBM) показывают высокую эффективность, но требуют дальнейшей адаптации для реальных буровых систем [5]. Обучение моделей на высокопроизводительных кластерах ИВЦ НГУ позволяет значительно ускорить процесс тренировки и тестирования моделей, что критически важно для обработки больших объемов геофизических данных [6]. Тестирование моделей на реальных данных ВИКПБ, а также их интеграция в производственные процессы, позволят повысить точность и оперативность бурения, что имеет значительный экономический эффект. Развитие вычислительных ресурсов, таких как использование облачных платформ (например, Google Colab) в дополнение к кластерам ИВЦ НГУ, и оптимизация алгоритмов для работы на ограниченных устройствах открывают новые возможности для практического внедрения.

Таким образом, применение методов машинного обучения, включая нейронные сети, обученные на кластерах ИВЦ НГУ, для анализа сигналов ВИКПБ является перспективным направлением, способным преодолеть ограничения традиционных подходов. Однако для достижения полной практической применимости необходимы дальнейшие исследования, направленные на тестирование моделей в реальных условиях, улучшение их устойчивости к шуму и оптимизацию для производственных нужд.

1. Даниловский К.Н., Москваев И.А. Эволюция приборов электромагнитного каротажа в процессе бурения (по материалам отечественной и зарубежной литературы). *Геофизические технологии*, 2022, №4, с. 4–22.
<https://doi.org/10.18303/2619-1563-2022-4-4>.
2. Антонов Ю.Н. Изопараметрическое каротажное зондирование (обоснование ВИКИЗ). *Геология и геофизика*, 1980, №6, с. 81–91.
3. Еремин В.Н., Волканин Ю.М., Тарасов А.В. Аппаратурно-методическое обеспечение электромагнитного каротажа в процессе бурения. *Каротажник*, 2013, вып. 8 (226), с. 62–69.
4. Яновская Т.Б., Порохова Л.Н. Обратные задачи геофизики. СПб., Изд-во СПб. ун-та, 2004, 216 с.
5. Информационные технологии. Научный инжиниринг: Материалы 62-й Междунар. науч. студ. конф. 17–23 апреля 2024 г. / Новосиб. гос. ун-т. — Новосибирск: ИПЦ НГУ, 2024. — 358 с. Быстрая симуляция гальванических и электромагнитных сигналов в одномерных моделях околоскважинного пространства с помощью нейронных сетей / А.К. Бобровский, с. 109.
6. Быкова Г.П. Программный модуль быстрого определения расстояния до границы пласта в горизонтальном окончании скважины по данным высокочастотного электромагнитного каротажа в процессе бурения. *Информационные технологии. Цифровые технологии нефтегазовой индустрии: Материалы 60-й Междунар. науч. студ. конф. 10–20 апреля 2022 г.* / Новосиб. гос. ун-т. — Новосибирск: ИПЦ НГУ, 2022, с. 82.
7. Еремин В.Н., Волканин Ю.М., Тарасов А.В. Аппаратурно-методическое обеспечение электромагнитного каротажа в процессе бурения. *Каротажник*, 2013, вып. 8 (226), с. 62–69.
8. Карпов, К.А. Строительство нефтяных и газовых скважин: учебное пособие для вузов / К.А. Карпов. — 4-е стер. — Санкт-Петербург: Лань, 2021. — 188 с. — ISBN 978-5-8114-8671-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/179154> (дата обращения: 12.12.2024). Режим доступа: для авториз. пользователей.
9. Власов А.А., Быкова Г.П. Исследование эквивалентных моделей быстрым симулятором сигналов модуля высокочастотного индукционного каротажа в процессе бурения телеметрической системы "Луч" для горизонтальных скважин. *Геология, геофизика и разработка нефтяных и газовых месторождений*, 2024, №6 (390), с. 25–33.
10. Эпов М.И., Глинских В.Н., Сухорукова К.В., Никитенко М.Н., Еремин В.Н. Численное моделирование и инверсия данных электромагнитного каротажа в процессе бурения и шаблонирования нефтегазовой скважины. *Геология и геофизика*, 2015, т. 56, №8, с. 1520–1529.

4.3 Подробное описание работы, включая используемые алгоритмы

В рамках исследования разработан программный модуль на основе методов машинного обучения (ML) для количественной интерпретации сигналов высокочастотного индукционного каротажа (ВИКПБ) в горизонтальных нефтегазовых скважинах. Целью

работы являлось решение прямых и обратных задач геофизики для определения параметров окружающей среды, таких как удельное электрическое сопротивление пласта (R_{sfV}), вмещающего пласта (R_{fv}), расстояние до границы слоя (H) и зенитный угол ($Zenith$). Ниже описаны основные этапы работы, включая обработку данных, примененные алгоритмы и вычислительные ресурсы.

Для обучения моделей использовались два синтетических набора данных: `palette.csv` (1 073 664 записи), представляющий структурированную сетку параметров среды и сигналов зондов ВИКПБ (PL , PH , AN , AL), и `random.csv` (1 048 576 записей), содержащий рандомизированную выборку аналогичных данных. Предобработка данных включала следующие шаги:

- Логарифмическое преобразование численных признаков (R_{sfV} , R_{fv} , H , сигналы f) для стабилизации дисперсии и уменьшения влияния больших значений.
- Удаление выбросов методом межквартильного размаха (IQR) для повышения качества данных.
- Кодирование категориального признака $Zenith$ с использованием one-hot encoding для совместимости с ML-алгоритмами.
- Разделение данных на обучающую (80%) и тестовую (20%) выборки для последующей оценки моделей.

Обучение моделей проводилось на высокопроизводительных вычислительных кластерах Института вычислительного центра Новосибирского государственного университета (ИВЦ НГУ), оснащенных процессорами Intel Xeon и графическими ускорителями NVIDIA, что позволило обрабатывать большие объемы данных и проводить масштабные эксперименты с нейронными сетями. Дополнительно использовалась облачная платформа Google Colab с доступом к графическим процессорам NVIDIA Tesla T4, обеспечивающая гибкость при настройке гиперпараметров и тестировании моделей. Программная реализация осуществлялась на языке Python с использованием библиотек Pandas и NumPy для обработки данных, Scikit-learn для базовых функций машинного обучения, LightGBM, XGBoost и CatBoost для алгоритмов градиентного бустинга, а также Keras и TensorFlow для разработки нейронных сетей.

Для решения задачи интерпретации сигналов ВИКПБ исследовались два основных типа ML-алгоритмов: нейронные сети и методы градиентного бустинга. Также применялся ансамблевый подход Stacking Ensemble для комбинирования предсказаний. Сравнение производительности алгоритмов представлено в таблице 1.

Таблица 1 – сравнительная таблица столбцов с полученными данными после обучения

модель	Точность предсказания					Дополнительные характеристики	
	время предсказания одной модели (сек)	MAE	максимальная абсолютная ошибка	95-процентиль	процент данных, на которых ошибка меньше погрешности прибора (0.3 для фаз, 0.001 для амплитуд)	память, занимаемая моделью (МВ)	время обучения (сек)
LightGBM	0.00065	0.0503	8.1819	0.2168	96.79	27.9	3603.71
XGBoost	0.00037	0.0840	9.9303	0.5036	93.48	19	3034.89
CatBoost	0.000006	0.1365	14.3286	0.5887	89.21	11.4	1168.78
Gradient Boosting	0.004995	0.0869	0.5238	0.3519	93.52	2977.12	480.10
бустинг с ансамблем (Stacking Ensemble)		0.3005		0.5161	89.30	2588.01	1080.20

На начальном этапе были разработаны и обучены несколько полносвязных нейронных сетей с 3–5 скрытыми слоями, содержащими от 64 до 256 нейронов на слой. В скрытых слоях использовалась функция активации ReLU, а в выходном слое — линейная функция для регрессии численных значений сигналов f . Модели минимизировали среднюю абсолютную ошибку (MAE) с помощью оптимизатора Adam, с размером пакета (batch size) 128–512 и обучением в течение 50–100 эпох с применением механизма ранней остановки для предотвращения переобучения. Обучение нейронных сетей проводилось на кластерах ИВЦ НГУ, что позволило использовать вычислительные ресурсы для обработки больших объемов данных.

Для достижения высокой точности и оперативности тестировались три алгоритма градиентного бустинга: LightGBM, XGBoost и CatBoost. Гиперпараметры оптимизировались с использованием поиска по сетке (Grid Search) для максимизации точности и минимизации времени предсказания:

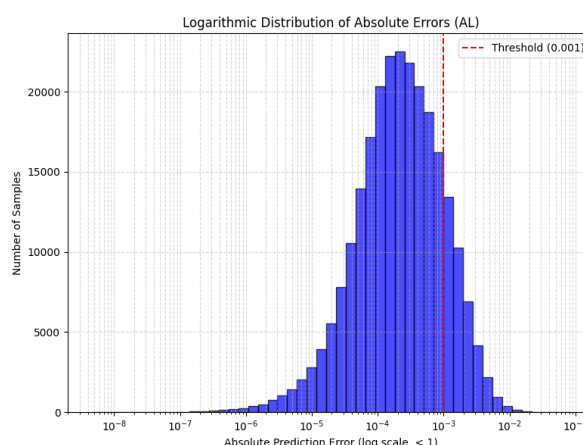
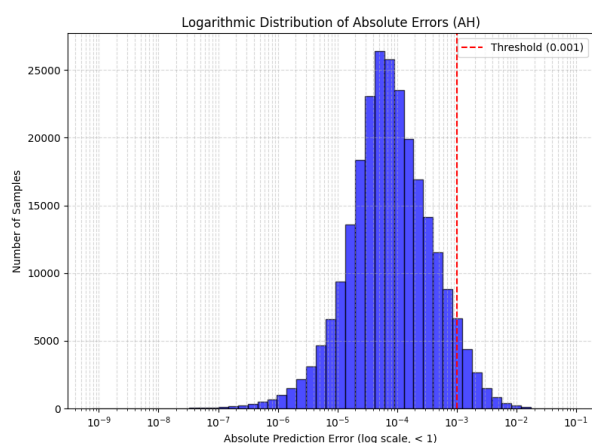
- LightGBM: max_depth=7, learning_rate=0.1, n_estimators=1000, num_leaves=31. Алгоритм использовал гистограммный подход для ускорения обучения и обработки категориальных признаков.

- XGBoost: max_depth=6, learning_rate=0.05, n_estimators=800. Отличался встроенной регуляризацией и способностью работать с разреженными данными.
- CatBoost: depth=6, learning_rate=0.1, iterations=1000. Обеспечивал автоматическую обработку категориальных признаков и устойчивость к переобучению.

Для повышения точности применялся ансамблевый метод Stacking Ensemble, комбинирующий предсказания LightGBM, XGBoost и CatBoost. В качестве мета-модели использовалась линейная регрессия. Процесс включал обучение базовых моделей на обучающей выборке, генерацию предсказаний как новых признаков и последующее обучение мета-модели для получения финального прогноза.

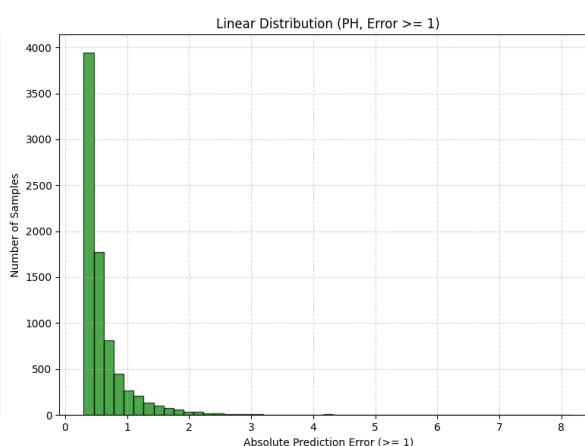
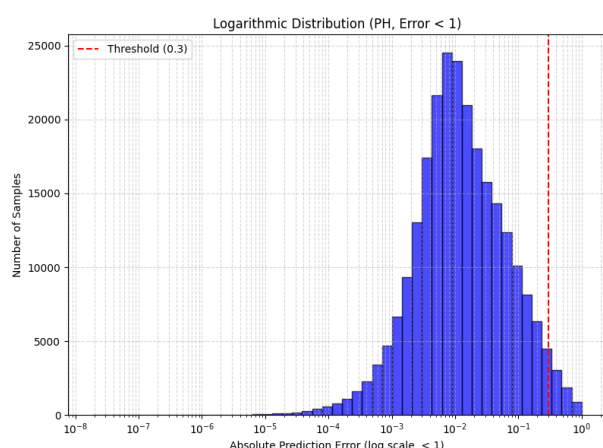
4.4 Полученные результаты

После обучения моделей LightGBM для всех зондов (АН, АЛ, РН и PL), финальные результаты предсказаний представлены на рисунке 2. Эти модели были протестированы на синтетических данных, отражающих геологические условия Западной Сибири.

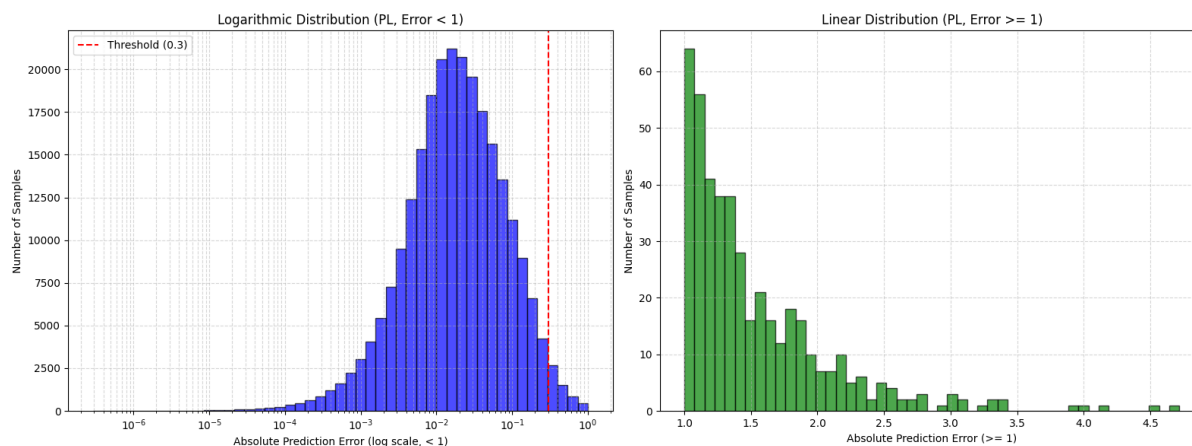


Абсолютные ошибки предсказания для АН

Абсолютные ошибки предсказания для АЛ



Абсолютные ошибки прогнозирования (логарифмическая и линейная шкалы) для РН



Абсолютные ошибки прогнозирования (логарифмическая и линейная шкалы) для PL

Рисунок 1 — абсолютные ошибки прогнозирования (логарифмическая и линейная шкалы) для PH, PL, AH и AL

После обучение моделей для их апробации был создан программный модуль на языке программирования Python, решающий обратную задачу с помощью алгоритма, описанного в [1]. Палеточная аппроксимация сигналов была заменена полученными моделями машинного обучения. Далее этот программный модуль был запущен на синтетических данных, для которых траектория и модель околоскважинного пространства представлены на рисунке 5.

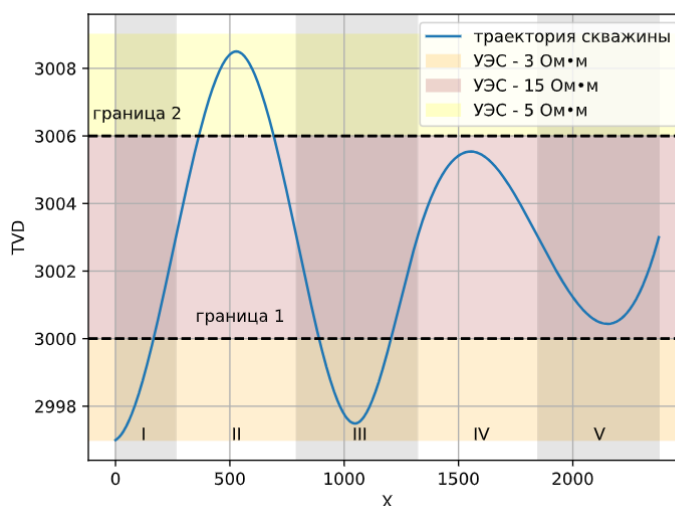


Рисунок 2 – траектория бурения и модель околоскважинного пространства [1]

Процесс тестирования был начат с подготовки синтетического набора данных, примером которого является *synthetic.csv*, содержащий наблюдаемые значения для PL, AL, AH и PH. Эти параметры представляют собой фазовые и амплитудные характеристики, полученные с помощью численного моделирования для синтетической модели, и служат в качестве входных данных для проверки обученных моделей. Прогностическую способность

решения обратной задачи с использованием обученных моделей планировалось проверить с использованием этих данных.

При решении обратной задачи использовалась численная инверсия, представленная как минимизация целевой функции, которая количественно определяет расхождение между наблюдаемыми и смоделированными (полученными с помощью моделей машинного обучения) сигналами. Следуя подходу, описанному в литературе, целевая функция J определяется как среднеквадратическая ошибка по n точкам данных:

$$J = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} \left[\frac{S_i^{\text{exp}} - M_i(m)}{\text{Err}(S_i^{\text{exp}})} \right]^2} \quad (1)$$

Где:

- St^{exp} - наблюдаемый сигнал (преобразованный из кажущегося сопротивления в амплитудное положение или разность фаз),
- $M_i(m)$ -это смоделированный сигнал для набора параметров $m = [Rsfv, Rfv, H, Zenith]$,
- $Err(St^{\text{exp}})$ это ошибка, определяемая как:
 $Err(x) = 0.034x + 0.0547$ для разности фаз x ,
 $Err(x) = 0.337(1 - x) + 0.0013$ для амплитудного отношения x .

Оптимизация была выполнена с использованием метода Нелдера-Мида из `scipy.optimize.minimize`, алгоритм без производных, подходящий для подхода деформируемого многогранника. Метод итеративно корректирует m для минимизации J , начиная с начального предположения $m_0 = [R^0sfv, R^0fv, H^0, Zenith^0]$.

Начальные параметры определялись графиком траектории, который отображает истинную вертикальную глубину (TVD) в зависимости от расстояния x , со слоями удельного сопротивления 3 Ом·м (0–1000), 15 Ом·м (1000–1500) и 5 Ом·м (1500–2000), ограниченными TVD = 3000 м и 3006 м.

Полученные результаты близки к ожидаемым. На рисунке 3 изображены графики предсказанного (синяя) и ожидаемого (оранжевая) Rfv, на рисунке 4 – Rsfv, на рисунке 5 – H. Видно, что Rfv, Rsfv, H близки к ожидаемым, в отличие от Zenith. Однако Zenith слабо влияет на сигналы, поэтому такой результат ожидаем. На рисунке 6 изображен график целевой функции подобранной модели. Модель считается подобранной, если целевая функция меньше 1. Видно, что практически во всех точках это целевая функция меньше 1.

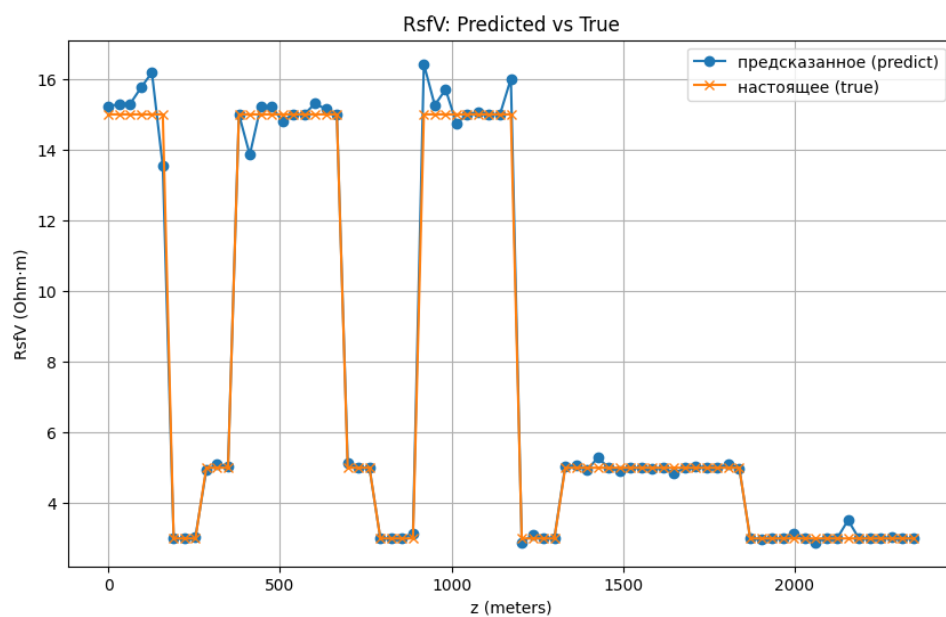


Рисунок 3 – Сравнение предсказанного и настоящего Rsfv

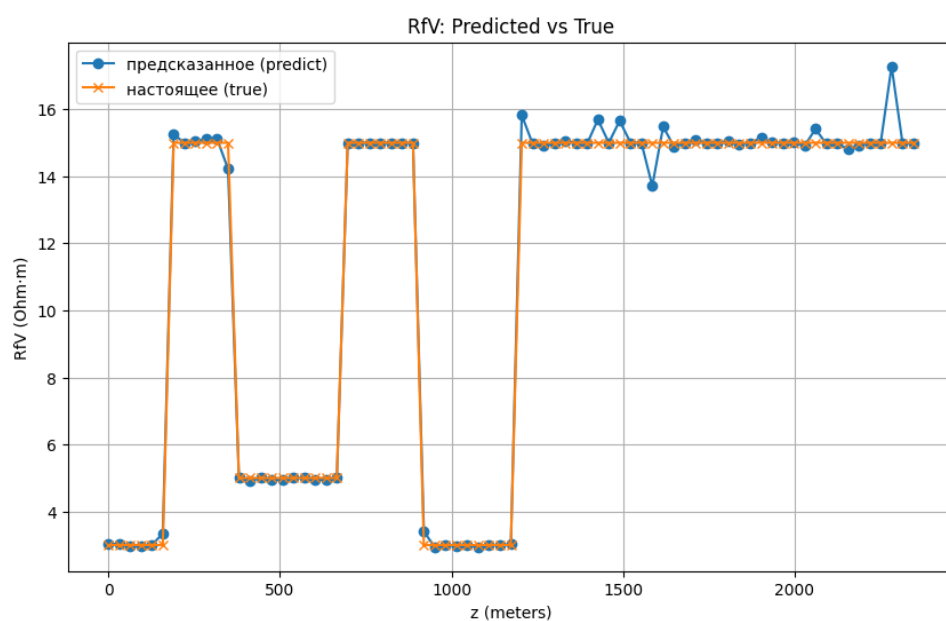


Рисунок 4 – Сравнение предсказанного и настоящего Rfv

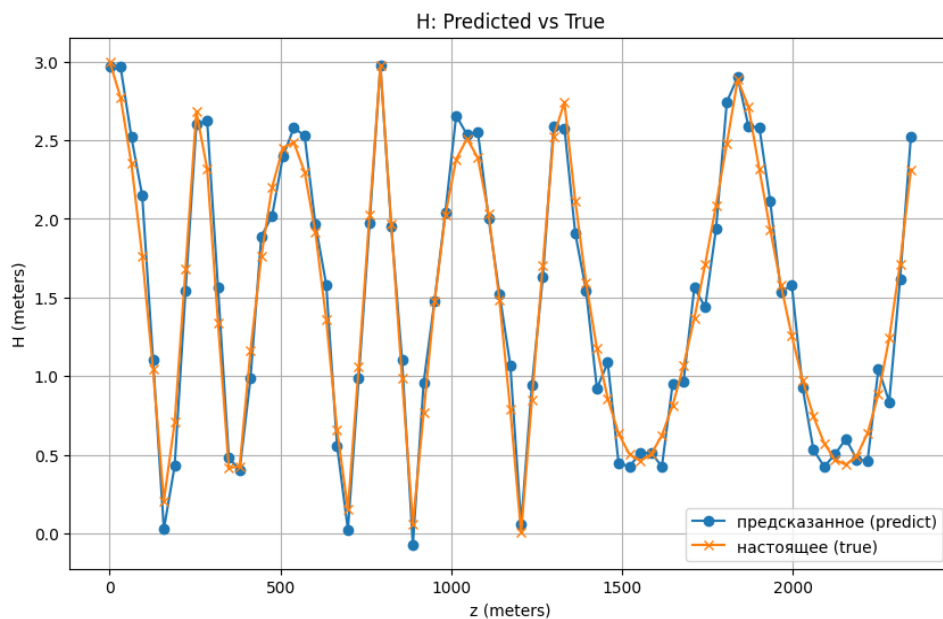


Рисунок 5 – Сравнение предсказанного и настоящего Н

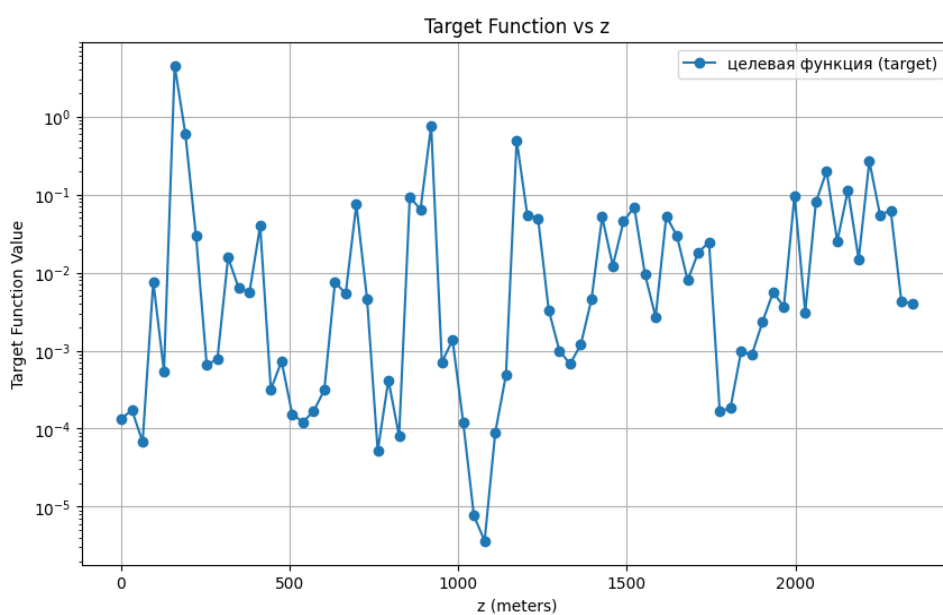


Рисунок 6 – Целевая функция предсказанной модели

1. Власов А.А., Быкова Г.П. Исследование эквивалентных моделей быстрым симулятором сигналов модуля высокочастотного индукционного каротажа в процессе бурения телеметрической системы "Луч" для горизонтальных скважин. *Геология, геофизика и разработка нефтяных и газовых месторождений*, 2024, №6 (390), с. 25–33.

5. Эффект от использования кластера в достижении целей работы

Использование вычислительного кластера ИВЦ НГУ сыграло важную роль в достижении целей данной работы. Благодаря высокой производительности процессоров и доступу к графическим ускорителям удалось эффективно обучить модели машинного обучения, включая несколько моделей глубоких нейронных сетей, модели градиентного бустинга и LightGBM, на большом объеме синтетических данных. Это существенно сократило время обучения и повысило стабильность вычислительных процессов, особенно при настройке гиперпараметров и тестировании различных конфигураций моделей. Кластер также обеспечил возможность параллельной обработки данных, что было критически важно для реализации численной инверсии и оценки качества предсказаний. В результате были получены точные и надежные модели для интерпретации сигналов ВИКПБ, готовые к применению в условиях реального бурения.

6. Перечень публикаций, содержащих результаты работы

1. Работа была представлена на 63-й Международной научной студенческой конференции в НГУ с 16 по 22 апреля 2025 года в секции «Информационные технологии», подсекции «Нейронные сети и машинное обучение». Масуд Рофикул Ал. «Использование средств машинного обучения для моделирования сигналов прибора ВИКПБ при горизонтальном бурении». Тезисы опубликованы в сборнике конференции МНСК-2025.
2. Быкова Галина Павловна, Власов Александр Александрович, Масуд Рофикул Ал., Государственная регистрация программы для ЭВМ «БИС — быстрая интерпретация данных высокочастотного электромагнитного зондирования». Свидетельство № 2024687640