

УДК 550.837.82:004.85

Электроразведка методом ЗСБ и машинное обучение: новые возможности для прогнозирования геологических рисков

■ Давыденко С.А.^{1,2}, Поляков Д.В.¹, Мисюркеева Н.В.^{1,3}, Агафонов Ю.А.¹, Селяев В.А.¹, Агафонов Е.А.¹

¹ ООО «СИГМА-ГЕО», г. Иркутск.

² Институт солнечно-земной физики СО РАН, г. Иркутск.

³ Институт земной коры СО РАН, г. Иркутск.

Аннотация. В работе представлен комплексный подход к прогнозированию геологических рисков на основе совместного анализа данных электроразведки методом зондирования становлением поля в ближней зоне (ЗСБ), сейсморазведки 3D и скважинной информации с применением методов машинного обучения с учителем. Показана эволюция методов машинного обучения в геофизике — от первых нейросетевых моделей до современных ансамблевых алгоритмов, таких как случайный лес (Random Forest). Подробно описана методология прогнозирования, включающая формирование многомерного признакового пространства (геоэлектрические параметры, их градиенты, структурные характеристики), подготовку эталонных выборок по скважинам, обучение моделей и оценку качества с использованием ООВ-оценки и кросс-валидации с группировкой по скважинам. Все вычислительные процедуры реализованы в специализированном программном комплексе X-Search (разработка ООО «СИГМА-ГЕО»), обеспечивающем сквозной цикл обработки и интерпретации данных. Эффективность подхода продемонстрирована на двух практически значимых задачах: прогнозирование осложнений бурения (поглощений бурового раствора) на газоконденсатном месторождении в Восточной Сибири и прогнозирование распространения многолетнемерзлых пород (ММП) на участках в Якутии. Для поглощений достигнута ООВ-точность 87,53%, для ММП — 98,13%. Показано, что существенный вклад в прогноз вносят геоэлектрические параметры (проводимость и её градиент), что подтверждает высокую чувствительность электромагнитных методов к флюидонасыщению и трещиноватости. Разработанный подход может быть рекомендован для использования в производственных геофизических и геологических организациях при решении широкого круга прогнозных задач.

Ключевые слова: электроразведка, зондирование становлением поля (ЗСБ), машинное обучение (ML), случайный лес (Random Forest), прогнозирование осложнений бурения, поглощения бурового раствора, многолетнемерзлые породы (ММП), комплексная интерпретация, программный комплекс X-Search, нефтегазовая геология, геокриология.

Актуальность

Геофизические исследования являются ключевым инструментом при изучении строения земной коры, поиске и разведке полезных ископаемых, а также при решении инженерно-геологических задач. За последние десятилетия объёмы получаемых данных колоссально выросли: сейсморазведка 3D, электроразведка, грави- и

магниторазведка, результаты исследования скважин, генерируют терабайты информации. Однако традиционные методы интерпретации, основанные на эмпирических корреляциях и упрощённых аналитических моделях, часто не справляются с нелинейностью, многомерностью и гетерогенностью геологических сред.

В этих условиях методы машинного

обучения (МО) становятся не просто полезным дополнением, а необходимым инструментом, позволяющим выявлять скрытые закономерности, строить прогнозные модели с высокой точностью и количественно оценивать неопределённости. Особую ценность МО представляет при комплексировании данных различных геофизических методов, когда необходимо совместно анализировать электрические, сейсмические, плотностные и магнитные характеристики среды [6].

Таким образом, решение перечисленных задач требует разработки и практической апробации комплексного подхода, базирующегося на совместном анализе данных электроразведки методом зондирования становлением поля в ближней зоне (ЗСБ) и сейсморазведки с применением методов машинного обучения с учителем (в частности, алгоритма случайного леса). Предлагаемый подход не является данью современным тенденциям, а представляет собой объективную необходимость, обусловленную сложностью и многомерностью геолого-геофизической информации.

Особое внимание уделено эволюции методов машинного обучения в геофизике, методологии прогнозирования на основе случайного леса, а также перспективам дальнейшего развития интеллектуальных технологий — от глубокого обучения до создания цифровых двойников месторождений.

Эволюция методов машинного обучения в геофизике

Применение методов искусственного интеллекта в геофизике имеет более чем полувековую историю. 1960–1980-е годы: появление персептрона [9] и алгоритма обратного распространения ошибки [10] заложило теоретические основы. В геофизике нейронные сети применялись для классификации типов пород по данным каротажа и интерпретации сейс-

мических атрибутов. 1990-е годы: активно развиваются методы опорных векторов [11], ансамблевые методы — бэггинг и бустинг. Наряду с методами обучения с учителем развиваются и методы кластеризации без учителя, в частности алгоритм k-means++ [2], который находит применение для выделения однородных групп объектов по совокупности геофизических признаков [6]. Появляются первые коммерческие системы автоматической интерпретации сейсмических данных. 2000-е годы: предложен алгоритм случайного леса [1], быстро завоевавший популярность благодаря устойчивости к переобучению, встроенной оценке важности признаков и способности работать с разнородными данными. 2010-е годы — настоящее время: эпоха глубокого обучения. Многослойные нейронные сети, свёрточные сети, рекуррентные сети (LSTM) [12,13] находят широкое применение в обработке сейсмических кубов, автоматическом выделении горизонтов, инверсии электромагнитных данных. В бурении активно внедряются системы прогнозирования осложнений в реальном времени [3,5]. Прогнозирование зон геологических осложнений при бурении с использованием искусственного интеллекта на основе скважинных и сейсморазведочных данных [4].

Методология прогнозирования на основе случайного леса

Случайный лес (Random Forest) — ансамблевый алгоритм машинного обучения, формирующий множество решающих деревьев на основе бутстрап-выборок и случайного подмножества признаков на каждом разбиении. Ключевые преимущества метода применительно к геологическим задачам: устойчивость к переобучению, способность работать с разнородными данными без сложной предобработки, встроенная оценка важности признаков и

внутренняя ООВ-оценка (Out-of-Bag), не требующая отдельной валидационной выборки.

Все вычислительные процедуры — от формирования многомерного признакового пространства и подготовки эталонных выборок по скважинам до обучения моделей и построения прогнозных 3D-кубов — реализованы в специализированном программном комплексе X-Search (разработка ООО «СИГМА-ГЕО»). Комплекс обеспечивает сквозной цикл обработки и интерпретации геофизических данных: загрузку исходных скважинных материалов и наборов признаков, представленных в виде кубов как с регулярной, так и с нерегулярной сеткой, обучение моделей и построение прогнозных трёхмерных моделей. Архитектура комплекса предусматривает расширяемый функционал — набор алгоритмов, методов оценки качества и форматов данных может пополняться под новые задачи, а активная разработка ведётся в направлении интеграции дополнительных геологических и геофизических параметров, развития средств визуализации и автоматизации интерпретацион-

ной цепочки.

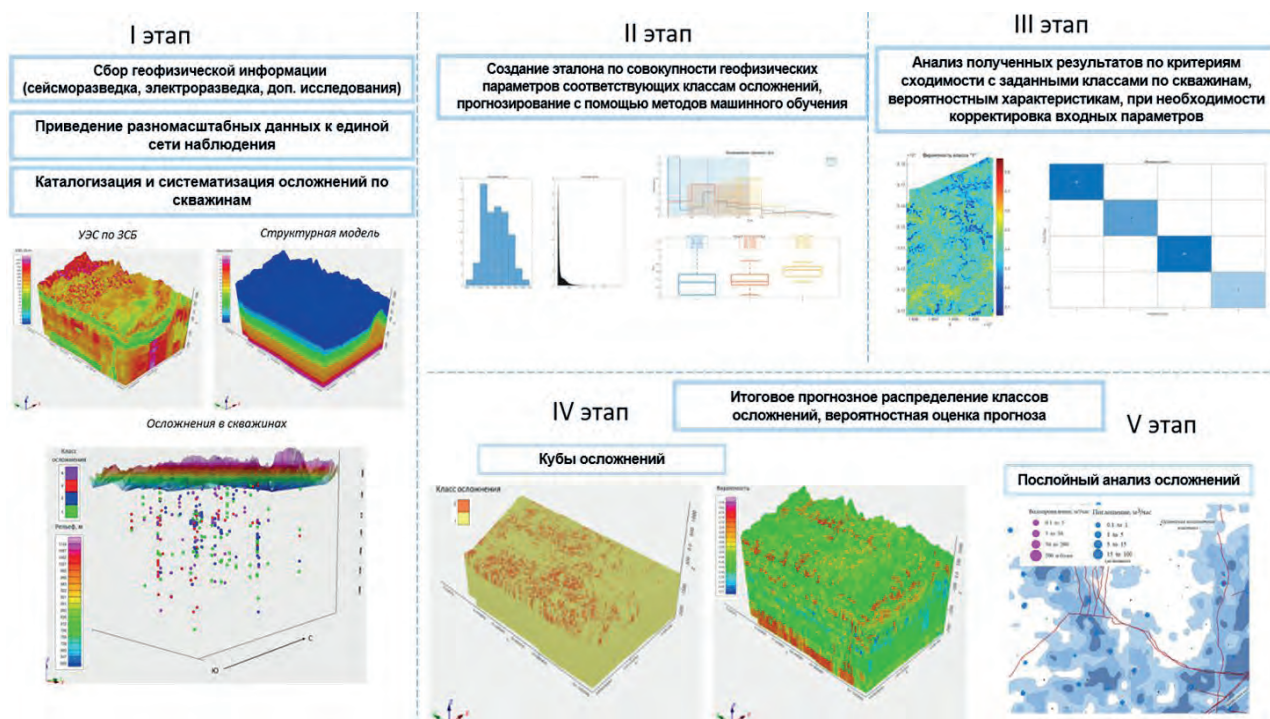
В основе двух представленных примеров лежат данные электроразведочных работ методом ЗСБ, площадных 3D-сейсморазведочных работ и скважинная информация. Совместное использование этих данных позволяет получать:

- распределение удельного электрического сопротивления (УЭС) по данным ЗСБ;
- структурный каркас по данным сейсморазведки (возможно также привлечение сейсмических атрибутов);
- скважинную информацию об искоемых объектах (интервалах осложнений, глубине подошвы ММП и т.д.).

Обработка и инверсия данных ЗСБ выполнялись с использованием программного обеспечения TEM-Processing и Model 5 [7, 8]. Сейсмические кубы, структурные построения и данные бурения использовались для формирования признакового пространства и валидации прогнозных моделей.

Схематично процесс прогнозирования, проиллюстрированный на примере горно-геологических условий (ГГУ), показан на Рис 1.

▼ Рис. 1. Общая схема прогнозирования ГГУ с помощью машинного обучения.



Обобщенная последовательность действий при прогнозировании ГГУ:

1. Систематизация и классификация: унификация всех известных осложнений по типам в поисковых, поисково-оценочных и разведочных скважинах, а также кустовым площадкам, пробуренным на территории исследования.

2. Интеграция данных в единое геоинформационное пространство: приведение разнородных геофизических данных (сейсморазведка, электроразведка) к общей 3D-сетке с равным шагом.

3. Формирование обучающей выборки: выделение эталонных значений геофизических параметров в точках скважин, где были зафиксированы осложнения.

4. Геостатистический анализ и прогноз: применение алгоритмов машинного обучения для построения прогнозной модели распространения осложнений/других параметров в межскважинном пространстве на основе эталонной выборки.

5. Анализ распределения вероятностей: оценка неопределенности и достоверности полученного прогнозного решения.

6. Верификация модели: проверка полученных результатов на соответствие критериям сходимости и известным данным.

7. Построение итоговых прогнозных карт в целевых интервалах исследования

Формирование признакового пространства

На основе априорных геологических представлений и работ предыдущих исследователей [14,15], был сформирован набор признаков для задачи прогнозирования поглощений бурового раствора: удельное электрическое сопротивление ($\text{Ом}\cdot\text{м}$), проводимость (См), градиенты (градиент проводимости, градиент УЭС, градиент кровли, градиент толщин) —

характеризуют зоны контрастных изменений, связанные с тектоническими нарушениями, фациальными переходами, трещиноватостью.

Для каждого узла сетки был сформирован набор признаков на основе данных малоглубинных исследований ЗСБ, загружены скважинные данные. При прогнозировании ММП были использованы следующие признаки: удельное электрическое сопротивление, проводимость, градиент проводимости, абсолютная глубина.

Процесс включает загрузку скважинных данных, извлечение значений признаков из 3D-кубов в точках инклинометрии и обучение модели с оптимизацией гиперпараметров. Для оценки прогнозной способности модели и выбора оптимальных гиперпараметров используется кросс-валидация с группировкой по скважинам (Group K-Fold). Это означает, что данные одной скважины (все точки её инклинометрии) целиком попадали либо в обучающую, либо в валидационную подвыборку.

Такой подход исключает утечку данных (data leakage), когда пространственно близкие точки из одной скважины могли бы одновременно оказаться и в обучении, и в контроле. Количество фолдов: 5. Для каждой задачи выполнялось 5 последовательных разбиений, после чего вычислялись среднее значение метрики и стандартное отклонение.

После выбора оптимальных гиперпараметров итоговая модель обучается на 100% доступных скважинных данных и используется для построения прогноза с использованием всей имеющейся информации.

Оценка качества классификационных моделей выполняется с использованием точности (Accuracy), матрицы неточностей (Confusion Matrix), ООВ-точности и кросс-валидации.

В данном случае приведена средняя

точность (Accuracy) по фолдам кросс-валидации. Формат записи «значение $\pm$ отклонение» означает: значение — средняя точность по 5 фолдам; отклонение — стандартное отклонение точности между фолдами. Например, для задачи ММП: $94,02\% \pm 1,64\%$ означает, что средняя точность составила $94,02\%$, а стандартное отклонение — $1,64$ процентных пункта. Небольшое отклонение свидетельствует о стабильности модели на разных подвыборках.

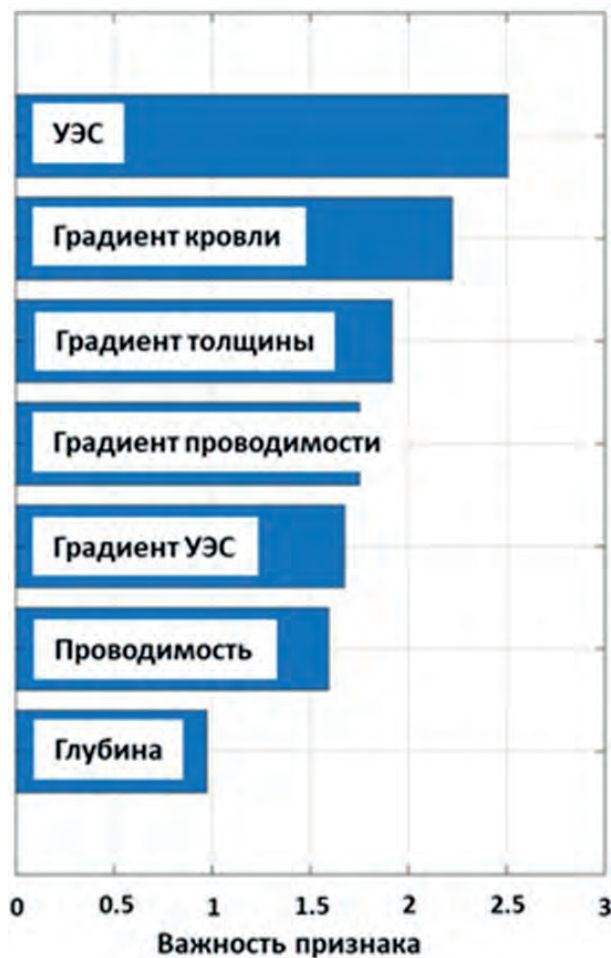
Оценка качества регрессионных моделей выполняется с использованием MAE, RMSE и Bias, а также кросс-валидации и OOB-оценки.

Способ расчёта важности признаков: важность признаков рассчитывалась с использованием встроенного в Random Forest показателя Mean Decrease in Impurity (MDI) — среднего уменьшения неопределённости (критерия Джини) по всем деревьям ансамбля.

Для каждого признака суммировались уменьшения неопределённости во всех узлах, где он использовался для разбиения, с последующей нормировкой на общее количество деревьев. В задаче прогнозирования поглощений суммарный вклад признаков, производных от данных ЗСБ (проводимость S и её полный градиент), вносит существенный вклад в итоговый результат (Рис. 2).

Для задач с учителем формируется обучающая выборка, где для каждой точки 3D-сетки известна метка класса. Для осложнений бурения метка определяется по интервалам с зафиксированными осложнениями в скважинах. Для ММП — по глубине подошвы ММП: выше подошвы — класс 1 (ММП), ниже — класс 0 (отсутствие ММП).

В задаче прогнозирования поглощений на газоконденсатном месторождении использованы данные по более чем 100 скважинам (поисково-разведочным и эксплуатационным с кустовыми ответ-



▲ Рис. 2. Гистограмма оценки вклада каждого из признаков в результат прогнозирования осложнений ГГУ.

влениями). Соотношение классов (наличие/отсутствие поглощения) в выборке является несимметричным: доля интервалов с зафиксированными поглощениями составляет порядка 10–15% от общего числа точек. Для борьбы с дисбалансом классов применялась стратифицированная разбивка выборки и настройка весов классов в алгоритме Random Forest.

В задаче прогнозирования ММП использованы кубы УЭС и проводимости (сетка $300 \times 300 \times 10$ м), данные бурения по 31 скважине с зафиксированной глубиной подошвы ММП. Формирование классов выполнялось следующим образом: каждая точка вертикального столбца (с шагом 10 м) относилась к классу 1 (ММП), если её абсолютная глубина была выше подошвы ММП, и к классу 0 (отсутствие ММП) — если ниже. Таким

образом, задача сводилась к бинарной классификации каждой точки, а поверхность подошвы ММП строилась как граница между классами.

Прогнозирование осложнений бурения на газоконденсатном месторождении

Рассматриваемое газоконденсатное месторождение — одно из крупнейших на востоке России. Разрез представлен сложным чередованием карбонатных, терригенных и галогенных пород кембрия с развитием зон аномально высокого пластового давления (АВПД). При бурении часто возникают поглощения бурового раствора. Использованы данные сейсморазведки 3D ($>5000 \text{ км}^2$), электроразведки ЗСБ ($>3000 \text{ км}^2$) и данные по осложнениям из >100 скважин.

По результатам прогнозирования построена модель бинарной классификации. Получены следующие показатели: ООВ-точность 87,53%, кросс-валидация $73,85\% \pm 2,49\%$.

Наибольший вклад в прогноз вносят проводимость и её градиент (суммар-

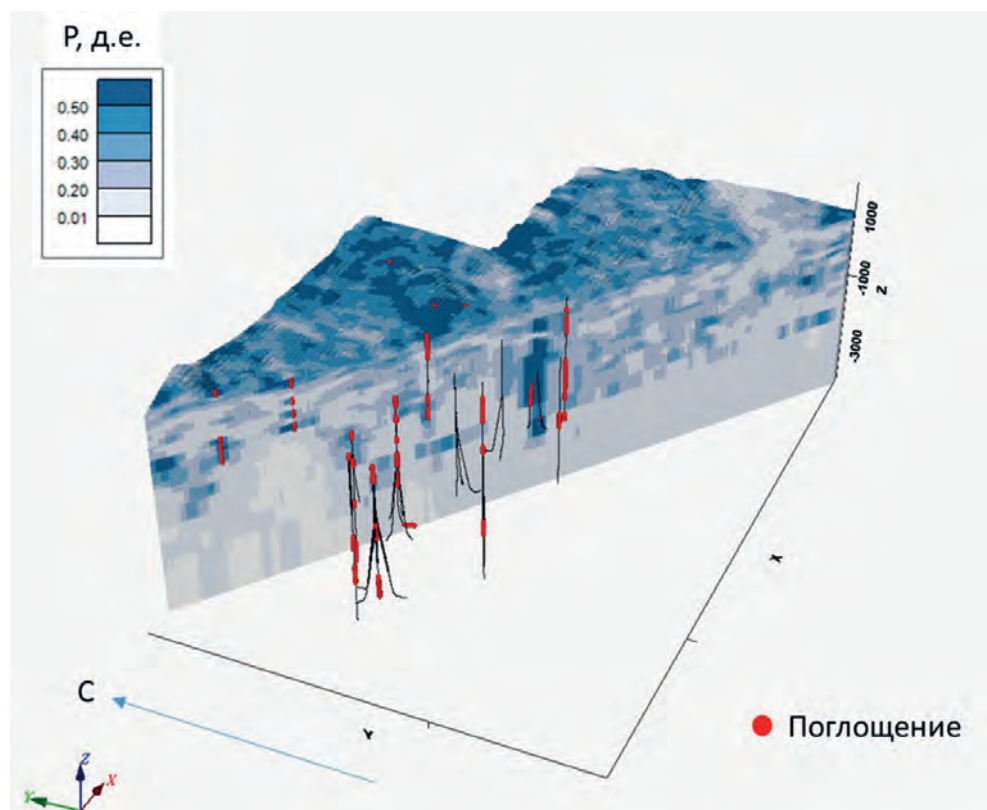
но $>60\%$), что подтверждает чувствительность электромагнитных методов к флюидонасыщению и трещиноватости. Прогнозные кубы вероятности позволяют оконтуривать зоны риска и оптимизировать траектории новых скважин (Рис 3.). Помимо поглощений, разработанная методика может быть адаптирована для прогнозирования газо- и рапопроявлений, что расширяет область её применения.

Прогнозирование распространения многолетнемерзлых пород (ммп)

Участки исследования расположены в Республике Саха (Якутия), где в 2016–2022 гг. выполнены работы методом ЗСБ в малоглубинном варианте. Основная задача — определение таликовых зон и подошвы ММП. Использованы кубы УЭС и проводимости (сетка $300 \times 300 \times 10 \text{ м}$), данные бурения по 31 скважине с зафиксированной глубиной подошвы ММП.

Получены следующие показатели: ООВ-точность 98,13%, кросс-валидация $94,02\% \pm 1,64\%$. Построена поверхность подошвы ММП двумя способами: по ми-

► Рис. 3.
Фрагмент
прогнозного куба
вероятности
осложнений по
поглощению с
визуализацией
скважин.



нимальной отметке 1-го класса (подошва ММП) и по максимальной отметке 0-го класса (кровля таликов).

Контрольные данные для расчёта MAE и RMSE: из 31 скважины часть (около 25%) была исключена из обучения и использовалась как независимая тестовая выборка. MAE (средняя абсолютная ошибка) и RMSE (среднеквадратическая ошибка) рассчитывались как разница между прогнозной глубиной подошвы

ММП (из построенной поверхности) и фактической глубиной, зафиксированной в скважине. Полученные значения: MAE = 7–12 м, RMSE = 10–18 м, что находится в пределах погрешности исходных геофизических данных.

Пример прогноза в формате плана и разреза, детализируют структуру ММП, выделяют зоны прерывистого и сплошного распространения, а также области сквозных таликов (Рис. 4).

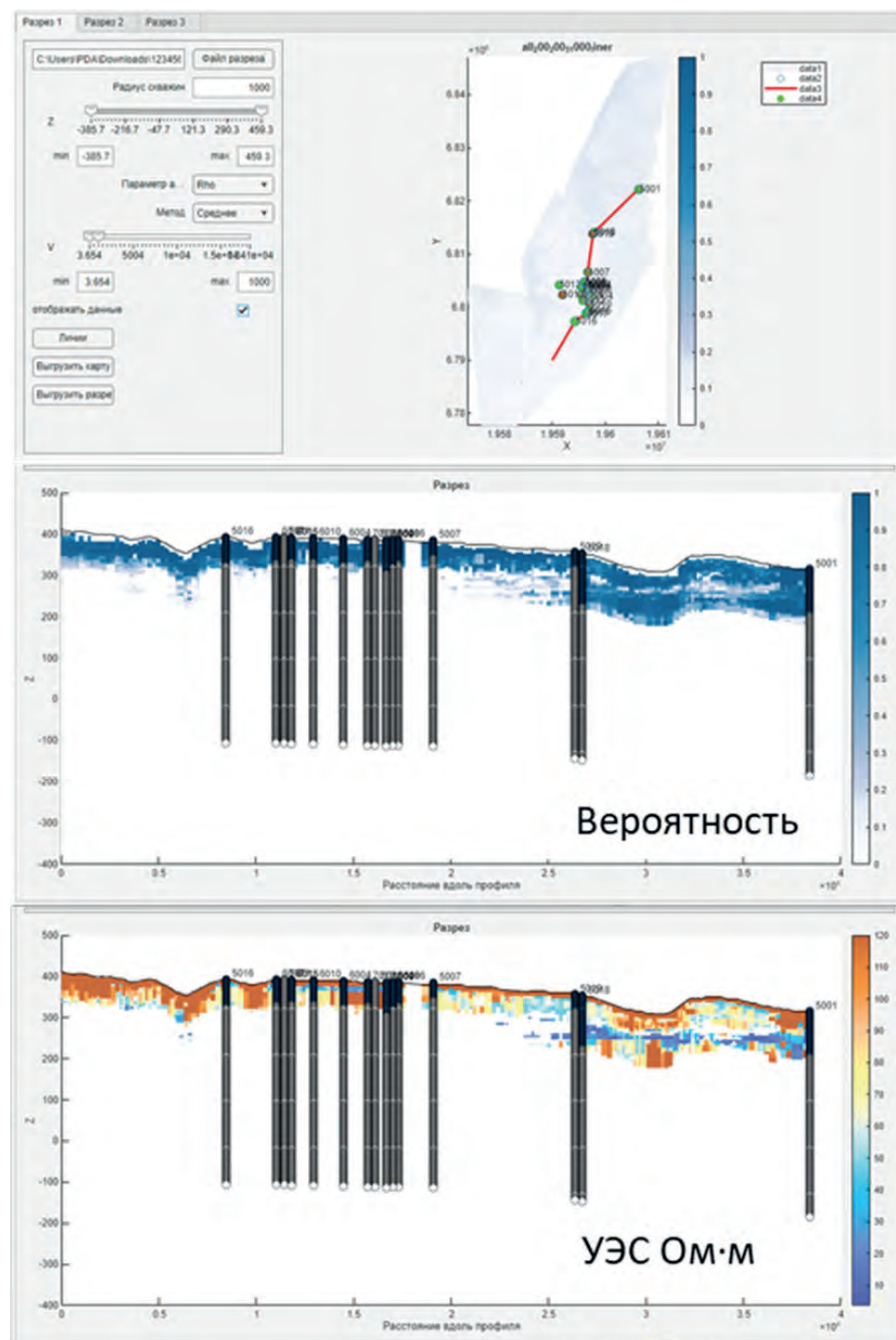


Рис. 4. Окно интерфейса программы с результатами прогнозирования распространения ММП, с визуализацией скважинных данных.

Обсуждение результатов и перспективы

Представленные примеры демонстрируют, что методы машинного обучения, реализованные в программном комплексе X-Search, повышают точность прогнозирования при комплексировании геофизических данных, дают независимый прогноз на основе скрытых зависимостей. Ключевые факторы успеха: многомерность признаков пространства (использование не только первичных параметров, но и их градиентов), ансамблевые методы (устойчивость к шуму и переобучению), интеграция разнородных данных.

Основные ограничения: дефицит эталонных данных, неоднородность данных (разновременные съёмки, разные методики инверсии), интерпретируемость сложных моделей.

Пути решения: использование трансферного обучения и полуавтоматической разметки, калибровка через опорные скважины, применение методов объяснимого ИИ.

Перспективы развития: интеграция геологических моделей (литология, пористость, давление); глубокое обучение для анализа 3D-кубов; прогнозирование в реальном времени с использованием буровых параметров; создание цифровых двойников месторождений; автоматизация всей цепочки интерпретации в рамках единого программного комплекса X-Search. Реализация этих направлений уже ведётся в рамках развития программного комплекса X-Search.

Выводы

Разработан и апробирован комплексный подход к прогнозированию геологических рисков на основе методов машинного обучения (случайный лес) с использованием данных электроразведки ЗСБ и сейсморазведки, данных скважин. Подход показал высокую эф-

фективность на двух практически значимых задачах. Все расчёты выполнены в специализированном программном комплексе X-Search, разработанном в ООО «СИГМА-ГЕО», который обеспечивает полный цикл обработки и интерпретации данных.

В задаче прогнозирования осложнений бурения по поглощениям достигнута ООВ-точность 87,5% и кросс-валидация $73,85\% \pm 2,49\%$. Наибольший вклад вносят геоэлектрические параметры (проводимость и её градиент, $>60\%$).

В задаче прогнозирования ММП получена ООВ-точность 98,13% и кросс-валидация $94,02\% \pm 1,64\%$. Ошибки по сравнению с эталонными скважинами (MAE = 7–12 м) находятся в пределах погрешности исходных данных.

Показана устойчивость алгоритма случайного леса к неоднородности исходных данных, связанной с разновременностью съёмок и различиями в методиках инверсии, что подтверждает его пригодность для работы в реальных производственных условиях.

Разработанный подход может быть рекомендован для использования в производственных геофизических и геологических организациях при решении широкого круга прогнозных задач.

Литература

1. Breiman L. Random Forests // Machine Learning. 2001. Vol. 45, No. 1. P. 5–32.
2. Arthur D., Vassilvitskii S. k-means++: The Advantages of Careful Seeding // SODA '07. 2007. P. 1027–1035.
3. Guoli Huang, Sarah Kanaan Hamzah, Data-driven prediction of rate of penetration (ROP) in drilling operations using advanced machine learning models // Scientific Reports. 2025. DOI: 10.1038/s41598-025-17588-9.
4. Бугаев А.С., Ерохин Г.Н., Рябых С.А., Смирнов А.С. Прогноз зон флюидо-рапопроявлений методами искусственно-

го интеллекта на основе сейсмических атрибутов нового поколения RTH и данных бурения на Ковыктинском ГКМ // Доклады РАН. Науки о Земле. 2024. Т. 514, № 1. С. 131–140.

5. Mohammad J Aljubran, Jothibas Ramasamy Deep Learning and Time-Series Analysis for the Early Detection of Lost Circulation Incidents During Drilling Operations // IEEE Access. 2021. DOI: 10.1109/ACCESS.2021.3082557.

6. Давыденко С.А., Мисюркеева Н.В., Давыденко А.Ю., Агафонов Ю.А. и др. Комплексный подход к интерпретации геофизических данных при уточнении модели кристаллического фундамента в Лено-Тунгусской нефтегазоносной провинции // (в печати, 2026).

7. Емельянов В.С., Кердан А.В., Мурзина Е.В., Шелохов И.А. Свидетельство о регистрации программы ЭВМ № 2023663275 Model 5. 2023.

8. Шарлов М.В. Интегрированная система обработки и инверсии данных нестационарных электромагнитных зондирований // Приборы и системы разведочной геофизики. 2017. Т. 60. № 2. С. 24–35.

9. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain // Psychological Review. 1958. Vol. 65, No. 6. P. 386–408.

10. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // Nature. 1986. Vol. 323. P. 533–536.

11. Vapnik V. The Nature of Statistical Learning Theory. – Springer, 1995.

12. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. – 2nd ed. – Springer, 2009.

13. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press, 2016.

14. Вахромеев А.Г. Закономерности формирования и локализации место-

рождений промышленных рассолов в карбонатных каверново-трещинных резервуарах кембрия юга Сибирской платформы (по данным глубокого бурения, испытания скважин и полевой геофизики): Монография. Иркутск: Изд-во ИрННТУ, 2015. 248 с.

15. Ильин А.И., Вахромеев А.Г., Сверкунов С.А., Поспеев А.В., Горлов И.В. Пути прогноза горно-геологических условий бурения на Ковыктинском газоконденсатном месторождении // Известия Сибирского отделения секции наук о Земле Российской академии естественных наук. Геология, поиски и разведка рудных месторождений. 2016. № 4 (57)