

ГЕОФИЗИЧЕСКИЕ МЕТОДЫ ПОИСКОВ И РАЗВЕДКИ  
GEOPHYSICAL METHODS OF PROSPECTING AND EXPLORATION

УДК 004.94:534.2

**АНАЛИЗ ПОДХОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ИНТЕРПРЕТАЦИИ  
АКУСТИЧЕСКИХ ПОЛЕЙ, ПОЛУЧЕННЫХ МОДЕЛИРОВАНИЕМ ДАННЫХ  
СКВАЖИННОЙ ШУМОМЕТРИИ**

*Н.В. МУТОВКИН*

*Московский физико-технический институт  
9, Институтский пер., г. Долгопрудный 141701, Россия  
e-mail: mutovkin@phystech.edu*

Оценка фазового состава флюида в скважине на основе анализа частот радиальных резонансных мод, возбуждаемых акустическим шумом в зоне притока, является перспективным методом интерпретации результатов пассивной шумометрии. Машинное обучение позволяет учитывать многие факторы, влияющие на спектр измеряемого сигнала, выделяя из них именно те, которые связаны с изменением фазового состава. Для построения наилучшей модели в работе рассмотрены такие подходы машинного обучения, как линейная регрессия с различными вариантами регуляризации, байесовская регрессия, нейронная сеть, методы опорных векторов, решающего дерева, случайного леса и градиентного бустинга. Наборы данных для обучения и тестирования алгоритма получены на основе рассчитанных по двумерной математической модели сценариев с различными значениями параметров пласта и соотношения объемных долей флюидов, заполняющих скважину. Проверено влияние на точность оценки фазового состава различных факторов, в числе которых наличие корпуса акустического прибора, посторонний шум в сигнале и формы спектра сигнала. Показано, что при отсутствии искажений данных можно построить модели, обеспечивающие абсолютную ошибку в оценке фазового состава порядка 1% после зоны притока флюида и порядка 5% в зоне до притока.

**К л ю ч е в ы е с л о в а:** акустический шум; интерпретация; машинное обучение; линейная регрессия; метод опорных векторов; случайный лес; градиентный бустинг; нейронная сеть.

<https://doi.org/10.32454/0016-7762-2019-6-73-79>

**ANALYSIS OF MACHINE LEARNING APPROACHES FOR THE INTERPRETATION  
OF ACOUSTIC FIELDS OBTAINED BY WELL NOISE DATA MODELLING**

*NIKITA V. MUTOVKIN*

*<sup>1</sup>Moscow Institute of Physics and Technology  
9, Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russia  
e-mail: mutovkin@phystech.edu*

Assessing the phase composition of the fluid in a well based analysis of the frequencies of the radial resonance modes excited by acoustic noise in the inflow zone is a promising method for interpreting the results of passive noise metering. Machine learning makes it possible to take into account many factors affecting the spectrum of the measured signal, extracting from them exactly those factors associated with a change in phase composition. In order to build the best model, machine learning approaches such as linear regression with different variants of regularisation, Bayesian regression, neural net, methods of supporting vectors, decision tree, random forest and gradient boosting are considered. Data sets for training and testing the algorithm were obtained on the basis of scenarios calculated using a two-dimensional mathematical model with the different values of the bed parameters and ratio of volume fractions of the well filling fluids. The effect on the assessment accuracy of the phase composition of various factors, including the presence of acoustic device housing, the foreign noise in the signal and the shape of the signal spectrum, was checked. It is shown that in the absence of data distortion, it is possible to build models that provide an absolute error in the assessment of the phase composition about 1% after the zone of fluid inflow and about 5% in the zone before the inflow.

**Keywords:** acoustic noise; interpretation; machine learning; linear regression; reference vector method; random forest; gradient boosting; neural net.



Измерение и последующая интерпретация акустического шума в скважине играет важную роль при мониторинге добычи нефти и газа [1]. Одним из источников сигнала является шум, генерируемый при фильтрации флюида в пористой среде. Природа возникновения такого шума обсуждалась во многих работах, как теоретических, так и экспериментальных [2–6]. На основе анализа характеристик таких шумов возможно оценить границы потока в пласте. Прямое моделирование [7–8] акустического поля, порожденного шумом такого типа, показало, что в спектре регистрируемого в скважине сигнала наблюдаются резонансные моды, зависящие от таких факторов как параметры пластов, геометрия скважины, геометрия акустического прибора и свойства флюида, заполняющего скважину.

В [9] предложен подход интерпретации акустического поля для решения задачи оценки фазового состава флюида в скважине. В качестве примера приводится ситуация с прорывом потока воды из пласта в нефтедобывающую скважину, в результате чего в стволе меняется состав двухфазного потока (нефть + вода). Соответственно, необходимо

оценить соотношения долей воды и нефти в стволе до и после зоны притока на основе записанного прибором в скважине акустического сигнала. При этом предполагается, что доля воды после и в зоне притока одинакова и далее в качестве неизвестной считается доля воды после притока. Подход основан на чувствительности резонансной картины сигналов в скважине к соотношению долей воды и нефти в стволе скважины (рис. 1). Задача оценки соотношения долей нефти и воды рассматривается как задача регрессии и решается с помощью метода гребневой регрессии.

Цель настоящей работы — улучшить точность и надежность предсказания при проведении интерпретации с использованием того же признакового вектора как и в [9], но рассмотрев другие методы машинного обучения, в том числе линейную регрессию с регуляризацией, метод опорных векторов, методы основанные на решающих деревьях и нейронные сети [10–11]. Кроме того, исследуется влияние факторов, затрудняющих интерпретацию акустического поля в скважине, в числе которых учет корпуса акустического прибора, форма спектра источника шума и случайный шум.

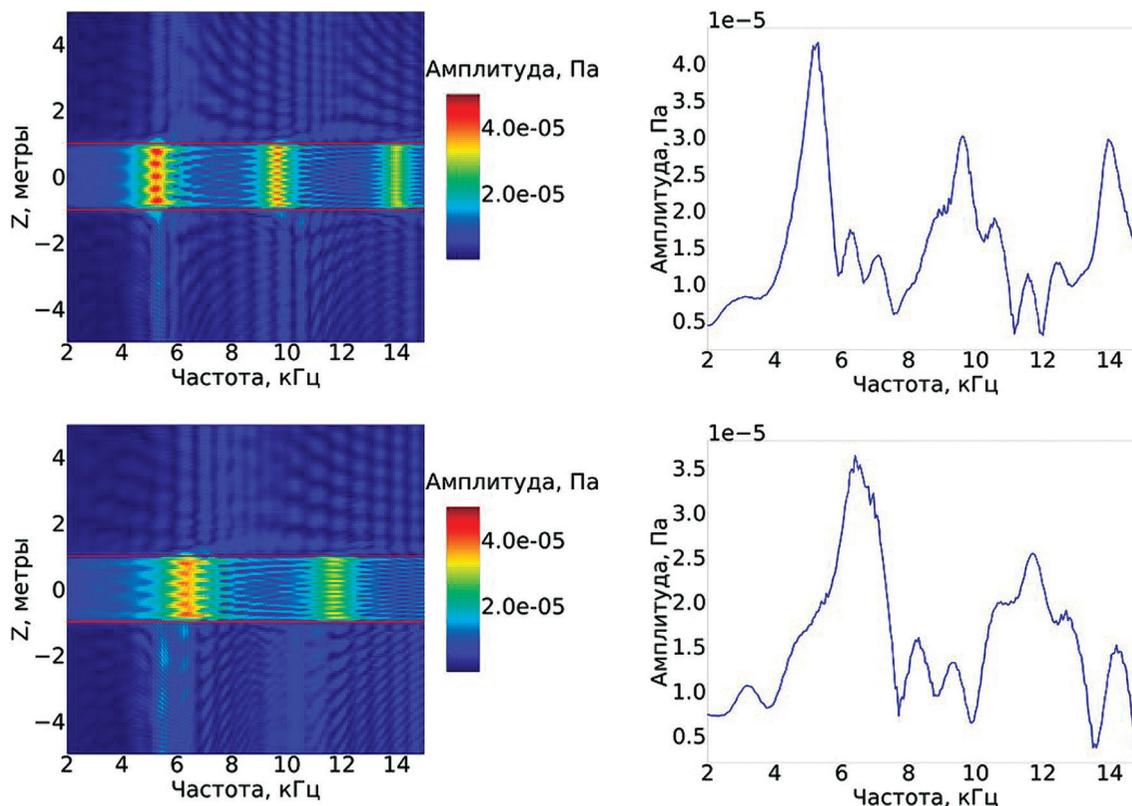


Рис. 1. Рассчитанные спектры акустического поля в скважине при разных соотношениях долей воды и нефти: сверху — 40% воды, снизу — 80% воды, слева — распределения амплитуд спектров сигналов по глубине вдоль ствола скважины, справа — амплитуды спектра сигнала в середине зоны притока. Красные линии соответствуют границам зоны притока.



### Материалы и методы

Алгоритм интерпретации акустического поля в скважине использует методы машинного обучения с учителем. Вначале формируется набор сценариев для модели с различными свойствами пласта и флюида, заполняющего скважину. Для каждого из сформированных сценариев выполняется расчет по двумерной численной модели [12], по результату которого записываются сигналы акустического давления в точках вдоль ствола скважины. На основе их спектров (рис. 1) составляется набор входных данных из признаков векторов для машинного обучения. Целевыми переменными в задаче являются числовые значения долей воды до и после зоны притока.

Конфигурация для поставленной задачи рассматривается в двух вариантах: с учетом и без учета корпуса акустического прибора, считывающего сигнал. Таким образом, мы получаем два однотипных набора данных, что позволит нам оценить влияние прибора на точность интерпретации.

Параметры модели подразделяются на фиксированные и варьируемые. К фиксированным относятся: геометрические размеры, свойства пластов, нефти, воды и материала корпуса акустического прибора. В качестве варьируемых параметров рассматриваются доли воды в скважине до и после зоны притока, ширина зоны притока, меняющаяся в диапазоне от 0,5 м до 3,5 м с шагом 0,5 м и один из параметров породы, для которого было выбрано 3 значения. Для долей воды были взяты значения от 10% до 90% с шагом в 10% и с условием, что после зоны притока доля воды выше, чем до нее.

Для каждого сценария мы формируем признаковый вектор. Его компоненты вычисляются по амплитудам частот в спектре (рис. 1). Опишем процесс этого вычисления. После расчета очередного сценария мы получаем сигналы акустического давления  $p(t)$  в точках сетки  $z_i$  (рис. 2). В реальных скважинах регистрация акустического сигнала выполняется с некоторым шагом по стволу, мы фиксируем этот шаг  $d = 1$  м. Каждому сценарию

мы назначаем три варианта координат точек регистрации, выбирая случайным образом координату первой точки слева от  $z = 0$ . Значения сигнала в точках регистрации вычисляются усреднением по  $N_w$  сеточным узлам  $z_j$ , в которых записаны данные, попавшим в окно шириной  $w = 0,06$  м.

Так как реальные данные шумометрии искажены посторонними сигналами, в каждой точке регистрации к акустическому сигналу добавляется искусственный шум:

$$p_n(z_r, t_j) = p(z_r, t_j) + 0,2C_N \max_{z_r, t_j} p(z_r, t) \xi(t_j).$$

Здесь,  $C_N$  — относительная амплитуда постороннего шума, меняющаяся от 0 до 0,25 (25%);  $\xi(t_j)$  — равномерно распределенная случайная величина на отрезке  $[-1, 1]$ . Максимум, используемый для нормировки, берется по всем точкам регистрации и по всем моментам времени  $t_j$ . Эмпирический коэффициент 0,2 в формуле возник из-за того, что в зоне притока на резонансных частотах уровень сигнала возрастает приблизительно в 5 раз по сравнению с фоновым уровнем сигнала вне притока (рис. 1).

Для вычисления спектров применяется быстрое преобразование Фурье  $FFT$  и обычным усреднением формируются 65 частотных компонент, лежащих в диапазоне от 2 до 15 кГц с шагом 200 Гц.

Чтобы получить окончательный признаковый вектор, точки регистрации группируются по их расположению: до притока, в зоне притока и после притока. В нашей постановке доли воды после притока и в притоке предполагаются одинаковыми, поэтому точки регистрации, лежащие после зоны притока, исключаются. Ширина зоны притока при этом считается известной. Компоненты признакового вектора, соответствующие зоне до притока рассчитываются усреднением по точкам регистрации, лежащим в этой зоне. Для получения компонент признакового вектора в зоне притока выбирается одна точка регистрации с максимальной интенсивностью.

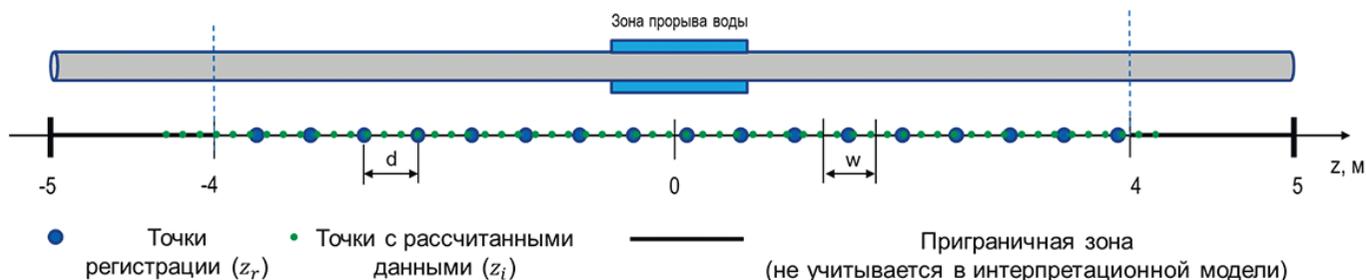


Рис. 2. Расположение точек регистрации



Перебор всех комбинаций варьируемых параметров в заданных наборах значений позволяет сгенерировать 2268 различных сценариев.

Задача оценки доли воды рассматривается как задача регрессии. Для построения наиболее точных моделей были рассмотрены следующие подходы: линейная регрессия (ЛР1), гребневая регрессия (ГР), лассо регрессия (ЛР2), elastic net (EN), байесовская регрессия (БР), решающее дерево (РД), случайный лес (СЛ), градиентный бустинг (ГБ), метод опорных векторов (МОВ), нейронная сеть с одним скрытым слоем (НС) [10–11].

Набор данных разбивается на обучающую и тестовую выборку в соотношении 60% на 40%. На обучающей выборке проводится кросс-валидация. Метрикой качества предсказания является величина RMSE,  $\sqrt{(y_{true} - y_{pred})^2}$ , где  $y_{true}$  и  $y_{pred}$  — истинные и предсказанные доли воды соответственно. Поскольку доля воды уже является нормированной величиной (0% ÷ 100%), используемая величина RMSE служит мерой абсолютной ошибки предсказания. Рассмотренный выше алгоритм реализован на языке программирования Python с использованием библиотеки Scikit-learn [13].

Отметим, что обучение моделей, основанных на линейной регрессии, гребневой регрессии, лассо регрессии, elastic net, решающем дереве или методе байесовской регрессии занимает несколько секунд на обычном ноутбуке. Для методов градиентного бустинга, случайного леса и опорных векторов время обучения модели увеличивается до 3–5 минут. Наибольшее время требуется для обучения нейронной сети — 10 минут (4000 эпох).

### Результаты и обсуждение

Сначала изучим влияние постороннего шума на точность оценки для случая прямоугольного

спектра. На рис. 3 сравниваются ошибки оценок (RMSE) для разных методов машинного обучения при разном уровне постороннего шума.

По приведенным диаграммам можно видеть, что методы, основанные на линейной регрессии, а также методы байесовской регрессии и опорных векторов демонстрируют схожие результаты по точности оценки долей воды до и после зоны притока. Таким образом, использование регуляризаторов (L2, L1 и их смеси) в линейной регрессии слабо влияет на точность предсказания долей воды. Метод с одним решающим деревом показывает чуть более слабые результаты по точности при наличии постороннего шума. В то же время, методы градиентного бустинга и случайного леса, а также нейронная сеть дают существенное улучшение точности, особенно после притока.

Далее, сравним эффективность методов машинного обучения для разных форм спектра сигнала источника без учета постороннего шума. В [7–9] рассмотрен прямоугольный спектр в интервале 2–15 кГц. В этой работе добавлены наборы данных с убывающим степенным спектром и спектром с локальным максимумом. Результаты тестирования на этих наборах данных с различными формами спектра сигнала от источника показали, что для всех типов спектра RMSE примерно одинакова. При этом наилучшие результаты демонстрируют случайный лес, градиентный бустинг и нейронная сеть со значением RMSE в отсутствие постороннего шума от 0,7% до 1,5% после зоны притока и от 4% до 6% — до зоны притока. Таким образом, форма спектра сигнала источника оказывает слабое влияние на ошибку оценки доли воды в скважине.

Проанализируем влияние наличия корпуса акустического прибора на точность оценки фазового состава. Тестирование на наборе данных с учетом

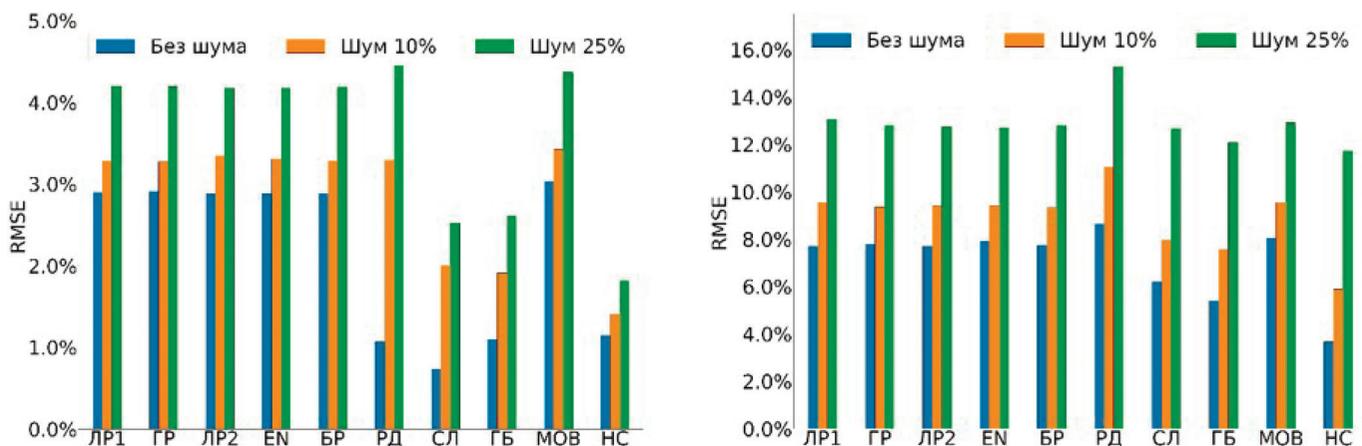


Рис. 3. RMSE для прямоугольного спектра при разном уровне постороннего шума: слева — после зоны притока, справа — до зоны притока. Расшифровка названий методов представлена в разделе «Материалы и методы»

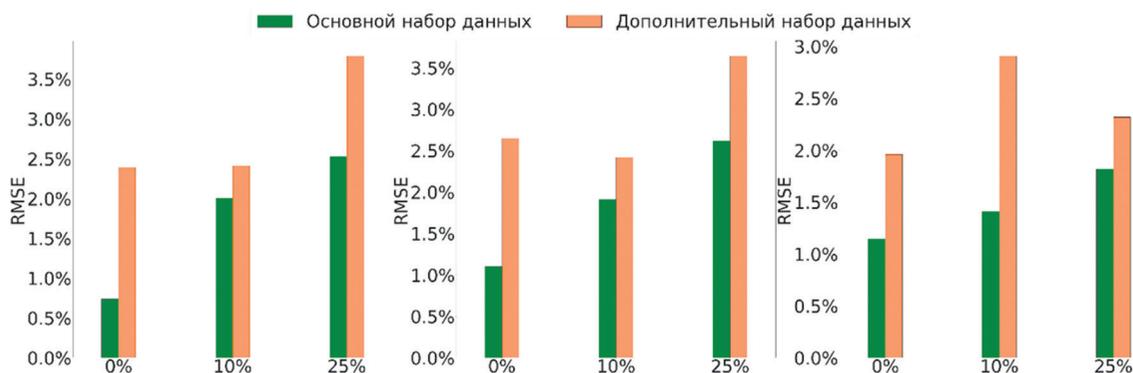


Рис. 4. Сравнение RMSE для основного и дополнительного наборов данных (после зоны притока): слева — случайный лес, посередине — градиентный бустинг, справа — нейронная сеть

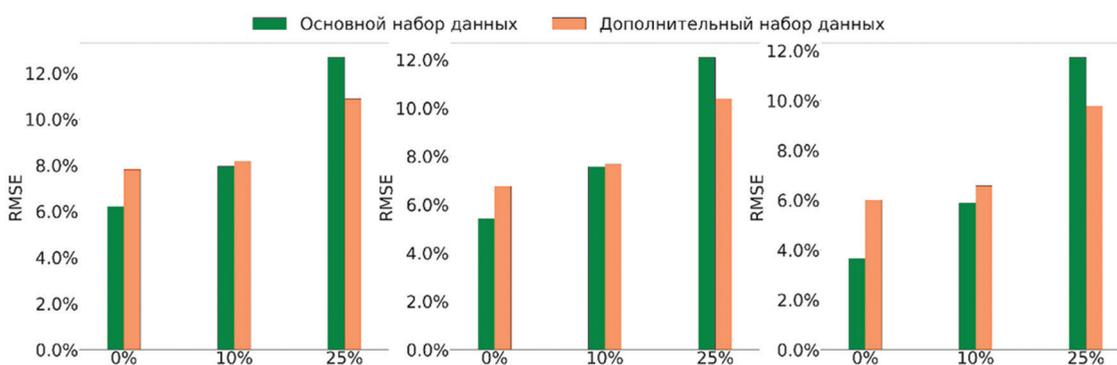


Рис. 5. Сравнение RMSE для основного и дополнительного наборов данных (до зоны притока): слева — случайный лес, посередине — градиентный бустинг, справа — нейронная сеть

прибора в скважине показывает, что в отсутствии постороннего шума RMSE увеличивается приблизительно от 1% до 2% после зоны притока и от 4–6% до 5–7% до зоны притока по сравнению с данными без учета прибора. Добавление постороннего шума увеличивает это отличие до 3 процентных пунктов до зоны притока. Поэтому, можно заключить, что учет прибора приводит к вполне приемлемому ухудшению точности оценки фазового состава.

В заключение изучим насколько падает точность интерпретации для случая, когда тестовая и обучающая выборки формируются по отличающимся наборам варьируемых параметров. Для этого был сгенерирован дополнительный набор данных аналогично ранее приведенной процедуре. В нем доли воды варьируются от 13% до 91% с шагом 13%, а ширина зоны прорыва — от 0,8 до 3,2 м с шагом 0,8 м; для параметра породы взято 2 значения (отличающихся от исходных в обучающей выборке). При переборе всех комбинаций варьируемых параметров в этом наборе данных с учетом того, что каждому сценарию назначается три варианта координат точек регистрации, получается 504 различных сценария. В качестве обучающей выборки для алгоритма машинного обучения берутся

те же 60% от ранее сформированного основного набора данных. Тестовой выборкой служит либо оставшиеся 40% основного набора, либо дополнительный набор данных.

На рис. 4, рис. 5 сравниваются точности оценок долей воды на основном и дополнительном тестовых наборах данных. Из них видно, что после зоны притока RMSE на дополнительном наборе данных увеличивается примерно на 0,5–1,5 процентных пункта по сравнению с основным, например для нейронной сети в отсутствие постороннего шума ошибка увеличивается с 1,1% до 1,9%. Интересно отметить, что до зоны притока наблюдается разное поведение RMSE при разных уровнях постороннего шума: ошибка возрастает или падает для постороннего шума 0% и 25%, соответственно.

Исходя из приведенных результатов, можно сделать вывод, что построенные модели машинного обучения демонстрируют хорошую точность предсказания и на наборе данных, посчитанном с другими значениями физических параметров.

### Заключение

В работе рассмотрены различные модели машинного обучения с целью улучшения точности предложенного в [9] подхода интерпретации аку-



стического шума, вызываемого фильтрацией флюида через пористую среду.

В качестве данных для обучения и тестирования использованы акустические поля, рассчитанные прямым моделированием для большого количества сценариев. При этом наборы данных формировались с учетом или без учета акустического прибора, с различной формой спектра источника, и с добавлением искусственного постороннего шума к сигналу. Для тестирования был также построен набор данных на основе сценариев с отличающимися от обучающей выборки варьируемыми параметрами.

Из десяти моделей машинного обучения отобраны те, которые показали наилучший результат, а именно: градиентный бустинг, случайный лес и нейронная сеть. В отсутствии постороннего шума в сигнале прибора соотношения долей воды и нефти оцениваются с абсолютной ошибкой в доле воды 1% после зоны притока и 4–6% в зоне до притока. Добавление постороннего шума до 25% в интерпретационную модель может увеличить ошибку до 2.5% после зоны притока и до 13% до притока.

Изучение влияния факторов, затрудняющих интерпретацию регистрируемого шума, позволило выявить, что форма спектра сигнала от источника почти не влияет на точность оценки фазового

состава (в работе она предполагается известной и одинаковой для обучающей и тестовой выборки), в то время как наличие корпуса акустического прибора в скважине увеличивает абсолютную величину ошибки предсказания на 0,5–3 процентных пункта в зависимости от уровня постороннего шума, искажающего сигнал.

Построенные модели также показывают хорошие результаты интерпретации на наборе данных, для формирования которого использовались сценарии с другим набором значений долей воды в скважине и другими значениями ширины зоны прорыва воды, по сравнению с обучающими данными.

Полученные результаты свидетельствуют о том, что для предложенного в [9] подхода можно построить достаточно точные предсказательные модели, причем устойчивые к ряду факторов, затрудняющих интерпретацию. Это позволяет рассматривать данный подход как основу для разработки методов оценки параметров фазового состава флюида в скважине на основе полевых данных.

Автор выражает благодарность Д.Н. Михайлову и И.Л. Софронову за плодотворные обсуждения и помощь в ходе написания статьи, а также компании Шлюмберге за возможность опубликовать представленные результаты.

## ЛИТЕРАТУРА

1. Ипатов А.И., Кременецкий М.И. Геофизический и гидродинамический контроль разработки месторождений углеводородов. М.: НИЦ «Регулярная и хаотическая динамика», 2006. 780 с.
2. Николаев С.А., Овчинников М.Н. Генерация звука фильтрационным потоком в пористых средах // Акустический журнал. 1992 Т. 38. № 1. С. 114–118.
3. Красновидов Е.Ю. Создание методики акустико-гидродинамических исследований пористых сред и скважин. Дис. канд. тех. наук. Москва, 2005. 171 с.
4. DiCarlo D.A., Cidoncha J.I.G., Hickey C. Acoustic measurements of pore-scale displacements // Geophysical Research Letters, 2003. Vol. 30. N 17. P. HLS 2-1–2-5. DOI: 10.1029/2003GL017811.
5. Mikhailov D.N., Sergeev S.I. Investigation parameters for sound induced by fluid displacement in rock samples // Water Resources Research, 2019. Vol. 55. N 5. P. 4220–4232. DOI: 10.1029/2018WR024168.
6. Марфин Е.А., Гайфутдинов Р.Р., Метелев И.С. Исследование интенсивности акустической эмиссии при фильтрации газа через пористые среды // Геомодель, Геленджик, 2018. DOI: 10.3997/2214-4609.201802405.
7. Мутовкин Н.В., Михайлов Д.Н., Софронов И.Л. Моделирование акустических полей, генерируемых фильтрационным потоком в околоскважинной зоне // Математическое моделирование, 2019. Т. 31, № 6. С. 95–106. DOI: 10.1134/S0234087919060066.
8. Мутовкин Н.В., Михайлов Д.Н., Софронов И.Л. Анализ результатов моделирования акустических полей, возбуждаемых течением флюида в пласте // Геомодель, Геленджик, 2018. DOI: 10.3997/2214-4609.201802388.
9. Мутовкин Н.В., Михайлов Д.Н., Софронов И.Л. Оценка изменения фазового состава флюида в скважине на основе данных пассивной акустической шумометрии // Российская нефтегазовая техническая конференция SPE, Москва, 2019. ID: SPE-196845-RU. 16 с.
10. Hastie T., Tibshirani R., Friedman J.H. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics, 2009. 745 p.
11. Bishop M. Pattern recognition and machine learning. Springer-Verlag New York, 2006. 738 p.
12. Plyushchenkov B.D., Turchaninov V.I. Acoustic Logging Modeling by Refined Biot's Equations // Int. J. Mod. Phys. C., 2000. Vol. 11. N 2. P. 365–397.
13. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research, 2011. Vol. 12. P. 2825–2830.



REFERENCES

1. Ipatov A.I., Kremenetskiy M.I. *Geofizicheskiy i gidrodinamicheskiy kontrol' razrabotki mestorozhdeniy uglevodorodov*. Moscow, NITs «Regulyarnaya i khaoticheskaya dinamika», 2006, 780 p. (In Russian)
  2. Nikolaev S.A., Ovchinnikov M.N. Generatsiia zvuka filtratsionnym potokom v poristyykh sredakh. *Akusticheskii zhurnal*, 1992, Vol. 38, no 1, pp. 114–118. (In Russian)
  3. Krasnovidov E.Yu. Sozdanie metodiki akustiko-gidrodinamicheskikh issledovaniy poristyykh sred i skvazhin. Diss. kand. tech. nauk. Moscow, 2005, 171 p. (In Russian)
  4. DiCarlo D.A., Cidoncha J.I.G., Hickey C. Acoustic measurements of pore-scale displacements. *Geophysical Research Letters*, 2003, Vol. 30, no 17, P. HLS 2-1–2-5. DOI: 10.1029/2003GL017811.
  5. Mikhailov D.N., Sergeev S.I. Investigation parameters for sound induced by fluid displacement in rock samples. *Water Resources Research*, 2019, Vol. 55, no 5, P. 4220–4232. DOI: 10.1029/2018WR024168.
  6. Marfin E.A., Gaifutdinov R.R., Metelev I.S. Investigation of the Intensity of Acoustic Emission During Gas Filtration Through Porous Media. *Geomodel, Gelendzhik*, 2018, DOI: 10.3997/2214-4609.201802405. (In Russian)
  7. Mutovkin N.V., Mikhailov D.N., Sofronov I.L. Modeling of acoustic fields excited by fluid flow in near-wellbore zone. *Mathematical Models and Computer Simulations*, 2019, Vol. 31, no 6. pp. 95–106. DOI: 10.1134/S0234087919060066.
  8. Mutovkin N.V., Mikhailov D.N., Sofronov I.L. Analysis of Modeling of Acoustic Fields Excited by the Flow Noise in the Formation Inflow Zones. *Geomodel, Gelendzhik*, 2018. DOI: 10.3997/2214-4609.201802388. (In Russian).
  9. Mutovkin N.V., Mikhailov D.N., Sofronov I.L. Estimation of fluid phase composition variation along the wellbore by analyzing passive acoustic logging data. *SPE Russian Petroleum Technology Conference*, Moscow, 2019. ID: SPE-196845-MS. 15 p.
  10. Hastie T., Tibshirani R., Friedman J.H. *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics, 2009, 745 p.
  11. Bishop M. *Pattern recognition and machine learning*. Springer-Verlag New York, 2006, 738 p.
  12. Plyushchenkov B.D. Turchaninov V.I. 2000. Acoustic Logging Modeling by Refined Biot's Equations. *Int. J. Mod. Phys. C.*, 2000, Vol. 11, no 2, pp. 365–397.
  13. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011, Vol. 12, pp. 2825–2830.
-