

ПРИМЕНЕНИЕ МЕТОДА КЛАССИФИКАЦИИ RANDOM FOREST ДЛЯ РЕШЕНИЯ ЗАДАЧИ ФАЦИАЛЬНОГО РАЙОНИРОВАНИЯ ПО ДАННЫМ СЕЙСМОРАЗВЕДКИ

© А.В. Буторин,
Г.В. Мохов, 2021

А.В. Буторин*, Г.В. Мохов

Научно-Технический Центр «Газпром нефти» (ООО «Газпромнефть НТЦ»), РФ, Санкт-Петербург

Электронный адрес: Butorin.av@gazpromneft-ntc.ru

Введение. Фациальное районирование по данным сейсморазведки является актуальной задачей динамического анализа. Современные подходы к решению этой задачи довольно многочисленны и используют различные алгоритмы. Наиболее распространенной методикой является кластеризация по форме отражения. Данная методика относится к классу задач обучения «без учителя», то есть кластеризация производится без учета априорной информации, выделение сейсмofаций основано на внутренней структуре данных, и ключевой особенностью являются различия волнового пакета внутри целевого интервала. Такой подход требует в дальнейшем увязки результатов кластеризации и геологической информации, что является его недостатком. Другим направлением решения данной задачи является применение алгоритмов обучения «с учителем». К данной категории относятся различные методы классификации, которые относятся к категории машинного обучения. Отличием данного направления от традиционных подходов сейсмofациального анализа является учет геологической информации на стадии вычислений.

Цель. В данной работе рассматриваются результаты исследования, выполненного с целью изучения фациального строения отложений тюменской свиты на группе месторождений в Ханты-Мансийском автономном округе. Тюменская свита характеризуется преобладанием русловых фаций, связанных с развитием сложных речных систем, которые отчетливо проявляются в динамических характеристиках волнового поля. Усложняющим фактором при изучении данных отложений выступает достаточно низкая изученность скважинными данными, что затрудняет геологическую интерпретацию получаемых результатов.

Материалы и методы. Для решения поставленной задачи авторами использован метод классификации Random Forest. Применение метода рассмотрено в рамках кластера, состоящего из трех сейсмических съемок, полученных в разное время. Для обучения использована экспертная разметка по площади на основании изучения распределения амплитуд вдоль отражающего горизонта.

Результаты. В результате выполненного исследования на значительной площади получена вероятностная оценка распространения русловых фаций, с которыми связана перспективность данного типа отложений на изучаемой территории. Таким образом, авторами была разработана методика, позволяющая получить оценку вероятности наличия определенной фации с использованием данных сейсморазведки.

Заключение. Выполненное исследование показывает возможность использования метода классификации Random Forest для решения задачи фациального районирования.

Ключевые слова: сейсморазведка, динамическая интерпретация, классификация

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов

Для цитирования: Буторин А.В., Мохов Г.В. Применение метода классификации Random Forest для решения задачи фациального районирования по данным сейсморазведки. ПРОНЕФТЬ. Профессионально о нефти. 2021;6(3):23–29. <https://doi.org/10.51890/2587-7399-2021-6-3-23-29>

Статья поступила в редакцию 17.05.2021

Принята к публикации 04.06.2021

Опубликована 29.09.2021

APPLICATION OF THE RANDOM FOREST CLASSIFICATION METHOD FOR FACIES ZONING
ACCORDING TO SEISMIC DATA

Aleksander V. Butorin*, Grigory V. Mokhov

Gazpromneft STC LLC, RF, Saint Petersburg

E-mail: Butorin.av@gazpromneft-ntc.ru

Background. Facial zoning based on seismic data is an important task of dynamic analysis. There are numerous approaches of solving this problem using various algorithms. The most common method is clustering by reflection shape. This approach belongs to unsupervised learning algorithms, due to the mapping of seismic facies is based on the internal data structure and the key feature is the change in the wave packet within the target interval. The disadvantage of this method is requirement of further tying clustering results and geological information.

Another way of directed solution of this problem is the use of supervised learning algorithms. This category includes various classification methods that relate to the category of machine learning. In comparison to traditional approaches of seismic facial analysis, this method accounts geological information at the computation stage.

Aim. This paper considers the results of a research carried out with the study of the facies structure of the Tyumen formation at a group of fields in the Khanty-Mansiysk Autonomous Region. The Tyumen formation is characterized by the predominance of channel facies associated with the development of complex river systems, which are clearly observed in the dynamic characteristics of the wave field. A complicating factor in the study of these deposits is the rather low coverage of well data, which makes difficult the geological interpretation of the results obtained.

Materials and methods. The authors used the Random Forest classification method to deal with the assigned task. The application of the method is considered on the cluster consisting of three seismic surveys obtained at different times. For training, expert marking by area was used based on the distribution of amplitudes along the reflecting horizon.

Results. As a result of the research, a probabilistic assessment of the distribution of channel facies was obtained, that is related to the perspective of this type of deposits in the study area. Thus, the authors have developed a methodology that gives an opportunity to obtain an estimate of the probability of the presence of a certain facies using seismic data.

Conclusions. The performed study shows the possibility of using the Random Forest classification method to solve the problem of facial zoning.

Keywords: seismic, dynamic analysis, classification

Conflict of interest: the authors declare no conflict of interest.

For citation: Butorin A.V., Mokhov G.V. Application of the Random Forest classification method for facies zoning according to seismic data. PRONEFT. Professionally about oil. 2021;6(3):23–29. <https://doi.org/10.51890/2587-7399-2021-6-3-23-29>

Manuscript received 17.05.2021

Accepted 04.06.2021

Published 29.09.2021

ВВЕДЕНИЕ

Одной из задач качественной интерпретации данных сейсморазведки является фациальное районирование изучаемой территории. Существует множество подходов к решению данной задачи. Распространенным инструментом является кластеризация по форме отражения с использованием нейронных сетей либо метода K-средних. В этом случае на выходе алгоритма оценивается карта кластеров, в которые объединены схожие по форме отражения трассы. В дальнейшем требуется геологическая интерпретация полученного результата — отождествление различных форм отражения с фациальными обстановками, определенными в скважинах. Данный подход относится к задаче обучения «без учителя», то есть кластеризация производится без учета априорной информации, что является негативной стороной данного подхода. Другим инструментом является оконтуривание объектов или обстановок в полях динамических атрибутов (амплитудных или частотных). Негативным моментом данного подхода является значительная зависимость от интерпретатора.

В рамках данного исследования рассматривается подход к фациальному анализу, основанный на применении алгоритмов обучения «с учителем», который сочетает в себе обе упомянутых выше технологии. Подобные алгоритмы относятся к методам классификации и требуют на входе априорную ин-

формацию в виде выделенных фациальных обстановок в некоторой области изучаемой территории.

Объектом в данном исследовании выступают отложения тюменской свиты группы месторождений, расположенных на территории Ханты-Мансийского автономного округа. Характерной особенностью данных отложений является наличие обстановок, связанных с развитием русловых фаций, которые определяют распространение фильтрационно-емкостных свойств разреза. Таким образом, была выдвинута гипотеза, что картирование русловых систем тюменской свиты может быть выполнено с использованием алгоритмов обучения «с учителем».

Для проверки гипотезы использованы данные сейсморазведки МОВ ОГТ 3D по трем площадям, расположенным по соседству. Данные сейсморазведки были получены в разное время, обрабатывались и интерпретировались раздельно друг от друга. Данное обстоятельство позволяет оценить устойчивость алгоритма к различиям во входных данных.

На **рис. 1** показана карта среднеквадратичных амплитуд в окне 20 мс от отражающего горизонта Ю3. Распределение амплитуд позволяет установить наличие русловых тел, проявляющихся в пониженных значениях амплитуд. Данная карта использована для получения обучающего набора — по участкам пониженных значений амплитуд в центральной части площади куба 1 закартировано развитие системы русел. Обучающий участок

представляет собой индексную карту [1, 2], где 1 (серый цвет) — отсутствие русла, а 2 (зеленый цвет) — наличие русла.

МЕТОДИКА

В настоящее время алгоритмы машинного обучения эффективно используются для решения различных задач интерпретации данных сейсморазведки. В данном исследовании рассматривается задача классификации волнового поля по имеющейся обучающей выборке. Одним из распространенных методов решения подобных задач является алгоритм Random Forest, который был предложен в 2001 году [1] и заключается в использовании ансамбля решающих деревьев.

В качестве исходных данных рассматриваются пары объектов (x, y) , где x — вектор признаков (значения амплитуд), а y — метка класса (значение с обучающей карты). Основным элементом Random Forest является дерево решений. При построении дерева решений на каждом шаге выбирается признак из вектора x и значение порога, который делит входную выборку на две части — данный процесс является ветвлением дерева решений. Разделение может происходить до момента, когда в каждой полученной подвыборке остаются только элементы, относящиеся к одному классу, что будет давать ответ дерева решений. Подобный алгоритм однозначно классифицирует входную выборку данных, но при этом является «переобученным», то есть будет давать ошибки на выборке, не входящей в обучающий массив. Ключевая идея метода Random Forest заключается в использовании ансамбля (множества) деревьев решения, при этом в рамках алгоритма происходит осреднение ответов, полученных каждым отдельным деревом. Подобный подход позволяет получить эффективный инструмент, более устойчивый к переобучению.

В исследовании использована реализация метода Random Forest на языке Python, представленная в библиотеке scikit learn (<https://scikit-learn.org>). В рамках алгоритма выполнялось тестирование различного количества деревьев решения в ансамбле, а также ограничение на их глубину (количество ветвлений).

РЕЗУЛЬТАТЫ

Для тестирования гипотезы о применимости метода Random Forest для классификации русловых тел по волновому полю использо-

вались имеющиеся данные сейсморазведки по трем площадям, суммарная площадь кубов составила 3350 кв. км. Для обучения классификатора использовался участок площадью 75 кв. км, в рамках которого по распределению амплитуд были о контурены наблюдающиеся русловые тела (рис. 1). Таким образом, площадь обучения составила около 2% от общей изучаемой площади.

В ПРЕДЛАГАЕМОМ АЛГОРИТМЕ МАШИННОГО ОБУЧЕНИЯ «С УЧИТЕЛЕМ» ИСПОЛЬЗОВАЛАСЬ ВЫБОРКА ДАННЫХ НА ПЛОЩАДИ МЕНЕЕ 2% ОТ ОБЩЕЙ ТЕРРИТОРИИ, В РЕЗУЛЬТАТЕ ПОЛУЧЕНА КАРТА ВЕРОЯТНОСТИ ФАЦИЙ ПЛОЩАДЬЮ БОЛЕЕ 3000 КВ. КМ.

Индексная карта, использованная для обучения, представляет из себя массив из значений 1 и 2, где 1 — зона отсутствия русел, 2 — зона русел. Вектор признаков, по которым выполнялась классификация, представляет собой значения амплитуд для каждой трассы, в заданном по времени окне. В ходе тестирования определено оптимальное временное окно, составляющее 20 мс относительно целевого отражающего горизонта, данное значение сопоставимо с доминантным периодом волнового поля и позволя-

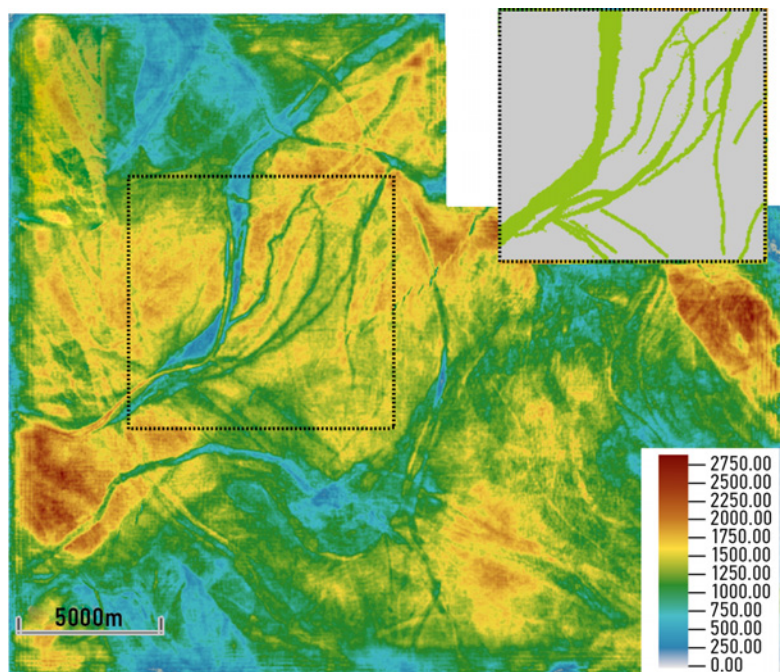


Рис. 1. Карта среднеквадратичных амплитуд вдоль ОГ Ю3 по кубу 1. Справа сверху показан проинтерпретированный участок, использованный для обучения. Составлено авторами

Fig. 1. A map of the root-mean-square amplitudes along the U3 horizon on the cube 1. The interpreted section used for training is shown at the top right. Made by the authors

Таблица 1. Оценка качества классификации
Table 1. Estimation of classification quality

	Precision, %	Recall, %	F1-score. y.e.
«Не русло»	0,95	0,91	0,93
«Русло»	0,66	0,80	0,72

ет изучить изменение формы отражения по площади.

Таким образом, была сформирована выборка для обучения алгоритма Random Forest, представляющая из себя набор амплитуд в окне 20 мс (10 карт с шагом 2 мс) для каждой трассы и соответствующее значение класса «русло» (2), «не русло» (1). Сформированный массив подавался на вход алгоритма Random Forest, при этом вся выборка случайным образом разделялась на две равные части — одна использовалась для обучения классификатора, вторая использовалась для оценки качества и не учитывалась при обучении. В рамках алгоритма использовался ансамбль из 15 решающих деревьев, ограниченных по глубине ветвления 20 узлами.

В ходе исследования тестировались различные подходы к формированию обучающей выборки и настройкам алгоритма. Наилучший результат был получен при использовании окна снятия амплитуд 20 мс вниз от имеющегося отражающего горизонта. Полученный классификатор на отложенной выборке характеризуется точностью (precision) на уровне 66 % и чувствительностью (recall) 80 % для фации «русло» (табл. 1).

В дальнейшем классификатор использовался для всех трасс куба 1 с целью получения карты вероятности фации «русло». Как видно из рис. 2, применение классификатора позволило детально закартировать систему русел в центральной части площади, в том числе в областях, которые не входили в размеченный участок для обучения.

Наличие карты вероятности позволяет оценить различия в форме трассы при условии различной вероятности наличия фации «русло». Для этого выполнено осреднение трасс, характеризующихся вероятностью фации русло в интервалах [0,7; 1], [0,45; 0,55] и [0; 0,1]. Как видно из рис. 3, формы трасс для различных диапазонов вероятности отличаются друг от друга достаточно слабо, при этом чем выше вероятность, тем ниже уровень амплитуд.

Используя обученный классификатор были проанализированы два других сейсмических куба, расположенных на изучаемой территории. При этом необходимо отметить, что эти кубы не учитывались при обучении классификатора, что позволяет протестировать устойчивость алгоритма. В результате классификации были получены карты вероятности фации «русло» на оставшихся частях изучаемой площади. В целом полученный

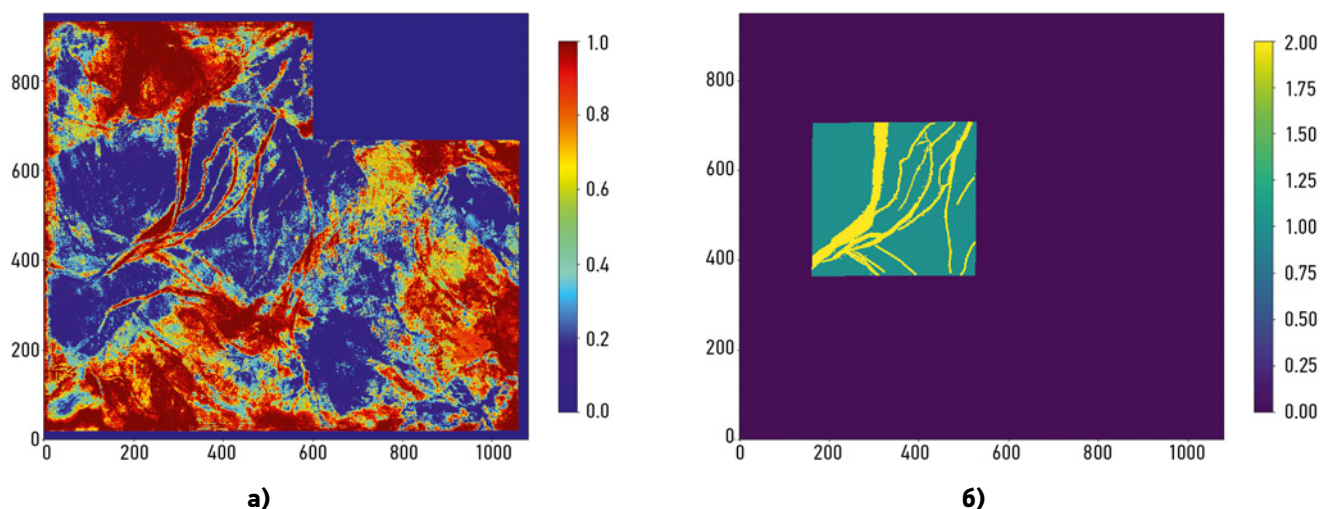


Рис. 2. Результат применения классификатора к сейсмическим данным: а) карта вероятности фации «русло»; б) размеченный участок, использованный для обучения классификатора (справа). Составлено авторами
Fig. 2. The result of applying the classifier to seismic data: а) the probability map of the "channel" facies; б) the marked part used for training (on the right). Made by the authors

результат позволяет получить корректные результаты с позиции геологической интерпретации — зоны высокой вероятности фации «русло» выстраиваются в линейные аномалии, отражая характер развития русловой

системы в период формирования отложений тюменской свиты (рис. 4). В дальнейшем все полученные результаты объединены в единую карту вероятности фации «русло» для изучаемой территории

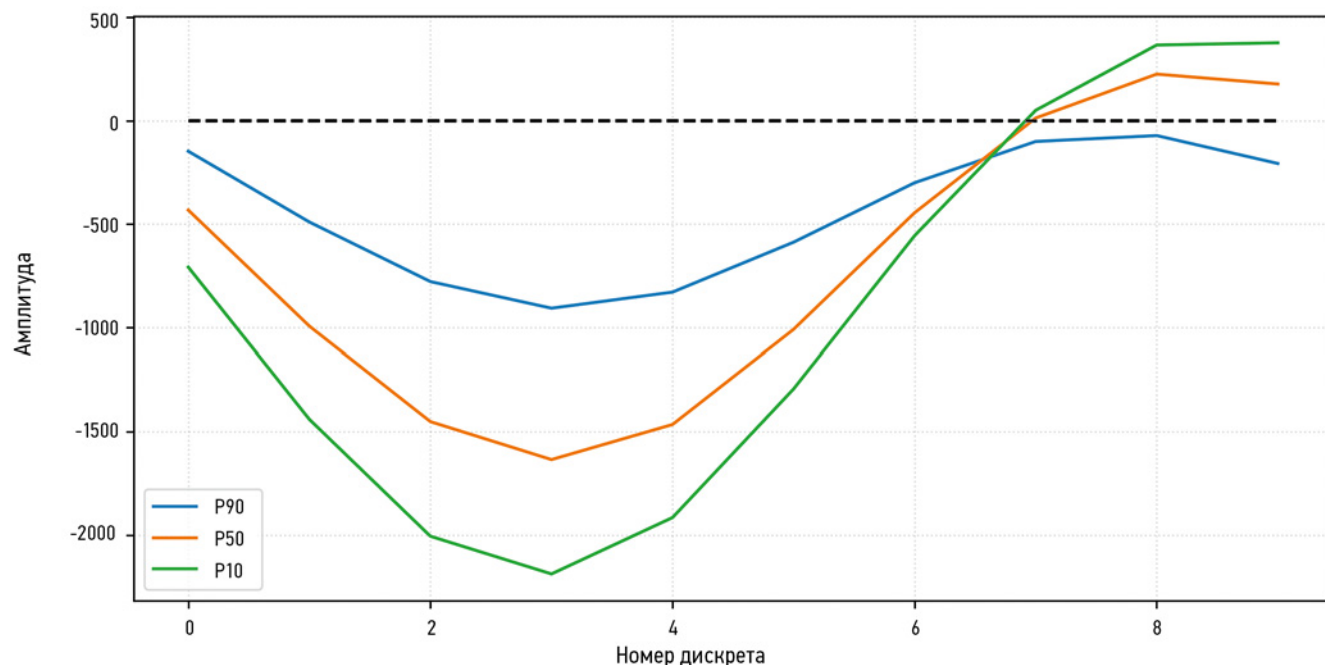


Рис. 3. Характерные формы сейсмических трасс для различных диапазонов вероятности наличия фации «русло»: синий цвет — вероятность выше 0,9, оранжевый — вероятность от 0,45 до 0,55, зеленый — вероятность ниже 0,1. Составлено авторами
Fig. 3. Characteristic shapes of seismic traces for different ranges of probability of the presence of the “channel” facies: blue — probability higher than 0.9, orange-probability from 0.45 to 0.55, green-probability lower than 0.1. Made by the authors

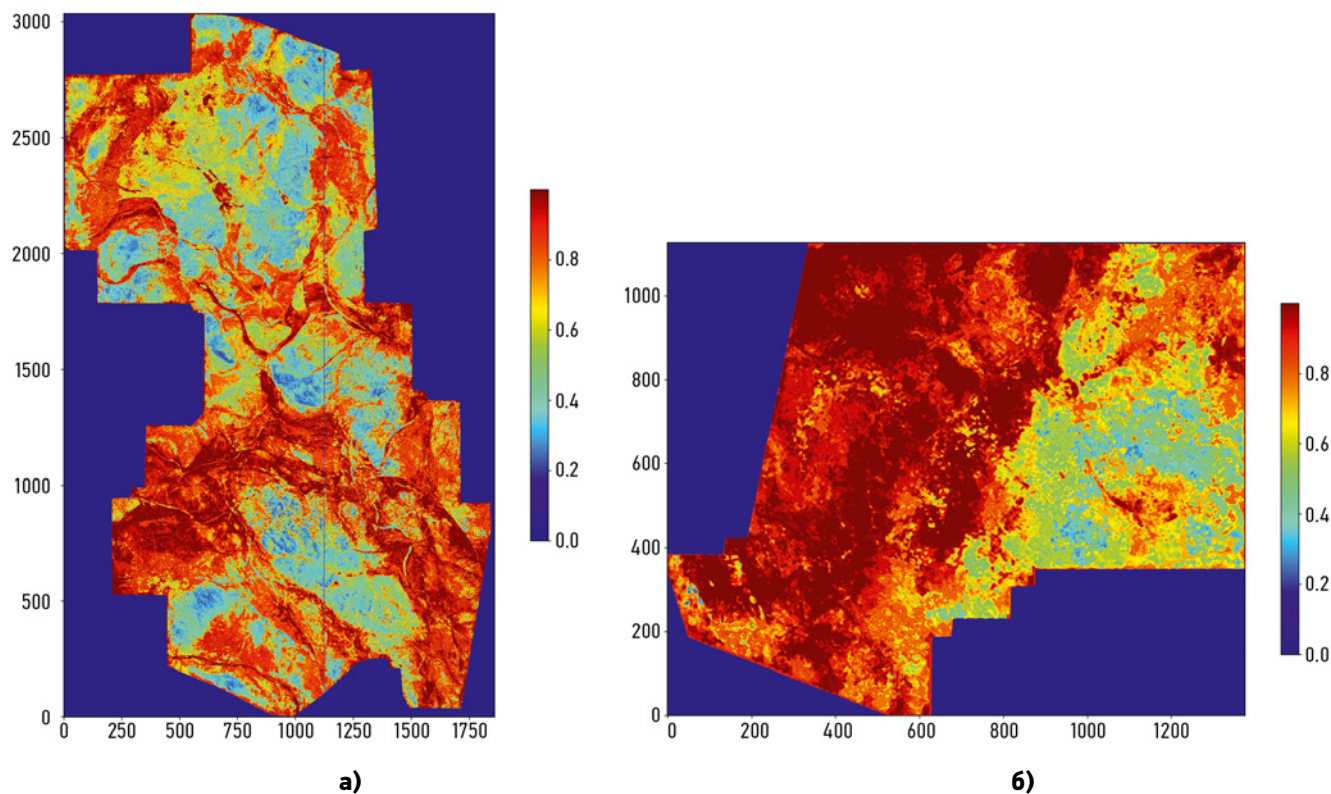


Рис. 4. Карты вероятности фации «русло»: а) для куба 2; б) для куба 3. Составлено авторами
Fig. 4. Probability maps of the “channel” facies: а) for cube 2; б) for cube 3. Made by the authors

(рис. 5). Необходимо отметить высокую степень согласованности значений по границам объединения, все наблюдаемые аномалии вероятности логично переходят из одного куба в другой. Таким образом, применение классификатора позволило получить оценку вероятности развития русловых систем в отложениях тюменской свиты. При этом для обучения алгоритма использовалось около 2 % всей площади и данные только из одного куба.

Анализ полученной карты позволяет сделать предположение о развитии на большей части территории притоков основного русла, проходящего в субмеридиональном направлении на площади куба 3 (восточная часть). Информация о фациальном строении тюменских отложений позволяет в дальнейшем оценивать перспективность различных участков площади. Так, прошлые исследования [2] показывают взаимосвязь наличия рус-

ловой системы и фильтрационно-емкостных характеристик разреза. Таким образом, вблизи русловой системы вероятность наличия коллектора выше, чем на удалении. Данный критерий может быть использован для ранжирования территории по перспективности при дальнейшем планировании геолого-разведочных мероприятий.

ЗАКЛЮЧЕНИЕ

В результате проведенного исследования подтверждена гипотеза о возможном использовании методов классификации («обучение с учителем») для картирования фациальных обстановок тюменской свиты. Получена карта вероятности фаций, связанных с развитием русловых систем на территории более 3000 кв. км. Чувствительность классификатора составила 80 %,

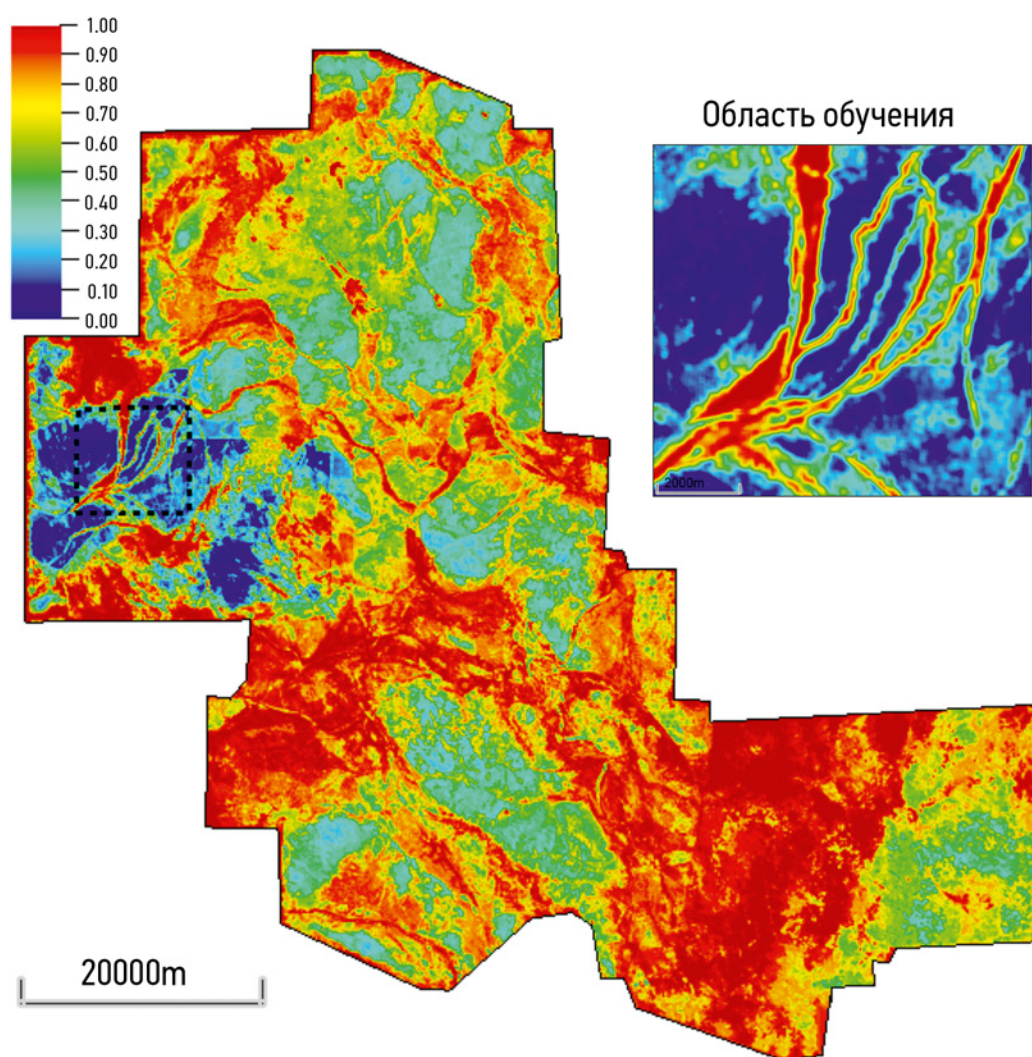


Рис. 5. Итоговая карта вероятности фации «русло». Справа сверху показан участок, использованный для обучения классификатора. Составлено авторами

Fig. 5. The final probability map of the "channel" facies. The section used for training is shown at the top right. Made by the authors

что позволяет говорить о высокой достоверности полученного результата. При этом для обучения использовалась незначительная выборка размеченных данных площадью менее 2% от общей изучаемой территории. По результатам классификации проанализировано влияние отложений типа «русло» на форму отражения в целевом интервале. Наличие отложений русловой системы при-

водит к снижению уровня амплитуд. Однако отличия в форме отражения являются незначительными, что создает неопределенности в работе классификатора. Результаты исследования позволяют выполнить районирование по фациальному признаку, что, в свою очередь, дает информацию о перспективности отложений тюменской свиты на изучаемой территории.

Список литературы

1. Breiman L. Random forests // *Machine Learning*. 2001, vol. 45, no. 1, pp. 5–32.
2. Буторин А.В., Зиннурова Р.Р., Митяев М.Ю., Онегов А.В., Шарифуллин И.Ф., Виноходов М.А. Оценка потенциала тюменской свиты в пределах Ноябрьского региона Западной Сибири // *Нефтяное хозяйство*. — 2015. — № 12. — С. 41–43.

References

1. Breiman L. Random forests. *Machine Learning*. 2001, vol. 45, no. 1, pp. 5–32.
2. Butorin A.V., Zinnurova R.R., Mityaev M.U., Onegov A.V., Sharifullin I.F., Vinohodov M.A. Estimating the potential of the Tumen formation in the Noyabrsk region of Western Siberia. *Neftyanoe khozyaistvo*, 2015, no. 12, pp. 41–43. (In Russ.)

ВКЛАД АВТОРОВ / AUTHOR CONTRIBUTIONS

А.В. Буторин — разработал концепцию статьи, подготовил текст статьи, окончательно утвердил публикуемую версию статьи и согласен принять на себя ответственность за все аспекты работы.

Г.В. Мохов — участвовал в подготовке текста статьи, осуществлял геологическую интерпретацию получаемых данных.

Aleksander V. Butorin — developed the concept of the article, prepared the text of the article, finally approved the published version of the article and agrees to take responsibility for all aspects of the work.

Grigory V. Mokhov — participated in the preparation of the text of the article, carried out the geological interpretation of the data obtained.

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Александр Васильевич Буторин* — кандидат геолого-минералогических наук, руководитель по развитию дисциплины «Сейсморазведка», ООО «Газпромнефть НТЦ» 190000, Санкт-Петербург, Набережная реки Мойки, д. 75–79.
e-mail: butorin.av@gazpromneft-ntc.ru
SPIN-код: 8474-6120
Author ID: 877389
Web of Science Researcher ID: B-7405-2019
ORCID: <https://orcid.org/0000-0002-6074-1439>
Scopus Author ID: 56370048400

Григорий Владимирович Мохов — руководитель направления, ООО «Газпромнефть НТЦ» 190000, Санкт-Петербург, Набережная реки Мойки, д. 75–79.
e-mail: mokhov.gv@gazpromneft-ntc.ru

Aleksander V. Butorin* — Cand. of Sci (Geophysics), Head of seismic discipline, JST "Gazpromneft Science & Technology Center" 75–79 liter D Moika River emb., 190000, St. Petersburg, Russia.
e-mail: butorin.av@gazpromneft-ntc.ru
SPIN-code: 8474-6120
Author ID: 877389
Web of Science Researcher ID: B-7405-2019
ORCID: <https://orcid.org/0000-0002-6074-1439>
Scopus Author ID: 56370048400

Grigory V. Mokhov — Project manager, JST "Gazpromneft Science & Technology Center" 75–79 liter D Moika River emb., 190000, St. Petersburg, Russia
e-mail: mokhov.gv@gazpromneft-ntc.ru

* Автор, ответственный за переписку / Corresponding author