

УДК 622.276:550.832.92.05

Статья / Article

© ПНИПУ / PNRPU, 2019

ИСПОЛЬЗОВАНИЕ ДИСКРИМИНАНТНОГО АНАЛИЗА ДЛЯ ИНТЕРПРЕТАЦИИ ДАННЫХ ГАЗОВОГО КАРОТАЖА НА ПРИМЕРЕ ПАВЛОВСКОГО НЕФТЯНОГО МЕСТОРОЖДЕНИЯ

М.Н. Расторгуев

Пермский инженерно-технический центр «Геофизика» (614015, Россия, г. Пермь, ул. Петропавловская, 16а)

USING DISCRIMINANT ANALYSIS FOR THE INTERPRETATION OF GAS LOGGING DATA ON THE EXAMPLE OF THE PAVLOV OIL FIELD

Mikhail N. Rastorguev

ПИТС «Geofizika» (16a, Petropavlovskaya st., Perm, 614015, Russian Federation)

Получена / Received: 12.10.2018. Принята / Accepted: 14.01.2019. Опубликовано / Published: 29.03.2019

Ключевые слова:

месторождение, нефтенасыщенность, коллектор, газовый каротаж, геолого-технологические исследования, геофизические исследования скважин (ГИС), хроматография, люминесцентно-битуминологический анализ, гамма-каротаж, нейтрон-нейтронный каротаж, углеводородные газы, буровой шлам, буровой раствор, линейный дискриминантный анализ, математическая статистика, вероятность.

Key words:

field, oil saturation, reservoir, gas logging, geological and technological research, well logging, chromatography, luminescent-bitumen analysis, gamma-ray logging, neutron-neutron logging, hydrocarbon gases, drilling sludge, drilling mud, linear discriminant analysis, mathematical statistics, probability.

Статья посвящена разработке новой методики интерпретации газового каротажа по материалам пробуренных скважин Павловского нефтяного месторождения в нижне-средневизейском терригенном нефтегазоносном комплексе в интервале тульского терригенного горизонта C_{11l} с применением элементов математической статистики – пошагового дискриминантного анализа, для проведения которого в качестве переменных были использованы данные газового хроматографического анализа – суммарное газосодержание нефтенасыщенных углеводородов в буровом растворе и покомпонентный состав газовой смеси, а также данные люминесцентно-битуминологического анализа бурового шлама. Кроме того, были использованы данные геофизического исследования скважин (ГИС) – гамма-каротажа, нейтрон-нейтронного каротажа по тепловым нейтронам и разница между диаметром долота и диаметром скважины. По результатам анализа была рассчитана вероятность отнесения наблюдений к нефтенасыщенным песчаникам, которая учитывает показатели газового каротажа и ГИС. Также при разработке методики, кроме нефтенасыщенных песчаников, исследовались породы, попадающиеся в данном нефтегазоносном комплексе: аргиллиты, алевролиты, глинистые песчанники и нефтенасыщенные песчанники. Для визуализации результатов была построена геолого-геофизическая диаграмма, на которой были нанесены: насыщение пород согласно геолого-технологическим исследованиям (ГТИ), насыщение пород согласно ГИС и литологическая колонка по данным ГТИ и ГИС, а также все переменные, которые участвовали в дискриминантном анализе. На основании проведенного анализа были выделены три области значений: 1) область с наблюдениями, относящимися к нефтенасыщенным породам; 2) область с наблюдениями, относящимися к нефтенасыщенным породам и 3) область с наблюдениями неясного насыщения с показателями, занимающими промежуточное положение (переходная зона – возможно, нефтеводонасыщенные или промытые пласты).

The article is devoted to the development of new methodology of gas-logging interpretation based on materials from drilled wells of the Pavlovsky oil field in the Lower-Middle-Visayan terrigenous oil and gas complex in the interval of the Tula terrigenous horizon C_{11l} , with using elements of mathematical statistics - stepwise discriminant analysis, for which the gas chromatographic data (the total gas content of hydrocarbons in the drill mud and component composition of the gas-air mixture) as well as data of luminescent-bitumen analysis of drill cuttings were used as variables. In addition, logging data were used - gamma-ray logging, neutron-neutron logging for thermal neutrons and the difference between the bit diameter and the well diameter. Based on the results of the analysis, the probability of attributing observations to oil-saturated sandstones, which takes into account gas logging and well logging, was calculated. In addition to oil-saturated sandstones, rocks occurring in this oil-gas-bearing complex were studied: mudstones, aleurolites, clayey sandstones and non-oil saturated sandstones. To visualize the results, a geological and geophysical diagram was plotted on which were shown: rock saturation according to well survey, rock saturation according to well logging and lithological column according to well survey and logging, as well as all variables that participated in the discriminant analysis. Based on the discriminatory analysis, three ranges of values were identified: 1) an area with observations related to oil-saturated rocks; 2) an area with observations related to non-oil-saturated rocks and 3) an area with observations of unclear saturation with indicators that are intermediate (the transition zone is possibly oil-saturated or washed reservoirs).

Расторгуев Михаил Николаевич – инженер-интерпретатор группы геолого-технологических исследований (тел.: +007 342 212 14 15, e-mail: mrastorguev@pitc.ru).

Mikhail N. Rastorguev – interpreter engineer of geological and technological research group (tel.: +007 342 212 14 15, e-mail: mrastorguev@pitc.ru).

Введение

В последние годы на месторождениях Пермского края приходится тратить все больше сил по увеличению, а в некоторых случаях и просто по удержанию на текущем уровне, добычи углеводородов. Этой проблемой сейчас озабочены все нефтедобывающие и нефтесервисные компании Пермского края. Роль геолого-технологических исследований (ГТИ) в обеспечении оптимизации процесса строительства скважин для извлечения нефти и газа все больше увеличивается. Расширяется круг задач и требований к качеству проведения исследований, оперативности получения информации, наполняемости комплекса ГТИ.

На данный момент геолого-технологические исследования скважин являются составной частью геолого-геофизических исследований нефтяных и газовых скважин и предназначены для осуществления контроля за состоянием скважины на всех этапах ее строительства и ввода в эксплуатацию с целью изучения геологического разреза, достижения высоких технико-экономических показателей, а также обеспечения выполнения природоохранных требований [1].

ГТИ проводятся непосредственно в процессе бурения скважины, без простоя в работе буровой бригады и бурового оборудования; решают комплекс геологических и технологических задач, направленных на оперативное выделение в разрезе бурящейся скважины перспективных на нефть и газ пластов-коллекторов, изучают их фильтрационно-емкостные свойства и характер насыщения, оптимизацию отбора керны, экспрессное опробование и изучение методами ГИС выделенных объектов, обеспечение безаварийной проводки скважин и оптимизацию режима бурения. Газовый каротаж входит в комплекс ГТИ и составляет его существенную часть [2–3].

Можно разделить все научно-технические разработки, проводимые в сфере геолого-

технологических исследований, на две группы: улучшение характеристик оборудования (в том числе датчиков) и повышение возможностей интерпретации данных ГТИ. Разработка и введение нового оборудования в работу станций ГТИ важны, но требуют больших материальных затрат по сравнению с улучшением методики интерпретации данных, собираемых оборудованием, установленным сейчас на станциях ГТИ. Поэтому предлагается улучшать качество интерпретации как менее затратный процесс. В частности, для повышения геологической эффективности газового каротажа возможно использование методов математической статистики – пошагового дискриминантного анализа, ход и результаты которого описаны в этой статье.

Общая геологическая характеристика Павловского месторождения

Геологический разрез Павловского месторождения изучен по материалам структурных, поисково-разведочных и добывающих скважин на глубину 2243 м и представлен породами от четвертичного до верхнерифейского возраста.

В основу расчленения геологического разреза положена Унифицированная стратиграфическая схема Русской платформы, утвержденная в 1988 году. К сожалению, не во всех пробуренных скважинах в достаточном объеме имеется керновый материал с палеонтологической характеристикой пород, поэтому выделение стратиграфических горизонтов проводилось на основании сопоставления разрезов скважин Павловского месторождения с разрезами других площадей (Красноярско-Куединской, Батырбайской, Таныпской и др.).

В современном региональном тектоническом плане территория Павловского месторождения приурочена к Чернушинской валообразной зоне, осложняющей северный склон Башкирского свода (рис. 1). Ширина вала на юге – 25–30 км, на севере, в районе Таныпского поднятия, – 10–15 км. Вал прослеживается по всем горизонтам верхнего палеозоя.

Павловское месторождение связано с верхнедевонскими рифогенными массивами в пределах бортовой зоны Камско-Кинельской системы впадин, которые образуют поднятия: Березовское, Деткинское, Барановское, Улыкское, Павловское, Южно-Павловское, Григорьевское.

К настоящему времени из семи нефтегазоносных комплексов, выделяемых в

разрезе осадочного чехла Пермского Прикамья на Павловском месторождении, промышленно нефтеносны: верхнедевонско-турнейский карбонатный (пласты T_2 , T_1), нижне-средневизейский терригенный (пласты Мл, Бб₂, Бб₁, Тл_{2-б}, Тл_{2-а}), окско-серпуховско-башкирский карбонатный (пласты Бш₂, Бш₁), московский терригенно-карбонатный (пласт В₃В₄).

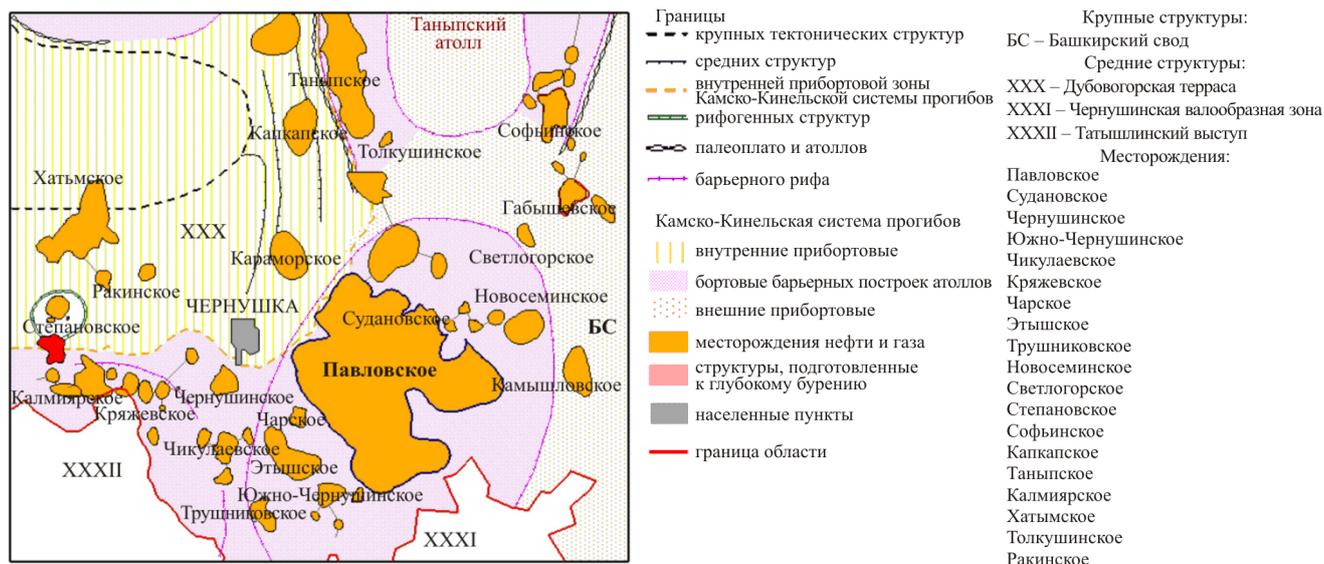


Рис. 1. Выкопировка из сводной схемы тектонического районирования Пермского края

Разработка многомерных моделей выделения нефтенасыщенных песчаников

Методики интерпретации газового каротажа описаны в [1]. Первичные данные, полученные со скважины, обрабатываются в центре интерпретации. Интерпретатор, изучая относительный состав газа и абсолютный процент углеводородных газов в газозудной смеси, поступающей вместе с буровым раствором из скважины, дает заключение о насыщении пластов. При этом возможны ошибки, связанные с различиями в подходе к интерпретации, т.е. человеческим фактором. Предлагается использовать математические методы для того, чтобы создать единую для всех методику интерпретации, которая будет учитывать все необходимые параметры для разделения пластов по типу насыщения флюидами и,

таким образом, математически объяснит интерпретацию газового каротажа, т.е. выделение нефтенасыщенных пород в исследуемых интервалах. Одним из возможных инструментов для этой цели может служить линейный дискриминантный анализ.

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы). Определим сначала понятие разделения (дискриминации) и покажем, чем оно отличается от близкого к нему понятия классификации. Предположим, что имеются две группы проб сланцев, о которых заранее известно, что они образовались в пресноводном и морском бассейнах. Это можно определить на основании исследования остатков ископаемых организмов. Ископаемые остатки в пресноводных и морских бассейнах различаются, что

позволяет легко отличать сланцы друг от друга, но на практике встречаются образцы сланцев без ископаемых остатков, поэтому необходимо найти другой критерий разделения этих групп сланцев. В пробах измерено некоторое число геохимических характеристик, а именно содержания ванадия, бора, железа и других элементов. Задача состоит в нахождении такой линейной комбинации этих переменных, которая даст максимально возможное различие между двумя ранее определенными группами. Если нам удастся найти такую функцию, то мы сможем использовать ее для отнесения новых образцов к той или другой исходной группе. Иными словами, новые образцы сланца, не содержащие диагностических ископаемых остатков, можно будет разделить на морские и пресноводные на основе линейной дискриминантной функции, построенной по их геохимическим компонентам. (Эта задача рассматривалась Поттером, Шимпом и Уиттерсом [4–5]).

Задачу классификации можно проиллюстрировать на аналогичном примере. Предположим, что мы собрали большую коллекцию образцов сланцев, каждый из которых был подвергнут геохимическому анализу. Можно ли на основе значений измеренных переменных осуществить разделение выборки на относительно однородные группы (кластеры), отличающиеся друг от друга? Численные методы решения такого рода задач достаточно хорошо разработаны и принадлежат к разделу науки, называемому таксономией. Существует несколько явных различий между этими методами и методами дискриминантного анализа. Классификация внутренне замкнута, т.е., в отличие от дискриминантного анализа, она не зависит от априорных сведений о соотношении между пробами. В дискриминантном анализе число групп задается заранее, в то время как число кластеров, которые получаются в результате классификации, не может быть заранее определено. Каждая проба из исходного множества в дискриминантном анализе

принадлежит к одной из заданных групп. В большинстве задач классификации проба может войти в любую из групп, возникающих в результате классификации. Другие различия станут очевидными при рассмотрении этих двух процедур. В результате кластерного анализа сланцев пробы распределяются по группам. Представляет интерес проведение геологического осмысливания найденных таким образом групп [5].

Таким же образом можно описать и задачу, рассматриваемую в настоящей статье: для того чтобы определить, какие переменные относят определенный интервал к нефтенасыщенным породам, были собраны данные о следующих переменных:

1. $G_{\text{сум}}$, абс. % – суммарное газосодержание нефтенасыщенных углеводородов в буровом растворе;
2. C_1 , отн. % – относительное содержание метана (CH_4) в газовой смеси;
3. C_2 , отн. % – относительное содержание этана (C_2H_6) в газовой смеси;
4. C_3 , отн. % – относительное содержание пропана (C_3H_8) в газовой смеси;
5. C_4 , отн. % – относительное содержание бутана (C_4H_{10}) в газовой смеси;
6. C_5 , отн. % – относительное содержание пентана (C_5H_{12}) в газовой смеси;
7. ЛБА, баллы – данные люминесцентно-битуринологического анализа шлама;
8. ГК, мкР/ч – показания гамма-каротажа;
9. ННКТ, усл. ед. – показания нейтрон-нейтронного каротажа по тепловым нейтронам;
10. ДС, мм – разница между диаметром долота и диаметром скважины.

Данные ГИС (гамма-каротаж, нейтрон-нейтронный каротаж по тепловым нейтронам и диаметр скважины) были добавлены, чтобы разделить более точно выборку в общей сложности на пять классов: аргиллиты, алевролиты, глинистые песчаники, нефтенасыщенные песчаники, нефтенасыщенные песчаники.

Затем можно использовать дискриминантный анализ для определения такого решающего

правила (дискриминантной функции), которое позволило бы отнести определенный интервал к определенному классу.

Для исследования были взяты данные газового каротажа скважин Павловского месторождения в нижне-средневизейском терригенном нефтегазоносном комплексе в интервале тульского терригенного горизонта C_{1II} . После сбора данных

количество наблюдений в различных классах сильно разнилось, поэтому была проведена выборка случайным образом, для того чтобы уравнять количество наблюдений во всех классах (общее число наблюдений $n = 1535$).

На первом этапе попытки разделения необходимо сравнить средние значения в классах (табл. 1).

Таблица 1

Средние значения переменных в классах $\left(\frac{x \pm \sigma}{\min - \max} \right)$

Показатель	Аргиллит	Алевролит	Песчаник глинистый	Песчаник нефтенасыщенный	Песчаник нефтенасыщенный
$\Gamma_{\text{сум}}$, абс. %	$\frac{0,040 \pm 0,074}{0,001 - 0,388}$	$\frac{0,050 \pm 0,098}{0,001 - 0,551}$	$\frac{0,100 \pm 0,375}{0,001 - 2,599}$	$\frac{0,039 \pm 0,079}{0,001 - 0,396}$	$\frac{0,184 \pm 0,359}{0,004 - 2,899}$
C_1 , отн. %	$\frac{70,359 \pm 18,638}{25,698 - 100,000}$	$\frac{66,701 \pm 14,776}{34,712 - 94,820}$	$\frac{66,358 \pm 13,032}{36,034 - 96,748}$	$\frac{69,026 \pm 14,918}{28,974 - 100,00}$	$\frac{61,648 \pm 13,938}{23,155 - 91,704}$
C_2 , отн. %	$\frac{15,881 \pm 11,685}{0,001 - 52,936}$	$\frac{17,439 \pm 8,831}{0,001 - 45,052}$	$\frac{18,348 \pm 8,795}{0,001 - 55,847}$	$\frac{16,969 \pm 9,415}{0,001 - 45,240}$	$\frac{18,146 \pm 6,863}{4,552 - 39,829}$
C_3 , отн. %	$\frac{7,746 \pm 7,378}{0,001 - 29,060}$	$\frac{9,955 \pm 6,500}{0,001 - 27,323}$	$\frac{10,845 \pm 7,365}{0,001 - 37,057}$	$\frac{10,190 \pm 8,014}{0,001 - 41,480}$	$\frac{12,628 \pm 6,981}{0,001 - 41,287}$
C_4 , отн. %	$\frac{4,309 \pm 4,843}{0,001 - 20,670}$	$\frac{4,029 \pm 3,393}{0,001 - 14,429}$	$\frac{3,253 \pm 3,258}{0,00 - 14,205}$	$\frac{2,684 \pm 3,195}{0,001 - 18,533}$	$\frac{5,158 \pm 4,608}{0,001 - 22,977}$
C_5 , отн. %	$\frac{1,687 \pm 2,632}{0,001 - 13,265}$	$\frac{1,896 \pm 3,764}{0,001 - 41,186}$	$\frac{1,216 \pm 2,039}{0,001 - 9,792}$	$\frac{1,160 \pm 1,817}{0,001 - 10,325}$	$\frac{2,397 \pm 3,276}{0,001 - 15,194}$
ННКТ, усл. ед.	$\frac{2,397 \pm 2,216}{0,573 - 16,418}$	$\frac{7,540 \pm 2,552}{4,000 - 14,976}$	$\frac{9,751 \pm 5,120}{2,459 - 22,353}$	$\frac{8,243 \pm 4,902}{2,807 - 23,236}$	$\frac{8,835 \pm 5,606}{2,532 - 22,406}$
ГК, мкР/ч	$\frac{10,620 \pm 3,793}{0,930 - 20,441}$	$\frac{14,200 \pm 2,475}{10,119 - 26,025}$	$\frac{7,014 \pm 1,597}{4,26 - 10,048}$	$\frac{2,836 \pm 0,812}{1,065 - 4,495}$	$\frac{2,807 \pm 0,791}{0,910 - 4,485}$
ДС, мм	$\frac{17,697 \pm 24,779}{-8,385 - 151,745}$	$\frac{4,823 \pm 6,430}{-14,000 - 29,000}$	$\frac{1,111 \pm 4,442}{-7,400 - 39,635}$	$\frac{-1,267 \pm 2,631}{-7,148 - 26,000}$	$\frac{-1,428 \pm 2,986}{-13,000 - 30,000}$

Отсюда видно, что флюиды из нефтенасыщенных интервалов имеют более «тяжелый» состав газа (преобладание сложных углеводородов по сравнению с остальными интервалами). Самые высокие значения $\Gamma_{\text{сум}}$ (0,184) соответствуют нефтенасыщенному песчанику, меньшие значения (0,100) – глинистому песчанику, скорее всего, эти показания связаны с переходной зоной между нефтенасыщенным песчаником и аргиллитами. Покомпонентный состав газа в нефтенасыщенных песчаниках указывает на высокое содержание «тяжелых» углеводородных газов и низкое

содержание метана в нефтенасыщенных песчаниках и обратные характеристики у остальных классов, например, аргиллит характеризуется самым высоким относительным содержанием метана и низкими относительными содержаниями газов (от этана до пентана). Самые высокие средние значения переменной ДС относятся к аргиллитам, что, вероятно, связано с вывалами со стенок скважины, значения ДС в нефтенасыщенных и нефтенасыщенных песчаниках самые низкие, что можно объяснить образованием глинистой корочки.

Таким образом, для выделения нефтенасыщенных интервалов надо использовать совместно все эти переменные. Данная задача была решена с применением элементов математической статистики – пошагового дискриминантного анализа, ход работы и результаты которого описаны в настоящей статье.

Выполненный анализ средних значений и плотностей распределений изученных переменных показал, что ни по одной из переменных не представляется возможным выполнить разделение на вышеприведенные классы. Поэтому для комплексного (совместного) использования изучаемых показателей будем использовать линейный дискриминантный анализ (ЛДА). Его возможности для решения аналогичных задач приведены в работах [6–21].

Простая линейная дискриминантная функция осуществляет преобразование исходного множества измерений, входящих в выборку, в единственное дискриминантное число. Это число, или преобразованная переменная, обуславливает положение образца на прямой, определенной дискриминантной функцией. Поэтому мы можем представлять себе дискриминантную функцию как способ преобразования многомерной задачи в одномерную [22, 23].

Дискриминантный анализ основан на нахождении преобразования, которое дает минимум отношения разности многомерных средних значений для некоторой пары групп к многомерной дисперсии в пределах двух групп. Если мы изобразим наши две группы совокупностями точек в многомерном пространстве, то легко найти такое направление, вдоль которого эти совокупности явно разделяются и в то же время имеют наименьшую выпуклость. Покажем на графике возможность разделения нефтенасыщенных песчаников и аргиллитов по двум наиболее информативным параметрам (рис. 2). Если использовать переменные «ГК, мкР/ч» и «ЛБА, баллы», то провести удовлетворительное разделение групп А (нефтенасыщенные песчаники) и

В (аргиллиты) не удастся. Однако можно найти направление, вдоль которого разделение совокупностей очевидно, а выпуклость минимальна. Координаты точек этого направления задаются уравнением линейной дискриминантной функции. Указаны перекрытия распределений для групп А и В по осям «ГК, мкР/ч» и «ЛБА, баллы»; проектирование на дискриминантную линию позволяет различить две группы [5].

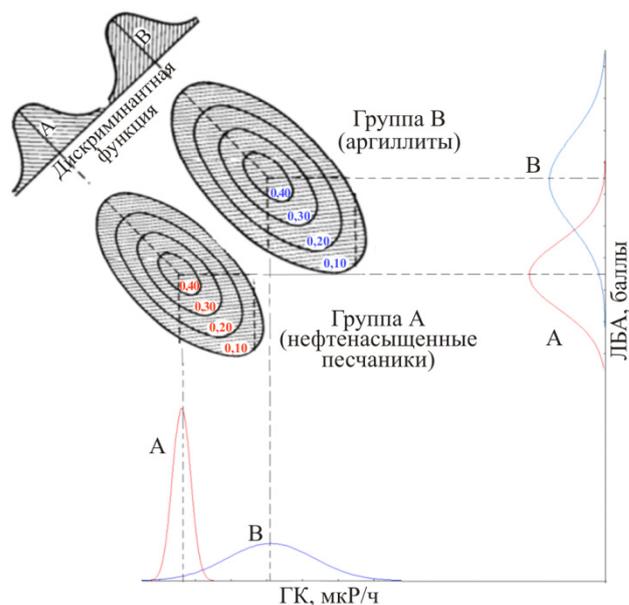


Рис. 2. Графическое представление двух двумерных распределений

Дискриминантный анализ преследует такие цели:

1. Определение дискриминантных функций или линейных комбинаций независимых переменных, которые наилучшим образом различают (дискриминируют) категории (группы) зависимой переменной.
2. Проверка существования между группами значимых различий с точки зрения независимых переменных.
3. Определение предикторов, вносящих наибольший вклад в межгрупповые различия.
4. Отнесение случаев к одной из групп (классификация) исходя из значений предикторов.
5. Оценка точности классификации данных на группы [24–29].

Дискриминантная функция – выведенная посредством дискриминантного анализа линейная комбинация независимых переменных, с помощью которой можно наилучшим образом различить (дискриминировать) категории зависимой переменной [21].

Метод дискриминантного анализа описывается числом категорий, имеющих у зависимой переменной. Если она имеет две категории, то метод называют дискриминантным анализом для двух групп (two-group discriminant analysis). Если анализируют три категории или больше, то метод называют множественным дискриминантным анализом (multiple discriminant analysis). Главное отличие между ними заключается в том, что при наличии двух групп можно вывести только одну дискриминантную функцию. Используя множественный дискриминантный анализ, можно вычислить несколько функций [30, 31].

С помощью ЛДА представляется возможным построение оптимальных поверхностей (дискриминантных функций) Z в пространстве признаков, разделяющих все пространство на области, соответствующие объектам разных классов. Эти поверхности (функции) служат границами между областями и обеспечивают оптимальное разделение объектов, относящихся к разным классам. Правило классификации в этом случае заключается в определении по величине дискриминантной функции принадлежности объекта к той или иной из выделенных областей. Принадлежность к соответствующей области означает принадлежность соответствующему классу [5].

Количество дискриминантных функций определяется как $K-1$ или $P-1$, где K – число групп, а P – число переменных. Как правило, для определения количества необходимых дискриминантных функций выбирают наименьшее из этих чисел [32–34]. В рассматриваемом анализе используется 10 переменных и 5 групп, т.е. дискриминантных функций должно быть четыре.

Выполнение дискриминантного анализа включает следующие стадии: формулирование

проблемы, вычисление коэффициентов дискриминантной функции, определение значимости, интерпретация и проверка достоверности.

Первый шаг дискриминантного анализа – формулирование проблемы путем определения целей, зависимой переменной и независимых переменных. Зависимая переменная должна состоять из двух или более взаимоисключающих и взаимно исчерпывающих категорий.

Если зависимая переменная измерена с помощью интервальной или относительной шкалы, то ее следует в первую очередь перевести в статус категориальной. Следующий шаг – разделение выборки на две части. Одна из них – анализируемая выборка (analysis sample) – используется для вычисления дискриминантной функции. Другая часть – проверочная выборка (validation sample) – предназначена для проверки дискриминантной функции. Это называется двойной перекрестной проверкой [34–38].

Часто распределение количества случаев в анализируемой и проверочной выборках является из распределения в общей выборке. Например, если общая выборка содержит 50 % нефтенасыщенных и 50 % ненасыщенных интервалов, то анализируемая и проверочная выборки должны каждая содержать 50 % нефтенасыщенных и 50 % ненасыщенных интервалов. В другом случае, если выборка содержит 25 % нефтенасыщенных и 75 % ненасыщенных интервалов, следует выбрать анализируемую и проверочную выборки таким образом, чтобы их распределение отражали аналогичную картину (25 против 75 %).

И наконец, проверку достоверности дискриминантной функции предлагают выполнять неоднократно. Каждый раз выборку следует разбивать на две части: для анализа и проверки. Вычисляют дискриминантную функцию и выполняют анализ достоверности модели. Таким образом, оценка достоверности основана на ряде испытаний.

После определения анализируемой выборки мы можем вычислить коэффициенты дискриминантной функции, используя две методики. Прямой метод (*direct method*) – вычисление дискриминантной функции при одновременном введении всех предикторов.

В этом случае учитывается каждая независимая переменная. При этом ее дискриминирующая сила не принимается во внимание. Этот метод больше подходит к ситуации, когда аналитик, исходя из результатов предыдущего исследования или теоретической модели, хочет, чтобы в основе различия лежали все предикторы.

При пошаговом дискриминантном анализе (*stepwise discriminant analysis*) предикторы вводят последовательно, исходя из их способности различить (дискриминировать) группы. Этот метод лучше применять в ситуации, когда исследователь хочет отобрать подмножество предикторов для включения их в дискриминантную функцию. Есть две разновидности пошагового анализа:

1. Пошаговый анализ с включением. В пошаговом анализе дискриминантных функций модель дискриминации строится по шагам. Точнее, на каждом шаге просматриваются все переменные и находится та из них, которая вносит наибольший вклад в различие между совокупностями. Эта переменная должна быть включена в модель на данном шаге, и происходит переход к следующему шагу.

2. Пошаговый анализ с исключением. Можно также двигаться в обратном направлении, в этом случае все переменные будут сначала включены в модель, а затем на каждом шаге будут устраняться переменные, вносящие малый вклад в предсказания. Тогда в качестве результата успешного анализа можно сохранить только важные переменные в модели, т.е. те переменные, чей вклад в дискриминацию больше остальных [39–40].

Математически дискриминантные функции можно описать следующим уравнением:

$$Z_{ik} = a + b_1 y_{i1} + b_2 y_{i2} + \dots + b_p y_{ip},$$

где a – константа; b – стандартизованный коэффициент (со средним 0 и выборочной дисперсией 1), позволяет оценить вклад каждой переменной в данную дискриминантную функцию [7].

Рассмотрим возможность использования метода линейной дискриминантной функции на примере ряда скважин Павловского месторождения в интервале ниже-средне-визейского терригенного нефтегазоносного комплекса. Полученные в каждой точке значения параметров газового каротажа были сгруппированы в пять классов: аргиллиты, алевролиты, песчаники, песчаники глинистые и песчаники нефтенасыщенные. Группирование было проведено согласно данным интерпретации ГТИ.

Таким образом, были получены группирующая переменная (порода) и независимые переменные (параметры газового каротажа – $\Gamma_{\text{сум}}$, C_1 , C_2 , C_3 , C_4 , C_5 , ЛБА и данные ГИС – ГК, ННКТ, ДС), которые будут использованы для дискриминации. Так как все эти переменные имеют разные единицы измерения, то возможны ситуации, когда более значимый параметр, с небольшим значением величины, перекрывается менее значимым, у которого значение больше. Таким образом мы получаем информацию плохого качества, т.е. данные могут оказаться несравнимыми. Для исключения такой ситуации была проведена нормализация данных – преобразование формальных параметров или критериев оценки эффективности объекта, выражаемых в общем случае в различных единицах, к безразмерному виду с целью их сопоставления и сравнительной оценки. Для нормализации была использована следующая формула:

$$X'_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}.$$

Далее, чтобы проследить, что происходит на каждом шаге дискриминантного анализа, был выполнен пошаговый анализ с

включением, позволяющий вводить переменные в модель постепенно, одну за другой, каждый раз выбирая ту, которая вносит наибольший вклад в дискриминацию. Данный анализ будет производиться до тех пор, пока не произойдет одно из представленных ниже событий:

1. Все переменные введены или отброшены.

2. Достигнуто максимальное число шагов.

3. Нет других переменных вне модели, имеющих большее значение статистики F , чем заданное значение F -включить, равное 1, и когда в модели нет других переменных, имеющих меньшее значение F , чем значение F -исключить, равное 0. При пошаговом анализе с включением отбираются переменные для включения, дающие наиболее значащий единственный (дополнительный) вклад в дискриминацию между совокупностями, т.е. выбираются переменные с наибольшим значением F (большим, чем соответствующее значение F -включить, равное 1). При выполнении шагов с исключением отбираются для исключения наименее значимые переменные, т.е. переменные с наименьшим значением F (меньшим, чем соответствующее значение F -исключить, равное 0).

4. Какая-либо переменная на следующем шаге имеет значение толерантности меньше, чем значение толерантности, равное 0,01. На каждом шаге для каждой переменной вычисляется множественная корреляция (R^2) со всеми другими переменными, которые были включены в модель. Таким образом, значение толерантности переменной вычисляется как $1 - R^2$, поэтому значение толерантности является мерой избыточности переменной. Также необходимо отметить, что когда одна включенная переменная или более становится слишком избыточной, то матрица дисперсий/ковариаций для переменных, включенных в модель, может оказаться необратимой и дискриминантный анализ не может быть выполнен. Следовательно, граничное значение толерантности было задано 0,01, ведь если переменная,

включенная в модель, сократима с другими переменными более чем на 99 %, то ее практический вклад в улучшение качества дискриминации весьма незначителен. Более важно то, что если задать значительно меньшее значение толерантности, то ошибки округления могут привести к неустойчивым результатам [4].

Результаты проведенного пошагового дискриминантного анализа представлены в табл. 2. Были получены критерии значимости, которые приведены ниже.

Критерий лямбда Уилкса является стандартной статистикой, используемой для обозначения статистической значимости мощности дискриминации в текущей модели. Ее значение меняется от 1 (нет никакой дискриминации) до 0 (полная дискриминация).

Таблица 2

Результаты анализа дискриминантных функций

Переменная	Лямбда Уилкса	Частная лямбда	F -исключить (2,447)	p -уровень
ГК	0,390462	0,232090	1258,947	0,000000
ЛБА	0,118121	0,767203	115,457	0,000000
ННКТ	0,107498	0,843015	70,856	0,000000
ДС	0,102912	0,880579	51,602	0,000000
C_4	0,092164	0,983277	6,471	0,000037
C_2	0,092219	0,982682	6,706	0,000024
C_1	0,091740	0,987821	4,691	0,000915
$\Gamma_{\text{сум}}$	0,091702	0,988229	4,532	0,001216
C_3	0,091327	0,992283	2,959	0,018916

Частная лямбда Уилкса – это статистика лямбда Уилкса для одиночного вклада соответствующей переменной в дискриминацию между совокупностями. Это значение можно рассматривать как аналог частного коэффициента корреляции. Лямбда со значением 0 обозначает полную дискриминацию, следовательно, чем ниже ее значение, тем больше одиночный вклад соответствующей переменной в степень дискриминации.

Таким образом, по результатам пошагового дискриминантного анализа можно сделать вывод, что наибольший вклад в общую дискриминацию вносит переменная ГК – это связано с выделением

глинистых интервалов от интервалов коллекторов (показания ГК повышены в аргиллитах и понижены в песчаниках). Второй по значимости переменной является ЛБА – этот параметр, скорее всего, выделяет интервалы нефтенасыщенных песчаников, так как свечение шлама при люминесцентно-битуминологическом анализе является прямым признаком наличия нефти в пласте. Третьим значимым параметром является ННКТ – это, скорее всего, связано с отделением аргиллитов от остальных классов. Четвертым по значимости параметром является ДС – он, скорее всего, также отделяет интервалы аргиллитов от других классов, но и вносит вклад в выделение песчаников с коллекторскими свойствами. Это видно, если сравнить средние значения ДС в разных классах (см. табл. 1) – песчаники с хорошими коллекторскими свойствами при бурении покрываются глинистой коркой, и происходит сужение скважины. Остальные переменные практически не влияют на дискриминацию данных по классам, показывая низкие значения вкладов в дискриминацию.

Чтобы узнать, как четыре переменные разделяют различные классы, найдем действительную дискриминантную функцию. С помощью канонического анализа вычислим различные независимые (ортогональные) дискриминирующие функции. Каждая последующая дискриминантная функция будет вносить все меньший и меньший вклад в общую дискриминацию. Максимальное число оцениваемых функций равно числу переменных или числу классов минус один, в зависимости от того, какое число меньше. В нашем случае оцениваются четыре дискриминирующие функции. Сначала определим, являются ли обе дискриминантные функции (корни) статистически значимыми (табл. 3).

В табл. 3 приведен отчет о пошаговом критерии с включением для всех канонических корней. Первая строка содержит критерий значимости для всех

корней, вторая – данные о значимости корней, оставшихся после удаления первого корня и т.д. Таким образом, в этой таблице показано, как много канонических корней (дискриминирующих функций) следует интерпретировать. В нашем случае получено четыре статистически значимые дискриминантные функции, которые можно использовать для разделения на классы.

Таблица 3

Критерии χ^2 последовательных корней

Корни исключенные	Лямбда Уилкса	χ^2	p-уровень
0	0,090622	3666,409	0,000000
1	0,511245	1024,473	0,000000
2	0,771769	395,600	0,000000
3	0,981939	27,831	0,000101

Для того чтобы вычислить веса дискриминантной функции и записать дискриминантные функции, узнаем исходные коэффициенты для канонических переменных (табл. 4).

Таблица 4

Исходные коэффициенты для канонических переменных

Параметр	Z ₁	Z ₂	Z ₃	Z ₄
ГК	0,44205	0,141440	-0,03656	0,00967
ЛБА	-0,23902	0,320572	-0,64354	-0,09412
ННКТ	-0,03586	0,144381	0,05269	-0,11102
ДС	-0,00210	-0,044725	-0,03096	-0,05547
C ₄	0,05188	-0,105805	-0,07426	-0,11248
C ₂	0,04584	-0,048629	0,01273	-0,17474
C ₁	0,03131	-0,047732	0,02560	-0,15797
Г _{сум}	-0,17723	0,206584	-0,89272	-0,52816
C ₃	0,01530	-0,031519	0,04850	-0,14971
Константа	-5,99788	2,447964	-1,31175	16,69311
Кумулятивная доля	0,85293	0,946577	0,99662	1,00000

$$Z_1 = -5,998 + 0,442ГК - 0,239ЛБА - 0,036ННКТ - 0,002ДС + 0,052C_4 + 0,046C_2 + 0,031C_1 - 0,177Г_{сум} + 0,015C_3;$$

$$Z_2 = 2,448 + 0,141ГК + 0,321ЛБА + 0,144ННКТ - 0,045ДС - 0,106C_4 - 0,049C_2 - 0,048C_1 + 0,207Г_{сум} - 0,032C_3;$$

$$Z_3 = -1,312 - 0,037ГК - 0,644ЛБА + \\ + 0,053ННКТ - 0,031ДС - 0,074C_4 + \\ + 0,013C_2 + 0,026C_1 - 0,893Г_{\text{сум}} + 0,049C_3;$$

$$Z_4 = 16,693 + 0,010ГК - 0,094ЛБА - \\ - 0,111ННКТ - 0,055ДС - 0,112C_4 - \\ - 0,175C_2 - 0,158C_1 - 0,528Г_{\text{сум}} - 0,150C_3.$$

Первая дискриминантная функция использует наиболее информативные переменные ГК, ЛБА и $G_{\text{сум}}$ (самые высокие исходные коэффициенты, см. табл. 3). Вторая и третья дискриминантные функции взвешиваются наиболее тяжело переменными ЛБА и $G_{\text{сум}}$. Четвертая функция взвешивается наиболее тяжело переменной $G_{\text{сум}}$. Другие переменные также дают вклад в эти функции, но меньший.

В табл. 4 приведены собственные значения (корни) для каждой дискриминантной функции и кумулятивная доля объясненной дисперсии, накопленной каждой функцией. Как вы можете видеть, первая функция ответственна за 85,29 % объясненной дисперсии, т.е. 85,29 % всей дискриминирующей мощности объясняется этой функцией. Таким образом, ясно, что эта первая функция наиболее важна.

Теперь мы знаем, какие переменные участвуют в дискриминации между различными классами. Следующая задача заключается в том, чтобы определить природу дискриминации для каждого канонического корня. Канонические средние первого шага представлены в табл. 5.

Таблица 5

Канонические средние
переменных первого шага

Класс	Z_1	Z_2	Z_3	Z_4
Аргиллиты	1,75708	-1,20724	-0,320186	-0,039617
Алевролиты	3,10252	0,75046	0,118458	0,117952
Песчаники глинистые	-0,35301	0,52841	0,124485	-0,248422
Песчаники	-1,99839	-0,36809	0,792809	0,083623
Песчаники нефтенасыщенные	-2,52714	0,32116	-0,771621	0,086569

Очевидно, что первая дискриминантная функция отделяет главным образом

аргиллиты и алевролиты (т.е. плотные глинистые породы) от других классов – песчаников и нефтенасыщенных песчаников (средние канонических переменных сильно различаются – от -2,52714 у нефтенасыщенных песчаников до 3,10252 у алевролитов, при этом глинистые песчаники располагаются примерно посередине этого диапазона со значениями -0,35301, но все равно больше тяготеют к песчаникам). Вторая дискриминантная функция, по-видимому, предназначена для разделения аргиллитов и алевролитов, однако, как и следовало ожидать, основываясь на рассмотренных ранее собственных значениях, качество теперь немного хуже и будет ухудшаться с каждой новой функцией. Третья дискриминантная функция, скорее всего, делит нефтенасыщенные песчаники и нефтенасыщенные песчаники, а четвертая, по-видимому, служит для выделения глинистых песчаников.

Быстрый способ визуализации этих результатов заключается в выводе на экран диаграммы рассеяния дискриминантных функций (рис. 3–5).

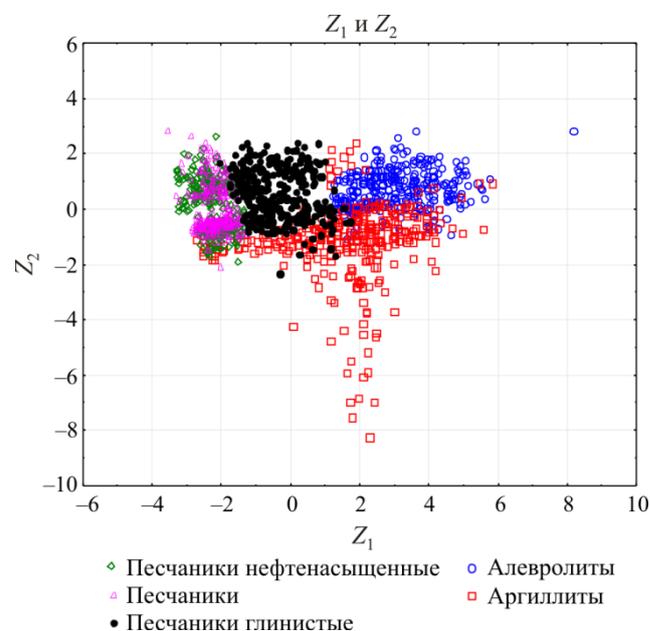


Рис. 3. Диаграммы рассеяния
для канонических значений

Отсюда видно, что наилучшее разделение получено по функциям 1 и 2. Видно, что

ненефтенасыщенные и нефтенасыщенные песчаники представлены на диаграмме в левой части, и они отделены от аргиллитов и алевролитов глинистыми песчаниками. Поэтому первая дискриминантная функция главным образом делит ненефтенасыщенные песчаники и нефтенасыщенные песчаники от алевролитов и аргиллитов. Также видна некоторая дискриминация между классами аргиллитов, алевролитов, глинистых песчаников. Однако дискриминация здесь не настолько отчетлива, как для первой канонической функции (корня).

Для подведения итогов заметим, что наиболее ясная дискриминация возможна с использованием первой дискриминантной функции. Эта функция отмечена отрицательными коэффициентами для наличия свечения шлама при ЛБА, суммарного содержания нефтенасыщенных углеводородов в буровом растворе и показаний НКТ и положительными весами для показаний ГК, ДС и относительного содержания метана, этана, пропана и бутана. Таким образом, чем больше показания ГК и ДС и меньше суммарное содержание нефтенасыщенных углеводородов в буровом растворе, а также если отсутствует свечение шлама, тем более вероятно, что это аргиллиты и алевролиты, т.е. породы-неколлекторы. Проанализировав коэффициенты третьей дискриминационной функции, видно, что с ее помощью можно разделить песчаники и нефтенасыщенные песчаники. Это обусловлено тем, что чем больше суммарное содержание нефтенасыщенных углеводородов в буровом растворе и если присутствует свечение шлама, тем вероятнее, что это нефтенасыщенные песчаники.

Также с помощью дискриминантного анализа были получены апостериорные вероятности для каждого класса. Данные значения показывают вероятность того, что наблюдение принадлежит к определенному классу. Для того чтобы графически изобразить эти вероятности, был создан геолого-геофизический планшет, на который были нанесены: насыщение пород согласно

ГТИ, насыщение пород согласно ГИС и литологическая колонка по данным ГТИ и ГИС, а также все переменные, которые участвовали в дискриминантном анализе (рис. 4).

На этом фрагменте можно увидеть следующую картину: напротив нефтенасыщенных пород кривая вероятности находится в правой части (значения 0,4–1,0), т.е. вероятность отнесения этого интервала к классу нефтенасыщенных пород высока. Напротив плотных пород кривая вероятности находится в левой части (значения около 0), т.е. вероятность отнесения этого интервала к классу нефтенасыщенных пород очень низка. Напротив некоторых интервалов кривая вероятности показывает промежуточные значения, т.е. это интервалы, на которые интерпретатору надо обратить внимание.

Заключение

Таким образом, использование пошагового дискриминантного анализа позволило определить значение вероятности отнесения наблюдений к классу пород – аргиллиты, алевролиты, глинистые песчаники, ненефтенасыщенные песчаники, нефтенасыщенные песчаники. Анализ распределения геолого-геофизических данных совместно с вычисленной вероятностью отнесения наблюдений к классу нефтенасыщенных песчаников позволил выделить следующие области: первая зона – со значениями вероятности около нуля (зона ненефтенасыщенных пород), вторая зона – со значениями вероятности от 0,4 до 1 (зона нефтенасыщенных песчаников) и третья зона – с промежуточными значениями вероятности (зона неясного насыщения). Эта зона, на которую должен обратить особое внимание интерпретатор, так как она может содержать нефтенасыщенные песчаники, но не полностью себя проявила из-за отсутствия каких-нибудь данных, например, свечения шлама при ЛБА. Для интерпретации этих неясных интервалов необходимо воспользоваться дополнительными данными, например методами ГИС или результатами отбора и описания керна, если таковые имеются.

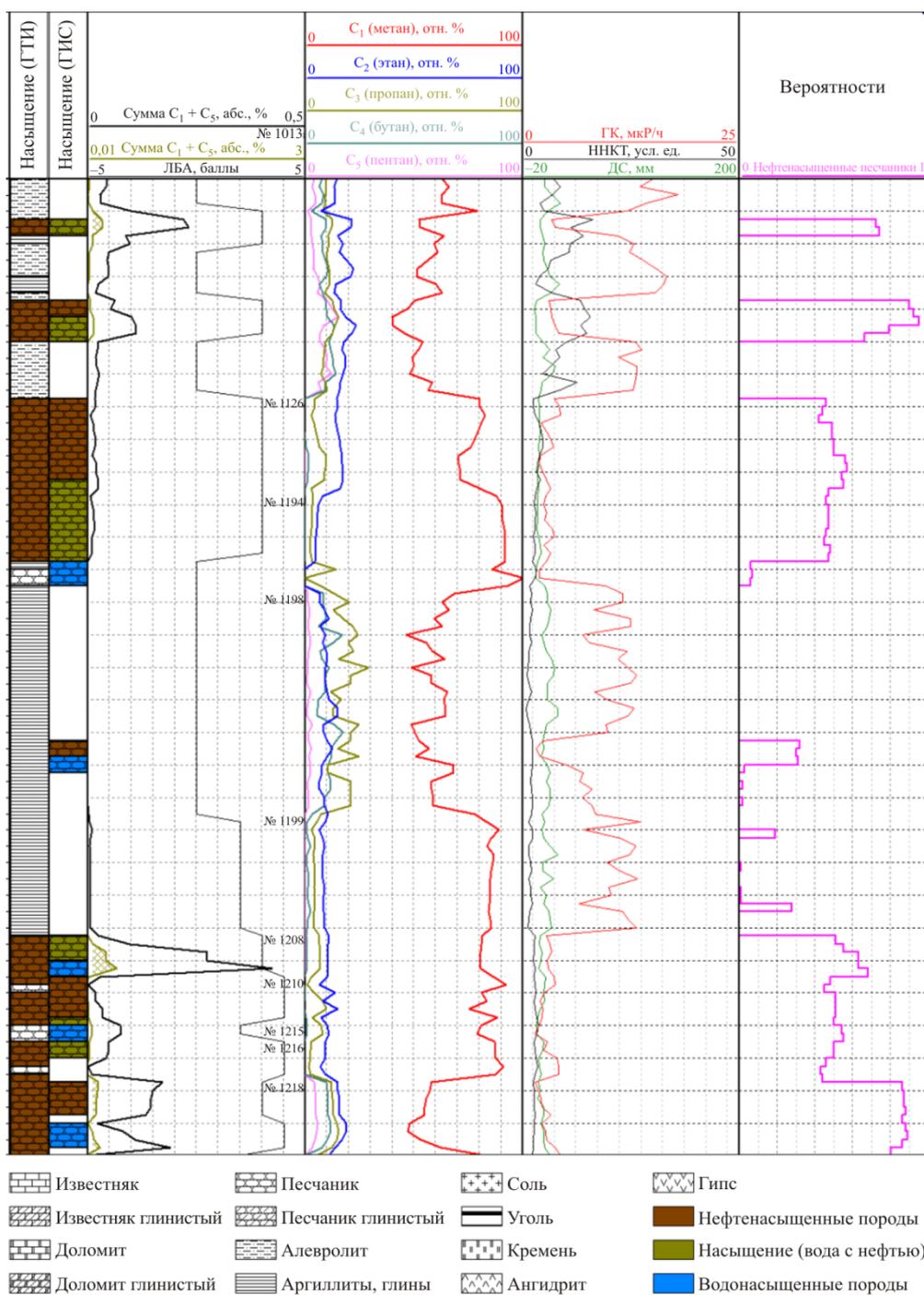


Рис. 4. Зависимости изменения вероятности наличия нефтяных пластов от геолого-геофизических параметров

Библиографический список

1. Геофизические исследования и работы в скважинах: в 7 т. Т. 7. Геолого-технологические исследования в скважинах / сост. С.Н. Шматченко. – Уфа: Информ реклама, 2010. – 248 с.
 2. ГОСТ Р 53375-2009. Скважины нефтяные и газовые. Геолого-технологические

исследования. Общие требования. М.: Стандартинформ, 2009. – 24 с.
 3. РД 153-39.0-069-01. Техническая инструкция по проведению геолого-технологических исследований нефтяных и газовых скважин. – М., 2001.
 4. Potter P.E., Skimp N.F., Witters I. Trace elements in marine and fresh-water argillaceous sediments // Geochimica et Cosmochimica. –

1963. – Vol 27, iss 6. – P. 669–694.
DOI: 10.1016/0016-7037(63)90019-X

5. Davis J.C. Statistics and data analysis in geology. – 3rd ed. – John Wiley & Sons, 2002. – 656 p.

6. Statsoft [Электронный ресурс]: электрон. учеб. по статистике. – URL: <http://statsoft.ru/home/textbook/default.htm> (дата обращения: 08.05.2018).

7. Кривошеков С.Н., Галкин В.И., Козлова И.А. Определение перспективных участков геолого-разведочных работ на нефть вероятностно-статистическими методами на примере территории Пермского края // Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело. – 2012. – № 4. – С. 7–15.

8. Матерон Ж. Основы прикладной геостатистики. – М., 1968. – 408 с.

9. Соснин Н.Е. Разработка статистических моделей для прогноза нефтегазоносности (на примере терригенных девонских отложений Северо-Татарского свода) // Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело. – 2012. – № 5. – С. 16–25.

10. Денисов С.В., Исмагилов Р.Н., Сидельников К.А. Результаты применения метода последовательного прогнозирования вероятностей для дискриминантного анализа простых моделей пласта // Известия Тульского государственного университета. Естественные науки. – 2009. – С. 205–210.

11. Денисов С.В., Исмагилов Р.Н., Сидельников К.А. Результаты применения метода последовательного прогнозирования вероятностей для дискриминантного анализа сложных моделей пласта // Известия Тульского государственного университета. Естественные науки. – 2009. – С. 211–216.

12. Применение математических методов при анализе геологической информации (с использованием компьютерных технологий): учеб. пособие / сост. И.М. Михалевич, С.П. Примина. – Иркутск: Иркут. гос. ун-т, 2006. – Ч. III. – 115 с.

13. Поротников А.В., Попов М.П., Горбунова Н.П. Применение линейного дискриминантного анализа для автоматического определения происхождения изумруда по данным рентгенофлюоресцентного анализа // Труды института геологии и геохимии им. академика А.Н. Заварицкого. – Екатеринбург: Институт геологии и геохимии им. академика А.Н. Заварицкого, 2013. – С. 353–355.

14. Кравцов Б.А., Милютин Л.И. Возможности применения многомерной классификации при изучении популяций древесных растений // Пространственно-временная структура лесных биогеоценозов. – Новосибирск: Наука, 1981. – С. 47–65.

15. Урбах В.Ю. Дискриминантный анализ и его применение в биологической систематике и медицинской диагностике // Применение математических методов в биологии. – Л., 1964. – Вып. 3. – С. 67–87

16. Каган Е.С., Морозова И.С. Изучение факторов оптимизации познавательной деятельности студентов с помощью методов кластерного и дискриминантного анализов // Сибирская психология сегодня. – Кемерово: Кузбассвузиздат, 2002. – С. 36–41.

17. Гиниятуллин К.Г., Валеева А.А., Смирнова Е.В. Использование кластерного и дискриминантного анализов для диагностики литологической неоднородности почвообразующей породы по гранулометрическому составу. – М.: Наука, 2017. – С. 946–953.
DOI: 10.7868/S0032180X17080044

18. Dat Thanh Tran, Moncef Gabbouj, Alexandros Iosifidis. Multilinear class-specific discriminant analysis // Pattern Recognition Letters. – 2017. – Vol. 100, iss. 1. – P. 131–136.
DOI: 10.1016/j.patrec.2017.10.027

19. Yong Wang, Jian-Bin Xie, Yi Wu. Orthogonal discriminant analysis revisited // Pattern Recognition Letters. – 2016. – Vol. 84. – P. 149–155. DOI: 10.1016/j.patrec.2016.09.010

20. Andrewartha H.G., Birch L.C. The distribution and abundance of animals. – Chicago: Univ. Press, 1954. – 782 p.

21. Cacoullos T. Discriminant analysis and applications. – London: Academic Press, Inc. ltd, 1972. – 434 p.

22. Факторный, дискриминантный и кластерный анализ / Дж.О. Ким, Ч.У. Мьюллер, У.Р. Клекка [и др.]. – М.: Финансы и статистика, 1989. – 215 с.

23. Горбань А.Н. Обучение нейронных сетей. – М.: ParaGraph, 1990. – 160 с.

24. Андерсен Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963. – 500 с.

25. Morrison D.F. Multivariate statistical methods. – 2nd ed. – New York: McGraw-Hill, Inc., 1976. – 415 p.

26. Morrison D.F. Applied statistical methods. – Prentice-Hall, Inc, Englewood Cliffs, N.J., 1983. – 562 p.

27. Anderson T.W. An introduction to multivariate statistical analysis. – New York: John Wiley and Sons, Inc., 1958. – 374 p.

28. Cooley W.W., Lohnes P.R. Multivariate data analysis. – New York: John Wiley and Sons, Inc., 1971. – 364 p.

29. Fisher R.A. The precision of discriminant function // *Annals of Human Genetics*. – 940. – Vol. 10, iss. 1. – P. 422–429. DOI: 10.1111/j.1469-1809.1940.tb02264.x

30. Боровиков В.П. Статистика для студентов и инженеров. – М.: КомпьютерПресс, 2001. – 301 с.

31. Гмурман В.С. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1972. – 368 с.

32. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240 с.

33. Machine learning, neural and statistical classification / ed. D. Mitchie [et al.]. – Ellis Horwood, Chichester, 1994. – 304 p.

34. Li C.C. Introduction to experimental statistics. – New York: McGraw-Hill, Inc., 1964. – 460 p.

35. Marriot F.H.C. The interpretation of multiple observations. – London Academic Press, Inc., Ltd., 1974. – 117 p.

36. Efron M.A. Multiple regression analysis // *Mathematical Methods for Digital Computers*. – 1960. – Vol. 1. – P. 191–203.

37. Айвазян С.А., Степанов В.С. Инструменты статистического анализа данных // *Мир ПК*. – 1997. – № 8.

38. Прикладная статистика. Классификация и снижение размерностей / С.А. Айвазян, В.М. Буштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.

39. Бикел П., Доксам К. Математическая статистика. – М.: Финансы и статистика, 1983. – Вып. 1. – 278 с.; Вып. 2. – 254 с.

40. Михалевич И.М., Алферова М.А., Рожкова Н.Ю. Основы прикладной статистики: в 3 т. – Иркутск: РИО Гос. ин-та усовершенствования врачей, 2008. – Т. 3. – 92 с.

References

1. Shmatchenko S.N. Geofizicheskie issledovaniia i raboty v skvazhinakh. Vol.7: Geologo-tekhnologicheskie issledovaniia v skvazhinakh [Geophysical surveys in wells. Vol.7: Geological and technological research in wells]. Ufa, Inform Reklama, 2010, 248 p.

2. GOST R 53375-2009. Skvazhiny nefianye i gazovye. Geologo-tekhnologicheskie issledovaniia. Obshchie trebovaniia [Oil and gas wells. Geological-technological logging. General requirements]. Moscow, Standartinform, 2009, 24 p.

3. RD 153-39.0-069-01. Tekhnicheskaia instruksiia po provedeniiu geologo-tekhnologicheskikh issledovaniï nefianykh i gazovykh skvazhin [Technical instruction for conducting geological and technological studies of oil and gas wells]. Moscow, 2001.

4. Potter P.E., Skimp N.F., Witters I. Trace elements in marine and fresh-water argillaceous sediments. *Geochimica et Cosmochimica*, 1963, vol.27, iss.6, pp.669-694. DOI: 10.1016/0016-7037(63)90019-X

5. Davis J.C. Statistics and data analysis in geology. 3rd ed. John Wiley & Sons, 2002. 656 p.

6. Statsoft, available at: URL: <http://statsoft.ru/home/textbook/default.htm> (accessed 08 May 2018).

7. Krivoshchekov S.N., Galkin V.I., Kozlova I.A. Determination of potentially oil bearing areas by behavioral method by the example of Perm region (krai). *Perm Journal of Petroleum and Mining Engineering*, 2012, no.4, pp.7-15.

8. Materon Zh. Osnovy prikladnoi geostatistiki [Basics of applied geostatistics]. Moscow, 1968, 408 p.

9. Sosnin N.E. Development of statistical models for predicting oil-and-gas content (on the example of terrigenous devonian sediments of North Tatar arch). *Perm Journal of Petroleum and Mining Engineering*, 2012, no.5, pp.16-25.

10. Denisov S.V., Ismagilov R.N., Sidelnikov K.A. Rezultaty primeneniia metoda posledovatel'nogo prognozirovaniia veroiatnostei dlia diskriminantnogo analiza prostykh modelei plasta [The results of applying the method of successive prediction of probabilities for discriminant analysis of simple reservoir models]. *Izvestiia Tul'skogo gosudarstvennogo universiteta. Estestvennye nauki*, 2009, pp.205-210.

11. Denisov S.V., Ismagilov R.N., Sidelnikov K.A. Rezultaty primeneniia metoda posledovatel'nogo prognozirovaniia veroiatnostei dlia diskriminantnogo analiza slozhnykh modelei plasta [The results of applying the method of successive prediction of probabilities for discriminant analysis of complicated reservoir models]. *Izvestiia Tul'skogo gosudarstvennogo universiteta. Estestvennye nauki*, 2009, pp.211-216.

12. Mikhalevich I.M., Primina S.P. Primeniye matematicheskikh metodov pri analize geologicheskoi informatsii (s ispolzovaniem kompiuternykh tekhnologii) [The use of mathematical methods in the analysis of geological information (using computer technology)]. Irkutsk, Irkutskii gosudarstvennyi universitet, 2006, part III, 115 p.

13. Porotnikov A.V., Popov M.P., Gorbunova N.P. Primeniye lineinogo diskriminantnogo analiza dlia avtomaticheskogo opredeleniia proiskhozhdeniia izumruda po dannym rentgenofluorestsennogo analiza [The use of linear discriminant analysis to automatically determine the origin of emerald according to X-ray fluorescence analysis]. *Trudy instituta geologii i geokhimii im. Akademika A.N. Zavaritskogo*. Ekaterinburg, Institut geologii i geokhimii im. Akademika A.N. Zavaritskogo, 2013, pp.353-355.

14. Kravtsov B.A., Miliutin L.I. Vozmozhnosti primeneniia mnogomernoi klassifikatsii pri izuchenii populiatsii

drevesnykh rastenii [The possibility of using multidimensional classification in studying populations of woody plants]. *Prostranstvenno-vremennaiia struktura lesnykh biogeotsenozov*. Novosibirsk, Nauka, 1981, pp.47-65.

15. Urbakh V.Iu. Diskriminantnyi analiz i ego primeniye v biologicheskoi sistematike i meditsinskoi diagnostike [Discriminant analysis and its application in biological systematic and medical diagnostics]. *Primeniye matematicheskikh metodov v biologii*. Leningrad, 1964, iss.3, pp.67-87

16. Kagan E.S., Morozova I.S. Izuchenie faktorov optimizatsii poznavatelnoi deiatelnosti studentov s pomoshchiu metodov klasternogo i diskriminantnogo analizov [Studying the factors of optimization of students' cognitive activity using the methods of cluster and discriminant analyzes]. *Sibirskaiia psikhologiya segodnia*. Kemerovo: Kuzbassvuzizdat, 2002, pp.36-41.

17. Giniyatullin K.G., Valeeva A.A., Smirnova E.V. Ispolzovanie klasternogo i diskriminantnogo analizov dlia diagnostiki litologicheskoi neodnorodnosti pochvoobrazuiushchei porody po granulometricheskomu sostavu [Application of cluster and discriminant analyses to diagnose lithological heterogeneity of the parent material according to its particle-size distribution]. Moscow, Nauka, 2017, pp.946-953. DOI: 10.7868/S0032180X17080044

18. Dat Thanh Tran, Moncef Gabbouj, Alexandros Iosifidis. Multilinear class-specific discriminant analysis. *Pattern Recognition Letters*, 2017, vol.100, iss.1, pp.131-136. DOI: 10.1016/j.patrec.2017.10.027

19. Yong Wang, Jian-Bin Xie, Yi Wu. Orthogonal discriminant analysis revisited. *Pattern Recognition Letters*, 2016, vol.84, pp.149-155. DOI: 10.1016/j.patrec.2016.09.010

20. Andrewartha H.G., Birch L.C. The distribution and abundance of animals. Chicago, Univ. Press, 1954, 782 p.

21. Cacoullos T. Discriminant analysis and applications. London, Academic press, Inc. Ltd, 1972, 434 p.

22. Kim Dzh.O., Miuller Ch.U, Klekka U.R. et al. Faktorny, diskriminantnyi i klasternyi analiz

[Factor, discriminant and cluster analysis]. Moscow, Finansy i statistika, 1989, 215 p.

23. Gorban A.N. Obuchenie neironnykh setei [Neural Network Training]. Moscow, ParaGraph, 1990, 160 p.

24. Andersen T. Vvedenie v mnogomernyi statisticheskii analiz [Introduction to multivariate statistical analysis]. Moscow, Fizmatgiz, 1963, 500 p.

25. Morrison D.F. Multivariate statistical methods. 2nd ed. New York, McGraw-Hill, Inc., 1976, 415 p.

26. Morrison D.F. Applied statistical methods. Prentice-Hall, Inc, Engle-wood Cliffs, N.J., 1983, 562 p.

27. Anderson T.W. An introduction to multivariate statistical analysis. New York, John Wiley and Sons, Inc., 1958, 374 p.

28. Cooley W.W., Lohnes P.R. Multivariate data analysis. New York, John Wiley and Sons, Inc., 1971, 364 p.

29. Fisher R.A. The precision of discriminant functions. *Annals of Human Genetics*, 1940, vol.10, iss.1, pp.422-429. DOI: 10.1111/j.1469-1809.1940.tb02264.x

30. Borovikov V.P. Statistika dlia studentov i inzhenerov [Statistica for engineering students]. Moscow, KompiuterPress, 2001, 301 p.

31. Gmurman V.S. Teoriia veroiatnostei i matematicheskaia statistika [Theory of probability and mathematical statistics]. Moscow, Vysshiaia shkola, 1972, 368 p.

32. Aivazian S.A., Bezhaeva Z.I., Staroverov O.V. Klassifikatsiia mnogomernykh nabliudenii [Classification of multidimensional observations]. Moscow, Statistika, 1974, 240 p.

33. Machine learning, neural and statistical classification. Ed. D. Mitchie et al. Ellis Horwood, Chichester, 1994, 304 p.

34. Li C.C. Introduction to experimental statistics. New York, McGraw-Hill, Inc., 1964, 460 p.

35. Marriot F.H.C. The interpretation of multiple observations. London, Academic Press, Inc., Ltd., 1974, 117 p.

36. Efroimson M.A. Multiple regression analysis. *Mathematical Methods for Digital Computers*, 1960, vol.1, pp.191-203.

37. Aivazian S.A., Stepanov V.S. Instrumenty statisticheskogo analiza dannykh [Statistical Data Analysis Tools]. *Mir PK*, 1997, no.8.

38. Aivazian S.A., Bushtaber V.M., Eniukov I.S., Meshalkin L.D. Prikladnaia statistika. klassifikatsiia i snizhenie razmernostei [Application statistics. Classification and reduction of dimensions]. Moscow, Finansy i statistika, 1989, 607 p.

39. Bikel P., Doksam K. Matematicheskaia statistika [Mathematical statistics]. Moscow, Finansy i statistika, 1983, iss.1, 278 p.; iss.2, 254 p.

40. Mikhalevich I.M., Alferova M.A., Rozhkova N.Iu. Osnovy prikladnoi statistiki [Basics of applied statistics]. Irkutsk, RIO gosudarstvennogo instituta usovershenstvovaniia vrachei, 2008, vol.3, 92 p.

Просьба сослаться на эту статью в русскоязычных источниках следующим образом:

Расторгуев М.Н. Использование дискриминантного анализа для интерпретации данных газового каротажа на примере Павловского нефтяного месторождения // Вестник Пермского национального исследовательского политехнического университета. Геология. Нефтегазовое и горное дело. – 2019. – Т.19, №1. – С.39–55. DOI: 10.15593/2224-9923/2019.1.4

Please cite this article in English as:

Rastorguev M.N. Using discriminant analysis for the interpretation of gas logging data on the example of the Pavlov oil field. *Perm Journal of Petroleum and Mining Engineering*, 2019, vol.19, no.1, pp.39-55. DOI: 10.15593/2224-9923/2019.1.4