

УДК 622.276.344:004.9

ПРИМЕНЕНИЕ РЕГРЕССИОННОГО АНАЛИЗА ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ РАБОТЫ НЕФТЯНЫХ СКВАЖИН С ПАРАФИНИСТОЙ НЕФТЬЮ

Вирстюк Анастасия Юрьевна¹,
Nastua5.1991@mail.ru

Микшина Виктория Степановна¹,
mikshinavs@gmail.com

¹ Сургутский государственный университет,
Россия, 628412, г. Сургут, пр. Ленина, 1.

Актуальность исследования обусловлена необходимостью своевременного и быстрого анализа состояния нефтяного фонда скважин, работа которого осложнена парафинистостью. Рост числа показателей работы нефтяной скважины в связи с усложнением условий нефтедобычи приводит к трудностям их интерпретации специалистами геофизических служб и, как результат, – к неверному принятию решений об эффективности работы нефтяных скважин. Существующие модели анализа эффективности работы нефтяных скважин включают либо небольшое количество параметров работы скважины, что делает их недостаточно комплексными, либо избыточное количество показателей, что делает модели сложными и ресурсоемкими.

Цель: определить геолого-физические показатели, оказывающие существенное влияние на парафинистость нефти для создания регрессионной модели прогнозирования эффективности работы нефтяной скважины с точки зрения геофизики.

Объекты: нефтяные скважины и эффективность их работы с учетом парафинистости нефти.

Методы: регрессионный анализ, оценка адекватности модели по t -критерию Стьюдента, коэффициенту детерминации, F -критерию Фишера, проверка несмещенности и эффективности оценок с помощью условий Гаусса–Маркова, оценка автокорреляции остатков с помощью статистики Дарбина–Уотсона, непараметрическая статистика, анализ остатков регрессии.

Результаты. Определены форма и вид регрессионной зависимости между парафинистостью нефти, пластовой температурой и вязкостью нефти в пластовых условиях. Доказана адекватность полученной модели. Полученная математическая модель в совокупности с моделью анализа технологической эффективности работы нефтяных скважин может стать основой для систем поддержки принятия решений специалистов геофизических служб с целью комплексного анализа эффективности работы нефтяных скважин.

Ключевые слова:

Регрессионный анализ, эффективность работы нефтяной скважины, парафинистость, остатки регрессии, адекватность модели.

Введение

Повышение эффективности эксплуатации скважин является одной из важнейших задач нефтегазодобывающих предприятий. В настоящее время большинство нефтяных месторождений страны находятся на поздней и завершающей стадиях разработки [1]. Они характеризуются увеличением доли трудноизвлекаемых запасов. На последних стадиях разработки осуществляется отбор 10–20 % извлекаемых запасов месторождения, но из-за падения общей добычи нефти и усложнения условий её залегания эти стадии требуют особого внимания со стороны специалистов геофизических служб.

Разработка трудноизвлекаемых запасов нефти требует больших финансовых затрат и характеризуется усложнением процесса нефтедобычи вследствие того, что «существующие технологии не отвечают геологическим особенностям пласта» [2].

Трудноизвлекаемые нефти определяются большим числом геологофизических характеристик нефти. Ключевым фактором отнесения нефти к трудноизвлекаемым является высокое содержание парафинов [3].

Парафиновые отложения осложняют добычу и транспортировку нефти [4]. Поэтому своевременный анализ работы нефтяных скважин с парафинистой нефтью позволит специалистам геофизических служб быстро реагировать на изменения условий их эксплуатации и принимать верные решения относительно эффективности их работы.

Можно выделить две группы показателей нефтяных скважин, влияющих на эффективность их работы [5]:

- геолого-физические, характеризующие особенности строения залежи (сжимаемость, эффективная толщина пластов, пористость, проницаемость и т. п.);
- технологические показатели эксплуатации скважин (к примеру, время работы скважин), в том числе параметры работы нагнетательных скважин, оказывающих влияние на нефтяные скважины (радиус влияния, суммарный объем закачанной жидкости по интерферирующим скважинам и т. п.).

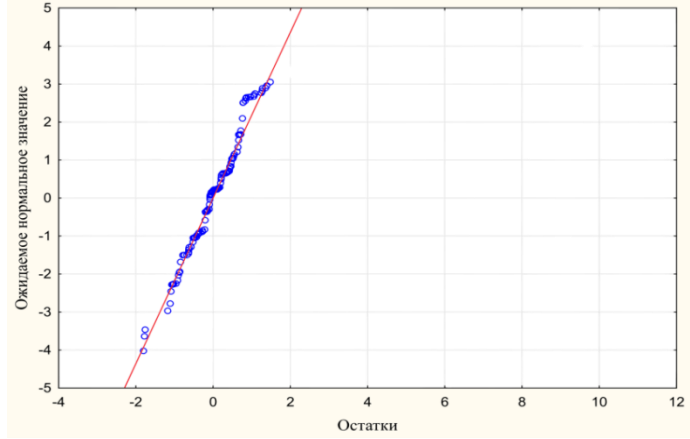
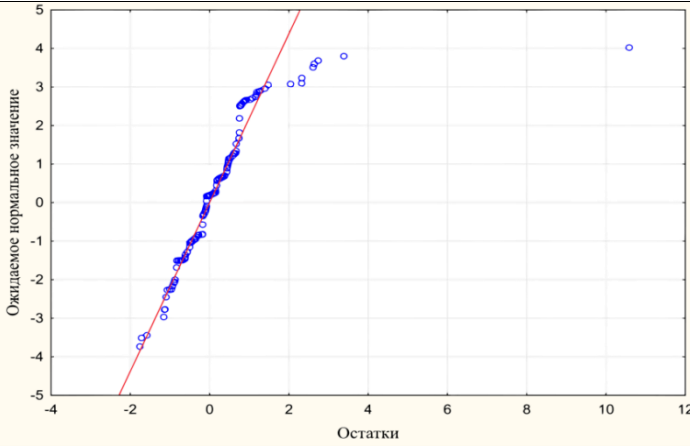
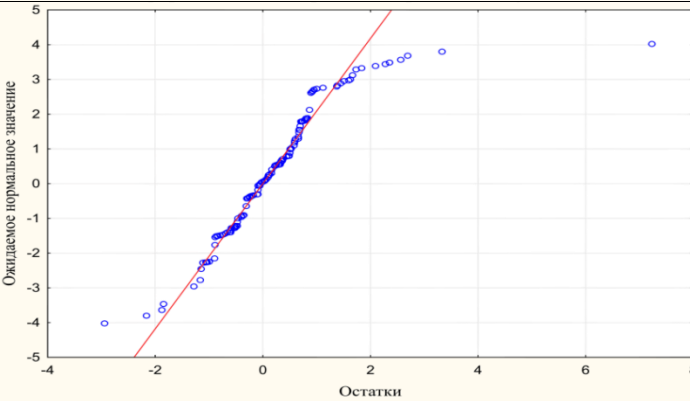
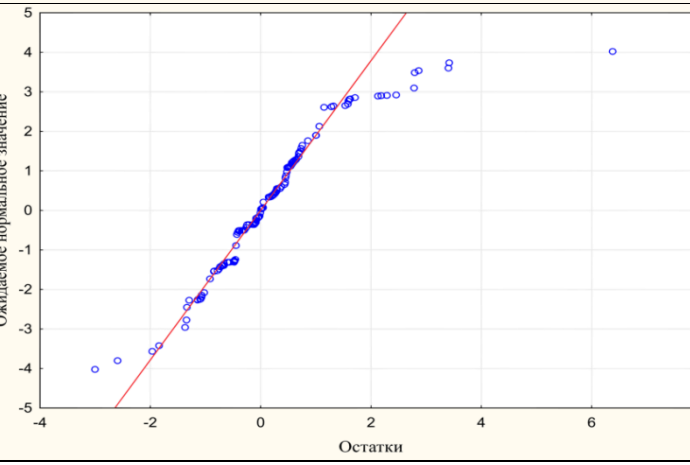
Статистический анализ работы 200 нефтяных скважин показал, что ключевым параметром геолого-физической группы является «содержание парафина в нефти» (G_8). Наибольшее влияние оказывают следующие показатели: пластовая температура (G_2), плотность (G_5) и вязкость нефти в пластовых условиях (G_6) [6, 7].

Подбор вида уравнения регрессии для построения модели эффективности работы нефтяных скважин

Для установления зависимости между показателями требуется составить уравнение регрессии. Независимых переменных в каждом случае несколько, поэтому следует рассматривать уравнение множественной регрессии. Различают линейные, нелинейные, сводящиеся к линейным, и внутренне нелинейные виды уравнений [8].

Таблица 1. Результаты применения линеаризующего преобразования к исходным данным

Table 1. Results of applying a linearization transform to the source data

Линеаризующее преобразование, значение R^2 Linearizing conversion, R^2 value	График остатков на нормальной вероятностной бумаге Plot of residuals on normal probabilistic paper
X^2 $R^2 = 0,89$	
X^3 $R^2 = 0,40$	
e^x $R^2 = 0,71$	
$1/X$ $R^2 = 0,57$	

Имеющиеся данные по каждому из этих показателей не соответствуют нормальному закону распределения, поэтому в качестве модели регрессионного уравнения следует выбрать нелинейную, но сводящуюся к линейной. Зависимости между показателями работы скважин, в том числе и нефтяных, характеризуются сложностью. Большая часть разработанных моделей носит непараметрический характер, что отражено в работах [9–13].

На первом этапе при построении нелинейной множественной модели регрессии требуется провести отбор значимых показателей с помощью пошаговой регрессии, а именно метода последовательного включения.

Суть метода последовательного включения заключается в том, что в модель последовательно добавляются независимые переменные, которые имеют наибольший коэффициент корреляции с зависимой переменной [14].

Для построения нелинейного, но сводящегося к линейному, уравнения регрессии сначала требуется провести линеаризующее преобразование.

Лучшее линеаризующее преобразование было выбрано на основании анализа графиков остатков на нормальной вероятностной бумаге, построенных с помощью модуля Fixed Nonlinear Regression программного пакета для статистического анализа *STATISTICA*. Обработка большого количества данных с помощью пакетов для статистического анализа подробно описана в работах [15–17]. Примененное линеаризующее преобразование, полученное значение коэффициента детерминации и соответствующий график остатков регрессии представлены в табл. 1.

Анализ распределения остатков регрессии показал, что наиболее близким к нормальному закону распределения является уравнение регрессии с линеаризующим преобразованием переменных X^2 , поскольку значения остатков лучше ложатся на прямую линию. Таким образом, между зависимым и независимыми показателями имеет место квадратичная зависимость, представленная в общем виде:

$$y = \sum_{i,j=1}^n a_{ij}x_i x_j,$$

где y – зависимая переменная; x_i, x_j – независимые переменные; a_{ij} – коэффициенты регрессии.

Линеаризующее преобразование позволило достичь наиболее высокого коэффициента детерминации. Чем выше коэффициент детерминации, тем лучше полученная модель согласуется с данными наблюдений. В табл. 2 представлены коэффициенты регрессии, стандартные ошибки коэффициентов уравнения регрессии, а также значения t -критерия Стьюдента, полученные при построении уравнения с линеаризующим преобразованием X^2 .

Далее необходимо провести проверку адекватности полученной модели.

Проверка адекватности регрессионной модели анализа эффективности работы нефтяных скважин

Проверка адекватности моделей начинается с проверки значимости каждого коэффициента регрессии

[18]. Следует отметить, что оценка нелинейных по параметрам уравнений осуществляется обычным методом наименьших квадратов, но примененным не к исходным, а к преобразованным данным.

Таблица 2. Результаты регрессионного анализа для зависимого показателя «Содержание парафина в нефти»

Table 2. Results of regression analysis for the dependent indicator «Oil paraffinicity»

Переменная Variable	Коэффициент уравнения регрессии b Regression equation coefficient b	Стандартная ошибка отклонения b Standard deflection error b	Критерий Стьюдента Student t-test
b_0	0,231	0,0054	43,118
G_2^2	–0,113	0,0084	–13,414
G_6^2	–0,113	0,0038	–29,405
G_2	0,229	0,0180	12,705
G_6	0,189	0,010	19,227
G_5^2	–0,011	0,0086	–1,263

Для проверки значимости коэффициентов регрессии применяется t -критерий Стьюдента:

$$t_{\text{эмп}} = \frac{|b_i|}{\sqrt{\sigma_{b_i}^2}}$$

где $\sigma_{b_i}^2$ – дисперсия коэффициента регрессии, определяемая по формуле:

$$\sigma_{b_i}^2 = \frac{\sigma_Y^2}{k},$$

где k – число факторов в уравнении; σ_Y^2 – дисперсия результативного фактора.

Параметр модели признается статистически значимым, если:

$$t_{\text{эмп}} > t_{\text{теор}}(\alpha, \mu = n - k - 1),$$

где μ – число степеней свободы; α – уровень значимости; n – число наблюдений.

Для построения регрессионной модели была проанализирована работа 200 скважин. Число степеней свободы μ для $t_{\text{теор}}$ равно 195.

Табличное значение коэффициента Стьюдента при $\alpha = 0,05$ и $\mu = 195$ равно 1,96.

Поскольку $t_{\text{эмп}}$ для квадрата плотности нефти меньше табличного значения, то данный показатель признается незначимым и не может войти в уравнение регрессии.

В табл. 2 представлены стандартные ошибки коэффициентов уравнения регрессии. Стандартная ошибка – это стандартное отклонение оценок, которые будут получены при многократной случайной выборке данного размера из одной и той же совокупности. Чем меньше стандартная ошибка, тем более достоверной является оценка. Из табл. 2 видно, что для всех показателей стандартная ошибка достаточно мала, что говорит о достоверности полученных оценок.

Соответственно, в регрессионное уравнение будут включены следующие показатели: пластовая температура, вязкость нефти в пластовых условиях, а также квадраты этих показателей.

Для проверки качества полученного уравнения регрессии можно воспользоваться коэффициентом детерминации.

Коэффициент детерминации R^2 равен 0,89, то есть близок к 1, соответственно полученное уравнение регрессии наилучшим образом описывает зависимость переменных.

Но данный коэффициент имеет недостаток: включение нового фактора в уравнение регрессии, даже незначительного, автоматически увеличивает значение R^2 [16]. Для исправления данной ситуации применяется скорректированный коэффициент детерминации R_{adj}^2 :

$$R_{adj}^2 = 1 - \frac{n - 1}{n - p - 1} (1 - R^2),$$

где n – число наблюдений; p – число факторов в уравнении регрессии.

Для полученной модели скорректированный коэффициент детерминации равен 0,88.

Для того чтобы оценить статистическую значимость уравнения регрессии и полученного коэффициента детерминации, необходимо воспользоваться F -критерием Фишера:

$$F = \frac{\rho^2}{1 - \rho^2} \frac{n - p - 1}{p} = 192,15,$$

где p – число независимых переменных в уравнении регрессии; ρ^2 – индекс корреляции; n – число наблюдений.

В свою очередь индекс корреляции определяется по формуле:

$$\rho = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}},$$

где \hat{y} – расчетные значения зависимой переменной по нелинейной регрессии.

Индекс корреляции между y и x совпадает с линейным коэффициентом корреляции между y и преобразованным x [19].

В силу того, что табличное значение F -критерия Фишера равно 4,36, что намного меньше рассчитанного значения, нулевую гипотезу о статистической незначимости уравнения регрессии можно отвергнуть.

Для получения наилучших оценок параметров линейной множественной регрессии необходимо проверить их на несмещенность и эффективность с помощью условий Гаусса–Маркова [8]:

1. Математическое ожидание случайного члена равно нулю в любом наблюдении.
2. Дисперсия случайного члена постоянна для всех наблюдений.
3. Значения случайного члена в любых наблюдениях не коррелируют между собой.
4. Случайный член должен быть распределен независимо от объясняющих переменных.

Условия будут проверены на основании анализа остатков регрессии.

Для подтверждения или опровержения 1-го и 4-го условий Гаусса–Маркова необходимо проанализировать графики зависимости остатков от каждой независимой переменной (в данном случае от пластовой

температуры и вязкости нефти в пластовых условиях), что представлено на рис. 1, 2.

На рис. 1, 2 видно, что математическое ожидание остатков равно нулю и остатки независимы относительно независимых показателей, поскольку коэффициент корреляции в первом и во втором случаях равен 0 (строка *correlation* в заголовках графиков).

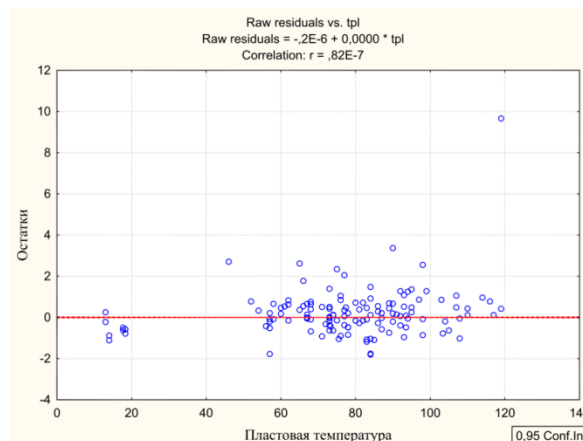


Рис. 1. График зависимости остатков от пластовой температуры

Fig. 1. Plot of residuals dependence on reservoir temperature

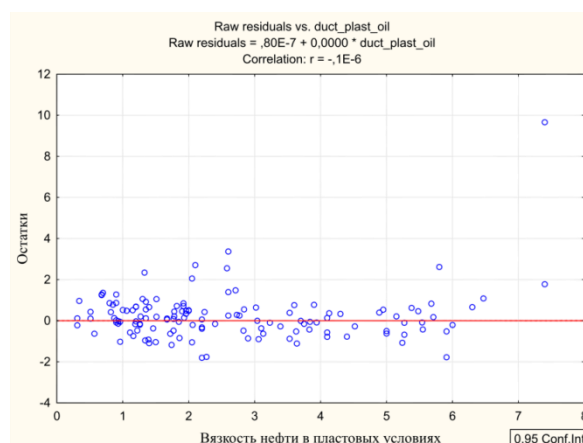


Рис. 2. График зависимости остатков от вязкости нефти

Fig. 2. Plot of residuals dependence on oil viscosity

Для выявления нестабильности ошибки уравнения регрессии, то есть для проверки 2-го условия Гаусса–Маркова, необходимо проанализировать зависимость остатков от предсказанных значений зависимого показателя (в данном случае парафинистости нефти). Зависимость представлена на рис. 3.

Видно, что линия дисперсий остатков параллельна оси X . Дисперсия случайного возмущения не изменяется, так как с увеличением значений зависимой переменной линия дисперсии остается на том же уровне.

Для проверки 3-го условия Гаусса–Маркова, а именно анализа автокорреляции остатков, необходимо рассчитать статистику Дарбина–Уотсона.

Проверка автокорреляции остатков необходима для того, чтобы исключить однонаправленное воздействие на объясняемую переменную неучтенных в модели факторов.

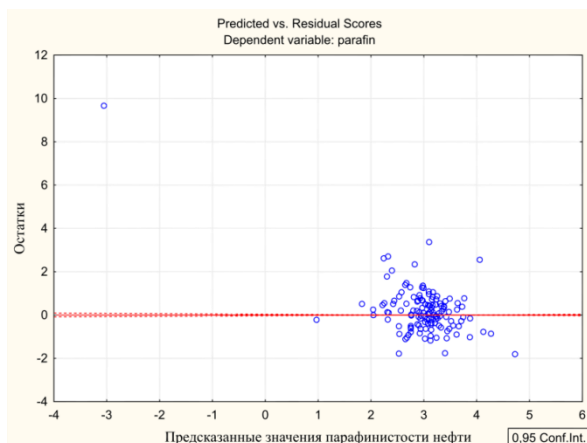


Рис. 3. График зависимости предсказанных значений зависимого показателя от остатков

Fig. 3. Plot of dependence of the dependent indicator predicted values on residuals

Статистика Дарбина–Уотсона характеризует наличие или отсутствие сериальной корреляции [20, 21].

Значение данной статистики находится в промежутке от 0 до 4. В случае отсутствия автокорреляции статистика Дарбина–Уотсона близка к 2. Близость к 0 говорит о положительной автокорреляции, к 4 – об отрицательной.

Значение статистики Дарбина–Уотсона, рассчитанное в пакете *STATISTICA*, равно 1,759; значение сериальной корреляции – 0,121.

Сериальная корреляция минимальна, значение статистики Дарбина–Уотсона стремится к 2, поэтому автокорреляция остатков отсутствует.

Для дополнительного анализа остатков регрессии необходимо проверить наличие выбросов. Анализ выбросов был проведен на основании показателя Кука. Наблюдение может исказить оценки коэффициентов регрессии, если обладает большим потенциалом влияния. Большое значение показателя Кука указывает на сильно влияющие наблюдения. Для рассматриваемых статистических данных среднее значение показателя Кука незначительно и составляет 0,000197, что подтверждает отсутствие выбросов.

На основании проведенного регрессионного анализа можно сделать вывод, что уравнение регрессии, описывающее зависимость «содержания парафина в нефти» от «пластовой температуры» G_2 и «вязкости нефти в пластовых условиях» G_6 имеет вид:

$$G_8 = 0,231 + 0,189 * G_6 + 0,229 * G_2 - 0,113 * (G_2^2 + G_6^2).$$

Полученная зависимость позволит своевременно оценивать и прогнозировать изменения парафинистости

сти нефти в нефтяной скважине с целью её корректировки для упрощения добычи и транспортировки и, как следствие, повышения эффективности работы скважин.

Заключение

Мировые запасы трудноизвлекаемой нефти с особыми условиями (вязкая, тяжелая, парафинистая и т. п.) значительно превышают запасы легких нефтей. В частности, парафинистая нефть встречается в 84 % случаев. Поэтому своевременный анализ парафинистости нефти позволит повысить эффективность работы нефтяных скважин. Этим и объясняется необходимость разработки математической модели, связывающей данный показатель с наиболее значимыми геолого-физическими характеристиками.

В ходе анализа работы 200 нефтяных скважин с парафинистой нефтью была разработана соответствующая математическая модель.

Адекватность полученной математической модели была подтверждена следующим образом:

1. Оценки параметров множественной регрессии, полученные методом наименьших квадратов, являются согласованными, несмещенными и эффективными по Гауссу–Маркову.
2. Коэффициенты регрессии являются значимыми по t -критерию Стьюдента.
3. Стандартные ошибки коэффициентов уравнения регрессии минимальны. Изменяются в пределах от 0,05 до 0,1.
4. Значения коэффициента детерминации и скорректированного коэффициента детерминации достаточно высоки: $R^2=0,89$; $R_{adj}^2=0,88$.

Оценка статистической значимости уравнения регрессии подтверждена по F -критерию Фишера. Значение $F=192,15$.

Автокорреляция остатков отсутствует. Значение статистики Дарбина–Уотсона равно 1,76.

Выбросы на основании показателя Кука не выявлены. Среднее значение показателя Кука составляет 0,000197.

Предложенная модель анализа эффективности работы нефтяных скважин с парафинистой нефтью может стать основой для системы поддержки принятия решений специалистов геофизических служб при анализе работы нефтяных скважин.

Статья написана при финансовой поддержке гранта «Математическое моделирование процессов нефтепереработки и нефтехимии на основе динамических моделей в терминах смесей непрерывного состава», проект № 18-47-860003. Руководитель проекта – В.С. Микшина.

СПИСОК ЛИТЕРАТУРЫ

1. Уметбаев В.В. Повышение эффективности эксплуатации скважин с применением растворителей АСПО: дис. ...канд. тех. наук. – Уфа, 2003. – 151 с.
2. Шарф И.В., Борзенкова Д.Н. Трудноизвлекаемые запасы нефти: понятие, классификационные подходы и стимулирование разработки // Фундаментальные исследования. – 2015. – № 2-16. – С. 3593–3597.
3. Яценко И.Г. О роли трудноизвлекаемых нефтей как источника углеводородов в будущем на основе информационно-вычислительной системы по нефтехимической геологии Музея нефтей // Культурное наследие и информационные технологии на постсоветском пространстве: Материалы международной научно-практической конференции / под ред. И.Б. Лаптева. – Минск, 10–14 мая 2011. – Минск: Белпринт, 2011. – С. 39–41.
4. Ильин А.Н., Полищук Ю.М., Яценко И.Г. Высокопарафинистые нефти: закономерности пространственных и временных

- изменений их свойств // Электронный научный журнал «Нефтегазовое дело». – 2007. – №1. URL: http://ogbus.ru/files/ogbus/authors/Iiin/Iiin_1.pdf (дата обращения: 27.08.2018).
5. Вирстюк А.Ю., Микшина В.С., Назина Н.Б. Оценка эффективности технологического процесса в нефтяной промышленности // Инновационные, информационные и коммуникационные технологии: сборник трудов XV Международной научно-практической конференции / под ред. С.У. Увайсова. – М.: Ассоциация выпускников и сотрудников ВВИА им. проф. Жуковского, 2019. – С. 288–295.
 6. Вирстюк А.Ю., Микшина В.С. Математические модели анализа эффективности работы добывающих скважин // Информационная безопасность регионов России (ИБРР-2017). Юбилейная X Санкт-Петербургская межрегиональная конференция. – СПб., 1–3 ноября 2017. – С. 376–380.
 7. Вирстюк А.Ю., Микшина В.С. Применение факторного анализа для редукции признакового пространства нагнетательных скважин // Вестник кибернетики. – 2018. – № 2 (30). – С. 172–178.
 8. Шанченко Н.И. Лекции по эконометрике. – Ульяновск: Ульяновский государственный технический университет, 2008. – 139 с.
 9. Watheq J. Al-Mudhafar, Ahmed Khalil Al-Khazraji. Non-parametric adaptive regression splines for multisource permeability modeling in a sandstone oil reservoir // Offshore Technology Conference. – Kuala Lumpur, Malaysia, 2016. – P. 2906–2918.
 10. Watheq J. Al-Mudhafar, Ahmed Khalil Al-Khazraji. Efficient reservoir modeling-statistical approach for evaluation and development of future waterdrive undersaturated oil reservoir performance // SPE Latin American and Caribbean Petroleum Engineering Conference 2014. – Maracaibo, Venezuela, 2014. – V. 3. – P. 1859–1872.
 11. Kettaneh N., Berglund A. PCA and PLS with very large data sets // Computational Statistics & Data Analysis. – 2005. – V. 48. – P. 69–85.
 12. Basilevsky A. Statistical factor analysis and related methods: theory and applications. – New York: John Wiley & Sons, Inc., 2008. – 737 p.
 13. Lolon E., Hamidieh K., Weijers L. Evaluating the relationship between well parameters and production using multivariate statistical models: a Middle Bakken and Three Forks case history // SPE Hydraulic Fracturing Technology Conference. – The Woodlands, Texas, 2016. – P. 303–331. DOI:10.2118/179171-MS.
 14. Воскобойников Ю.Е. Эконометрика в Excel (парный и множественный регрессионный анализ). – Новосибирск: Новосибирский государственный архитектурно-строительный университет, 2005. – 77 с.
 15. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. Data mining, inference, and prediction. Second ed. – Switzerland: Springer, 2017. – 764 p.
 16. Heumann C., Schomaker M. Introduction to statistics and data analysis with exercises, solutions and applications in R. – Switzerland: Springer, 2013. – 426 p.
 17. Joaquim P. Marques de Sá. Applied statistics using SPSS, STATISTICA, MATLAB and R. – Switzerland: Springer Publishing Company, 2007. – 505 p.
 18. Минина И.Д. Статистика. Ч. 1. Теория статистики. – Пенза: РИО ПГСХА, 2013. – 225 с.
 19. Бушин П.Я. Эконометрика. Практикум по решению и анализу задач. – Хабаровск: Хабаровский государственный университет экономики и права, 2005. – 88 с.
 20. Кабанов С.В. Использование пакета Statistica 5.0 для статистической обработки опытных данных. – Саратов: Саратовский государственный аграрный университет, 2001. – 200 с.
 21. Халафян А.А. STATISTICA 6. Статистический анализ данных. – М.: ООО «Бином – Пресс», 2007. – 512 с.

Поступила 02.12.2019 г.

Информация об авторах

Вирстюк А.Ю., аспирант кафедры информатики и вычислительной техники Сургутского государственного университета.

Микшина В.С., кандидат технических наук, профессор кафедры информатики и вычислительной техники Сургутского государственного университета.

UDC 622.276.344:004.9

APPLICATION OF REGRESSION ANALYSIS TO EVALUATE THE EFFICIENCY OF OIL WELL OPERATING WITH THE PARAFFIN OIL

Anastasiya Yu. Virstyuk¹,
Nastua5.1991@mail.ru

Victoria S. Mikshina¹,
mikshinavs@gmail.com

¹ Surgut State University,
1, Lenin avenue, Surgut, 628412, Russia.

The relevance of the research is caused by the need for timely and rapid analysis of the state of oil wells which work is complicated by paraffinicity. The increase of the number of oil well indicators due to complicated production conditions leads to the difficulties in their interpretation by the geophysical services specialists, and as a result to incorrect decisions about the operating efficiency of oil wells. The existing models for evaluation of the oil wells operating efficiency include either a small number of well operation indicators which makes them insufficiently complex, or an excessive number of indicators which makes the models difficult and resource-intensive.

The main aim of the research is to define the geological and physical indicators that have a significant impact on the oil paraffinicity to create a regression model for predicting oil well operating efficiency from a geological point of view.

Objects of researches are oil wells and oil well operating efficiency with regard to oil paraffinicity.

Methods: regression analysis, evaluation of the model adequacy using Student's *t*-criterion, determination coefficient, Fisher's *F*-criterion, checking unbiasedness and efficiency of estimates using Gauss–Markov conditions, estimating residual autocorrelation using Durbin–Watson statistics, non-parametric statistics, regression residual analysis.

Results. The authors have determined the form and type of regression dependence between oil paraffinicity, reservoir temperature and oil viscosity in reservoir conditions. The adequacy of the obtained model is proved. The resulting mathematical model in conjunction with the model of the technological oil well efficiency can be the basis for decision support systems of specialists of geophysical services for complex analysis of the oil well operating efficiency.

Key words:

Regression analysis, oil well operating efficiency, paraffinicity, regression residual, model adequacy.

The article was written with the financial support of the grant «Mathematical modeling of refining and petrochemical processes based on dynamic models in terms of mixtures of continuous composition», project no. 18-47-860003. Project Manager – V.S. Mikshina.

REFERENCES

1. Umetbaev V.V. *Povyshenie effektivnosti ekspluatatsii skvazhin s primeneniem rastvoriteley ASPO*. Dis. Kand. nauk [Improving the efficiency of well operation with the use of paraffin solvents. Cand. Diss.]. Ufa, 2003. 151 p.
2. Sharf I.V., Borzenkova D.N. *Trudnoizvlekaemye zapasy neftei: ponyatie, klassifikatsionnye podkhody i stimulirovanie razrabotki* [Hard-to-recover oil reserves: concept, classification approaches and development stimulation]. *Fundamentalnye issledovaniya*, 2016, no. 2-16, pp. 3593–3597.
3. Yashchenko I.G. O roli trudnoizvlekaemykh neftey kak istochnike uglevodorodov v budushchem na osnove informatsionno-vychislitel'noy sistemy po neftekhimicheskoy geologii Muzeya neftey [On the role of hard-to-recover oils as a source of hydrocarbons in future based on the information system of petrochemical Geology Museum of Oil]. *Materialy mezhdunarodnoy nauchno-prakticheskoy konferentsii. Kulturnoe nasledie i informatsionnye tekhnologii na postsovet'skom prostranstve* [Materials of the international scientific and practical conference]. Minsk, 10–14 May 2011. Minsk, Belprint Publ., 2011. pp. 39–41.
4. Il'in A.N., Polishchuk Yu.M., Yashchenko I.G. Vysokoparafinistye nefti: zakonomernosti prostranstvennykh i vremennykh izmeneniy ikh svoystv [Highly paraffinic oils: regularities of spatial and temporal changes in their properties]. *Neftegazovoe delo*, 2007, no. 1. Available at: http://ogbus.ru/files/ogbus/authors/Iliin/Iliin_1.pdf (accessed 27 August 2018).
5. Virstyuk A.Yu., Mikshina V.S., Nazina N.B. Otsenka effektivnosti tekhnologicheskogo protsessa v neftyanoy promyshlennosti [Evaluation of oil well operating efficiency]. *Innovatsionnye, informatsionnye i kommunikatsionnye tekhnologii. Sbornik trudov XV Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Innovative, Information and Communication Technologies: Proceedings of the XV International Scientific and Practical Conference]. Ed. by S.U. Uvaysov. Moscow, Assotsiatsiya vypusknikov i sotrudnikov VVIA im. prof. Zhukovskogo Publ., 2019. pp. 288–295.
6. Virstyuk A.Yu., Mikshina V.S. Matematicheskie modeli analiza effektivnosti raboty dobyvayushchikh skvazhin [Mathematical models for analyzing the performance of production wells]. *Informatsionnaya bezopasnost' regionov Rossii (IBRR-2017). Yubileynaya X Sankt-Peterburgskaya mezhdunarodnaya konferentsiya* [Information Security of Russian Regions (ISRR-2017). Anniversary X St. Petersburg Interregional Conference]. St-Petersburg, 1–3 November 2017. pp. 376–380.
7. Virstyuk A.Yu., Mikshina V.S. Application of factor analysis to reduce the feature space of injection wells. *Vestnik kibernetiki*, 2018, no. 2 (30), pp. 172–178. In Rus.
8. Shanchenko N.I. *Lektsii po ekonometrike* [Lectures on econometrics]. Ulyanovsk, Ulyanovsk State Technical University Publ., 2008. 139 p.
9. Watheq J. Al-Mudhafar, Ahmed Khalil Al-Khazraji. Non-parametric adaptive regression splines for multisource permeability modeling in a sandstone oil reservoir. *Offshore Technology Conference*. Kuala Lumpur, Malaysia, 2016. pp. 2906–2918.
10. Watheq J. Al-Mudhafar, Ahmed Khalil Al-Khazraji. Efficient reservoir modeling-statistical approach for evaluation and development of future waterdrive undersaturated oil reservoir performance. *SPE Latin American and Caribbean Petroleum Engineering Conference*. Maracaibo, Venezuela, 2014. Vol. 3, pp.1859–1872.
11. Kettaneh N., Berglund A. PCA and PLS with very large data sets. *Computational Statistics & Data Analysis*, 2005, vol. 48, pp. 69–85.

12. Basilevsky A. *Statistical factor analysis and related methods: theory and applications*. New York, John Wiley & Sons, Inc, 2008. 737 p.
13. Lolon E., Hamidieh K., Weijers L. Evaluating the relationship between well parameters and production using multivariate statistical models: a Middle Bakken and Three Forks case history. *SPE Hydraulic Fracturing Technology Conference*. The Woodlands, Texas, 2016. pp. 303–331. DOI: 10.2118/179171-MS.
14. Voskoboynikov Yu.E. *Ekonometrika v Excel (parny i mnozhestvennyy regressionnyy analiz)* [Excel econometrics (paired and multiple regression analysis)]. Novosibirsk, Novosibirsk State Architecture Building University, 2005. 77 p.
15. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning. Data mining, inference, and prediction*. Second ed. Switzerland, Springer, 2017. 764 p.
16. Heumann C., Schomaker M. *Introduction to statistics and data analysis with exercises, solutions and applications in R*. Switzerland, Springer, 2013. 426 p.
17. Joaquim P. Marques de Sá. *Applied statistics using SPSS, STATISTICA, MATLAB and R*. Switzerland, Springer Publ. Company, 2007. 505 p.
18. Minina I.D. *Statistika. Ch. 1. Teoriya statistiki* [Statistics. P. 1. The theory of statistics]. Penza, RIO PGSHA Publ., 2013. 225 p.
19. Bushin P.Ya. *Ekonometrika. Praktikum po resheniyu i analizu zadach* [Econometrics. Workshop on solving and analyzing problems]. Khabarovsk, Khabarovsk State University of Economics and Law Publ., 2005. 88 p.
20. Kabanov S.V. *Ispolzovanie paketa Statistica 5.0 dlya statisticheskoy obrabotki opytных dannыkh* [Using Statistica 5.0 package for statistical processing of experimental data]. Saratov, Saratov State Agrarian University Publ., 2001. 200 p.
21. Halafyan A.A. *STATISTICA 6. Statisticheskyy analiz dannыkh* [STATISTICA 6. Statistical data analysis]. Moscow, Binom-Press, 2007. 512 p.

Received: 2 December 2019.

Information about the authors

Anastasiya Yu. Virstyuk, postgraduate, Surgut State University.

Victoria S. Mikshina, Cand. Sc., professor, Surgut State University.