

Моделирование технологических режимов и прогноз показателей газоконденсатных скважин на основе методов машинного обучения и кластерного анализа

Е. С. Рублева

ООО «Газпром ВНИИГАЗ»

В работе представлен подход к моделированию гидродинамических показателей работы газоконденсатных скважин на основе гибридного ансамблирования методов машинного обучения и физически обоснованной кластеризации. Для исключения эффекта утечки данных из пространства входных признаков извлечены прямо вычисляемые гидродинамические параметры (число Рейнольдса, скорость потока, коэффициент Фаннинга). Проведена робастная фильтрация аномалий и кластеризация промысловых данных методом K-Means, подтвержденная физико-гидродинамической интерпретацией и критерием Краскела—Уоллиса. Сформирован гибридный ансамбль градиентного бустинга с наложением монотонных физических ограничений. Продемонстрировано статистически значимое по критерию Уилкоксона преимущество предложенного алгоритма над единичными моделями машинного обучения: относительная ошибка прогностической модели дебита снижена с 23.16% до 19.70% при коэффициенте детерминации равном 0.898. Для инверсной задачи восстановления забойного давления достигнута средняя абсолютная ошибка 0.517 атм.

Газоконденсатная скважина; дебит газа; забойное давление; машинное обучение; K-Means; PCA; градиентный бустинг; физически информированные модели; критерий Уилкоксона.

ВВЕДЕНИЕ

Повышение точности прогнозирования технологических параметров работы газоконденсатных скважин является критически важной задачей при управлении разработкой месторождений. Традиционные гидродинамические модели, основанные на физических уравнениях притока и многофазной фильтрации, требуют высокой степени детализации пластовых систем и трудоемкой адаптации. С другой стороны, прямое применение алгоритмов машинного обучения часто сталкивается с проблемой несбалансированности режимов эксплуатации скважин и риском некорректной интерпретации результатов из-за эффекта «утечки данных», когда в число входных признаков попадают производные величины, математически сопряженные с искомым целевым показателем.

Целью данного исследования является разработка и статистическая верификация гибридного алгоритма прогнозирования дебита и инверсного восстановления забойного давления газоконденсатных скважин, сочетающего робастную кластеризацию технологических режимов, физическую регуляризацию и ансамблирование моделей градиентного бустинга.

МЕТОДЫ ИССЛЕДОВАНИЯ

Подготовка данных и устранение эффекта утечки. Исходный массив промысловых измерений включал 60 443 записи. На первом этапе выполнен анализ физической независимости признаков пространства. Внедрение методов машинного обучения в задачи нефтегазового

Рублева Е. С. Моделирование технологических режимов и прогноз показателей газоконденсатных скважин на основе методов машинного обучения и кластерного анализа // СИИТ. 2026. Т. 8, № 4(28). С. 145-150. DOI: [10.54708/SIIT-2026-no4-p145](https://doi.org/10.54708/SIIT-2026-no4-p145). EDN: UWKXHM.

Rubleva E. S. "Modeling of Operating Modes and Performance Forecasting for Gas Condensate Wells Based on Machine Learning Methods and Cluster Analysis" // SIIT. 2026. Vol. 8, no. 4(28), pp. 145-150. DOI: [10.54708/SIIT-2026-no4-p145](https://doi.org/10.54708/SIIT-2026-no4-p145). EDN: UWKXHM. (In Russian).

комплекса требует строгого контроля корректности признакового пространства для предотвращения завышения точности [Дми21].

Такие параметры, как скорость газожидкостного потока V , число Рейнольдса Re и коэффициент гидравлического сопротивления Фаннинга K_{fan} , вычисляются непосредственно через искомый объемный дебит Q :

$$V = \frac{4Q}{\pi d^2}, \quad Re = \frac{\rho V d}{\mu}. \quad (1)$$

Включение указанных параметров в вектор входных признаков приводит к искусственному завышению метрик точности ($R^2 \rightarrow 0.999$) за счет прямого восстановления алгебраической зависимости [Goo16]. Для обеспечения физической корректности модели из вектора признаков были исключены расчетные параметры V , Re и K_{fan} . В финальный набор вошли 13 измеряемых промысловых и PVT-показателей [Ани3]: устьевое давление ($P_{гол}$), давление на штуцере ($P_{дикт}$), затрубное давление ($P_{затр}$), среднее пластовое давление ($P_{сред}$), забойное давление ($P_{заб}$), депрессия (ΔP), устьевая температура ($T_{гол}$), сверхсжимаемость (Z), объемный коэффициент (BG), плотность флюида, динамическая вязкость (μ), глубина скважины по стволу (MD) и диаметр НКТ (d).

После удаления пропущенных значений и робастной фильтрации аномалий с помощью алгоритма IsolationForest (доля загрязнения 3%) итоговая обучающая выборка составила 58 629 записей.

Физико-математический алгоритм кластеризации. Для учета нелинейности процессов фильтрации в различных технологических режимах выполнено робастное масштабирование признаков с последующим применением алгоритма K -Means [Has09]. Для обоснования оптимального числа кластеров K рассчитаны метрики силуэта, Дэвиса–Болдина и Калински–Харабаша в диапазоне $K \in [2, 6]$.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

Обоснование структуры кластеризации

В табл. 1 приведены результаты расчета критериев качества кластеризации.

Таблица 1

Статистическое обоснование выбора числа кластеров

№ п/п	K (число кластеров)	Индекс силуэта	Индекс Дэвиса–Болдина	Индекс Калински–Харабаша
1	2	0.475	0.955	40 558.8
2	3	0.457	0.868	38 841.5
3	4	0.444	0.866	39 368.8
4	5	0.336	0.961	40 465.2
5	6	0.349	0.966	39 549.4

На основе компромисса между геометрической метрикой Дэвиса–Болдина (минимум 0.866 достигается при $K = 4$) и физической интерпретируемостью режимов в качестве оптимальной выбрана 4-кластерная структура ($K = 4$). Метод главных компонент (РСА) показал, что первые две компоненты описывают 80.2% общей дисперсии (РСА1 — 56.1%, РСА2 — 24.1%), формируя четко разделенные области режимов работы (рис. 1).

Статистическая верификация разделимости режимов

Для проверки гипотезы о том, что выделенные кластеры соответствуют статически различным гидродинамическим состояниям скважин, проведен непараметрический H -тест Краскела–Уоллиса [Kru52] (табл. 2).

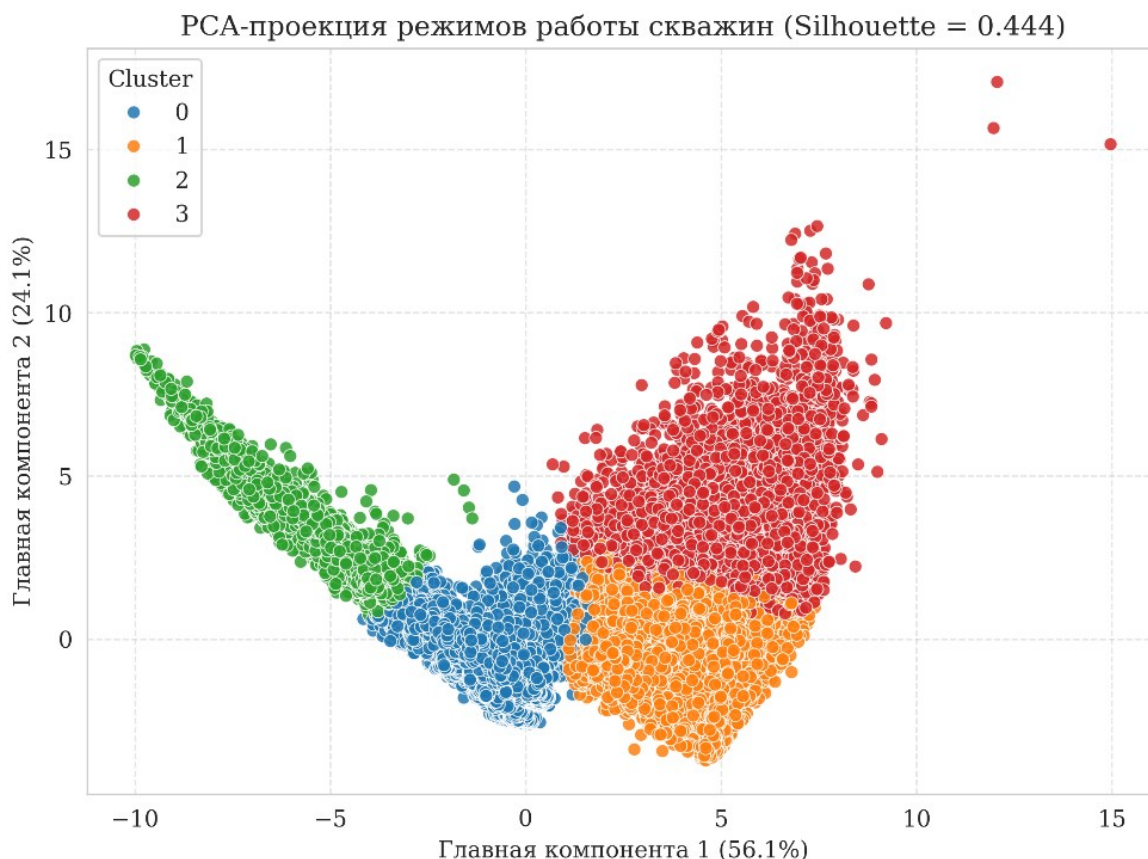


Рис. 1. Двумерная проекция выделенных технологических режимов работы газоконденсатных скважин в пространстве главных компонент ($K = 4$, индекс силуэта = 0.444)

Таблица 2

Результаты H -теста Краскела–Уоллиса ($K = 4$)

№ п/п	Параметр	H -статистика	p -value	Статистическая значимость ($p < 0.05$)
1	Депрессия (ΔP)	31 829.61	< 0.001	да
2	Забойное давление ($P_{\text{заб}}$)	33 355.95	< 0.001	да
3	Дебит переведенный (Q)	26 063.42	< 0.001	да
4	Динамическая вязкость (μ)	29 718.84	< 0.001	да
5	Сверхсжимаемость (Z)	35 807.77	< 0.001	да
6	Среднее давление ($P_{\text{сред}}$)	29 718.64	< 0.001	да
7	Устьевая температура ($T_{\text{гол}}$)	31 420.92	< 0.001	да

Все проанализированные параметры демонстрируют статистическую значимость ($p < 0.001$), что подтверждает объективность выделения режимов. Усредненные физико-гидродинамические характеристики кластеров приведены в табл. 3.

Таблица 3

Средние физико-гидродинамические характеристики кластеров

№ п/п	Кластер	ΔP , атм	$P_{\text{заб}}$, атм	Дебит Q , усл. ед.	μ , Па · с	Z	$P_{\text{сред}}$, Па	$T_{\text{гол}}$, К
1	0	4.30	85.59	0.1001	0.0140	1.0015	7.84×10^6	285.58
2	1	36.37	165.99	0.0229	0.0149	0.9402	1.39×10^7	301.98
3	2	4.03	22.91	0.1239	0.0120	1.0350	2.07×10^6	281.18
4	3	115.10	141.50	0.0265	0.0146	0.9649	1.16×10^7	295.04

По полученным кластерам можно сделать следующие выводы:

Кластер 0 характеризует средний стационарный режим при умеренном забойном давлении.

Кластер 1 представляет высоконапорный режим с высоким пластовым давлением (3.9 МПа) и низким относительным дебитом.

Кластер 2 показывает высокодебитный режим при истощенном забойном давлении (22.91 МПа).

Кластер 3 – режим глубоких депрессий ($\Delta P \approx 115.1$ МПа), характерный для скважин с повышенными фильтрационными сопротивлениями призабойной зоны.

Сравнительный анализ точности моделей

На основе выделенных кластеров сформирован гибридный ансамбль. В задачах прогнозирования параметров скважин широко применяются алгоритмы ансамблирования и нейронные сети [Ха626, Евс20]. Для прямого прогноза дебита введено монотонное ограничение:

$$\frac{\partial Q}{\partial (\Delta P)} \geq 0. \quad (2)$$

Оценка точности проводилась на отложенной тестовой выборке (20% от общего объема) [Пис24, Wil45]. В качестве базовой модели выступал аналогичный градиентный бустинг, обученный на всем массиве данных без кластеризации (табл. 4).

Таблица 4

Сравнительный анализ точности прогнозирования

№ п/п	Модель / Задача	RMSE	MAE	MAPE, %	R2
1	Базовая модель (единый ML без утечек)	0.021	0.014	23.159	0.884
2	Предложенный алгоритм (прямой прогноз дебита Q)	0.019	0.013	19.704	0.898
3	Предложенный алгоритм (инверсное $P_{заб}$)	1.168	0.517	0.581	0.999

Для доказательства статистической значимости превосходства гибридного алгоритма над базовым применен парный критерий Уилкоксона для модулей ошибок:

$$W = 41\,462\,225.50, \quad p\text{-value} = 1.6137 \times 10^{-83} \ll 0.05.$$

Полученное значение $p < 0.05$ подтверждает гипотезу о статистической неслучайности снижения ошибки при использовании предложенного алгоритма.

Сопоставление фактических и прогнозируемых значений на тестовой выборке для прямого и инверсного алгоритмов приведено на рис. 2.

ЗАКЛЮЧЕНИЕ

1. Исключение производных расчетных величин (Re , V , K_{fan}) из числа входных признаков позволило ликвидировать эффект утечки данных и сформировать объективную базовую модель с реальной точностью $R^2 = 0.884$ (MAPE = 23.16%).

2. Робастная кластеризация методом K -Means ($K = 4$) обеспечила разбиение массива промысловых данных на четыре статистически различимых по критерию Краскела–Уоллиса ($p < 0.001$) технологических режима работы скважин.

3. Построение гибридного ансамбля с учетом монотонных физических ограничений позволило снизить среднюю относительную ошибку прогноза дебита до 19.70% (улучшение на 3.45% в абсолютном выражении по сравнению с базовым ML). Статистическое преимущество подтверждено критерием Уилкоксона ($p = 1.61 \times 10^{-83}$).

Разработанный инверсный модуль обеспечивает высокоточное восстановление забойного давления $P_{\text{заб}}$ со средней абсолютной ошибкой 0.517 МПа ($R^2 = 0.999$), что позволяет использовать модель в системах оперативного мониторинга разработки газоконденсатных месторождений.

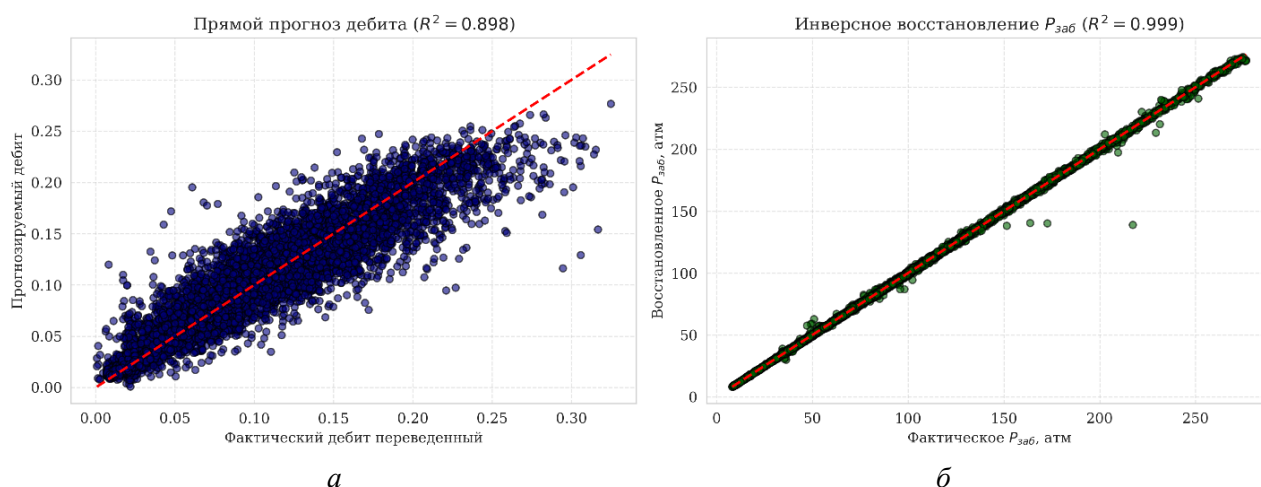


Рис. 2. Сопоставление фактических и рассчитанных гибридным алгоритмом показателей на тестовой выборке: *а* — прямой прогноз дебита Q ($R^2 = 0.898$); *б* — инверсное восстановление забойного давления $P_{\text{заб}}$ ($R^2 = 0.999$)

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

- [Дми21] Дмитриевский А. Н. Анализ рисков при использовании технологий искусственного интеллекта в нефтегазодобывающем комплексе // Автоматизация, телемеханизация и связь в нефтяной промышленности. 2021. № 7(576). С. 17–27. [QPMNQE](#).
- [Goo16] Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016. 800 p.
- [Ани24] Аникеев Д. П., Аникеева Э. С. Применение нейросетей для определения параметров PVT-моделей, используемых при решении задач гидродинамического моделирования // Актуальные проблемы нефти и газа. 2024. № 3. [FHHNCR](#).
- [Has09] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009. 745 p. [10.1007/978-0-387-84858-7](#).
- [Kru52] Kruskal W. H., Wallis W. A. Use of ranks in one-criterion variance analysis // Journal of the American Statistical Association. 1952. Vol. 47, No. 260. P. 583–621. [10.1080/01621459.1952.10483441](#).
- [Ха626] Хабибуллин А. И., Муравьева Е. А. Алгоритм экспертной системы для управления жизненным циклом нефтяных скважин // СИИТ. 2026. Т. 8, № 3(27). С. 71–79. [SBYISM](#).
- [Евс20] Евсюткин И. В., Марков Н. Г. Глубокие искусственные нейронные сети для прогноза значений дебитов добывающих скважин // Известия ТПУ. Инжиниринг георесурсов. 2020. Т. 331, № 11. С. 142–150. [KLOYSS](#).
- [Пис24] Пискунов С. А., Давуди Ш. Прогнозирование дебита горизонтальных скважин с применением модели машинного обучения // Известия ТПУ. Инжиниринг георесурсов. 2024. № 5. [OSZSPB](#).
- [Wil45] Wilcoxon F. Individual comparisons by ranking methods // Biometrics Bulletin. 1945. Vol. 1, No. 6. P. 80–83. [10.2307/3001968](#).
- Dmitrievsky A. N. Risk analysis when using artificial intelligence technologies in the oil and gas production sector. *Automation, Telemechanization and Communication in Oil Industry*, 2021, no. 7(576), pp. 17–27. (In Russian). [QPMNQE](#).
- Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016. 800 p.
- Anikeev D.P., Anikeeva E.S. Application of neural networks to determine the parameters of PVT models used in solving hydrodynamic modeling problems. *Actual Problems of Oil and Gas*, 2024, no. 3, pp. 45–54. (In Russian). [FHHNCR](#).
- Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009. 745 p. [10.1007/978-0-387-84858-7](#).
- Kruskal W.H., Wallis W.A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 1952, vol. 47, no. 260, pp. 583–621. [10.1080/01621459.1952.10483441](#).
- Khabibullin A.I., Muravyeva E.A. Algorithm of an expert system for managing the life cycle of oil wells. *SIIT*, 2026, vol. 8, no. 3(27), pp. 71–79. (In Russian). [SBYISM](#).
- Evsyutkin I.V., Markov N.G. Deep artificial neural networks for forecasting production well flow rates. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2020, vol. 331, no. 11, pp. 142–150. (In Russian). [KLOYSS](#).
- Piskunov S.A., Davoodi Sh. Horizontal well production forecasting using a machine learning model. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2024, vol. 335, no. 5, pp. 112–121. (In Russian). [OSZSPB](#).
- Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1945, vol. 1, no. 6, pp. 80–83. [10.2307/3001968](#).

ОБ АВТОРАХ | ABOUT THE AUTHORS

РУБЛЕВА Екатерина Сергеевна
ООО «Газпром ВНИИГАЗ», Россия.
kateledjer@yandex.ru ORCID: 0009-0002-3650-8343.
Специалист.

RUBLEVA Ekaterina Sergeevna
LLC "Gazprom VNIIGAZ", Russia.
kateledjer@yandex.ru ORCID: 0009-0002-3650-8343.
Specialist.

МЕТАДАННЫЕ | METADATA

Заглавие: Моделирование технологических режимов и прогноз показателей газоконденсатных скважин на основе методов машинного обучения и кластерного анализа.

Авторы: Рублева Е. С.

Аннотация: В работе представлен подход к моделированию гидродинамических показателей работы газоконденсатных скважин на основе гибридного ансамблирования методов машинного обучения и физически обоснованной кластеризации. Для исключения эффекта утечки данных из пространства входных признаков извлечены прямо вычисляемые гидродинамические параметры (число Рейнольдса, скорость потока, коэффициент Фаннинга). Проведена робастная фильтрация аномалий и кластеризация промысловых данных методом K-Means, подтвержденная физико-гидродинамической интерпретацией и критерием Краскела—Уоллиса. Сформирован гибридный ансамбль градиентного бустинга с наложением монотонных физических ограничений. Продemonстрировано статистически значимое по критерию Уилкоксона преимущество предложенного алгоритма над едиными моделями машинного обучения: относительная ошибка прогностической модели дебита снижена с 23.16% до 19.70% при коэффициенте детерминации равном 0.898. Для инверсной задачи восстановления забойного давления достигнута средняя абсолютная ошибка 0.517 атм.

Ключевые слова: Газоконденсатная скважина; дебит газа; забойное давление; машинное обучение; K-Means; PCA; градиентный бустинг; физически информированные модели; критерий Уилкоксона.

Язык: Русский.

Статья поступила в редакцию 11 августа 2026 г.

Title: Modeling of Operating Modes and Performance Forecasting for Gas Condensate Wells Based on Machine Learning Methods and Cluster Analysis.

Authors: Rubleva E. S.

Abstract: This paper presents an approach to modeling the hydrodynamic performance of gas condensate wells based on a hybrid ensemble of machine learning methods and physically informed clustering. To eliminate the risk of data leakage, directly calculated hydrodynamic parameters (Reynolds number, flow rate, Fanning friction factor) were removed from the input feature space. Robust outlier filtering and K-Means clustering of field data were performed, validated by physical-hydrodynamic interpretation and the Kruskal–Wallis test. A hybrid ensemble using gradient boosting with imposed monotonic physical constraints was developed. A statistically significant advantage of the proposed algorithm over unified machine learning models according to the Wilcoxon test was demonstrated: the relative error of the flow rate forecasting model was reduced from 23.16% to 19.70%, with a coefficient of determination of 0.898. For the inverse problem of bottomhole pressure reconstruction, a mean absolute error of 0.517 atm was achieved.

Key words: Gas condensate well; gas production rate; bottom-hole pressure; machine learning; K-Means; PCA; gradient boosting; physics-informed neural networks/models; Wilcoxon test.

Language: Russian.

The article was received by the editors on 11 August 2026.