REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

PREDICTION OF POROUS-MEDIUM COLMATATION USING MACHINE LEARNING METHODS

Specialty:	2525.01 – Oil and gas field development and operation
Field of science:	Technical sciences
Applicant:	**Timur Elmanovich Abdulmutalibov**

Baku - 2026

The dissertation was performed at the Department of Oil and Gas Engineering of Azerbaijan State Oil and Industry University.

Scientific supervisor:	PhD in Technical Sciences, Associate Professor **Yelena Yevgenyevna Shmoncheva**
Official opponents:	Doctor of Technical Sciences, Associate professor **Nahida Ismat Huseynova**
	Doctor of Technical Sciences, Associate Professor **Mubariz Sevdimali Khalilov**
	PhD in Technical Sciences **Ulviyya Tahir Ahmadova**

Dissertation council FD 2.03 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Azerbaijan State Oil and Industry University.

Chairman of the Dissertation Council: ____________	Doctor of Technical Sciences, Professor **Elman Kheyrulla Iskandarov**
Scientific secretary of the Dissertation Council: ____________	PhD in Technical Sciences, Associate Professor **Gullu Valeh Cabbarova**
Chairman of the Scientific Seminar: ____________	Doctor of Technical Sciences, Associate Professor **Natig Neyman Hamidov**

I confirm the signatures:

Scientific Secretary of ASOIU, PhD in Technical Sciences,

Associate Professor ________________ **Inglab Namig Aliyev**

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ AZƏRBAYCAN DÖVLƏT NEFT VƏ SƏNAYE UNİVERSİTETİ PUBLİK HÜQUQİ ŞƏXSİ

GENERAL CHARACTERISTICS OF THE WORK

Relevance and degree of development of the topic. Colmatation of the porous medium in the near-wellbore formation zone is among the most widespread causes of deterioration of the filtration connection between the reservoir and the well. Accumulation of solid particles, formation of mineral and organic deposits, migration of fine dispersed material, development of a filter cake, changes in wettability and biological processes reduce the effective cross-section of pore channels. As a result, hydrodynamic resistance increases, permeability decreases, production-well productivity deteriorates and injection-well injectivity declines.

The practical complexity of the problem is that colmatation develops gradually and, at an early stage, rarely appears as an unequivocally identifiable failure. In production data it is reflected indirectly through changes in flow rate, pressure, water cut, gas–oil ratio, productivity index, injectivity, pressure drawdown and additional resistance in the near-wellbore zone. Similar changes may also result from reservoir depletion, operating-mode changes, well shutdowns, phase transformations and technological interventions. Therefore, no single indicator can provide sufficient proof of colmatation; several indicators and the operating history must be interpreted jointly.

Laboratory core studies, well testing, pressure-buildup analysis and skin-factor estimation can diagnose formation damage, but they usually characterize a condition that has already formed. To prevent sustained productivity loss, retrospective diagnosis must be complemented by early risk assessment. This is particularly important for large well stocks, where frequent special studies of every well are constrained by economic and organizational factors.

Modern production-accounting systems generate large time-series datasets. Combined with machine-learning methods, these data make it possible to identify parameter combinations that precede deterioration of filtration properties. However, direct application of algorithms without accounting for the temporal direction of the process, the physical meaning of the features and the quality of the source data may overestimate accuracy. The scientific task therefore includes not only selecting an algorithm, but also constructing a physically justified target variable, a causally correct feature space, a temporal validation scheme and an engineer-interpretable probabilistic output.

Analysis of previous studies showed that formation-damage mechanisms and diagnostic methods have been investigated in considerable detail, while machine-learning methods are actively used to forecast production rate, pressure, complications and technological indicators. At the same time, an integrated methodology connecting production time series, diagnostic parameters of the near-wellbore zone, early forecasting of colmatation risk and procedures for engineering use of the results remains insufficiently developed. This determines the relevance of the dissertation [1–3, 5–9, 11].

Published studies commonly use physical colmatation models to describe particle transport and retention, changes in porosity and permeability, deposit formation and additional hydrodynamic resistance. Their advantage is causal interpretability, but practical application requires pore-space parameters, kinetic coefficients and particle properties that are rarely measured during routine operation. Statistical and analytical methods are easier to apply to field data, but mainly register deviations that have already occurred. Machine-learning methods can account for multidimensional and nonlinear dynamics, although without temporal validation and engineering constraints on features they may provide an overoptimistic estimate of performance.

A combined approach was therefore adopted. Physical understanding is used to select diagnostic indicators and explain the output, while production time series are used to form predictive features. This approach does not equate a decrease in production rate with colmatation and does not treat the calculated signal as a final diagnosis. Its purpose is to detect early a stable combination of changes that requires verification against the operating history, current regime, well-test data and the condition of the near-wellbore zone.

A specific feature of this approach is that the model output is not interpreted as an independent technological parameter, but as a risk signal that must be compared with the well operating history. At early stages of colmatation, changes in production rate, pressure, gas–oil ratio and water cut may be weak and unstable. Therefore, the forecast should not rely on an abrupt change in one indicator, but on the coherent temporal dynamics of several parameters. This formulation reduces the risk of erroneous diagnosis and, at the same time, indicates which groups of parameters should be checked first by the engineer.

The notion of risk adopted in the dissertation does not replace conventional diagnosis; rather, it complements it. It allows well tests, operating-regime data and information on the near-wellbore-zone condition to be used more purposefully. This is especially important for large well stocks, where frequent special studies of every object are practically difficult. Early risk assessment helps identify the most problematic wells first and substantiate the order of additional investigations and preventive actions.

Aim and objectives of the study. The aim is to develop and validate a methodology for early forecasting of colmatation risk in the porous medium of the near-wellbore formation zone using production time series and machine-learning methods, in order to improve well-condition monitoring and support technological decisions in oil and gas field development and operation.

The following objectives were addressed:

1. to analyse colmatation mechanisms and determine their influence on reservoir filtration and storage properties, production-well productivity and injection-well injectivity;
2. to substantiate a system of diagnostic parameters characterizing the near-wellbore zone, including permeability, skin factor, reservoir pressure, productivity index, injectivity, pressure drawdown and derived operating indicators;
3. to form a production time-series database reflecting well-performance dynamics and changes in filtration characteristics;
4. to develop a colmatation index linking actual well productivity with its calculated potential;
5. to construct a causally correct feature space incorporating lagged parameters, rolling statistics, rates of change and operating ratios;

6. to select and temporally evaluate candidate models and choose an L1-regularized logistic-regression model for the final pipeline, forecasting the event IC < 0.90 over a 14-day horizon;

7. to evaluate ranking performance, calibration, sensitivity and inter-well transferability using an independent G12 test period and an external G09 sample;

8. to develop a practical procedure for using the forecast in well-stock monitoring and generating signals that require engineering verification.

Object of the study. Production and injection wells in oil and gas fields where the filtration properties of the near-wellbore zone deteriorate.

Subject of the study. Methods for forming diagnostic indicators and forecasting porous-medium colmatation risk from production time-series analysis.

Research methods. The study used field and hydrodynamic analysis, statistical and correlation analysis, causally correct feature engineering, L1- and L2-regularized logistic regression, a linear support-vector machine and random forest. Chronological splitting with two 30-day protective gaps, rolling-origin validation with an expanding training window, median imputation, robust scaling, automatic class balancing, Platt probability calibration, ablation analysis, block bootstrap and local interpretation of individual forecasts were applied.

The principal data-processing requirement was preservation of temporal causality. For a forecast date t, only information available by that date was used, while the target was defined in the future interval (t; t+14]. The final model configuration, calibrator and operational threshold were selected using the training and validation data; the test sample was not used for tuning.

Reliability of the results is supported by actual production data, physical-consistency checks, strict future-event labelling, chronological splitting with protective gaps, an independent G12 test sample, comparison of candidate models, analysis of consecutive temporal subperiods, block-based uncertainty assessment and external transfer of the fixed G12 pipeline to G09 without retraining on G09.

Main provisions submitted for defence:

1. A methodology for early forecasting of colmatation risk that combines production time series, diagnostic parameters and a probabilistic model.

2. The colmatation index as an engineering-interpretable measure of deviation of actual well productivity from its calculated potential.

3. A causally correct feature space formed exclusively from information available at the forecast date.

4. The final L1-regularized logistic model with Platt calibration and an operational threshold selected on the pre-event validation subset subject to $FPR \leq 0.15$.

5. A framework for assessing stability and engineering interpretability using ablation analysis, non-zero coefficients, local feature contributions and evaluation across consecutive temporal subperiods.

6. A daily-use and transfer procedure including data control, calculation of 19 features, calibrated probability estimation, alarm generation and mandatory engineering verification.

Scientific novelty of the study:

1. A strict target-variable scheme was proposed: an event is recorded when the colmatation index IC falls below 0.90 in the future interval (t; t+14], with the current date and all future information excluded from the features.

2. A compact set of 19 causally admissible features was constructed, combining operating regime, oil, gas and water rates, water cut, gas ratios, calculated potential, current and previous IC, IC change and observation-irregularity indicators.

3. L1-regularized logistic regression with $C = 0.1$ was justified as the final model because it provides embedded coefficient sparsity, a probabilistic output and direct engineering interpretation under a limited time-series sample.

4. A sequential procedure was developed for model selection, Platt calibration and operational-threshold selection on the pre-event validation subset, followed by independent G12 testing and external transfer to G09.

5. Ranking, calibration, feature-set sensitivity, temporal heterogeneity and inter-well transferability were evaluated jointly, allowing the practical limits of the model to be quantified.

Theoretical and practical significance. The theoretical significance lies in formalizing colmatation-related deterioration as a future probabilistic event linked to the dynamics of the engineering-interpretable IC index. The practical significance is the possibility of daily risk ranking over a 14-day horizon while controlling false early warnings.

The developed pipeline can serve as an additional analytical module in field monitoring. The primary threshold $\tau^* = 0.4446$ generates a signal for verification, but is neither an automatic diagnosis nor a command for technological intervention. The engineer compares the warning with operating conditions, measurement quality, rates, pressure, water cut, gas ratios and intervention history.

Operational use includes daily data updating, calculation of a fixed feature set, median imputation and scaling with training parameters, L1-model scoring, Platt calibration, threshold comparison and alarm logging. Distribution changes trigger calibration monitoring, threshold review and, where necessary, retraining.

Personal contribution of the applicant. The applicant formulated the problem, systematized colmatation mechanisms, prepared and integrated the source production data, calculated the colmatation index, formed temporal features and target variables, built and evaluated the models, interpreted the results and developed the operational procedure. In co-authored works, the applicant performed the data analysis, principal calculations and synthesis of the dissertation results.

Organization where the work was carried out. Department of Petroleum Engineering, Azerbaijan State Oil and Industry University.

Approbation of the work. The main provisions and results were presented and discussed at the following scientific events:

1. International Scientific and Technical Conference "Modern Technologies in the Oil and Gas Industry – 2023", dedicated to the 75th anniversary of Ufa State Petroleum Technological University, Ufa, 2023;

2. VII International Scientific and Practical Conference "Bulatov Readings", Krasnodar, 2023;

3. Republican Scientific Conference of Young Researchers and Doctoral Students dedicated to the 100th anniversary of the birth of National Leader Heydar Aliyev, Baku, 4–5 May 2023;

4. VI All-Russian Scientific and Practical Conference "Oil and Gas Engineering, Technosphere Safety and Rational Use of Natural Resources: Current Realities", Makhachkala, 16–17 May 2023;

5. Scientific and Practical Conference "Heydar Aliyev and the Oil Strategy of Azerbaijan: Achievements of Oil and Gas Geology and Geotechnology", Baku, 23–26 May 2023;

6. International Scientific and Practical Conference "Field Development Technologies and Process Modelling in Oil and Gas Production", dedicated to Academician A.Kh. Mirzajanzade, Ufa, 2023;

7. SPE Caspian Technical Conference and Exhibition, Baku, 21–23 November 2023.

Publications. Eleven scientific works have been published on the dissertation topic, including four journal articles, six conference papers and one Eurasian patent. Two articles were published in journals indexed in both Scopus and Web of Science, and one conference paper was indexed in Scopus.

Structure and scope of the dissertation. The dissertation consists of an introduction, four chapters, conclusions, a list of references, and two appendices. The work includes 9 figures, 25 graphs, 51 tables, a list of references comprising 150 sources, and 2 appendices. The number of characters in the dissertation is as follows: Introduction — 13,657; Chapter I — 52,546; Chapter II — 90,708; Chapter III — 41,856; Chapter IV — 33,121; Conclusions — 2,334. The total number of characters is 234,222.

MAIN CONTENT OF THE WORK

The introduction substantiates the relevance of early forecasting of colmatation in the near-wellbore formation zone, formulates the aim and objectives, and defines the object, subject and methods. It is shown that machine learning has engineering value only when the target variable is physically justified, temporal causality is strictly observed and the probabilistic output is clearly interpretable.

Chapter I systematizes concepts of porous-medium colmatation and near-wellbore formation damage. The term "colmatation" describes reduction of porous-medium flow capacity due to plugging or narrowing of pore channels, whereas formation damage is a broader concept including mechanical, physicochemical, phase, thermal and geomechanical causes of productivity deterioration. The skin factor reflects the total additional resistance near the well but does not identify its specific mechanism.[1]

Mechanical colmatation is associated with particle invasion and deposition, sieving at pore throats, bridging and migration of fines. Chemical colmatation results from salt precipitation and products of water incompatibility. Organic deposits include asphaltenes, resins and paraffins. Colloidal and emulsion processes depend on dispersion stability, salinity, acidity and fluid composition. Biological colmatation develops because of microbial growth and biofilm formation. In a real reservoir, these mechanisms often act jointly and intensify one another.[2]

The chapter reviews models of particle transport and deposition, deep-bed filtration, permeability alteration and additional resistance. Physical models provide causal explanations but require parameters that are not always available under field conditions. Data-driven models can identify nonlinear relationships, but their results depend on labelling quality, sample size and validation correctness (Figure 1).

[1] Civan F. Reservoir Formation Damage: Fundamentals, Modeling, Assessment, and Mitigation. 4th ed. Amsterdam: Elsevier, 2023.
[2] Khilar K.C., Fogler H.S. Migrations of Fines in Porous Media. Dordrecht: Kluwer Academic Publishers, 1998. DOI: 10.1007/978-94-017-0967-6.

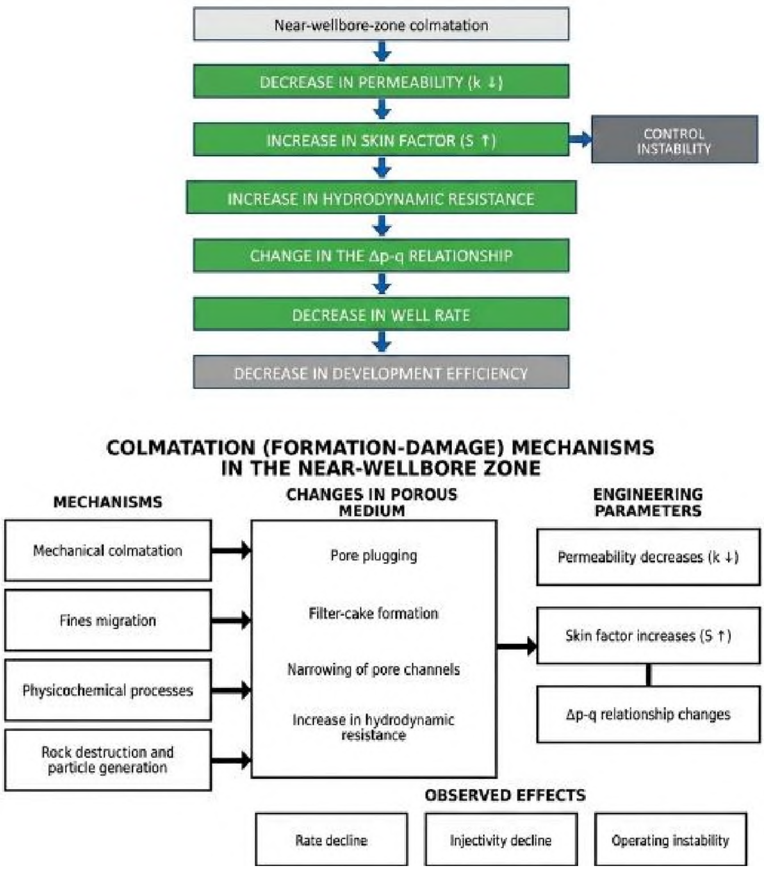


Figure 1 – Main colmatation mechanisms and their relationship with filtration parameters

A hybrid approach is therefore justified: process physics defines admissible features, constraints and the meaning of the target variable, while machine learning estimates the latent state and the probability of its deterioration.

Mechanism analysis showed that identical field manifestations may have different causes. Productivity decline may result from mechanical plugging, migration of indigenous fines, clay swelling, salt or paraffin deposition, emulsion or water blocking, and changes in operating regime. Diagnosis therefore cannot rely on a single parameter. Distinguishing the causes requires joint analysis of production rate, pressure, water cut, gas–oil ratio, productivity index, injectivity and additional resistance, together with records of workovers, treatments and operating-mode changes.

The diagnostic methods considered were divided into direct and indirect methods. Direct laboratory methods assess changes in permeability, pore-space structure, deposit composition and return permeability. Well tests and pressure-buildup analysis provide an integral assessment of the reservoir–well system, but require special operations and do not provide continuous observation. Production time series are diagnostically less specific, but are regularly available and therefore suitable for early monitoring. On this basis, the model is required to estimate risk probability, while the final conclusion is made by an engineer after the signal has been verified.

Core-flow tests directly determine changes in permeability and return permeability. X-ray tomography, microscopy and nuclear magnetic resonance characterize pore space and fluid distribution. Hydrodynamic well tests provide an integral assessment of the reservoir–well system. Production time series are less directly interpretable than special studies, but their continuous accumulation makes them suitable for early screening of a well stock.

The principal forecasting requirements were defined: the target variable must represent deterioration of the filtration connection; features must be available at the forecast time; training and validation samples must be separated chronologically; risk probabilities must be calibrated; and the result must have an engineering explanation. The review led to the following general workflow: data preparation – diagnostic index – temporal features – model – calibration – alarm threshold – engineering verification (Figure 2).

Chapter II transitions from general mechanisms to a system of engineering parameters. Diagnosis of the near-wellbore zone should account for permeability, skin factor, reservoir and bottomhole pressure, productivity index or injectivity, pressure drawdown, actual flow rate and operating changes. None of these indicators alone is sufficient to establish colmatation, but their coherent dynamics can reveal the formation of a persistent filtration restriction.[3]

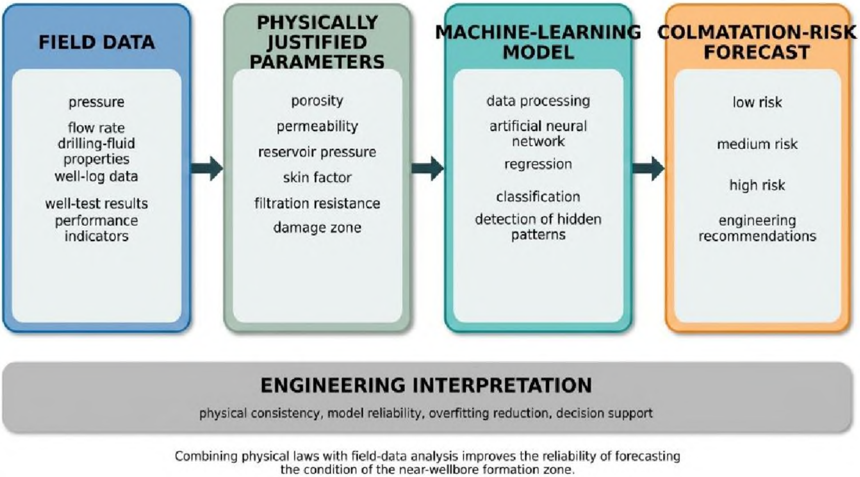


Figure 2 – Hybrid colmatation-forecasting scheme based on physical parameters and field data

The role of controllable technological factors was examined: drilling-fluid composition and dispersion, filter-cake quality, fluid compatibility with reservoir minerals, solids concentration, contact time, hole cleaning and reservoir-opening regime. Lower filtration and formation of a thin, low-permeability but removable cake reduce particle-invasion depth. At the same time, excessive viscosity, particle aggregation or system instability may impair cuttings transport and increase invasion risk [1, 2, 5–7].

Nanoparticles were considered separately as a means of controlling cake structure and interfacial interaction. Their use should be assessed not merely by the presence of a nanophase, but by a set of criteria: filtration reduction, dispersion stability, compatibility with the rock, retention of rheological properties and absence of large

---

[3] Bourdet D. Well Test Analysis: The Use of Advanced Interpretation Models. Amsterdam: Elsevier, 2002. 438 p.

aggregates. This emphasizes the need for a technological, rather than only laboratory, assessment of the anti-colmatation effect.

As an intermediate stage, the possibility of restoring integral reservoir parameters from interpretation data and artificial-neural-network models was demonstrated. For six wells, estimates of permeability, skin factor and initial reservoir pressure were compared. The smallest relative error was obtained for pressure, while higher errors were obtained for permeability and skin factor. These results confirm that models can be used as an auxiliary assessment tool, but do not eliminate the need to control source data and compare the output with hydrodynamic interpretation [3, 8].

The diagnostic system is built around parameters with direct physical meaning. Permeability characterizes porous-medium conductivity; productivity index and injectivity reflect the actual relationship between rate and pressure drawdown; the skin factor summarizes additional resistance in the near-wellbore zone. Reservoir and bottomhole pressures help separate general reservoir depletion from local inflow deterioration. Where regular well tests are unavailable, calculated and reconstructed indicators are used, but their uncertainty is considered during interpretation and must not be concealed by high formal accuracy of the predictive model.

The calculated well potential is determined for comparable operating conditions and serves as the reference for assessing deviation of actual productivity. It is not an invariant nameplate value: it depends on pressure, production regime, fluid properties and the available operating history. Before feature construction, units and timestamps are checked, missing values are processed, and shutdowns, workovers and abrupt technological changes are identified. Values obtained under non-comparable regimes must not be used automatically as evidence of formation damage.

The analysis produced a physically consistent workflow: production data undergo quality control; diagnostic and temporal features are then calculated; the model produces a risk probability; and, after comparison with the operational threshold, the result is sent to the engineer. The workflow emphasizes that machine learning does not automatically identify a colmatation mechanism. A high risk is a reason to check the operating regime, intervention history, latest study results and additional diagnostic measures.

Features are formed only from information known at the forecast time. Previous parameter values, changes over one or more time steps, rolling means and standard deviations, ratios of actual to calculated rates and operating-stability indicators are used. This prevents future information from entering the training sample. Windows of 3, 7 and 14 days represent, respectively, short-term response, weekly dynamics and more persistent regime change; using them jointly helps distinguish a single outlier from a developing trend.

In Chapter III, colmatation is treated as a latent operating state manifested by sustained deterioration of actual productivity relative to calculated potential. The quantitative analysis combines production data from well G12: actual and theoretical oil rates, water and gas rates, water cut, gas–oil ratio, wellhead pressure, choke position, gas-lift rate and results of the latest well studies. Timestamps were synchronized, measurement units unified, missing values processed, shutdowns marked and physically impossible values checked.[4]

Well G12 was selected because a sufficiently long observation series and a set of operating indicators were available for testing the computational procedure. This object was used to verify the sequence of index calculation, risk-event labelling and feature-space construction. The analysis focused not only on the absolute level of the index, but also on the persistence of its decrease, the subsequent recovery rate and joint dynamics with other operating parameters. This made it possible to distinguish short-term technological changes from a more persistent filtration restriction.

The colmatation index was adopted as the key diagnostic indicator:

$$IC(t) = Qactual(t) / Qtheoretical(t),$$

where Qactual(t) is the actual oil rate and Qtheoretical(t) is the calculated well potential on the same date. Values close to one indicate

[4] Tariq Z., Abdulraheem A., Mahmoud M. et al. Data-driven analytics and machine learning applications in petroleum engineering. Journal of Petroleum Exploration and Production Technology. 2021. Vol. 11. P. 4339–4374.

agreement between actual operation and the calculated level. A decrease in the index indicates increasing discrepancy, but does not by itself prove colmatation. The following relative change was also used to analyse dynamics:

$$\Delta IC(t) = [IC(t) - IC(t-1)] / IC(t-1) \times 100\%.$$

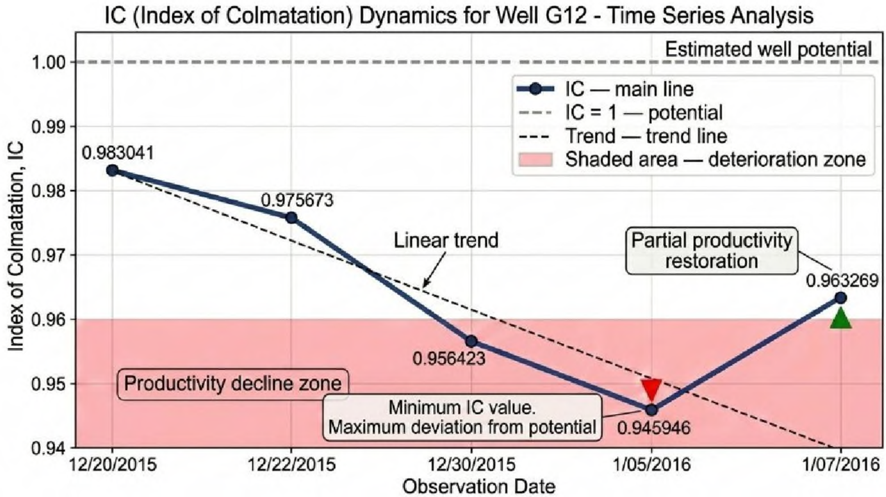
For the first five observation dates, the index changed from 0.983 to 0.946 and then partially recovered to 0.963. This example demonstrates the indicator's sensitivity to operating-mode changes and, at the same time, the need to analyse a long time series and additional parameters (Table 1, Graph 1).

Table 1 – Example calculation of the colmatation index

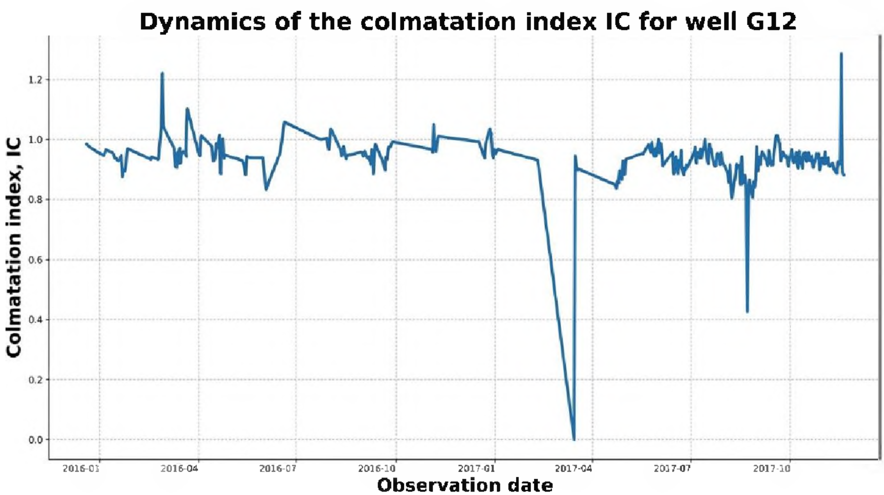
Date	Qactual, bbl/day	Qtheoretical, bbl/day	IC
20.12.2015	8521	8668	0.983
22.12.2015	9425	9660	0.976
30.12.2015	6738	7045	0.956
05.01.2016	6790	7178	0.946
07.01.2016	7238	7514	0.963

The complete series covers 2015–2017. At the beginning of the observations, actual oil rate reached approximately 8,000–9,500 stb/day, then changed unevenly and declined in later periods. At most points, the colmatation index remained close to one, but the series included several deep and prolonged drops. Some extreme values could be associated with shutdowns or peculiarities in calculating the theoretical rate and were therefore not automatically interpreted as physical colmatation (Graph 2).

The dataset contains 308 consistent observations. The mean index was 0.932 and the median 0.938. Approximately 17.3% of observations were below the adopted diagnostic level, and 78 local drops were identified. Sustained decline intervals lasted from 3 to 13 days. The most pronounced episode occurred in August 2017, when the minimum index reached 0.428.



Graph 1 – Illustration of calculation and dynamics of the colmatation index for the first dates of well G12



Graph 2 – Dynamics of the colmatation index for well G12 over the observation period

Rank correlation analysis showed a positive relationship of the index with gas rate and wellhead pressure and a negative relationship with water cut, water rate, choke opening and gas-lift rate. Correlations were not treated as evidence of causality, but helped direct feature construction and identify parameters accompanying the deviation of actual productivity from its calculated level (Table 2).

Table 2 – Relationship between operating parameters and the colmatation index

Parameter	Rank correlation coefficient with IC
Gas rate	0.419
Wellhead pressure	0.391
Gas–oil ratio	0.245
Gas-lift rate	–0.210
Choke opening	–0.289
Water rate	–0.366
Water cut	–0.419

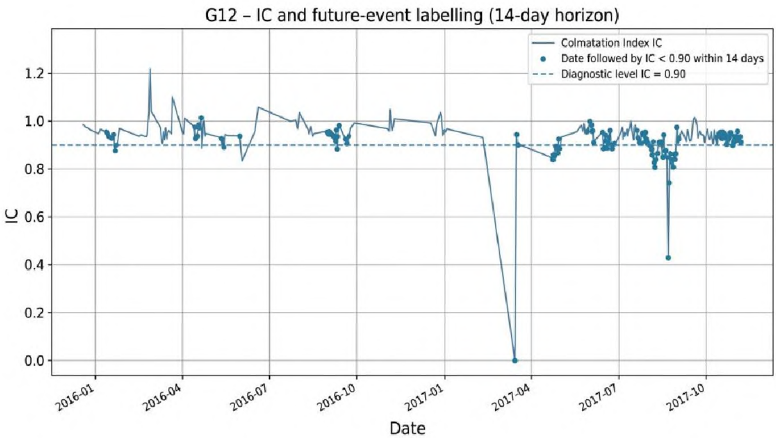
The analysis used 308 consistent observations obtained after production and calculated data were joined by date. In addition to the overall index level, local drops, duration below the diagnostic level and recovery rate were considered. A short-term decrease coinciding with a shutdown or abrupt choke change was interpreted differently from gradual deterioration under a relatively stable regime. This made it possible to form risk events without claiming that every index decrease had a single physical cause.

Correlation analysis was used only as a preliminary step. The positive relationship of the index with gas rate and wellhead pressure, and the negative relationship with water cut, water rate and choke opening, are consistent with changes in the inflow regime but do not establish causality. Features were therefore not selected solely by pairwise correlation. Their physical meaning, availability before the forecast date, temporal stability and absence of a direct relationship with the future label were also considered.

An integrated dataset was formed containing the original channels, the colmatation index, its previous values, rates of change, rolling means and standard deviations over 3-, 7- and 14-day windows, data-quality features and binary event indicators. This structure enables transition from engineering analysis to a causally correct forecasting problem.

Chapter IV forms the target variable in a causally correct manner. For each date t, the label is positive when the colmatation index falls below 0.90 at least once during the following 14 days, that is, within (t; t+14]. The current date is excluded from the target window and future-period indicators are not used as features.[5]

The final forecast horizon is fixed at 14 days. A supplementary label based on a relative IC decrease was used only for an auxiliary check and was not included in the final operational pipeline. This separation prevents the main early-warning task from being mixed with alternative labelling variants (Graph 3).



Graph 3 – IC dynamics and future-event labelling for well G12 at a 14-day horizon

---

[5] Bergmeir C., Benítez J.M. On the use of cross-validation for time series predictor evaluation. Information Sciences. 2012. Vol. 191. P. 192–213. DOI: 10.1016/j.ins.2011.12.028.

Area under the precision–recall curve (PR-AUC) was adopted as the primary performance metric and area under the ROC curve (ROC-AUC) as a supplementary metric. The Brier score and expected calibration error (ECE) were used for probabilistic assessment; precision, recall and false-positive rate (FPR) were used for the operating point.

The final feature space contains 19 variables: observation day, on-stream hours, choke, wellhead pressure, actual oil rate, gas and water rates, NGOR, water cut, GLR, calculated oil rate, calculated gas–oil ratio and water cut, current IC, IC change, previous IC, a proxy productivity index, relative GLR change and days since the previous observation. Future labels and service fields were excluded.

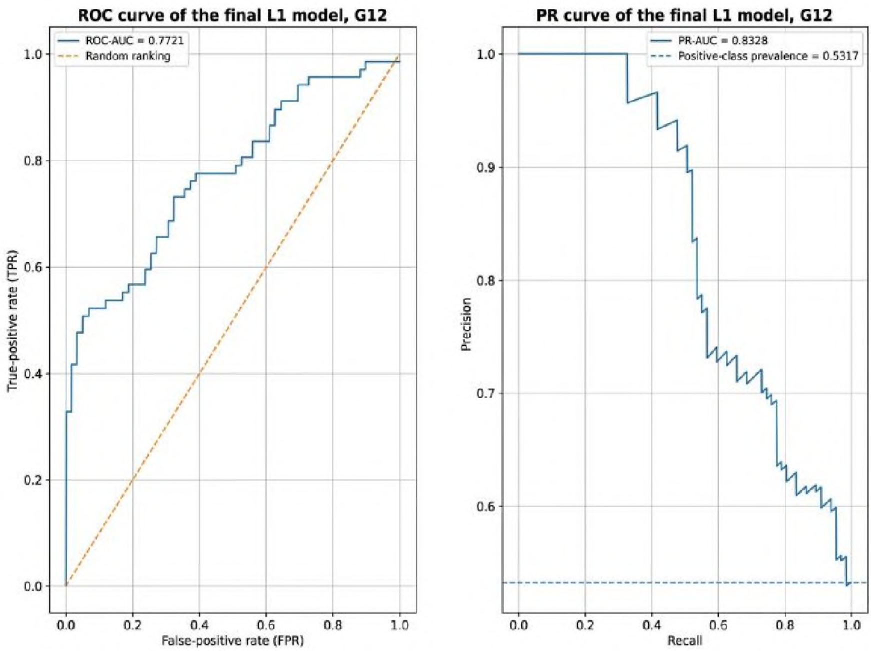
After strict labelling, G12 data were split chronologically: training — 20 December 2015 to 1 December 2016; validation — 1 January to 3 June 2017; test — 4 July to 6 November 2017. Thirty-day protective gaps were placed between training and validation and between validation and test. The final parts contain 108, 35 and 126 observations, respectively (Table 3).

Table 3 – Chronological split of the G12 series with protective gaps

Sample	Start	End
Training	20.12.2015	01.12.2016
Validation	01.01.2017	03.06.2017
Test	04.07.2017	06.11.2017

L1- and L2-regularized logistic models, a linear support-vector machine and random forest were compared on the validation sample. L1-regularized logistic regression with C = 0.1 achieved the highest PR-AUC: PR-AUC = 0.8355 and ROC-AUC = 0.7961. Random forest using the full set of 157 features achieved PR-AUC = 0.4249 and did not provide a benefit commensurate with its complexity.

The final model was fixed before the test data were examined. On the independent G12 test interval, PR-AUC was 0.8328 and ROC-AUC was 0.7721, compared with a PR-AUC baseline of 0.5317. Thus, the model retained ranking ability on the subsequent period and substantially outperformed trivial ranking by positive-class prevalence (Graph 4).



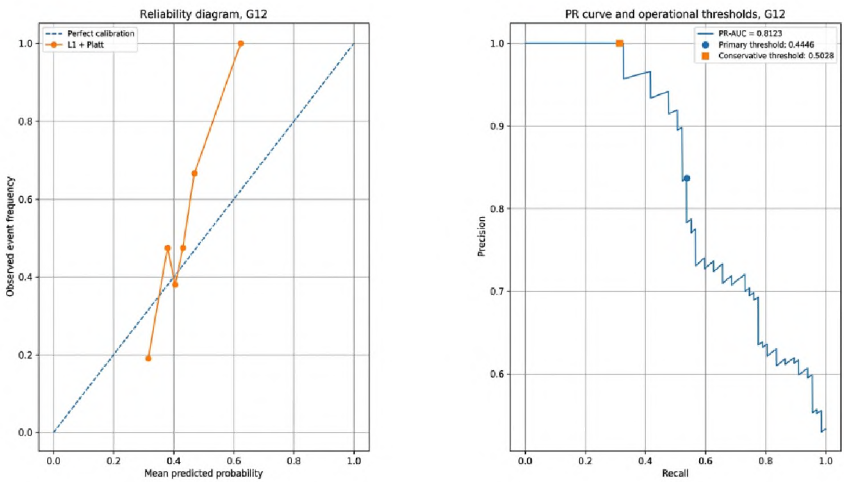
Graph 4 – ROC and PR curves of the final L1-regularized model on the G12 test sample

Rolling-origin validation with an expanding training window and a protective gap was used for additional temporal evaluation. Imputation, scaling, model fitting and calibration were estimated only on the corresponding earlier data. This scheme reproduces the real arrival of observations and prevents leakage from the future.

Probability calibration was performed after model selection. Raw scores, Platt scaling and isotonic regression were compared. Platt scaling was selected for the final pipeline: on the test sample, the Brier score was 0.2248 and ECE was 0.1350. Isotonic calibration produced zero ECE on the small validation sample but worsened test PR-AUC, ROC-AUC and ECE, indicating overfitting.

The primary operational threshold $\tau^* = 0.4446$ was selected on the pre-event validation subset, where current IC had not yet fallen below 0.90, by maximizing recall subject to FPR $\leq$ 0.15. A conservative control point $\tau = 0.5028$, selected on the full validation sample, was also retained. The threshold is applied to the predicted probability p(t), whereas IC < 0.90 defines the observed event.

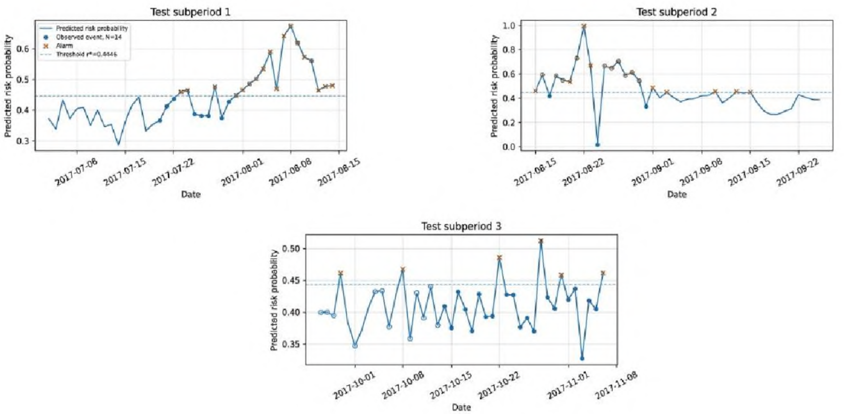
On the full G12 test sample, $\tau^* = 0.4446$ identified 36 of 67 positive observations, generated 7 false alarms and missed 31 observations; precision was 0.8372, recall 0.5373 and FPR 0.1186. In the pre-event subset, 14 of 44 positive observations were identified with 6 false alarms; FPR was 0.1034. The conservative threshold eliminated false alarms but reduced recall to 0.3134 (Graph 5).



Graph 5 – Probability calibration and operational thresholds of the final G12 model

Ablation analysis showed that L1 regularization retained four non-zero coefficients out of 19 inputs: current IC, water rate, relative GLR change and previous IC. Removing current IC reduced PR-AUC to 0.8054, while removing both current and previous IC reduced PR-AUC to 0.5924 and ROC-AUC to 0.4908. The principal predictive contribution therefore comes from the IC family, while operating features provide limited additional information in the available sample.

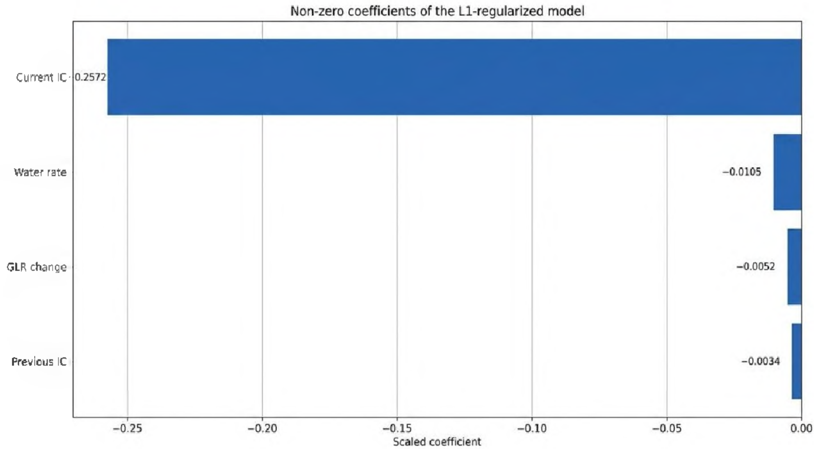
Performance was temporally heterogeneous. Across three consecutive test subperiods, recall was 0.6923, 0.8235 and 0.1667, while FPR was 0, 0.2000 and 0.1111, respectively. A block bootstrap with a 14-day block yielded a wide 95% PR-AUC interval of 0.5565–0.9598, so a single point estimate must be interpreted cautiously (Graph 6).



Graph 6 – Risk probability and alarms across consecutive G12 test subperiods

The coefficient sign indicates the direction of influence after robust scaling. The negative coefficient of current IC means that lower IC increases predicted risk. The much smaller coefficients of water rate, GLR change and previous IC supplement the principal signal. A non-zero coefficient does not establish physical causality; it describes the feature contribution to the model decision (Graph 7).

Local interpretation was performed for a correct early warning on 23 July 2017 and a missed positive observation on 16 October 2017. In the first case, probability 0.4604 exceeded the threshold 11 days before the event; in the second, probability 0.4322 remained below the threshold although the event occurred 12 days later. The comparison shows why the operational threshold must be interpreted together with engineering context.

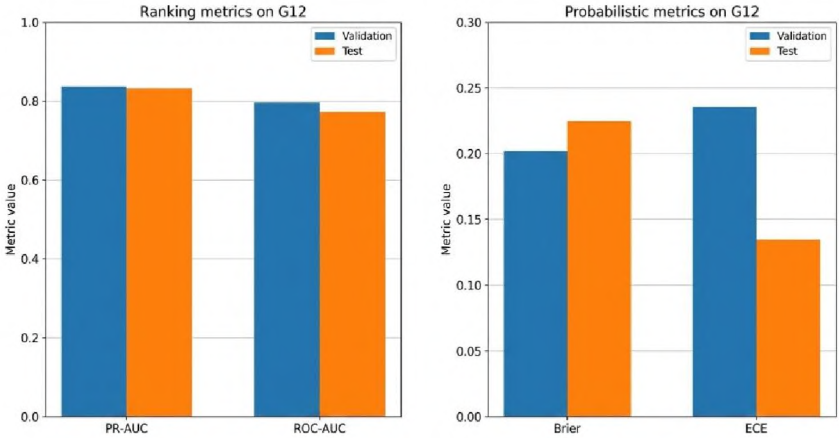


Graph 7 – Non-zero coefficients of the final L1-regularized model

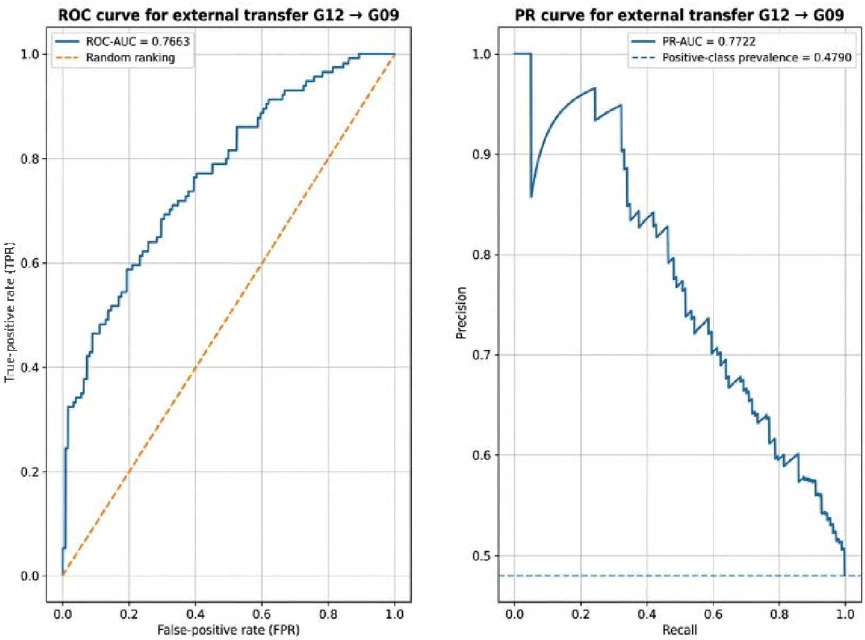
Validation-to-test comparison showed only small changes in ranking metrics: PR-AUC decreased from 0.8355 to 0.8328 and ROC-AUC from 0.7961 to 0.7721. The Brier score increased from 0.2024 to 0.2248, whereas ECE decreased from 0.2358 to 0.1350. Ranking therefore remained stable, but probabilistic quality must be monitored separately and cannot be inferred from ROC or PR curves alone (Graph 8).

For external evaluation, the fixed G12 pipeline was applied to G09 without retraining, changing the 19 features or selecting a new threshold. The external sample contains 238 observations, including 114 positives. On the full sample, PR-AUC was 0.7722, ROC-AUC 0.7663, the Brier score 0.2167 and ECE 0.1128 (Graph 9).

When the primary threshold $\tau^* = 0.4446$ was transferred to the full G09 sample, 59 events were identified, 21 false alarms were generated and 55 events were missed; precision was 0.7375, recall 0.5175 and FPR 0.1694. In the pre-event subset with current IC ≥ 0.90, recall was 0.2763 and FPR 0.1271.



Graph 8 – Comparison of ranking and probabilistic metrics on G12 validation and test samples



Graph 9 – ROC and PR curves for external transfer of the final model G12 → G09

The results show that overall ranking ability is retained on another well, but the operating point does not fully preserve the FPR constraint on the complete external sample. Transfer to a new object must therefore include checks of data composition, feature distributions, calibration and false-alarm costs; local recalibration or threshold reselection may be required. G09 data were used only for external evaluation and did not participate in development of the G12 model. This preserves independence of the test and distinguishes transfer of the fixed scoring function from transfer of the general methodology. Universality of an unchanged model for all wells is not claimed.

A daily operational pipeline was developed: load and validate new data; calculate the 19 features in a fixed order; apply median imputation and robust scaling using training parameters; obtain the L1-model score; apply Platt calibration; compare probability with $\tau^* = 0.4446$; and record the warning in an alarm log (Table 4).

A warning is not a final diagnosis. The engineer checks current and previous IC, rates, pressure, choke, water cut, gas ratios, missing values and recent operations, and then decides whether to continue observation, perform additional investigation, adjust the regime or plan remedial action.

To reduce isolated and repeated signals, operational logic may require confirmation in n of m consecutive calculations, hysteresis between alarm activation and clearance, and a cooldown period before repeated notification. These rules are tuned on a separate validation period and do not change the target-event definition.

Monitoring includes PR-AUC, ROC-AUC, Brier score, ECE, pre-event recall and FPR, alarm rate and input-feature distributions. Calibration deterioration triggers recalibration; loss of ranking ability triggers renewed temporal evaluation and retraining while retaining the horizon, target definition and causal structure.

Table 4 – Procedure for daily use of the forecast

Stage	Main content	Result
Data control	Completeness, units, dates, missing values and regime changes	Validated time series
Feature calculation	Fixed set of 19 causally admissible features	Feature vector
Forecast	L1 logistic model and Platt calibration	14-day risk probability
Threshold and alarm	Comparison with $\tau^* = 0.4446$ and alarm logging	Signal for engineering review
Engineering decision	Review of operating and diagnostic data	Observation, investigation or intervention

MAIN CONCLUSIONS AND RECOMMENDATIONS

1. The study has shown that colmatation of the near-wellbore zone is one of the factors that can impair the filtration connection between the reservoir and the well and reduce production efficiency. Therefore, timely control of the well condition requires not only subsequent diagnostics, but also an early assessment of possible risk [1, 6].

2. To quantitatively assess changes in well performance, the colmatation index IC = Qactual/Qtheoretical was used. It was established that this indicator can be applied as an engineering indicator of the deviation of actual productivity from the calculated potential; however, it should not be regarded as the sole and direct evidence of colmatation [6, 10].

3. A target variable for early risk prediction was formed: a risk event was defined as a decrease in IC below 0.90 within the following 14 days. This forecast horizon was selected as a practically convenient interval that makes it possible to draw attention in advance to a potential deterioration in the well condition [10].

4. Based on production time series, a set of features was formed, including oil, gas, and water flow rates, water cut, gas ratios, calculated potential, current and previous IC values, as well as indicators characterizing changes in the well operating regime. When constructing the features, special attention was paid to ensuring that no future information entered the model [3, 8, 10].

5. Based on the results of model comparison, logistic regression with L1 regularization and probability calibration was selected for practical application. In the test interval of well G12, the model demonstrated its ability to identify periods of increased risk and rank observations according to the degree of unfavorable conditions [8, 10].

6. It was established that the dynamics of the colmatation index have the greatest influence on the forecast; however, the operating parameters of the well are also important for the engineering interpretation of the result. This indicates that the forecast should be based not on a single indicator, but on a combination of field features [6, 10].

7. Testing on well G09 showed that the proposed approach can be applied to another well; however, such transfer requires additional verification of data quality, the risk threshold, and model calibration. Therefore, universal application of the model without adaptation is not recommended [10].

8. A procedure for the practical use of the forecast in the monitoring system was developed: regular data updating, feature calculation, risk probability estimation, recording of warnings, and mandatory engineering verification of the obtained signal [8, 10].

9. It is recommended to use the proposed methodology as a tool for supporting engineering decision-making: for ranking wells by risk level, selecting objects for additional diagnostics, analyzing the operating regime, and planning preventive measures [3, 8, 10].

THE MAIN RESULTS OF THE DISSERTATION WERE PUBLISHED IN THE FOLLOWING WORKS

1. Абдулмуталибов, Т.Э. Исследование механизмов повреждения пласта в песчанике // Материалы Республиканской научной конференции молодых исследователей и докторантов, посвящённой 100-летию со дня рождения Гейдара Алиева. Баку, 4–5 мая 2023. С. 1–5.

2. Абдулмуталибов, Т.Э., Шмончева, Е.Е. Уменьшение повреждения пласта за счёт проектирования и выбора буровых растворов // Нефтегазовое дело, техносферная безопасность, рациональное природопользование: современные реалии: материалы VI Всероссийской научно-практической конференции. Махачкала: ДГТУ, 16–17 мая 2023.

3. Абдулмуталибов, Т.Э., Шмончева, Е.Е., Джаббарова, Г.В. Возможность использования больших данных и машинного обучения для повышения эффективности бурения // Гейдар Алиев и нефтяная стратегия Азербайджана: достижения нефтегазовой геологии и геотехнологий. Баку, 23–26 мая 2023. С. 1002–1007.

4. Абдулмуталибов, Т.Э., Джаббарова, Г.В. Нанотехнология в буровых растворах // Булатовские чтения: материалы VII Международной научно-практической конференции. Краснодар: Издательский Дом – Юг, 2023. Т. 1. С. 328–329.

5. Абдулмуталибов, Т.Э., Шмончева, Е.Е. Экспериментальное исследование эрозии образцов соленосных пород // Современные технологии в нефтегазовом деле – 2023: материалы международной научно-технической конференции. Уфа: УГНТУ, 2023. С. 542–545.

6. Абдулмуталибов, Т.Э., Джаббарова, Г.В. Оценка начального давления, проницаемости и скин-фактора пласта с использованием методов машинного обучения // Технологии разработки месторождений и моделирование процессов в нефтегазодобыче. Уфа: УГНТУ, 2023. С. 224–225.

7. Abdulmutalibov, T.E., Shmoncheva, Y.Y., Jabbarova, G.V. Drilling fluids in complicated conditions: a review. Nafta-Gaz. 2023. No. 10. P. 660–669. DOI: 10.18668/NG.2023.10.03.

8. Abdulmutalibov, T.E., Jabbarova, G.V., Shmoncheva, Y.Y. Advancements in applications of machine learning for formation damage predictions. SPE Caspian Technical Conference and Exhibition. Baku, 21–23 November 2023. DOI: 10.2118/217610-MS.

9. Абдулмуталибов, Т.Э., Салаев, М.Т., Шмончева, Е.Е., Джаббарова, Г.В., Кузнецов, В.А. Превентор универсальный. Евразийский патент № 045267B1. 2023.

10. Abdulmutalibov, T.E., Malikov, H.Kh., Jabbarova, G.V., Tashmenov, R.R. Forecasting the risk of formation damage (colmatation) with a 14-day horizon using calibrated probabilistic models. Journal of SRI "Geotechnological Problems of Oil, Gas and Chemistry". 2025. Issue 2. P. 106–120.

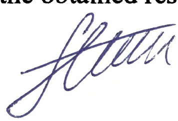
11. Abdulmutalibov, T.E., Jabbarova, G.V., Shmoncheva, Y.Y. Universal BOP with built-in grinding tool. Nafta-Gaz. 2025. No. 7. P. 474–481. DOI: 10.18668/NG.2025.07.06.

Personal contribution of the applicant to the published works:

Work [1] was completed independently by the applicant, without co-authors.

In works [2, 3, 5, 7, 8, 9, 11], the contribution of the authors was equal.

In works [4, 6, 10], the applicant's contribution consisted in formulating the research problem, participating in the research, and summarizing and analyzing the obtained results.



The defense will be held at 11:00 on September 29, 2026, at the meeting of the Dissertation Council FD 2.03 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and Industry University.

Address: AZ 1010, Baku city, Azadliq Avenue, 34

Dissertation is accessible at the Azerbaijan State Oil and Industry University Library.

Electronic versions of dissertation and its abstract are available on the official website of the Azerbaijan State Oil and Industry University.

Abstract was sent to the required addresses on August "28" 2026.

Signed to print: 27.08.2026
Paper format: A5
Volume: 36113
Number of hard copies: 20