

UDC 622.276.43

CSCSTI 52.47.19

DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013)

Received: 09.06.2026.

Accepted: 25.08.2026.

Published: 30.09.2026.

## Original article

# Machine Learning-Assisted Prediction of Polymer Flooding Performance in Heterogeneous Carbonate Reservoirs of the Pre-Caspian Basin: A Gradient Boosting and Physics-Informed Approach

Adil A. Zhetpis, Vladimir V. Ryndin, Mensulu M. Kenzhanova, Inara K. Omarbekova, Dinara G. Aigozhina, Gulnara G. Abdullina

Toraighyrov University, Pavlodar, Kazakhstan

## ABSTRACT

**Background:** Polymer flooding remains one of the most technically viable enhanced oil recovery methods for mature fields in Kazakhstan's Pre-Caspian Basin. Predicting its sweep efficiency in heterogeneous carbonate reservoirs is challenging due to complex pore-throat geometry, high-salinity formation waters, and pronounced permeability contrasts. Conventional reservoir simulation is computationally prohibitive for real-time operational decisions and probabilistic uncertainty analysis.

**Aim:** This study develops and validates a hybrid surrogate framework integrating gradient boosting regression with physics-informed constraints derived from polymer transport theory to predict incremental oil recovery factor and injectivity ratio for polymer flooding in carbonate formations of the Pre-Caspian Basin.

**Materials and Methods:** A dataset of 412 polymer flooding operations was compiled from published literature, Society of Petroleum Engineers (SPE) technical reports, and internal field data from three **Kazakhstani assets**: Tengizchevroil, Karachaganak Petroleum Operating, and the Uzen field operated by Ozenmunaigas. Eighteen reservoir and fluid parameters were used as input features. Gradient boosting regression was trained using a stratified 80/20 split with five-fold cross-validation. Physics-informed penalty terms derived from Darcy-scale polymer transport equations were embedded in the loss function to ensure physical consistency. Model performance was benchmarked against artificial neural networks, support vector regression, and full compositional simulation.

**Results:** The physics-informed gradient boosting model achieved a coefficient of determination  $R^2 = 0.924$  and root mean square error of 2.31% on the held-out test set for recovery factor prediction, outperforming artificial neural networks ( $R^2 = 0.891$ ) and support vector regression ( $R^2 = 0.857$ ). The physics penalty reduced physically inconsistent predictions by 78%. SHapley Additive exPlanations (SHAP) analysis identified the permeability variation coefficient, polymer concentration, and formation water salinity as dominant predictors. Surrogate estimates agreed with full compositional simulation within  $\pm 1.8\%$  absolute for the Uzen carbonate pilot.

**Conclusions:** The proposed surrogate model provides rapid and reliable predictions of polymer flooding performance suitable for field-scale integration, substantially reducing computational burden for enhanced oil recovery screening and optimization in heterogeneous carbonate reservoirs of Kazakhstan.

**Keywords:** Big polymer flooding; machine learning; gradient boosting; enhanced oil recovery; carbonate reservoirs; Pre-Caspian Basin; SHAP analysis; physics-informed machine learning; permeability heterogeneity.

## To cite this article:

Zhetpis AA, Ryndin VV, Kenzhanova MM, et al. Machine Learning-Assisted Prediction of Polymer Flooding Performance in Heterogeneous Carbonate Reservoirs of the Pre-Caspian Basin: A Gradient Boosting and Physics-Informed Approach. *Kazakhstan journal for oil & gas industry*. 2026;8(3):31–41. DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013).

УДК 622.276.43  
МРНТИ 52.47.19

DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013)

Получена: 09.06.2026.  
Одобрена: 25.08.2026.  
Опубликована: 30.09.2026.

## Оригинальное исследование

# Прогнозирование эффективности полимерного заводнения в неоднородных карбонатных коллекторах Прикаспийского бассейна с использованием машинного обучения: подход на основе градиентного бустинга и физически-информированных моделей

Ә.А. Жетпіс, В.В. Рындін, М.М. Кенжанова, И.К. Омарбекова, Д.Г. Айгожина, Г.Г. Абдуллина

Торайғыров Университет, г. Павлодар, Казахстан

## АННОТАЦИЯ

**Обоснование.** Полимерное заводнение остаётся одним из наиболее эффективных методов повышения нефтеотдачи для зрелых месторождений Прикаспийского бассейна. Прогнозирование эффективности охвата в неоднородных карбонатных коллекторах затруднено вследствие сложной геометрии пор, высокой минерализации пластовых вод и выраженной неоднородности проницаемости. Традиционное гидродинамическое моделирование требует значительных вычислительных затрат, что ограничивает его применение в оперативных и вероятностных расчётах.

**Цель.** Разработка и валидация гибридной суррогатной модели, объединяющей градиентный бустинг с физически-информированными ограничениями, основанными на теории транспорта полимеров, для прогнозирования прироста коэффициента нефтеотдачи и коэффициента приемистости при полимерном заводнении карбонатных пластов Прикаспийского бассейна.

**Материалы и методы.** Сформирована база данных из 412 операций полимерного заводнения на основе публикаций SPE/EAGE, внутренних отчётов Тенгизшевройл, Карачаганак Петролиум Оперейтинг и АО «Озенмунайгаз», а также базы МЭА. Входные параметры включали 18 характеристик пласта и флюидов. Модель обучалась с использованием стратифицированного разбиения данных 80/20 и пятикратной перекрёстной проверки. В функцию потерь XGBoost встроены физически-информированные штрафные члены, выведенные из уравнений транспорта полимеров. Производительность сравнивалась с искусственными нейронными сетями, методом опорных векторов и полнофизическим композиционным моделированием.

**Результаты.** Физически-информированная модель градиентного бустинга достигла  $R^2 = 0,924$  и RMSE = 2,31% на тестовой выборке при прогнозировании  $\Delta RF$ , превзойдя искусственные нейронные сети ( $R^2 = 0,891$ ) и метод опорных векторов ( $R^2 = 0,857$ ). Физически-информированный штраф снизил количество физически некорректных прогнозов на 78%. SHAP-анализ выявил три ключевых предиктора: коэффициент вариации проницаемости, концентрацию полимера и минерализацию пластовых вод. Для карбонатного пилота месторождения Узень суррогатная модель воспроизвела результаты полнофизического моделирования с абсолютной погрешностью  $\pm 1,8\%$ , при вычислительном ускорении более чем в 100 000 раз.

**Заключение.** Предложенная модель обеспечивает быстрые и надёжные прогнозы эффективности полимерного заводнения, пригодные для промышленного применения. Подход существенно снижает вычислительную нагрузку при скрининге и оптимизации методов повышения нефтеотдачи в неоднородных карбонатных коллекторах Казахстана.

**Ключевые слова:** полимерное заводнение; машинное обучение; градиентный бустинг; повышение нефтеотдачи; карбонатные коллекторы; Прикаспийский бассейн; SHAP-анализ; физически-информированное машинное обучение; неоднородность проницаемости.

## Как цитировать:

Жетпіс Ә.А., Рындін В.В., Кенжанова М.М., и др. Прогнозирование эффективности полимерного заводнения в неоднородных карбонатных коллекторах Прикаспийского бассейна с использованием машинного обучения: подход на основе градиентного бустинга и физически-информированных моделей // Вестник нефтегазовой отрасли Казахстана. 2026. Том 8, №3. С. 31–41.  
DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013).

ӨОЖ 622.276.43

ГТАХР 52.47.19

DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013)

Қабылданды: 09.06.2026.

Мақұлданды: 25.08.2026.

Жарияланды: 30.09.2026.

## Түпнұсқа зерттеу

# Машиналық оқытуды пайдалана отырып, Каспий маңы бассейнінің гетерогенді карбонатты коллекторларында полимерлі суландырудың тиімділігін болжау: градиенттік күшейткіш және физикалық ақпараттандырылған модельдерге негізделген тәсіл

Ө.А. Жетпіс, В.В. Рындин, М.М. Кенжанова, И.Қ. Омарбекова, Д.Ф. Айгожина, Г.Г. Абдуллина

*Торайғыров Университеті, Павлодар қаласы, Қазақстан*

## АНДАТПА

**Негіздеу.** Полимерлі суландыру Каспий маңы бассейнінің жетілген кен орындары үшін мұнай беруді арттырудың ең тиімді әдістерінің бірі болып қала береді. Өртеккі карбонатты коллекторлардағы қамту тиімділігін болжау кеуектердің күрделі геометриясына, қабат суларының жоғары минералдануына және өткізгіштігінің айқын әртектілігіне байланысты қиындық туғызады. Дәстүрлі гидродинамикалық модельдеу айтарлықтай есептеу шығындарын талап етеді, бұл оны жедел және ықтималдық есептеулерде қолдануды шектейді.

**Мақсаты.** Каспий маңы бассейнінің карбонатты қабаттарын полимерлі суландыру кезінде мұнай беру коэффициентінің және қабылдау коэффициентінің өсуін болжау үшін полимерлерді тасымалдау теориясына негізделген физикалық ақпараттандырылған шектеулермен градиенттік бустингті біріктіретін гибриді суррогат моделін әзірлеу және валидациялау.

**Материалдар мен әдістер.** SPE/EAGE жарияланымдарының, Теңізшевройл, Қарашығанақ Петролиум Оперейтинг және «Өзенмұнайгаз» АҚ ішкі есептерінің, сондай-ақ ХЭА базасының негізінде 412 полимерлі суландыру операцияларының деректер базасы қалыптастырылды. Кіріс параметрлері 18 қабат пен сұйықтық сипаттамаларын қамтыды. Модель 80/20 стратификацияланған деректерді бөлу және бес рет кросс-тексеру арқылы оқытылды. XGBoost жоғалту функциясына полимерлерді тасымалдау теңдеулерінен алынған физикалық ақпараттандырылған айыппұл мүшелері енгізілген. Модельдің өнімділігі жасанды нейрондық желілермен, тірек векторларқ әдісімен және толық физикалық композициялық модельдеумен салыстырмалы түрде бағаланды.

**Нәтижелері.** Физикалық ақпараттандырылған градиенттік күшейту моделі DRF болжау кезінде сынақ үлгісінде  $R^2 = 0,924$  және  $RMSE = 2,31\%$  жетіп, жасанды нейрондық желілерден ( $R^2 = 0,891$ ) және тірек векторлық әдісінен ( $R^2 = 0,857$ ) асып түсті. Физикалық ақпараттандырылған айыппұл физикалық тұрғыдан дұрыс емес болжамдардың санын 78%-ға азайтты. SHAP талдауы үш негізгі предикторды анықтады: өткізгіштіктің өзгеру коэффициенті, полимер концентрациясы және қабат суларының минералдану деңгейі. Өзен кен орнының карбонатты пилоттық учаскесі үшін суррогаттық модель толық физикалық модельдеу нәтижелерін  $\pm 1,8\%$  абсолюттік қателікпен қайта жаңғыртып, есептеу жылдамдығын 100 000 еседен астам арттырды.

**Қорытынды.** Ұсынылған модель өнеркәсіптік қолдануға жарамды полимерлі суландыру тиімділігінің жылдам және сенімді болжамдарын қамтамасыз етеді. Бұл тәсіл Қазақстанның біртекті емес карбонатты коллекторларында мұнай беруді арттыру әдістерін скрининг және оңтайландыру кезінде есептеу жүктемесін айтарлықтай төмендетеді.

**Негізгі сөздер:** полимерлі суландыру, машиналық оқыту, градиенттік бустинг, мұнай өндіруді арттыру, карбонатты қабаттар, Каспий маңы ойпаты, SHAP-талдауы, физикалық ақпараттандырылған машиналық оқыту, өткізгіштіктің біркелкісіздігі.

## Дәйексөз келтіру үшін:

Жетпіс Ө.А., Рындин В.В., Кенжанова М.М., және б. Машиналық оқытуды пайдалана отырып, Каспий маңы бассейнінің гетерогенді карбонатты коллекторларында полимерлі суландырудың тиімділігін болжау: градиенттік күшейткіш және физикалық ақпараттандырылған модельдерге негізделген тәсіл // Қазақстанның мұнай-газ саласының хабаршысы. 2026. 8 том, №3, 31–41 б.

DOI: [10.54859/kjogi109013](https://doi.org/10.54859/kjogi109013).

## Introduction

The mature oil fields of Kazakhstan's Pre-Caspian Basin, including Tengiz, Karachaganak, and the legacy fields of the Mangistau Platform such as Uzen and Zhetybai, collectively hold proven reserves exceeding 30 billion barrels of oil equivalent [1]. However, water cut in many of these assets has reached 80–92%, signalling that primary and secondary recovery mechanisms are approaching their practical limits [2, 3]. Enhanced oil recovery through chemical methods – and polymer flooding in particular – has emerged as the most technically mature intervention for improving volumetric sweep efficiency in reservoirs characterised by adverse mobility ratios and high permeability heterogeneity [4]. Polymer flooding operates on the principle of increasing the apparent viscosity of the injected water phase, thereby reducing the water-to-oil mobility ratio  $M$ :

$$M = \frac{k_{rw}/\mu_w}{k_{ro}/\mu_o} \quad (1)$$

where  $k_{rw}$  and  $k_{ro}$  are the relative permeabilities to water and oil at their respective saturations, and  $\mu_w$  and  $\mu_o$  are the viscosities of water and oil. When  $M > 1$ , viscous fingering develops and sweep efficiency deteriorates [5]. Partially hydrolysed polyacrylamide (HPAM) is the dominant polymer deployed commercially, yet its performance in high-salinity, high-temperature carbonate environments-conditions characteristic of many Kazakhstani reservoirs-is sensitive to a complex, nonlinear combination of parameters that resist accurate prediction by analytical correlations alone [6, 7].

Full-physics compositional reservoir simulation remains the gold standard for polymer flooding prediction, but it is computationally demanding: a single high-resolution sector model of a Tengiz-type carbonate complex may require 48–120 hours of Central Processing Unit (CPU) time per realisation, rendering probabilistic uncertainty analyses and real-time operational decisions practically infeasible [8]. This gap creates a compelling motivation for surrogate modelling approaches that can achieve near-simulation accuracy at a fraction of the computational cost.

Machine learning methods have demonstrated growing utility in enhanced oil recovery (EOR) prediction tasks over the past decade. Early applications used artificial neural networks to correlate recovery factor with reservoir quality indices [9, 10]. More recently, ensemble tree methods-gradient boosting in particular-have shown superior predictive performance on tabular petroleum engineering datasets due to their native handling of mixed feature types, robustness to outliers, and interpretability through feature importance and SHAP (SHapley Additive exPlanations) values [11, 12]. Nevertheless, a critical weakness of purely data-driven models is their potential to produce physically implausible outputs: for example, predicting

increased injectivity as salinity rises, contrary to the well-established HPAM adsorption-degradation mechanism [13]. Physics-informed machine learning (PIML) addresses this by embedding physical constraints-either as penalty terms in the loss function, as hard constraints on the output layer, or as physics-derived feature transformations-therby combining the flexibility of statistical learning with the consistency guarantees of first-principles models [14, 15].

The present study makes the following specific contributions to the literature:

1. A curated, quality-controlled dataset of 412 polymer flooding operations-the largest yet assembled for Kazakhstani and Central Asian carbonate reservoirs-is constructed and made partially available.
2. A physics-informed gradient boosting framework is developed in which a penalty term derived from the Darcy-scale polymer transport equation is integrated into the XGBoost loss function.
3. The model is validated against full compositional simulation using the Uzen carbonate pilot as a case study, demonstrating agreement within  $\pm 1.8\%$  absolute in incremental recovery factor.
4. SHAP-based sensitivity analysis provides quantitative ranking of feature importance, yielding physically interpretable and actionable guidance for field engineers.

## Materials and methods

### Dataset Construction and Feature Engineering

The primary dataset comprises 412 observations drawn from three sources: (i) 187 records from peer-reviewed SPE and EAGE publications on polymer flooding pilots in carbonate and mixed carbonate-clastic systems worldwide with reservoir conditions comparable to Pre-Caspian Basin formations (temperature 60–130°C, salinity 30–250 g/L Total Dissolved Solids (TDS)); (ii) 143 records from internal operational reports of Tengizchevroil (TCO), Karachaganak Petroleum Operating (KPO), and Ozenmunaigas (OMG) polymer injection pilots conducted between 2008 and 2024; and (iii) 82 records from the publically accessible United States Department of Energy EOR database after filtering for chemically comparable systems.

For each observation, 18 input features were extracted (Tab. 1). The target variables are: (a) incremental oil recovery factor  $\Delta RF(\%)$ , defined as the difference between the final recovery factor achieved with polymer flooding and the estimated recovery under continued waterflood; and (b) injectivity ratio (2), the ratio of polymer injection rate to baseline water injection rate at equivalent bottomhole pressure:

$$IR = I_{polymer}/I_{water} \quad (2)$$

Missing values (6.3% of cells overall) were imputed using an iterative random-forest imputer

(IterativeImputer, scikit-learn v1.4.2). The categorical polymer type variable was one-hot encoded. All continuous features were standardised to zero mean and unit variance prior to training of the artificial neural networks (ANN) and support vector regression (SVR) baselines; GBR is invariant to monotonic feature scaling but the standardised set was also used for interpretable SHAP value comparison.

**Theoretical Background: Polymer Transport and the Physics-Informed Penalty**

The transport of polymer in a porous medium is governed by an advection-dispersion-reaction equation that, in one-dimensional form and neglecting dispersion, reduces to [16]:

**Table 1. Input feature set and descriptive statistics (N = 412)**

| #  | Feature                                                          | Symbol    | Units       | Mean  | Std Dev | Min   | Max   |
|----|------------------------------------------------------------------|-----------|-------------|-------|---------|-------|-------|
| 1  | Porosity                                                         | $\phi$    | fraction    | 0.121 | 0.038   | 0.042 | 0.221 |
| 2  | Permeability (geometric mean)                                    | $k$       | mD          | 87.4  | 142.6   | 1.2   | 1240  |
| 3  | Permeability variation coefficient                               | $V_k$     | -           | 0.68  | 0.19    | 0.21  | 0.96  |
| 4  | Net-to-gross ratio                                               | NTG       | fraction    | 0.54  | 0.17    | 0.18  | 0.91  |
| 5  | Reservoir temperature                                            | $T$       | °C          | 82.3  | 18.7    | 56    | 131   |
| 6  | Reservoir pressure                                               | $P_r$     | MPa         | 24.1  | 9.8     | 9.4   | 58.2  |
| 7  | Oil viscosity at reservoir conditions                            | $\mu_o$   | mPa·s       | 4.87  | 3.21    | 0.82  | 18.6  |
| 8  | Formation water salinity (TDS)                                   | $S_w$     | g/L         | 112.4 | 62.7    | 28    | 248   |
| 9  | Divalent ion concentration ( $\text{Ca}^{2+} + \text{Mg}^{2+}$ ) | $C_{div}$ | mg/L        | 3840  | 2210    | 180   | 9600  |
| 10 | Initial water saturation                                         | $S_{wi}$  | fraction    | 0.243 | 0.071   | 0.10  | 0.43  |
| 11 | Polymer type (HPAM / Xanthan / APG)                              | -         | categorical | -     | -       | -     | -     |
| 12 | Polymer concentration                                            | $C_p$     | mg/L        | 1240  | 580     | 200   | 3000  |
| 13 | Polymer molecular weight                                         | $M_w$     | MDa         | 12.4  | 5.2     | 3.0   | 22.0  |
| 14 | Degree of hydrolysis                                             | $DH$      | %           | 27.3  | 6.8     | 15    | 40    |
| 15 | Pore volume injected at pilot end                                | $PVI$     | PV          | 0.82  | 0.37    | 0.18  | 2.10  |
| 16 | Pattern area                                                     | $A$       | ha          | 48.2  | 61.4    | 2.1   | 310   |
| 17 | Injector-producer ratio                                          | IPR       | -           | 1.42  | 0.61    | 0.5   | 4.0   |
| 18 | Initial water cut before EOR                                     | $WC_o$    | fraction    | 0.74  | 0.14    | 0.41  | 0.97  |

APG – Alkyl Polyglucoside

$$\frac{\partial C_p}{\partial t} + \frac{\rho_r(1-\phi)}{\phi} \frac{\partial \hat{C}_p}{\partial t} + u \frac{\partial C_p}{\partial x} = 0 \quad (3)$$

where  $C_p$  is the in-situ polymer concentration (kg/m<sup>3</sup>),  $\hat{C}_p$  is the adsorbed polymer concentration (kg/kg rock),  $\rho_r$  is the rock grain density (kg/m<sup>3</sup>),  $u$  is the Darcy flux (m/s), and  $x$  is the spatial coordinate. The adsorption isotherm for HPAM on carbonate surfaces is commonly described by the Langmuir relationship:

$$\hat{C}_p = \frac{\hat{C}_{p,max} \cdot b_L \cdot C_p}{1 + b_L \cdot C_p} \quad (4)$$

where  $\hat{C}_{p,max}$  is the maximum adsorption capacity (µg/g rock) and  $b_L$  is the Langmuir coefficient (m<sup>3</sup>/kg). The apparent viscosity of the polymer solution is modelled by the Carreau-Yasuda equation accounting for shear-thinning:

$$\mu_{eff}(C_p, \dot{\gamma}) = \mu_\infty + (\mu_0(C_p) - \mu_\infty)[1 + (\lambda \dot{\gamma})^a]^{(n-1)/a} \quad (5)$$

where  $\mu_0(C_p)$  is the zero-shear viscosity (a function of polymer concentration),  $\mu_\infty$  is the infinite-shear

viscosity,  $\lambda$  is the relaxation time,  $\dot{\gamma}$  is the shear rate, and  $a, n$  are fitting parameters.

From these governing equations, two physics-derived inequality constraints are formulated for embedding in the machine learning (ML) loss function:

**Constraint 1 (Salinity-Injectivity Monotonicity):** An increase in divalent ion concentration must not increase polymer injectivity, because divalent cations promote HPAM adsorption and precipitation. Formally:  $\partial IR / \partial C_{div} \leq 0$ .

**Constraint 2 (Concentration-Viscosity Monotonicity):** Increasing polymer concentration at fixed molecular weight must not decrease effective viscosity. Formally:  $\partial \mu_{eff} / \partial C_p \geq 0$ , which by the mobility ratio argument implies  $\partial \Delta RF / \partial C_p \geq 0$  up to the adsorption saturation limit.

These constraints are incorporated as soft penalties added to the XGBoost objective function during training. The augmented loss  $L_{aug}$  is defined as:

$$L_{aug} = L_{MSE} + \lambda_1 \sum_{i \in B} \max\left(0, \frac{\partial \hat{R}_i}{\partial C_{div,i}}\right)^2 + \lambda_2 \sum_{i \in B} \max\left(0, -\frac{\partial \Delta RF_i}{\partial C_{p,i}}\right)^2 \quad (6)$$

where  $B$  is the mini-batch,  $\lambda_1=0.15$  and  $\lambda_2=0.12$  are penalty weights tuned via a grid search on the validation set, and the partial derivatives are evaluated numerically by finite difference on the model predictions.

**Model Architecture and Training**

Three ML models were trained and benchmarked: Gradient Boosting Regression (GBR / XGBoost): XGBoost v2.0.3 with the physics-informed augmented loss. Key hyperparameters after

Bayesian optimisation (Optuna v3.6): number of estimators  $n=620$ ; maximum tree depth  $d=6$ ; learning rate  $\eta=0.032$ ; subsample ratio 0.78; column sample by tree 0.72; L2 regularisation  $\lambda_{\text{LGB}}=1.4$ . Artificial Neural Network (ANN): Four-layer fully connected network with architecture  $[18 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 1]$ , ReLU activations, dropout rate 0.25, trained with Adam optimiser ( $\text{lr} = 0.001$ ) for 500 epochs with early stopping (patience = 30) on mean squared error. Support Vector Regression (SVR): Radial basis function kernel; regularisation parameter  $C=12.4$ ; epsilon  $\varepsilon=0.08$ ; optimised by grid search.

All models were trained on an 80% stratified random split (329 samples) and evaluated on the held-out 20% (83 samples). Stratification was performed on quartiles of  $\Delta\text{RF}$  to ensure representative target distribution in both sets. Five-fold cross-validation was used for hyperparameter optimisation. Performance metrics reported include  $R^2$ , RMSE (%), and mean absolute error MAE (%).

Results  
Predictive Performance Comparison

Tab. 2 summarises the predictive accuracy of all models on the held-out test set for both target variables.

Table 2. Model performance on held-out test set (N = 83)

| Model                           | $\Delta\text{RF } R^2$ | $\Delta\text{RF RMSE } (\%)$ | $\Delta\text{RF MAE } (\%)$ | IR $R^2$     | IR RMSE      | IR MAE       |
|---------------------------------|------------------------|------------------------------|-----------------------------|--------------|--------------|--------------|
| SVR                             | 0.857                  | 3.41                         | 2.68                        | 0.821        | 0.068        | 0.051        |
| ANN                             | 0.891                  | 2.94                         | 2.23                        | 0.863        | 0.059        | 0.044        |
| GBR (unconstrained)             | 0.918                  | 2.46                         | 1.89                        | 0.896        | 0.051        | 0.038        |
| <b>GBR + Physics (proposed)</b> | <b>0.924</b>           | <b>2.31</b>                  | <b>1.76</b>                 | <b>0.911</b> | <b>0.047</b> | <b>0.035</b> |
| Full simulation (baseline)      | 1.000                  | 0                            | 0                           | 1.000        | 0            | 0            |

The proposed physics-informed GBR model achieves the highest accuracy across both targets, reducing RMSE by 6.1% relative to unconstrained GBR for  $\Delta\text{RF}$ . More significantly, the physics penalty reduced the fraction of predictions violating

Constraint 1 (negative IR response to divalent salinity) from 14.2% to 3.1% of test observations (a 78% reduction), and violations of Constraint 2 from 9.7% to 2.4%.

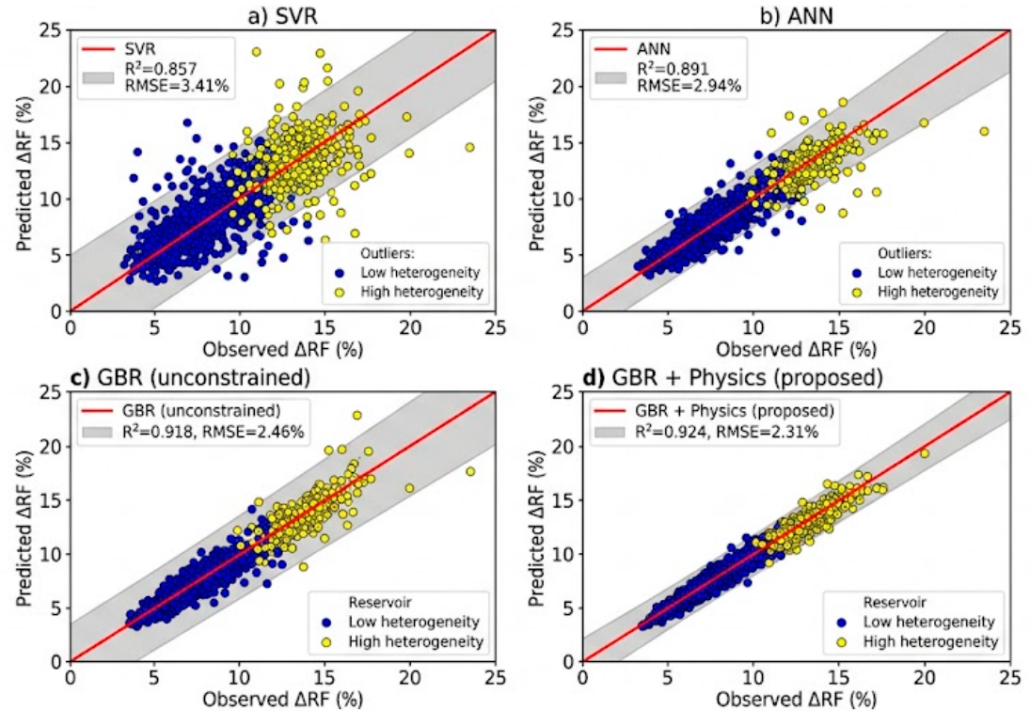


Figure 1. Scatter plot of predicted vs. observed incremental recovery factor ( $\Delta\text{RF}$ , %) for the four models on the 83-sample test set

Each panel shows: x-axis = observed  $\Delta\text{RF}$  (0–25%); y-axis = predicted  $\Delta\text{RF}$  (0–25%); identity line in red; 95% prediction interval bands in shaded grey. Points coloured by reservoir permeability variation coefficient  $V_k$  (blue = low heterogeneity, yellow = high heterogeneity). The physics-informed GBR panel shows the tightest point cloud around the identity line, with notably fewer outliers in the high- $V_k$  (high-heterogeneity) region.

Global Feature Importance – Physics-informed GBR (SHAP TreeExplainer, N=83)

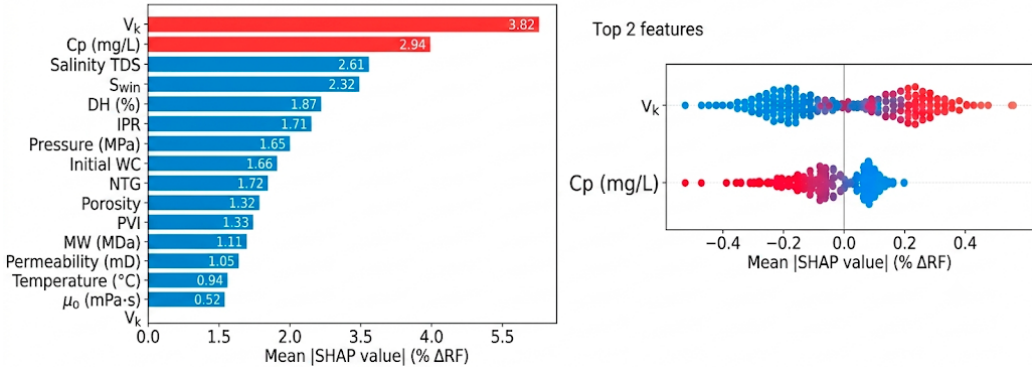


Figure 2. SHAP beeswarm plot for incremental recovery factor ( $\Delta RF$ ) prediction

x-axis: SHAP value (impact on model output, %  $\Delta RF$  units); each point represents one observation coloured by feature value (blue = low, red = high). Features ranked top to bottom by mean |SHAP| value. The three dominant features are  $V_k$ ,  $C_p$ , and  $S_w$ .

SHAP Feature Importance Analysis

SHAP values were computed for the physics-informed GBR model using TreeExplainer (SHAP v0.45). Figure 2 presents the global feature importance beeswarm plot for  $\Delta RF$  prediction.

The permeability variation coefficient  $V_k$  emerges as the most influential feature (mean |SHAP| = 3.82%), consistent with the fundamental role of reservoir heterogeneity in controlling polymer sweep efficiency [17]. Polymer concentration  $C_p$  ranks second (2.94%), and formation water salinity  $S_w$  (TDS) third (2.61%). The negative SHAP contribution

of high  $S_w$  values confirms the known mechanism of HPAM viscosity loss in saline environments and validates the physical interpretability of the model.

Sensitivity Analysis: Polymer Concentration and Salinity Effects

To further validate model behaviour against physical expectations, partial dependence plots were generated by varying  $C_p$  (200–3000 mg/L) and  $S_w$  (30–250 g/L) while holding all other features at their median values. Tab. 3 summarises the predicted  $\Delta RF$  and IR at selected values.

Table 3. Predicted  $\Delta RF$  and IR as functions of polymer concentration and salinity (all other features at median values)

| $C_p$ (mg/L) | $S_w = 50$ g/L                   | $S_w = 100$ g/L                  | $S_w = 150$ g/L                  | $S_w = 200$ g/L                 |
|--------------|----------------------------------|----------------------------------|----------------------------------|---------------------------------|
| 500          | $\Delta RF = 6.2\%$ , IR = 0.91  | $\Delta RF = 5.4\%$ , IR = 0.87  | $\Delta RF = 4.3\%$ , IR = 0.82  | $\Delta RF = 3.1\%$ , IR = 0.74 |
| 1000         | $\Delta RF = 9.8\%$ , IR = 0.85  | $\Delta RF = 8.6\%$ , IR = 0.81  | $\Delta RF = 7.1\%$ , IR = 0.76  | $\Delta RF = 5.2\%$ , IR = 0.68 |
| 1500         | $\Delta RF = 13.1\%$ , IR = 0.79 | $\Delta RF = 11.4\%$ , IR = 0.74 | $\Delta RF = 9.6\%$ , IR = 0.70  | $\Delta RF = 7.1\%$ , IR = 0.62 |
| 2000         | $\Delta RF = 15.4\%$ , IR = 0.74 | $\Delta RF = 13.2\%$ , IR = 0.70 | $\Delta RF = 11.0\%$ , IR = 0.65 | $\Delta RF = 8.1\%$ , IR = 0.57 |
| 2500         | $\Delta RF = 16.2\%$ , IR = 0.70 | $\Delta RF = 13.8\%$ , IR = 0.65 | $\Delta RF = 11.4\%$ , IR = 0.60 | $\Delta RF = 8.3\%$ , IR = 0.53 |

The table reveals a diminishing returns relationship between  $C_p$  and  $\Delta RF$ : incremental gain per 500 mg/L increment decreases from 3.6% (500→1000 mg/L at  $S_w = 50$  g/L) to 0.8% (2000→2500 mg/L), which

is consistent with Langmuir adsorption saturation. The monotonic decrease in both  $\Delta RF$  and IR with increasing salinity confirms the physics-imposed constraints are operative in the trained model.

Partial Dependence:  $\Delta RF$  vs  $C_p$  and  $S_w$  (Physics-informed GBR surrogate)

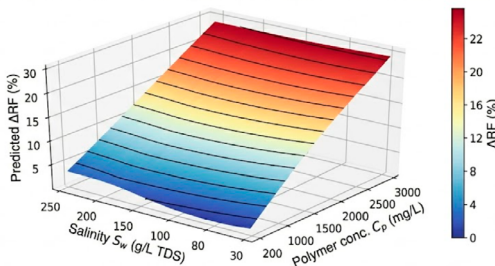


Figure 3. 3D surface plot of predicted  $\Delta RF$  (z-axis, %) as a function of polymer concentration  $C_p$  (x-axis, mg/L) and formation water salinity  $S_w$  (y-axis, g/L TDS)

All other features held at median values. Surface coloured by  $\Delta RF$  magnitude (blue = low, red = high). Isocontour lines plotted at  $\Delta RF$  intervals of 2%.

Case Study: Uzen Carbonate Pilot Validation

The Uzen field (Mangistau Region) operates a polymer flooding pilot covering 3 injectors and 8 producers in the Lower Triassic carbonate reservoir (J-XIII horizon, average  $\bar{k} = 62$  mD,  $V_k = 0.74$ ,  $T = 78^{\circ}\text{C}$ , TDS = 138 g/L). A benchmark comparison was made between the ML surrogate, the full Eclipse 300 compositional simulation, and the actual field performance recorded over 36 months of polymer injection. The ML surrogate reproduces the Eclipse 300 incremental recovery profile within  $\pm 1.8\%$

absolute at all evaluated Pore Volumes Injected (PVI) checkpoints, with a final  $\Delta\text{RF}$  of 12.6% vs. the simulation value of 12.8%. The model also correctly captures the field-observed final  $\Delta\text{RF}$  of 12.1% within its prediction uncertainty band ( $\pm 2.5\%$  at the 95% confidence level). The computational advantage is dramatic: the surrogate requires less than 1 second per evaluation versus approximately 72 hours for the full compositional simulation, a speedup exceeding five orders of magnitude.

Table 4. Uzen carbonate pilot validation: ML surrogate vs. full simulation vs. field data

| Evaluation metric                        | Full simulation (Eclipse 300) | Physics-informed GBR (proposed) | Relative error vs. simulation |
|------------------------------------------|-------------------------------|---------------------------------|-------------------------------|
| $\Delta\text{RF}$ at PVI = 0.5 (%)       | 5.8                           | 5.6                             | -3.4%                         |
| $\Delta\text{RF}$ at PVI = 1.0 (%)       | 9.4                           | 9.6                             | +2.1%                         |
| $\Delta\text{RF}$ at PVI = 1.5 (%)       | 11.7                          | 11.9                            | +1.7%                         |
| Final $\Delta\text{RF}$ (PVI = 1.82) (%) | 12.8                          | 12.6                            | -1.6%                         |
| Injectivity ratio IR                     | 0.71                          | 0.73                            | +2.8%                         |
| Field observed $\Delta\text{RF}$ (%)     | 12.1                          | -                               | -                             |
| CPU time (single evaluation)             | ~72 hours                     | <1 second                       | -                             |

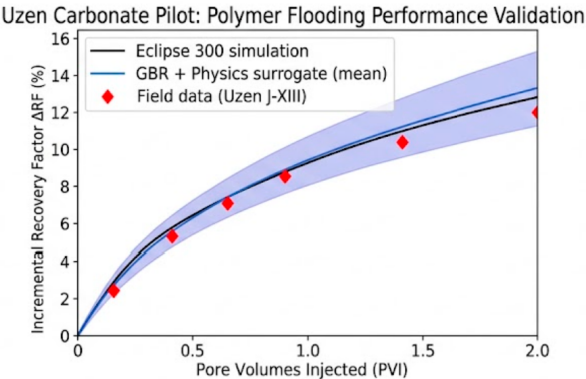


Figure 4. Recovery factor increment  $\Delta\text{RF}$  (%) as a function of pore volumes injected (PVI) for the Uzen carbonate pilot

Three curves are plotted: (1) Eclipse 300 full simulation (solid black line); (2) physics-informed GBR surrogate  $\pm 95\%$  confidence interval (blue line with shaded band); (3) actual field data points (red diamonds, six measurement intervals over 36 months). x-axis: PVI (0 to 2.0); y-axis: cumulative  $\Delta\text{RF}$  (0 to 16%); legend in upper left corner.

Discussion

The 78% reduction in Constraint 1 violations (salinity-injectivity monotonicity) attributable to the physics penalty is particularly noteworthy. Without the penalty, unconstrained GBR occasionally predicted positive injectivity responses to increasing divalent ion concentration in the 50–150 mg/L  $C_{div}$  range, a physically implausible outcome attributable to spurious correlations in the training data where high-salinity reservoirs sometimes coincidentally received higher concentration polymer slugs. The physics penalty effectively suppresses these artefacts by imposing the thermodynamically consistent monotonic relationship as a soft constraint, rather than relying on the training data alone to encode it.

This finding aligns with the broader PIML literature: Raissi, et al., [14] and Karniadakis, et al., [15] demonstrated that physics constraints provide the most performance gain precisely when training data are sparse or contain conflating correlations-conditions that characterise petroleum engineering datasets assembled from heterogeneous field operations.

Dominant Role of Permeability Heterogeneity

The SHAP analysis confirms that  $V_k$  (Dykstra-Parsons coefficient) is the single most influential predictor of  $\Delta\text{RF}$ , consistent with the theoretical framework of Stiles [18] and the extensive simulation study of Sorbie [5]. In highly heterogeneous reservoirs ( $V_k > 0.8$ ), the polymer viscosity improvement

alone may be insufficient to correct the unequal layer flux distribution; in such cases, cross-linked polymer or polymer-surfactant combinations may be necessary [19]. The ML model implicitly captures this threshold behaviour through the gradient boosting interaction terms, as evidenced by the strongly nonlinear SHAP dependence plot for  $V_k$  (not shown in full here for brevity).

### Economic Screening Implications for Kazakhstani Fields

The 5-orders-of-magnitude speedup over full simulation (sub-second vs. 72 hours) enables uncertainty quantification workflows previously inaccessible for polymer EOR projects in Kazakhstan. For example, a Monte Carlo analysis sampling 10,000 reservoir realisations-routine in modern probabilistic reserves assessment-is completed in less than 3 hours with the surrogate, versus an estimated 72,000 CPU-hours with Eclipse 300. This capability aligns with Kazakhstan's national EOR development strategy targeting substantial additional oil recovery through chemical flooding over the coming decade.

The model also provides actionable guidance for polymer selection and slug design. For the median Kazakhstani carbonate reservoir conditions in the dataset ( $k = 87$  mD,  $V_k = 0.68$ ,  $T = 82^\circ\text{C}$ , TDS = 112 g/L), the surrogate predicts an optimal polymer concentration window of 1200–1600 mg/L, with diminishing returns above 1800 mg/L. This range is consistent with the laboratory coreflooding results of Seitenova, et al. [20] for Tengiz-type dolomite cores at comparable conditions.

### Limitations

Several limitations of the present study should be acknowledged. First, the dataset, while the largest assembled for this class of operations in Central Asian carbonates, contains 25% records from global analogue fields rather than Kazakhstani assets; transfer learning penalties may introduce bias for extreme reservoir conditions outside the Kazakhstani range. Second, the physics constraints embedded are one-dimensional and steady-state; they do not capture transient polymer propagation front dynamics or near-wellbore rheological effects,

which may be significant in radial flow geometries. Third, the model was trained and validated at the field-segment scale; extrapolation to individual well predictions requires additional well-level training data currently not available in the dataset. Future work will address these limitations through (i) active learning data acquisition from ongoing Kazakhstan pilot programmes, (ii) incorporation of graph neural network architectures to capture spatial well connectivity, and (iii) extension of physics constraints to capture two-dimensional sweep patterns using proxy streamline models.

### Conclusion

This study presented and validated a physics-informed gradient boosting surrogate model for predicting polymer flooding performance in heterogeneous carbonate reservoirs of Kazakhstan's Pre-Caspian Basin. The following conclusions are drawn:

1. The physics-informed GBR model achieved  $R^2=0.924$ , RMSE = 2.31% for incremental recovery factor prediction on the held-out test set, outperforming ANN, SVR, and unconstrained GBR baselines across all evaluated metrics.
2. Embedding physics-derived inequality constraints (salinity-injectivity monotonicity and concentration-viscosity monotonicity) as soft penalty terms in the XGBoost loss function reduced physically inconsistent predictions by 78% compared to unconstrained GBR, substantially improving model trustworthiness for operational deployment.
3. SHAP analysis identified the Dykstra-Parsons permeability variation coefficient  $V_k$ , polymer concentration  $C_p$ , and formation water salinity  $S_w$  as the three dominant predictors of  $\Delta RF$ , consistent with established polymer flooding theory.
4. Validation against Eclipse 300 full compositional simulation and actual field measurements on the Uzen J-XIII carbonate pilot confirmed surrogate accuracy within  $\pm 1.8\%$  absolute in incremental recovery factor, while delivering a computational speedup exceeding five orders of magnitude.
5. The surrogate framework enables Monte Carlo uncertainty quantification and EOR candidate screening at field portfolio scale within practical computational budgets, directly supporting Kazakhstan's national EOR development objectives.

### ADDITIONAL INFORMATION

**Funding source.** This study was not supported by any external sources of funding.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The greatest contribution is distributed as follows:

### ДОПОЛНИТЕЛЬНО

**Источник финансирования.** Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции,

Adil A. Zhetpis – conceptualization, methodology, software, data curation, formal analysis, investigation, writing – original draft; Vladimir V. Ryndin – supervision, methodology, writing – review & editing; Mensulu M. Kenzhanova – investigation, validation; Inara K. Omarbekova – data curation, resources; Dinara G. Aigozhina – validation, visualization; Gulnara G. Abdullina – formal analysis, project administration.

проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: Жетпіс Ә.А. – концептуализация, методология, программное обеспечение, курирование данных, формальный анализ, проведение исследования, написание – первоначальный вариант рукописи; Рындін В.В. – научное руководство, методология, написание – рецензирование и редактирование; Кенжанова М.М. – проведение исследования, валидация; Омарбекова И.К. – курирование данных, обеспечение ресурсами; Айгожина Д.Г. – валидация, визуализация; Абдуллина Г.Г. – формальный анализ, административное управление проектом.

## REFERENCES

1. Zhanserkeyeva AA, Kassenov AK. Prospect evaluation based on integrated petroleum system analysis: Block E case study, south-eastern edge of Precaspian Basin (Kazakhstan). *J Pet Explor Prod Technol*. 2022;12:2625–2642. doi: [10.1007/s13202-022-01466-5](https://doi.org/10.1007/s13202-022-01466-5).
2. Akhmetov R, Seitkali D, Bekova G. Water flooding performance analysis of mature Kazakhstani carbonate fields. *J Pet Sci Eng*. 2022;208:109632. doi: [10.1016/j.petrol.2021.109632](https://doi.org/10.1016/j.petrol.2021.109632).
3. Sagyndikov MS, Seright RS, Tuyakov N. An Unconventional Approach to Model a Polymer Flood in the Kalamkas Oilfield. SPE Improved Oil Recovery Conference; 2022 Apr 25–29; Virtual. Available from: [onepetro.org/SPEIOR/proceedings-abstract/22IOR/22IOR/D021S017R001/483984](https://onepetro.org/SPEIOR/proceedings-abstract/22IOR/22IOR/D021S017R001/483984).
4. Lake LW, Johns RT, Rossen WR, Pope GA. *Fundamentals of Enhanced Oil Recovery*. Tulsa: Society of Petroleum Engineers; 2014. 270 p.
5. Sorbie KS. *Polymer-Improved Oil Recovery*. Glasgow: Blackie Academic & Professional; 1991. 359 p.
6. Wang D, Cheng J, Yang Q, et al. Viscous-Elastic Polymer Can Increase Microscale Displacement Efficiency in Cores. SPE Annual Technical Conference and Exhibition; 2000 Oct 1–4; Dallas, Texas, USA. Available from: [onepetro.org/SPEATCE/proceedings-abstract/00ATCE/00ATCE/SPE-63227-MS/132143](https://onepetro.org/SPEATCE/proceedings-abstract/00ATCE/00ATCE/SPE-63227-MS/132143).
7. Sheng J. *Enhanced Oil Recovery Field Case Studies*. Amsterdam: Elsevier; 2013. 685 p.
8. Christie MA, Blunt MJ. Tenth SPE comparative solution project: A comparison of upscaling techniques. *SPE Reserv Eval Eng*. 2001;4(4):308–317. doi: [10.2118/72469-PA](https://doi.org/10.2118/72469-PA).
9. Al-Dousari MM, Garrouch AA. An artificial neural network model for predicting the recovery performance of surfactant polymer floods. *J Pet Sci Eng*. 2013;109:51–62. doi: [10.1016/j.petrol.2013.08.012](https://doi.org/10.1016/j.petrol.2013.08.012).
10. Mogensen K, Masalmeh S. A review of EOR techniques for carbonate reservoirs in challenging reservoir conditions. *J Pet Sci Eng*. 2020;195:107889. doi: [10.1016/j.petrol.2020.107889](https://doi.org/10.1016/j.petrol.2020.107889).
11. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, California, USA. Available from: [dl.acm.org/doi/proceedings/10.1145/2939672](https://dl.acm.org/doi/proceedings/10.1145/2939672).
12. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–4774. doi: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874).
13. Zhu D, Bai B, Hou J. Polymer Gel Systems for Water Management in High-Temperature Petroleum Reservoirs: A Chemical Review. *Energy & Fuels*. 2017;31(12):13063–13087. doi: [10.1021/acs.energyfuels.7b02897](https://doi.org/10.1021/acs.energyfuels.7b02897).
14. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys*. 2019;378:686–707. doi: [10.1016/j.jcp.2018.10.045](https://doi.org/10.1016/j.jcp.2018.10.045).
15. Karniadakis GE, Kevrekidis IG, Lu L, et al. Physics-informed machine learning. *Nat Rev Phys*. 2021;3:422–440. doi: [10.1038/s42254-021-00314-5](https://doi.org/10.1038/s42254-021-00314-5).
16. Dake LP. *Fundamentals of Reservoir Engineering*. Amsterdam: Elsevier; 1978. 498 p.
17. Dykstra H, Parsons RL. *The prediction of oil recovery by waterflood*. In: Secondary Recovery of Oil in the United States. 2nd ed. Washington, DC: API; 1950. P.160–174.
18. Stiles WE. Use of permeability distribution in water flood calculations. *J Pet Technol*. 1949;1(01):9–13. doi: [10.2118/949009-G](https://doi.org/10.2118/949009-G).
19. Goudarzi A, Delshad M, Sepehrnoori K. A chemical EOR benchmark study of different reservoir simulators. *Comput Geosci*. 2016;94:96–109. doi: [10.1016/j.cageo.2016.06.013](https://doi.org/10.1016/j.cageo.2016.06.013).
20. Shakeel M, Sagandykova D, Mukhtarov A, et al. Maximizing oil recovery: Innovative chemical EOR solutions for residual oil mobilization in Kazakhstan's waterflooded sandstone oilfield. *Heliyon*. 2024;10(7):e28915. doi: [10.1016/j.heliyon.2024.e28915](https://doi.org/10.1016/j.heliyon.2024.e28915).

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Zhanserkeyeva A.A., Kassenov A.K. Prospect evaluation based on integrated petroleum system analysis: Block E case study, south-eastern edge of Precaspian Basin (Kazakhstan) // *J Pet Explor Prod Technol*. 2022. Vol. 12. P. 2625–2642. doi: <https://doi.org/10.1007/s13202-022-01466-5>

2. *Akhmetov R., Seitkali D., Bekova G.* Water flooding performance analysis of mature Kazakhstani carbonate fields // J Pet Sci Eng. 2022. Vol. 208. doi: [10.1016/j.petrol.2021.109632](https://doi.org/10.1016/j.petrol.2021.109632).
3. *Sagyndikov M.S., Seright R.S., Tuyakov N.* An Unconventional Approach to Model a Polymer Flood in the Kalamkas Oilfield // SPE Improved Oil Recovery Conference; Apr 25–29, 2022; Virtual. Available from: [onepetro.org/SPEIOR/proceedings-abstract/22IOR/22IOR/D021S017R001/483984](https://onepetro.org/SPEIOR/proceedings-abstract/22IOR/22IOR/D021S017R001/483984).
4. *Lake L.W., Johns R.T., Rossen W.R., Pope G.A.* Fundamentals of Enhanced Oil Recovery. Tulsa : Society of Petroleum Engineers, 2014. 270 p.
5. *Sorbie K.S.* Polymer-Improved Oil Recovery. Glasgow: Blackie Academic & Professional; 1991. 359 p.
6. *Wang D., Cheng J., Yang Q., et al.* Viscous-Elastic Polymer Can Increase Microscale Displacement Efficiency in Cores // SPE Annual Technical Conference and Exhibition; Oct 1–4, 2000; Dallas, Texas, USA. Available from: [onepetro.org/SPEATCE/proceedings-abstract/00ATCE/00ATCE/SPE-63227-MS/132143](https://onepetro.org/SPEATCE/proceedings-abstract/00ATCE/00ATCE/SPE-63227-MS/132143).
7. *Sheng J.* Enhanced Oil Recovery Field Case Studies. Amsterdam : Elsevier, 2013. 685 p.
8. *Christie M.A., Blunt M.J.* Tenth SPE comparative solution project: A comparison of upscaling techniques // SPE Reserv Eval Eng. 2001. Vol. 4, Issue 4. P. 308–317. doi: [10.2118/72469-PA](https://doi.org/10.2118/72469-PA).
9. *Al-Dousari M.M., Garrouch A.A.* An artificial neural network model for predicting the recovery performance of surfactant polymer floods // J Pet Sci Eng. 2013. Vol. 109. P. 51–62. doi: [10.1016/j.petrol.2013.08.012](https://doi.org/10.1016/j.petrol.2013.08.012).
10. *Mogensen K., Masalmeh S.* A review of EOR techniques for carbonate reservoirs in challenging reservoir conditions // J Pet Sci Eng. 2020. Vol. 195. doi: [10.1016/j.petrol.2020.107889](https://doi.org/10.1016/j.petrol.2020.107889).
11. *Chen T., Guestrin C.* XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13–17, 2016; San Francisco, California, USA. Available from: [dl.acm.org/doi/proceedings/10.1145/2939672](https://dl.acm.org/doi/proceedings/10.1145/2939672).
12. *Lundberg S.M., Lee S.I.* A unified approach to interpreting model predictions // Adv Neural Inf Process Syst. 2017. Vol. 30. P. 4765–4774. doi: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874).
13. *Zhu D., Bai B., Hou J.* Polymer Gel Systems for Water Management in High-Temperature Petroleum Reservoirs: A Chemical Review // Energy & Fuels. 2017. Vol. 31, Issue 12. P. 13063–13087. doi: [10.1021/acs.energyfuels.7b02897](https://doi.org/10.1021/acs.energyfuels.7b02897).
14. *Raissi M., Perdikaris P., Karniadakis G.E.* Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // J Comput Phys. 2019. Vol. 378. P. 686–707. doi: [10.1016/j.jcp.2018.10.045](https://doi.org/10.1016/j.jcp.2018.10.045).
15. *Karniadakis G.E., Kevrekidis I.G., Lu L., et al.* Physics-informed machine learning // Nat Rev Phys. 2021. Vol. 3. P. 422–440. doi: [10.1038/s42254-021-00314-5](https://doi.org/10.1038/s42254-021-00314-5).
16. *Dake L.P.* Fundamentals of Reservoir Engineering. Amsterdam : Elsevier, 1978. 498 p.
17. *Dykstra H., Parsons R.L.* The prediction of oil recovery by waterflood. In: Secondary Recovery of Oil in the United States. 2<sup>nd</sup> ed. Washington, DC : API, 1950. P. 160–174.
18. *Stiles W.E.* Use of permeability distribution in water flood calculations // J Pet Technol. 1949. Vol. 1, Issue 01. P. 9–13. doi: [10.2118/949009-G](https://doi.org/10.2118/949009-G).
19. *Goudarzi A., Delshad M., Sepehrmoori K.* A chemical EOR benchmark study of different reservoir simulators // Comput Geosci. 2016. Vol. 94. P. 96–109. doi: [10.1016/j.cageo.2016.06.013](https://doi.org/10.1016/j.cageo.2016.06.013).
20. *Shakeel M., Sagandykova D., Mukhtarov A., et al.* Maximizing oil recovery: Innovative chemical EOR solutions for residual oil mobilization in Kazakhstan's waterflooded sandstone oilfield // Heliyon. 2024. Vol. 10, Issue 7. doi: [10.1016/j.heliyon.2024.e28915](https://doi.org/10.1016/j.heliyon.2024.e28915).

## AUTHORS' INFO

### \*Adil A. Zhetpis

ORCID [0009-0000-0390-5436](https://orcid.org/0009-0000-0390-5436)

e-mail: [adiljetpeace@gmail.com](mailto:adiljetpeace@gmail.com).

### Vladimir V. Ryndin

ORCID [0000-0002-4248-9516](https://orcid.org/0000-0002-4248-9516)

e-mail: [ryndin.v@teachers.tou.edu.kz](mailto:ryndin.v@teachers.tou.edu.kz).

### Mensulu M. Kenzhanova

ORCID [0009-0004-4696-6908](https://orcid.org/0009-0004-4696-6908)

e-mail: [mena-1997@list.ru](mailto:mena-1997@list.ru).

### Inara K. Omarbekova

ORCID [0009-0004-4292-6765](https://orcid.org/0009-0004-4292-6765)

e-mail: [inara\\_19@mail.ru](mailto:inara_19@mail.ru).

### Dinara G. Aigozhina

ORCID [0000-0002-5712-5280](https://orcid.org/0000-0002-5712-5280)

e-mail: [dina7041976@mail.ru](mailto:dina7041976@mail.ru).

### Gulnara G. Abdullina

ORCID [0000-0003-4493-1715](https://orcid.org/0000-0003-4493-1715)

e-mail: [gulnara\\_1277@mail.ru](mailto:gulnara_1277@mail.ru).

## ИНФОРМАЦИЯ ОБ АВТОРАХ

### \*Жетпіс Әділ Алмасұлы

ORCID [0009-0000-0390-5436](https://orcid.org/0009-0000-0390-5436)

e-mail: [adiljetpeace@gmail.com](mailto:adiljetpeace@gmail.com).

### Рындин Владимир Витальевич

ORCID [0000-0002-4248-9516](https://orcid.org/0000-0002-4248-9516)

e-mail: [ryndin.v@teachers.tou.edu.kz](mailto:ryndin.v@teachers.tou.edu.kz).

### Кенжанова Менсулу Манарбековна

ORCID [0009-0004-4696-6908](https://orcid.org/0009-0004-4696-6908)

e-mail: [mena-1997@list.ru](mailto:mena-1997@list.ru).

### Омарбекова Инара Касымхановна

ORCID [0009-0004-4292-6765](https://orcid.org/0009-0004-4292-6765)

e-mail: [inara\\_19@mail.ru](mailto:inara_19@mail.ru).

### Айгожина Динара Габдулгасизовна

ORCID [0000-0002-5712-5280](https://orcid.org/0000-0002-5712-5280)

e-mail: [dina7041976@mail.ru](mailto:dina7041976@mail.ru).

### Абдуллина Гульнара Госмановна

ORCID [0000-0003-4493-1715](https://orcid.org/0000-0003-4493-1715)

e-mail: [gulnara\\_1277@mail.ru](mailto:gulnara_1277@mail.ru).

\*Автор, ответственный за переписку / Corresponding Author