

Article

Research on Control Factors and Parameter Optimization of Surfactant Flooding in Low-Permeability Reservoirs Using Random Forest Algorithm

Yangnan Shangguan ^{1,2}, Chunling Gao ^{1,2}, Junhong Jia ^{1,2}, Jinghua Wang ^{1,2}, Guowei Yuan ^{1,2}, Huilin Wang ^{1,2}, Jiangping Wu ^{1,2}, Ke Wu ^{1,2}, Yun Bai ^{1,2}, Hengye Liu ^{3,*} and Yujie Bai ³ 

¹ Research Institute of Exploration and Development, PetroChina Changqing Oilfield Company, Xi'an 710018, China

² National Engineering Laboratory for Exploration and Development of Low-Permeability Oil & Gas Fields, Xi'an 710018, China

³ Key Laboratory of Enhanced Oil Recovery, Northeast Petroleum University, Ministry of Education, Daqing 163318, China

* Correspondence: 15512466108@163.com

Abstract

As oil and gas development increasingly targets low and ultra-low permeability reservoirs, conventional recovery techniques often prove insufficient for mobilizing residual oil. Surfactant flooding, a key chemical enhanced oil recovery (EOR) technology, thus requires careful system optimization and mechanistic investigation. This study focuses on low-permeability reservoirs in the Changqing Oilfield, evaluating three surfactant systems—YHS-Z1 (a 7:3 mass ratio blend of hydroxypropyl sulfobetaine and cocamide), YHS-Z2 (a polyether carboxylate, a nonionic-anionic composite) and a middle-phase microemulsion system (Heavy alkylbenzene sulfonate and hydroxysulfobetaine were combined with a mass ratio of 7:3)—through a series of experiments including interfacial tension measurement, contact angle analysis, static and dynamic oil displacement tests, as well as emulsion transport/retention index assessments, to comprehensively characterize their oil displacement properties. Based on the experimental data, this study constructed four classical regression models: Ridge Regression, Random Forest (RF), Gradient Boosting Regression (GBR), and Support Vector Regression (SVR), and conducted a comparative analysis of their predictive performance. The results demonstrate that the Random Forest (RF) model achieved the optimal prediction performance, with a Mean Absolute Error (MAE) of 1.8245, a Mean Absolute Percentage Error (MAPE) of 4.78%, and a coefficient of determination (R^2) of 0.9428 on the training set. Further analysis using the SHapley Additive exPlanations (SHAP) algorithm revealed that the retention index is the primary global factor (accounting for 49.79% of the variance), while significant intergroup differences exist in the primary factors across different surfactant systems. Concurrently, single-factor and multi-factor sensitivity analyses were conducted to elucidate synergistic effects and threshold behaviors among parameters. The optimal parameter combination, identified via a random search method, achieved a predicted recovery factor of 45.61%, representing a 6.57% improvement over the highest experimental value. This study demonstrates that machine learning methods can effectively identify the dominant factors in oil displacement and enable synergistic parameter optimization, thereby providing a theoretical foundation for the efficient development of surfactant flooding in low-permeability reservoirs.



Academic Editor: Dicho Stratiev

Received: 26 February 2026

Revised: 25 March 2026

Accepted: 25 March 2026

Published: 29 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: low permeability reservoir; surfactant flooding; random forest algorithm; main controlling factors

1. Introduction

Low-permeability tight oil reservoirs represent a critical component of global unconventional oil and gas resources, and their cost-effective development is of great significance for ensuring energy security. According to International Energy Agency statistics, the global recoverable resources of low-permeability tight oil reservoirs exceed 400 billion barrels, accounting for approximately 30% of the world's total crude oil resources. In China, low-permeability tight oil resources are widely distributed, with substantial reserves discovered in basins such as Ordos, Songliao, Junggar, and Sichuan, demonstrating huge development potential [1]. However, such reservoirs generally exhibit characteristics including fine pore throats, low permeability, and poor connectivity, leading to unsatisfactory performance in traditional water flooding development, with the average crude oil recovery factor typically below 15% [2].

Surfactant flooding is a core technology in chemical flooding. It offers significant advantages for enhancing recovery factor in low-permeability tight reservoirs. Its effectiveness stems from synergistic mechanisms, including ultra-low interfacial tension, wetting reversal, and emulsification transport [3]. In recent years, researchers have developed various novel surfactant systems. These systems are tailored to the unique characteristics of low-permeability reservoirs. For the ultra-low permeability reservoirs in the Jiyuan area of the Ordos Basin, Liu X et al. [4] conducted a systematic evaluation. They focused on the interfacial activity, emulsification capacity, and wettability modification ability of stearamide-based bio-nano surfactants. Through oil displacement experiments with different reservoir cores, they clarified the factors influencing surfactant displacement efficiency. They also demonstrated the synergistic interaction between wettability and emulsification. Pal et al. [5] synthesized twin surfactants. These surfactants showed outstanding performance in achieving ultra-low interfacial tension. They could also transform rock wettability from neutral to hydrophilic. Shaikhah et al. [6] carried out a systematic review of recent advances in bio-surfactants for enhanced oil recovery. Their review covered production strategies, characterization methods for action mechanisms, and application advantages and disadvantages. They also introduced optimization concepts such as hydrophilic-lipophilic deviation (HLD) to guide molecular design. Kamal et al. [7] provided another systematic review. They pointed out that the final oil displacement effect is determined by multiple factors. These factors include interfacial tension, the degree of wettability modification, emulsion stability, and migration and retention behavior within porous media. The investigation of surfactant interfacial behavior in porous media has long been a key focus in this field. Standnes et al. [8] experimentally confirmed that surfactants can alter rock wettability from oil-wet to neutral-wet or water-wet, thereby enhancing microscopic displacement efficiency. The study by Ahmadi et al. [9] further demonstrated that in low-permeability tight reservoirs, the transport and retention behavior of surfactants decisively influences macroscopic oil displacement performance. Zhang et al. [10] developed a novel formulation composed of anionic and nonionic surfactants and found that higher salinity environments facilitated further interfacial tension reduction, providing new insights for optimizing surfactant concentration in high-salinity reservoirs. Moreover, the stability of emulsified phases and their transport capability within pore throats have also been identified as key factors affecting the oil displacement outcome [11].

Regarding research on adaptability in low-permeability reservoirs, Hirasaki et al. [12] systematically summarized recent advances in surfactant flooding, with particular emphasis on application challenges under complex reservoir conditions. Ahmadi and Shadizadeh [13] investigated the implementation of natural surfactants in aqueous solutions for EOR applications, conducting core flooding experiments to evaluate the effectiveness of the introduced surfactants as EOR agents in carbonate reservoirs. They found that the interfacial tension of the surfactants plays a decisive role in the oil displacement outcome. Ramatou et al. [14] combining phase behavior design with pore-scale visualization, differentiated the combined chemical and mechanical mechanisms and directly visualized the oil displacement process by Winsor III microemulsions, providing new insights for advancing EOR applications at the field scale.

However, complex relationships exist between the aforementioned key parameters and the recovery factor, making it difficult for traditional prediction models based on linear assumptions to accurately describe these complexities. Ahmadi et al. [9] demonstrated that numerical models considering capillary pressure effects still have significant limitations in low-permeability reservoirs. Taber et al. [15] established screening criteria for EOR technologies as early as 1997. Most traditional methods rely on empirical models or numerical simulations, but they exhibit evident shortcomings in prediction accuracy—such as high dependence on experimental data and long calculation time. Wang et al. [16] in their study of the Daqing Oilfield, indicated that improving sweep efficiency is crucial for enhancing oil recovery, but this requires more accurate predictive models.

With the advancement of artificial intelligence technology, machine learning offers a new approach to addressing this challenge. Tariq et al. [17] systematically reviewed the application of machine learning in petroleum engineering, pointing out the significant potential of ML techniques for solving problems in nearly all areas of the oil and gas industry. Mohaghegh [18] pioneered the introduction of intelligent data analysis methods into petroleum engineering, laying the foundation for subsequent research. In chemical flooding prediction, Teixeira and Secchi [19] proposed using machine learning models, specifically artificial neural networks, to represent the nonlinear dynamic behavior of reservoirs. The proposed method was applied to data generated by a synthetic reservoir simulation model, showing promising results. Wu et al. [20] applied a two-step algorithm combining hyperparameter optimization and reconstruction to predict chemical flooding oil/water relative permeability, achieving superior accuracy compared to traditional methods.

In recent years, more advanced machine learning algorithms have been introduced into the field. Cheraghi et al. [21] compared the performance of various machine learning algorithms for EOR prediction, screening applicable ML models and validating their reliability. Ahmadi et al. [22] utilized Support Vector Machines to establish a prediction model for minimum miscibility pressure, providing a new method for chemical flooding optimization. Meng et al. [23] applied machine learning to predict the performance of CO₂ flooding in low-permeability reservoirs, validating the applicability of data-driven models in complex reservoirs. Chaikine et al. [24] developed a production prediction model for fractured horizontal wells based on machine learning methods, demonstrating its potential for practical production forecasting.

Regarding in-depth research on surfactant flooding mechanisms, Alvarado and Manrique [25] provided a systematic update reviewing EOR technologies, emphasizing the need for surfactant formulation optimization. Talebian et al. [26] investigated foam-assisted EOR techniques, offering new perspectives for surfactant application under complex conditions. Zhou et al. experimentally studied the oil displacement effectiveness of surfactants in ultra-low permeability reservoirs, screening polymers suitable for low-permeability, high-salination formations and identifying surfactants with good performance, compatibility,

and low cost. Furthermore, Sheng [27] systematically analyzed the mechanisms of chemical EOR, providing a comprehensive theoretical framework for surfactant selection.

Nevertheless, existing research still has the following shortcomings: Firstly, systematic experimental data for specific surfactant systems tailored to low-permeability tight reservoirs are relatively scarce. Secondly, when constructing predictive models, key characteristics such as surfactant emulsion transport and retention behavior are not comprehensively considered. Additionally, comparative studies on different machine learning algorithms for predicting surfactant flooding performance are not yet in-depth. In response, this study focuses on typical low-permeability tight reservoirs. Through systematic experimental testing, it acquires multidimensional performance parameters for three surfactant systems: YHS-Z1, YHS-Z2, and a middle-phase microemulsion. A dataset containing seven key features, such as interfacial tension, contact angle, and emulsion transport index, is constructed. The predictive performances of five machine learning algorithms—Ridge Regression, Random Forest, Gradient Boosting, Support Vector Regression—are systematically compared. The aim is to establish accurate predictive models linking surfactant performance parameters with the recovery factor, providing theoretical foundation and practical guidance for technical decision-making regarding surfactant flooding in low-permeability tight reservoirs, as detailed in Section 4.2. It is hypothesized that the surfactant flooding effect in low-permeability reservoirs is jointly controlled by multiple physical and chemical parameters, and the machine learning model can efficiently mine the nonlinear mapping relationship between these control factors and the flooding development effect, so as to realize high-precision prediction of surfactant flooding performance and optimal selection of key parameters; among different machine learning algorithms, the ensemble learning model has stronger adaptability and higher prediction accuracy for the complex nonlinear system of low-permeability reservoir surfactant flooding. The research results not only fill the gap in the study of surfactant flooding mechanisms in low-permeability reservoirs but also provide a reliable technical means for optimizing surfactant formulation and injection parameters, which is of great significance for promoting the efficient development of low-permeability oil resources and ensuring energy supply stability.

2. Experiment and Simulation

2.1. Experimental Materials

The crude oil used in the experiments was obtained from the Changqing Oilfield (Xi'an, China), with its key properties summarized in Table 1.

Table 1. Characteristics of oil.

Project	Value
viscosity (63 °C, mPa·s)	4.97
original gas oil ratio (m ³ /t)	8.2
density (20 °C, g/cm ³)	0.923
saturated hydrocarbon (%)	67.41
aromatic hydrocarbon (%)	17.02
the resin constituent in the crude oil (%)	8.77
asphaltene (%)	6.8

The crude oil sample comprised 67.41% saturates, 17.02% aromatics, 6.8% asphaltenes, and 8.77% resins. The mineralization of the formation water was 24.6 g/L, was classified as

the Na_2SO_4 type, exhibited a pH of 8.15, and its detailed ionic composition is provided in Table 2.

Table 2. Simulate the ion content of formation water.

Water Source	$\text{Na}^+ + \text{K}^+$ (mg/L)	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	Cl^- (mg/L)	SO_4^{2-} (mg/L)	HCO_3^- (mg/L)	Mineralization (g/L)	Water Type
original formation water	7465	566	266	7077	7120	1638	24.6	Na_2SO_4

This study focuses on identifying the dominant factors influencing the enhanced oil recovery (EOR) performance of three distinct surfactant systems. Among them, “YHS-Z1” features hydroxypropyl sulfobetaine as its primary agent and Cocamide as its auxiliary agent, with a mass ratio of 7:3, forming an amphoteric-ionic + nonionic composite formulation. “YHS-Z2” primarily consists of polyether carboxylate, a nonionic-7 anionic composite surfactant; The medium-phase microemulsion is a stable system formed by “water-surfactant-auxiliary agent,” appearing as an oily liquid at room temperature. Its main agent is heavy alkylbenzene sulfonate with hydroxysulfobetaine as the auxiliary agent, in a mass ratio of 7:3. The molecular structures of the compounds contained in the above chemical formulations are shown in Figure 1. Considering the impact of concentration variations on the solution’s performance characteristics, the prepared formation water solution was adjusted to common field operation concentrations of 0.05%, 0.1%, 0.15%, 0.2%, and 0.25%. The specific concentration ranges and components are shown in Table 3.

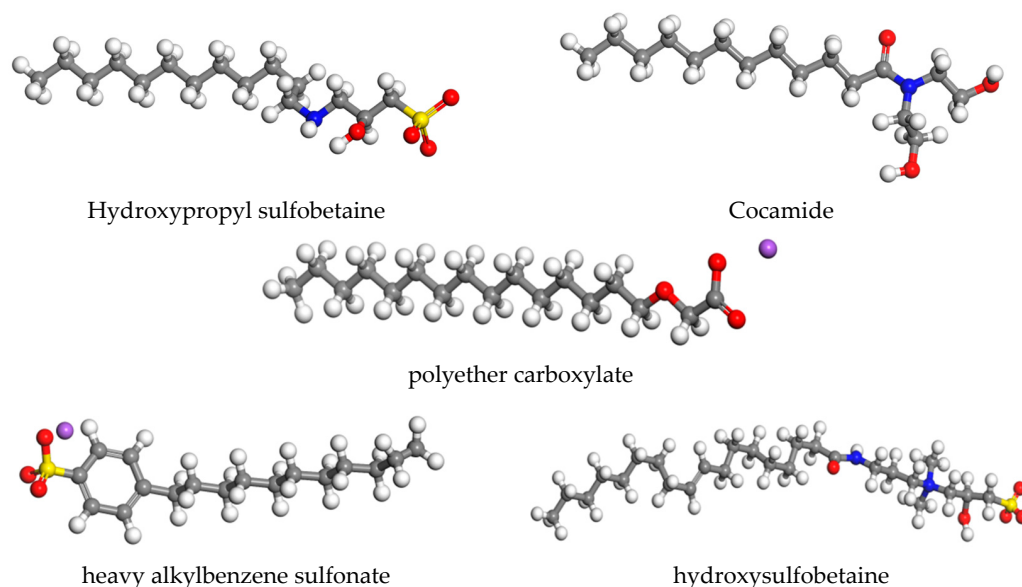


Figure 1. Molecular structure of chemical agents.

The specific cores employed in the experiment were supplied by Changqing Oilfield’s First Oil Production Plant. This batch of cores, classified as feldspathic sandstone, is visually presented in Figure 2. Prior to experimentation, the cores were subjected to oil-washing treatment using xylene, followed by drying. The porosity and permeability of the cores were measured using an overburden pore-permeability instrument. The core length and dry weight were recorded. Subsequently, the cores were saturated with formation water for 48 h, after which the wet weight was measured to calculate the effective porosity. The corresponding data are summarized in Table 4.

Table 3. Specific concentration ranges and components.

Surfactant System	Main Surfactant	Auxiliary Agent	Mass Ratio (Main:Auxiliary)	System Type	State at Room Temperature	Tested Concentration Range
YHS-Z1	Hydroxypropyl sulfobetaine	Cocamide	7:3	Amphoteric-ionic + nonionic composite	—	0.05%, 0.1%, 0.15%, 0.2%, 0.25%
YHS-Z2	Polyether carboxylate	—	—	Nonionic-anionic composite	—	0.05%, 0.1%, 0.15%, 0.2%, 0.25%
Middle-phase microemulsion	Heavy alkylbenzene sulfonate	Hydroxysulfobetaine	7:3	Water-surfactant-adjuvant stable system	Oily liquid	0.05%, 0.1%, 0.15%, 0.2%, 0.25%

**Figure 2.** Core sample diagram.**Table 4.** Core physical property data table.

Core ID	Dimensions	Permeability ($\times 10^{-3} \mu\text{m}^2$)	Porosity (%)
10	$\phi 25 \text{ mm} \times 98 \text{ mm}$	9.29	12.92
11	$\phi 25 \text{ mm} \times 75 \text{ mm}$	11.967	13.69
12	$\phi 25 \text{ mm} \times 72 \text{ mm}$	13.275	12.58
13	$\phi 25 \text{ mm} \times 65 \text{ mm}$	14.293	10.10
14	$\phi 25 \text{ mm} \times 66 \text{ mm}$	14.071	8.41
15	$\phi 25 \text{ mm} \times 67 \text{ mm}$	8.227	11.275

The emulsion migration index/retention index test utilized artificial cores with physical properties closely resembling natural cores, featuring a permeability of $50 \times 10^{-3} \mu\text{m}^2$.

2.2. Interfacial Tension Measurement

The interfacial tension between crude oil and surfactant solutions at varying concentrations, as well as formation water, was measured using a TX-500C spinning drop tensiometer (manufactured by Beijing Shengwei Jiye Technology Co., Ltd., Beijing, China, with a measurement range of 10^{-5} to 10^2 mN/m). Prior to testing, the chemical agents and crude oil were preheated in an oven. The experiment was conducted at a set temperature of 63°C . The instrument was operated at a rotational speed of 6000 rpm for 45 min before image capture commenced to calculate interfacial tension values. Equilibrium was deemed achieved when three consecutive measurements showed dimensional differences less than 0.01 readings, at which point the readings were recorded.

2.3. Contact Angle Measurement

This study employed an advanced static/dynamic contact angle analyzer (manufactured by Shanghai Solon Information Technology Co., Ltd., Shanghai, China, measurement

range: 0–100°, accuracy: $\pm 0.1^\circ$) to measure the contact angles of different concentrations of chemical solutions and formation water on rock surfaces, with a specific focus on investigating the effect of middle-phase microemulsion on rock wettability. Prior to testing, the chemical solutions were preheated to 63 °C. The sessile drop method was adopted for the measurements: after polishing the rock surface to a smooth finish, a 4 μ L droplet was dispensed using the instrument's automatic dispensing system. Images were captured after the droplet stabilized, and each droplet was measured three times to minimize experimental error. The final contact angle value was taken as the average of the three repeated measurements.

2.4. Static Oil Washing Experiment

The ability of different surfactants to actively strip crude oil was evaluated through a crude oil-rock separation test. To simulate the wettability and oil-saturated state of the reservoir, representative quartz sand, with a composition closely matching that of the reservoir, was used. The static oil-washing efficiency of each chemical solution was tested across five increasing concentrations. The specific experimental procedure was as follows:

(1) Dry the washed quartz sand and mix it with crude oil at a ratio of 6:1 to form oil sands, then age the mixture at 63 °C for 24 h.

(2) Upon completion of the aging process, approximately $m_1 = 15$ g of the oil-sand (containing 2.14 g of crude oil) was accurately weighed and placed into individual plastic tubes. A middle-phase microemulsion solution was then added to each tube at a solution-to-oil-sand mass ratio of 2:1. This procedure was repeated to prepare mixtures of the same oil-sand with five different chemical solution concentrations. All plastic tubes were subsequently placed in a constant-temperature oven and maintained at 63 °C for 24 h.

After this incubation period, the mixtures in the tubes were filtered. The separated oil-sand was collected and transferred to an oven for 24 h of drying. The mass of the dried oil-sand together with the centrifuge tube was recorded as m_2 . The static oil displacement efficiency was calculated using Equation (1).

$$\eta = \frac{m_1 - (m_2 - m_0)}{m_1 \times p} \times 100\% \quad (1)$$

In the equation, η is the static oil separation efficiency, %; m_0 is the mass of the centrifuge tube, g; p is the oil content, % ($m_1 \times p = 2.14$ g).

2.5. Dynamic Washing Oil Experiment

The experimental procedure for the dynamic oil displacement test is outlined as follows:

1. The natural core plugs were cleaned and dried.
2. The core was placed in a core holder and preheated to 63 °C. Confining pressure was applied, maintained approximately 2 MPa above the injection pressure. Crude oil was injected at a constant rate of 0.05 mL/min to saturate the core. The volume of water displaced was measured using a graduated cylinder at the outlet until no more water production was observed, at which point injection was stopped. The total volume of displaced water equated to the volume of oil saturating the core. The oil-saturated core was then aged for 24 h at the reservoir temperature.
3. The displacement fluid was switched to formation water and injected at a rate of 0.05 mL/min. The prepared formation water was continuously injected until the water cut at the outlet exceeded 98%. Throughout this process, the volumes of oil produced, water produced, and total liquid produced at the outlet were recorded. The oil recovery factor curve was plotted based on the data.

4. Steps 1–3 were repeated, replacing the displacement fluid with different concentrations of surfactant solutions.

2.6. Emulsification Migration/Retention Index Experiment

Artificial cores with pore-throat characteristics comparable to those of the low-permeability tight reservoirs in the Changqing Oilfield were selected to conduct core flooding experiments. These tests aimed to investigate the oil-water migration behaviors within low-permeability cores under different chemical flooding systems, using the Emulsification Migration Index and Retention Index as key metrics. The study focused on analyzing the retention characteristics of emulsified droplets in pores and their influence law of the occurrence state of residual oil, thereby revealing the synergistic regulation mechanism of emulsification strength-water flooding parameters. The specific experimental procedure was as follows:

The experimental setup was assembled as illustrated in Figure 3. Components including a constant-rate pump, identical piston containers (pre-filled and purged with formation water, experimental oil, and chemical flooding solutions), a pressure monitoring system, a core holder, pipeline, and valves were connected and placed inside a constant-temperature oven. The system was maintained at reservoir temperature for over 3 h to allow the formation water to adequately saturate the core and to ensure the entire core assembly reached a uniform experimental temperature.

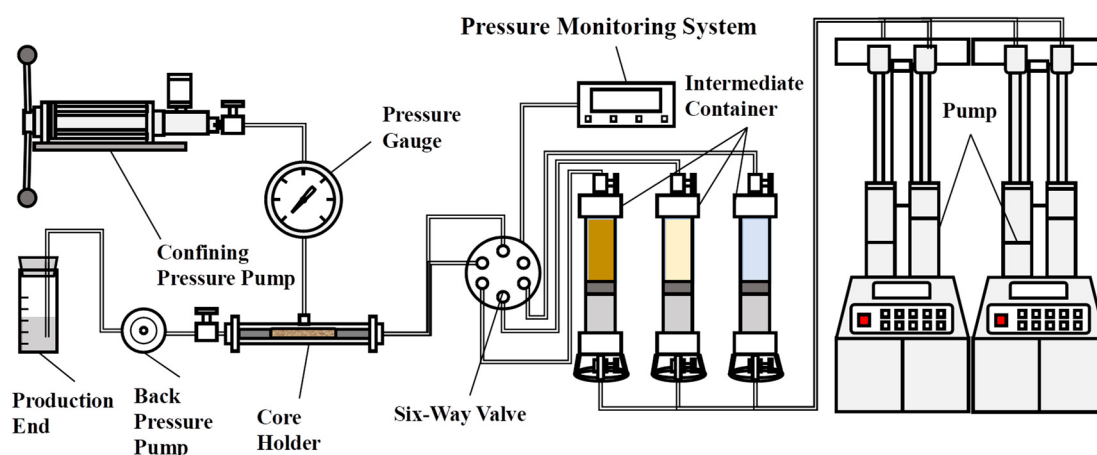


Figure 3. Emulsification migration index device diagram.

Close the exit valve of the piston container filled with chemical EOR solution, and the inlet and outlet valves of the core. Open the inlet and outlet valves of the piston container filled with formation water and experimental oil, the inlet valve of the piston container filled with chemical EOR solution, and the vent valve. The physical diagram is shown in Figure 4.

Two constant-rate pumps were started and maintained at a unified flow rate of 0.05 mL/min. Once the oil-water ratio flowing from the vent valve stabilized, the core inlet and outlet valves were opened while the vent valve was closed. This allowed 0.75 PV of experimental oil and 0.75 PV of formation water to be co-injected into the core simultaneously, each at a rate of 0.05 mL/min.



Figure 4. Physical diagram of the emulsification migration index device.

The pressure reading from the pressure sensor was recorded after every 0.25 PV of experimental oil injected. The average pressure during the injection from 0.5 PV to 0.75 PV of experimental oil was calculated and denoted as p_{wo} .

The outlet valve of the piston container filled with the chemical flooding system solution was opened, while the outlet valve of the container filled with formation water was closed. This enabled the simultaneous co-injection of 5 PV of experimental oil and 5 PV of the chemical flooding system solution into the core, each at a rate of 0.05 mL/min.

The pressure reading was recorded after every 0.25 PV of experimental oil injected. The average pressure during the injection from 4.5 PV to 5 PV of experimental oil was calculated and denoted as p_{em} .

The outlet valve of the piston container filled with formation water was opened, and the outlet valve of the container holding the chemical flooding system solution was closed. This allowed 1.5 PV of experimental oil and 1.5 PV of formation water to be co-injected into the core simultaneously, each at a rate of 0.05 mL/min.

The pressure reading was recorded after every 0.25 PV of experimental oil injected. The average pressure during the injection from 1 PV to 1.5 PV of experimental oil was calculated and denoted as p_{ewo} .

The emulsion migration index was calculated using Equation (2):

$$T_{ei} = \frac{p_{wo}\mu_s}{p_{em}\mu_w} \quad (2)$$

where:

T_{ei} —Emulsion transport index;

p_{wo} —The average pressure of oil-water two-phase flow, MPa.

p_{em} —Emulsion seepage average pressure, MPa.

μ_s —Apparent viscosity of chemical flooding system solution at target reservoir temperature, mPa·s.

μ_w —Apparent viscosity of formation water at target reservoir temperature, mPa·s.

The emulsion retention index is calculated according to Equation (3):

$$R_{ei} = \frac{p_{ewo}}{p_{wo}} \quad (3)$$

In the formula:

R_{ei} —Emulsion retention index.

p_{ewo} —The average pressure of oil-water two-phase seepage after emulsion formation, MPa.

3. Model Construction

3.1. Construction of Machine Learning Model Framework

In this paper, four classical regression models are selected: Ridge Regression, Random Forest (RF), Gradient Boosting Regression (GBR), and Support Vector Regression (SVR). These models cover linear, nonlinear, and ensemble learning types. The core principles and application scenarios of each model are described below, and multi-model comparison is adopted to ensure the reliability and generalization ability of the prediction results.

(1) Data and Preprocessing

① Data Set Construction

Three chemical agents (YHS-Z1, YHS-Z2, and middle-phase microemulsion) were used as experimental objects. Five concentration gradients ranging from 0.05 to 0.25 were set for each agent, and three parallel experiments were performed for each gradient. Finally, 15 groups of valid samples were screened after excluding outliers.

Input characteristics (7 items): Agent type (categorical variable, coded as 1, 2, 3), concentration (%), contact angle ($^{\circ}$), interfacial tension (mN/m), static oil displacement efficiency (%), emulsification rate, transport index, retention index.

The target variable (1 item): natural core recovery factor (%), which is directly measured by core displacement experiment to reflect the final mining effect of chemical agent.

② Data pre-processing

To eliminate the influence of dimensional differences on model training, standardization was applied to normalize continuous features (concentration, contact angle, etc.). The formula is as follows:

$$x_{scaled} = \frac{x - \mu}{\sigma} \quad (4)$$

where x is the original feature value, μ is the feature mean, and σ is the feature standard deviation. To alleviate the curse of dimensionality common in small-sample datasets, the Pearson correlation coefficient between each feature and the target variable was calculated. Features with an absolute correlation coefficient greater than 0.3 were selected as core features, reducing interference from redundant features and improving model stability under small-sample conditions.

(2) Model Training and Evaluation Process

① Data Splitting: Given that only 15 samples were available, the conventional 7:3 train-test split would lead to an extremely small test set and large evaluation bias. Therefore, Leave-One-Out (LOO) cross-validation was used instead: One sample was taken as the validation set, and the remaining 14 samples as the training set. The process was repeated 15 times to complete full-sample validation, maximizing the use of limited data and ensuring reliable evaluation results.

② Model Training: Initial parameters suitable for small-sample scenarios were set for the four models, and parameters were optimized via Leave-One-Out cross-validation. The final key parameters are as follows:

Ridge Regression: regularization coefficient $\alpha = 5.0$ (balances linear fitting and overfitting).

Random Forest (RF): number of decision trees = 50, max depth = 3, minimum samples per leaf = 1 (restricts complexity to avoid overfitting on small samples). The principle of random forest is shown in Figure 5.

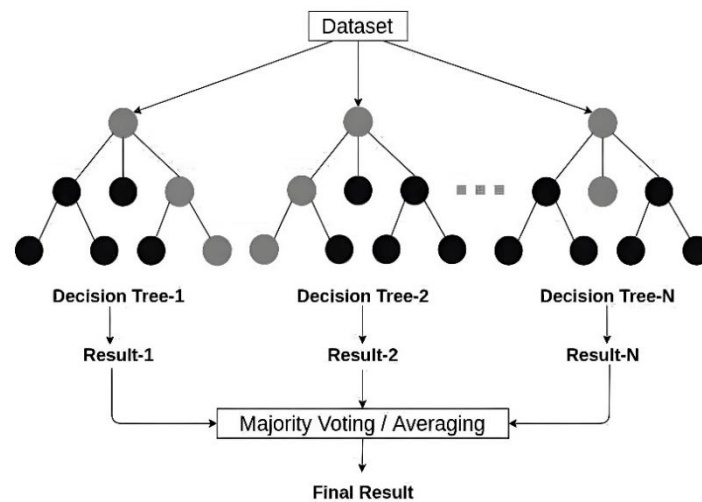


Figure 5. Random forest prediction flow chart.

Gradient Boosting Regression (GBR): number of learners = 50, max depth = 2, learning rate = 0.1 (gradually improves fitting performance).

Support Vector Regression (SVR): RBF kernel, penalty coefficient $C = 5.0$, gamma = 'scale', epsilon = 0.1 (adapted for nonlinear fitting with small samples).

③ Performance Evaluation: Since only one sample exists in each LOO validation set, the coefficient of determination (R^2) may become invalid due to a zero denominator. Therefore, MAE, MAPE, and RMSE were used as the core evaluation metrics, while R^2 was only used as a reference for the fitting performance on the training set. The formulas are as follows:

Mean Absolute Error (MAE): Computes the average absolute magnitude of prediction errors, robust to outliers compared to RMSE. Units match the target variable.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Mean Absolute Percentage Error (MAPE): A dimensionless metric that quantifies the relative percentage of prediction errors (%):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (6)$$

Root Mean Square Error (RMSE): Sensitive to larger errors, representing the square root of the average squared deviations:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Coefficient of Determination (R^2): Measures the goodness-of-fit of the model to the data. A value closer to 1 indicates a better fit:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

In the formula:

n is the number of test samples.

y_i is the real recovery.

$\hat{y}_i$ is predict recovery.

$\bar{y}$ is average of real recovery.

3.2. Feature Importance Analysis Based on SHAP Algorithm

The SHAP (SHapley Additive exPlanations) algorithm was employed. SHAP was selected over other feature importance methods (e.g., Gini importance) because it provides both global and local interpretability for non-linear models like Random Forest. Unlike Gini importance, which only quantifies the overall contribution of each feature to model performance, SHAP decomposes individual predictions into additive feature contributions, enabling us to explain both the aggregate behavior of the model and the specific drivers of each sample's recovery ratio prediction. This dual interpretability is critical for deriving physically meaningful insights into the key factors controlling enhanced oil recovery in low-permeability reservoirs.

Based on the Shapley value from cooperative game theory, it quantifies the contribution of each feature to the prediction outcome. The global importance ranking of features was obtained by calculating the mean absolute value of the SHAP values across all samples. The formula is as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [f(S \cup \{i\}) - f(S)] \quad (9)$$

In the formula:

ϕ_i : The SHAP value of the feature i.

F : The complete set of all features.

S : The subset of features excluding feature i.

n : The total number of features.

$f(S)$: The model's predicted output when using only the feature subset S .

$f(S \cup \{i\})$: The model's predicted output after adding feature i to the subset S .

Combination coefficient $\frac{|S|!(n-|S|-1)!}{n!}$ is used to distribute the contribution weight of all subsets fairly.

4. Results and Discussion

4.1. Study on Surfactant Flooding Characteristic Parameters of Low Permeability Reservoir

Figure 6 illustrates the wettability alterations of YHS-Z1, YHS-Z2 and the middle-phase microemulsion system with increasing concentration at 63 °C and atmospheric pressure.

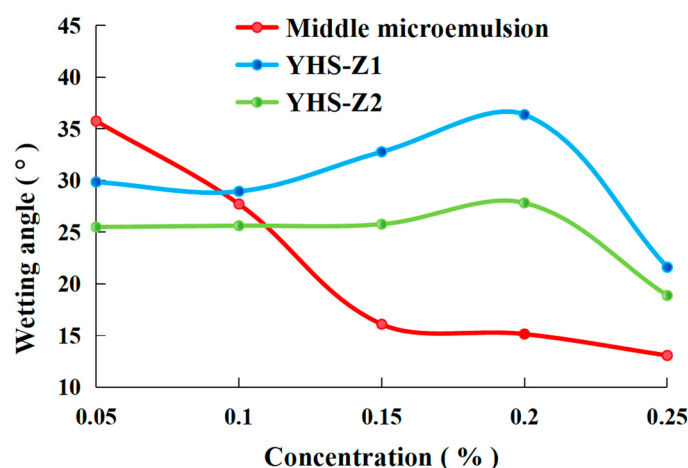


Figure 6. Wettability change.

As shown in Figure 6, all three surfactants demonstrate a consistent trend of decreasing contact angle with increasing concentration from 0.05% to 0.25%, though significant differences exist in their wettability alteration capabilities and characteristics. The middle-phase microemulsion exhibits the most effective wettability modification, attributable to the synergistic adsorption between heavy alkylbenzene sulfonate and hydroxysulfobetaine. YHS-Z2 maintains consistently low contact angles, indicating stable wettability improvement due to the uniform adsorption film formed by its polyether segments. YHS-Z1 shows a fluctuating decline in contact angle, resulting from localized aggregation of zwitterionic molecules creating heterogeneous adsorption, yet still achieves an overall enhancement in hydrophilicity. The fundamental mechanism enabling these surfactants to alter rock wettability lies in their amphiphilic molecular structures. The hydrophilic groups competitively adsorb with silicon-oxygen bonds on rock surfaces and hydrophobic sites of feldspar minerals, gradually displacing crude oil molecules adhered to the rock surface. This process enhances the rock's affinity for water phase wetting. Increasing surfactant concentration further amplifies this effect by increasing the adsorption capacity and coverage of amphiphilic molecules on the rock surface, thereby strengthening the water-wetting modification.

Figure 7 depicts the variation in oil-water interfacial tension with increasing concentration for YHS-Z1, YHS-Z2 and the middle-phase microemulsion system at 63 °C and atmospheric pressure.

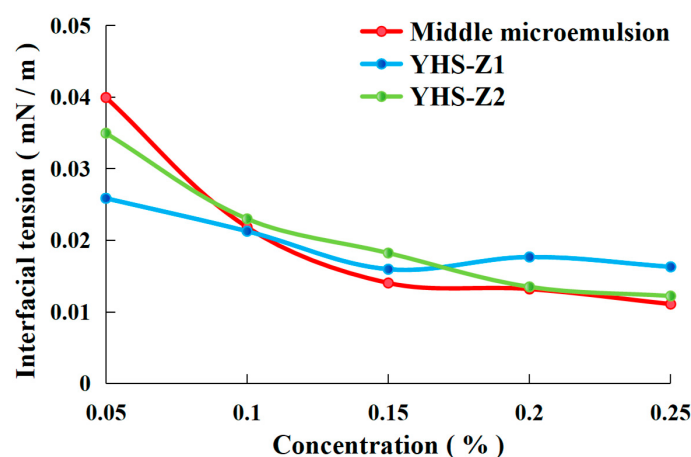


Figure 7. Change of interfacial tension.

As observed in Figure 7 as the concentration of the three surfactants—YHS-Z1, YHS-Z2, and the middle-phase microemulsion—increases from low to high, the oil-water interfacial tension exhibits a continuous decline, gradually plateauing in the later stages. The control group (formation water) exhibited a high oil-water interfacial tension of 54 mN/m. In contrast, the addition of any of the three surfactants led to a sharp reduction in IFT, with higher concentrations resulting in lower values. Ultimately, all systems achieved ultra-low interfacial tension below 0.02 mN/m, with the middle-phase microemulsion showing the most significant reduction. This behavior is attributed to the amphiphilic structure shared by all three surfactant molecules. At the oil-water interface, their hydrophobic tails orient towards the oil phase and their hydrophilic heads towards the water phase. This interfacial adsorption disrupts the original force balance between oil and water molecules, weakening the intermolecular attractions and thereby reducing the interfacial tension. As concentration increases, more surfactant molecules accumulate at the interface, further packing into the interfacial layer and “diluting” the molecular resistance across it, leading to a continued decrease in IFT. The eventual plateau arises from the interfacial adsorption approaching a saturated state.

Figure 8 demonstrates the variation in static oil displacement efficiency with increasing concentration for YHS-Z1, YHS-Z2 and the middle-phase microemulsion system at 63 °C and atmospheric pressure.

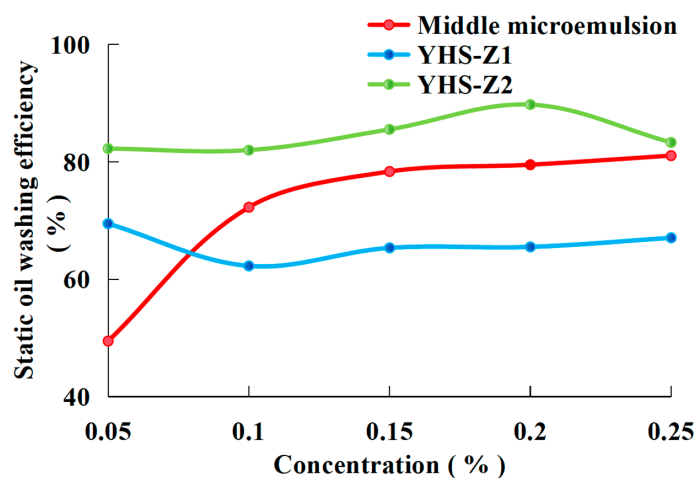


Figure 8. Static oil washing efficiency.

As shown in Figure 8, the static oil displacement efficiency of all three surfactants increases as their concentration rises from 0.05% to 0.25%. YHS-Z2 demonstrates the optimal performance, maintaining efficiency above 82% across the entire concentration range. This is attributed to the uniform adsorption film formed by its nonionic-anionic composite polyether carboxylate structure, which effectively weakens the molecular interactions between crude oil and the quartz sand surface, reduces adhesion resistance, and regulates surface hydrophilicity. The middle-phase microemulsion exhibits the most significant improvement in static efficiency with increasing concentration, due to the synergistic adsorption of heavy alkylbenzene sulfonate and hydroxulfobetaine enhancing the coverage of amphiphilic molecules. This progressively shifts the rock surface toward a more water-wet state, improving active displacement capability. YHS-Z1 shows relatively lower and more fluctuating static efficiency, resulting from localized aggregation of zwitterionic molecules leading to heterogeneous adsorption. Nevertheless, its betaine groups still achieve crude oil displacement through competitive adsorption.

Figure 9 presents the dynamic oil displacement efficiency of YHS-Z1, YHS-Z2 and the middle-phase microemulsion system as a function of increasing concentration at 63 °C under a confining pressure of 2 MPa.

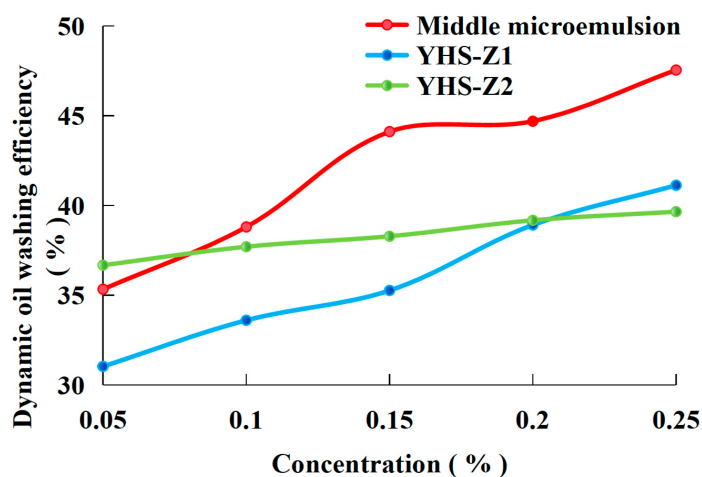


Figure 9. Dynamic oil washing efficiency.

As shown in Figure 9, the dynamic oil displacement efficiency of all three surfactants exhibits a continuous increasing trend with rising concentration, significantly outperforming the control group (formation water dynamic efficiency of 25.85%). The middle-phase microemulsion demonstrates the best performance, with its enhanced displacement effect on core crude oil becoming more pronounced at higher concentrations. This superiority stems from its more effective alteration of rock surface wettability and greater reduction of oil-water interfacial tension at elevated concentrations. These improved interfacial properties substantially weaken crude oil adhesion on pore surfaces, reducing both “pore throat resistance” and “interfacial resistance,” thereby facilitating crude oil stripping and transport under displacement pressure. YHS-Z2 shows a more gradual increase in dynamic efficiency, as the steric hindrance of its polyether chains, while maintaining emulsion stability, provides weaker interfacial activity compared to the middle-phase microemulsion, resulting in limited reduction of flow resistance. YHS-Z1 exhibits the lowest dynamic displacement efficiency due to its relatively poor interfacial tension reduction capability and insufficient control over emulsion stability by the zwitterionic-nonionic composite system, which restricts crude oil transport efficiency.

Overall, static oil displacement efficiency serves as the foundation for dynamic displacement. All three surfactants enhance efficiency through the synergistic mechanism of “wettability alteration-interfacial tension reduction.” However, the middle-phase microemulsion achieves the optimal balance between static and dynamic efficiency at high concentrations, leveraging the superior interfacial properties and emulsion transport advantages of its composite system.

Figure 10 presents the variations in transport index and retention index for YHS-Z1, YHS-Z2 and the middle-phase microemulsion system with increasing concentration.

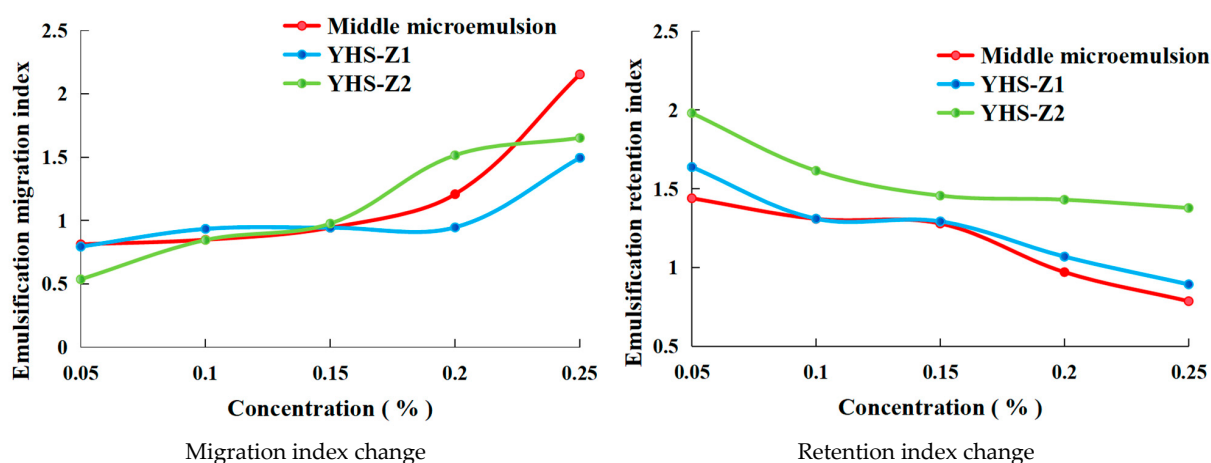


Figure 10. The change of emulsification migration/retention index of three different surfactants with the increase of concentration.

As shown in Figure 10, the emulsion transport index of all three surfactants exhibits a continuous increasing trend with rising concentration, while the retention index shows a consistent decreasing pattern. Higher concentrations result in significantly enhanced emulsion transport indices, indicating improved passage capability of emulsion droplets through rock pores.

The increase in surfactant concentration enhances their adsorption density at the oil-water interface. This elevated adsorption density contributes to improved emulsion transport through two primary mechanisms: First, the substantial reduction in interfacial tension (as previously demonstrated in interfacial tension measurements showing a continuous decrease with concentration) promotes the formation of finer, more uniformly

distributed emulsion droplets from crude oil, thereby enhancing their mobility through porous media (manifested as an increase in the transport index). Second, the formation of a compact, high-strength interfacial film effectively suppresses droplet coalescence, emulsion breakdown, and mechanical entrapment at pore throats, significantly reducing emulsion retention tendency (reflected in the decreasing retention index).

Among the three systems, the middle-phase microemulsion achieves the highest transport index and the lowest retention index at 0.25% concentration. This superior performance originates from the stable middle-phase microemulsion structure formed by the composite system (heavy alkylbenzene sulfonate and hydroxysulfobetaine). This unique structure provides ultra-low interfacial tension and optimal phase behavior characteristics, enabling minimal flow resistance and maximum transport efficiency of emulsions through pore spaces. Consequently, it maintains excellent transport capacity while minimizing retention losses within the porous medium.

4.2. Study on the Main Controlling Factors of Surfactant Flooding in Low Permeability Reservoirs Based on Random Forest Algorithm

The MAE, MAPE, and RMSE of the four models were calculated using Leave-One-Out cross-validation, which is suitable for the small dataset of 15 samples, and the training-set R^2 was supplemented as a reference for fitting performance. The results are listed in Table 5, which intuitively compares the error distribution and stability of each model.

Table 5. Comparative analysis of model performance.

Model	MAE	MAPE (%)	RMSE	Training-Set R^2
Ridge Regression (Ridge)	3.2157	7.65	3.8942	0.8154
Random Forest (RF)	1.8245	4.78	2.3158	0.9428
Gradient Boosting Regression (GBR)	2.1589	5.32	2.6781	0.9215
Support Vector Regression (SVR)	2.8976	6.89	3.4521	0.8579

As observed in the Random Forest (RF) model achieves the best performance: its MAE (1.8245), MAPE (4.78%), and RMSE (2.3158) are the lowest among all models, and its R^2 (0.9428) is the highest, indicating the highest prediction accuracy for recovery factor and superior capability in fitting the nonlinear relationship between parameters and recovery factor.

Table 5 shows that the Random Forest (RF) model achieves the best performance: its MAE (1.8245), MAPE (4.78%), and RMSE (2.3158) are the lowest among all models, and its R^2 (0.9428) is the highest, indicating the highest prediction accuracy for recovery factor and superior capability in fitting the nonlinear relationship between parameters and recovery factor.

At the global level (across all samples), the ranking of feature importance and their corresponding contribution percentages are summarized in Table 6. These results can be intuitively corroborated by the visualizations in Figures 11 and 12, illustrating the contribution of each feature.

As shown in Table 6, the Retention Index emerged as the dominant global controlling factor, contributing nearly 50% to the model's output. This indicates that the degree of chemical retention within core pores directly determines crude oil transport efficiency. Insufficient retention leads to an inadequate effective chemical concentration, whereas excessive retention may cause pore blockage; both scenarios result in reduced oil recovery. The Emulsification Rate and Transport Index were identified as secondary factors, collectively accounting for approximately 21% of the total impact. The emulsification rate reflects

the ability of chemicals to disperse crude oil, while the transport index characterizes the flow behavior of the chemical–oil system. Both are key parameters influencing oil recovery. In contrast, the Contact Angle demonstrated the least influential effect with merely 3.41% contribution, suggesting that under the experimental conditions, chemical-induced wettability alteration contributed less to recovery improvement compared to retention, emulsification, and transport mechanisms.

Table 6. Feature Importance (SHAP)—Overall ranking.

Factors	Importance of SHAP	Proportion of Impact (%)
Retention Index	1.8689	49.79
Emulsification Rate	0.4019	10.71
Transport Index	0.3852	10.26
Static Oil Displacement Efficiency (%)	0.3689	9.83
Interfacial Tension (mN/m)	0.3099	8.26
Concentration (%)	0.2911	7.76
Contact Angle (°)	0.1280	3.41

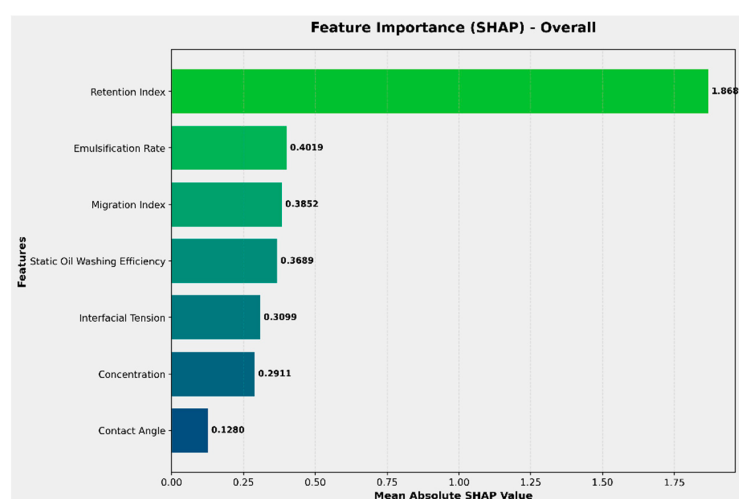


Figure 11. Feature Importance (SHAP)—Overall.

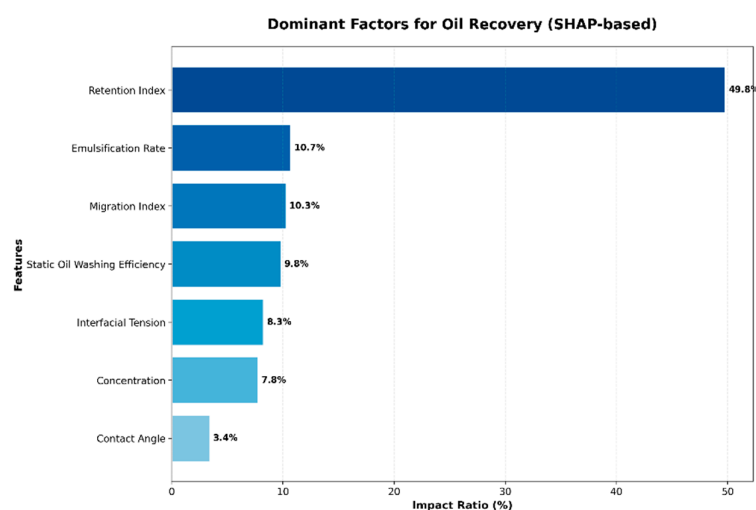


Figure 12. Dominant Factors for Oil Recovery (SHAP-based).

Due to differences in their compositional makeup, YSH-Z1, YSH-Z2, and the middle-phase microemulsion (ME) were divided into two groups: Z1Z2 (including YSH-Z1 and YSH-Z2) and ME. Feature importance was calculated separately for each group to explore inter-group variations. Using SHAP values, the relative contribution of each feature was calculated as its percentage of the total SHAP importance across all features, according to the formula: $\text{Contribution (\%)} = (\text{SHAP importance of the feature} / \text{Total SHAP importance of all features}) \times 100\%$. The proportions of each feature were visualized using bar charts to identify the main controlling factors. As shown in Figures 13–16, these results reveal differences in the action mechanisms of the various chemical agents.

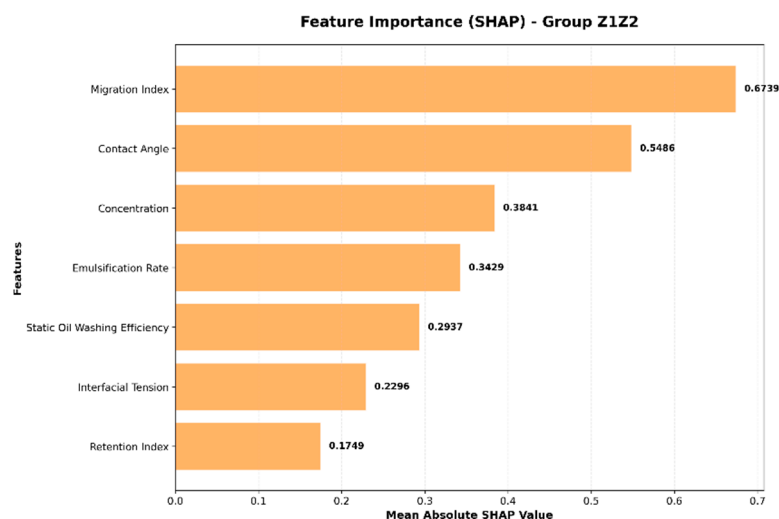


Figure 13. Feature Importance (SHAP)—Group Z1Z2.

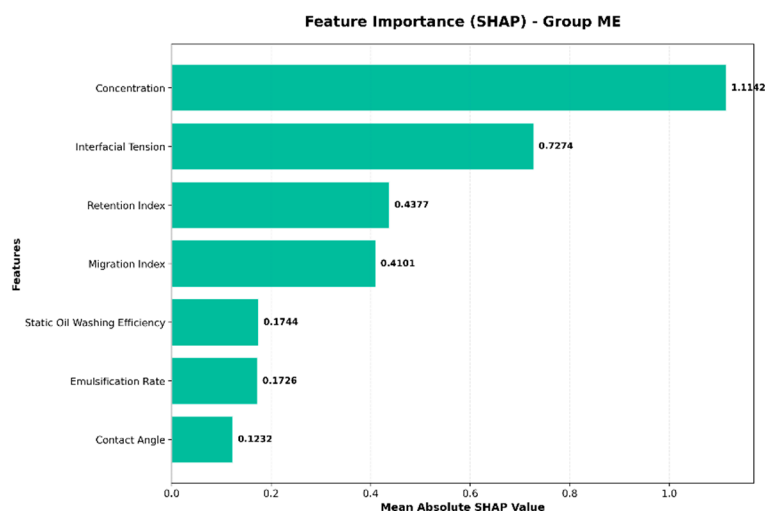


Figure 14. Feature Importance (SHAP)-Group ME.

Z1Z2 Group: The Transport Index (26.78%) and Contact Angle (20.60%) were identified as the most influential factors, suggesting that YSH-Z1 and YSH-Z2 primarily improve oil recovery by enhancing fluid mobility and modifying rock wettability, while concentration played a relatively minor role (11.01%). **ME Group:** In contrast, concentration (29.59%) and Retention Index (22.55%) were the dominant factors, indicating that the middle-phase microemulsion's effectiveness is highly dependent on concentration. Low concentrations result in insufficient interfacial activity, while high concentrations may cause retention-induced pore blockage, highlighting the need for precise concentration control.

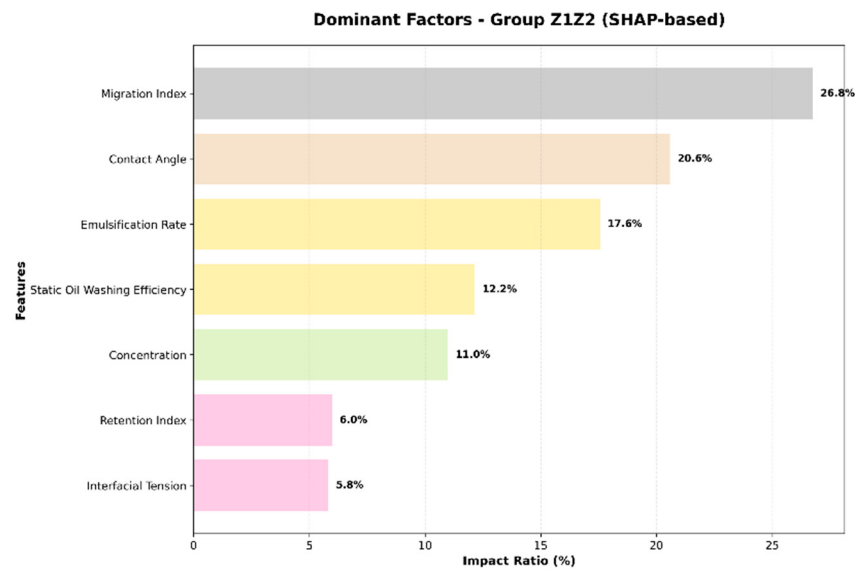


Figure 15. Dominant Factors -Group Z1Z2 (SHAP-based).

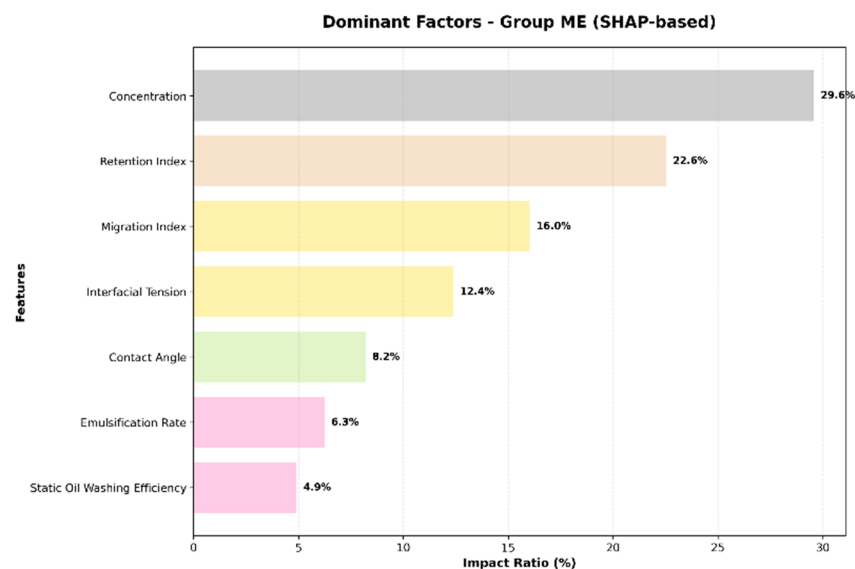


Figure 16. Dominant Factors-Group ME (SHAP-based).

The chemical concentration was fixed at the optimal value of 0.2258%, determined via grid search across the range of 0.05–0.25% using random forest (RF) predictions. For the remaining features, 50 uniformly distributed sampling points were generated within their experimental ranges, and the corresponding recovery factors were predicted. Response curves showing the relationship between each feature and recovery were plotted, as shown in Figure 17.

Retention Index: Within 0.8–1.0, recovery increased linearly with higher retention, but plateaued beyond 1.0, suggesting that moderate retention enhances efficiency, while excessive retention provides no further benefit.

Emulsification Rate: Recovery peaked at values between 2.0 and 3.0, with rates of 42.5–43%, while values outside this range led reduced efficiency.

Interfacial Tension: Recovery increased as tension decreased from 0.05 to 0.15 mN/m, but declined at higher values, consistent with the notion that lower interfacial tension enhances oil mobilization, although excessively low values may cause over-emulsification and impair flow.

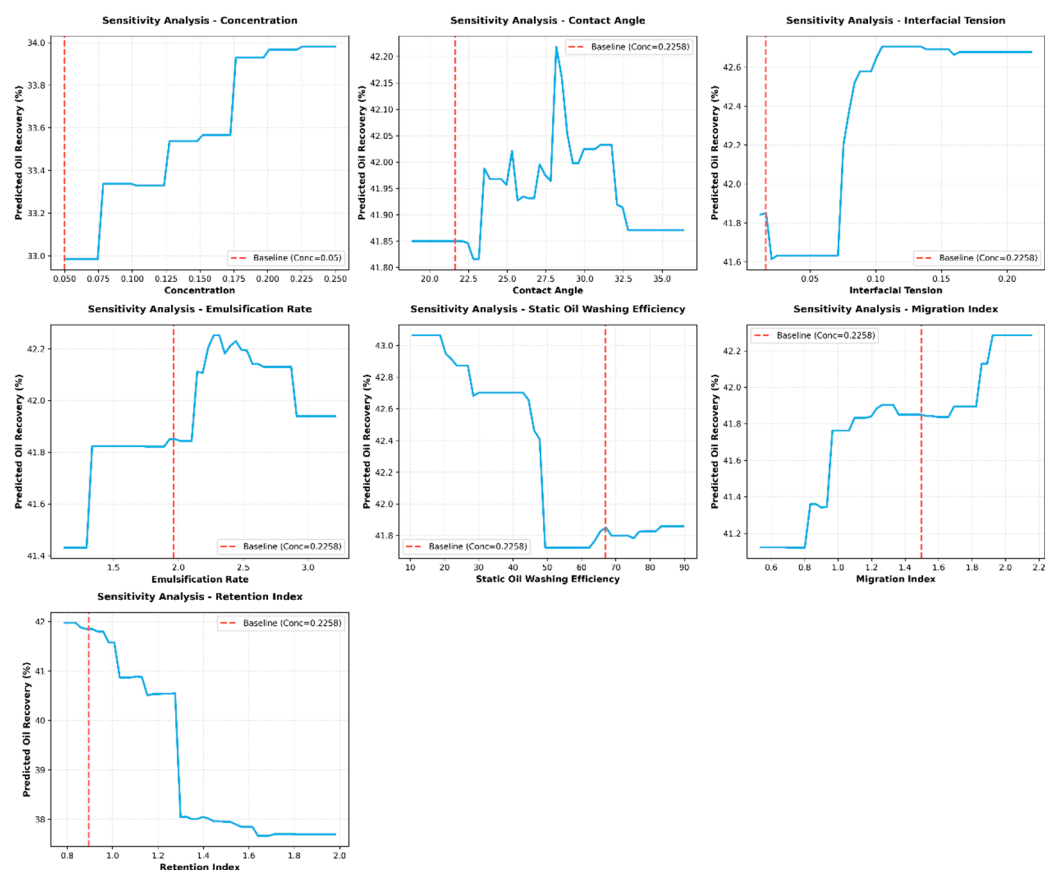


Figure 17. OFAT Sensitivity Analysis.

The two most influential features—Retention Index and Emulsification Rate—were selected to construct a parameter grid, and a 3D interaction plot was generated Figure 18.

3D Interaction: Retention Index vs Emulsification Rate (Opt Conc=0.2258)

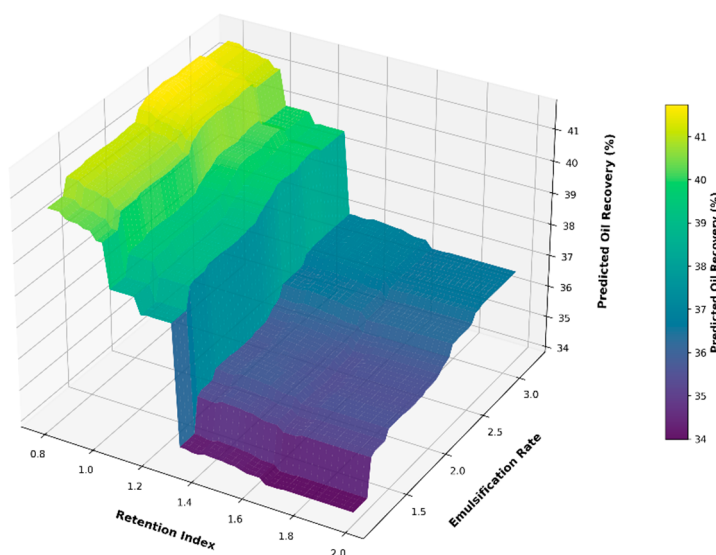


Figure 18. Multifactor Interaction Analysis.

The results indicate that the recovery factor peaked at 43.0–43.5% when the Retention Index was between 0.8 and 1.0 and the Emulsification Rate ranged from 2.5 to 3.0. This highlights a synergistic interaction between the two factors: moderate retention extends

the chemical agent's residence time, while a high emulsification rate enhances crude oil dispersion. Together, these effects led to a significant improvement in oil recovery.

In contrast, recovery remained below 41.0% when the Retention Index exceeded 1.2 or the Emulsification Rate fell below 1.5, even when the other parameter was optimized. These findings underscore the importance of synergistic optimization across multiple parameters, demonstrating that optimizing a single factor alone is insufficient to achieve maximum recovery.

Based on the results of the Random Forest (RF) model and sensitivity analysis, appropriate ranges were established for each feature (e.g., concentration: 0.05–0.25%; contact angle: 20–60°). A random search approach was used to generate 20,000 parameter combinations. The recovery factor for each combination was predicted using the RF model, and the parameter set yielding the highest predicted recovery was selected. The results are shown in Table 7 and Figure 19, and the distribution of the optimal parameter set can be analyzed.

Table 7. Optimal Parameter Data.

Factor	Value
Concentration (%)	0.2925
Contact Angle (°)	28.2338
Interfacial Tension (mN/m)	0.1140
Static Oil Displacement Efficiency (%)	16.4859
Emulsification Rate	2.9834
Transport Index	2.4314
Retention Index	0.8105
Predicted Recovery Factor (%)	45.61%

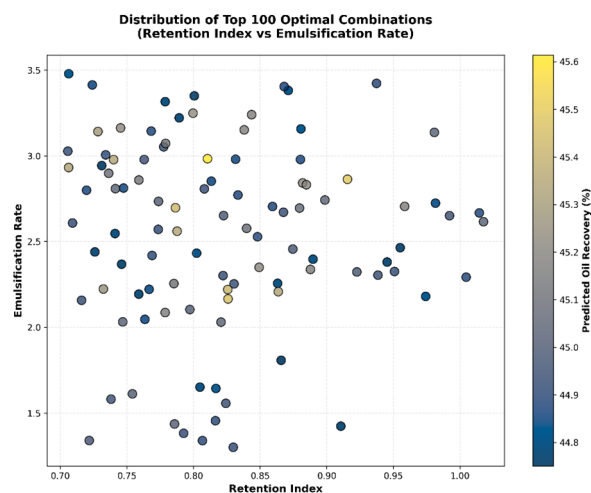


Figure 19. Distribution of Top 100 Optimal Combinations. (Retention Index vs. Emulsification Rate).

Despite the use of leave-one-out cross-validation (LOO) to fully exploit the 15 available experimental samples and mitigate overfitting during model training, the limited sample size constitutes a primary limitation of this study. On one hand, the small dataset fails to comprehensively cover the full range of variations in key factors governing surfactant flooding in low-permeability reservoirs, which imposes constraints on the model's generalizability and necessitates further validation for field operating conditions outside the tested parameter intervals. On the other hand, the restricted sample volume prevents a thorough exploration of the complex non-linear coupling interactions between features, leaving

some low-probability parameter combinations and their correlations with displacement performance inadequately identified. Future work will expand the dataset through additional laboratory core-flooding experiments and synthetic data generation techniques (e.g., SMOTE), and iteratively optimize the model using field test data from Changqing Oilfield to enhance its predictive accuracy and applicability across diverse geological settings and surfactant systems.

To validate the predictive accuracy of the algorithm, a field test was conducted in Block X of the Changqing Oilfield. The recovery factor increased from 39% to 45%, The specific data are shown in Table 8, confirming the accuracy of the algorithm’s predictions and demonstrating its practical utility in field applications.

Table 8. Baseline Production Data.

Block No.	Injection Rate (m³/d)	Oil Production Rate (t/d)	Cumulative Injection Volume (×10⁴ m³)	Original Oil Production Rate (t/d)	Average Stimulation Duration (d)	Statistical Period	Cumulative Additional Oil Production (t)
X-1	120	22.1	131.4	15.0	1095	2022–2025	7774.5
X-2	115	21.3	125.9	14.8	1095	2022–2025	7168.5
X-3	125	23.2	138.1	15.2	1095	2022–2025	8551.0
X-4	118	21.7	129.2	14.9	1095	2022–2025	7456.5
X-5	122	22.5	133.6	15.1	1095	2022–2025	7996.5
Average	120	21.96	131.6	15.0	1095	2022–2025	7789.4

5. Conclusions

1. Surfactants realize efficient stripping of residual oil on low-permeability rock surfaces through synergistic physical effects: their amphiphilic molecules competitively adsorb on the rock surface to enhance hydrophilicity and reduce the adhesion of crude oil, while lowering oil-water interfacial tension to overcome capillary trapping, and the formed fine stable emulsion droplets with high transport and low retention avoid secondary oil trapping, thus significantly improving microscopic displacement efficiency.
 2. Feature importance varies globally and across surfactant groups. Globally, the Retention Index is the dominant factor (49.79% contribution), directly determining crude oil transport efficiency. For the Z1Z2 group (YHS-Z1 and YHS-Z2), the Transport Index (26.78%) and Contact Angle (20.60%) are key, while the middle-phase microemulsion (ME) group is most influenced by Concentration (29.59%) and Retention Index (22.55%), requiring targeted optimization strategies.
 3. Both single and combined parameters exhibit threshold and synergistic effects on recovery. For example, the Retention Index enhances recovery linearly within 0.8–1.0 but plateaus beyond 1.0, and the Emulsification Rate peaks at 2.0–3.0. Their synergistic optimization (Retention Index: 0.8–1.0, Emulsification Rate: 2.5–3.0) achieves the best recovery, highlighting the necessity of multi-parameter coordination.
 4. The optimal parameter combination identified via random search (Concentration: 0.2925%, Contact Angle: 28.2338°, Interfacial Tension: 0.1140 mN/m, etc.) predicts a recovery factor of 45.61%, a 6.57% improvement over the highest experimental value. This provides a quantitative basis for surfactant formulation adjustment and field application.
- Practically, the optimal parameter combination and key controlling factors identified can directly guide surfactant formulation adjustment and injection scheme optimization in low-permeability reservoir development, reducing field trial-and-error costs.
- This study is based on a dataset of 15 samples, which is relatively small for machine learning applications. While Leave-One-Out Cross-Validation (LOOCV) was employed to

maximize the use of limited data, the small sample size may limit the generalizability of the model. Future work will focus on incorporating synthetic data generation techniques (e.g., SMOTE) to augment the dataset or collecting larger field datasets to improve the robustness and predictive power of the model. Future research will incorporate fracture-related parameters into the RF model and expand the study scope to reservoirs with different geological characteristics, further improving the technology's adaptability and accuracy.

Author Contributions: Conceptualization, Y.S.; Methodology, Y.S.; Formal analysis, J.J.; Investigation, J.W. (Jinghua Wang); Resources, J.J., J.W. (Jinghua Wang), G.Y., H.W., J.W. (Jiangping Wu), K.W., Y.B. (Yun Bai), H.L. and Y.B. (Yujie Bai); Data curation, C.G., J.J., J.W. (Jinghua Wang), G.Y., H.W., J.W. (Jiangping Wu), K.W., Y.B. (Yun Bai), H.L. and Y.B. (Yujie Bai); Writing—original draft, C.G.; Writing—review & editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major National Science and Technology Program for Oil and Gas (grant number: 2025ZD1405400).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Yangnan Shangguan, Chunling Gao, Junhong Jia, Jinghua Wang, Guowei Yuan, Huilin Wang, Jiangping Wu, Ke Wu and Yun Bai were employed by the company Research Institute of Exploration and Development, PetroChina Changqing Oilfield Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Jia, C.; Zou, C.; Li, J.; Li, D.; Zheng, M. Evaluation criteria, major types, characteristics and resource prospects of tight oil in China. *Pet. Res.* **2016**, *1*, 1–9. [\[CrossRef\]](#)
2. Lake, L.W.; Johns, R.T.; Rossen, W.R.; Pope, G.A. *Fundamentals of Enhanced Oil Recovery*; Society of Petroleum Engineers: Richardson, TX, USA, 2014.
3. Sheng, J.J. Status of surfactant EOR technology. *Petroleum* **2015**, *1*, 97–105. [\[CrossRef\]](#)
4. Liu, X.; Li, C.; Zhang, Y.; Liu, Y.; Yao, Q.; Wang, K.; Sheng, X.; Han, T. Research on Surfactant Flooding Technology for Ultra-Low Permeability Reservoirs, Ordos Basin. In *Proceedings of the International Field Exploration and Development Conference*; Springer Nature: Singapore, 2024; pp. 1345–1363.
5. Pal, N.; Saxena, N.; Mandal, A. Studies on the physicochemical properties of synthesized tailor-made gemini surfactants for application in enhanced oil recovery. *J. Mol. Liq.* **2018**, *258*, 211–224. [\[CrossRef\]](#)
6. Shaikhah, D.; Loise, V.; Angelico, R.; Porto, M.; Calandra, P.; Abe, A.A.; Testa, F.; Bartucca, C.; Rossi, C.O.; Caputo, P. New trends in biosurfactants: From renewable origin to green enhanced oil recovery applications. *Molecules* **2024**, *29*, 301. [\[CrossRef\]](#)
7. Kamal, M.S.; Hussein, I.A.; Sultan, A.S. Review on surfactant flooding: Phase behavior, retention, IFT, and field applications. *Energy Fuels* **2017**, *31*, 7701–7720. [\[CrossRef\]](#)
8. Standnes, D.C.; Austad, T. Wettability alteration in chalk: 2. Mechanism for wettability alteration from oil-wet to water-wet using surfactants. *J. Pet. Sci. Eng.* **2000**, *28*, 123–143. [\[CrossRef\]](#)
9. Ahmadi, M.A.; Bahadori, A.; Shadizadeh, S.R. A rigorous model to predict the amount of surfactant required for desired recovery in surfactant flooding process. *Fuel* **2015**, *139*, 421–428.
10. Zhang, J.; Zhang, G.; Ge, J.; Feng, A.; Jiang, P.; Li, R.; Zhang, Y.; Fu, X. Laboratory studies of depressurization with a high concentration of surfactant in low-permeability reservoirs. *J. Dispers. Sci. Technol.* **2012**, *33*, 1589–1595. [\[CrossRef\]](#)
11. Hou, J.; Liu, Z.; Zhang, S.; Yue, X.; Yang, J. The role of viscoelasticity of alkali/surfactant/polymer solutions in enhanced oil recovery. *J. Pet. Sci. Eng.* **2005**, *47*, 219–235. [\[CrossRef\]](#)
12. Hirasaki, G.J.; Miller, C.A.; Puerto, M. Recent advances in surfactant EOR. *SPE J.* **2011**, *16*, 889–907. [\[CrossRef\]](#)
13. Ahmadi, M.A.; Shadizadeh, S.R. Implementation of a high-performance surfactant for enhanced oil recovery from carbonate reservoirs. *J. Pet. Sci. Eng.* **2013**, *110*, 66–73. [\[CrossRef\]](#)
14. Ramatou, I.I.; Liu, Z.; Li, Y.; Wang, W.; Cao, J. Pore-scale Mechanisms of Microemulsion Driven Residual Oil in surfactant flooding systems: Phase behavior, Micromodel Visualization, and oil recovery. *SSRN* **2025**, 1–22. [\[CrossRef\]](#)

15. Taber, J.J.; Martin, F.D.; Seright, R.S. EOR screening criteria revisited-Part 1: Introduction to screening criteria and enhanced recovery field projects. *SPE Reserv. Eng.* **1997**, *12*, 189–198. [[CrossRef](#)]
16. Wang, D.; Han, P.; Shao, Z.; Chen, J.; Seright, R.S. Sweep improvement options for the Daqing oil field. *SPE Reserv. Eval. Eng.* **2009**, *12*, 25–34.
17. Tariq, Z.; Aljawad, M.S.; Hasan, A.; Murtaza, M.; Mohammed, E.; El-Husseiny, A.; Alarifi, S.A.; Mahmoud, M.; Abdulraheem, A. A systematic review of data science and machine learning applications to the oil and gas industry. *J. Pet. Explor. Prod. Technol.* **2021**, *11*, 4339–4374. [[CrossRef](#)]
18. Mohaghegh, S.D. Recent developments in application of artificial intelligence in petroleum engineering. *J. Pet. Technol.* **2005**, *57*, 86–91. [[CrossRef](#)]
19. Teixeira, A.F.; Secchi, A.R. Machine learning models to support reservoir production optimization. *IFAC-Pap.* **2019**, *52*, 498–501. [[CrossRef](#)]
20. Wu, D.; Xue, X.; Zhou, L. Improving the robustness of tree-based prediction model of oil/water relative permeability through hyperparameter optimization and re-ensemble algorithm. *Fuel* **2025**, *385*, 134146. [[CrossRef](#)]
21. Cheraghi, Y.; Kord, S.; Mashayekhizadeh, V. Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities. *J. Pet. Sci. Eng.* **2021**, *205*, 108761. [[CrossRef](#)]
22. Bian, X.Q.; Han, B.; Du, Z.M.; Jaubert, J.-N.; Li, M.J. Integrating support vector regression with genetic algorithm for CO₂-oil minimum miscibility pressure (MMP) in pure and impure CO₂ streams. *Fuel* **2016**, *182*, 550–557. [[CrossRef](#)]
23. Meng, S.; Fu, Q.; Tao, J.; Liang, L.; Xu, J. Predicting CO₂-EOR and storage in low-permeability reservoirs with deep learning-based surrogate flow models. *Geoenergy Sci. Eng.* **2024**, *233*, 212467. [[CrossRef](#)]
24. Chaikine, I.A.; Gates, I.D. A machine learning model for predicting multi-stage horizontal well production. *J. Pet. Sci. Eng.* **2021**, *198*, 108133. [[CrossRef](#)]
25. Alvarado, V.; Manrique, E. Enhanced oil recovery: An update review. *Energies* **2010**, *3*, 1529–1575. [[CrossRef](#)]
26. Talebian, S.H.; Masoudi, R.; Tan, I.M.; Zitha, P.L.J. Foam assisted CO₂-EOR: A review of concept, challenges, and future prospects. *J. Pet. Sci. Eng.* **2014**, *120*, 202–215. [[CrossRef](#)]
27. Sheng, J.J. *Modern Chemical Enhanced Oil Recovery: Theory and Practice*; Elsevier: Amsterdam, The Netherlands, 2010.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.