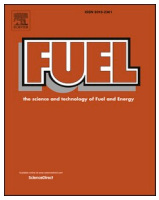
Fuel 404 (2026) 136353

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel



Full Length Article

Oil-bearing evaluation of different shale lithofacies utilizing integrated optimization regression model and multifractal dimension: the Fengcheng Formation in Mahu Sag, China

Zaiquan Yang [a], Dongxia Chen [a,*], Xianglu Tang [a], Yuchao Wang [a], Zhenxue Jiang [a], Liliang Huang [b], Leilei Yang [a], Zhiye Gao [a]

[a] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China
[b] Institute of Exploration and Development, PetroChina Xinjiang Oilfield Company, Karamay 834000, China

ARTICLE INFO

Keywords:
Fengcheng Formation
Lithofacies
Oiliness
Artificial intelligence
Multi-fractal dimension

ABSTRACT

The development of continental shale oil in the Fengcheng Formation, Mahu Sag, is hindered by complex lithofacies and strong reservoir heterogeneity, leading to unclear lithofacies distribution patterns in individual wells and ambiguous mechanisms governing oil-bearing property variations. To address these challenges, this study innovatively integrates artificial intelligence (AI) with multifractal theory. Leveraging scanning electron microscopy (SEM), nuclear magnetic resonance (NMR), multi-stage pyrolysis, and oil saturation data, we developed an intelligent lithofacies identification model to elucidate the differential oil-bearing mechanisms across lithofacies. By correlating logging parameters with fractal dimensions, we established a Full well section "lithofacies-pore structure-oil content" characterization model. Key findings include: 1. An ensemble learning model, optimized via Bayesian optimization and an improved grey wolf optimizer (IGWO), achieves continuous and accurate lithofacies identification in single wells. 2. Mineral composition and reservoir heterogeneity jointly control storage capacity. Carbonate minerals enhance oil-bearing properties through dissolution-induced pore enlargement, whereas quartz and feldspar exhibit dual roles—suppressing storage at low concentrations (<30 %) but optimizing pore networks via rigid scaffolding and dissolution at higher concentrations. 3. Multifractal analysis quantifies the "macropores govern free oil, micropores dominate adsorbed oil" mechanism. 4. A well-scale lithofacies-oil content model is established by integrating logging responses with fractal parameters. This study constructs an intelligent characterization framework for the shale oil, providing a standardized AI-driven solution for sweet-spot prediction and economic development.

1. Introduction

Unconventional hydrocarbon resources constitute an indispensable component of the global energy supply, playing a pivotal role in oil and gas exploration. Among these resources, shale oil and gas have emerged as a top priority for exploration and development [1,2]. Since the early 21st century, rising global energy demand has driven intensified efforts to explore and exploit unconventional reservoirs. By 2023, the shale oil industry achieved significant breakthroughs worldwide [3,4]. The United States has led this progress through innovations in hydraulic fracturing and horizontal drilling, unlocking vast reserves and reshaping global energy dynamics, with profound implications for international supply chains [5,6]. Meanwhile, China has attained critical milestones in shale oil exploration, supported by focused research and substantial technological investments. Major discoveries have been reported in the Sichuan, Songliao, Ordos, and Junggar Basins [7–10]. The Junggar Basin, a cornerstone of China's hydrocarbon exploration, exhibits exceptional promise, particularly in the Jimsar Sag, Mahu Sag, and Chepaizi Uplift, where robust oil production potential has been identified [11]. Among these, the Mahu Sag stands out as the most prospective yet underexplored region. Current research focuses on deciphering shale oil accumulation mechanisms and addressing the complexity of lithofacies, including heterogeneity in oil content and its governing factors.

Previous studies on shale oil in the Mahu Sag have primarily focused on three key aspects: 1. Pore Structure Characterization of Shale Reservoirs. Liu et al. [12] employed scanning electron microscopy (SEM), X-

* Corresponding author.
E-mail address: lindachen@cup.edu.cn (D. Chen).

https://doi.org/10.1016/j.fuel.2025.136353
Received 4 January 2025; Received in revised form 23 April 2025; Accepted 21 July 2025
Available online 28 July 2025

ray diffraction (XRD), and high-pressure mercury intrusion porosimetry (MICP) to compare pore structures between shale and tight sandstone reservoirs, establishing sweet spot evaluation criteria for shale and tight oil reservoirs. Zhang et al. [13] analyzed the relationship between mineral composition and pore structure using similar experimental methods. Li et al. [14] expanded on this work by incorporating rock pyrolysis and nitrogen adsorption experiments, reaching conclusions consistent with Zhang et al. Huang et al. [15] further integrated 1D and 2D nuclear magnetic resonance (NMR) to investigate movable fluid distribution, providing deeper insights into mineral-pore interactions. Tang et al. [16] supplemented these analyses with carbon–oxygen isotopes and fluid inclusion studies, examining pore and fracture development characteristics while discussing mineralogical controls on reservoir heterogeneity. 2. Occurrence State and Mechanisms of Shale Oil. Gong et al. [17] utilized total organic carbon (TOC), reservoir properties, and oil saturation indices to differentiate adsorbed and free oil contents in laminated and interbedded shale types. Gao et al. [18] applied confocal laser scanning microscopy and oil saturation experiments to characterize submicron-scale light and heavy hydrocarbon components, highlighting the significant influence of temperature and pressure on heavy fractions. Jia et al. [19] combined field-emission SEM (FE-SEM), XRD, TOC, HPMI, and confocal microscopy with PCA to evaluate how reservoir structural parameters affect free and adsorbed oil across lithofacies. 3. Lithofacies Classification and Oil-Bearing Properties. Wang et al. [20] classified shale lithofacies using SEM, NMR, and XRD, discussing the storage mechanisms of free and adsorbed oil in different pore systems. Tang et al. [21] adopted comparable methodologies to categorize lithofacies, analyzing pore space variations and mineralogical controls on reservoir capacity. These studies collectively advance the understanding of Mahu Sag shale oil reservoirs but underscore persistent challenges in linking lithofacies heterogeneity to oil content variability.

Previous studies on the Fengcheng Formation shale in the Mahu Sag have established a preliminary lithofacies classification system and elucidated reservoir space characteristics and shale oil occurrence mechanisms through various experimental approaches. However, two major limitations persist in current research: First, most studies have focused on core-scale analysis, while systematic investigations of lithofacies distribution patterns, reservoir heterogeneity, and oil occurrence characteristics at the single-well scale remain insufficient. This knowledge gap has hindered the understanding of vertical oil-bearing property distribution in individual wells. Second, conventional XRD analysis is constrained by discrete sampling depths, making it incapable of achieving continuous, high-resolution lithofacies identification across entire well sections. More importantly, this limitation prevents quantitative characterization of the dynamic coupling relationships among lithofacies, pore structure, and oil-bearing properties along vertical profiles. To address these challenges, an integrated approach incorporating artificial intelligence and advanced analytical methods is required for comprehensive reservoir characterization.

Extensive research has established mature AI applications across petroleum engineering domains, particularly in reservoir evaluation, hydrocarbon development, and enhanced oil recovery (EOR). Three advancements characterize reservoir characterization: First, machine learning algorithms (Artificial Neural Networks, Random Forests, Support Vector Machines) have achieved sub-meter-scale precision in predicting critical parameters like porosity and permeability through integrated analysis of well logs, core samples, and seismic data, significantly enhancing static reservoir model accuracy [22–24]. Second, deep learning architectures (LSTM, GRU) have demonstrated exceptional capability in analyzing dynamic production data to predict spatiotemporal pressure evolution and remaining oil distribution, providing proactive guidance for development planning [25,26]. Third, convolutional neural networks (CNN) and generative adversarial networks (GAN) have revolutionized 3D reservoir heterogeneity modeling, overcoming traditional geostatistical limitations in characterizing complex pore-fracture networks within carbonate and shale formations [27,28]. In hydrocarbon development and EOR, AI technologies have made transformative contributions. Neural networks have successfully established explicit mathematical relationships (e.g., pressure-solubility correlations) to predict CO_2 solubility in ionic liquids, enabling efficient solvent screening for low-energy, high-capture-efficiency systems [29,30]. For CO_2 flooding operations, neural network models have proven effective in predicting diffusion coefficients, minimum miscibility pressure, and critical injection rates [31–33], while reinforcement learning algorithms have optimized injection-production parameters to enhance displacement efficiency. These data-driven modeling approaches have replaced traditional empirical formulas and trial-and-error experimentation, not only addressing challenges in complex physical mechanisms and multi-field coupling but also establishing comprehensive intelligent workflows from static parameter prediction to dynamic process optimization. This technological advancement provides critical support for cost reduction and green transformation in the petroleum industry. However, AI applications in Mahu Sag remain in their infancy. While Chen et al. [34] implemented feedforward neural networks for shale reservoir fracture density prediction and Wang et al. [35] applied deep learning methods to optimize gravel parameter calculations from image logs, systematic research is notably absent in several critical areas. Key knowledge gaps persist in intelligent lithofacies identification and dynamic oil-bearing property prediction.

Although previous studies have established preliminary lithofacies classification systems and oil-bearing evaluation frameworks for the Mahu Sag shale through extensive experimental techniques [22–35], two fundamental limitations remain unresolved. First, the discrete sampling nature of core-based analyses leads to insufficient resolution for well-scale lithofacies classification, failing to quantify gradual mineralogical transitions and their continuous petrophysical responses along vertical profiles. Second, existing investigations inadequately address reservoir heterogeneity controls on oil occurrence, particularly the differential roles of nano- versus micro-scale pore-throat systems, with systematic understanding of pore-size-dependent oil distribution patterns still lacking.

Therefore, this study employs the Bagging ensemble learning algorithm to integrate XRD analysis results with elemental logging parameters. The model parameters are adaptively optimized through a Bayesian optimization framework and an Improved Grey Wolf Optimizer (IGWO), enabling accurate identification of shale reservoir lithofacies. Additionally, by coupling multiple experimental techniques including multi-temperature pyrolysis, nuclear magnetic resonance (NMR), and oil saturation tests. we established a computational formula to quantify oil-bearing properties across different pore sizes in the reservoir, thereby characterizing the distribution of oil content within distinct lithofacies. Furthermore, leveraging multifractal dimension theory, we analyzed the controlling effects of reservoir heterogeneity on oil-bearing properties. Combined with correlation analyses of well-logging parameters (GR, RT, DEN), we developed a single-well conversion model for reservoir oil-bearing properties. Ultimately, this approach culminates in an integrated intelligent characterization technology for “lithofacies–pore size–oil-bearing properties” across entire well sections. The proposed method has been successfully applied to shale reservoirs in the Mahu Depression. Moreover, its conceptual framework can be extended to other basins, offering a novel solution for precise lithofacies characterization and accurate oil-bearing property evaluation in shale reservoirs.

2. Materials and methods

2.1. Construction of the Bayes – IGWO-Bagging regression model

This study employs AI as the core methodology for the refined characterization of shale lithology. Our approach integrates X-ray diffraction (XRD) analysis with elemental logging data to enhance the

precision of elemental log interpretation, thereby ensuring accurate parameter determination throughout the entire well section. The methodology is based on a backpropagation (BP) neural network architecture, augmented with two optimization algorithms – Bayesian optimization and improved grey wolf optimization (IGWO) – to achieve superior data processing accuracy. The BP neural network serves as the primary model within this regression framework [36,37]. By combining the Bayesian optimization model with the IGWO algorithm, we achieve automated parameter adjustment in the primary network (Fig. 1).

The Bayesian optimization technique is a globally applicable strategy rooted in Bayesian statistics principles [38]. It effectively combines the Gaussian process model with an acquisition function to strike a balance between leveraging known information and exploring unknown domains, ultimately minimizing the number of function evaluations. The application of Bayesian optimization is employed to optimize the number of hidden layers and the number of nodes within each hidden layer in a backpropagation neural network. The GWO algorithm is a metaheuristic optimization method inspired by gray wolves' social hierarchy and hunting strategy. The algorithm categorizes the individuals within the Wolf pack into multiple tiers, designating the top three wolves as Alpha, Beta, and Delta, respectively, while classifying the remaining ones as Omega [39–41]. The three dominant wolves in the pack lead and direct other members to update their positions, aiming to approach and capture the prey. In essence, they strive to find the globally optimal solution. The optimization process comprises the subsequent procedures. The A-IGWO algorithm represents a significant advancement over GWO by incorporating an adaptive adjustment mechanism and enhancing the balance between exploration and exploitation. This results in a more effective avoidance of local optima than the original GWO, leading to accelerated convergence and improved solution quality [42,43].

2.2. Nuclear magnetic resonance and the multi-temperature step pyrolysis

Nuclear magnetic resonance (NMR) leverages the behaviour of atomic nuclei in magnetic fields to investigate the microstructure of matter. The technique is particularly well-suited for investigating nuclei possessing non-zero spin quantum numbers, such as the hydrogen nucleus [44]. When a sample is subjected to a strong, static magnetic field, the energy levels of specific nuclei undergo splitting. By applying a radio frequency pulse (20 MHz) that matches the energy level difference, the nucleus is excited to a higher energy state. After the cessation of the RF pulse, the nucleus returns to its original energy level through relaxation. The T_2 relaxation timewas measured according to ASTM D8539-23 [45]. This parameter exhibits high sensitivity to pore structure due to molecular motion and interactions within pores. The distribution of T_2 relaxation times provides insights into pore size, network properties, and fluid dynamics.

The MicroMR12-040 V system an integrated cabinet, a magnet cabinet, and a monitor. The probe coil and primary frequency (20 MHz) for low-field NMR instrumentation. Prior to testing, the system was calibrated using a standard deuterium oxide (D_2O) reference sample. Experiments were conducted in a temperature-controlled environment (22–28 °C). T_2 spectra acquisition parameters included a 90° pulse duration of 10 μs, 3000 echoes, and an echo spacing of 0.2 ms. The instrument is primarily utilized for measuring the T_2 spectrum of the

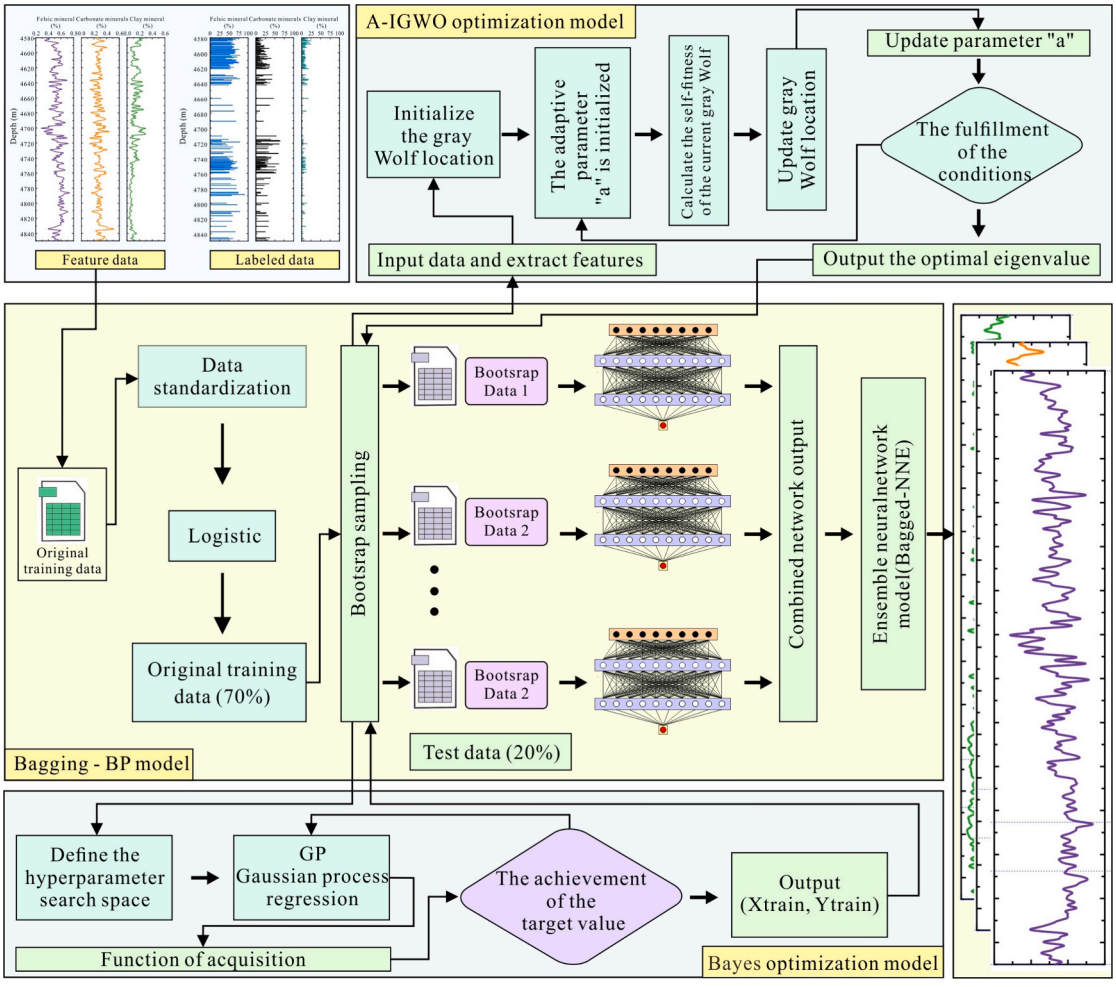


Fig. 1. Automated workflow for lithofacies prediction: Bayesian optimization and adaptive IGWO-driven parameter tuning in Bagging-BP ensemble models.

sample, and the experiment typically operates within a controlled temperature range of 22 °C to 28 °C.

The multi-temperature step pyrolysis experiment was performed according to ASTM UOP649-10 [46] for assessing hydrocarbon generation potential in geological samples. Rock specimens were heated in an oxygen-free environment using a linear temperature program. Continuous thermal desorption-thermal cracking analysis quantified gaseous (C_1-C_5) and liquid hydrocarbon products through flame ionization detection [47]. The ROCK-EVAL6 pyrolyzer system comprising pyrolysis/oxidation furnaces, NDIR sensors, and automated sampling was pre-calibrated using certified shale reference materials. Temperature control and data acquisition ensured $S_{1\text{-}1}$, $S_{1\text{-}2}$, $S_{2\text{-}1}$, $S_{2\text{-}2}$ parameters met reproducibility limits.

2.3. The content of adsorbed oil and free oil in pores of different lithofacies

In this approach, we integrated a set of parameters derived from multi-temperature-order pyrolysis experiments to compute the oil content at various pore sizes, encompassing oil saturation, $S_{1\text{-}1}$, $S_{1\text{-}2}$, $S_{2\text{-}1}$, $S_{2\text{-}2}$, and pertinent data from nuclear magnetic resonance (NMR) investigations. The determination of oil saturation is based on the method of utilizing anhydrous ethanol to extract the aqueous phase from the sample, thereby obtaining the saturation of all hydrocarbons in the overall sample, including light hydrocarbons, resins, putrescine, and asphaltenes. The multi-temperature stage pyrolysis technique involves heating the sample at various temperature stages to analyze the content of different components. The sample is hypothesised to exclusively consist of two fluids, oil and water, occupying the entire pore space. Following this principle, we initially determine the total hydrocarbon content in the sample by multiplying the oil saturation with the pore volume. However, this approach must differentiate between light (free hydrocarbons) and heavy (adsorbed hydrocarbons). The distribution of light and heavy hydrocarbons in the samples was determined to overcome this limitation based on experimental results obtained from multi-temperature step pyrolysis. We further analyzed the signal intensity distribution at various pore sizes in combination with NMR techniques, allowing us to characterize both the free and adsorbed oil within different pore size ranges. In summary, the pore volume occupied by hydrocarbons derived from oil saturation, in conjunction with multi-temperature step pyrolysis parameters, was employed for determining the distribution of oil content at various pore sizes using the percentage of signal intensity from different pore sizes in the NMR data. This approach enables a more accurate evaluation of the oil content within different pore size ranges, providing a new perspective for understanding the distribution of hydrocarbons in pore media (Table 1).

Distribution of liberated oil across different pore sizes:

$$O_i = S_o \times V_p \times \frac{(S_{1-1} + S_{1-2})}{S_T} \times \frac{h_i}{h_t} \tag{1}$$

Distribution of sorbent oil across different pore sizes:

$$O_i = S_o \times V_p \times \frac{(S_{2-1} + S_{2-2})}{S_T} \times \frac{h_i}{h_t} \tag{2}$$

2.4. Multifractal dimension principle

The multifractal dimension is determined in this study using the box-counting method [48,49]. The preprocessing of NMR data is necessary to calculate multifractal dimensions on equally spaced data. The processing proceeds as follows: the ε aperture distribution interval of the NMR test is logarithmically transformed to yield a dimensionless interval Q.

$$y_i = \lg\left(\frac{\rho_i}{\rho_b}\right) \quad i = 1, 2, 3, \cdots, n \tag{3}$$

y_i: The value of the converted pore size measured by NMR experiment; ρ_i: The aperture measured by NMR experiments, nm. ρ_b: The smallest aperture measured by NMR experiments, nm.

After transformation, the dimensionless data set is interpolated to ensure that each interval contains at least one substance. The technique employed for interpolation is the Lagrange polynomial interpolation method (Fig. 2).

$$\begin{cases} P_n(\rho) = \sum_{i=0}^{n} L_i(\rho) \times y_i \\ L_i(x) = \dfrac{(\rho-\rho_0)(\rho-\rho_1)\cdots(\rho-\rho_{i-1})(\rho-\rho_{i+1})}{(\rho_i-\rho_0)(\rho_i-\rho_1)\cdots(\rho_i-\rho_{i-1})(\rho_i-\rho_{i+1})\cdots(\rho_i-\rho_n)} \end{cases} \quad i = 1, 2, 3, \cdots, n \tag{4}$$

After evenly dividing each interval, it can be subdivided into M(ε) sub-intervals of equal length. The ith subinterval is measured as follows:

$$f_i(q,\varepsilon) = \frac{P_i(\varepsilon)^q}{\sum_{i=1}^{M(\varepsilon)} P_i(\varepsilon)^q} \tag{5}$$

Style:$f_i(q,\varepsilon)$: The measurement of the ith subinterval. $\sum_{i=1}^{M(\varepsilon)} p_i(\varepsilon)^q$: The cumulative probabilities of q-order for all subinterval values.

Multifractal singular exponent α(q):

$$\alpha(q) = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{M(\varepsilon)} f_i(q,\varepsilon) \lg p_i(\varepsilon)}{\lg \varepsilon} \tag{6}$$

Multifractal spectral function F(α(q)) obtained by multifractal singular exponent α(q):

Table 1
Parameters of Quantitative Oil-Bearing Characterization Formulas in Reservoirs.

Serial number	Depth (m)	Lithofacies	Porosity (%)	Oil saturation (%)	Water saturation (%)	Multiple temperature pyrolysis				
						$S_{1\text{-}1}$	$S_{1\text{-}2}$	$S_{2\text{-}1}$	$S_{2\text{-}2}$	S_T
104	4753.45	Mixed-felsic shale lithofacies	5.3	13.3	76.5	0.01	0.27	0.37	0.36	1.01
56	4610.89		4.8	7.8	82	0.01	0.25	0.49	0.35	1.11
1–86	4810.24		4.9	5.7	86	0.01	0.91	0.25	0.05	1.23
1–29	4676.47		4.7	19.3	70.4	0.01	0.73	0.66	0.33	1.73
1–16	4612.19	Felsic shale lithofacies	7.8	21.4	13.6	0.01	0.51	0.51	0.9	2.74
a	4613.31		8.1	23.6	12.5	0.01	0.46	1.91	0.64	4.23
33	4761.32	Calcareous-felsic shale lithofacies	4.4	41.2	47.9	0.02	1.24	2.92	4.02	8.21
4	4846.57		4.3	45.1	43.7	0.01	1.97	3.68	1.47	7.13
40–2	4733.05	Felsic-calcareous shale lithofacies	7.6	55.7	34.6	0.1	2.06	3.37	2.17	7.75
e	4751.36		5.9	54.6	35.6	0.07	1.47	2.89	2.24	6.92
39–1	4737.99		5.9	52.2	43.1	0.02	1.9	2.08	1.38	4.59
1–63	4799.98	Carbonate/felsic mixed shale lithofacies	4.5	33.6	55.9	0.03	1.29	1.96	0.58	3.87
30	4682.42		4.6	32.5	56.8	0.11	1.56	1.7	1.11	4.48
j1	4635.08		5.2	37.8	51.5	0.02	1.13	1.88	1.05	4.09

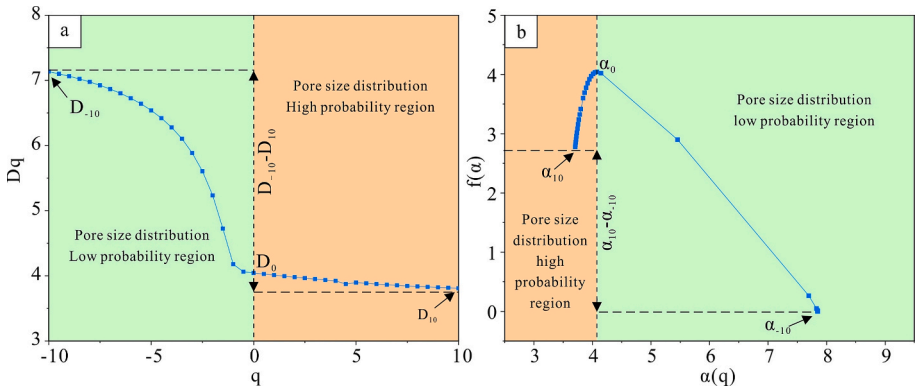


Fig. 2. Multifractal characterization of pore systems: (a) Generalized dimension spectrum (Dq – q). (b) Multifractal spectrum function (f(α) – α(q)).

$$F(\alpha(q)) = \lim_{\varepsilon\to 0}\frac{\sum_{i=1}^{M(\varepsilon)} f_i(q,\varepsilon)\lg f_i(q,\varepsilon)}{\lg\varepsilon} \tag{7}$$

The generalized dimension spectrum Dq can be derived, in conclusion (Fig. 2).

$$\begin{cases} Dq = \dfrac{\lim_{\varepsilon=0}\sum_{i=1}^{M(\varepsilon)}[P_i(\varepsilon)\ln P_i(\varepsilon)]}{\ln\varepsilon}, q = 1 \\ Dq = \dfrac{1}{q-1}\lim_{\varepsilon\to 0}\dfrac{\lg[\sum_{i=1}^{M(\varepsilon)} P_i(\varepsilon)^q]}{\lg\varepsilon}, q \neq 1 \end{cases} \tag{8}$$

q = 0, Dq is denoted as: capacity dimension; q = 1, Dq is denoted as: Information entropy dimension, q = 2, Dq is denoted as: correlation dimension.

3. Geological background

The Junggar Basin, located in northern Xinjiang, China, along the eastern margin of the Kazakhstan Plate [50], comprises six primary tectonic units: the Western Uplift, Wulungu Depression, Luliang Uplift, Eastern Uplift, Central Depression, and North Piedmont Thrust Belt [51]. The study area, Mahu Sag, lies within the western Central Depression adjacent to the Halaalat Mountains and features a Permian succession consisting of (from base to top) the Jiamuhe Formation (P_1j), Fengcheng Formation (P_1f), Xiazijie Formation (P_2x), Lower Wuerhe Formation (P_2w), and Upper Wuerhe Formation (P_3w) [52]. The Fengcheng Formation exhibits a northwest-thickening (800–1800 m) trough deposition and is subdivided into three lithologically distinct members: the lower P_1f_1 (volcaniclastic rocks), middle P_1f_2 (organic-rich shales with dolomitic mudstones), and upper P_1f_3 (terrigenous clastics). This formation represents semi-deep to deep lacustrine deposition in an arid

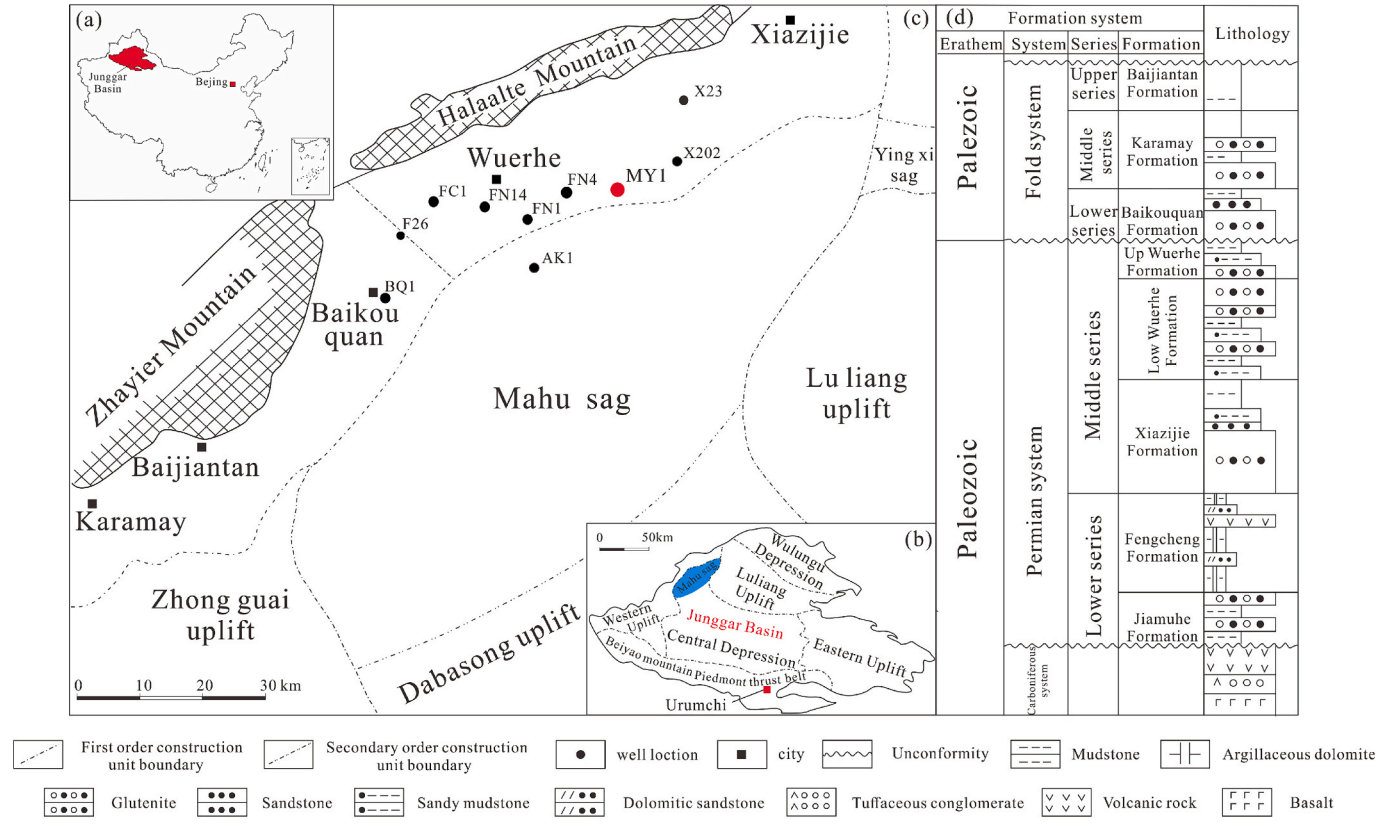


Fig. 3. (a) Structural map showing the location of Junggar Basin. (b) Tectonic map of the Junggar Basin showing the location of the Mahu Sag. (c) The geological structures and well locations near the Mahu Sag (d) Stratigraphic columns of the Permian System in Paleozoic erathem and Triassic in Mesozoic erathem.

to semi-arid, hypersaline reducing environment, characterized by well-developed laminations and pronounced rhythmic bedding [53], reflecting complex sedimentary conditions that present significant challenges for reservoir characterization (Fig. 3).

4. Results and discussion

4.1. Lithofacies types and distribution characteristics

The shale facies development within the Fengcheng Formation of the Mahu Depression exhibits considerable diversity. This study utilizes whole-rock analyses from well MY-1 as a case example. The results indicate that feldspathic quartz predominates among the mineral constituents of the Fengcheng Formation, followed by carbonate minerals. The feldspar content varied between 28.2 % and 91.0 %, with an average of 56.93 %. The carbonate mineral content ranged from 3.0 % to 64.6 %, averaging 26.77 %. Meanwhile, the clay mineral content ranged from 1.0 % to 25.0 %, with a mean value of 7.84 % (Fig. 4). Utilizing the tri-terminal meta-classification method, the lithology of the Fengcheng Formation was systematically categorized into four primary groups and sixteen subgroups based on the interrelationships among feldspathic, carbonate, and clay minerals. The 16 sub-lithofacies are listed below: Felsic shale lithofacies, Mixed-felsic shale lithofacies, Calcareous-felsic shale lithofacies, Clayey-felsic shale lithofacies, Calcareous shale lithofacies, Mixed-calcareous shale lithofacies, Felsic-calcareous shale lithofacies, Clayey-calcareous shale lithofacies, Carbonate/clay mixed shale lithofacies, Mixed shale lithofacies, Carbonate/felsic mixed shale lithofacies, Clay/felsic mixed shale lithofacies, Clayey shale lithofacies, Mixed-clayey shale lithofacies, Felsic-clayey shale lithofacies, Calcareous-clayey shale lithofacies. Utilizing the comprehensive whole-rock data analyzed, the principal lithofacies characterized within the Windy City Formation comprise Felsic shale lithofacies, Mixed-felsic shale lithofacies, Calcareous-felsic shale lithofacies, Carbonate/felsic mixed shale lithofacies, and Felsic-calcareous shale lithofacies, with Mixed-felsic shale lithofacies and Calcareous-felsic shale lithofacies representing the predominant volumetric contributions (Fig. 4).

In the Fengcheng Formation, the Felsic-calcareous shale and Calcareous-felsic shale lithofacies exhibit laminated structures, characterized by the presence of calcareous and feldspathic bands. Felsic shale lithofacies samples primarily consist of quartz and feldspar, as the low calcium content results in an absence of distinct layered structures. The mixed-felsic shale lithofacies phase, while containing a small amount of calcium, exhibits a dispersed distribution that precludes the formation of distinct bands. The Carbonate-felsic mixed shale lithofacies exhibits faint banding, which is not pronounced (Fig. 5).

In identifying and predicting lithology for well MY-1, this study primarily employs the results of reservoir mineral fractions derived from elemental logging analyses. By leveraging the Logistic-Bayes-A-IGWO-Bagging-BP regression model delineated in section 2.1, alongside empirical XRD data, we undertook a meticulous reassessment of locations within well MY-1 where elemental logging calculations demonstrated inconsistencies, ultimately yielding highly refined and intricate characteristics of petrographic distribution. The comprehensive assessment of the test set outcomes demonstrates that predictions produced by the regression model exhibit significantly improved accuracy. As depicted in Fig. 6, the shale facies of the Fengcheng Formation reveal pronounced stratigraphic heterogeneity along the vertical axis.

Within the comprehensive stratigraphic framework of the single well, the Fengcheng Formation predominantly encompasses Mixed-felsic shale lithofacies and Calcareous-felsic shale lithofacies, reflecting significant geological characteristics. In the P_1f^3, Mixed-felsic shale lithofacies dominance is distinctly evident. The calcareous-felsic shale lithofacies predominate, while the carbonate/felsic mixed shale lithofacies develop to a lesser degree in the lower-middle sections of P_1f^2 and P_1f^1. In the lower-middle segment of the P_1f^2, Felsic-calcareous shale lithofacies are predominantly observed. Within the central stratigraphic interval of the Fengcheng Formation, Felsic shale lithofacies are comparatively underrepresented and predominantly manifest as

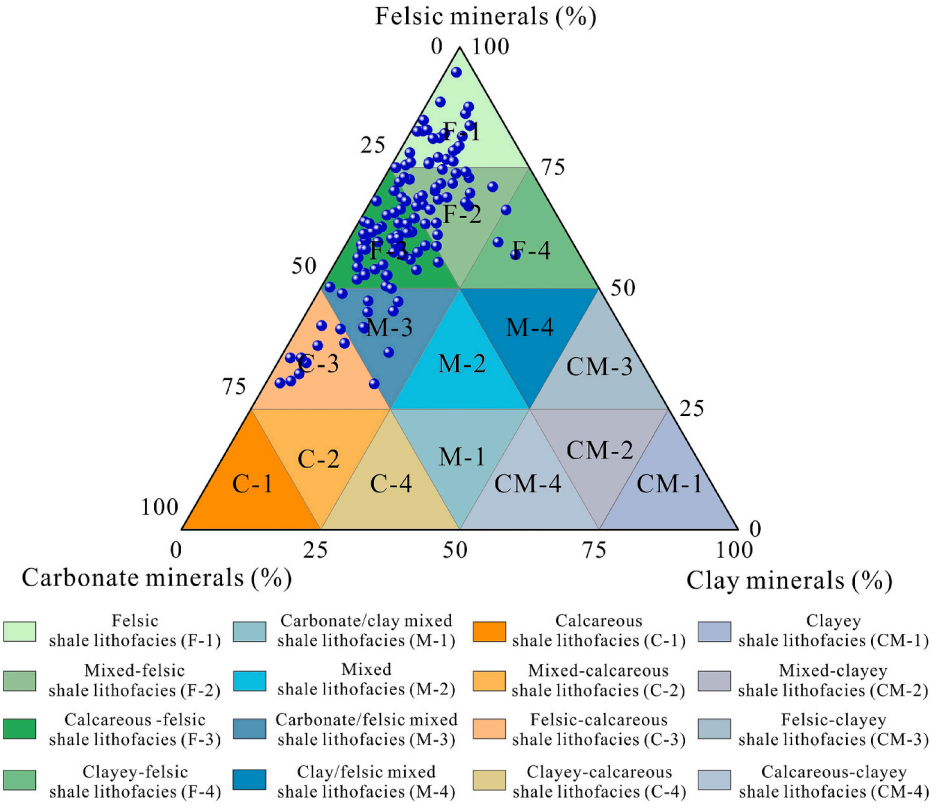


Fig. 4. Ternary lithofacies classification of the Fengcheng Formation shale based on Felsic-Carbonate-Clay mineral assemblages.

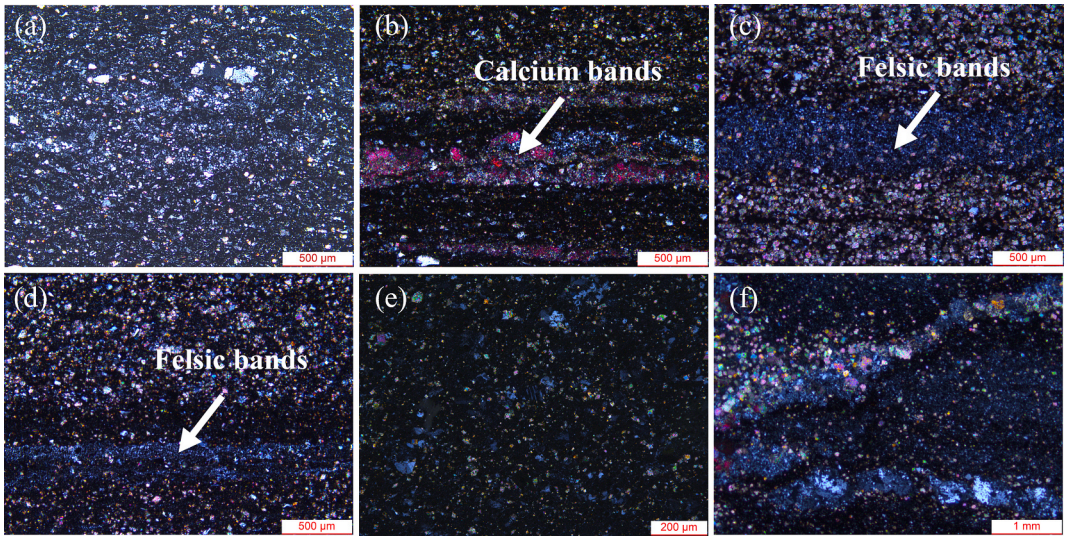


Fig. 5. Sedimentary structure characteristics of shale lithofacies in Fengcheng Formation(a) Felsic shale lithofacies (F-1); (b) Calcareous-felsic shale lithofacies (F-3); (c) Felsic-calcareous shale lithofacies (C-3); (d) Felsic-calcareous shale lithofacies (e) Mixed-felsic shale lithofacies (F-2); (f) Carbonate/felsic mixed shale lithofacies (M-3).

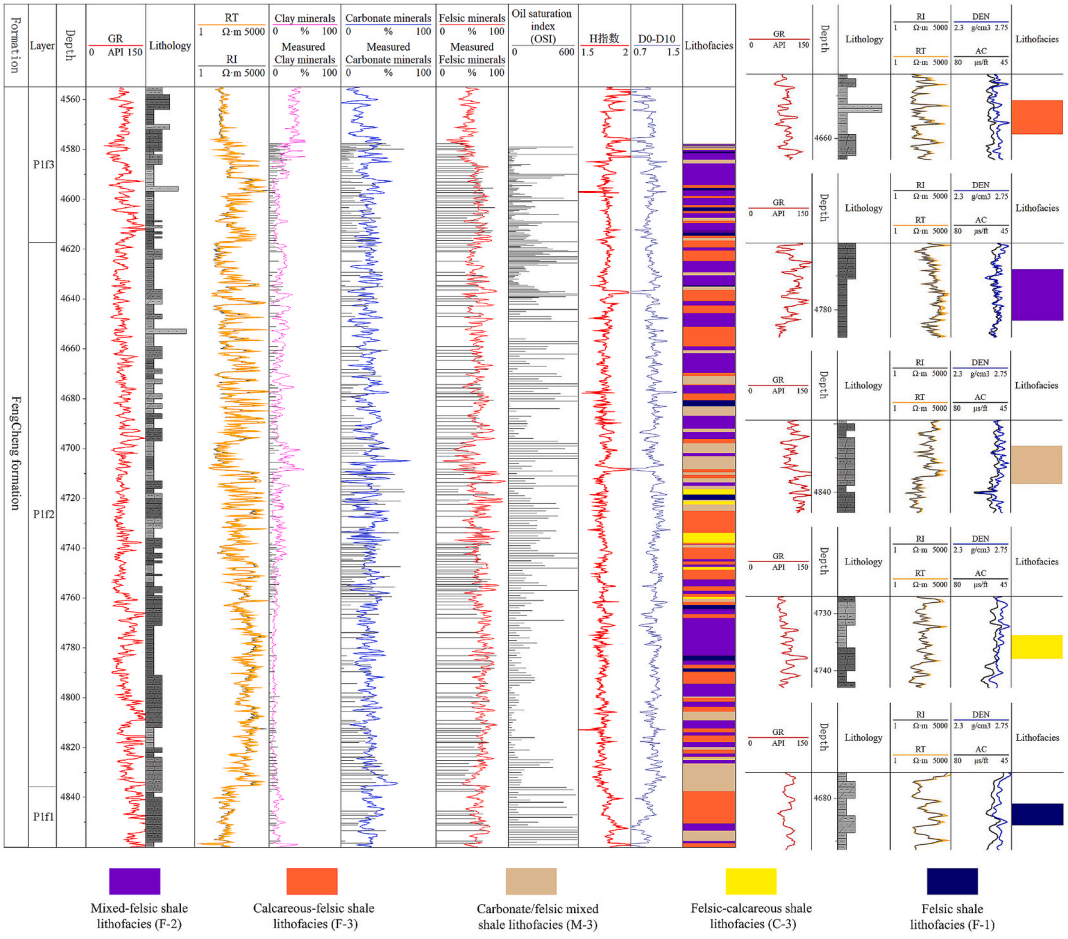


Fig. 6. Lithofacies distribution in MY-1 Well predicted by Bayesian-IGWO-optimized Bagging-BP model.

intercalations characterized by a heterogeneous spatial distribution.

4.2. *Characterization of reservoir space in different lithofacies*

The intergranular porosity within the Fengcheng Formation primarily arises from the contraction of pore spaces between mineral particles during mechanical compaction and residual pore spaces that persist following deposition and diagenesis. These intergranular pores can be categorized into primary and secondary types based on their diagenetic history. The Fengcheng Formation predominantly features

primary intergranular and relatively well-developed dissolution pores within the grains. However, while localized occurrences of organic matter pores are noted, their overall development remains limited. Furthermore, the pore structure exhibits variability across different lithofacies.

The felsic shale lithofacies, subjected to significant burial depths, experience marked compaction within the reservoir space, predominantly exhibiting slit-like intergranular pores as illustrated in their pore architecture (Fig. 7a). The calcareous-felsic shale lithofacies are distinctly characterized by wedge-shaped and irregular secondary intergranular pores alongside a relatively limited development of intragranular dissolution pores (Fig. 7b, c). In the mixed-felsic shale lithofacies, intragranular dissolution pores are prominently developed, with localized occurrences of clay minerals; the intercrystalline pores associated with these minerals can be observed within these formations (Fig. 6d). The felsic-calcareous shale lithofacies demonstrate a relative enrichment of carbonate minerals within the matrix, highlighted by ovoid and reniform intergranular pores formed through dissolution processes. Organic porosity primarily manifests as reniform organic pores, some partially filled with bitumen (Fig. 7 e-h). The carbonate/felsic mixed shale lithofacies exhibit lower quantities of feldspar and carbonate minerals but show a higher relative abundance of clay minerals within the reservoir; local intercrystalline pores associated with clay minerals are also visible (Fig. 7i).

The pore size distribution within the shale reservoirs of the Fengcheng Formation exhibits notable heterogeneity across diverse lithofacies, and this variability persists even within uniform lithofacies. As delineated in Fig. 8, the pore sizes across the lithofacies of the Fengcheng Formation are predominantly concentrated between 5 nm and 1 μm. Specifically, the feldspathic shale and calcareous feldspathic shale lithofacies exhibit an unimodal distribution of pore sizes, unlike other lithofacies, which present a bimodal distribution pattern. The primary peak is associated with pore sizes ranging from 5 to 100 nm within the bimodal distribution. In contrast, a secondary peak extends from 100 nm to 10 μm, with a transitional zone observed between 100 nm and 1 μm. Categorized following the pore development characteristics of the Fengcheng Formation, pores are classified into three distinct categories: micropores (10–100 nm), mesopores (100 nm–1 μm), and macropores (>1μm). Notably, the feldspathic and calcareous feldspathic shale lithofacies are distinguished by a higher prevalence of mesopores than micropores and macropores. In contrast, the remaining lithofacies are predominantly characterized by micropores, with less pronounced development of mesopores. Even within the mixed feldspathic shale lithofacies, discernible variations in pore size distribution are evident; these discrepancies may be ascribed to the intraformational heterogeneity inherent within the same lithofacies.

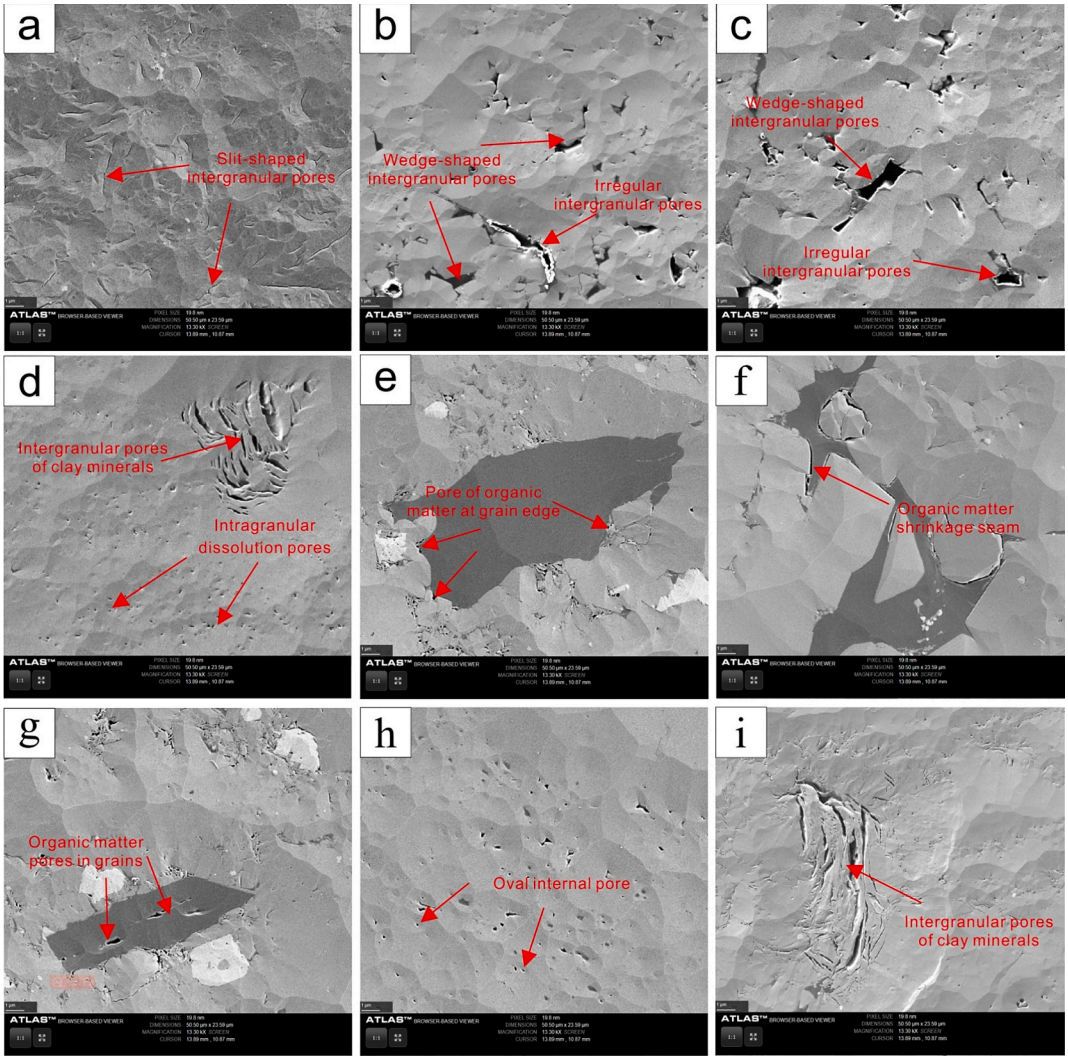


Fig. 7. The types of pores that occur in different lithofacies (a) The felsic shale lithofacies; (b, c)The calcareous-felsic shale lithofacies; (d) The mixed-felsic shale lithofacies; (e,h) The felsic-calcareous shale lithofacies; (i) The carbonate/felsic mixed shale lithofacies.

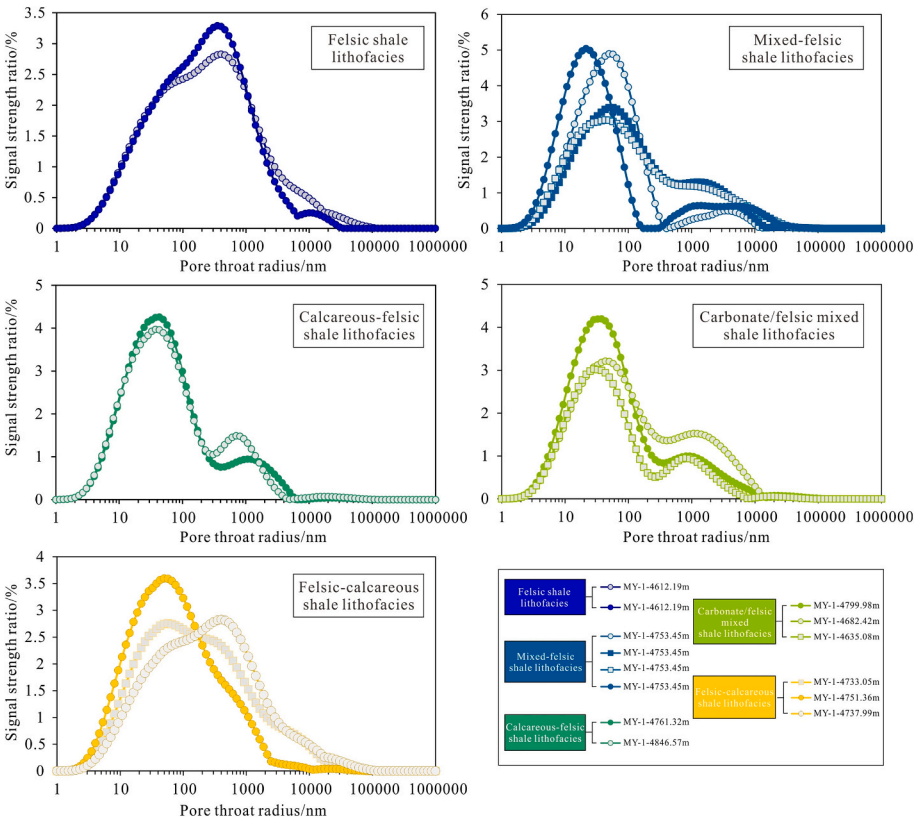


Fig. 8. Pore size distribution of different lithologies in the Fengcheng Formation.

4.3. *Oil-bearing characteristics of different lithofacies*

The TOC content of shale samples from the Fengcheng Formation varied between 0.17 and 1.59 %, with an average of 0.69 %. An elevated calcium concentration in shale lithology positively correlates with the TOC concentration within its reservoir. Shale phases with high carbonate mineral content also have relatively high TOC content. In contrast, lithologies characterized by a relatively high felsic and clay minerals content tend to have lower internal TOC levels (Fig. 9a). This study rigorously examines the analysis of free and adsorbed oils by employing hydrocarbon fractions derived from multi-temperature step pyrolysis experiments conducted at precisely defined temperature

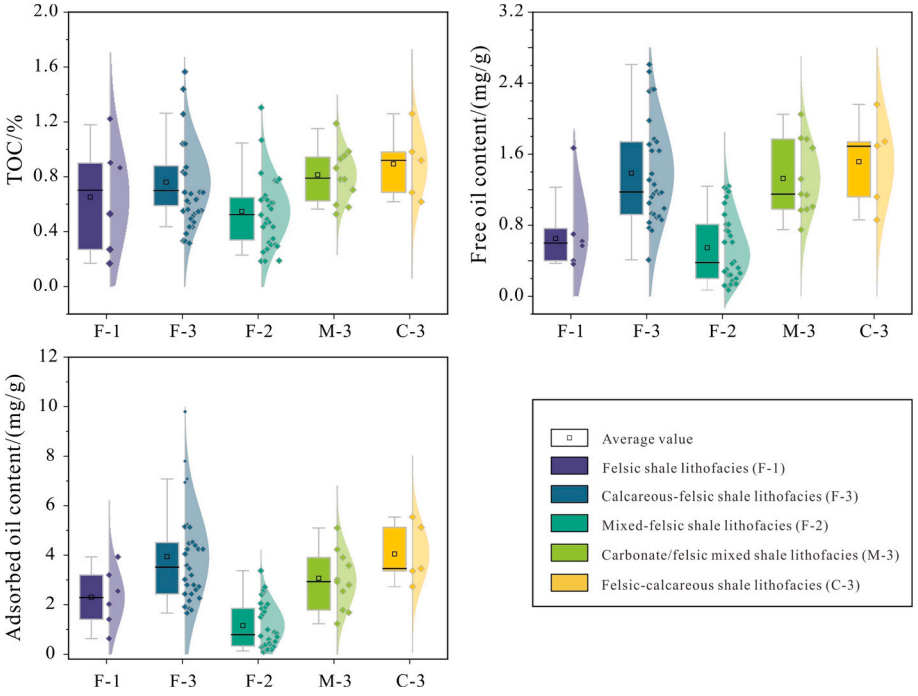


Fig. 9. Distribution characteristics of TOC, free oil and adsorbed oil in different lithofacies.

intervals: $S_{1\text{-}1}$, $S_{1\text{-}2}$, $S_{2\text{-}1}$, and $S_{2\text{-}2}$. The aggregate of $S_{1\text{-}1}$ and $S_{1\text{-}2}$ is classified as the free oil fraction, whereas the aggregate of $S_{2\text{-}1}$ and $S_{2\text{-}2}$ is categorized as the adsorbed oil fraction. The concentration of free oil within the shale lithofacies of the Fengcheng Formation is delineated by a range from 0.07 to 2.61, with an average of 1.03; in contrast, the concentration of adsorbed oil is quantified between 0.31 and 9.8, resulting in an average value of 2.67. Among the five shale lithofacies, free oil is mainly found in Calcareous-felsic shale lithofacies, Carbonate/felsic mixed shale lithofacies and Felsic-calcareous shale lithofacies. Free oil content is comparatively low in both Felsic shale lithofacies and Mixed-felsic shale lithofacies (Fig. 9b). Adsorbed oils exhibit characteristics similar to free oils and serve as the primary reservoir for adsorbed oils in lithofacies with high carbonate mineral content (Fig. 9c). The abundance of adsorbed oil is comparatively lower in Felsic shale lithofacies and Mixed-felsic shale lithofacies.

The oil content calculation formula formulated for various pore sizes across distinct petrographic phases elucidates pronounced disparities in pore size distribution, both among samples of different petrographic phases and within samples of the same phase. This heterogeneity in pore size distribution subsequently governs the spatial allocation of free and adsorbed oil within these petrographic matrices. The shale oil content is markedly enhanced in regions where the pore size exhibits significant development within the lithology. In contrast, it is substantially reduced in areas characterized by poorly developed pore spaces (Fig. 10).

Lithofacies with high calcium content, including Calcareous-felsic shale lithofacies, Carbonate/felsic mixed shale lithofacies, and Felsic-calcareous shale lithofacies, possess elevated shale oil content. The primary reservoir space for shale oil is within the pore size range of 10–1000 nm, with peak oil content occurring at an optimal pore diameter of approximately 50 nm. As depicted in Fig. 10, the overall concentration of free oil remains relatively low, while the shale reservoir exhibits a comparatively high content of adsorbed oil. The characteristics of free oil and adsorbed oil occurrence exhibit congruence with the pore structure features inherent to the petrographic phase, both

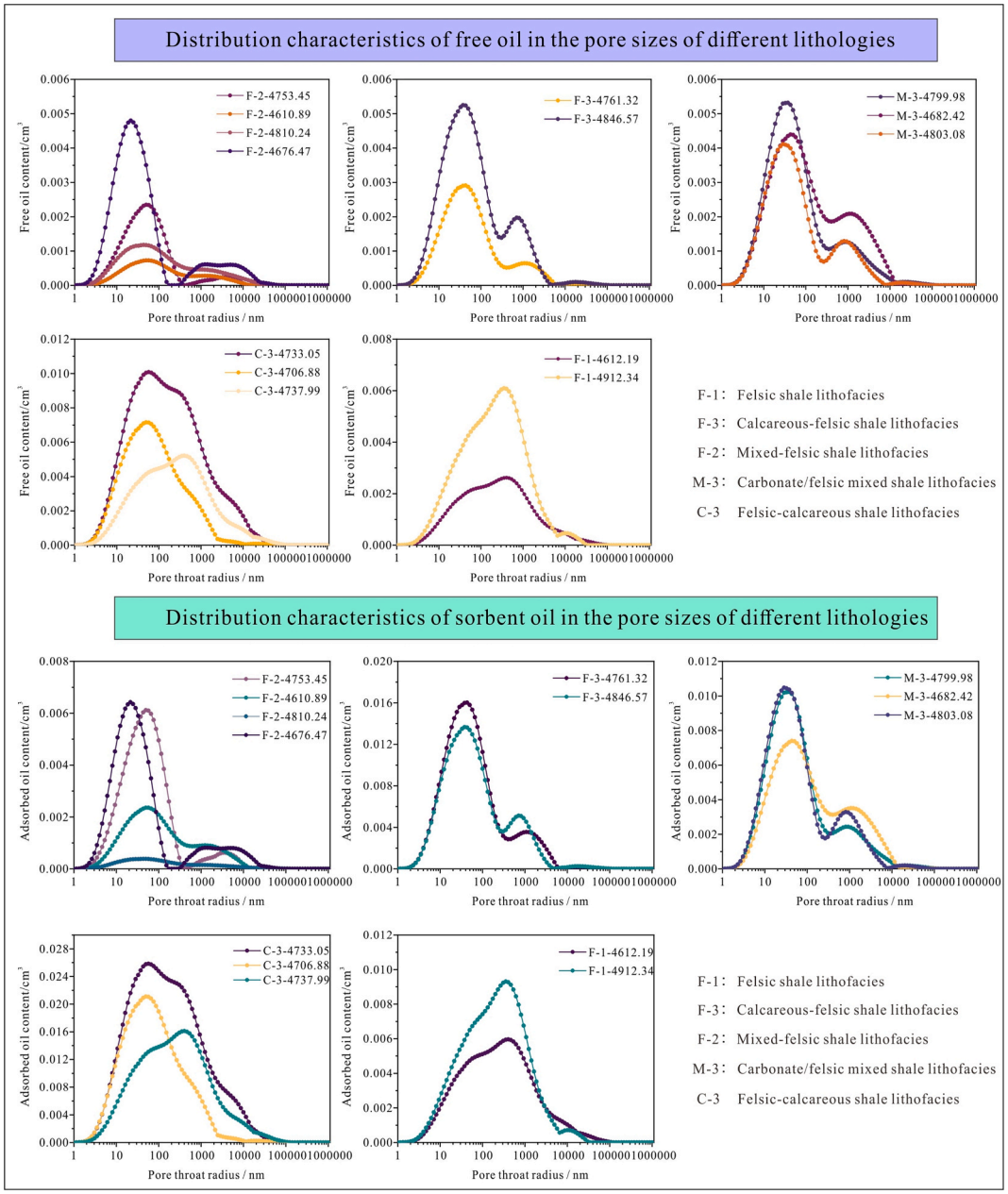


Fig. 10. Oil-bearing distribution characteristics of different lithofacies and different pore sizes.

revealing a propensity for more developed pore regions to correlate with elevated oil content: specifically, the maximum free oil concentration ranges from 0.005 to 0.01 cm^3, whereas the peak bound oil concentration spans from 0.01 to 0.025 cm^3. There is a notable reduction in overall shale oil content for lithofacies characterized by low calcium content, specifically the mixed-felsic shale lithofacies and felsic shale lithofacies. The primary reservoir space for shale oil in the mixed-felsic shale lithofacies is characterized by pore diameters of 10–100 nm. Additionally, shale oil is present in pores measuring 1–10 μm, but its concentration remains relatively low. The pore size distribution in the mixed-felsic shale lithofacies exhibits a broad range, with the primary reservoir space for shale oil spanning from 50 to 5000 nm. Nevertheless, shale oil concentration is significantly lower than that found in calcareous shale lithofacies.

4.4. The reliability analysis of artificial intelligence prediction models

The AI-based lithofacies identification method established in Section 2.1 successfully predicts vertical lithofacies distribution in individual wells, but requires further validation of model stability and accuracy. Performance analysis reveals significant differences in prediction capability across mineralogical classes: clay minerals show optimal fitting during training (MAE = 0.006, RMSE = 0.017, R^2 = 0.988), followed by carbonates (MAE = 0.013, RMSE = 0.028, R^2 = 0.976), while felsic minerals exhibit relatively weaker performance (MAE = 0.020, RMSE = 0.053, R^2 = 0.929), suggesting either greater complexity in felsic mineral signatures or insufficient capture of nonlinear relationships (Table 2).

During testing, the model demonstrates good stability and generalization capacity, with carbonates maintaining the most robust performance (minimal error increase), clay minerals showing moderate degradation, and felsic minerals presenting the most significant generalization challenges—indicating potential discrepancies between training and test data distributions. These results confirm the model's overall reliability for lithofacies characterization while highlighting felsic mineral prediction as the key target for future optimization through expanded training datasets or enhanced feature engineering to better capture geological variability.

As evidenced in Fig. 11, the prediction results for all three mineral types demonstrate generally favorable accuracy, with minimal deviation between predicted and actual values. The felsic mineral predictions show isolated outliers near the 30 % concentration range in the training set, suggesting potential data collection anomalies in this interval, while systematic deviations of multiple data points from the baseline in the 50–70 % mid-to-high concentration range of the test set indicate inadequate capture of nonlinear responses, resulting in significantly lower test R^2 (0.929) compared to training performance (0.988).

Carbonate minerals exhibit intermediate performance, with high slope values for both training and test regression lines, though slight deviations from the baseline occur in the 20–40 % mid-concentration range of the test set, likely attributable to localized data distribution differences. Clay minerals show excellent training set agreement (R2 = 0.988) but notable performance degradation in testing (R2 = 0.854), particularly evidenced by wider scatter in the 4–16 % concentration range, potentially due to introduced noise or cross-interference from mixed mineral compositions in the test data. Despite these observed deviations, the overall prediction accuracy remains sufficiently robust for practical lithofacies identification applications, with all mineral types maintaining acceptable error margins (MAE < 0.020 for felsics, <0.013 for carbonates, and <0.006 for clays in training). These results validate the model's operational reliability while highlighting specific concentration ranges requiring potential refinement through additional data sampling or feature engineering to address the identified nonlinear response characteristics.

Table 2
Performance evaluation indices of neural network models.

Data	Metrics	Felsic minerals	Carbonate minerals	Clay minerals
Training	MAE	0.020	0.013	0.006
	RMSE	0.053	0.028	0.017
	R^2	0.9289	0.9764	0.9883
Testing	MAE	0.052	0.072	0.066
	RMSE	0.067	0.090	0.082
	R^2	0.8287	0.8767	0.8535

4.5. Controls on vertical lithofacies variability

The adaptive AI prediction model developed in Section 2.1 (Fig. 6) reveals significant vertical lithofacies variations within individual wells. These observed differences primarily originate from paleoenvironmental shifts during Fengcheng Formation deposition, where changing sediment provenance and associated mineralogical transformations induced progressive vertical lithofacies changes [54]. To systematically interpret this heterogeneity, we employ elemental proxies as depositional environment indicators to explain the lithofacies distribution patterns.

The Fengcheng Formation predominantly developed under suboxic to anoxic conditions throughout its depositional history [55]. Stratigraphic thickness analysis identifies the second member (P_1f_2) as the primary interval of interest. During P_1f_2 deposition, systematic environmental changes occurred: paleo-water depth progressively decreased (La), water salinity increased (Sr/Ba), dissolved oxygen content declined (V/Cr), culminating in fully reducing conditions (Fig. 12). The P_1f_2 interval subsequently experienced oxygen replenishment, transitioning to oxic conditions coinciding with peak biological productivity (Al/Ba). This combination of elevated primary productivity and oxygen availability enhanced both hydrocarbon source potential and carbonate mineral content within P_1f_2 reservoirs, ultimately increasing TOC accumulation. Consequently, calcareous feldspathic shales and carbonate/feldspar mixed shales emerged as the dominant lithofacies (Fig. 12), while other lithologies formed through fluctuating paleo-water depth, salinity, and redox conditions. These variations resulted from seasonal environmental changes coupled with dynamic lake-level fluctuations and water mass energy variations during deposition [56].

4.6. Controls on oil-bearing property variations in lithofacies

The quantitative oil-bearing characterization model established in Section 2.3 (Fig. 10) reveals significant intra-lithofacies variations in hydrocarbon saturation. These observed differences primarily stem from pore-size distribution heterogeneity within individual lithofacies units, where subtle variations in pore throat radii (50–800 nm) and connectivity create distinct fluid storage and migration pathways [57]. This section examines oil content differences within equivalent lithofacies, using the feldspar-calcareous shale facies as an example. Mineral content analysis of five samples from the same lithofacies but different depths reveals distinct variations: differences exist between the same mineral types within the lithofacies, while variations between different minerals are even more pronounced (Fig. 13).

To investigate these observed differences, we conducted Ropscan mineral analysis on a sample from 4717.4 m depth, revealing non-uniform vertical distribution of individual minerals (Fig. 14). This mineralogical heterogeneity directly influences pore structure distribution along the vertical profile [58]. Detailed examination of a selected area through superimposed pore-mineral mapping clearly demonstrates distinct pore architectures between different mineral boundaries. The preservation of pore structures varies significantly depending on mineral mechanical properties – brittle minerals maintain better pore structure integrity, while ductile minerals exhibit poorer pore preservation [59]. These two key observations explain the oil content

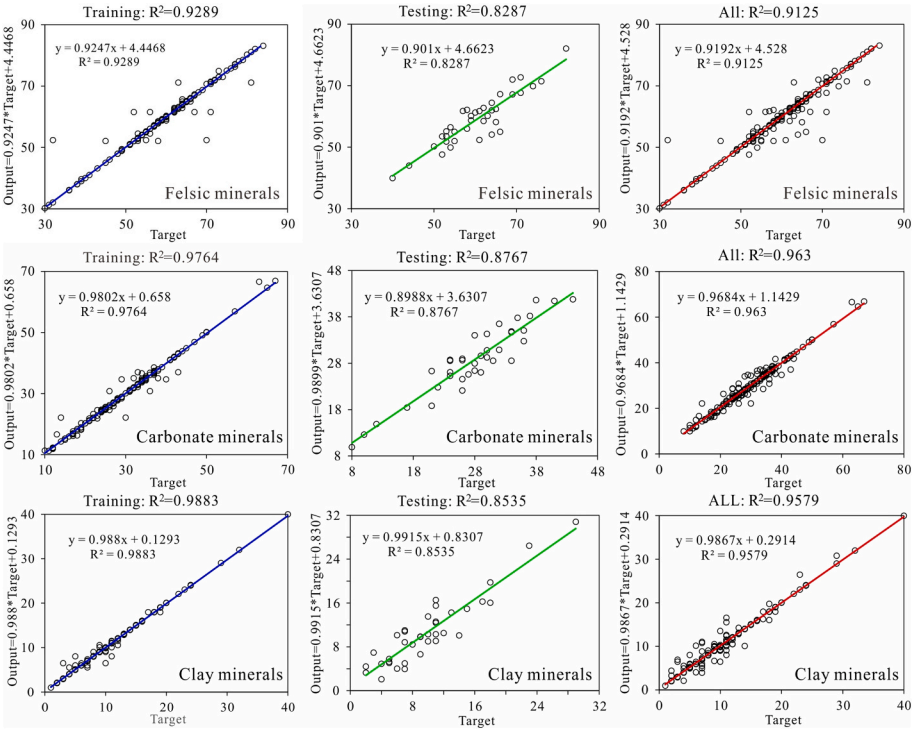


Fig. 11. Regression Analysis of Predicted and Actual Mineral Content for Felsic, Carbonate and Clay Minerals.

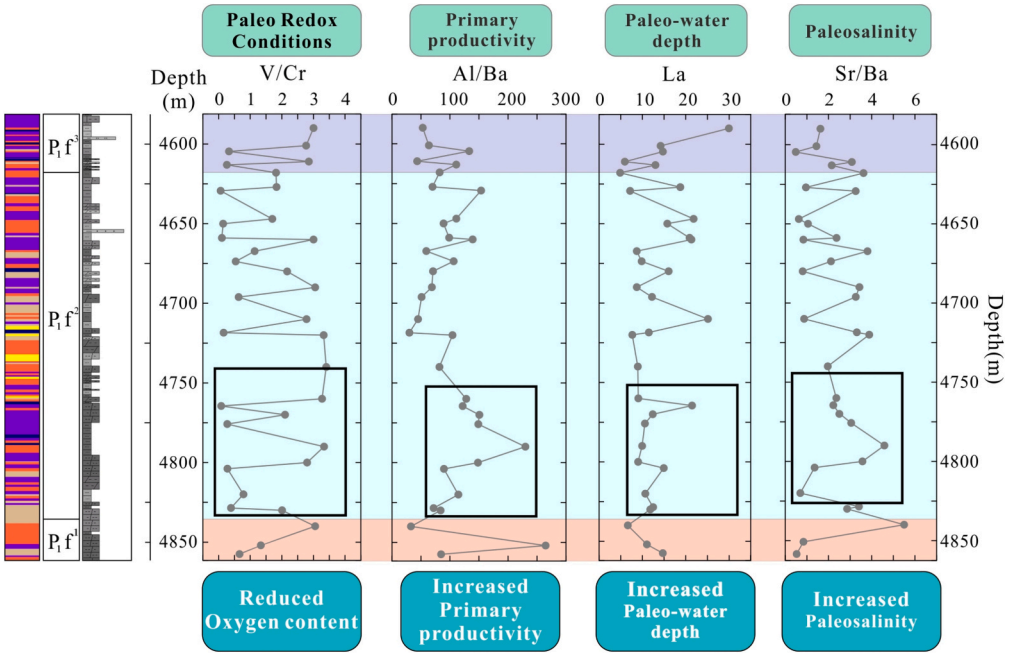


Fig. 12. Element Distribution of Different Lithofacies Reflecting Depositional Environment in Well MY-1.

variations within the same lithofacies: (1) differences in mineral composition percentages and (2) spatial distribution patterns of minerals collectively serve as indirect but critical factors causing oil saturation differences at varying depths within identical lithologies.

The calculated results presented in Fig. 10 reveal significant differences in oil-bearing properties among various lithofacies. These variations arise from two primary factors: (1) compositional differences in mineral components between lithofacies, which affect their respective capacities for storing free versus adsorbed oil, and (2) variations in pore size distribution and heterogeneity levels within each lithofacies. We analyze these two aspects separately by examining the relationship between mineral composition and oil content, and the correlation between multifractal dimension theory and oil-bearing characteristics.

The analysis of the relationship between carbonate minerals and both free and adsorbed oil across various lithofacies demonstrates a positive correlation. As the concentration of carbonate minerals increases, there is a corresponding and significant rise in free and adsorbed oil levels within these lithofacies. Shale lithofacies with elevated calcium content primarily characterize the P_1f^2 stratigraphic interval and exhibit a high TOC concentration within its reservoirs [60].

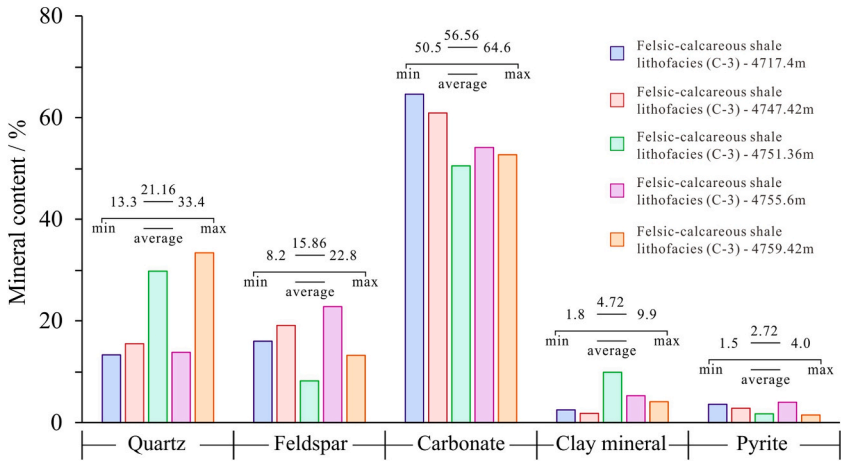


Fig. 13. Distribution of mineral components across varying depths within the same lithofacies.

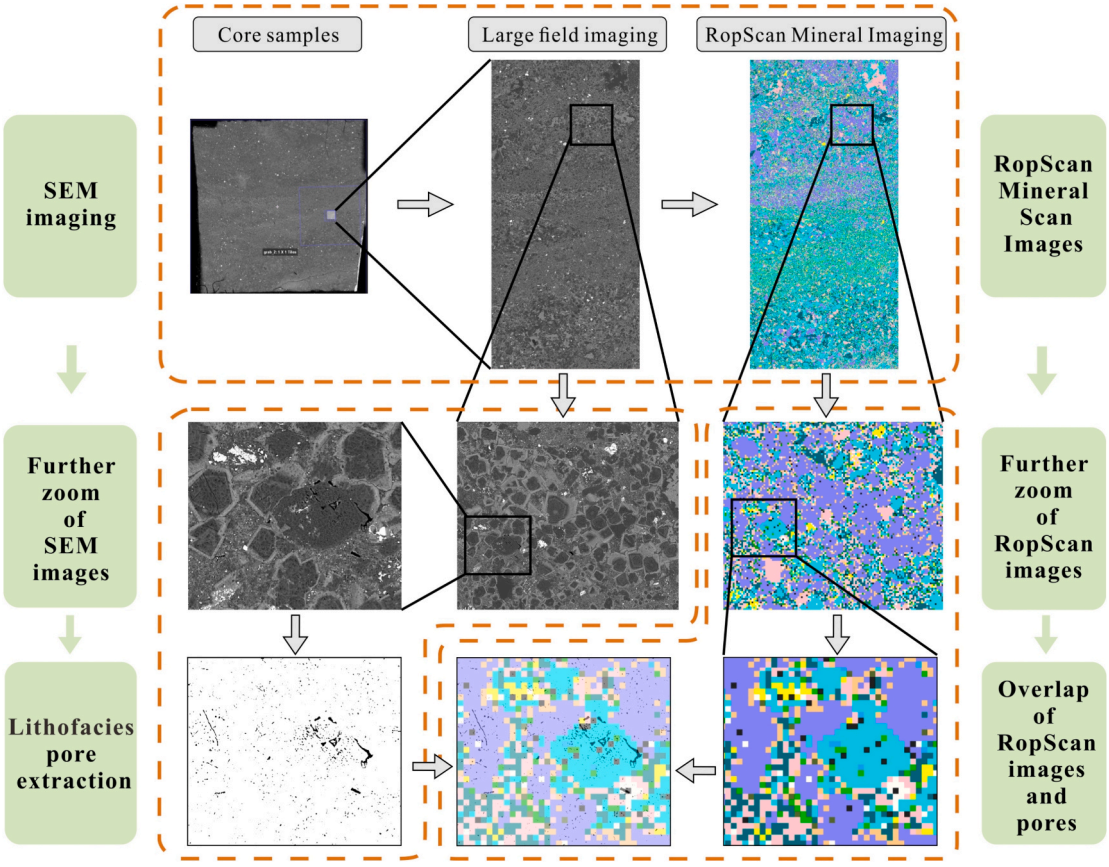


Fig. 14. Element Distribution of Different Lithofacies Reflecting Depositional Environment in Well MY-1.

Consequently, the lithofacies enriched in calcium demonstrate a robust hydrocarbon-generating potential. On one hand, lithofacies exhibiting significant hydrocarbon-generating potential are characterized by relatively elevated oil content within their reservoirs. On the other hand, heightened concentrations of carbonate minerals facilitate the release of organic acids during the maturation phase of organic matter and simultaneously accelerate the dissolution of these minerals during periods of deep fluid activity [61]. This process substantially enhances the porosity of carbonate minerals and markedly contributes to hydrocarbon concentration (Fig. 15).

Quartz, esteemed for its stability and brittleness, exerts a negligible influence on reducing porosity within the reservoir. Instead, it primarily upholds the integrity of pore space by providing critical structural support [62]. The correlation analysis between quartz content and free oil indicates that the impact of quartz on free oil can be distinctly classified into two distinct phases. In the initial phase, when quartz content is below 30 %, it inhibits the presence of free oil. The inhibitory effect is primarily attributed to the low concentration of quartz, which provides minimal structural support to the granular framework [63]. Within a single sample, a diminished quartz particle concentration correlates with a proportionally elevated content of clay minerals and other minor constituents. The excessive clay mineral content can lead to pore constriction, consequently causing a decrement in reservoir porosity and a concomitant reduction in permeability. In the second

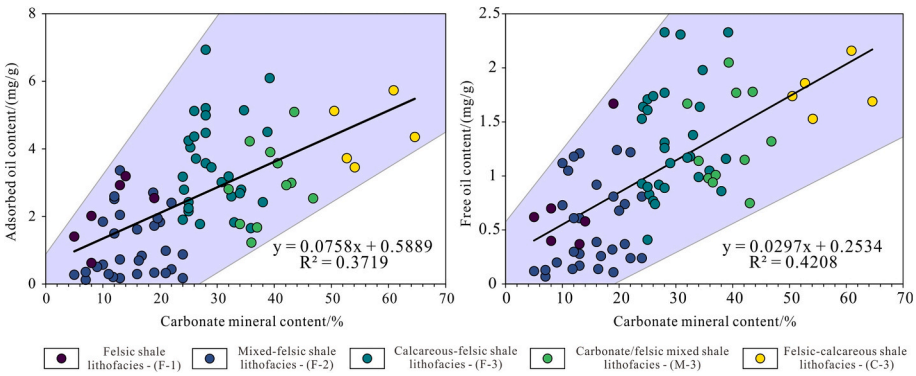


Fig. 15. Correlation between carbonate mineral content and free oil and adsorbed oil.

phase, where the quartz content surpasses 30 %, it exerts a facilitative effect on the retention of free oil within the reservoir. With the increment in quartz particle concentration, their contribution to the structural integrity of the reservoir framework is enhanced, consequently attenuating the impact of clay minerals. In contrast to lithofacies enriched in carbonate minerals, those with elevated quartz content, exemplified by the Felsic shale lithofacies, exhibit a comparatively reduced hydrocarbon-generating potential. The shale oil within these reservoirs is hypothesized to have been generated within carbonate lithofacies and then migrated over short distances to the quartz-rich lithofacies (Fig. 16).

The data in Fig. 16 reveals an inverse relationship between quartz and feldspar concentrations within the feldspathic lithologies of the Fengcheng Formation. Notably, elevated feldspar concentrations are inversely associated with diminished quartz levels, and this relationship is reciprocal. This study on the relationship between feldspar content and the occurrence of free hydrocarbons revealed that a significant threshold of feldspar content substantially affects the distribution of free hydrocarbons. Empirical data indicate that feldspar content exerts a limited effect on the sequestration of free oil when it is below the 30 % threshold. This phenomenon can be attributed to suboptimal feldspar levels, which lead to increased quartz content within the reservoir following the principle of inverse proportionality. As a result, the impact of quartz on free hydrocarbons becomes predominant. Conversely, an enhancement in the feldspar content beyond the 30 % threshold is associated with a marked increase in the quantity of free hydrocarbons. One contributing factor is the enhanced compaction associated with increased burial depth during the process of the diagenesis of Fengcheng rocks [64]. This compaction, when quartz content is low, results in a decrease in porosity. Nevertheless, reduced quartz content is counterbalanced by an elevated feldspar content due to the compensatory relationship. The higher feldspar content alleviates the effects of compaction, thus preserving a portion of the porosity. On the other hand, the sedimentary environment of the Fengcheng Formation is

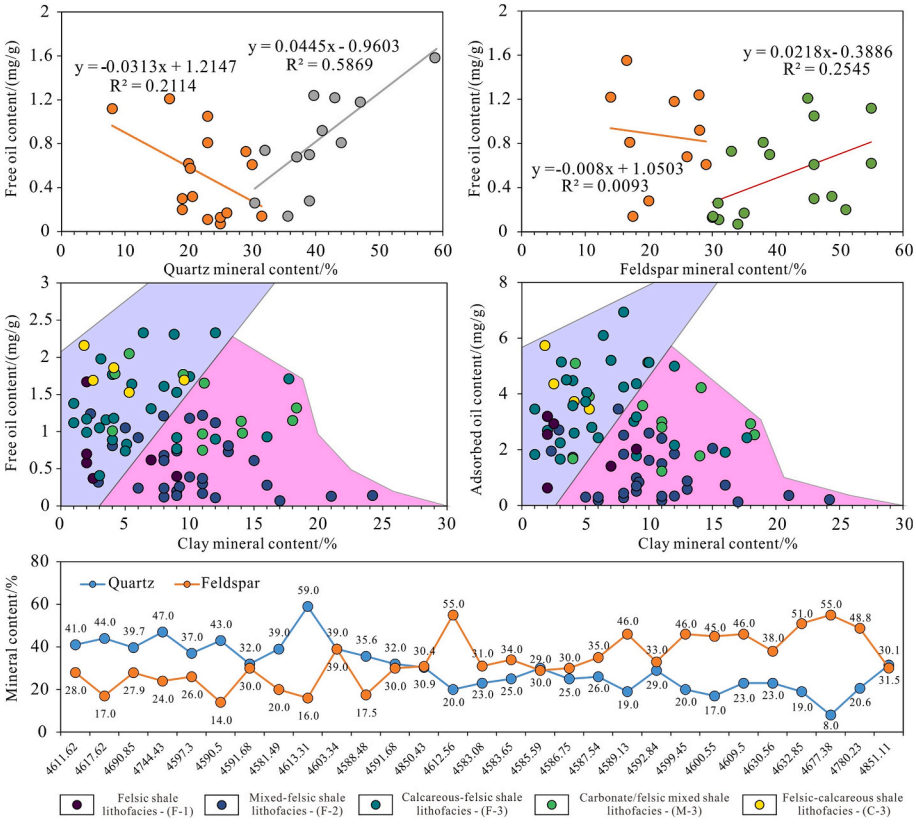


Fig. 16. Correlation between mineral composition and free oil and adsorbed oil.

characterized as a saline water environment [65]. During the maturation of hydrocarbons within source rocks, organic acids are liberated. These acids engage in complexation with metal cations within the feldspar, thereby catalyzing the dissolution of feldspar. The dissolution-induced porosity that emerges is beneficial for the sequestration of free hydrocarbons.

The relationship between the concentration of clay minerals and the quantities of free and bound oil within the reservoir demonstrates a distinct segmented correlation. As the concentration of clay minerals falls below 10 %, there is a corresponding increase in free and bound oil levels, which correlates with the gradual rise in clay mineral content. This phenomenon can be ascribed to the incomplete sealing of pore spaces at reduced clay mineral concentrations, in conjunction with the heightened adsorptive affinity of clay minerals for shale oil relative to other mineral species [66]. Consequently, the increased concentration of clay minerals significantly enhances their adsorptive affinity for shale oil, thereby increasing the total oil content within the reservoir. Conversely, beyond a threshold of 10 % clay mineral concentration, the quantities of free and bound oil diminish as the concentration of clay minerals rises. The phenomenon is primarily driven by the elevated concentrations of clay minerals, which induce the formation of blockages within the shale reservoir matrix. This obstruction diminishes the pore space, significantly reducing reservoir porosity and permeability. The blockage induced by high clay mineral concentrations within the reservoir matrix disrupts the internal connectivity, attenuating the shale oil's enrichment capacity.

This study employs multi-fractal dimension analysis to investigate the influence of lithofacies pore size variability on the oil-bearing capacity of shales, with a focus on 10 representative samples from the five principal lithofacies of the Fengcheng Formation. Fig. 17 illustrates the double logarithmic plot of the logarithmic mass distribution function u (q, e) against the cell side length C for a quintessential sample within each of the five lithofacies, delineating their logarithmic correlation. The statistical moment order q ranges from 10 to −10, with an incremental step of 1. The analysis delineates that within the q range of 0 to 10, log u(q,e) escalates with the escalation of loge, while within the q range of −10 to 0, log u(q,e) corresponds with a decrease in loge. Throughout the q spectrum from −10 to 10, the scatter plots manifest a pronounced linear correlation, a trend that is similarly exhibited by the additional nine samples (Fig. 17). This regularity substantiates the robust multifractal characteristics within samples of the five lithofacies [67,68].

Samples from diverse lithofacies within the Fengcheng Formation exhibit multifractal behaviour, underscoring the need for a thorough

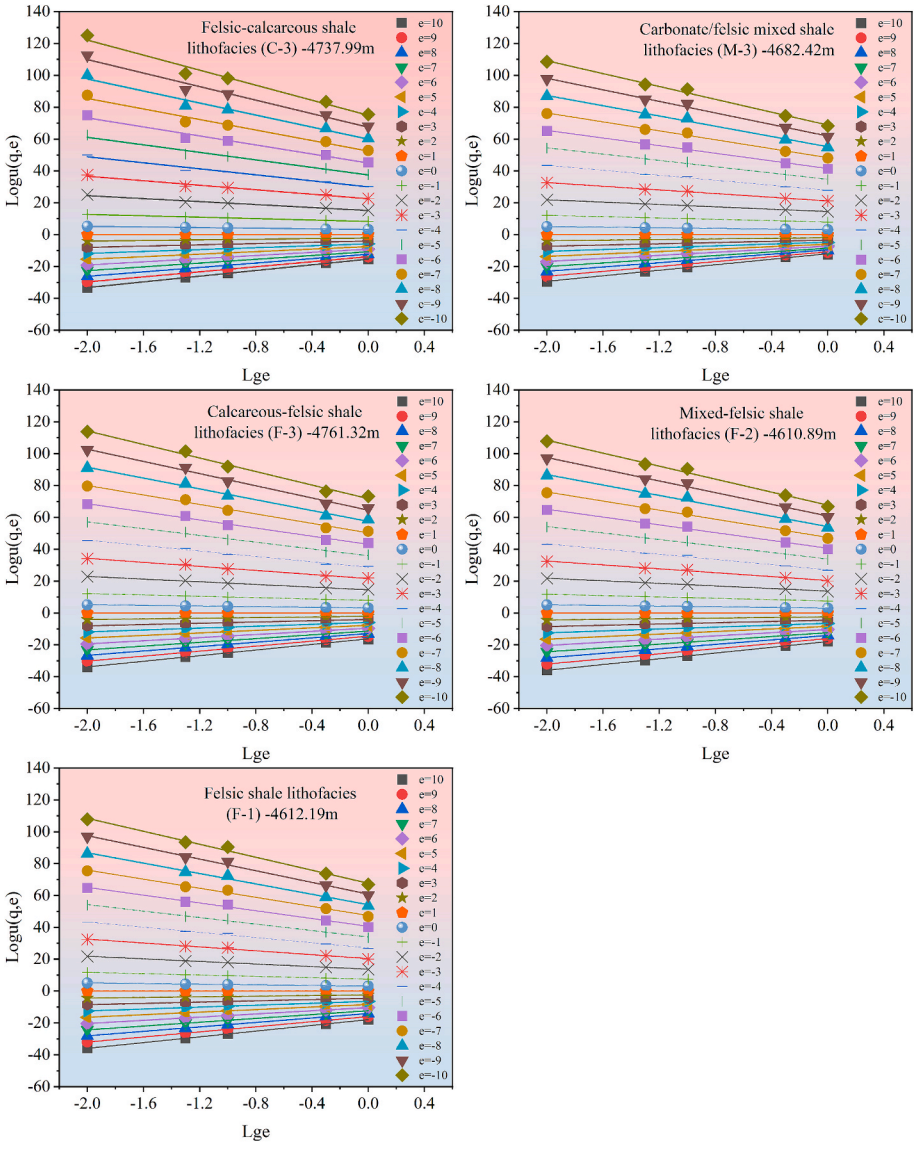


Fig. 17. Log-log plots of the mass partition function u(q, ε) versus the cell side length (ε) for the different shale lithofacies.

analytical examination. Fig. 18a depicts the generalized fractal dimension spectrum across the q range from −10 to 10, and Fig. 18b delineates the multifractal singularity spectrum over the same range. The generalized fractal dimension spectrum manifests an inverse 'S'-shaped profile, with the left side of q = 0 indicating relatively low probability in pore size distribution and the right side signifying regions of relatively high probability [69]. Drawing from nuclear magnetic resonance (NMR) analysis of the Fengcheng Formation reservoirs, the high probability region for pore size distribution is approximately 50 nm to 1 μm, in contrast to the low-probability region, which extends beyond one μm. Consequently, the high-probability region can be classified as the domain of meso- to micropore distributions, while the low-probability region is characterized as the domain of macropore distributions.

Employing multifractal dimension theory, we have derived the generalized fractal dimension spectrum q ~ Dq and the multifractal spectrum α ~ f(α) for various lithofacies. Within the generalized fractal dimension spectrum, the value of Dq diminishes as the order q ascends, attaining its nadir at q = 10, with an inflexion point occurring at q = 0. A more pronounced decline in D(q) with descending order q signifies higher heterogeneity. In the multifractal spectrum, the magnitude of f(α) diminishes progressively on either side of its peak (When α = 0, the corresponding value of f(α)), with the right side of the peak representing areas of reduced probability in pore size distribution (the macropore domain), and the left side indicating areas of increased probability (the meso- to micropore domain) [70]. The breadth of the x-axis in the multifractal spectrum (α_{max}-α_{min}) indicates reservoir heterogeneity. As depicted in Fig. 18a, among the distinct lithofacies of the Fengcheng Formation, the reduction in D(q) with increasing order q within the meso- to micropore range is relatively modest, implying a lower level of heterogeneity. Conversely, within the macropore range, the more pronounced decline in Dq suggests a more pronounced heterogeneity.

Based on multifractal spectrum analysis, samples are bifurcated into two principal categories: one category is predominantly influenced by larger pore sizes, which are indicative of low-probability regions, and the other is governed by meso- to micropore sizes, associated with high-probability areas [71,72]. The disparity between the maximum and minimum values of the multifractal dimension denoted as Δf, indicates the distribution characteristics within the pore space. A negative Δf, signifying the prevalence of low-probability regions, correlates with a rightward-hook shape in the f(α) curve; in contrast, a positive Δf, indicating the dominance of high-probability areas, is associated with a leftward-hook shape [73]. Within the Fengcheng Formation samples, three manifest a leftward-hook shape, suggesting that meso- to micropore sizes (high-probability regions) are the predominant influence. The remaining samples are marked by large pore areas (low-probability regions), aligning with the generalized fractal dimension spectrum's indication that samples with reduced heterogeneity in low-probability regions conform to this pattern. Consequently, it is apparent that the uniformity of larger pore sizes predominantly dictates the heterogeneity within shale reservoirs.

By employing multifractal theory, a comprehensive set of parameters can be derived to evaluate the heterogeneity of reservoir pore-throats. The interval between D_{-10}-D_{10} within the generalized fractal dimension spectrum represents the extensive range of pore size distribution probabilities, indicating the overall uniformity of reservoir pore sizes; a smaller value signifies a more homogeneous distribution [74]. As demonstrated in Table 3, there is minimal variation in the overall homogeneity of pore sizes across the five predominant lithofacies within the Fengcheng Formation. However, two mixed lithofacies—specifically, mixed feldspathic shale and calcareous feldspathic mixed lithofacies—exhibit greater heterogeneity in their pore size distributions. The differential D_{-10}-D_0 within the generalized fractal dimension spectrum delineates the extent of low-probability regions, characterizing heterogeneity among larger pore-throat domains in the Fengcheng Formation; an increased value correlates with heightened heterogeneity. Conversely, the differential D_0–D_{10} defines high-probability regions and characterizes meso- to micro-pore-throat domain heterogeneities; a larger value indicates greater variability. Notably, both mixed feldspathic shale and calcareous feldspathic mixed lithofacies demonstrate enhanced heterogeneity in medium to large pore-throat domains while exhibiting relatively subdued differences in meso- to micro-pore-throat region heterogeneities, as detailed in Table 3.

As elaborated in Section 4.5, lithofacies are classified into two primary categories based on carbonate content: those exhibiting elevated carbonate levels and those with comparatively lower carbonate concentrations. Fig. 18 illustrates that within the meso- to micro-pore spectrum, the concentration of free oil among the five predominant lithofacies of the Fengcheng Formation exhibits a progressive decline as the differential between D_0–D_{10} increases. Conversely, lithofacies enriched in carbonate minerals show an observed increase in adsorbed oil concentration corresponding to an expansion of the D_0–D_{10} differential. In contrast, lithofacies characterized by lower carbonate mineral content display an inverse relationship; specifically, their free oil concentration diminishes as the D_0–D_{10} differential widens. This phenomenon can be primarily attributed to heightened heterogeneity within meso- to micro-pores, leading to diminished connectivity in reservoirs, which reduces shale oil flow capacity and decreases free oil concentrations. Regarding adsorbed oil, increased complexity within meso- to micro-pores coupled with a greater specific surface area enhances adsorption potential relative to free oil, thereby elevating adsorbed oil concentrations. The analysis of hydrocarbon content across various lithofacies suggests that those with higher carbonate levels tend to

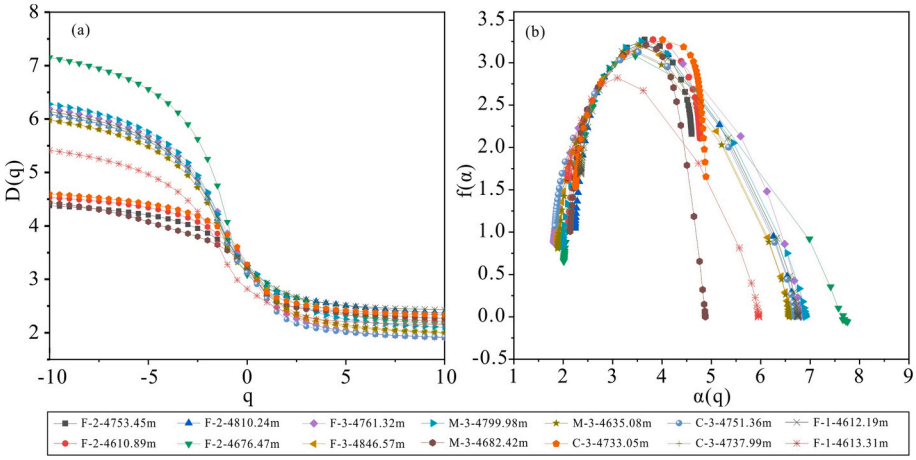


Fig. 18. Generalized dimension spectrum (a) and multifractal singularity spectrum (b) for the different lithofacies.

Table 3
Parameters of multiple fractal dimensions within different lithologies.

Serial number	Depth (m)	Lithofacies	Free oil (cm^3)	Absorbed oil (cm^3)	D_{-10}	D_{10}	α_{10}	α_{-10}	D_0	D_1
104	4753.45	Mixed-felsic shale lithofacies	0.28	0.73	4.38	1.91	1.81	4.60	3.27	2.83
56	4610.89		0.26	0.84	4.53	2.21	2.10	4.77	3.27	2.71
1–86	4810.24		0.92	0.3	6.10	2.37	2.24	6.71	3.21	2.83
1–29	4676.47		0.74	0.99	7.15	2.16	2.01	7.87	3.08	2.83
1–16	4612.19	Felsic shale lithofacies	0.52	1.41	6.14	2.43	2.35	6.75	3.21	2.93
a	4613.31		0.47	2.55	5.41	2.15	2.10	5.96	2.82	2.57
33	4761.32	Calcareous-felsic shale lithofacies	1.26	6.94	6.19	1.91	1.81	6.81	3.21	2.69
4	4846.57		1.98	5.15	5.98	2.00	1.88	6.58	3.21	2.80
40–2	4733.05	Felsic-calcareous shale lithofacies	2.16	5.54	4.60	2.33	2.25	4.89	3.27	2.76
e	4751.36		1.54	5.13	6.09	1.92	1.83	6.70	3.12	2.63
39–1	4737.99		1.92	3.46	6.14	2.43	2.35	6.75	3.21	2.93
1–63	4799.98	Carbonate/felsic mixed shale lithofacies	1.32	2.54	6.28	2.10	1.98	6.91	3.25	2.93
30	4682.42		1.67	2.81	4.43	2.27	2.14	4.88	3.21	2.71
j1	4635.08		1.15	2.93	5.98	2.01	1.89	6.58	3.12	2.76
Serial number	Depth (m)	Lithofacies	OSI (cm^3)	D_2	D_{-10}-D_{10}	D_{-10}-D_0	D_0–D_{10}	D_0-D_1	α_{-10}-α_{10}	Hurst
104	4753.45	Mixed-felsic shale lithofacies	121.74	2.44	2.47	1.11	1.36	0.44	2.79	1.72
56	4610.89		71.08	2.51	2.32	1.26	1.06	0.56	2.68	1.75
1–86	4810.24		158.62	2.69	3.72	2.88	0.84	0.38	4.47	1.85
1–29	4676.47		137.04	2.71	4.99	4.07	0.92	0.25	5.86	1.86
1–16	4612.19	Felsic shale lithofacies	192.59	2.74	3.71	2.93	0.78	0.28	5.00	1.87
a	4613.31		125.78	2.38	3.26	2.59	0.67	0.25	4.70	1.69
33	4761.32	Calcareous-felsic shale lithofacies	197.66	2.35	4.28	2.98	1.30	0.52	4.93	1.67
4	4846.57		158.40	2.42	3.99	2.77	1.22	0.41	2.74	1.71
40–2	4733.05	Felsic-calcareous shale lithofacies	219.98	2.60	2.27	1.33	0.94	0.52	4.69	1.80
e	4751.36		167.32	2.25	4.17	2.96	1.20	0.49	2.64	1.63
39–1	4737.99		252.68	2.74	3.71	2.93	0.78	0.28	4.86	1.87
1–63	4799.98	Carbonate/felsic mixed shale lithofacies	216.35	2.64	4.17	3.02	1.15	0.32	4.40	1.82
30	4682.42		307.14	2.52	2.17	1.22	0.94	0.50	4.40	1.76
j1	4635.08		145.57	2.46	3.97	2.86	1.12	0.36	3.86	1.73

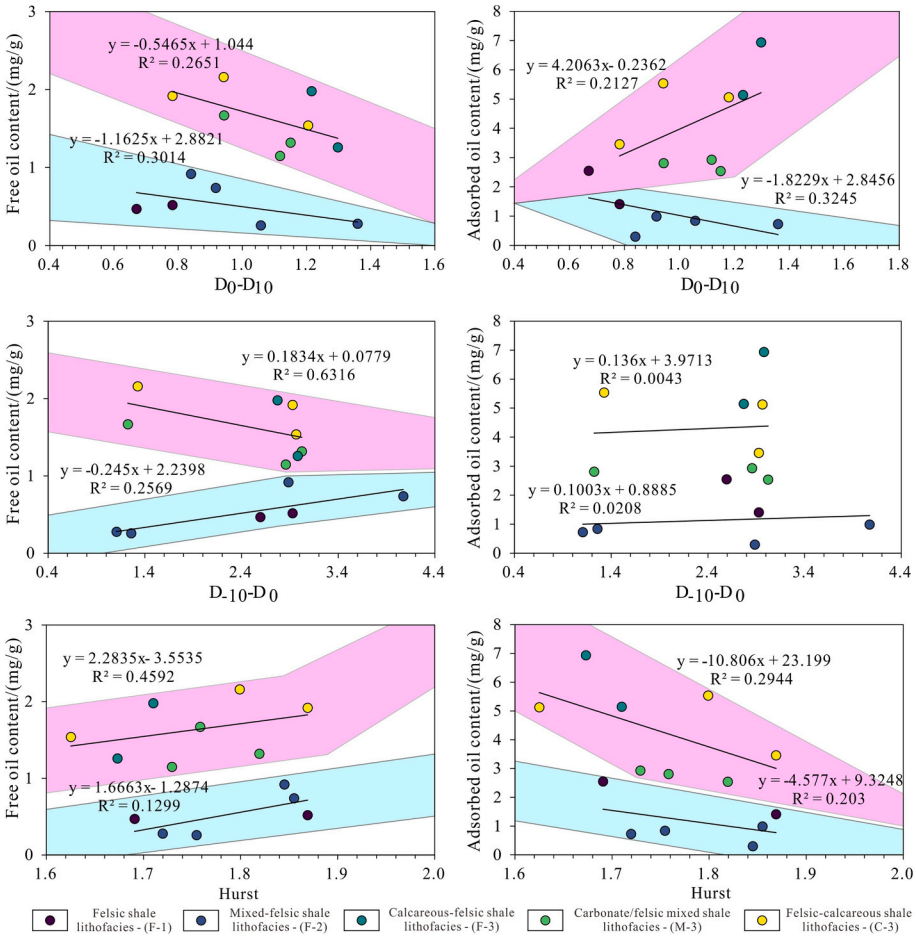


Fig. 19. The correlation between multifractal dimension parameters and free oil and adsorbed oil.

contain more abundant shale oil, likely due to superior ancient productivity and enhanced hydrocarbon-generating potential compared to their lower-carbonate counterparts. Consequently, it is plausible that shale oils found within low-carbonate lithofacies may originate from high-carbonate sources. Thus, intensified heterogeneity within meso- to micro-pores characterized by complex pore structures hinders shale oil ingress into these reservoirs while concurrently reducing adsorbed oil concentrations in low-carbonate lithofacies (Fig. 19).

In the macro-pore-throat domains, a notable decrease in free oil concentration is observed within lithofacies characterized by elevated carbonate mineral content as the D_{-10}-D_0 differential increases. Conversely, lithofacies with lower carbonate content exhibit an increase in free oil concentration. However, no significant correlation between adsorbed oil and lithofacies is detected. This dichotomy can be primarily attributed to the heightened heterogeneity present within the macro-pore throats of carbonate-rich lithofacies, which—being predominant hydrocarbon-generating lithotypes among the five studied—compromises reservoir connectivity and enhances shale oil's adsorptive affinity, thereby reducing free oil concentrations. In contrast, for lithofacies with diminished carbonate content, increased heterogeneity in their macro-pore throats enhances reservoir storage capacity, facilitating greater volumes of free oil sequestration. Regarding these larger pores, it is noteworthy that shale oil's adsorptive capacity on pore walls is significantly less than that of free oil due to larger pore-throat radii, thus rendering correlations with lithofacies less pronounced.

The connectivity of reservoirs is rigorously assessed through the Hurst exponent, mathematically defined as $(D_2 + 1)/2$, where D_2 represents the correlation dimension—a metric that quantifies the interconnectedness among various pore-throats. An elevated Hurst exponent correlates with improved pore-throat connectivity. Fig. 18 illustrates a positive correlation between the Hurst exponent and free oil concentration, alongside an inverse relationship with adsorbed oil concentration across the five primary lithofacies of the Fengcheng Formation. This trend contrasts with the internal heterogeneity of the reservoir. The observed phenomenon primarily results from enhanced reservoir connectivity, which significantly facilitates short-range translocation and coalescence of shale oil. Consequently, robust internal connectivity within the reservoir provides effective migration pathways for carbonate-rich lithofacies while offering substantial storage capacity for those with lower carbonate content.

4.7. Quantitative characterization model of oil-bearing properties by pore size

Based on the above analysis, we have clarified the influence of different lithofacies and that of the same lithofacies on oil content. However, a detailed analysis of the vertical distribution of lithofacies and variations in their oil content has not yet been conducted. Therefore, this study utilizes parameters derived from multifractal dimensions and the Oil Saturation Index (OSI) to characterize the vertical distribution features of different lithofacies and their associated oil content. Using parameters obtained from multifractal dimensions along with their correlation to oil content, we selected Hurst exponent and heterogeneity parameters for pore sizes ranging from D0 to D10 nanometers, establishing correlations with actual logging curves (CNL, DEN, CALI, RI, RXO, GR, AC, RT). Fig. 20 shows a significant correlation between the Hurst exponent and DEN, RI, RXO, GR, and RT. Additionally, D_0–D_{10} exhibits notable correlations with DEN, RI, GR, RT, and SP.

Therefore, using the method of multiple linear regression, the functional relationship between the Hurst index and D_0–D_{10} and the logging curve is established:

$$H = -1.09 \times DEN - 0.00028 \times RI + 0.00014 \times RXO + 0.0006 \times GR + 0.00014RT + 4.59 \quad R^2 = 0.62 \tag{29}$$

$$D_0 - D_{10} = 1.88 \times DEN + 0.0001 \times RI - 0.0021 \times GR - 0.000015 \times RT - 3.75 \quad R^2 = 0.34 \tag{30}$$

According to Fig. 21, there is a clear correspondence between the oil content of different lithofacies and their characteristics. Lithofacies with higher calcium content exhibit relatively high Oil Saturation Index (OSI) values, larger Hurst exponents and smaller D_0–D_{10} values. This observation is generally consistent with the results discussed earlier.

4.8. Deployment methodology of the model in practical applications

The proposed AI-based lithofacies prediction framework has demonstrated promising theoretical performance, and its field deployment requires three critical implementation phases. First, data accuracy and standardization: The model requires XRD-derived mineralogical data as primary input, necessitating rigorous data standardization including outlier removal and noise reduction. For cost-efficient field applications where XRD analysis is impractical, elemental logging data can be inverted to estimate required mineral compositions for dataset construction. Second, model architecture deployment: The validated model should be installed at well sites for real-time processing of logging data to generate lithofacies predictions, with outputs cross-verified against geologist-identified lithofacies for quality control. Third, cloud-based continuous learning: A distributed training platform could be established to systematically incorporate field verification results, enabling periodic model parameter updates to maintain prediction accuracy through operational feedback loops.

4.9. The model for the oil-bearing property of different lithofacies

Based on the characterization of reservoir space across different lithofacies and the analysis of their oil content, it is evident that within the shale lithofacies of the Fengcheng Formation, those with higher calcium content primarily function as generating spaces for shale oil and exhibit superior oil content. Conversely, lithofacies with lower calcium content mainly act as reservoir spaces for shale oil; however, their oil content is influenced by the heterogeneity of pore structures within the reservoir. Notably, greater heterogeneity in macropores enhances the enrichment potential of shale oil (Fig. 22).

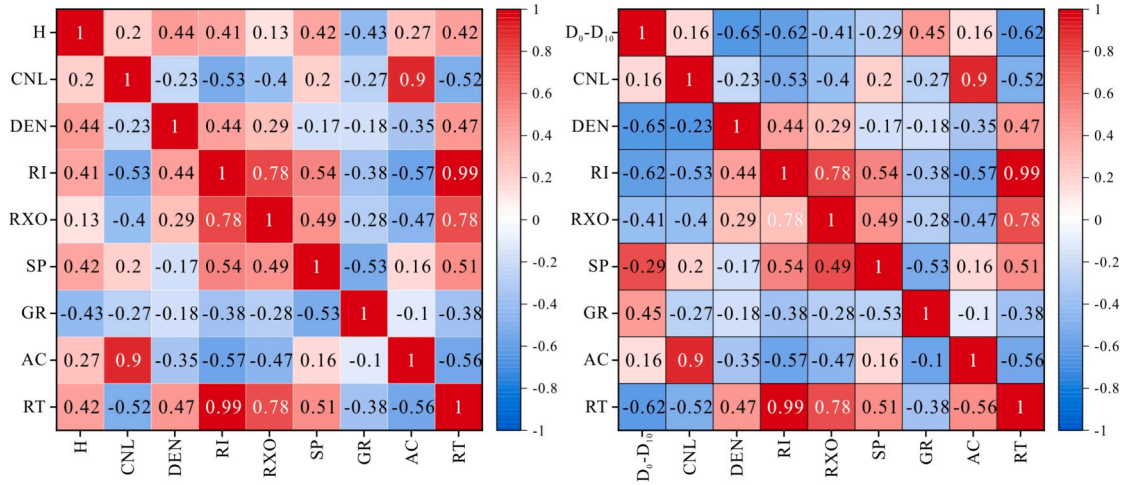


Fig. 20. Correlation between Hurst index, D0–D10 and logging parameters.

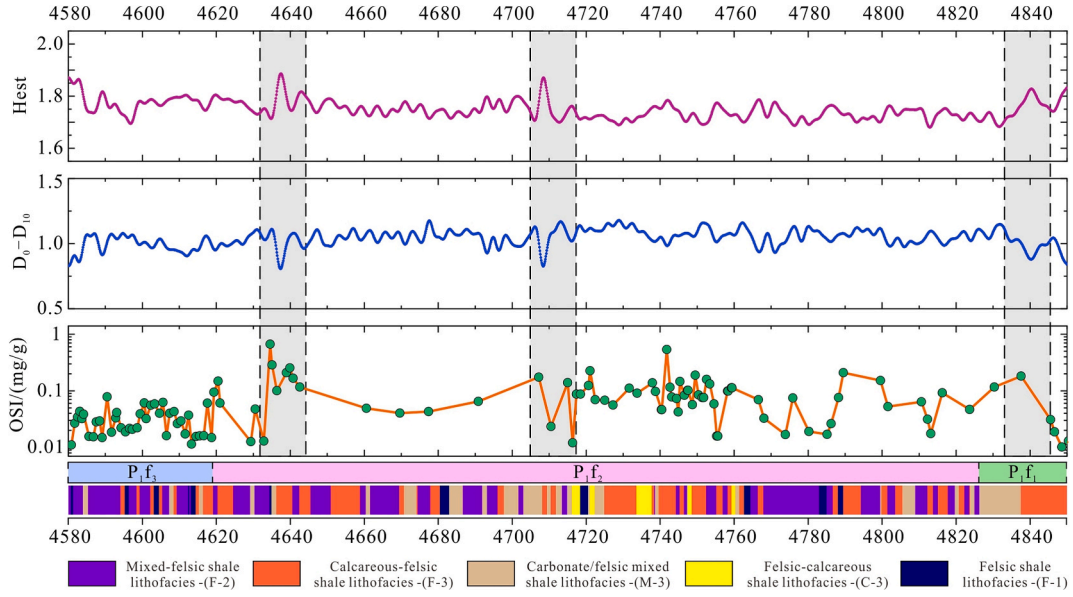


Fig. 21. The Hurst index, D0–D10, and OSI distribution characteristics of different lithofacies.

5. Study limitations and future recommendations

The current AI prediction work has only been conducted at the single-well scale with satisfactory results. However, characterizing the planar distribution across the entire study area solely based on well data has limitations. The next step requires integrating seismic data with multimodal theory. We will combine logging data with seismic attributes using transfer learning, incorporating lithofacies structural features and descriptions to achieve accurate planar lithofacies characterization across the study area. This will also help validate the single-well lithofacies distribution model.

6. Conclusion

This study innovatively integrates artificial intelligence algorithms with multiscale experimental characterization techniques. Based on elemental logging and XRD data, we developed an intelligent lithofacies identification model with adaptive optimization capabilities, systematically revealing the intrinsic relationships between mineral composition, pore structure characteristics, and shale oil occurrence patterns across different lithofacies. Furthermore, by combining multifractal dimension theory with well logging parameters, we established a comprehensive intelligent quantitative characterization model for "lithology-pore structure-oil content" across entire well sections. Specific conclusions are provided below:

1. This study developed an integrated Bayesian-IGWO ensemble learning model that achieves accurate vertical lithofacies identification with superior resolution compared to conventional XRD methods, demonstrating reliable performance for heterogeneous reservoir evaluation
2. Shale oil distribution is controlled by mineral composition and pore structure. Carbonates increase porosity through dissolution while quartz and feldspar optimize pore space. Clays exhibit dual adsorption and blockage effects. Macropores govern free oil mobility whereas micropores control adsorbed oil retention
3. This study integrates multi-stage pyrolysis with NMR to establish a quantitative method for oil content characterization across nano-to-micron pore systems. Using Hurst index and multifractal parameters, we developed a logging response model that confirms superior pore-throat connectivity and oil-bearing potential in calcareous

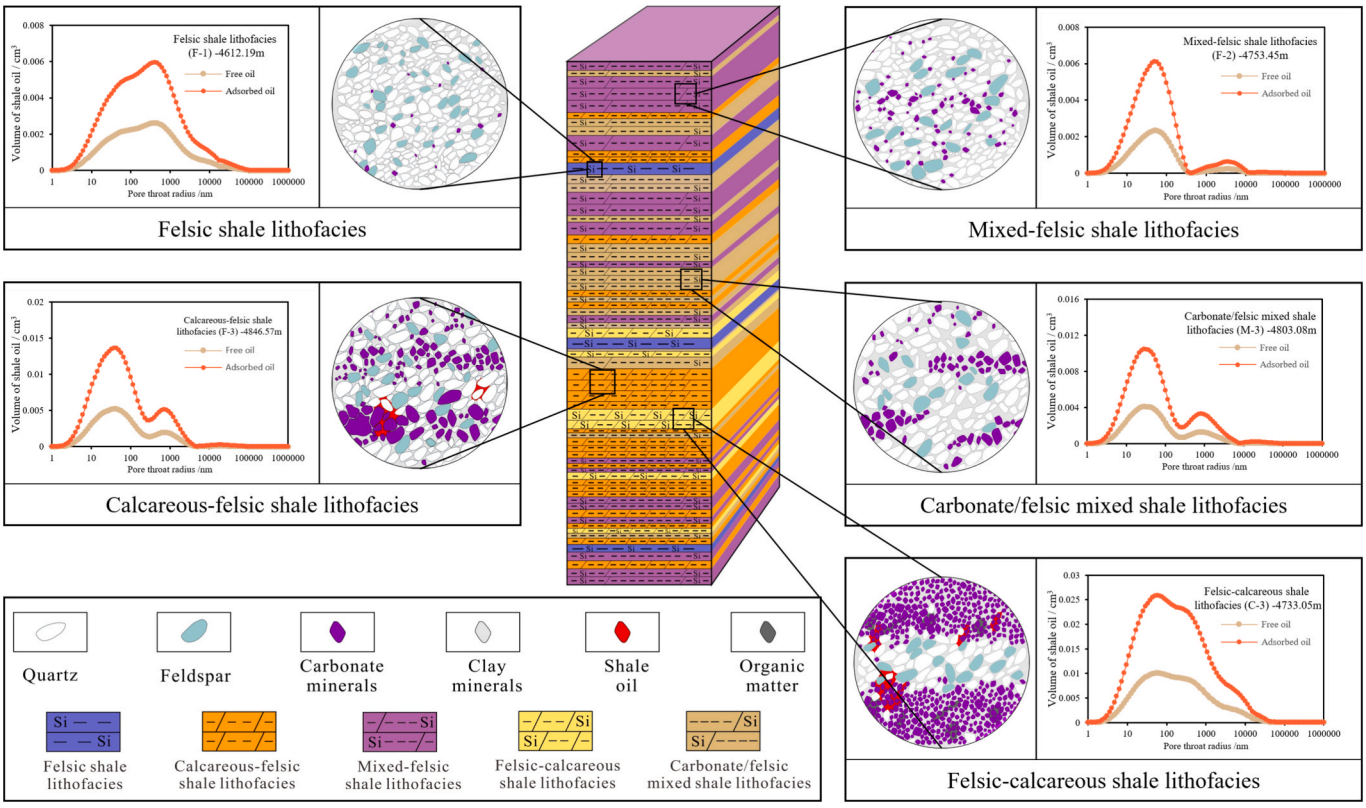


Fig. 22. Pattern of oil-bearing characteristics of different lithofacies in Fengcheng Formation of Mahu Sag.

lithofacies. The AI-enhanced geostatistical approach provides a standardized intelligent solution for sweet spot prediction.

CRediT authorship contribution statement

Zaiquan Yang: Writing – review & editing, Validation, Resources, Methodology, Formal analysis, Conceptualization, Writing – original draft, Supervision, Project administration, Funding acquisition, Data curation, Visualization, Software, Investigation. **Dongxia Chen:** Validation, Software, Methodology, Formal analysis, Writing – original draft, Resources, Investigation, Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition, Data curation. **Xianglu Tang:** Software, Methodology, Funding acquisition, Validation, Resources, Investigation, Formal analysis, Conceptualization, Visualization, Supervision, Project administration, Data curation. **Yuchao Wang:** Validation, Data curation, Writing – original draft, Software, Writing – review & editing, Resources. **Zhenxue Jiang:** Visualization, Project administration, Formal analysis, Validation, Funding acquisition, Supervision. **Liliang Huang:** Supervision, Methodology, Formal analysis, Resources, Funding acquisition. **Leilei Yang:** Supervision, Visualization, Funding acquisition, Software, Formal analysis. **Zhiye Gao:** Methodology, Software, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is financially supported by the Strategic Cooperation Technology Projects of CNPC and CUPB (ZLZX2020-01-05). In addition, the authors would like to express their sincere thanks to the PetroChina Xinjiang Oilfield Company for their assistance in providing the information and for their technical input to this work.

Appendix

1. The formula for BP neural network:

Step 1: Network initialization. The weights and biases of the network are initialized. Typically, the weights are initialized to small random values, while the bias can be initialized to 0 or small values.

$$W^{[l]}\ N(0,\sigma^2) \tag{1}$$

$$b^{[l]} = 0 \tag{2}$$

Step 2: Hidden layer output calculation. The data is propagated forward through the network layers. σ is the activation function. The output $a^{[l]}$ of each layer is computed:

$$z^{[l]} = W^{[l]} \bullet a^{[l-1]} + b^{[l]} \tag{3}$$

$$a^{[l]} = \sigma\left(z^{[l]}\right) \tag{4}$$

Step 3: Output layer input calculation. The output prediction for linear activation is obtained directly by calculating the output layer, where the output value equals the input value without any transformation.

The input of the output layer $z^{[output]}$:

$$z^{[output]} = W^{[output]} \bullet a^{[lasthiddenlayer]} + b^{[output]} \tag{5}$$

The output of the output layer $\widehat{y}$:

$$\widehat{y} = z^{[output]} \tag{6}$$

Step 4: Error calculating. The mean square error (MSE) serves as the designated loss function.

$$\text{Loss} = \frac{1}{2m}\sum_{i=1}^{m}(\widehat{y}_i - y_i)^2 \tag{7}$$

Step 5: Update weight. The weights are updated utilizing the gradient descent methodology.

$$W^{[l]} = W^{[l]} - \alpha\frac{\partial Loss}{\partial W^{[l]}} \tag{8}$$

Step 6: Update bias. The bias is updated through the utilization of gradient descent.

$$b^{[l]} = b^{[l]} - \alpha\frac{\partial Loss}{\partial b^{[l]}} \tag{9}$$

2. Optimization process of the Bayesian algorithm

Step 1: Define the objective function with the Gaussian process (GP) prior. The objective function is defined as "f(x)". The variable x in this scenario represents a vector. The vector x, in this case, represents two hyperparameters of the neural network: the number of nodes in the hidden layer (h) and the number of hidden layers (l). The objective function value corresponds to the validation set loss of the neural network given a specific hyperparameter configuration. In contrast, the prior distribution of the objective function f(x) is modelled as a Gaussian process.

$$\begin{cases} f(x) = NetworkLoss(h, l; D_{Train}, D_{val}) \\ \quad f(x)\ GP(m(x), k(x, x')\ ) \end{cases} \tag{10}$$

The D_{Train} and D_{val} respectively denote the training set and test set in the main code.

The mean function m(x) is typically assumed to be zero (or another simple form). In contrast, the covariance function or kernel function k(x, x') defines the covariance between any two points in the input space, such as the radial basis function (RBF).

Step 2: Update the Gaussian process (GP) posterior. The purpose of posterior updating is to enhance or revise the understanding and estimation of the behaviour of the objective function with new observations, specifically, performance measures obtained under new hyperparameter configurations. Simultaneously, it can enhance the subsequent parameter selection process with greater intelligence and efficiency. The posterior probability $p(f|D)$ represents the distribution of function f based on observation D.

According to Bayes' theorem, the posterior probability can be computed using the following formula:

$$p(f|D) = \frac{p(D|f) \bullet p(f)}{p(D)} \tag{11}$$

The prior probability distribution of the objective function f, denoted as $p(f)$, is commonly assumed to follow a Gaussian process. $p(D|f)$ represents the likelihood of the observed data D given the function f. The marginal probability of the data, $p(D)$, serves as a normalization constant to ensure that the posterior distribution maintains its validity as a proper probability distribution.

Step 3: Function of acquisition. The next query point is determined using the acquisition function EI (expected improvement) based on the current posterior distribution.

$$EI(x) = \int_{-\infty}^{\infty}\max(0, f_{min} - f(x)p(f(x)|D\ )df(x) \tag{12}$$

Step 4: Optimize acquisition function and data update. The next query point is determined by selecting the point that maximizes the acquisition function. The objective function is evaluated at the new point Xnext, new observations Ynext are obtained, and the dataset is updated.

$$\begin{cases} x_{next} = argmax(x; D) \\ D \leftarrow D \bigcup \{(x_{next}, y_{next})\ \} \end{cases} \tag{13}$$

Step 5: Steps 2 to 4 should be repeated until certain termination conditions are met, such as reaching the maximum number of iterations or when EI falls below a specific threshold.

3. Optimization process of the GWO algorithm

Step 1: Surround prey stage (initialize position). The grey Wolf, acting as the algorithm's agent, initiates at this stage and determines an approximate location for the prey, which refers to the potential minimum region of the loss function. The position xi of each Wolf is initialized, representing a collection of network weights and biases. The positions are initialized randomly.

$$x_i = lb + (ub + lb) \times rand() \tag{14}$$

The terms lb and ub denote the lower and upper limits of weight and bias, respectively.

Step 2: Hunting (tracking) prey stage (updating position) The grey Wolf updates its position at this stage, moving closer to the target based on the guidance of the lead Wolf, which represents the optimal solution. The update of the grey Wolf position is contingent upon the positions of Alpha, Beta, and Delta Wolves. The positions of all non-leader wolves (Omega wolves) are updated based on the positions of the three leader wolves.

$$x_{new} = \frac{1}{3}\left(x_{new,\alpha} + x_{new,\beta} + x_{new,\delta}\right) \tag{15}$$

Step 3: Stage of attacking prey (fine adjustment of position) The grey wolves employ a strategic approach to their prey, meticulously identifying and honing in on the optimal solution during optimization.

$$\mathrm{a} = 2\left(1 - \frac{iter}{maxIter}\right) \tag{16}$$

Step 4: Comprehensive evaluation and iteration. The new weights and bias sets are assessed using the loss function after each position update.

$$f(x) = Loss(Output(w,b), Target) \tag{17}$$

4. Optimization process of the A-IGWO algorithm

Improvement point 1 The adaptive dynamic adjustment coefficient a. The coefficient a in the original GWO typically undergoes a linear decrease from 2 to 0. The value of a can be dynamically adjusted based on the performance of the current iteration in comparison to the historically optimal performance following the improvement. The extent to which the function value f(t) improves in comparison to the historical best value f* during the current iteration. The small proportion indicates that the new iteration does not bring significant improvement, resulting in a decrease in the value of A. This reduction enhances the algorithm's development capability and diminishes exploration behaviour, accelerating convergence [34].

$$a(t) = 2\exp\left(-\gamma\frac{f(t) - f^*}{f^*}\right) \tag{18}$$

The function $f(t)$ represents the value of the objective function at the current iteration. f^*: represents the optimal value of the objective function discovered thus far. γ: the adjustment constant, $f(t) - f^*/f^*$: Relative improvement ratio.

Improvement point 2: Enhance the balance between exploration and development. Random disturbance can effectively address GWO's challenges in certain optimization problems, particularly in ensuring algorithm efficiency and effectiveness when dealing with complex, multi-peak optimization environments.

$$x_{new} = x_{pred} - A_{pred} \bullet D_{pred} + \delta \bullet rand(size(x)) \tag{19}$$

x_{pred}: The current position of the lead Wolf. A_{pre}: Controls parameter, The coefficient a affects A_{pred}, which determines the direction and intensity of Wolf movement. D_{pred}: The disparity in distance between the present location and the leader Wolf's position. δ: The coefficient governing the intensity of the disturbance, The magnitude of the random disturbance is defined by δ. $rand(size(x))$: Generate standard normally distributed random numbers of the same size as X, which is used to implement random disturbances.

Improvement point 3: Avoiding local optima and improving convergence speed. The GWO algorithm exhibits strong global search capabilities but may still encounter challenges in finding optimal solutions when dealing with complex or multi-peak optimization problems. Incorporating **Improvement Point 3** enables the algorithm to escape from local optima and explore a wider range of possibilities.

$$x_{new} = x_{pred} - A_{pred} \bullet D_{pred} \bullet \left(1 + \epsilon \bullet \frac{\Delta f}{\langle f \rangle}\right) \tag{20}$$

Δf: The improvement of the optimal solution before and after iteration. $\langle f \rangle$ The average improvement magnitude of the historical optimum. ∈: Adjustment coefficient.

Data availability

https://data.mendeley.com/drafts/s5jcs8nnb5

References

[1] Merriam ER, Petty JT, Maloney KO, Young JA, Faulkner SP, Slonecker ET, et al. Brook trout distributional response to unconventional oil and gas development: Landscape context matters. Sci Total Environ 2018;628:338–49.

[2] Wang X, et al. Coupled pressure-driven flow and spontaneous imbibition in shale oil reservoirs. Phys Fluids 2023;35:4.

[3] Misch D, Groß D, Mahlstedt N, Makogon V, Sachsenhofer RF. Shale gas/shale oil potential of upper Visean Black Shales in the Dniepr-Donets Basin (Ukraine). Mar Pet Geol 2016;75:203–19.

[4] Muther T, Qureshi HA, Syed FI, Aziz H, Siyal A, Dahaghi AK, et al. Unconventional hydrocarbon resources: geological statistics, petrophysical characterization, and field development strategies. J Pet Explor Prod Technol 2022;12(6):1463–88.

[5] Wood. J.M, Euzen. T, Cesar. J, Jaime. Cesar, "Influence of faults on late-stage migration of methane-rich gas in a major unconventional hydrocarbon accumulation: Montney Formation, western Canada." Marine and Petroleum Geology 164, 106857(2024).

[6] Wang YH, Lei ZD, Zhou Q, Liu YS, Xu ZH, Wang YW, et al. A physical constraint-based machine learning model for shale oil production prediction. Phys Fluids 2024;36:8.

[7] Xiong Y, Li S, Li M, et al. Study on characteristics of imbibition and recovery predicted methods with different lithology structure. Fuel 2025;381:133525.

[8] Yang Y, Wei J, Yang E, et al. Unlocking continental shale oil potential: Microscopic insights into fluid saturation mechanisms via imbibition for future development strategies in the Songliao Basin. SPE J 2024;29(03):1574–88.

[9] Guo Q, Yao Y, Hou L, et al. Oil migration, retention, and differential accumulation in "sandwiched" lacustrine shale oil systems from the Chang 7 member of the Upper Triassic Yanchang Formation, Ordos Basin, China. Int J Coal Geol 2022;261:104077.

[10] Zhi D, Xiang B, Zhou N, et al. Contrasting shale oil accumulation in the upper and lower sweet spots of the lacustrine Permian Lucaogou Formation, Junggar Basin, China. Mar Pet Geol 2023;150:106178.

[11] ·He. W, Zhu. R, Cui. B, Zhang. S, Meng. Q, Bai. B, Feng. Z, Lei. Z, Wu. S, He. K, Liu. H, "The geoscience frontier of Gulong shale oil: revealing the role of continental shale from oil generation to production." Engineering 28 79-92(2023).

[12] Liu G, Tang Y, Liu K, et al. Comparison of pore structure characteristics of shale-oil and tight-oil reservoirs in the Fengcheng Formation in Mahu Sag. Energies 2024;17(16):4027.

[13] Zhang Z, Liu K, Wang Z, et al. Detailed characterization of pore results of continental shale reservoir in Fengcheng Formation, Mahu Sag. ACS Omega 2024;9(21):22923–40.

[14] Li C, Hu T, Cao T, et al. Pore Structure and Geochemical Characteristics of Alkaline Lacustrine Shale: the Fengcheng Formation of Mahu Sag, Junggar Basin. Minerals 2023;13(10):1248.

[15] Huang Y, Wang G, Zhang Y, et al. Logging evaluation of pore structure and reservoir quality in shale oil reservoir: the Fengcheng Formation in Mahu Sag, Junggar Basin, China. Mar Pet Geol 2023;156:106454.

[16] Tang Y, Lv Z, He W, et al. Characteristics and genesis of alkaline lacustrine tight oil reservoirs in the Permian Fengcheng Formation in the Mahu Sag, Junggar Basin, NW China. Minerals 2022;12(8):979.

[17] Gong D, Bai L, Gao Z, et al. Occurrence mechanisms of laminated-type and sandwich-type shale oil in the Fengcheng Formation of Mahu Sag, Junggar Basin. Energy Fuel 2023;37(18):13960–75.

[18] Gao Z, Duan L, Jiang Z, et al. Using laser scanning confocal microscopy combined with saturated oil experiment to investigate the pseudo in-situ occurrence mechanism of light and heavy components of shale oil in sub-micron scale. J Pet Sci Eng 2023;220:111234.

[19] Jia L, Yang Z, Gao Z, et al. The occurrence characteristics and enrichment patterns of shale oil in different lithofacies: a case study of the Fengcheng Formation in the Mahu Sag. ACS Omega 2024;9(46):46505–18.

[20] Wang Y, Chang XC, Liu ZQ, et al. Occurrence state and oil content evaluation of Permian Fengcheng Formation in the Hashan area as constrained by NMR and multistage Rock-Eval. Pet Sci 2023;20(3):1363–78.
[21] Tang Y, He W, Zheng M, et al. Shale lithofacies and its effect on reservoir formation in Lower Permian alkaline lacustrine Fengcheng Formation, Junggar Basin, NW China. Front Earth Sci 2022;10:930890.
[22] Alatefi S, Abdel Azim R, Alkouh A, et al. Integration of multiple bayesian optimized machine learning techniques and conventional well logs for accurate prediction of porosity in carbonate reservoirs. Processes 2023;11(5):1339.
[23] Ameur-Zaimeche O, Kechiched R, Heddam S, et al. Real-time porosity prediction using gas-while-drilling data and machine learning with reservoir associated gas: Case study for Hassi Messaoud field, Algeria. Mar Pet Geol 2022;140:105631.
[24] Male F, Duncan IJ. Lessons for machine learning from the analysis of porosity-permeability transforms for carbonate reservoirs. J Pet Sci Eng 2020;187:106825.
[25] Wei X, Huang W, Liu L, et al. Low-rank coalbed methane production capacity prediction method based on time-series deep learning. Energy 2024;311:133247.
[26] Liu X, Liu B, Wu W, et al. Spatial distribution prediction of pore pressure based on Mamba model. Front Earth Sci 2025;13:1530557.
[27] Li T, Wang Z, Wang R, et al. Pore type identification in carbonate rocks using convolutional neural network based on acoustic logging data. Neural Comput & Applic 2021;33(9):4151–63.
[28] Zhang T, Dong Y, Bai H, et al. Stochastic reconstruction of shale combining multi-scale generators and discriminators with attention mechanisms. Comput Geosci 2023;27(6):1041–65.
[29] Agwu OE, Alatefi S, Alkouh A, et al. Carbon capture using ionic liquids: an explicit data driven model for carbon (IV) Oxide solubility estimation. J Clean Prod 2024; 472:143508.
[30] Song Z, Shi H, Zhang X, et al. Prediction of CO2 solubility in ionic liquids using machine learning methods. Chem Eng Sci 2020;223:115752.
[31] Alatefi S, Agwu OE, Alkouh A. Explicit and explainable artificial intelligent model for prediction of CO2 molecular diffusion coefficient in heavy crude oils and bitumen. Results Eng 2024;24:103328.
[32] Sinha U, Dindoruk B, Soliman M. Prediction of CO2 minimum miscibility pressure using an augmented machine-learning-based model. SPE J 2021;26(04):1666–78.
[33] Ajoma E, Sungkachart T, Le-Hussain F. Effect of miscibility and injection rate on water-saturated CO2 injection. Energy 2021;217:119391.
[34] Chen G, Qi H, Yu J, et al. Application of a multi-layer feedforward neural network to predict fracture density in shale oil, Junggar Basin, China. Front Earth Sci 2023; 11:1114389.
[35] Wang L, Lu J, Luo Y, et al. An automated quantitative methodology for computing gravel parameters in imaging logging leveraging deep learning: a case analysis of the Baikouquan Formation within the Mahu Sag. Processes 2024;12(7):1337.
[36] Deo IK, Rajeev J. Predicting waves in fluids with deep neural network. Phys Fluids 2022;34:6.
[37] Mittal. S, Pathak. S, Dhawan. H, Upadhyayula. S, "A machine learning approach to improve ignition properties of high-ash Indian coals by solvent extraction and coal blending," Chemical Engineering Journal, 413, p.127385(2021).
[38] Shahriari. B, Swersky. K, Wang. Z, Adams. R.P, De. Freitas. N, "Taking the human out of the loop: A review of Bayesian optimization," Proceedings of the IEEE, 104 (1), pp.148-175(2015).
[39] Ghalambaz M, Yengejeh RJ, Davami AH. Building energy optimization using grey wolf optimizer (GWO). Case Stud Therm Eng 2021;27:101250.
[40] Meidani K, Hemmasian A, Mirjalili S, Barati FA. Adaptive grey wolf optimizer. Neural Comput & Applic 2022;34(10):7711–31.
[41] Tong SH, Guo MM, Tian Y, Le JL, Zhang DQ, Zhang H. Multi-objective optimization design of scramjet nozzle based on grey wolf optimization algorithm and kernel extreme learning machine surrogate model. Phys Fluids 2024;36:2.
[42] Rabie AH, Eltamaly AM. A new NEST-IGWO strategy for determining optimal IGWO control parameters. Neural Comput & Applic 2023;35(20):15143–65.
[43] Saleh AI, Hussien SA. Disease diagnosis based on improved Gray Wolf Optimization (IGWO) and ensemble classification. Ann Biomed Eng 2023;51(11): 2579–605.
[44] Edwards J C. A review of applications of NMR spectroscopy in the petroleum industry . 2011.
[45] Standard Guide for Measurement of Polyolefin Properties Using TD-NMR Relaxometry, ASTM D8539-23 ASTM International, 2023.
[46] Total Oxygen in Solid, Semi-solid, and High Boiling Point Liquid Hydrocarbons by Pyrolysis, ASTM UOP649-10 ASTM International, 2010.
[47] Li K, Wang Q, Ma H, et al. Effect of clay minerals and rock fabric on hydrocarbon generation and retention by thermal pyrolysis of Maoming oil shale. Processes 2023;11(3):894.
[48] Ferreiro. J.P, Vázquez. E.V, "Multifractal analysis of Hg pore size distributions in soils with contrasting structural stability," Geoderma, 160(1), pp.64-73(2010).
[49] Ge. X, Fan. Y, Li. J, Zahid. M. A, "Pore structure characterization and classification using multifractal theory—An application in Santanghu basin of western China". Journal of Petroleum Science and Engineering, 127, pp, 297-304(2015).
[50] Wenlong D, Gang GAO, Xincai YOU, et al. Genesis and distribution of oils in Mahu Sag, Junggar Basin, NW China. Pet Explor Dev 2023;50(4):840–50.
[51] Liu Y, Wang X, Wu K, et al. Late Carboniferous seismic and volcanic record in the northwestern margin of the Junggar Basin: Implication for the tectonic setting of the West Junggar. Gondw Res 2019;71:49–75.
[52] Jiang. F, Hu. T, Huang. L, Cheng. L, Jiang. Z, Huang. R, Zhang. C, Wu. G. Y, "Controlling factors and models of shale oil enrichment in Lower Permian Fengcheng Formation, Mahu Sag, Junggar Basin, NW China," Petroleum Exploration and Development, 50(4)pp.812-825(2023).
[53] Yu. K, Cao. Y, Qiu. L, Sun. P, "The hydrocarbon generation potential and migration in an alkaline evaporite basin: The Early Permian Fengcheng Formation in the Junggar Basin, northwestern China," Marine and Petroleum Geology, 98, pp.12-32 (2018).
[54] Zhu X, Zhang M, Zhu S, et al. Shale lithofacies and sedimentary environment of the third member, Shahejie Formation, Zhanhua Sag, Eastern China. Acta Geol Sin-Engl Ed 2022;96(3):1024–40.
[55] ·Chen. G, Tang. Y, Nan. Y, Yang. F, Wang. D, "Paleo-Sedimentary Environments and Controlling Factors for Enrichment of Organic Matter in Alkaline Lake Sediments: A Case Study of the Lower Permian Fengcheng Formation in Well F7 at the Western Slope of Mahu Sag, Junggar Basin," Processes, 11(8), p.2483(2023).
[56] Ai L, Liu S, Cong S, et al. Spatial variability of surface sediments in the Malacca strait and its implications for sedimentary environments. J Asian Earth Sci 2024; 259:105922.
[57] Wang X, Pu X, Hu Q, et al. Geological controlling factors of oil content in the Paleogene Kongdian Formation, Bohai Bay Basin. Energy Fuel 2024;38(16): 15235–47.
[58] Jia J, Gu J. Control factors and porosity evolution of high-quality sandstone reservoirs of Kela-2 gas field in Kuqa depression. Chin Sci Bull 2002;47:100–6.
[59] Si L, Wei J, Xi Y, et al. The influence of long-time water intrusion on the mineral and pore structure of coal. Fuel 2021;290:119848.
[60] Zhao K, Zhang C, Zhang L, et al. Petrology, Physical Properties and Geochemical Characteristics of Alkaline Lake Shale—Fengcheng Formation in Mahu Sag, Junggar Basin. Energies 2022;15(8):2959.
[61] Garcia-Valles M, Vendrell-Saz M, Pradell-Cara T. Organic geochemistry (Rock-Eval) and maturation rank of the Garumnian coal in the Central Pyrenees (Spain). Fuel 2000;79(5):505–13.
[62] Kanaya. T, Hirth, G, "Brittle to semibrittle transition in quartz sandstone: Energetics," Journal of Geophysical Research: Solid Earth, 123(1), pp.84-106 (2018).
[63] Dong T, Gao J, Li S, et al. Quartz cementation in the lower paleozoic shales, middle yangtze region, south china: implications for shale reservoir properties. J Earth Sci 2024;35(6):1918–33.
[64] Lai. J, Wang. G, Wang. S, Cao. J, Li. M, Pang. X, Zhou. Z, Fan. X, Dai. Q, Yang. L, He. Z, "Review of diagenetic facies in tight sandstones: Diagenesis, diagenetic minerals, and prediction via well logs," Earth-Science Reviews, 185, pp.234-258 (2018).
[65] Hou. M, Qu. J, Zha. M, Swennen. R, Ding. X, Imin. A, Liu. H, Bian. B, "Significant contribution of haloalkaliphilic cyanobacteria to organic matter in an ancient alkaline lacustrine source rock: A case study from the Permian Fengcheng Formation, Junggar Basin, China," Marine and Petroleum Geology, 138, p.105546 (2022).
[66] Wang. P, Jiang. Z, Ji. W, Zhang. C, Yuan. Y, Chen. L, Yin. L, "Heterogeneity of intergranular, intraparticle and organic pores in Longmaxi shale in Sichuan Basin, South China: Evidence from SEM digital images and fractal and multifractal geometries," Marine and Petroleum Geology, 72, pp.122-138(2016).
[67] Jouini MS, Alabere AO, Alsuwaidi M, et al. Experimental and digital investigations of heterogeneity in lower cretaceous carbonate reservoir using fractal and multifractal concepts. Sci Rep 2023;13(1):20306.
[68] Yan G, Qin Z, Marsh S, et al. Nanopore size distribution heterogeneity of organic-rich shale reservoirs using multifractal analysis and its influence on porosity–permeability variation. Energy Fuel 2021;35(17):13700–11.
[69] Li. Z, Zhang. J, Mo. X, Xu. X, Zhang. Y, Wang. D, Tong. Z, Tang. X, "Adsorption behaviors of near-critical carbon dioxide on organic-rich shales: Modeling, multifractality, and kinetics," Chemical Engineering Journal, 428, p.132526 (2022).
[70] Wang Y, Li W, Wang X, et al. Nanopore structure and multifractal characteristics of continental shale oil reservoir: a case study from Ziliujing shales in the Sichuan Basin. J Mar Sci Eng 2023;11(10):1989.
[71] He T, Zhou Y, Li Y, et al. Research on the microscopic pore-throat structure and reservoir quality of tight sandstone using fractal dimensions. Sci Rep 2024;14(1): 22825.
[72] Sijilmassi O, Alonso JML, Sevilla ADR, et al. Multifractal analysis of embryonic eye structures from female mice with dietary folic acid deficiency. Part I: Fractal dimension, lacunarity, divergence, and multifractal spectrum. Chaos Solitons Fractals 2020;138:109885.
[73] Barreira L, Pesin Y, Schmeling J. On a general concept of multifractality: multifractal spectra for dimensions, entropies, and Lyapunov exponents. Multifractal rigidity. Chaos: an InterdiscJ Nonlinear Sci 1997;7(1):27–38.
[74] Pachr M, Šimák V, Šumbera M, et al. Entropy, dimensions and other multifractal characteristics of multiplicity distributions. Mod Phys Lett A 1992;7(25):2333–9.