Petroleum Science 20 (2023) 2747–2758

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Integrated classification method of tight sandstone reservoir based on principal component analysis– simulated annealing genetic algorithm–fuzzy cluster means

Bo-Han Wu, Ran-Hong Xie*, Li-Zhi Xiao, Jiang-Feng Guo, Guo-Wen Jin, Jian-Wei Fu

National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China

ARTICLE INFO

Article history:
Received 13 August 2022
Received in revised form
12 December 2022
Accepted 14 April 2023
Available online 14 April 2023

Edited by Jie Hao

Keywords:
Tight sandstone
Integrated reservoir classification
Principal component analysis
Simulated annealing genetic algorithm
Fuzzy cluster means

ABSTRACT

In this research, an integrated classification method based on principal component analysis– simulated annealing genetic algorithm–fuzzy cluster means (PCA–SAGA–FCM) was proposed for the unsupervised classification of tight sandstone reservoirs which lack the prior information and core experiments. A variety of evaluation parameters were selected, including lithology characteristic parameters, poro-permeability quality characteristic parameters, engineering quality characteristic parameters, and pore structure characteristic parameters. The PCA was used to reduce the dimension of the evaluation parameters, and the low-dimensional data was used as input. The unsupervised reservoir classification of tight sandstone reservoir was carried out by the SAGA-FCM, the characteristics of reservoir at different categories were analyzed and compared with the lithological profiles. The analysis results of numerical simulation and actual logging data show that: 1) compared with FCM algorithm, SAGA–FCM has stronger stability and higher accuracy; 2) the proposed method can cluster the reservoir flexibly and effectively according to the degree of membership; 3) the results of reservoir integrated classification match well with the lithologic profile, which demonstrates the reliability of the classification method.
© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Tight oil is the oil that exists in the tight sandstone reservoir, carbonate reservoir, and other reservoirs with air permeability less than 1 mD and can be used for commercial exploitation through certain technical measures (National Energy Administration, 2017). Since the successful exploration and development of tight oil in some areas of the United States (Bakken tight oil, Eagle ford tight oil, etc.) in 2008, tight oil has become a hot-spot for oil and gas exploration as an important force to replace conventional oil and gas resources (Zou et al., 2018; Sun et al., 2010; Zhang et al., 2015). Influenced by the development of foreign tight oil and the current situation of domestic energy supply and demand, China has promoted the exploration and development of tight oil in recent years. The exploration and development of tight oil in China can be divided into two stages: the exploration stage before 2014; and the industrial production stage after 2014 (Zou and Qiu, 2021). By the end of 2019, the annual production of tight oil in China had reached 2.3×10^6 t (Zou and Qiu, 2021). Nowadays, rich tight oil and gas resources have been found in many areas in China, such as Ordos Basin and Songliao Basin (Sun et al., 2019; Xiao, 2015).

As the main reservoir of tight oil, tight sandstone reservoir has the characteristics of low pore, low permeability, and complex pore structure (Liu, 2021). The reservoir productivity is closely related to poro-permeability and pore structure characteristics (Liu, 2021). The research on integrated classification method of tight sandstone reservoir is of great significance for the exploration and development of oil and gas resources. Zhang et al. (2012) selected oil viscosity, clay mineral content, threshold pressure gradient, saturation of movable fluid, and throat radius to establish the "five factors integrated classification index", and classified the low-permeability reservoirs. In the same year, Zhao and Du (2012) evaluated the tight oil reservoir using "seven properties characteristics": lithology, poro-permeability property, petroliferous property, electrical property, source rock characteristics, brittleness, and in-situ stress anisotropy. Zou et al. (2014) evaluated the "sweet spots" of the unconventional reservoir using "six properties characteristics": lithology, poro-permeability property, petroliferous property,

* Corresponding author.
E-mail address: xieranhong@cup.edu.cn (R.-H. Xie).

https://doi.org/10.1016/j.petsci.2023.04.014
1995-8226/© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

electrical property, brittleness, and in-situ stress anisotropy. Xiao (2015) established the "seven factors integrated classification index" by adding the formation pressure coefficient and brittleness index (BI) on the basis of the "five factors integrated classification index". Wang et al. (2016) considered the reservoir pressure coefficient and established the "six factors integrated classification index" based on the "five factors integrated classification index". Hu (2017) classified the tight oil reservoir using the maximum throat radius, flow zone index (FZI), porosity, BI, the range of two-phase seepage zone, and threshold pressure gradient. On the basis of Zhang et al. (2012) and Xiao (2015), Zhang (2019) established the "eight factors integrated classification index" to classify the tight oil reservoir. Wang et al. (2020) evaluated the "sweet spots" of Triassic Chang 7 Formation in Ordos Basin using fuzzy cluster means and Bayes discrimination. Wei et al. (2021) evaluated the "sweet spots" of shale oil using the characterization parameters of petroliferous sweet spot, mobility sweet spot, poro-permeability sweet spot, and engineering sweet spot. Wang et al. (2021) classified and evaluated shale oil reservoir using total organic carbon content (TOC), porosity, oil content, free hydrocarbon content (S_1), oil saturation, BI, stress difference, and reservoir thickness. Zheng et al. (2021) used TOC, S_1, porosity, oil saturation, BI, and fracture pressure to establish the integrated evaluation standard of shale oil reservoir, and determined enriched layer of shale oil. Zhou et al. (2021) selected burial depth, TOC, brittle mineral content, pressure coefficient, vitrinite reflectance, porosity, and permeability to evaluate the favorable shale gas exploration areas using multi-parameter linear regression model. Xie et al. (2022) calculated the reservoir integrated evaluation factors using NMR T_2 distribution, matrix porosity and permeability, and formation micro-resistivity scanning imaging (FMI) logging data, and evaluated the reservoir effectiveness of the igneous reservoir. In the same year, Qi et al. (2022) established the "seven properties characteristics" evaluation model of shale oil reservoir, and evaluated the "sweet spots" of shale reservoir.

For the integrated classification of reservoir, the existing "integrated classification index" mainly reflect the lithologic characteristics, poro-permeability properties and engineering quality characteristics, but lack the characteristics of pore structure. Meanwhile, in the area lacking the prior information or core experiments, the classification standard is difficult to determine, and the method of "integrated classification index" is not applicable. To solve the aforementioned problems, based on previous studies, this paper proposed an integrated classification method of tight sandstone reservoir based on principal component analysis–simulated annealing genetic algorithm–fuzzy cluster means (PCA–SAGA–FCM). A variety of evaluation parameters, i.e., lithology characteristic evaluation parameters, poro-permeability quality characteristic evaluation parameters, engineering quality characteristic evaluation parameters, and pore structure characteristic evaluation parameters were selected. The unsupervised classification of tight sandstone reservoir was realized using the aforementioned evaluation parameters. The analysis results of numerical simulation and actual logging data verify the effectiveness of the classification method.

2. Methodology and theory

Fuzzy cluster means (FCM) is an unsupervised and flexible classification method, and the principles are shown in Appendix A. Although FCM has the advantages of unsupervised clustering and fast searching rate. However, FCM is a local search algorithm, and the selection of the value of the clustering center will affect the accuracy of the clustering results. If the clustering center value is input randomly, the clustering result tends to fall into the local optimization. In this study, the simulated annealing genetic algorithm (SAGA) (Wu et al., 2021) was selected to combine with the FCM to improve the global search ability and the stability of clustering results. Because the number of input parameters will affect the clustering speed, PCA (Appendix B) was used to reduce the dimension of the data when there are too many input parameters. The flow chart of PCA–SAGA–FCM is shown in Fig. 1.

1) Input the characteristic parameters, obtain the low dimensional data using PCA. And then the low dimensional data is used as the input of SAGA–FCM. In this study, the C_{cr} was selected to determine the number of principal components;
2) Initialization of the parameters of SAGA. The number of individuals in the population was 10. The maximum number of generations (g_{max}) was 10. The crossover probability and mutation probability were 0.6 and 0.01 respectively. T_c is the current temperature is T_c, T_0 is the initial temperature, and T_l is the minimum temperature. In this paper, T_0 and T_l were 100 °C and 1 °C respectively. Gradient of temperature reduction G_T was 0.8;
3) Initialization of the parameters of FCM. The number of categories m was 3 (in this paper, the reservoir was mainly divided into type I, type II and type III reservoirs). The maximum number of iterations was 20. The membership weighting factor b was 3. The minimum value of the change of fitness function (i.e., convergence accuracy) was 10^{-6}. The fitness function F is expressed as:

$$F = \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}{}^{b} d_{ij}{}^{2} \tag{1}$$

4) Generate the initial population and calculate the F value of individual. Each individual in the population contains the information of randomly generated clustering centers;
5) Initialize the evolution number gen = 1;
6) Complete the evolutionary process using SAGA, output the global optimal solution, i.e., the membership matrix $\boldsymbol{U}$ and the information of clustering center. For more details of SAGA, refer to Wu et al. (2021);
7) Complete the clustering according to the degree of membership.

3. Numerical simulation

In order to verify the stability of SAGA-FCM, 300 data points ($0 < x < 1$, $0 < y < 1$) were randomly constructed (Fig. 2). The classification results using FCM and SAGA–FCM for four times were randomly selected, are shown in Figs. 3 and 4. The triangles in figures are the center of different categories. The comparison of F values of different methods in 500 classifications are shown in Fig. 5.

Fig. 3 shows that, the classification results using FCM are affected by the selection of clustering center, and the results of each classification are different. Fig. 4 shows that, SAGA can effectively avoid the influence of improper selection of clustering centers on the classification results of FCM. The classification results using SAGA–FCM are stable and unchanged. In Fig. 5, the blue line represents the F value of FCM (F_{FCM}), and the red line represents the F value of SAGA–FCM (F_{SF}). It can be seen that the value of F_{FCM} oscillates violently and is greater than that of F_{SF}. The above analysis shows that SAGA–FCM has strong stability and high accuracy. The classification result of FCM is easy to fall into the local optimization, while SAGA–FCM can effectively overcome this problem and accurately unsupervised classification.

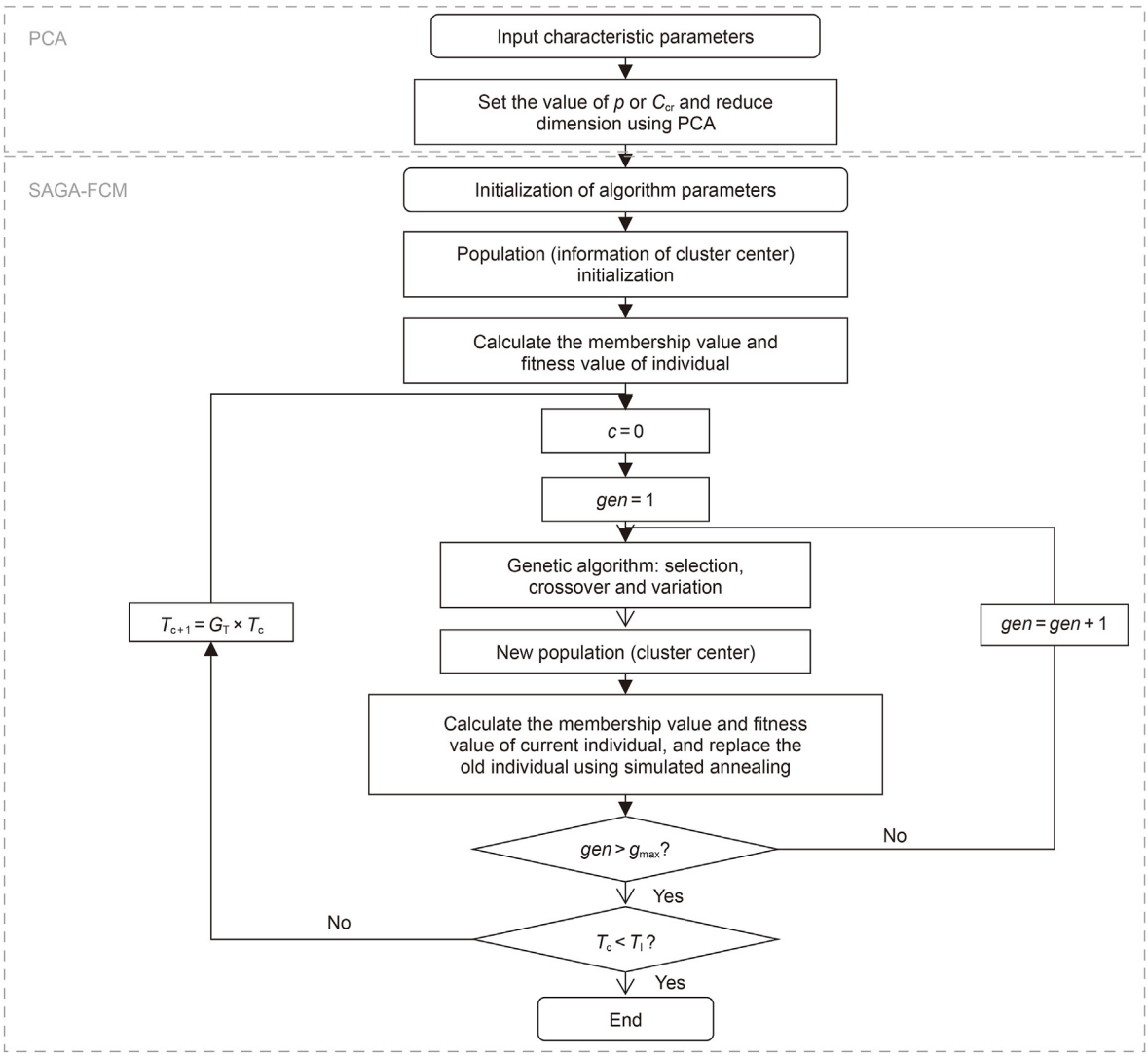

Fig. 1. Flow chart of the PCA-SAGA-FCM method.

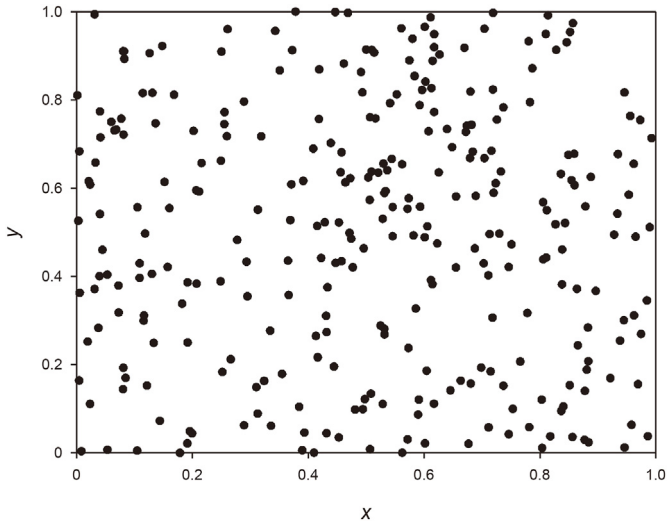

Fig. 2. The data points constructed randomly.

4. Calculation of evaluation parameters

In this paper, the tight sandstone reservoir section of well D (XX70–XX60 m), and well F (XX00–XX60 m) of Ordos Basin were selected. The evaluation parameters were selected from the four aspects, i.e., lithology characteristic, poro-permeability quality characteristic, engineering quality characteristic, and pore structure characteristic, to classify the reservoir without supervision. The logging data and the experimental results of core samples of well D and well F are shown in Figs. 6 and 7. Figs. 6 and 7 include: the depth; lithologic section; caliper (CAL) logs, gamma-ray (GR) logs, and spontaneous potential (SP) logs; acoustic (AC) logs, density (DEN) logs, and compensated neutron (CNL) logs; resistivity (AT10, AT20, AT30, AT60, AT90, RT) logs; NMR echo data; NMR T_2 distributions after inversion (Guo et al., 2019); array acoustic logs (compressional wave time difference (DT_c) and shear wave time difference (DT_s)); core porosity (*Cpor*); core permeability (*Cperm*); core oil saturation (CS_o); shale content (V_{sh}); photos of core samples; thin section images of core samples; scanning electron microscope (SEM) images of core samples; mercury injection capillary pressure (MICP) curves of core samples; pore size distribution (PSD) curves of core samples.

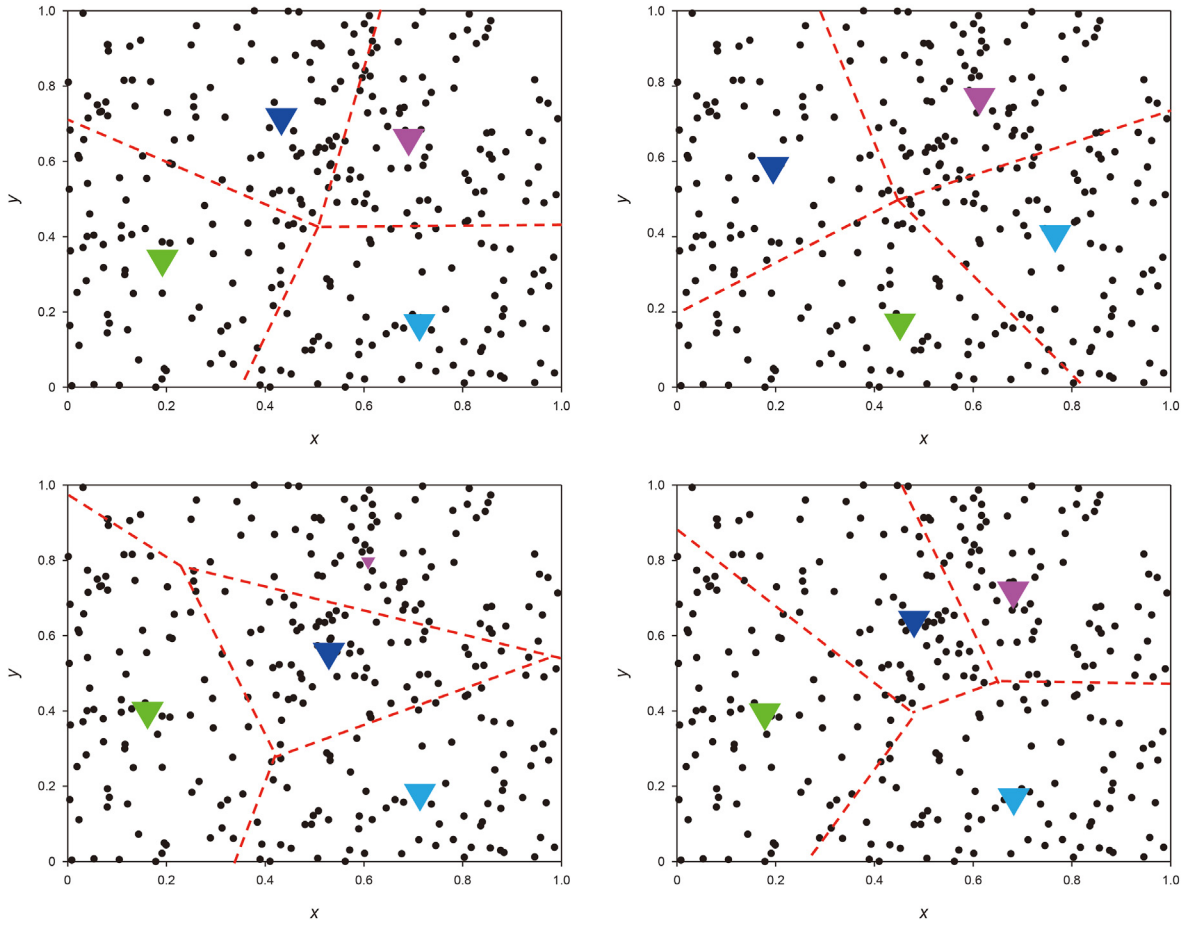

Fig. 3. Classification results of the data points using FCM (results of 4 times).

The lithology characteristic evaluation parameters include: reservoir thickness H and V_{sh}. H is the thickness of the interpretation reservoir section. V_{sh} can be calculated using GR logs (Hong, 2008). Since V_{sh} is negatively correlated with the reservoir quality, the reciprocal was taken as the input.

The poro-permeability quality characteristic evaluation parameters include: porosity φ, permeability k, and oil saturation S_o. According to the logging data, in this paper, NMR logging data was used to calculate the φ and k. The NMR porosity (POR_NMR) can be effectively calculated using NMR T_2 distribution (Dunn and Xie, 2010):

$$POR\_NMR = \sum_{i=1}^{N_{T_2}} f_i \tag{2}$$

where N_{T_2} is the number of T_2 components, f_i is the value of the ith porosity component.

The main models for predicting permeability using NMR data are the Timur-Coates (TC) model and the Schlumberger Doll Research Center (SDR) model (Timur, 1969; Coates and Dumanoir, 1973; Kenyon et al., 1988). Since the NMR saturation-centrifugation experimental data of core samples were not collected, in this paper, the SDR model was selected to calculate the permeability k_{SDR}:

$$k_{SDR} = A_1 \varphi^{B_1} {T_{2lm}}^{C_1} \tag{3}$$

where A_1, B_1, and C_1 are model coefficients, T_{2lm} is T_2 geometric mean (Eq. (4)) (Wu et al., 2020).

$$T_{2lm} = 10\text{^}\left(\sum_{i=1}^{N_{T_2}} f_i \log(T_{2,i}) \Big/ \sum_{i=1}^{N_{T_2}} f_i\right) \tag{4}$$

The well D selected in this paper mainly presents the lithologic characteristics of mainly fine sandstone and sand-shale interbedding (Fig. 6). The salinity of formation water of well D (XX70–XX60 m) was calculated using the SP logs, which is 10650 ppm (Hong, 2008). In this paper, the sedimentary characteristics and the salinity of formation water were considered, the Indonesian saturation equation (Eq. (5)) was used to calculate the oil saturation (S_o) of well D.

$$S_o = 1 - \left[\frac{{V_{sh}}^{0.5(2-V_{sh})}}{(R_{sh}/R_t)^{0.5}} + \left(\frac{R_t}{aR_w{*}\varphi^{-m}}\right)^{0.5}\right]^{-\frac{2}{n}} \tag{5}$$

where a is the lithology coefficient, m and n are cementation index and saturation index respectively, R_w is the formation water resistivity, R_t is the resistivity of original formation, R_{sh} is the resistivity of adjacent mudstone. Since the data of rock-electric experiments were not collected, the regional experienced values were selected for a, m, and n when calculating the oil saturation (Huang et al., 2015; Shi, 2012; Xu et al., 2021).

The engineering quality characteristic evaluation parameter is Brittleness index BI. In this paper, based on the collected array acoustic logging data, BI was calculated using Young's modulus (E)

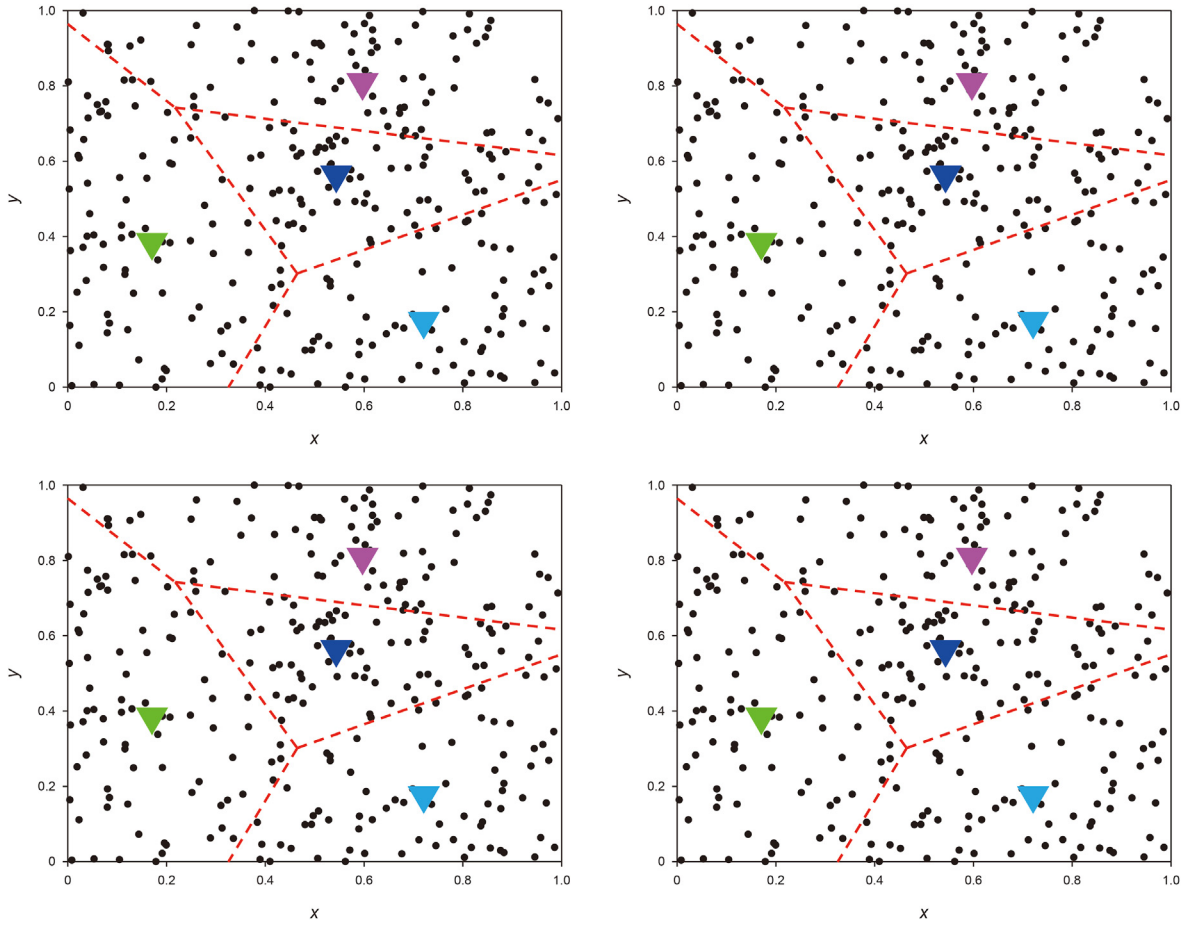

Fig. 4. Classification results of the data points using SAGA-FCM (results of 4 times).

and Poisson's ratio (v) (Eq. (6) - (8)). The larger the E and the smaller the v, the larger the BI, and the easier it is to form network fractures after fracturing (Lai et al., 2016).

$$E_{s\_n} = \frac{E_s - \min(E_s)}{\max(E_s) - \min(E_s)} \tag{6}$$

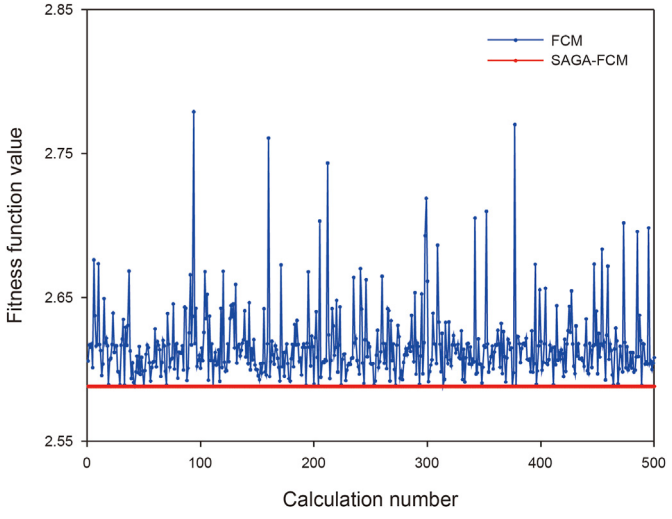

Fig. 5. Comparison of classification results using different methods.

$$v_{s\_n} = \frac{v_s - \min(v_s)}{\min(v_s) - \max(v_s)} \tag{7}$$

$$BI = \frac{E_{s\_n} + v_{s\_n}}{2} \times 100\% \tag{8}$$

where E_s is the static Young's modulus, v_s is the Poisson's ratio, E_{s_n} and v_{s_n} is the normalized value of E_s and v_s, respectively.

The dynamic Young's modulus (E_d) and dynamic Poisson's ratio (v_d) were calculated using density logs value (ρ_b), compressional wave time difference (DT_c), and shear wave time difference (DT_s) (Eqs. (9) - (10)) (Lai et al., 2016).

$$E_d = \frac{\rho_b}{DT_s^2}\left(\frac{3DT_s^2 - 4DT_c^2}{DT_s^2 - DT_c^2}\right) \tag{9}$$

$$v_d = \frac{DT_s^2 - 2DT_c^2}{2\left(DT_s^2 - DT_c^2\right)} \tag{10}$$

In practical application, the calculation of brittleness index using array acoustic logging data mainly includes the following steps.

1) Obtain the E_s and v_s of core samples using rock mechanics testing;
2) Obtain the ρ_b, DT_c, and DT_s of core samples;
3) Calculate the E_d and v_d of core samples using Eqs. (9) and (10);

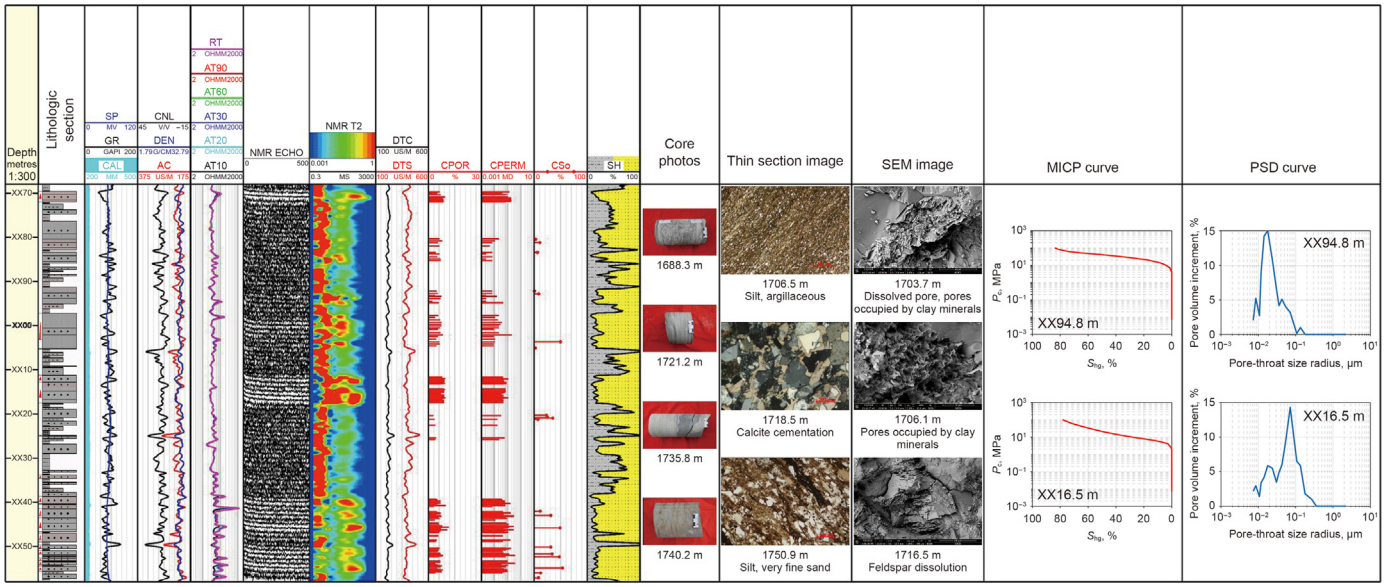

Fig. 6. Logging data and experimental results of core samples of well D. The first track is the depth. The second track is lithologic section. The third track includes CAL, GR, and SP logs. The fourth track includes AC, DEN, and CNL logs. The fifth track includes AT10, AT20, AT30, AT60, AT90, and RT logs. The sixth track is NMR echo data. The seventh track is NMR T_2 distribution. The eighth track includes DT_c and DT_s logs. The ninth track is *Cpor*. The tenth track is *Cperm*. The eleventh track is CS_o. The twelfth track is V_{sh}. The thirteenth track is the photos of core samples. The fourteenth track is the thin section image. The fifteenth track is the SEM image. The sixteenth track is the MICP curve. The seventeenth track is the PSD curve.

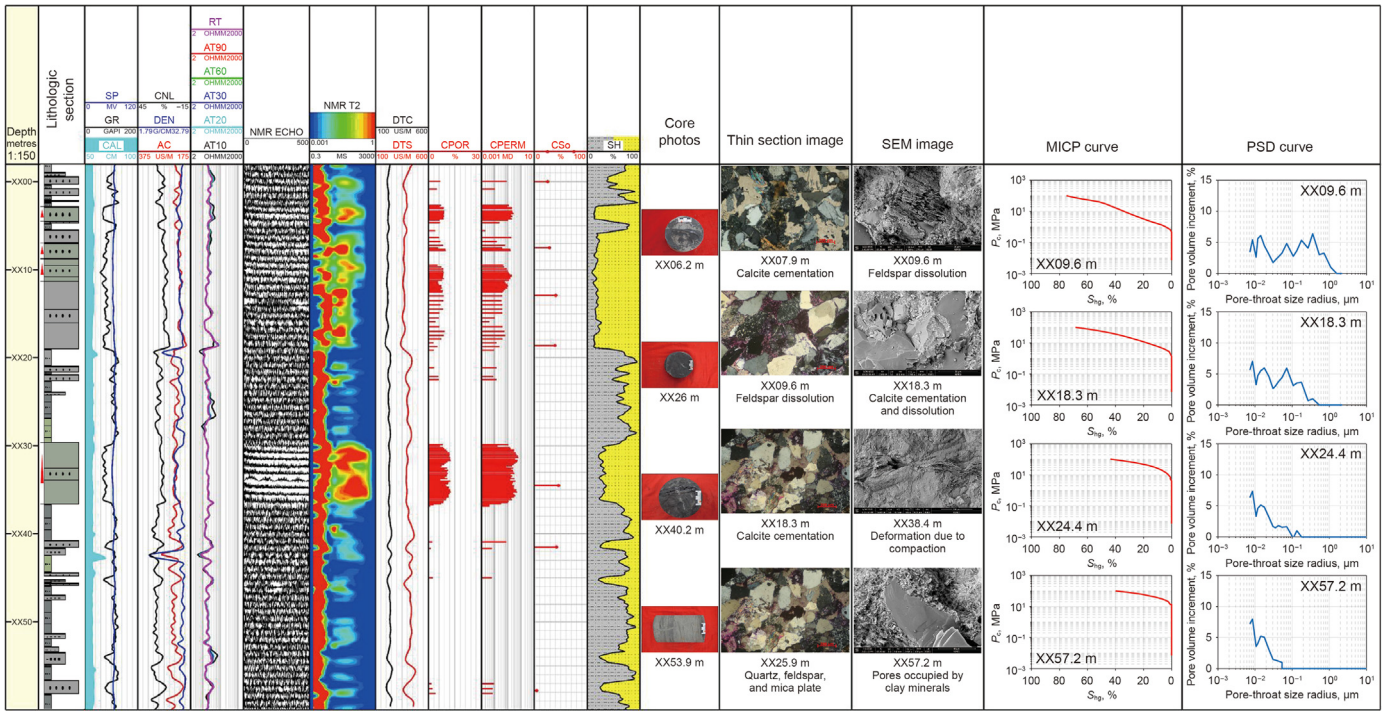

Fig. 7. Logging data and experimental results of core samples of well F.

4) Establish the conversion equations of between dynamic parameters (E_d, ν_d) and static parameters (E_s, ν_s);
5) Calculate the E_d and ν_d of the reservoir using well logging data;
6) Calculate the E_s and ν_s of the reservoir using E_d, ν_d, and the conversion equations;
7) Calculate the *BI* of the reservoir using Eqs. (6)–(8).

The pore structure characteristic evaluation parameters mainly calculated by the NMR T_2 distribution, including T_{2lm} and Gaussian characteristic parameter $w_{2*}\mu_2$. T_{2lm} represents the overall characteristics of PSD. $w_{2*}\mu_2$ represents the characteristics of large pores related to seepage capacity (Xu and Torres-verdín, 2013; Li et al., 2019).

The Gaussian characteristic parameters can be calculated by fitting the NMR T_2 distribution using Gaussian distribution function. The Gaussian distribution function *GF* can be expressed as:

$$GF(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{11}$$

where μ and σ are the mean value and standard deviation of Gaussian distribution, respectively.

For the fitting of the multimodal distribution curve, multiple Gaussian distribution function (*MGF*) (Eq. (12)) can be selected. The cumulative curve of multimodal distribution (Eq. (13)) is approached by the cumulative curve of *MGF* (*CGF*), and the optimal Gaussian characteristic parameters are obtained when the error between them is minimized (Fig. 8).

$$\begin{aligned}MGF(x,w_i,\mu_i,\sigma_i)=&w_1\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}+w_2\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}\\&+\cdots+w_i\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2}+\cdots\\&+w_n\frac{1}{\sqrt{2\pi}\sigma_n}e^{-\frac{1}{2}\left(\frac{x-\mu_n}{\sigma_n}\right)^2}\end{aligned} \tag{12}$$

$$\begin{aligned}CGF(x,w_i,\mu_i,\sigma_i)=&\frac{w_1}{2}\left[1+\frac{2}{\sqrt{\pi}}\int_0^{\frac{(x-\mu_1)}{\sqrt{2}\sigma_1}}e^{-t^2}dt\right]+\frac{w_2}{2}\left[1\right.\\&\left.+\frac{2}{\sqrt{\pi}}\int_0^{\frac{(x-\mu_2)}{\sqrt{2}\sigma_2}}e^{-t^2}dt\right]+\cdots+\frac{w_i}{2}\left[1\right.\\&\left.+\frac{2}{\sqrt{\pi}}\int_0^{\frac{(x-\mu_i)}{\sqrt{2}\sigma_i}}e^{-t^2}dt\right]+\cdots+\frac{w_n}{2}\left[1\right.\\&\left.+\frac{2}{\sqrt{\pi}}\int_0^{\frac{(x-\mu_n)}{\sqrt{2}\sigma_n}}e^{-t^2}dt\right]\end{aligned} \tag{13}$$

where w_i is the weight coefficient, which conforms to Eq. (14):

$$\sum_{i=1}^{n}w_i=1\quad 0<w_i<1 \tag{14}$$

In this paper, the double Gaussian distribution model was used to fit the NMR T_2 distribution based on the process shown in Fig. 8, and the Gaussian characteristic parameters were obtained. When fitting the NMR T_2 distribution, Eq. (15) was used for pretreatment:

$$\begin{cases}T_2{}' = \log(T_2)\\ f_i{}' = \dfrac{f_i}{\sum\limits_{i=1}^{\text{bin}} f_i}\end{cases} \tag{15}$$

where T_2' is the T_2 value after pretreatment, f_i' is the value of the normalized porosity component.

T_2' was used as the input of Eq. (13), and the optimal Gaussian distribution parameter was determined by calculating the minimum value of Eq. (16).

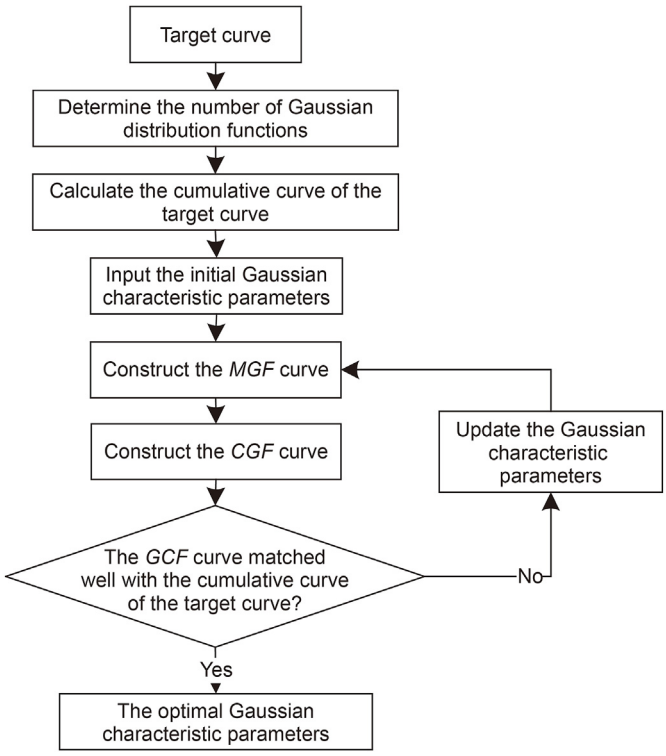

Fig. 8. Flow chart of curve fitting based on *MGF*.

$$f(P_G)=\left\|CGF(P_G)-C_{T_{2dist}}\right\|_2^2 \tag{16}$$

where P_G is the parameter of double Gaussian distribution model, including w_1, w_2, $\log(\mu_1)$, $\log(\mu_2)$, $\log(\sigma_1)$, and $\log(\sigma_2)$, $CGF(P_G)$ is the cumulative curve of double Gaussian distribution under the current model parameters, $C_{T_{2dist}}$ is the positive cumulative curve of T_2 distribution (i.e., from left to right). The optimal P_G (P_{GO}) was determined when the value of $f(P_G)$ reaches Eq. (17).

$$\min\left\|CGF(P_G)-C_{T_{2dist}}\right\|_2^2 \tag{17}$$

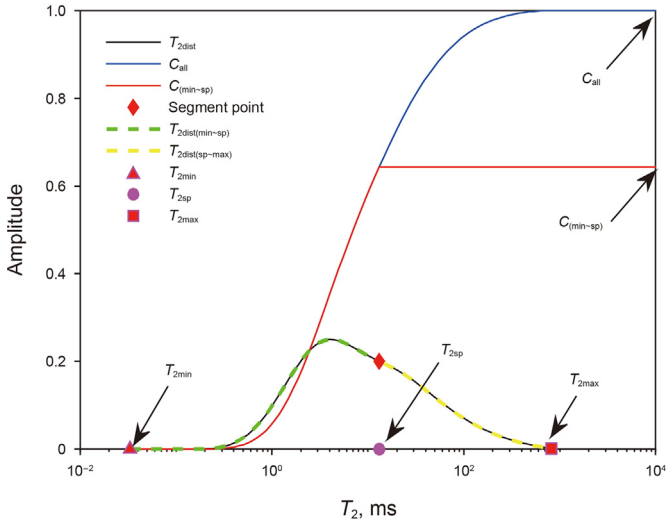

Fig. 9. Calculation of initial parameters of double Gaussian distribution model.

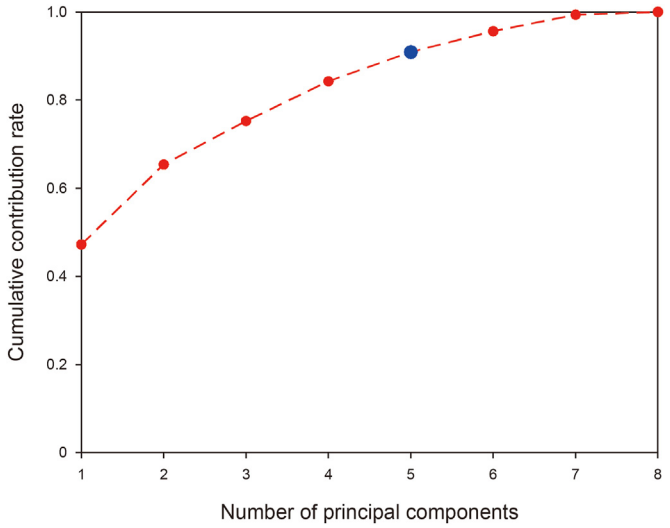

Fig. 10. Cumulative contribution rate under different number of principal components.

Since the calculation accuracy of P_{GO} is affected by the initial value of P_G. In this paper, the segmentation point between large and small pores (T_{2sp}) was obtained according to the method proposed by Wu et al. (2021). The initial weight coefficient, mean value and standard deviation of the large and small pores were calculated respectively (Eq. (18)), and then they were used as the initial parameters of Eqs. (12) and (13).

$$\begin{cases} w_1 = C_{(\min\sim sp)} \Big/ C_{all} \\ w_2 = 1 - w_1 \\ \mu_1 = \text{mean}\left(T_{2(\min\sim sp)}\right) \\ \mu_2 = \text{mean}\left(T_{2(sp\sim\max)}\right) \\ \sigma_1 = \text{std}\left(T_{2\text{dist}(\min\sim sp)}\right) \\ \sigma_2 = \text{std}\left(T_{2\text{dist}(sp\sim\max)}\right) \end{cases} \tag{18}$$

where $C_{(\min\sim sp)}$ is the amplitude value accumulated from the minimum T_2 to T_{2sp} in T_2 distribution, C_{all} is the overall cumulative amplitude value of T_2 distribution, $T_{2(\min\sim sp)}$ is the T_2 value from the minimum T_2 to T_{2sp}, $T_{2(sp\sim\max)}$ is the T_2 value from the T_{2sp} to the maximum T_2, $T_{2\text{dist}(\min\sim sp)}$ and $T_{2\text{dist}(sp\sim\max)}$ are the T_2 distributions in different T_2 intervals (Fig. 9).

The parameters of the double Gaussian distribution model have the following poro-permeability implications (Xu and Torres-verdín, 2013; Li et al., 2019).

1) Weight coefficient: w_1 and w_2 represent the volume fraction of small and large pores respectively. The small pores are mainly filled with the bound fluid. The large pores are mainly filled with the movable fluid, which controls the percolation capacity of the reservoir;
2) Mean value: μ_1 and μ_2 represents the mean value of small and large pores respectively. The values of μ_1 and μ_2 increases with the poro-permeability quality;
3) Standard deviation: σ_1 and σ_2 represents the standard deviation of the pore radius of small and large pores respectively. The standard deviation characterizes the uniformity of PSD.

Previous studies have indicated that the Gaussian characteristic parameters of NMR T_2 distribution can be used to quantitatively evaluate the reservoir parameters and characterize the reservoir pore structure (Li et al., 2019; Ge et al., 2014).

5. Integrated classification and pore structure evaluation

PCA-SAGA-FCM was used for the reservoir integrated classification. The H, SH, φ, k, S_o, BI, T_{2lm}, and $w_{2*}\mu_2$ were taken as the input evaluation parameters. The value of C_{cr} was set to 0.9. The cumulative contribution rate under different number of principal components is shown in Fig. 10. Fig. 10 shows that when the number of principal components is 5, the cumulative contribution rate reaches 0.9. In this paper, the first five principal components were selected as the input when using the aforementioned parameters for classification.

The classification results of well D and well F are shown in Figs. 11 and 12. The first to seventh tracks in Figs. 11 and 12 are the same as those in Figs. 6 and 7. The eighth to tenth track include the

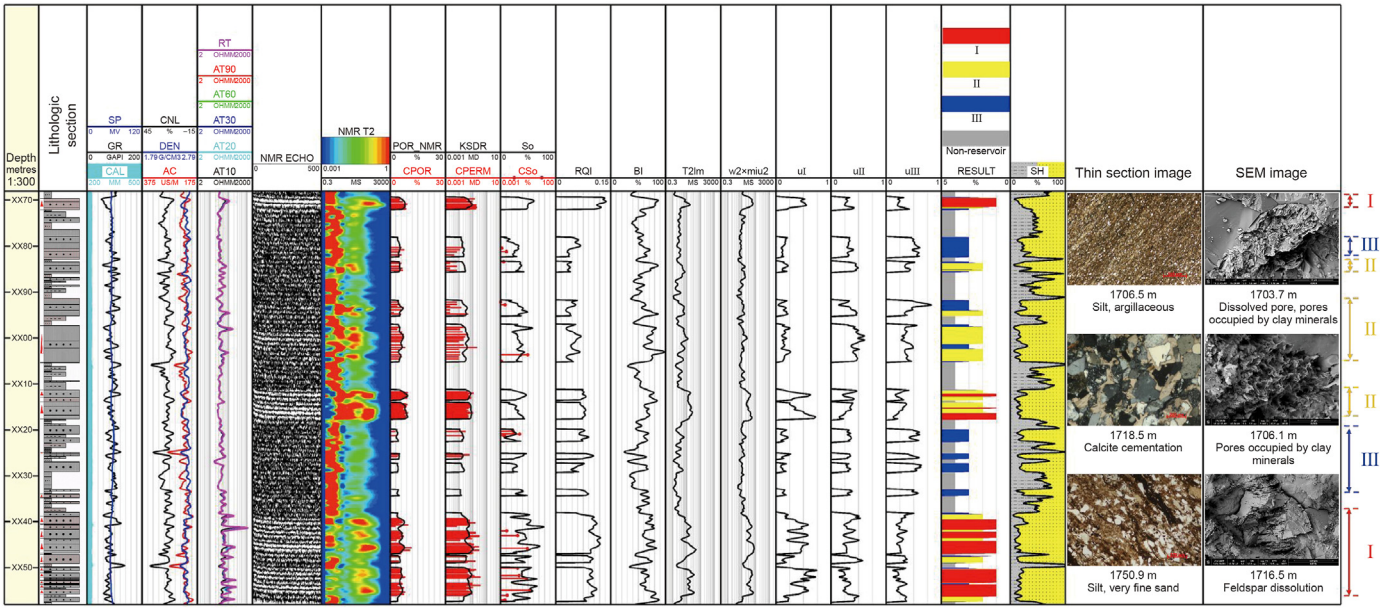

Fig. 11. Result of reservoir integrated classification of well D.

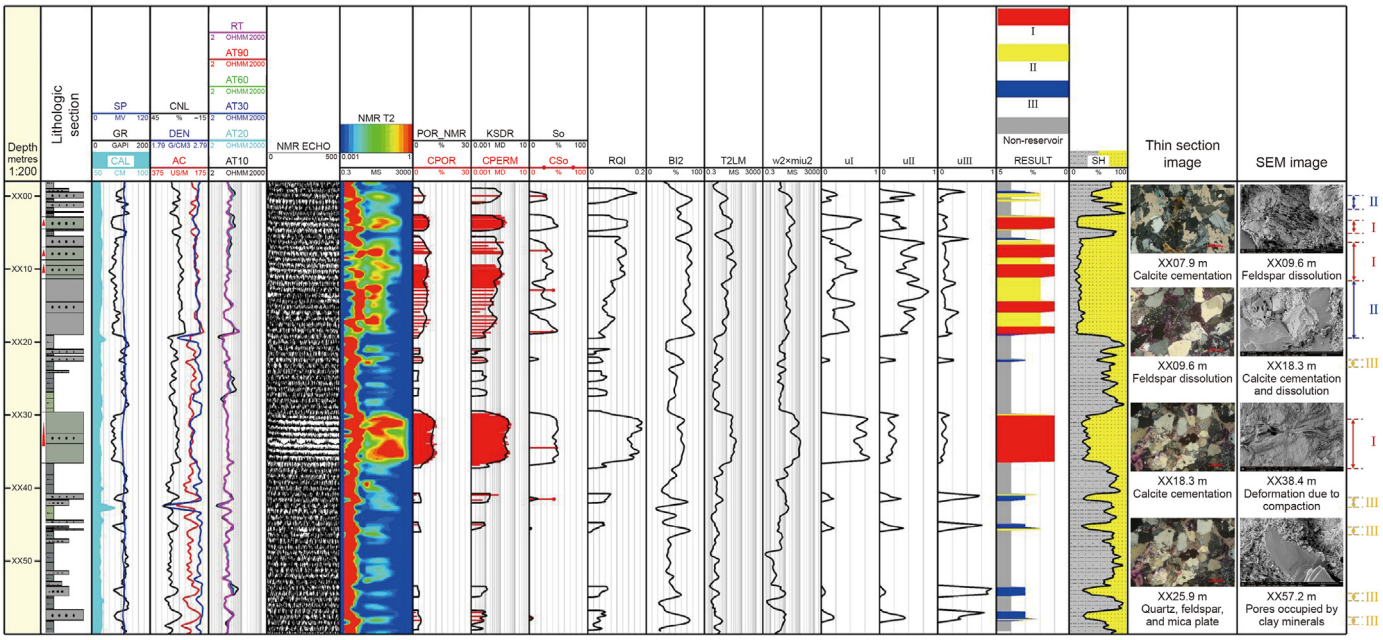

Fig. 12. Result of reservoir integrated classification of well F.

porosity, permeability and oil saturation of core samples (*Cpor*, *Cperm*, and CS_o), and the calculated porosity, permeability and oil saturation (*POR_NMR*, k_{SDR}, and S_o). The eleventh track is the reservoir quality index (*RQI*) (Wu et al., 2021). The twelfth track is the *BI*. The thirteenth and fourteenth tracks include T_{2lm} and $w_2{*}\mu_2$. The fifteenth to seventeenth tracks represent the degree of membership of type I reservoir, type II reservoir, and type III reservoir at each depth point (u_I, u_{II}, and u_{III}). The eighteenth track is the integrated classification results of the reservoir: red represents type I reservoir, yellow represents type II reservoir, and blue represents type III reservoir, and gray represents the non-reservoir. The nineteenth track is V_{sh}. The twentieth and twenty-first tracks include the thin section image and the SEM image.

The comparison of u_1, u_2, and u_3 are shown in Fig. 13. Figs. 11–13 show :

1) *POR_NMR*, k_{SDR}, and S_o matched well with *Cpor*, *Cperm*, and CS_o, indicating the reliability of the calculated results of porosity, permeability, and saturation.
2) The *BI* is negatively correlated with V_{sh}. With the increase of V_{sh}, the *BI* decreases.

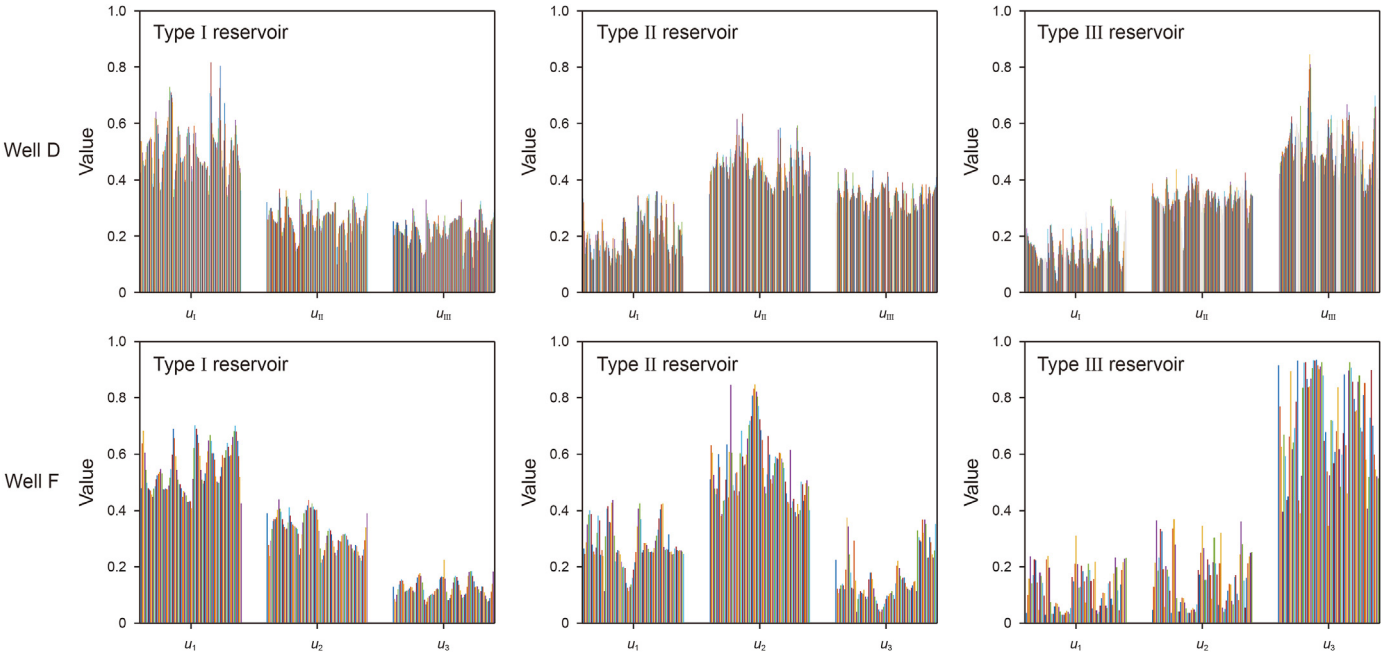

Fig. 13. Comparison of the degree of membership of different types of the reservoir.

3) T_{2lm} and $w_{2*}\mu_2$ are positively correlated with permeability. As the reservoir heterogeneity increases, the permeability, *RQI*, T_{2lm}, and $w_{2*}\mu_2$ decreases.
4) The cluster subsets all correspond well to the degree of membership. The u_1 is the largest in type I reservoir, u_2 is the largest in type II reservoir, and u_3 is the largest in type III reservoir. Flexible clustering of the reservoir can be effectively performed using the degree of membership (Ge et al., 2022).
5) Since the formation testing data were not collected, the classification results were compared with the lithologic section in this paper. The result of reservoir classification using PCA–SAGA–FCM matched well with the lithologic section. In the type I reservoir, the reservoir thickness is thick, the evaluation parameters are large, and the core descriptions all show oil-bearing characteristics, for example, XX38–XX57 m of well D, and XX30–XX36 m of well F. In the type II reservoir, the reservoir thickness is medium, the evaluation parameters are medium, and the core descriptions partially show oil-bearing characteristics, for example, XX92–XX05 m of well D, and XX12–XX18 m of well F. The type III reservoir is mainly composed of thin interbeds, with low evaluation parameter values and no oil-bearing features in core description, for example, XX27–XX29 m of well D, and XX54–XX55 m of well F.

6. Conclusions

In this paper, an integrated classification method based on PCA–SAGA–FCM was proposed. The lithology characteristic evaluation parameters, poro-permeability quality characteristic evaluation parameters, engineering quality characteristic evaluation parameters, and pore structure characteristic evaluation parameters were selected as inputs, and the unsupervised classification of the tight sandstone reservoir was realized. The conclusions of this paper are as follows.

1) Compared with FCM algorithm, SAGA–FCM has stronger stability and higher accuracy. SAGA–FCM can effectively overcome the problem of falling into local optimization, and can be used for unsupervised classification accurately.
2) The integrated classification results of the reservoir using PCA–SAGA–FCM matched well with the lithologic section, indicating the reliability of the classification method. In practical applications, accurate reservoir classification lays a foundation for the subsequent "sweet spots" evaluation.
3) The classification method proposed in this paper does not need prior information or core experimental data. Compared with the traditional method using "multi factors integrated classification index", the method proposed in this paper has certain advantages. In practical applications, the input parameters and the number of principal components can be determined according to the actual situation of the study area. The number of categories can be set in advance or determined according to some information criteria.

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

The work was funded by the National Natural Science Foundation of China (42174131) and the Strategic Cooperation Technology Projects of CNPC and CUPB (ZLZX2020-03).

Appendix A

Principal component analysis (PCA) is the transformation of linearly correlated data into linearly uncorrelated data using orthogonal transformation. The dimensionality of the data can be reduced by extracting the principal components of the original data. The steps of PCA include.

1) Input the sample dataset $\boldsymbol{X}$:

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1b} \\ x_{21} & x_{22} & \cdots & x_{2b} \\ \vdots & \vdots & & \vdots \\ x_{a1} & x_{a2} & \cdots & x_{ab} \end{bmatrix} \tag{A.1}$$

where a is the number of characteristic parameters in $\boldsymbol{X}$, b is the number of samples in $\boldsymbol{X}$.

2) Centralization of sample data :

$$\overline{\boldsymbol{X}}=\boldsymbol{X}-\begin{bmatrix} \overline{x_1} \\ \vdots \\ \overline{x_a} \end{bmatrix}=\boldsymbol{X}-\begin{bmatrix} \frac{1}{b}\sum_{k=1}^{b} x_{1k} \\ \vdots \\ \frac{1}{b}\sum_{k=1}^{b} x_{ak} \end{bmatrix} \tag{A.2}$$

where $\overline{\boldsymbol{X}}$ is dataset after centralization, $\overline{x_a}$ is the average value of each characteristic parameter.

3) Calculate the covariance ($\boldsymbol{Cov}$) :

$$\boldsymbol{Cov}=\frac{1}{j}\overline{\boldsymbol{X}}\overline{\boldsymbol{X}}^{\mathrm{T}} \tag{A.3}$$

4) Calculate the feature value λ and the corresponding feature vector e of $\boldsymbol{Cov}$:

$$\boldsymbol{Cov}=\boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^{-1} \tag{A.4}$$

where $\boldsymbol{\Lambda}$ a diagonal matrix consisting of λ in descending order, $\boldsymbol{E}$ an orthogonal matrix composed of e: $\boldsymbol{E}=[e_1, \ldots, e_a]^{\mathrm{T}}$.

5) Select the first p feature values and the corresponding feature vectors, and reduce the dimensionality of $(\boldsymbol{X})_{a\times b}$ to obtain $(\boldsymbol{X}')_{p\times b}$ (Eq. (A-5)). The value of p can be set artificially or calculated according to the cumulative contribution of principal components (Eq. (A-6)).

$$\boldsymbol{X}'=\boldsymbol{E}_{\boldsymbol{p}}\times\boldsymbol{X} \tag{A.5}$$

where $\boldsymbol{E}_{\boldsymbol{p}}$ is the orthogonal matrix composed of the selected feature vectors: $\boldsymbol{E}_{\boldsymbol{p}}=[e_1, \ldots, e_p]^{\mathrm{T}}$ $(p<a)$.

$$C_{\mathrm{cr}}=\sum_{l=1}^{p}\lambda_l \Big/ \sum_{l=1}^{a}\lambda_l \tag{A.6}$$

where C_{cr} the cumulative contribution of principal components, $0<C_{\mathrm{cr}}<1$. The larger the value of p, the larger the C_{cr}, the more information is retained.

Appendix B

FCM classifies the dataset according to degree of membership u. Set the dataset $\mathbf{Y} = [y_1, y_2, \ldots, y_i, \ldots, y_n]$ and the number of categories m ($2 \le m \le n$), the dataset of each category is expressed as $\{A_1, A_2, \ldots, A_j, \ldots, A_m\}$, and the clustering center of each category is expressed as $\{c_1, c_2, \ldots, c_j, \ldots, c_m\}$. The number of categories can be set in advance or determined according to some information criteria, such as Akaike information criterion (AIC) (Ge et al., 2022). The objective function is expressed as:

$$\min\left(\sum_{i=1}^{n}\sum_{j=1}^{m} {u_{ij}}^{b} {d_{ij}}^{2}\right) \tag{B.1}$$

where d_{ij} is the distance from the ith data to the jth clustering center (Eq. (B-2)), u_{ij} is the degree of membership of the ith data to category A_j (Eq. (B-3)), n is the number of data, b is the membership weighting factor.

$$d_{ij} = \left\|y_i - c_j\right\|_2 = \sqrt{\sum_{l=1}^{g}\left(y_{il} - c_{jl}\right)^2} \tag{B.2}$$

where g is the number of characteristic parameters in Y.

$$u_{ij} = \frac{{d_{ij}}^{-\frac{2}{b-1}}}{\sum_{l=1}^{m} {d_{il}}^{-\frac{2}{b-1}}} \tag{B.3}$$

u_{ij} should conform to the following equation:

$$\sum_{j=1}^{m} u_{ij} = 1 \quad u_{ij} \in [0, 1] \tag{B.4}$$

c_j can be calculated using u_{ij} and y_i

$$c_j = \sum_{i=1}^{n} \frac{{u_{ij}}^{b}}{\sum_{i=1}^{n} {u_{ij}}^{b}} y_i \tag{B.5}$$

The steps of FCM algorithm include.

1) Set the number of categories, the membership weighting factor and other parameters, and randomly initialize the location of clustering center;
2) Update the membership matrix **U** using Eqs. (B-2)–(B-4). **U** is composed of the degree of membership of each data;
3) Update the location of clustering center using Eq. (B-5);
4) Judge whether the maximum number of iterations or objective function value is reached, if not, go to step 2);
5) Output the information of clustering center and u_{ij}.

References

Coates, G.R., Dumanoir, J.L., 1973. A new approach to improved log-derived permeability. In: SPWLA 14th Annual Logging Symposium.

Dunn, K., Xie, R., 2010. Nuclear Magnetic Resonance Logging Theory and Applications. China University of Petroleum Press (in Chinese).

Ge, X., Fan, Y., Cao, Y., et al., 2014. Reservoir pore structure classification technology of carbonate rock based on NMR T_2 spectrum decomposition. Appl. Magn. Reson. 45 (2), 155–167. https://doi.org/10.1007/s00723-013-0511-5.

Ge, X., Xue, Z., Zhou, J., et al., 2022. An unsupervised clustering method for NMR transverse relaxation spectrums based on the Gaussian mixture model and its application. Petrol. Explor. Dev. 49 (2), 296–305. https://doi.org/10.1016/S1876-3804(22)60028-4.

Guo, J., Xie, R., Jin, G., 2019. An efficient method for NMR data compression based on fast singular value decomposition. Geosci. Rem. Sens. Lett. IEEE 16 (2), 301–305. https://doi.org/10.1109/LGRS.2018.2872111.

Hong, Y., 2008. Logging Principle and Comprehensive Interpretation. China University of Petroleum Press (in Chinese).

Huang, W., Zhang, X., Zhao, J., 2015. Fluild identification of tight reservoir in Chang 9 formation,Ordos Basin. J. NW Univ. 45 (5), 811–818. https://doi.org/10.16152/j.cnki.xdxbzr.2015-05-023 (in Chinese).

Hu, Y., 2017. Research on Tight Oil Reservoirs Geological Evaluation and Classification of the Chang7 Member in Southern Ordos Basin. Northwest University (in Chinese).

Kenyon, W.E., Day, P.I., Straley, C., et al., 1988. A three-part study of NMR longitudinal relaxation properties of water-saturated sandstones. SPE Form. Eval. 3 (3), 622–636. https://doi.org/10.2118/15643-PA.

Lai, J., Wang, G., Fan, Z., et al., 2016. Research progress in brittleness index evaluation methods with logging data in unconventional oil and gas reservoirs. Petroleum Science Bulletin 1 (3), 330–341. https://doi.org/10.3969/j.issn.2096-1693.2016.03.028 (in Chinese).

Li, C., Liu, M., Guo, B., 2019. Classification of tight sandstone reservoirs based on NMR logging. Appl. Geophys. 16 (4), 549–558. https://doi.org/10.1007/s11770-019-0793-y.

Liu, G., 2021. Challenges and countermeasures of log evaluation in unconventional petroleum exploration. Petrol. Explor. Dev. 48 (5), 891–902. https://doi.org/10.11698/PED.2021.05.02 (in Chinese).

National Energy Administration., 2017. Geological Evaluating Methods for Tight Oil: GB/T 34906-2017. Petroleum Industry Press, Beijing (in Chinese).

Qi, H., Su, J., Hu, X., et al., 2022. Study on well logging technology for the comprehensive evaluation of the “seven properties” of shale oil reservoirs—an example of shale oil in the Lucaogou formation in the Jimsar Sag, Junggar Basin. Front. Earth Sci. 9. https://doi.org/10.3389/feart.2021.827380.

Shi, Y., 2012. Hydrocarbon Potential and Oil Rich Domain Evaluation with Logging Method in Tow Permeability Lithologic Reservoir: Taking Chang 8 Member of Yanhang Formation in Jiyuan Region in Ordos Basin as an Example. Northwest University (in Chinese).

Sun, L., Fang, C., Li, F., et al., 2010. Petroleum exploration and development practices of sedimentary basins in China and research progress of sedimentology. Petrol. Explor. Dev. 37 (4), 385–396 (in Chinese).

Sun, L., Zou, C., Jia, A., et al., 2019. Development characteristics and orientation of tight oil and gas in China. Petrol. Explor. Dev. 46 (6), 1015–1026. https://doi.org/10.11698/PED.2019.06.01 (in Chinese).

Timur, A., 1969. Pulsed nuclear magnetic resonance studies of porosity, movable fluid, and permeability of sandstones. J. Petrol. Technol. 21 (6), 775–786. https://doi.org/10.2118/2045-PA.

Wang, F., Fu, Z., Wang, J., et al., 2021. Characteristics and classification evaluation of Gulong shale oil reservoir in Songliao Basin. Pet. Geol. Oilfield Dev. Daqing 40 (5), 144–156. https://doi.org/10.19597/J.ISSN.1000-3754.202107017 (in Chinese).

Wang, X., Ren, L., He, Y., et al., 2016. Definition of tight oil in Ordos Basin. Petroleum Geology and Recovery Efficiency 23 (1), 1–7. https://doi.org/10.13673/j.cnki.cn37-1359/te.2016.01.001 (in Chinese).

Wang, W., Kang, S., Gao, F., et al., 2020. Classification and evaluation of tight oil reservoirs based on fuzzy c-means clustering and bayes discrimination. Special Oil Gas Reservoirs 27 (5), 118–124. https://doi.org/10.3969/j.issn.1006-6535.2020.05.018 (in Chinese).

Wei, Y., Li, J., Lu, S., et al., 2021. Comprehensive evaluation method of sweet spot zone in lacustrine shale oil reservoir and its application: a case study of shale oil in lower1st member of the Shahejie formation in the Raoyang sag. J. China Inst. Min. Technol. 50 (5), 813–824. https://doi.org/10.13247/j.cnki.jcumt.001223 (in Chinese).

Wu, B., Xie, R., Liu, M., et al., 2021. Novel method for predicting mercury injection capillary pressure curves of tight sandstone reservoirs using NMR T_2 distributions. Energy Fuel. 35 (19), 15607–15617. https://doi.org/10.1021/acs.energyfuels.1c02146.

Wu, B., Xie, R., Wang, X., et al., 2020. Characterization of pore structure of tight sandstone reservoirs based on fractal analysis of NMR echo data. J. Nat. Gas Sci. Eng. 81, 103483. https://doi.org/10.1016/j.jngse.2020.103483.

Xiao, Q., 2015. The Reservoir Evaluation and Porous Flow Mechanism for Typical Tight Oilfields. University of Chinese Acedemy of Sciences (Institute of Porous Flow & Fluid Mechanics) (in Chinese).

Xie, W., Si, Z., Xu, F., et al., 2022. A method for reservoir effectiveness evaluations of altered igneous reservoirs based on logging data: a case study of the Nanpu structure in eastern China. Arabian J. Geosci. 15 (3), 255. https://doi.org/10.1007/s12517-022-09487-4.

Xu, C., Torres-verdín, C., 2013. Pore system characterization and petrophysical rock classification using a bimodal Gaussian density function. Math. Geosci. 45 (6), 753–771. https://doi.org/10.1007/s11004-013-9473-2.

Xu, J., He, Y., Ma, F., et al., 2021. Effective reservoir thickness of main oil layers in Dingbian Oilfield, Ordos Basin. Lithologic Reservoirs 33 (5), 107–119. https://doi.org/10.12108/yxyqc.20210510 (in Chinese).

Zhang, J., Bi, H., Xu, H., et al., 2015. New progress and reference significance of overseas tight oil exploration and development. Acta Pet. Sin. 36 (2), 127–137. https://doi.org/10.7623/syxb201502001 (in Chinese).

Zhang, P., 2019. Research of Microscopic Pore Structure and Reservoir Evaluation of Dingbian Chang 7 Tight Reservoir in Yanchang Oilfield. Northwest University (in Chinese).

Zhang, Z., Yang, Z., Liu, X., et al., 2012. A grading evaluation method for low-

permeability reservoirs and its application. Acta Pet. Sin. 33 (3), 437–441 (in Chinese).
Zhao, Z., Du, J., 2012. Tight Oil and Gas. Petroleum Industry Press (in Chinese).
Zheng, J., Wang, C., Zhang, H., et al., 2021. Logging evaluating methods of seven property parameters and enriched layers for Gulong shale oil reservoir in Songliao Basin. Pet. Geol. Oilfield Dev. Daqing 40 (5), 87–97. https://doi.org/10.19597/J.ISSN.1000-3754.202107003 (in Chinese).
Zhou, Y., Zhao, A., Yu, Q., et al., 2021. A new method for evaluating favorable shale gas exploration areas based on multi-linear regression analysis: a case study of marine shales of Wufeng-Longmaxi Formations, Upper Yangtze Region. Sediment. Geol. Tethyan Geol. 41 (3), 387–397. https://doi.org/10.19826/j.cnki.1009-3850.2021.05001 (in Chinese).
Zou, C., Qiu, Z., 2021. Preface: new advances in unconventional petroleum sedimentology in China. Acta Sedimentol. Sin. 39 (1), 1–8. https://doi.org/10.14027/j.issn.1000-0550.2021.001 (in Chinese).
Zou, C., Yang, Z., He, D., et al., 2018. Theory, technology and prospects of conventional and unconventional natural gas. Petrol. Explor. Dev. 45 (4), 604–618. https://doi.org/10.1016/S1876-3804(18)30066-1.
Zou, C., Yang, Z., Zhang, G., et al., 2014. Conventional and unconventional petroleum "orderly accumulation": concept and practical significance. Petrol. Explor. Dev. 41 (1), 14–30. https://doi.org/10.1016/S1876-3804(14)60002-1.