Original Paper

Multifractal estimation of NMR T_2 cut-off value in low-permeability rocks considering spectrum kurtosis: SMOTE-based oversampling integrated with machine learning

Xiao-Jun Chen [a, b], Rui-Xue Zhang [c], Xiao-Bo Zhao [a], Jun-Wei Yang [a], Zhang-Jian Lan [d], Cheng-Fei Luo [e], Jian-Chao Cai [a, f, *]

[a] MOE Key Laboratory of Tectonics and Petroleum Resources, Wuhan, 430074, Hubei, China
[b] Department of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M13 9PL, UK
[c] Jianghan Oilfied, SINOPEC, Qianjiang, 430063, Hubei, China
[d] Hainan Branch of CNOOC Ltd, Haikou, 570311, Hainan, China
[e] CNOOC EnerTech-Drilling & Production Co. CNOOC Central Laboratory (Zhanjiang), Zhanjiang, 524057, Guangdong, China
[f] National Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing, 102249, China

ARTICLE INFO

ABSTRACT

The transverse relaxation time (T_2) cut-off value plays a crucial role in nuclear magnetic resonance for identifying movable and immovable boundaries, evaluating permeability, and determining fluid saturation in petrophysical characterization of petroleum reservoirs. This study focuses on the systematic analysis of T_2 spectra and T_2 cut-off values in low-permeability reservoir rocks. Analysis of 36 low-permeability cores revealed a wide distribution of T_2 cut-off values, ranging from 7 to 50 ms. Additionally, the T_2 spectra exhibited multimodal characteristics, predominantly displaying unimodal and bimodal morphologies, with a few trimodal morphologies, which are inherently influenced by different pore types. Fractal characteristics of pore structure in fully water-saturated cores were captured through the T_2 spectra, which were calculated using generalized fractal and multifractal theories. To augment the limited dataset of 36 cores, the synthetic minority oversampling technique was employed. Models for evaluating the T_2 cut-off value were separately developed based on the classified T_2 spectra, considering the number of peaks, and utilizing generalized fractal dimensions at the weight <0 and the singular intensity range. The underlying mechanism is that the singular intensity and generalized fractal dimensions at the weight <0 can detect the T_2 spectral shift. However, the T_2 spectral shift has negligible effects on multifractal spectrum function difference and generalized fractal dimensions at the weight >0. The primary objective of this work is to gain insights into the relationship between the kurtosis of the T_2 spectrum and pore types, as well as to predict the T_2 cut-off value of low-permeability rocks using machine learning and data augmentation techniques.

1. Introduction

Deeply understanding the global pore structure of porous rocks is the crucial step to addressing microcosmic challenges in petroleum science (Dullien, 2012; Gao and Li, 2016; Blunt, 2017). Among various experimental methods, nuclear magnetic resonance (NMR) spectra of transverse relaxation time T_2 have emerged as a unique downhole-calibration technique, offering numerous opportunities to probe pore structure. This includes studying pore deformation and evolution, pore types, fluid distribution and migration, capillary pressure curve reconstruction, porosity inversion, as well as permeability and irreducible water saturation estimations in recent decades (Yan et al., 2019; Guo et al., 2020; Liu et al., 2020b; Tang et al., 2022; Zhang et al., 2022). NMR T_2 spectra provide rich quantitative information on pore structure, such as pore size distribution, porosity, and fluid saturation. This information can be obtained by combining T_2 spectra in fully water-saturated and centrifuged states (Li et al., 2020a; Wang et al., 2020; Yan et al.,

* Corresponding author.
  E-mail address: caijc@cup.edu.cn (J.-C. Cai).

2020; Liu et al., 2021b). Compared to other experimental methods like fluid-intrusion testing (Wang et al., 2019; Jiao et al., 2020; Liu et al., 2020a) and imaging (Blunt et al., 2013; Bultreys et al., 2016; Liu et al., 2016; Mehmani et al., 2020), NMR provides unique advantages in exploring various aspects of pore structure in porous rocks.

In NMR T_2 spectra, the "T_2 cut-off value" is a crucial parameter used to classify pore systems into different pore types (Yao et al., 2010). It separates pores into two categories: "free fluids" for pores with T_2 values higher than the cut-off value, where fluids can flow freely, and "bound fluids" for pores with T_2 values below the cut-off value, where fluids are trapped (Morriss et al., 1997). The T_2 cut-off value serves as a boundary for estimating permeability and irreducible water saturation and is determined by comparing the cumulative curves of NMR T_2 spectra in fully water-saturated and centrifuged states. However, experimental determination of the T_2 cut-off value is often challenging and time-consuming. In practical applications, an empirical T_2 cut-off value is commonly assigned to all samples of a specific lithology. For example, sandstone may be assigned a T_2 cut-off value of 33 ms, carbonate 90 ms, and limestone 100 ms (Coates et al., 1999; Hidajat et al., 2004). Nevertheless, increasing experimental evidence suggests that using a fixed T_2 cut-off value is inadequate for estimating reservoir properties, even within different porous rocks of the same lithology (Wang et al., 2004; Ge et al., 2011). For instance, the T_2 cut-off value ranges for coal and shale are approximately 2–40 ms and 0.1–3 ms, respectively (Yao et al., 2010; Testamanti and Rezaee, 2017; Li et al., 2020b).

Various estimation models have been proposed to determine the T_2 cut-off value, considering different factors related to the T_2 spectrum. Some models employ simple statistical methods, such as the weighted sum of T_2 amplitudes (Oraby et al., 1997), linear arithmetic average of T_2 cut-off values (Kleinberg and Boyd, 1997), T_2 geometric mean of the last peak (Mai and Kantzas, 2002), and magnetic susceptibility (Nicot et al., 2016). However, the T_2 cut-off value is influenced by several factors, including lithology (Ge et al., 2011), pore types (Westphal et al., 2005), wettability, and temperature (Godefroy et al., 2002).

The complete NMR T_2 spectrum represents a complex and highly disordered system with self-similar characteristics (Wang et al., 2018a). Therefore, it is more appropriate to evaluate the T_2 cut-off value based on the distribution characteristics of the entire T_2 spectrum, rather than relying solely on specific morphological parameters. Fractal theory, including monofractal and multifractal analysis, has demonstrated the feasibility of NMR T_2 spectrum analysis (Wang et al., 2018b; Guo et al., 2019; Wu et al., 2020; Liu et al., 2021a). While monofractal analysis has inherent limitations and cannot fully capture the complex distribution characteristics across different pore scales (Chen et al., 2017; Fu et al., 2019), multifractal and generalized fractals provide a hierarchical approach to determine the multi-scale and heterogeneous pore characteristics of the target of interest (Zhao et al., 2017, 2019).

Multifractal parameters play a significant role in determining the pore size distribution, which in turn is utilized to estimate the T_2 cut-off values in coal and shale reservoirs. Several studies have been conducted in this area. Ge et al. (2015) proposed a fractal T_2 cut-off model using multiple linear regression of generalized fractal dimensions for sandstones with complex pore structures. Xiao et al. (2018) developed new normal distribution functions to simulate fully water-saturated NMR T_2 spectra and estimated various T_2 cut-off values. Zheng et al. (2019) developed a new model for the T_2 cut-off value, which was tested on 15 samples of bituminous coals and anthracite coals using multiple linear fitting. Hu et al. (2020) introduced the flow zonation index into the T_2 cut-off fractal model using multiple linear regression for 10 tight sandstones in

the eastern Ordos Basin. Zhao et al. (2021) tested five predictive models developed by multiple linear fitting of multifractal parameters and T_2 cut-off values on four types of coal samples. Sun et al. (2021) found strong positive correlations between generalized fractal dimensions, singularity strength, and T_2 cut-off values in three ranks of coals. Liu et al. (2021b) observed a slight negative correlation between NMR T_2 cut-off values and fractal dimensions calculated from centrifugal T_2 spectra in oil shale reservoirs. Although the aforementioned studies have achieved significant success in estimating the T_2 cut-off value, the relationship between T_2 spectrum kurtosis and pore types remains unclear. Additionally, previous studies have mainly focused on shale, coal rock, and tight sandstone (Table 1), with limited attention given to the NMR morphological features and estimation of T_2 cut-off values in low-permeability sandstones. Due to the high cost of NMR experiments, the number of available samples is often limited (Table 1). Therefore, the utilization of the latest data augmentation methods is desirable for estimating NMR T_2 cut-off values.

In this study, an experimental investigation was conducted to analyze the kurtosis characteristics of T_2 spectra from 36 low-permeability cores and examined their relationships with different pore types. The multifractal characteristics of the T_2 spectra were also analyzed. The correlations between the T_2 cut-off value and various petrophysical parameters, geometric morphological parameters, and multifractal parameters were established. To enhance the dataset, the synthetic minority oversampling technique (SMOTE) was employed to augment the multifractal dataset with measured T_2 cut-off values. Subsequently, we compared and analyzed the performance of three commonly used machine learning algorithms for predicting the T_2 cut-off values. Furthermore, we employed the SHapley Additive exPlanation (SHAP) method to identify the main factors influencing the modeling of T_2 cut-off values. Additionally, we discussed the impact of the T_2 spectrum's location relative to the X-axis on multifractal spectra. The main objective of this study is to gain insights into the pore types and improve the prediction of T_2 cut-off values for low-permeability rocks by considering T_2 spectrum kurtosis and employing machine learning techniques with data augmentation.

2. Experimental and multifractal analysis

2.1. Theory of NMR, fractal, and multifractal

NMR T_2 time of pore fluids in porous media comprises three components: surface relaxation time T_{2S}, bulk relaxation time T_{2B}, and diffusion relaxation time T_{2D}, which satisfy the following relationship (Kleinberg and Boyd, 1997; Coates et al., 1999):

$$\frac{1}{T_2} = \frac{1}{T_{2B}} + \frac{1}{T_{2S}} + \frac{1}{T_{2D}} = \frac{1}{T_{2B}} + \rho_2 \left(\frac{S}{V}\right) + \frac{D(\gamma G T_E)^2}{12} \quad (1)$$

where ρ_2 represents the transverse surface relaxation rate, m/s; S/V is specific surface area, the ratio of pore surface area and pore volume; D represents diffusion coefficient, m²/s; γ represents the gyromagnetic ratio, MHz/T; G represents the field intensity gradient, T/µm; T_E is the echo interval, ms. For cores fully saturated with water, the free relaxation time of hydrogen nuclei in water is so long that $\frac{1}{T_{2B}} \approx 0$ in the uniform external magnetic field $G = 0$. Thus, T_2 can be simplified as (Daigle and Johnson, 2016):

$$\frac{1}{T_2} = \frac{1}{T_{2S}} = \rho_2 \frac{S}{V} \quad (2)$$

In the capillary bundle model, the specific surface area of the

Table 1
Summary of predicting NMR T_2 cut-off values for different reservoir rocks using fractal and multifractal theory.

Authors	Rock type	Number of cores	Location[a]	Range of T_2 cut-off value[b]	T_2 spectrum kurtosis	Methods for T_2 cut-off value models
Westphal et al. (2005)	Carbonate Rocks	36 (12-PP; 30-IP; 8-MP; 5-V; 6-FP)	USA; Canada; Abu Dhabi; Norway.	PP:60.4; IP:64.7; MP:68.7; V:442.9; FP:24.1.	N/A	N/A
Ge et al., (2015)	Siltstone	15	Guantao Formation, China	41.179 (7.243−69.327).	Bimodal	Multiple linear regression
Testamanti and Rezaee (2017)	Shales	4	Carynginia Formation, PB, Western Australia	0.24(n = 3); 0.26(n = 1).	Trimodal all	N/A
Wang et al., (2020)	SR, CR	8 (4-SR; 4-CR)	Shahejie Formation, BBB	SR: 5.61(1.55−9.64); CR: 3.985(5.57−1.86).	half of Unimodal, half of Bimodal	Multiple linear regression
Xiao et al., (2018)	Tight sandstones	36	Triassic tight sandstones, SOB	17.896 (6.31−50.1).	Unimodal; Bimodal; Trimodal.	Integrate with the simulated curve
Zheng et al. (2019)	Coals	15 (5-SJ; 5-NQ; 5-SQ)	SJB; NQB; SQB	SJ:4.396 (0.62−11.11); NQ:4.264 (2.02 −8.24); SQ:7.274 (1.82−12.66).	Trimodal mainly, Bimodal partly.	Multiple linear regression
Wang et al., (2020)	SR, CR	8 (4-SR; 4-CR)	Shahejie Formation, BBB	SR: 5.61(1.55−9.64); CR: 3.985(5.57−1.86).	Half of Unimodal, half of Bimodal.	Multiple linear regression
Hu et al. (2020)	Sandstone	10	Benxi, Taiyuan and Shanxi Formations, EOB	21.057 (9.72−35.16).	Unimodal mainly, Bimodal partly.	Multiple linear regression
Zhao et al. (2021)	Coals	32 (8-CZ; 8-GD; 8-GHS; 8-PDS)	CZ, GD, GHS PDS coal mines, China	CZ: 0.558 (0.37−0.74); GD: 0.51 (0.3−0.6); GHS: 0.524 (0.39−0.63); PDS: 0.161 (0.13 −0.21).	Trimodal mainly, Bimodal partly.	Multiple linear regression
Sun et al. (2021)	Coals	10- anthracite; 10- bitumite 10- lignite	Henan, Shanxi, and Inner Mongolia, China	0.38−0.62, 0.22−0.32, 0.11−0.23.	Depends on coal rank	BP neural network
This study	Low-permeability sandstones	36 cores with 525 augmented samples	Huangliu, Liushagang Formations, YSB, QDNB, BBWB, PRMB	24.08 (7−50).	Unimodal and Bimodal mainly, Trimodal rarely.	Machine Learning integrated with SMOTE

[a] Note: SJB–Southern Junggar Basin, NQB–Northern Qinshui Basin, SQB–Southern Qinshui Basin, EOB–eastern Ordos Basin, BBB–Bohai Bay Basin, PB–Perth Basin, CZ–Chengzhuang, GD–Guandi, GHS–Guhanshan, PDS–Pingdingshan, SR–Siliciclastic Rocks, CR–Carbonate Rocks, PP–Primary pores, IP–Intercrystalline pore, MP–Moldic Pore, V–Vuggy, FP–Infilled pore.
[b] Average (min–max).

porous medium is a function of mean pore radius r (Sigal, 2015):

$$\frac{1}{T_2} = \rho_2 \frac{F_s}{r} \tag{3}$$

where F_s represents the geometric factor, dimensionless. For cylindrical pores, F_s is equal to 2. Thus, T_2 spectrum is a function of pore radius, reflecting the pore size distribution (Fig. 1) (Yan et al., 2020). In the NMR experimental results, the T_2 spectrum obtained after porosity correction is a frequency histogram of T_2 where the y-axis is the porosity increment, which can be expressed as:

$$T_2 = f(r) = \{P_{2i}\} \tag{4}$$

T_2 is a discrete variable where P_{2i} represents the porosity increment at the corresponding time T_{2i}. Thus, the T_2 spectrum is a one-dimensional (1D) mass field of porosity increment. For a 1D mass field, the box-counting method was used to calculate the multifractal dimensions (Stanley and Meakin, 1988; Lopes and Betrouni, 2009). After a viewing scale ε was specified, the $\{P_{2i}\}$ was divided into N equal-size interval with a length size of ε. In i-th interval, the cumulative porosity increment $M_i(\varepsilon)$ was counted. The mass probability function $P_i(\varepsilon)$ of the i-th box can be calculated as:

$$P_i(\varepsilon) = \frac{M_i(\varepsilon)}{\sum_1^{N(\varepsilon)} M_i(\varepsilon)} \tag{5}$$

For an inhomogeneous 1D mass field satisfying a multifractal, $P_i(\varepsilon)$ satisfies a fractal up-scaling relationship with the box scale ε, which can be written as:

$$P_i(\varepsilon) \propto \varepsilon^{-a_i} \tag{6}$$

where a_i represents the Lipschitz-Hölder singularity exponent,

characterizing the singularity strength of the data distribution. The total number of sub-intervals with the same $P_i(\varepsilon)$ value was counted as $N_a(\varepsilon)$. Similarly, $N_a(\varepsilon)$ also satisfies the fractal scale law:

$$N_\alpha(\varepsilon) \propto \varepsilon^{-f(\alpha)} \tag{7}$$

where $f(\alpha)$ vs. α is the multifractal spectrum, describing how dense the distribution is. With the moment method, a partition function, $\chi_q(\varepsilon)$, is used:

$$\chi_q(\varepsilon) = \sum_{i=1}^{N} P_i^q(\varepsilon) \propto \varepsilon^{\tau(q)} \tag{8}$$

where $\tau(q)$ is the quality index, commonly ranging from −10 to 10 to reflect the contributions of $P_i(\varepsilon)$, which will be affected by sub-intervals with large $P_i(\varepsilon)$ at $q > 0$, and vice versa. $\tau(q)$ can be calculated by taking the log of Eq. (8) on both sides:

$$\tau(q) = -\lim_{\varepsilon \to 0} \frac{\log \chi(q,\varepsilon)}{\log \varepsilon} = -\lim_{\varepsilon \to 0} \frac{\log \sum_{i=1}^{N(\varepsilon)} P_i^q \varepsilon}{\log \varepsilon} \tag{9}$$

$f(\alpha)$ vs. α can be obtained by Legendre transform as (Chhabra and Jensen, 1989; Lopes and Betrouni, 2009):

$$\alpha = \frac{d(\tau(q))}{dq} = \frac{d}{dq}\left(\lim_{\varepsilon \to 0} \frac{\log X_q(\varepsilon)}{\log \varepsilon}\right) \tag{10}$$
$$f(\alpha) = \alpha q - \tau(q)$$

For the mathematically perfect fractal subjects (in 1D, 2D, and 3D), the multifractal spectrum of $f(\alpha)$ vs. α is a symmetric smooth convex curve. However, for the multifractal subjects, the multifractal spectrum shows a symmetrical characteristic (Cheng, 1999).
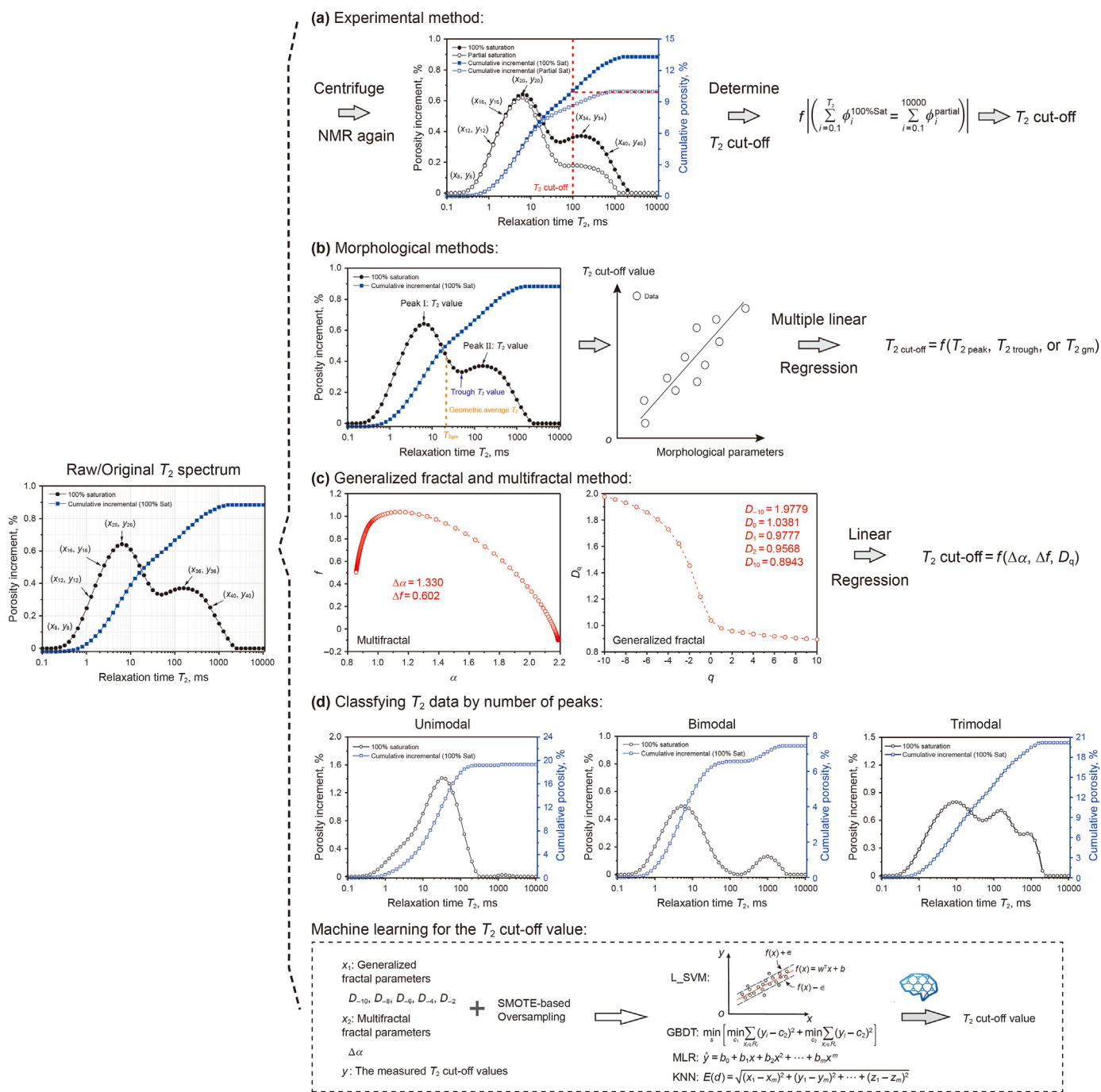
Fig. 1. Schematic diagram illustrating the conventional and developed methods for determining the T_2 cut-off value using the saturated T_2 spectrum, including (a) the conventional centrifugal experimental method with NMR test again; (b) the morphological method combined with the mathematical fitting; (c) fractal and multifractal methods with un-classified T_2 data; (d) the developed method combing multifractal analysis with SMOTE-based machine learning regressions on the classified T_2 data by curve kurtosis.

Generally, the singular intensity range $\Delta\alpha$ ($\alpha_{max} - \alpha_{min}$) and the corresponding multifractal spectrum function difference $\Delta f[= f(\alpha_{min}) - f(\alpha_{max})]$ are two key parameters to investigate the multifractal features, and to characterize the heterogeneity of pore information reflected by NMR T_2 spectrum.

Through the $\tau(q)$, the generalized fractal spectrum D_q can be defined as:

$$D_q = \frac{1}{q-1} \lim_{\varepsilon \to 0} \frac{\log \sum_{i=1}^{N} P_i^q(\varepsilon)}{\log \varepsilon} = \frac{\tau(q)}{q-1} (q \neq 1) \tag{11}$$

where $D_0 = - \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log \varepsilon}$ ($q = 0$), known as the capacity dimen-

sion, single fractal dimension; $D_1 = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{N(\varepsilon)} P_i(\varepsilon) \log P_i(\varepsilon)}{\log \varepsilon}$ ($q = 1$),

called as information dimension; $D_2 = \lim_{\varepsilon \to 0} \frac{\sum_{i=1}^{N(\varepsilon)} P_i^2(\varepsilon)}{\log \varepsilon}$ ($q = 2$), called

correlation dimension. Similarly, for the mathematically perfect fractal subjects (in 1D, 2D and 3D), the generalized fractal spectrum of D_q vs. q is a horizontal line. However, for the multifractal subjects, the generalized fractal spectrum commonly shows an S-shaped characteristic (Chen et al., 2017; Zhao et al., 2019). Next, pore system, morphological, and multifractal analysis of the T_2 spectrum were performed (Fig. 1).

2.2. Sampling and experiments

A total of 36 low-permeability sandstone cores were collected from various basins, formations, and wells. The specific sampling locations are depicted in Fig. 2. These core samples were obtained from 19 wells with available core data, and they were extracted from depths ranging from 1200 m to 4000 m. Each core sample has a diameter and length of 3.5 cm. At the Zhanjiang Experimental Center of China National Offshore Oil Corporation (CNOOC), all cores underwent analysis for petrophysical properties, including porosity and corrected permeability. Additionally, for this analysis, four conventional reservoir sandstones were selected, with low-permeability rocks constituting 89% of the samples.

To prevent any water-rock reactions, the cores were fully saturated with brine using a 1% KCl concentration. The saturation process was conducted at a pressure of 30 MPa for 48 h. The NMR experiments were carried out at a temperature of 28 °C and a relative humidity ranging from 50% to 70% RH. The experimental parameters for NMR included a waiting time of 3000 ms, an echo interval of 0.2 ms, and a total of 4096 echoes were collected. The T_2 spectrum inversion was performed using the Butler-Reeds-Dawson algorithm (Butler et al., 1981) based on the measured Carr-Purcell-

Meiboom-Gill echo trains (Ge et al., 2017). During the centrifugation experiments, the CSC-12 high-speed refrigerated centrifuge from Shanghai Lu Xiangyi centrifuge instrument Co., Ltd. Was utilized. The cores were subjected to centrifugation at a speed of 8000 rpm for 3 h, generating a centrifugal pressure of 2.425 MPa. These centrifugation conditions adhere to the guidelines specified in the National Standard SY/T 6490−2014.

3. Results and discussion

3.1. Characteristics of samples and NMR T_2 spectra

The petrophysical properties of the low-permeability rocks exhibit significant variability. The porosity values determined using helium-based measurements range from 7.40% to 26.8%, with an average value of 16.79%. The NMR-based porosity calculations yield a range of 7.46%−26.60%, with an average value of 16.77%. The scale factor F for the porosity component falls within the range of 22−24. The NMR-based porosity is in good agreement with the porosity measured from the cores (Fig. 3), indicating that all relaxation signals can be captured within the current echo interval, and the samples do not contain fast relaxation components (Ge et al., 2021). The permeability of the low-permeability rocks ranges from 0.02 mD to 30.10 mD, with an average value of 7.38 mD. In contrast, the permeability of the four conventional sandstones is 93.6 mD, 168.00 mD, 496.00 mD, and 702.00 mD, respectively. The T_2 cut-off values for the low-permeability rocks range from 7 ms to 50 ms, with an average value of 23.25 ms. For the four conventional cores, the T_2 cut-off values are 20 ms, 30 ms, 30 ms, and 40 ms. Importantly, the T_2 spectra of the low-permeability sandstones exhibit a
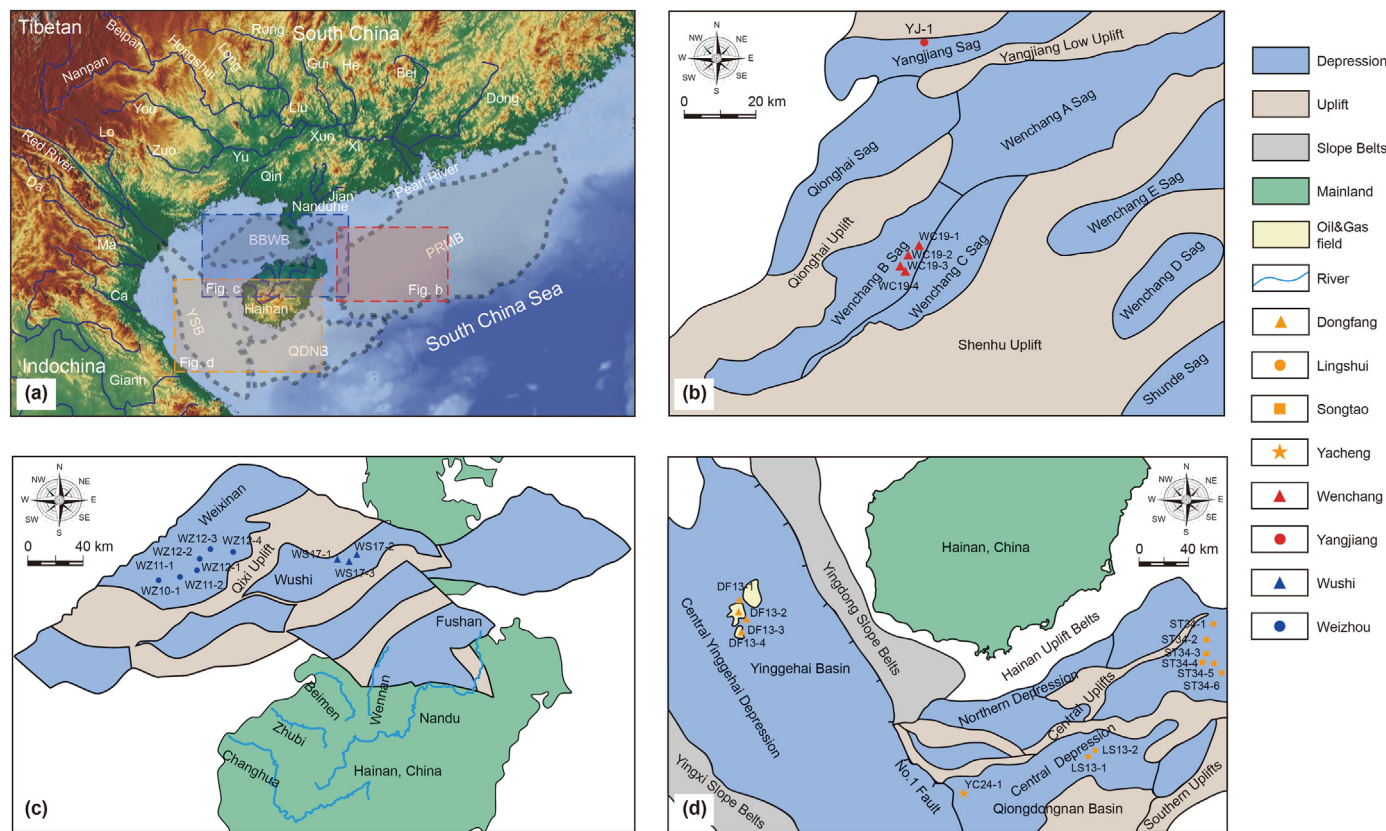

Fig. 2. Sampling location map of 36 low permeable cores, (a) Geographical location of the four basins sampled: Yinggehai Basin (YSB), Qiongdongnan Basin (QDNB), Beibuwan Basin (BBWB), Pearl River Mouth Basin (PRMB); (b) Structural map of Wenchang and Yangjiang Sags of PRMB; (c) Structural map of Weizhou and Wushi Sags of BBWB; (d) Structural map showing the Dongfang Sag of YSB, the Songtao, Yacheng and Lingshui Sags of QDNB.
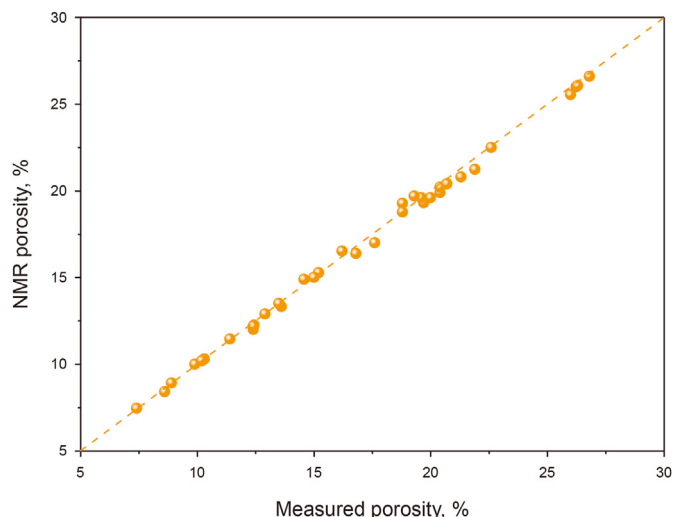
Fig. 3. Cross plot of NMR-based porosity and helium-based porosity where the dotted line is a 1:1 standard line.

multimodal morphological feature, including unimodal, bimodal, and trimodal modes, which are rarely observed in other rock types such as shale and coal (Fig. 4). Specifically, in this dataset, there are 13 samples with unimodal modes, 19 samples with bimodal modes, and four samples with trimodal modes (Fig. 4).

3.2. Pore system vs. T_2 spectrum kurtosis

The multimodal morphological feature observed in the T_2 spectra of low-permeability sandstones is closely related to the complex pore system present in these rocks. The 2D pore morphological characteristics were examined using thin sections, while the 3D pore size distribution characteristics were quantitatively analyzed using mercury intrusion capillary pressure (MICP). The complex diagenesis and deposition processes of low-permeability sandstones contribute to the presence of various pore types (Chen et al., 2018; Zhao et al., 2022a). The hybrid pore system in low-permeability sandstones comprises different pore types, as illustrated in Fig. 5. These pore types can be classified based on their origin and include primary, mold, dissolution, fracture, and clay-associated pores (Westphal et al., 2005). Each pore type exhibits distinct morphological features and size distributions. Furthermore, the T_2 spectrum can be converted into a pore size distribution, allowing for the identification of different pore types (Xiao et al., 2018). Based on the peak locations in the T_2 spectrum, pores can be categorized into four types: non-pore, micro-pore, meso-pore, and macro-pore regions. These regions correspond to grains and matrix (non-pore), clay-bound water (micro-pore), capillary-trapped water (meso-pore), and mobile water (macro-pore) in the NMR signal, respectively (Fig. 5) (Zhong et al., 2020).

The morphology of the T_2 spectrum (Fig. 6a) is determined by the combination of different pore types present in the rock. In the case of the trimodal type, the T_2 spectrum is influenced by the presence of extra-large dissolution pores along with primary pores and clay-associated pores, resulting in a trimodal distribution (Fig. 6c). The dissolution of feldspar grains leads to the precipitation of kaolinite clay minerals. In the T_2 spectrum, the T_2 cut-off value (=20 ms) is located between the first and second peaks, indicating that the determination of the T_2 cut-off value cannot solely rely on the positions of the three peaks. MICP analysis provides insights into the pore types within the rock, revealing an undulating characteristic. Specifically, at a mercury saturation of 36%, mercury fluid

enters the rock from the extra-large dissolution pores and moves into the primary pores, with a pore-throat radius boundary of 3.66 μm (Fig. 6b). As the mercury saturation increases to 69.5%, the mercury fluid further penetrates into the clay-associated pores, which have a smaller pore-throat radius boundary of 0.061 μm. This demonstrates the presence of different pore types and their connectivity within the rock, contributing to the complex pore structure observed in the T_2 spectrum and the determination of the T_2 cut-off value.

In the case of the bimodal mode, the combination of two different pore types leads to a bimodal T_2 spectrum. Two specific combinations of pore types result in the bimodal mode as observed in Fig. 6d, 6g. In one case, primary pores are combined with mold pores, along with clay-associated pores characterized by a pore-throat boundary of 1.63 μm (Fig. 6e, 6f). In the second case, moderate dissolution pores are combined with micro-fractures and clay-associated pores, with a pore-throat boundary of 3.64 μm (Fig. 6h, 6i). Interestingly, in both cases, none of the positions of the two T_2 cut-off values correspond to specific features such as two peaks or middle troughs, indicating that the determination of the T_2 cut-off value cannot be solely based on a simple morphological analysis of the T_2 spectrum. For the unimodal mode, which is commonly associated with clay-associated pores (Fig. 6l), this is confirmed by the MICP analysis where no fluctuations and pore boundaries are detected (Fig. 6k). The location of the T_2 cut-off value is found in the vicinity of the peak (Fig. 6j). However, it is worth noting that three out of the four conventional cores also exhibit a unimodal mode, where the dominant pore system is primarily composed of uniform intergranular primary pores. Therefore, the dominant pore system varies depending on the degree of permeability even in the unimodal mode.

3.3. NMR multifractal characteristics

The fractal analysis of the T_2 spectrum began by plotting T_q vs. q, which helps in detecting the fractal behavior of the 1D NMR T_2 distribution. For a 1D dataset with a single fractal distribution, such as the Cantor Set, the plot of T_q vs. q appears as a straight line. However, in the case of the tested T_2 data, the plots of T_q vs. q exhibit a curved pattern, indicating the presence of multifractal characteristics (Fig. 7). Subsequently, the generalized fractal and multifractal spectra were calculated for all the cores. The calculated results were categorized into three modes based on the number of peaks observed: unimodal, bimodal, and trimodal modes. In the generalized fractal spectra, the fractal dimension at $q = -10$ is referred to as $D_{\max}$, while the fractal dimension at $q = 10$ is referred to as $D_{\min}$. All the calculated fractal spectra were plotted and presented in Fig. 8.

Generalized fractal spectra from all cores exhibit an "S" shape (Fig. 8a, b, c). In these spectra, D_q shows a sharp decrease when q is less than 0, while it decreases more gradually when q is greater than 0. This pattern is consistent with observations in coal and shale studies (Testamanti and Rezaee, 2017; Zheng et al., 2019). The calculated average (min-max) distributions of D_q for different numbers of peaks are as follows: for the unimodal, $D_{\min}$ 2.162 (1.530–2.479), D_{-8} 2.115(1.496–2.428), D_0 0.989(0.923–1.066), D_1 0.905(0.865–0.984), D_2 0.876(0.800–0.968), $D_{\max}$ 0.805 (0.649–0.946); for the bimodal, $D_{\min}$ 2.009 (1.599–2.339), D_{-8} 1.965(1.564–2.287), D_0 1.001(0.884–1.079), D_1 0.921(0.815–1.014), D_2 0.892(0.780–0.997), $D_{\max}$ 0.837(0.697–0.959); for the trimodal, $D_{\min}$ 2.067(1.958–2.191), D_{-8} 2.021(1.916–2.141), D_0 1.052(1.038–1.066), D_1 0.969(0.940–0.995), D_2 0.943(0.906–0.986), $D_{\max}$ 0.880(0.817–0.966).

From the above D_q distributions, compared with the unimodal mode, the bimodal mode has a larger mean value and a wider range
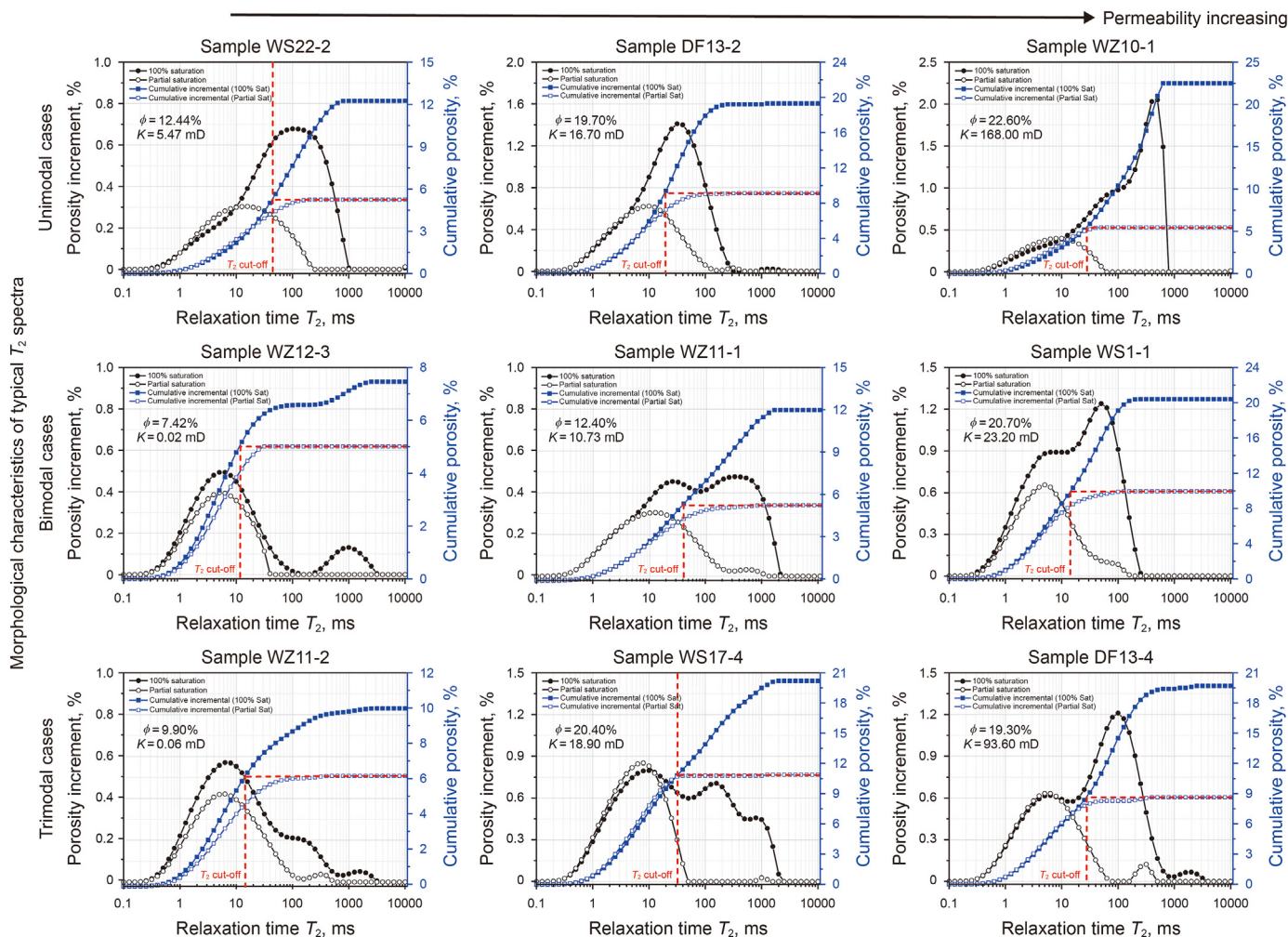
Fig. 4. Using saturated and centrifugal T_2 spectra to obtain T_2 cut-off value, typical experimental results of three types of saturated NMR T_2 spectra, including (a) unimodal mode of samples #WS22-2, #DF13-1, and #WZ10-1; (b) bimodal mode samples #WZ12-3, #WZ11-1, and #WS1-1; (c) trimodel mode samples #WZ11-2, #WS17-4, and #DF13-4. The permeability and porosity values of each type of NMR data vary greatly, indicating that physical properties are not the factor affecting the number of peaks of saturated T_2 spectrum.

of D_q (Fig. 8a, 8b). The maximum value of $D_{\min}$ in bimodal modes reaches as high as 2.479 but that of unimodal mode is only 2.339. This can be attributed to the greater variability and volatility in the morphology of the T_2 spectrum in the bimodal mode compared to the unimodal mode. However, the trimodal mode in this analysis does not follow this trend (Fig. 8c). This may be due to the relatively weaker fluctuation in the T_2 spectrum of the trimodal mode compared to the bimodal mode. Additionally, the limited number of samples ($n = 4$) in this study may not fully represent the range of the trimodal mode. Furthermore, the value of D_q ($q < 0$) is more sensitive to the shape of the T_2 spectrum and exhibits a wider range (1.5−2.5). In contrast, the variation range of D_q ($q > 0$) is very narrow (0.65−0.95). Therefore, D_q ($q < 0$) rather than D_q ($q > 0$) is expected to have more superior performance when examining the differences in T_2 cut-off values.

Multifractal spectra of $\alpha - f(\alpha)$ from all cores show a left-hook asymmetric bell-shaped characteristic, indicating a predominance of a high probability subset (Fig. 8d-f). The calculated 'average (min-max)' distributions of $\Delta\alpha$ and Δf follows: for the unimodal, $\Delta\alpha$ 1.597(0.890−2.063), Δf 0.445(0.060−0.802); for the bimodal, $\Delta\alpha$ 1.395(0.971−1.821), Δf 0.543(0.103−0.831); for the trimodal, $\Delta\alpha$

1.432(1.350−1.643); Δf 0.560(0.379−0.832). Compared with the unimodal mode, the bimodal mode has a relatively smaller mean value and a narrower range of $\Delta\alpha$ (Fig. 8e). The reason is attributed to the T_2 distribution in the unimodal mode having a more significant polarization. However, the bimodal mode has a slightly larger mean value of Δf. All Δf values are positive, indicating that the high probability subset is dominant, which is consistent with the generalized fractal analysis. The trimodal mode was not analyzed for comparison owing to the number of samples (Fig. 8f).

The T_2 cut-off value cannot be estimated by morphological parameters for low-permeability sandstones. The plots displaying the relationship between morphological parameters and T_2 cut-off value are presented in Fig. 9, where no significant correlations were observed. Notably, the geometric mean T_2 exhibits a higher correlation coefficient (=0.6903) compared to the arithmetic mean (Fig. 9a). This can be attributed to the fact that the T_2 cut-off value relies on the entire T_2 spectrum rather than a single peak or average value of the T_2 data. The morphological parameters fail to capture the full-size distribution characteristics of the T_2 spectrum. Consequently, obtaining the T_2 cut-off value cannot be obtained from the morphological parameters after curve normalization.
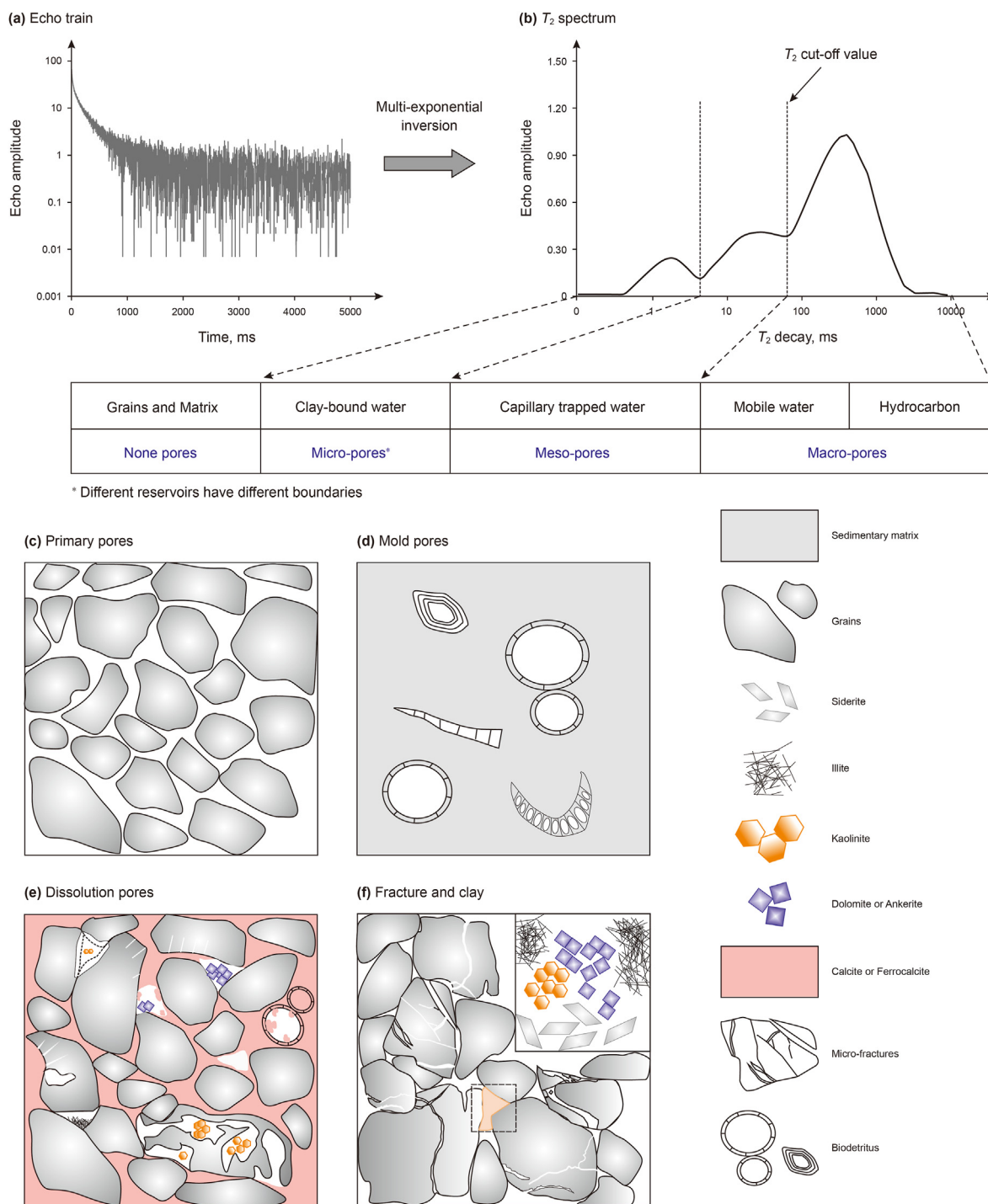
Fig. 5. NMR Data processing workflow and pore type classification according to pore size, the pore type in low permeable media classified according to the inherent genesis of pores, (a) Measured echo train data; (b) inverted T_2 spectrum where pore size type is classified according to the trough of T_2 spectrum in conventional method; (c)(d)(e)(f) the four common pore types classified according to their origins: primary pores, mold pores, dissolution pores, fracture and clay-associated pores.

3.4. SMOTE oversampling

Due to the substantial cost and resource-intensive nature of NMR experiments, the number of rock cores available for NMR testing is typically severely limited (e.g., Table 1). Consequently, this often leads to constraints on the quality of core-based NMR T_2 spectra. Specifically, the sample size is undersampled and insufficient for statistical analysis. In logistic regression analysis, a widely advocated minimum standard for sample size consideration is ten events per variable (EPV) (Peduzzi et al., 1996). When the total number of samples is less than ten times the number of feature vectors, overfitting becomes more likely during regression and classification using multiple or machine learning models. Supervised learning to underlying physical mechanisms from imbalanced datasets poses a challenging task (Camacho et al., 2022).

To mitigate the issue of overfitting with small NMR
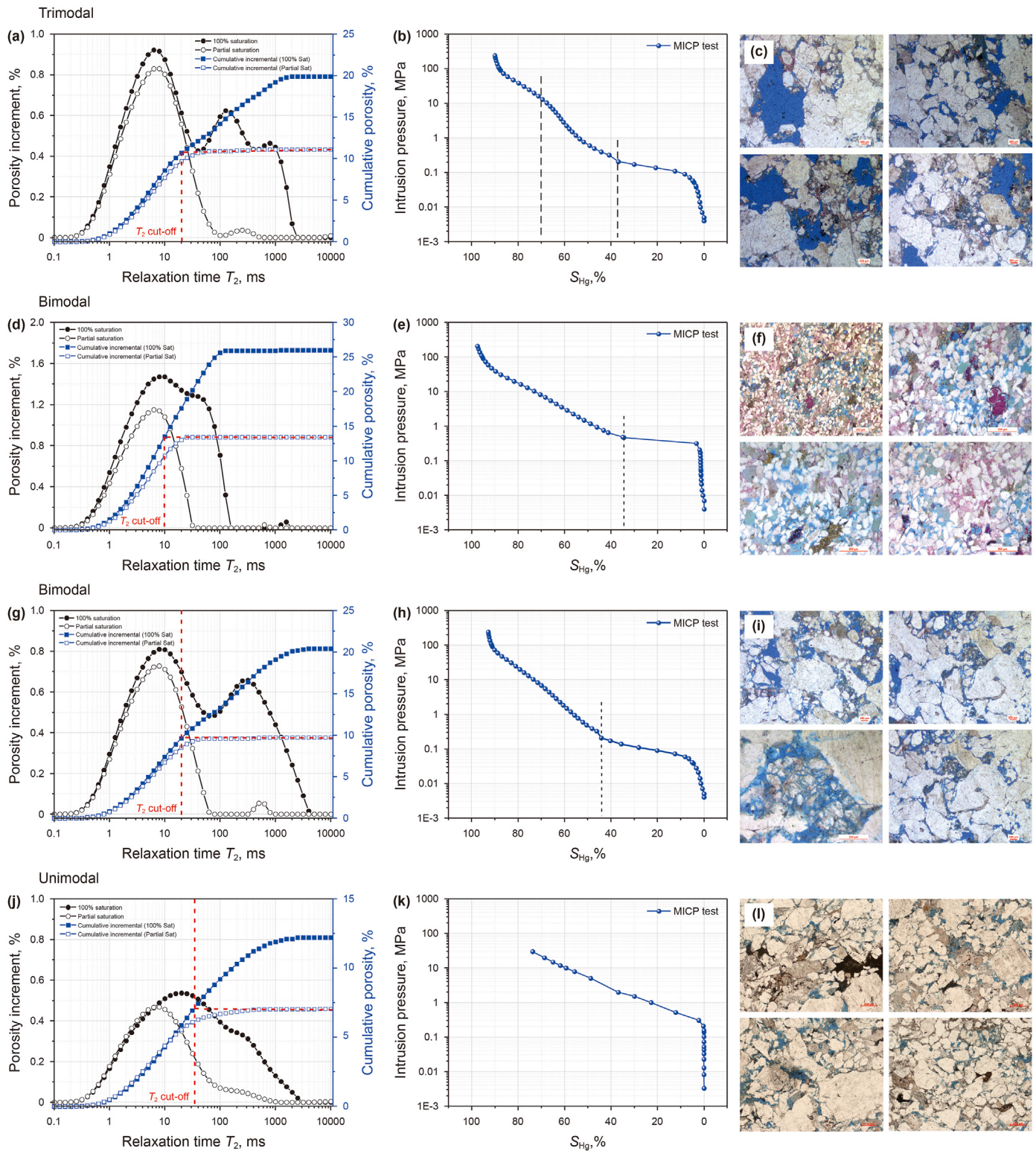
Fig. 6. The response relationship between the number of peaks in the saturated T_2 spectrum and the pore system or pore types where system are reflected by MICP in 3D space and by thin section in 2D plane, (a) (d) (g) (j) the tested NMR T_2 spectra in the saturated and centrifugal states from samples #WS17-3, #LS13-2, #WS11-1, #WZ12-1; (b) (e) (h) (k) the MICP tested results of four samples from (a) (d) (g) (j); (c) (f) (i) (l) 2D thin section for showing different pore types if four samples from (a) (d) (g) (j).

experimental datasets and develop a more robust prediction model for the T_2 cut-off value, data augmentation is introduced. This technique reliably expands the training datasets, enabling the prediction model to focus on deeper and more essential features,

thereby enhancing the model's robustness and generalization. Data augmentation, similar to actual oversampling, does not directly replicate observations, but instead introduces small perturbations to replicated data points (Douzas et al., 2018). For 2D or 3D image
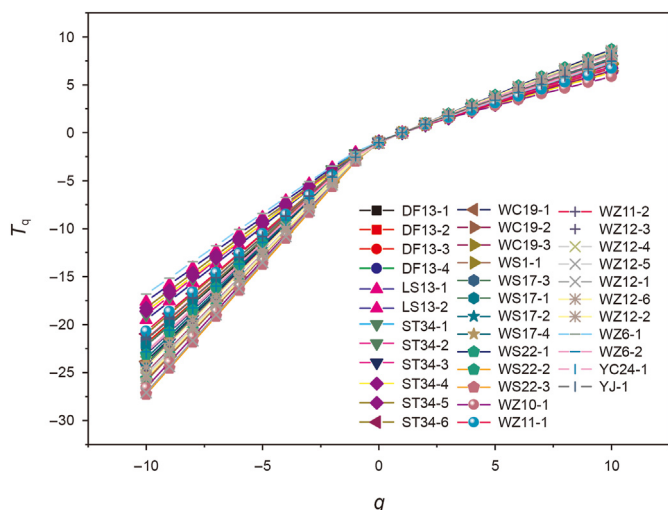
Fig. 7. The plots of quality index vs. *q* for all cores where the quality index between cores vary obviously at *q* < 0.

sets, data augmentation methods primarily involve geometric transformations such as rotation, flip, color transformation, cropping, scaling, translation, and noise, in order to generate similar yet distinct image sets. In the petroleum industry, image data augmentation is frequently employed for the transforming reservoir rocks images, including CT images, cast thin sections and scanning electron microscope (Karimpouli and Tahmasebi, 2019; Saxena et al., 2021; Chen et al., 2022). Among the algorithms used for data augmentation in 1D high-dimensional sample spaces (e.g., fractal parameter sets), SMOTE currently holds the most influence and serves as the established framework for most subsequent oversampling work. It effectively mitigates the risk of overfitting

associated with random oversampling (Douzas and Bacao, 2019; Douzas et al., 2021).

SMOTE performs data augmentation by generating synthetic data points based on existing raw data points, representing an advanced form of oversampling and a specific algorithm for data augmentation (Chawla et al., 2002). SMOTE consists of two main components: 1) selection rules for minority class instances, and 2) the mechanism for generating data after selecting these samples. Specifically, in the selection stage, different classes have been manually labeled. Regarding data generation, SMOTE follows these three steps to synthesize samples (Fig. 11):

(1) For each sample vector $\vec{a}$ in each mode of the NMR T_2 spectra, calculate its distance to the nearest k (= sample number) points samples in the same mode using the Euclidean distance.
(2) Determine the sampling ratio N (=2) based on the sample imbalance ratio. For each minority class sample vector $\vec{a}$, randomly select several samples from its k-nearest neighbors, assuming the selected neighbors are $\vec{b}$.
(3) For each randomly selected neighbor vector $\vec{b}$, generate a new sample vector $\vec{x}$ using the original sample vector $\vec{a}$ according to the following equation:

$$\vec{x} = \vec{a} + w \times \left(\vec{b} - \vec{a} \right), \; w = rand(0,1) \tag{12}$$

where $w(= 0.8)$ is a randomly generated weight from the range of 0–1. By employing SMOTE, the sample size of the multifractal parameter set with T_2 cut-off value increased from 36 to 561, resulting in new sample sizes of 172, 361, and 28 for the unimodal, bimodal, and trimodal distributions, respectively. The high-dimensional data was then visualized in 2D using the principal component analysis (PCA) method (Fig. 12). As depicted, SMOTE
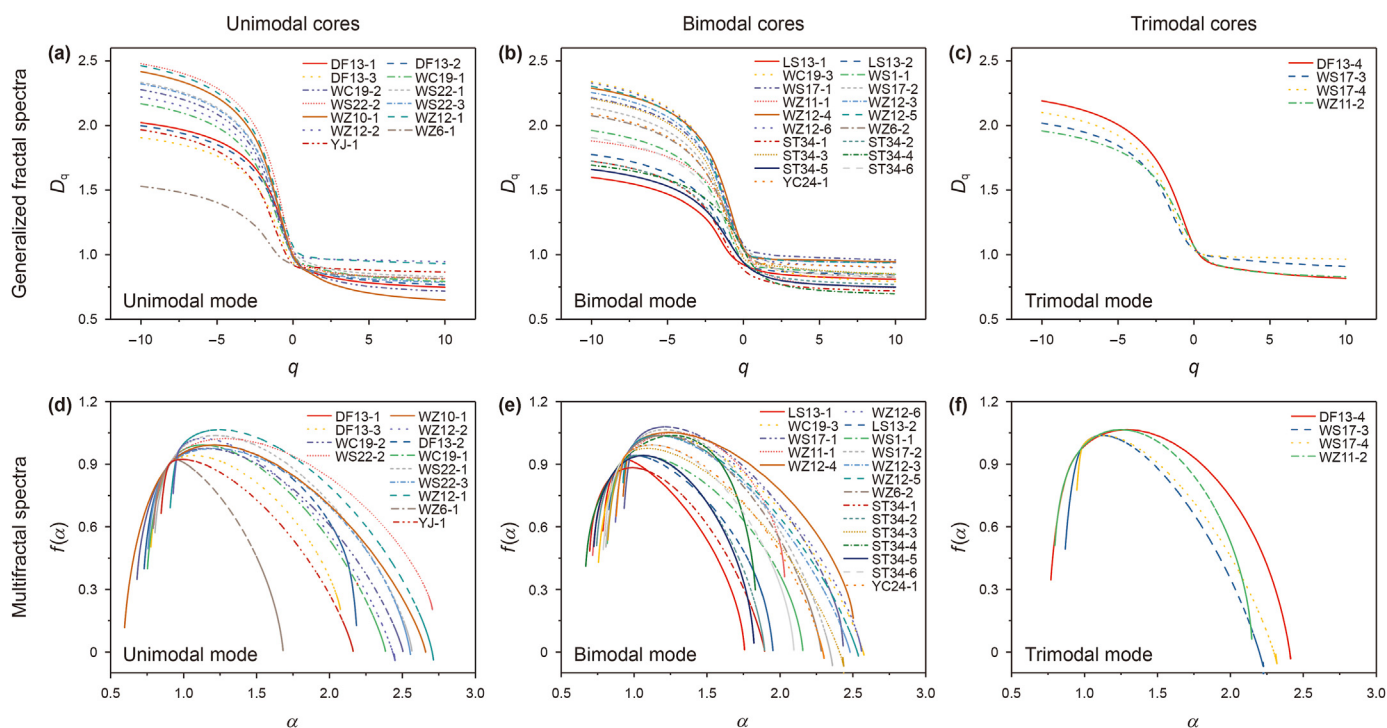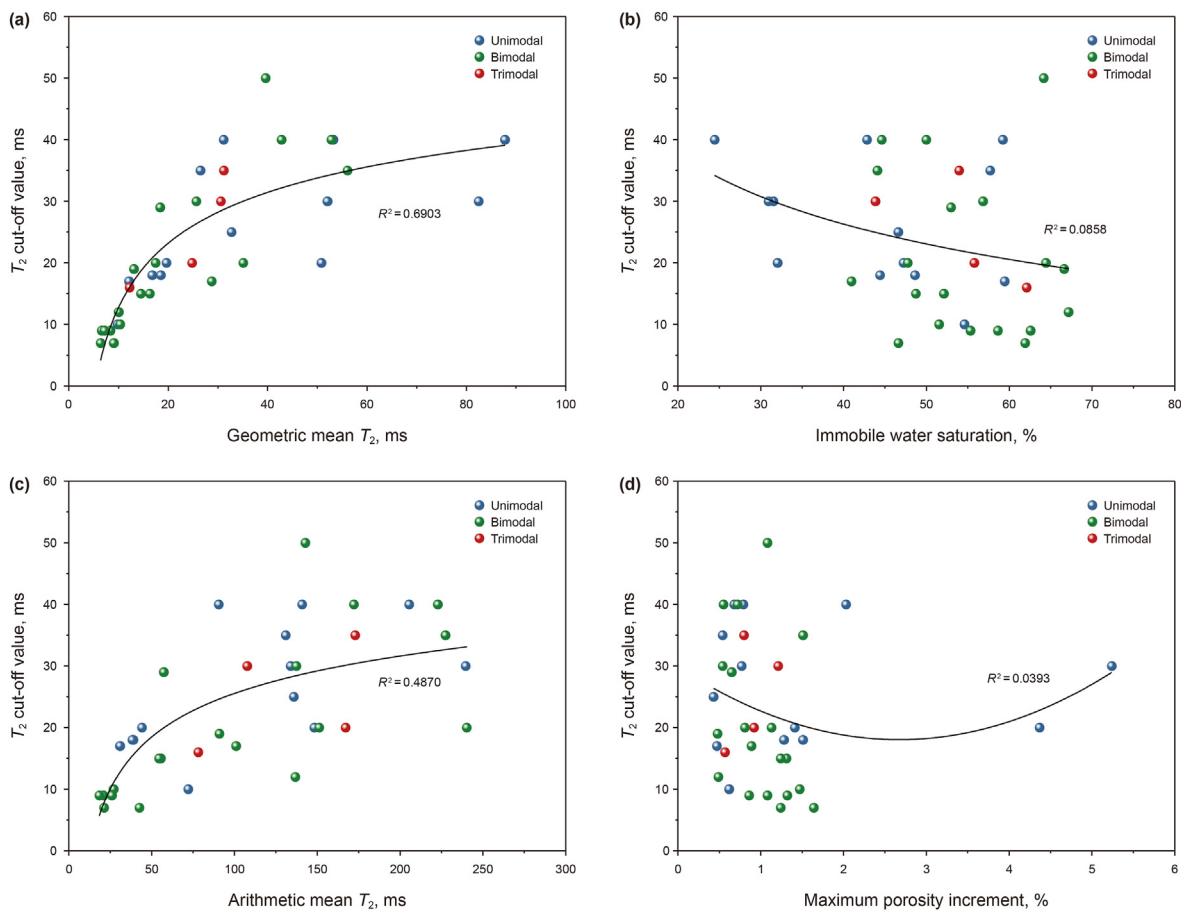

Fig. 8. Fractal spectra with *q* ranging from −10 to 10 of three types of T_2 spectral data (unimodal, bimodal, trimodel modes), including (a) (b) (c) generalized fractal spectra, and (d) (e) (f) multifractal spectra.

Fig. 9. Cross-plots of T_2 cut-off values versus geometric morphological parameters, including (a) geometric mean T_2 value; (b) Immobile water saturation; (c) Arithmetic mean T_2 value; (d) maximum porosity increment.
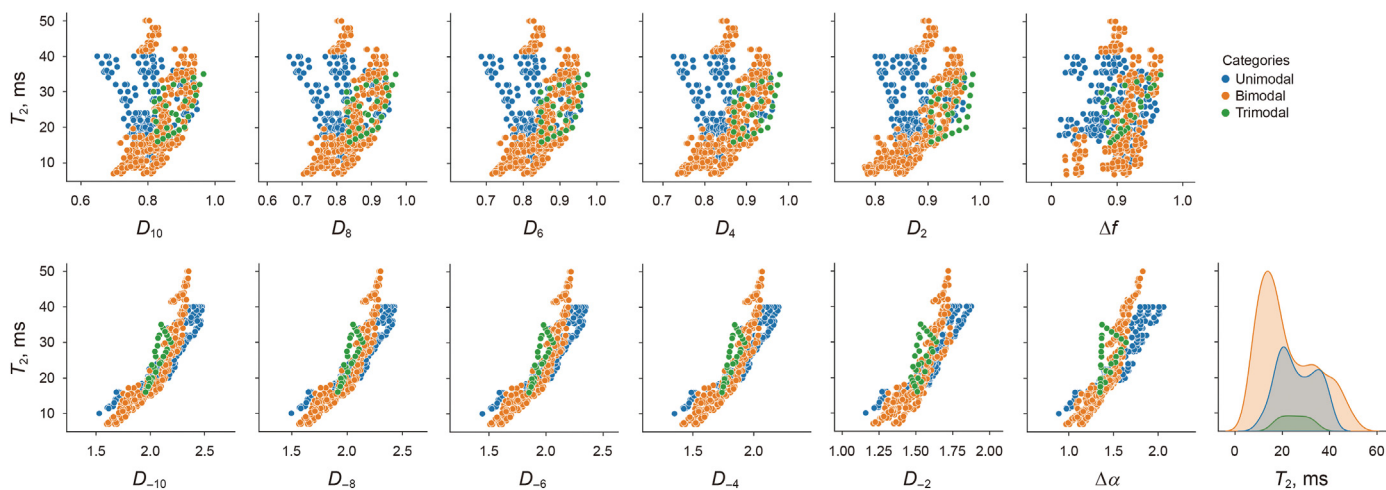
Fig. 10. Cross-plots of T_2 cut-off values versus multifractal parameters, including D_q with $q < 0$, D_q with $q > 0$, $\Delta\alpha$, Δf, and T_2 cut-off value distribution for the classified T_2 spectra including unimodal, and bimodal types.

enhances the sampling density of the dataset without altering the original mapping relationships between features and target variables. What needs to be explained here is the following: (1) The NMR data set used in this study is not divided into small sample sets separately. Therefore, by applying SMOTE, three types of NMR datasets, including unimodal, bimodal, and trimodal samples, were

simultaneously augmented. Typically, SMOTE does not alter the correlations between the original feature variables and the target variables. For a sample set that lacks intrinsic functional correlation, the sample sets after data augmentation by SMOTE will still lack intrinsic correlations. However, for a sample set that has an intrinsic functional correlation, it enhances the internal
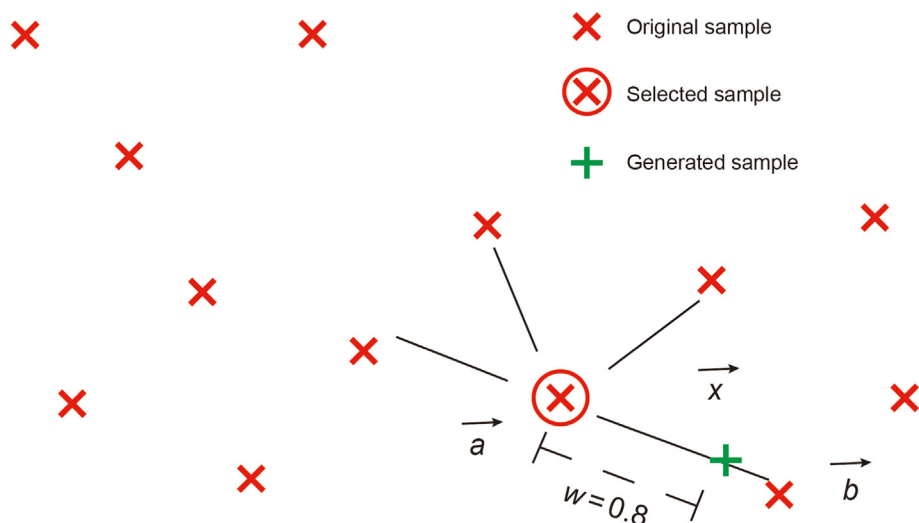
Fig. 11. Schematic diagram of the principle of generating new samples based on the original samples in the SMOTE oversampling method, where the nearest neighbor $k = 4$, and the weight $w = 0.8$.

correlations through data augmentation, making it detectable by the machine learning models. (2) If the newly generated data points in a specific sample set are scattered among other sample sets, those data points can be considered as noise, which can interfere with the identification of the above correlations (Douzas et al., 2018). To detect whether the generated data points are noise data, high-dimensional feature variables can be transformed into 2D variables using techniques such as PCA or t-SNE (t-distributed stochastic neighbor embedding). The result of this analysis is presented in Fig. 12. By employing SMOTE on this sample set, the newly generated data are valid data.

3.5. Machine learning for T_2 cut-off values

Machine learning allows for the discovery of implicit correlations between multiple variables and the target variable that cannot be captured by multiple linear regression. To assess the potential of machine learning in predicting the T_2 cut-off value, supervised machine learning regression was conducted using the augmented dataset. The samples were divided into two datasets for training purposes: 1) All samples ($N = 561$), and 2) The unimodal samples ($N = 172$) and the bimodal samples ($N = 361$), respectively. For each model training, the dataset was split into an 80% training set and a 20% test set.

Compared with morphological parameters, numerous multi-fractal parameters ($\Delta\alpha$, D_q ($q < 0$)) exhibited strong correlations with the T_2 cut-off values (Fig. 10). However, no significant correlations were observed for Δf and D_q ($q > 0$), (Fig. 10), which aligns with the T_q vs. q plots. The multifractal behaviors of cores varied significantly for $q < 0$ (Fig. 8). Additionally, the correlations differed between the unimodal and bimodal samples (Fig. 10). Notably, no significant correlations were observed for the trimodal mode, which is distinct from coal and shale reservoirs. $\Delta\alpha$ and D_q ($q < 0$) proved to be more sensitive to the morphological characteristics of the T_2 distribution compared to Δf and D_q ($q > 0$) (Fig. 10). Consequently, six sensitive parameters (D_{-10}, D_{-8}, D_{-6}, D_{-4}, D_{-2}, $\Delta\alpha$) were used as input variables (Fig. 12). To conduct the prediction of T_2 cut-off values, three commonly used regression models were employed: k-nearest neighbor (KNN), extreme gradient boosting (XGBoost), and gradient boosting decision tree (GBDT). Detailed methodology can be found in previous publications (Zhang and Zhou, 2007; Cai et al., 2020; Zhao et al., 2022b). Finally, the T_2

cut-off values, including both the unclassified dataset and the classified dataset by kurtosis, were compared with the measured results as shown in Fig. 13.

The correlation coefficients of the classified dataset (around 0.98 for unimodal, bimodal) were higher than that of the full-size unclassified dataset (approximately 0.89) (Fig. 13). Notably, all predicted data points closely align with the 1:1 line, indicating excellent agreement between the T_2 cut-off values calculated by machine learning regressions and those obtained from the centrifugation experiments with NMR tests. Among the three types of machine learning algorithms, XGBoost is the recommended model (Fig. 13). Besides, It is important to highlight that for T_2 cut-off values larger than 30 ms, the prediction results become more scattered, particularly for bimodal samples (Fig. 13). Additionally, special attention is required when dealing with trimodal mode in sandstones.

An error factor of ±5 ms can serve as a reasonable standard for low-permeability sandstones (Fig. 13), The combination of multi-fractal analysis with machine learning data augmentation through SMOTE proves to be an effective approach for determining the T_2 cut-off value in cases with limited NMR data. This method holds significant potential for T_2 NMR logging applications in low-permeability oil and gas reservoirs. It should be noted that the T_2 spectrum of low-permeability sandstone typically exhibits a single-peak or double-peak distribution, differs significantly from the typical three-peak distribution observed in coals and shales. Therefore, it is necessary to consider the kurtosis of the T_2 spectrum, augment the original sample size through data augmentation, and test various machine learning models.

There is no direct correlation between the T_2 cut-off value and the movable water saturation (Fig. 9b), since the fully-saturated T_2 spectrum is required. However, by combining the measured and predicted T_2 cut-off values with the fully-saturated T_2 spectrum, movable water saturation can be estimated. As demonstrated in Fig. 14, movable water saturation was reliably estimated using acceptable T_2 cut-off values. It is important to note that there is an error factor of 7% in the estimated water saturation due to linear interpolation of cumulative porosity at a specific T_2 time. Furthermore, although the trimodal T_2 cut-off values were not accurately estimated, they did not significantly impact the prediction of mobile water saturation. This may be attributed to the lower sensitivity of the T_2 cut-off value to saturation in trimodal samples
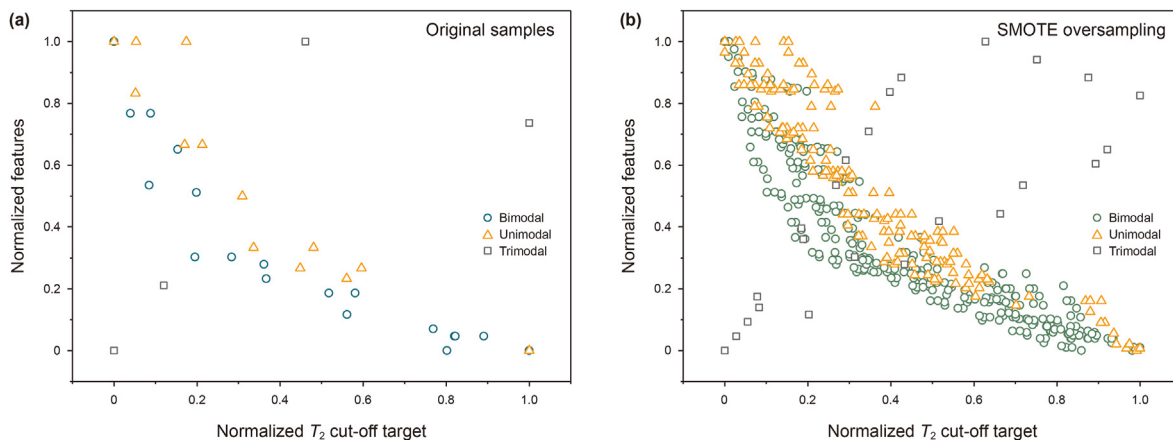
Fig. 12. Samples generated through oversampling multifractal dataset using SMOTE method.
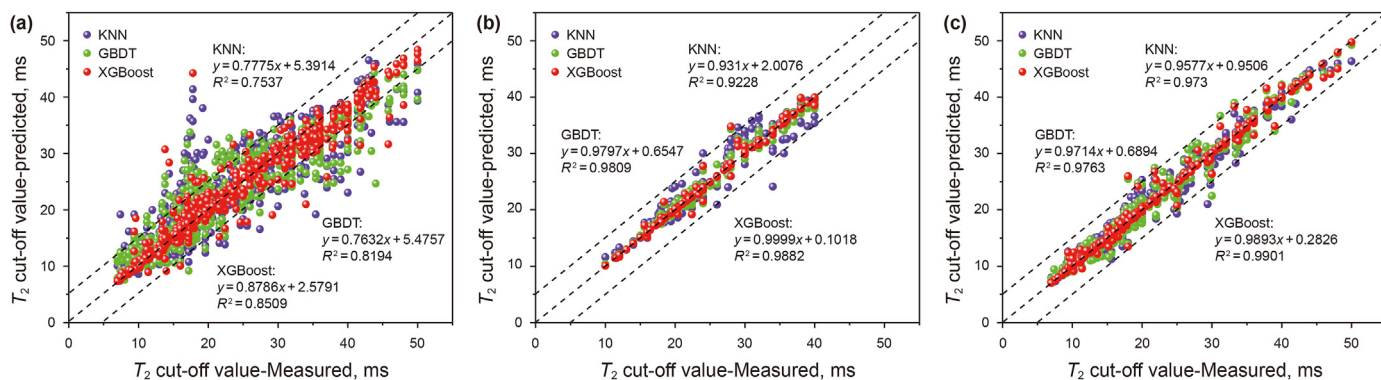

Fig. 13. Cross-plots of measured and predicted T_2 cut-off values by multifractal analysis with three machine learning models from (a) all samples; (b) the unimodal samples; (c) bimodal samples.
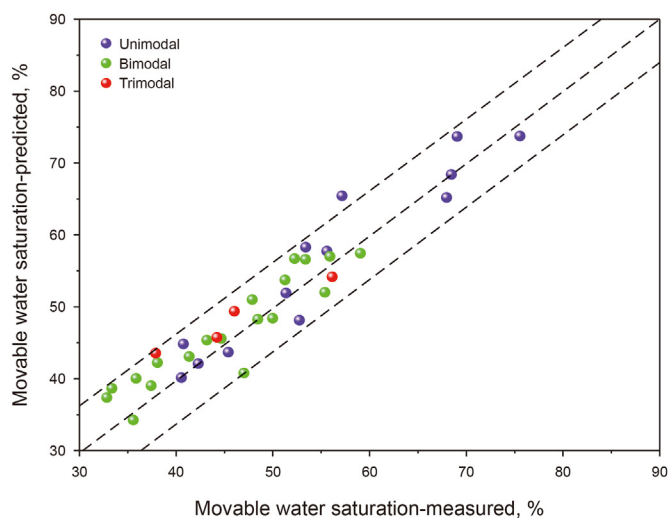

Fig. 14. Comparisons of measured and predicted movable water saturation by the measured and predicted T_2 cut-off values combined the full-size T_2 spectra at fully water-saturated sate from 36 cores.

compared to unimodal and bimodal samples.

SHAP (SHapley Additive exPlanation) was employed to address the interpretability of the XGBoost model. Derived from game theory, SHAP quantifies the importance of each feature by calculating the contribution of each feature to the target prediction based on the Shapley value (Shapley, 1953; Meng et al., 2022). The sum of the Shapley values for all features corresponds to the prediction output of the final model. The distribution of the global SHAP values for the six input features over the unimodal sample set was calculated (Fig. 15). In Fig. 15, the dotted line represents the base value, while the solid line represents the final output of the model. Among the features, $\Delta\alpha$ exhibit the most importance, followed by D_{-2}, D_{-10}. Notably, $\Delta\alpha$ provides a negative contribution approximately half of the samples. In addition, the SHAP method provides more detailed insights into the influence of specific features compared to the importance ranking map. Taking $\Delta\alpha$ as an example, this parameter exerted the most significant influence on the XGBoost model. When the value of $\Delta\alpha$ is smaller, it contributed negatively, resulting in a smaller model output, and vice versa (Fig. 16).

3.6. Effects of T2 spectrum position on multifractal analysis

Using the SHAP method, the importance of the eigenvectors was ranked. Among them, $\Delta\alpha$ and D_q ($q < 0$) were found to play a relatively significant role in predicting the T_2 cut-off values. The underlying mechanism was analyzed through multifractal theory. As known, the fractals and multifractals describe the structure, singularity, scale independence, order, self-similarity, vergence, anisotropy, self-organization, randomness, and criticality of objects of interest (Cheng, 1999). However, different fractal parameters exhibit different preferences in detecting these fractal features. In
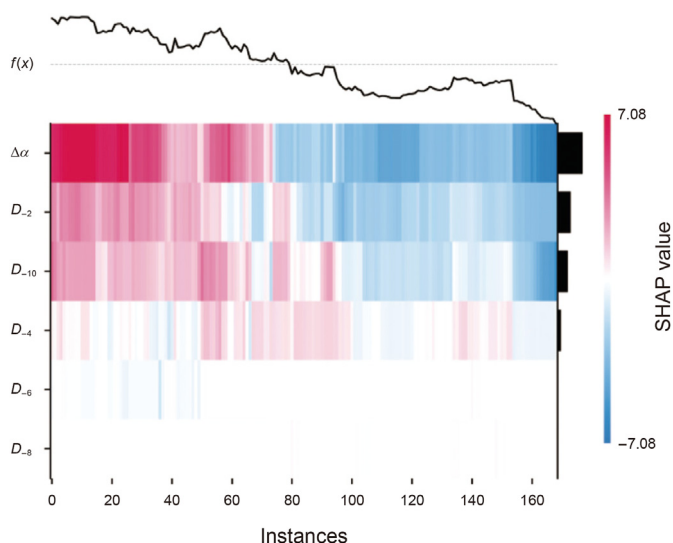
Fig. 15. The heatmap for a global interpretation by the SHAP value to all samples in predicting T_2 cut-off values.

cases where two T_2 spectra have the same shape but different positions relative to the X-axis, their T_2 cut-off values will differ significantly (Fig. 17a, d, g). However, if all the fractal parameters of these two T_2 spectra are equal, the robustness of multifractal parameters in predicting the T_2 cut-off value for a wide range of porous media diminishes (Hu et al., 2020). To illustrate this, the original T_2 spectrum was translated to generate two T_2 spectra with the same shape but different positions relative to the X-axis. The generalized fractal dimensions and multifractal dimensions are depicted in Fig. 17. The results demonstrate that $\Delta\alpha$ undergoes significant changes (increases), whereas Δf changes very little, for both unimodal (Fig. 17h) and bimodal (Fig. 17e) distributions, and even in the case of trimodal distributions (Fig. 17b). D_q ($q < 0$) exhibits substantial changes, while D_q ($q < 0$) changes only slightly (Fig. 17c, f, i). Hence, D_q ($q < 0$) and $\Delta\alpha$ can be used to accurately predict the T_2 cut-off value. The reason for this variation is provided from a multifractal theory perspective.

According to the multifractal principle, $\Delta\alpha$ reflects the unevenness of the distribution probability of T_2 value, while Δf indicates the structural changes among different probability subsets within the T_2 distribution (Stanley and Meakin, 1988; Ge et al., 2015). Specifically, larger values of $\Delta\alpha$ signify greater differences in the T_2 distribution, more pronounced polarization trends of each subset's

probability, and a more complex shape of the T_2 spectrum. Conversely, smaller values of $\Delta\alpha$ indicate smaller distribution differences in T_2 values, more concentrated and uniform distributions of each subset, and a higher degree of uniformity in the T_2 spectrum shape. Thus, $\Delta\alpha$ can quantitatively express the differences, non-uniformity, and polarization characteristics of the distribution of each subset within the T_2 spectrum shape, consequently reflecting the degree of fluctuation in the T_2 spectrum shape.

When $\Delta f < 0$, the multifractal spectrum of $\alpha - f(\alpha)$ exhibits a "right hook" shape, indicating that a subset of the T_2 distribution with a small probability dominates, and resulting in a relatively sharp T_2 spectrum. Conversely, when $\Delta f > 0$, the multifractal spectrum of $\alpha - f(\alpha)$ shows a "left hook" shape, signifying the dominance of the high probability subset and a sharper and rounder T_2 spectrum. Namely, Δf reflects the characteristics and relative proportion between T_2 spectral distributions. Larger Δf values correspond to more data points with larger actual T_2 values and uneven distribution. Thus, $\Delta\alpha$ detects the relative position of the T_2 spectrum and internal differences in pore information, while Δf mainly detects the smoothness and sharpness of the T_2 spectrum.

4. Conclusions

In this study, the pore types and NMR characteristics of 36 low-permeability reservoir sandstones were investigated by NMR experiments. The multifractal characteristic of the T_2 spectrum of the fully water-saturated medium was analyzed by using multifractal theory, and the correlations between T_2 cut-off values and porosity as well as permeability, morphological parameters, and multifractal parameters were discussed. The following conclusions were drawn:

(1) The T_2 spectra of low-permeability reservoir sandstones exhibited three modes of peak numbers: unimodal, bimodal, and trimodal, with the trimodal mode being relatively rare. The number of peaks in the T_2 spectrum was directly determined by pore types. The T_2 cut-off value ranged from 7 to 50 ms, with an average of 23.25 ms. This value could not be predicted by petrophysical parameters, curve morphological parameters, or simple mathematical statistical analysis.

(2) The T_2 spectra captures the multifractal behavior of the pore system within low-permeability sandstones. $\Delta\alpha$ and D_q ($q < 0$) were found to have a strongly correlated with the T_2 cut-off value. This correlation was attributed to $\Delta\alpha$ and $D_q(q < 0)$ detecting the position of the T_2 spectrum relative to the X-axis. However, the effects of the T_2 spectrum position
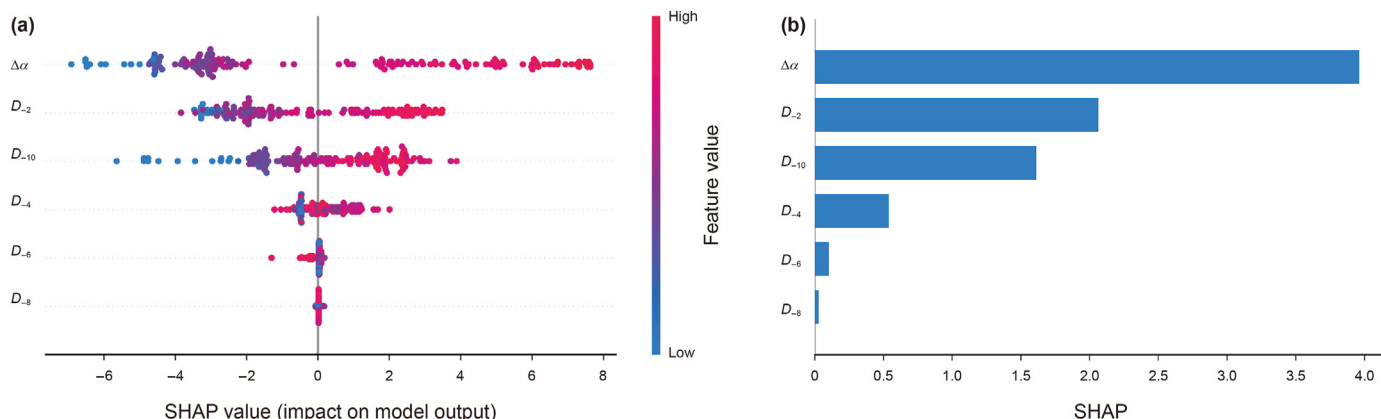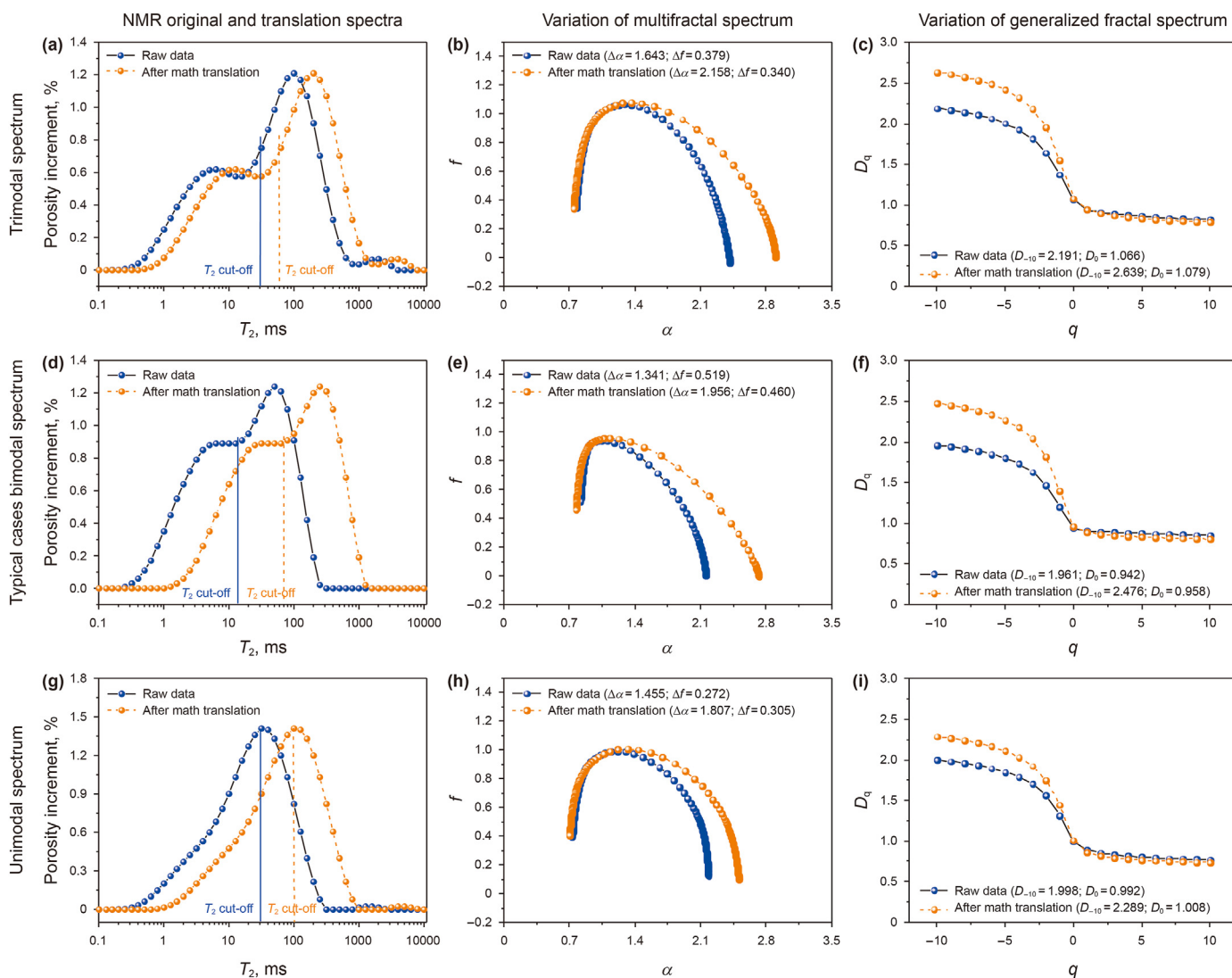

Fig. 16. (a) SHAP values assigned to different multifractal features; (b) Bar diagram showing the importance ranking of the six multifractal features.

Fig. 17. Effects of the location of T_2 spectrum relative to X-axis on calculations of multifractal analysis using math translation to produce original and translational NMR T_2 spectra from (a) sample # DF 13−4 as a trimodal case; (d) sample # WS1-1 as a bimodal case; (g) sample # DF13-2 as a unimodal case; (b) (e) (h) calculated multifractal spectra from (a) (d) (g); (c) (f) (i) calculated generalized fractal spectra from (a) (d) (g).

on D_q ($q > 0$) and Δf were minimal, leading to the inability of Δf and D_q ($q > 0$) to determine the T_2 cut-off value.

(3) The limitation of NMR data volume owing to experimental cost can be overcome by oversampling with the classic SMOTE technique. SMOTE can augment sample data without altering the underlying functional mapping relationships within NMR T_2 dataset.

(4) When classifying the T_2 spectra of low-permeability reservoir sandstones based on the number of peaks, machine learning models showed higher correlation coefficients when using six sensitive features of $\Delta \alpha$ and $D_q(q < 0)$ as inputs. The XGBoost model was recommended in determining T_2 cut-off values. $\Delta \alpha$ exhibited the greatest importance according to the SHAP method. No significant correlation between the T_2 cut-off value and movable water saturation was observed. Combining with the predicted T_2 cut-off value with the fully-saturated T_2 spectrum, movable water saturation can be reliably estimated.

(5) However, due to the limitations in experimental data, the multifractal method did not yield satisfactory performance

for the trimodal mode. Estimations became more scattered, especially for bimodal samples, when the T_2 cut-off value exceeded 30 ms. Considering the influences of random errors, an error factor of ± 5 ms can be utilized as a reliable boundaries for T_2 cut-off value estimation.

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.

acknowledged for funding Chen's research. The experimental data used are presented in the Figures.

References

Blunt, M.J., 2017. Multiphase Flow in Permeable Media: A Pore-Scale Perspective. Cambridge University Press. https://doi.org/10.1017/9781316145098.

Blunt, M.J., Bijeljic, B., Dong, H., et al., 2013. Pore-scale imaging and modelling. Adv. Water Resour. 51, 197–216. https://doi.org/10.1016/j.advwatres.2012.03.003.

Bultreys, T., De Boever, W., Cnudde, V., 2016. Imaging and image-based fluid transport modeling at the pore scale in geological materials: a practical introduction to the current state-of-the-art. Earth-Sci. Rev. 155, 93–128. https://doi.org/10.1016/j.earscirev.2016.02.001.

Butler, J.P., Reeds, J.A., Dawson, S.V., 1981. Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing. SIAM J. Numer. Anal. 18, 381–397. https://doi.org/10.1137/0718025.

Cai, J., Xu, K., Zhu, Y., et al., 2020. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. Appl. Energy 262, 114566. https://doi.org/10.1016/j.apenergy.2020.114566.

Camacho, L., Douzas, G., Bacao, F., 2022. Geometric SMOTE for regression. Expert Syst. Appl. 193, 116387. https://doi.org/10.1016/j.eswa.2021.116387.

Chawla, N.V., Bowyer, K.W., Hall, L.O., et al., 2002. SMOTE: synthetic minority oversampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, J., Huang, H., Cohn, A.G., et al., 2022. Machine learning-based classification of rock discontinuity trace: SMOTE oversampling integrated with GBT ensemble learning. Int. J. Min. Sci. Technol. 32, 309–322. https://doi.org/10.1016/j.ijmst.2021.08.004.

Chen, X., Yao, G., Cai, J., et al., 2017. Fractal and multifractal analysis of different hydraulic flow units based on micro-CT images. J. Nat. Gas Sci. Eng. 48, 145–156. https://doi.org/10.1016/j.jngse.2016.11.048.

Chen, X., Yao, G., Herrero-Bervera, E., et al., 2018. A new model of pore structure typing based on fractal geometry. Mar. Petrol. Geol. 98, 291–305. https://doi.org/10.1016/j.marpetgeo.2018.08.023.

Cheng, Q., 1999. Multifractality and spatial statistics. Comput. Geosci. 25, 949–961. https://doi.org/10.1016/S0098-3004(99)00060-6.

Chhabra, A., Jensen, R.V., 1989. Direct determination of the $f(\alpha)$ singularity spectrum. Phys. Rev. Lett. 62, 1327. https://doi.org/10.1103/PhysRevLett.62.1327.

Coates, G.R., Xiao, L., Prammer, M.G., 1999. NMR Logging: Principles and Applications. Haliburton Energy Services, Houston.

Daigle, H., Johnson, A., 2016. Combining mercury intrusion and nuclear magnetic resonance measurements using percolation theory. Transp. Porous Media 111, 669–679. https://doi.org/10.1007/s11242-015-0619-1.

Douzas, G., Bacao, F., 2019. Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE. Inf. Sci. 501, 118–135. https://doi.org/10.1016/j.ins.2019.06.007.

Douzas, G., Bacao, F., Last, F., 2018. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Inf. Sci. 465, 1–20. https://doi.org/10.1016/j.ins.2018.06.056.

Douzas, G., Rauch, R., Bacao, F., 2021. G-SOMO: an oversampling approach based on self-organized maps and geometric SMOTE. Expert Syst. Appl. 183, 115230. https://doi.org/10.1016/j.eswa.2021.115230.

Dullien, F.A., 2012. Porous Media: Fluid Transport and Pore Structure. Academic Press. https://doi.org/10.1016/0300-9467(81)80049-4.

Fu, H., Wang, W., Chen, X., et al., 2019. Fractal and multifractal analysis of fracture surfaces caused by hydrogen embrittlement in high-Mn twinning/transformation-induced plasticity steels. Appl. Surf. Sci. 470, 870–881. https://doi.org/10.1016/j.apsusc.2018.11.179.

Gao, H., Li, H.A., 2016. Pore structure characterization, permeability evaluation and enhanced gas recovery techniques of tight gas sandstones. J. Nat. Gas Sci. Eng. 28, 536–547. https://doi.org/10.1016/j.jngse.2015.12.018.

Ge, X., Fan, Y., Deng, S., 2011. Research on T_2 cutoff-value determination method for shaly sand based on experiments. Well Logging Technol. 35, 308–313.

Ge, X., Fan, Y., Zhu, X., et al., 2015. Determination of nuclear magnetic resonance T_2 cutoff value based on multifractal theory—an application in sandstone with complex pore structure. Geophysics 80, D11–D21. https://doi.org/10.1190/geo2014-0140.1.

Ge, X., Chen, H., Fan, Y., et al., 2017. An improved pulse sequence and inversion algorithm of T_2 spectrum. Comput. Phys. Commun. 212, 82–89. https://doi.org/10.1016/j.cpc.2016.10.012.

Ge, X., Fan, Y., Liu, J., et al., 2021. Numerical investigating the low field NMR response of representative pores at different pulse sequence parameters. Comput. Geosci. 151, 104761. https://doi.org/10.1016/j.cageo.2021.104761.

Godefroy, S., Fleury, M., Deflandre, F., et al., 2002. Temperature effect on NMR surface relaxation in rocks for well logging applications. J. Phys. Chem. B 106, 11183–11190. https://doi.org/10.1021/jp0213452.

Guo, J.-C., Zhou, H.-Y., Zeng, J., et al., 2020. Advances in low-field nuclear magnetic resonance (NMR) technologies applied for characterization of pore space inside rocks: a critical review. Petrol. Sci. 17, 1281–1297. https://doi.org/10.1007/s12182-020-00488-0.

Guo, X., Huang, Z., Zhao, L., et al., 2019. Pore structure and multi-fractal analysis of tight sandstone using MIP, NMR and NMRC methods: a case study from the Kuqa depression, China. J. Petrol. Sci. Eng. 178, 544–558. https://doi.org/

Hidajat, I., Mohanty, K., Flaum, M., et al., 2004. Study of vuggy carbonates using NMR and X-ray CT scanning. SPE Reservoir Eval. Eng. 7, 365–377. https://doi.org/10.2118/88995-pa.

Hu, Y., Guo, Y., Zhang, J., et al., 2020. A method to determine nuclear magnetic resonance T_2 cutoff value of tight sandstone reservoir based on multifractal analysis. Energy Sci. Eng. 8, 1135–1148. https://doi.org/10.1002/ese3.574.

Jiao, L., Andersen, P.Ø., Zhou, J., et al., 2020. Applications of mercury intrusion capillary pressure for pore structures: a review. Capillarity 3, 62–74. https://doi.org/10.46690/capi.2020.04.02.

Karimpouli, S., Tahmasebi, P., 2019. Image-based velocity estimation of rock using Convolutional Neural Networks. Neural Network. 111, 89–97. https://doi.org/10.1016/j.neunet.2018.12.006.

Kleinberg, R., Boyd, A., 1997. Tapered Cutoffs for Magnetic Resonance Bound Water Volume. SPE Annual Technical Conference and Exhibition, OnePetro. https://doi.org/10.2523/38737-ms.

Li, C., Liu, G., Cao, Z., et al., 2020a. Analysis of petrophysical characteristics and water movability of tight sandstone using low-field nuclear magnetic resonance. Nat. Resour. Res. 29, 2547–2573. https://doi.org/10.1007/s11053-019-09582-6.

Li, J., Jiang, C., Wang, M., et al., 2020b. Adsorbed and free hydrocarbons in unconventional shale reservoir: a new insight from NMR T1-T2 maps. Mar. Petrol. Geol. 116, 104311. https://doi.org/10.1016/j.marpetgeo.2020.104311.

Liu, J., Pereira, G.G., Liu, Q., et al., 2016. Computational challenges in the analyses of petrophysics using microtomography and upscaling: a review. Comput. Geosci. 89, 107–117. https://doi.org/10.1016/j.cageo.2016.01.014.

Liu, K., Mirzaei-Paiaman, A., Liu, B., et al., 2020a. A new model to estimate permeability using mercury injection capillary pressure data: application to carbonate and shale samples. J. Nat. Gas Sci. Eng. 84, 103691. https://doi.org/10.1016/j.jngse.2020.103691.

Liu, K., Jin, Z., Zeng, L., et al., 2021a. Determination of clay bound water in shales from NMR signals: the fractal theory. Energy Fuels 35, 18406–18413. https://doi.org/10.1021/acs.energyfuels.1c02285.

Liu, X., Jin, Z., Lai, J., et al., 2021b. Fractal behaviors of nmr saturated and centrifugal T2 spectra in oil shale reservoirs: the paleogene funing formation in subei basin, China. Mar. Petrol. Geol. 129, 105069. https://doi.org/10.1016/j.marpetgeo.2021.105069.

Liu, Z., Liu, D., Cai, Y., et al., 2020b. Application of nuclear magnetic resonance (NMR) in coalbed methane and shale reservoirs: a review. Int. J. Coal Geol. 218, 103261. https://doi.org/10.1016/j.coal.2019.103261.

Lopes, R., Betrouni, N., 2009. Fractal and multifractal analysis: a review. Med. Image Anal. 13, 634–649. https://doi.org/10.1016/j.media.2009.05.003.

Mai, A., Kantzas, A., 2002. An evaluation of the application of low field NMR in the characterization of carbonate reservoirs. In: SPE Annual Technical Conference and Exhibition. OnePetro. https://doi.org/10.2118/77401-ms.

Mehmani, A., Verma, R., Prodanović, M., 2020. Pore-scale modeling of carbonates. Mar. Petrol. Geol. 114, 104141. https://doi.org/10.1016/j.marpetgeo.2019.104141.

Meng, J., Zhou, Y.-J., Ye, T.-R., et al., 2022. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. Petrol. Sci. 20, 277–294. https://doi.org/10.1016/j.petsci.2022.09.003.

Morriss, C., Rossini, D., Straley, C., et al., 1997. Core analysis by low-field NMR. Log. Anal. 38.

Nicot, B., Ligneul, P., Akbar, M., 2016. T_2-cutoff determination using magnetic susceptibility measurements. Google Patents.

Oraby, M., Chafai, N., Hussing, R., et al., 1997. A new NMR interpretation technique using error minimization with variable T_2 cutoff. In: SPE Annual Technical Conference and Exhibition. OnePetro. doi:10.2118/38734-ms.

Peduzzi, P., Concato, J., Kemper, E., et al., 1996. A simulation study of the number of events per variable in logistic regression analysis. J. Clin. Epidemiol. 49, 1373–1379. https://doi.org/10.1016/S0895-4356(96)00236-3.

Saxena, N., Day-Stirrat, R.J., Hows, A., et al., 2021. Application of deep learning for semantic segmentation of sandstone thin sections. Comput. Geosci. 152, 104778. https://doi.org/10.1016/j.cageo.2021.104778.

Shapley, L., 1953. Quota Solutions Op N-Person Games1. Emil Artin and Marston Morse, p. 343. https://doi.org/10.1515/9781400881970-021.

Sigal, R.F., 2015. Pore-size distributions for organic-shale-reservoir rocks from nuclear-magnetic-resonance spectra combined with adsorption measurements. SPE J. 20, 824–830. https://doi.org/10.2118/174546-pa.

Stanley, H.E., Meakin, P., 1988. Multifractal phenomena in physics and chemistry. Nature 335, 405–409. https://doi.org/10.1038/335405a0.

Sun, Y., Zhai, C., Zhao, Y., et al., 2021. Multifractal analysis and neural network prediction of pore structures in coal reservoirs based on NMR T_2 spectra. Energy Fuels 35, 11306–11318. https://doi.org/10.1021/acs.energyfuels.1c01409.

Tang, M., Wang, C., Yang, H., et al., 2022. Experimental investigation on plugging performance of nanospheres in low-permeability reservoir with bottom water. Adv. Geo-Energy Res. 6, 95–103. https://doi.org/10.46690/ager.2022.02.02.

Testamanti, M.N., Rezaee, R., 2017. Determination of NMR T_2 cut-off for clay bound water in shales: a case study of Carynginia Formation, Perth Basin, Western Australia. J. Petrol. Sci. Eng. 149, 497–503. https://doi.org/10.1016/j.petrol.2016.10.066.

Wang, F., Yang, K., Cai, J., 2018a. Fractal characterization of tight oil reservoir pore structure using nuclear magnetic resonance and mercury intrusion porosimetry. Fractals 26, 1840017. https://doi.org/10.1142/s0218348x18400170.

Wang, F., Yang, K., You, J., et al., 2019. Analysis of pore size distribution and fractal dimension in tight sandstone with mercury intrusion porosimetry. Results Phys.

13, 102283. https://doi.org/10.1016/j.rinp.2019.102283.

Wang, G., Han, D., Qin, X., et al., 2020. A comprehensive method for studying pore structure and seepage characteristics of coal mass based on 3D CT reconstruction and NMR. Fuel 281, 118735. https://doi.org/10.1016/j.fuel.2020.118735.

Wang, Z.H., Zhang, C.G., Xiao, C.W., et al., 2004. Experimental study of T_2-cutoff values in low-permeability reservoirs. Prog. Geophys. 19, 652–655. https://doi.org/10.1016/S0960-0779(03)00420-X.

Wang, Z., Pan, M., Shi, Y., et al., 2018b. Fractal analysis of Donghetang sandstones using NMR measurements. Energy Fuels 32, 2973–2982. https://doi.org/10.1021/acs.energyfuels.7b03463.

Westphal, H., Surholt, I., Kiesl, C., et al., 2005. NMR measurements in carbonate rocks: problems and an approach to a solution. Pure Appl. Geophys. 162, 549–570. https://doi.org/10.1007/s00024-004-2621-3.

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, L., Li, J., Mao, Z., et al., 2018. A method to determine nuclear magnetic resonance (NMR) T_2 cutoff based on normal distribution simulation in tight sandstone reservoirs. Fuel 225, 472–482. https://doi.org/10.1016/j.fuel.2018.03.178.

Yan, J., He, X., Zhang, S., et al., 2020. Sensitive parameters of NMR T_2 spectrum and their application to pore structure characterization and evaluation in logging profile: a case study from Chang 7 in the Yanchang Formation, Heshui area, Ordos Basin, NW China. Mar. Petrol. Geol. 111, 230–239. https://doi.org/10.1016/j.marpetgeo.2019.08.025.

Yan, W., Sun, J., Golsanami, N., et al., 2019. Evaluation of wettabilities and pores in tight oil reservoirs by a new experimental design. Fuel 252, 272–280. https://doi.org/10.1016/j.fuel.2019.04.130.

Yao, Y., Liu, D., Che, Y., et al., 2010. Petrophysical characterization of coals by low-field nuclear magnetic resonance (NMR). Fuel 89, 1371–1380. https://doi.org/10.1016/j.fuel.2009.11.005.

Zhang, M.-L., Zhou, Z.-H., 2007. ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn. 40, 2038–2048. https://doi.org/10.1016/j.patcog.2006.12.019.

Zhang, Y., Li, Q., Hu, Q., et al., 2022. Pore wetting process characterization of Equal-Sized granular coals by using LF-NMR technology. Fuel 313, 122670. https://doi.org/10.1016/j.fuel.2021.122670.

Zhao, P., Wang, Z., Sun, Z., et al., 2017. Investigation on the pore structure and multifractal characteristics of tight oil reservoirs using NMR measurements: permian Lucaogou Formation in Jimusaer Sag, Junggar Basin. Mar. Petrol. Geol. 86, 1067–1081. https://doi.org/10.1016/j.marpetgeo.2017.07.011.

Zhao, P., Wang, X., Cai, J., et al., 2019. Multifractal analysis of pore structure of Middle Bakken formation using low temperature N_2 adsorption and NMR measurements. J. Petrol. Sci. Eng. 176, 312–320. https://doi.org/10.1016/j.petrol.2019.01.040.

Zhao, X., Yao, G., Chen, X., et al., 2022a. Diagenetic facies classification and characterization of a high-temperature and high-pressure tight gas sandstone reservoir: a case study in the Ledong area, Yinggehai Basin. Mar. Petrol. Geol. 140, 105665. https://doi.org/10.1016/j.marpetgeo.2022.105665.

Zhao, X., Chen, X., Huang, Q., et al., 2022b. Logging-data-driven permeability prediction in low-permeable sandstones based on machine learning with pattern visualization: a case study in Wenchang A Sag, Pearl River Mouth Basin. J. Petrol. Sci. Eng. 214, 110517. https://doi.org/10.1016/j.petrol.2022.110517.

Zhao, Y., Lin, B., Liu, T., et al., 2021. Multifractal analysis of coal pore structure based on NMR experiment: a new method for predicting T_2 cutoff value. Fuel 283, 119338. https://doi.org/10.1016/j.fuel.2020.119338.

Zheng, S., Yao, Y., Liu, D., et al., 2019. Nuclear magnetic resonance T_2 cutoffs of coals: a novel method by multifractal analysis theory. Fuel 241, 715–724. https://doi.org/10.1016/j.fuel.2018.12.044.

Zhong, J., Yan, R., Zhang, H., et al., 2020. A decomposition method of nuclear magnetic resonance T_2 spectrum for identifying fluid properties. Petrol. Explor. Dev. 47, 740–752. https://doi.org/10.1016/S1876-3804(20)60089-1.