GeoStorage

Vol. 2, No. 3, p. 216-229, 2026

Research article

# Prediction Method for CT Pore Extraction Thresholds in Tight Sandstone Based on NMR Porosity Constraint and Grayscale S-tatistical Features

Xi Wang[1,2], Gong Zhang[1,2,*], Dong Zhao[1,2]

[1] *Key Laboratory of Oil and Gas Resources and Exploration Technology, Ministry of Education, Yangtze University, Wuhan 430100, China*
[2] *College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*

**Keywords:**
Tight sandstone
digital core
grayscale statistics
pore extraction threshold
NMR

**Cited as:**
Wang X, Zhang G, Zhao D. 2026. Prediction Method for CT Pore Extraction Thresholds in Tight Sandstone Based on NMR Porosity Constraint and Grayscale Statistical Features. *GeoStorage*, 2(3), 216-229.
https://doi.org/10.46690/gs.2026.03.02

**Abstract:**
Three-dimensional core CT scanning is widely used for reservoir microstructure characterization. However, the accuracy of pore extraction is strongly influenced by manually selected grayscale segmentation thresholds, resulting in uncertainties in pore characterization. This study proposes a data-driven method for predicting CT pore extraction thresholds based on NMR porosity constraints and grayscale statistical features. The NMR-derived porosity of 11 tight sandstone cores was first used as a constraint, and five grayscale parameters, including mean, median, mode, standard deviation, and kurtosis, were extracted from CT images. Their relationships with the pore extraction threshold were analyzed, and a ridge regression model was developed for threshold prediction. Results show that the mean, median, and mode exhibit strong correlations with the threshold, with correlation coefficients exceeding 0.94, while kurtosis shows a correlation coefficient of 0.82. The proposed model achieves high prediction accuracy, with an $R^2$ of 0.984 and a mean absolute percentage error of 0.69% in the training dataset, and maintains good generalization performance with a cross-validation $R^2$ of 0.858. This method provides an effective approach for reducing subjective uncertainty in CT pore segmentation and improving the reliability of digital core characterization in tight sandstone reservoirs.

## 1 Introduction

With the increasing global energy demand, unconventional oil and gas resources are gradually becoming an important direction for energy development (Chen et al., 2021; Wang et al., 2022).Tight sandstone reservoirs, as a typical unconventional reservoir, show broad prospects for exploration and development. However, tight sandstone reservoirs generally have characteristics such as low porosity, low permeability, and complex mineral composition and pore structure, which pose a serious challenge to accurate reservoir evaluation (He et al., 2025; Xue et al., 2025).

Three-dimensional core CT technology enables non-destructive, high-resolution, and three-dimensional observation of the internal structure of rocks, and is an important means of characterizing tight sandstone reservoirs (Wang et al., 2023). Accurate extraction of porosity from reservoir core CT data is crucial for precise reservoir evaluation. The correct grayscale segmentation threshold for the core CT data directly determines the accuracy of porosity extraction.

Wildenschild and Sheppard (2013) proposed that the grayscale distribution of CT data of conventional sandstone with sufficiently large pore size and good image quality exhibits a distinct ”bimodal distribution,” with the two peaks corresponding to “pores” and “solid matrix,” respectively. Schlüter et al. (2014) pointed out in X-ray micro-CT image processing that, in the bimodal distribution of the CT grayscale histogram, the valley between the two peaks representing pores and the solid matrix is the key location for determining the optimal segmentation threshold. Desbois et al. (2016) used the Otsu

Yandy Scientific Press

* Corresponding author.
*E-mail address*: 1976475685@qq.com (Wang X); zhanggong@yangtzeu.edu.cn (Zhang G); zhaodong@yangtzeu.edu.cn (Zhao D)

Received Jun. 10, 2026; revised Jun. 17, 2026; accepted Jun. 24, 2026; available online Jul. 24, 2026.

algorithm to automatically locate the trough in the acquired two-dimensional image sequence by maximizing the inter-class variance between the foreground and background, thus segmenting pores and minerals. Arunpandian et al. (2018) proposed a threshold selection method combining class variance and maximum entropy (CME-CV) to find the optimal trough point for extracting pore structure in soil CT images. These methods are effective for extracting pores from core CT data that exhibit a bimodal distribution. However, the pore scale of tight sandstone is small, and its three-dimensional core CT data does not show a clear "bimodal distribution" and lacks obvious gray-scale distribution troughs, causing traditional threshold selection methods to fail. In recent years, some scholars have proposed methods for selecting pore extraction thresholds in tight sandstone based on human experience. Wang et al. (2024a), when pre-segmenting tight sandstone CT data using SLIC superpixels, found that the original grayscale distribution exhibited a single-peak tailing pattern. Traditional thresholding methods would result in a large number of missegments, necessitating the introduction of prior human knowledge to determine the pore extraction threshold. Deng et al. (2024) found that the grayscale distribution of CT data from tight sandstone reservoir cores exhibited a distinct single-peak morphology, rendering the conventional trough threshold selection method ineffective. They indicated that threshold selection required reliance on personal experience and mercury intrusion porosimetry, and that different operators' threshold selections could introduce ±10% pore volume errors. When Feng et al. (2025) conducted core CT experiments on tight sandstone, they found that the grayscale distribution of tight sandstone exhibited a unimodal to weak shoulder pattern, requiring repeated manual adjustment of the threshold to extract pores. These studies indicate that relying on personal experience to select porosity extraction thresholds is prone to introducing large errors, leading to a significant decrease in the accuracy of porosity extraction in tight sandstone reservoirs (Hao et al., 2023; Li et al., 2020).

In recent years, with the development of digital core image processing methods, multi-scale CT imaging, multi-threshold segmentation, machine learning, and deep learning methods have gradually been applied to pore identification in rock CT images. Hu et al. (2025) used micro-CT and nano-CT to obtain micro-scale and nano-scale grayscale images of tight sandstone, respectively, and adopted a multi-threshold image segmentation method to process grayscale images at different scales, achieving multi-scale pore-throat structure characterization of tight sandstone. Wang et al. (2025) combined micro-CT images with different resolutions and SEM images, and constructed multi-scale three-dimensional digital rocks through an associated image segmentation method to compensate for the difficulty of identifying nanoscale pores and clay micropores using single-resolution CT images. In terms of machine learning and deep learning segmentation, Wang et al. (2024b) compared the application effects of U-Net and its variants in segmenting microfractures, pore spaces, and solid matrices in digital rocks, indicating that deep learning methods can improve the multi-object segmentation capability of complex rock images. Hayatdavoudi et al. (2025) used several convolutional neural network architectures to segment sandstone micro-CT images and compared them with the traditional Otsu thresholding method, demonstrating the application potential of CNN methods in automatic digital rock image segmentation. In addition, Liu et al. (2025a) proposed a grayscale threshold segmentation method combining Retinex enhancement and Laplacian convolution to address the common problems of unimodal grayscale distribution and unclear pore-matrix boundaries in geomaterial images, which improved the pore identification performance of complex grayscale images. These studies have promoted the development of digital rock pore identification methods from the perspectives of multi-scale imaging, improved threshold segmentation, and deep learning-based automatic segmentation. However, most of these methods mainly focus on image segmentation algorithms themselves or multi-scale pore-throat structure characterization, while insufficient attention has been paid to how pore extraction thresholds under unimodal or weak-shoulder grayscale distributions in tight sandstone can be quantitatively constrained by experimental porosity.

To address the aforementioned issues, this study proposes a porosity extraction threshold prediction method based on the gray-scale distribution characteristics of core CT data, incorporating nuclear magnetic resonance (NMR) technology. First, the porosity of tight sandstone cores is measured using NMR to constrain the prediction accuracy of the porosity extraction threshold. Then, the gray-scale distribution characteristics of the tight sandstone core CT data, including mean, median, mode, standard deviation, and kurtosis, are calculated, and their relationship with the porosity extraction threshold is analyzed. Finally, a ridge regression model with L2 regularization is used to construct a multiple linear prediction model between the gray-scale distribution characteristics of tight sandstone CT data and the porosity extraction threshold. This study conducted NMR and 3D CT scanning experiments on 11 tight sandstone cores. The experimental results show that the proposed method can accurately predict the porosity extraction threshold of tight sandstone 3D CT data, providing technical support for the accurate evaluation of tight sandstone reservoirs.

## 2 Methods and principles

### 2.1 Nuclear magnetic resonance (NMR) technology

Three-dimensional core CT scanning technology uses continuous two-dimensional tomographic scanning of core samples and three-dimensional reconstruction algorithms to achieve three-dimensional visualization and quantitative analysis of information such as pore network and mineral distribution in the core. During core CT scanning, an X-ray beam with an initial intensity of penetrates the core sample along the path L, and its outgoing intensity follows the Beer-Lambert law (Equation 1). Components with higher density (such as rock skeletons) have a higher absorption degree of X-rays and a larger attenuation coefficient of X-rays, while low-density components such as pores and cracks have a lower absorption degree of X-rays and a smaller attenuation coefficient of X-rays (Juhos, 2022).

$$I = I_0 \exp\left(-\int_L \mu(x, y, z)\, dl\right) \tag{1}$$

Where $\mu$ is the linear attenuation coefficient, $I_0$ is the intensity of the incident X-ray signal emitted by the X-ray source, and $I$ is the intensity of the emitted X-ray received by the detector. The detector receives the X-rays that penetrate the core sample to obtain its two-dimensional projection image $P_\theta$. By acquiring the two-dimensional projection dataset $\{P_\theta(s)\}$ of the core sample at different angles $\theta$ , and using algorithms such as computer filtered back projection (FBP), a three-dimensional model $\mu(x, y, z)$ is reconstructed.

In 3D core CT scan data, the gray value of each voxel is proportional to the material density at its corresponding location. Voxels with high gray values correspond to high-density components such as rock skeletons and metallic minerals, while voxels with low gray values correspond to rock pores. Therefore, the porosity of 3D core CT data can be extracted using the gray-scale thresholding method (Equation 2).

$$g(i, j, k) = \begin{cases} 1, & f(i, j, k) \geq T \\ 0, & f(i, j, k) < T \end{cases} \tag{2}$$

Where $f(i, j, k)$ represents the gray value of the voxel at spatial location $(i, j, k)$ , and $g(i, j, k)$ is the component identification result after gray-scale thresholding (Liu et al., 2022).

Core NMR scanning technology calculates core porosity by measuring the difference in magnetic resonance signals between dried and saturated cores containing hydrogen fluid (Equation 3). Core NMR scanning technology is an important means of accurately measuring the core porosity of tight sandstone reservoirs. In this study, the porosity obtained by core NMR measurement is used as a reference for calibrating the porosity extraction threshold of core CT data.

$$\varphi_m = \int_{T_{2\min}}^{T_{2\max}} f(T_2)\, dT_2 \tag{3}$$

In the formula, $\varphi_m$ represents nuclear magnetic porosity, $T_{2min}$ represents the start time of the T2 distribution spectrum, $T_{2max}$ represents the end time of the T2 distribution spectrum, and $f(T_2)$ represents the T2 distribution spectrum obtained by inversion through the CPMG echo attenuation curve (Kenyon et al., 1988).

## 2.2 Calculate the grayscale distribution characteristics of core CT data

The signals acquired by 3D core CT technology follow a power-law distribution. Statistical analysis methods can be used to deeply explore the differences in gray-level distribution between pores and other components within the core. Therefore, this study selects five statistical features to characterize the relationship between the pore extraction threshold and the gray-level distribution. Their definitions and calculation formulas are shown in the table below:

The calculation process for the grayscale distribution characteristics of 3D core CT data of tight sandstone is as follows (Fig.2): First, the 3D core CT data is read, and the grayscale values of the voxels are divided into several intervals according to the data storage format. If the data is stored in 8-bit binary format, it is divided into 255 intervals; if the data is stored in 16-bit binary format, it is divided into 65,535 intervals. The latter method was used in the experimental data analysis process of this study. Then, the range of grayscale value of each voxel in the 3D data volume is determined, and the number of voxels in each grayscale range is counted. Finally, the statistical characteristics of the core grayscale distribution curve are calculated according to the formula, and the StandardScaler method (mean 0, standard deviation 1) is used to standardize all grayscale distribution characteristics to eliminate dimensional differences between grayscale distribution characteristics, providing a data basis for subsequent pore extraction threshold.

## 2.3 Prediction of pore extraction threshold for core CT data

### 2.3.1 Pore extraction threshold calibration method constrained by NMR porosity

To obtain the pore extraction thresholds for model training, this study used the porosity measured by nuclear magnetic resonance (NMR) to constrain and calibrate the grayscale segmentation threshold of three-dimensional core CT data. For any given grayscale threshold T, voxels with grayscale values less than or equal to the threshold were identified as pores, whereas voxels with grayscale values greater than the threshold were identified as the rock skeleton or mineral components. The binary segmentation function can be expressed as:

$$M_T(x, y, z) = \begin{cases} 1, & G(x, y, z) \geq T \\ 0, & G(x, y, z) < T \end{cases} \tag{4}$$

where $G(x, y, z)$ represents the grayscale value of the voxel at the spatial position $(x, y, z)$, and $M_T(x, y, z)$ represents the pore identification result under the threshold $T \cdot MT = 1$ denotes a pore voxel, whereas $MT = 0$ denotes a non-pore voxel.

Under the threshold T, the porosity calculated from the core CT data is given by:

$$\phi_{CT}(T) = \frac{N_p(T)}{N_{ROI}} \times 100\% \tag{5}$$

where $N_p(T)$ represents the number of voxels identified as pores under the threshold T, and $N_{ROI}$ represents the total number of voxels in the region of interest of the core involved in the calculation. Since the CT data used in this study were stored as 16-bit unsigned integers, the grayscale values ranged from 0 to 65535. Therefore, the threshold was searched within the effective grayscale range of the core data, and the corresponding $\phi_{CT}(T)$ values under different thresholds were calculated. Taking the porosity measured by NMR, $\phi_{NMR}$, as the constraint reference, the threshold that minimized the difference between the two porosity values was selected as the calibrated threshold of the sample:

$$T_{cal} = \mathrm{argmin}_T |\phi_{CT}(T) - \phi_{NMR}| \tag{6}$$

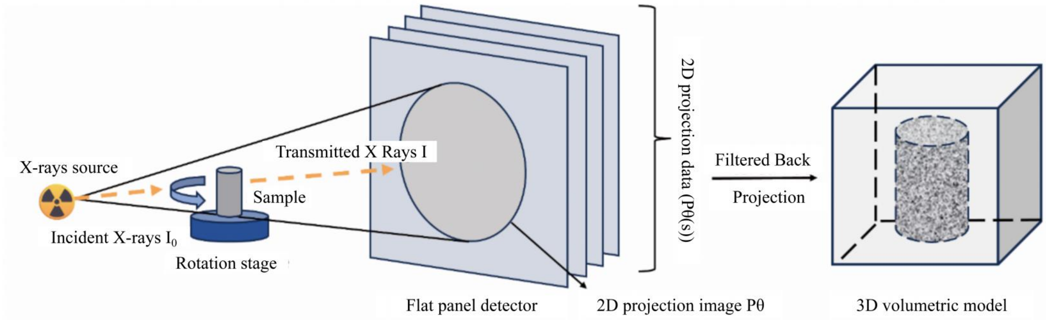


**Fig. 1** Schematic diagram of CT scan

**Tab. 1** Statistical characteristics and significance of grayscale images

| Formula | Description | Ref. |
|---|---|---|
| $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ | $x_i$: gray value of the $i$-th pixel; $N$: total number of pixels | (Pearson, 1893) |
| $M = \begin{cases} x_{\left(\frac{N+1}{2}\right)} \\ \dfrac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2} \end{cases}$ | $x_i$: gray value at the $i$-th position in ascending order; $N$: total number of pixels | (Pearson, 1894) |
| $\text{Mode} = \arg\max_x f(x)$ | $x$: gray value; $f(x)$: frequency of gray value $x$ in the image | (Pearson, 1894) |
| $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$ | $x_i$: gray value of the $i$-th pixel; $\mu$: mean gray value; $N$: total number of pixels | (Pearson, 1893) |
| $\beta = \dfrac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right)^2}$ | $\beta$: Kurtosis; $x_i$: gray value of the $i$-th pixel; $\mu$: mean gray value; $N$: total number of pixels | (Pearson, 1905) |

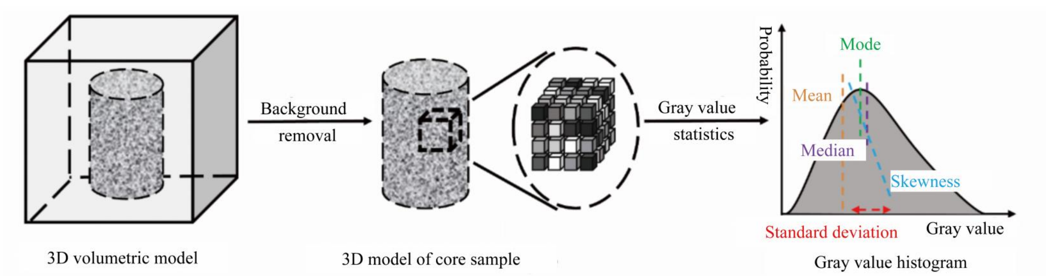


**Fig. 2** Schematic diagram of the calculation process of grayscale distribution characteristics of core CT data

where $T_{cal}$ represents the calibrated CT pore extraction threshold obtained under the constraint of NMR porosity. The essence of this process is to use the volume fraction of fluid-fillable pores measured by NMR to constrain the volume fraction obtained from CT grayscale threshold segmentation, so that the CT pore extraction result has a clear experimental measurement basis. Subsequently, the $T_{cal}$ corresponding to each core sample was used as the prediction target of the ridge regression model to establish a quantitative prediction relationship between grayscale distribution features and pore extraction thresholds.

Although both nuclear magnetic resonance (NMR) and X-ray computed tomography (CT) can be used to characterize rock pore structure, they correspond to different pore scales. The CT scan used in this paper has a resolution of 9.94 μm; therefore, CT images primarily reflect pores larger than the voxel scale, while pores smaller than the micrometer scale are difficult to directly identify. In contrast, NMR testing can characterize pore space over a wider scale range and is more suitable for reflecting the overall porosity development of rock cores. Therefore, NMR porosity is not a simple substitute for CT-visible porosity, but rather serves as a macroscopic pore constraint to guide the reasonable determination of CT thresholds.

This paper does not directly equate NMR-measured porosity with CT-extracted porosity, but rather uses it as a reference target in the threshold selection process: visible pores in CT images are extracted under different grayscale segmentation thresholds, and the consistency between the pore characterization results and NMR porosity is compared. Since CT cannot resolve micropores below the voxel scale, the goal of threshold selection is not to force both methods to obtain identical porosity values, but rather to find a grayscale threshold that most reasonably characterizes the degree of porosity development at the sample scale while minimizing segmentation bias. In other words, NMR porosity acts as a "constraint" and "calibration reference," while CT provides spatial distribution information of the pore space.

This method avoids simply equating porosity results at different testing scales. Instead, it leverages the constraint of NMR on the overall porosity development and the spatial resolution of CT on the pore space distribution to achieve information complementarity between the two testing methods. In this way, while fully considering the limitations of CT resolution, the objectivity and reliability of the porosity extraction threshold determination can be improved, thereby enhancing the stability and interpretability of subsequent digital core characterization results.

### 2.3.2 Pore extraction threshold prediction method

Since the grayscale distribution curve of CT data from tight sandstone reservoir cores exhibits a unimodal distribution, the traditional "bimodal method" cannot be used to select the porosity extraction threshold from 3D CT data (Fig.3). Therefore, this study utilizes the grayscale distribution characteristics of CT data from tight sandstone cores to establish a prediction model, and constructs a multiple regression prediction model for the porosity extraction threshold based on the statistical characteristics of its grayscale distribution curve (Formula 7).

$$T_{cal} = \beta_0 + \beta_1\tilde{x}_1 + \beta_2\tilde{x}_2 + \beta_3\tilde{x}_3 + \beta_4\tilde{x}_4 + \beta_5\tilde{x}_5 + \varepsilon \tag{7}$$

where $T_{cal}$ represents the calibrated CT pore extraction threshold obtained under the constraint of NMR porosity. This calibrated threshold was used as the prediction target of the subsequent ridge regression model to establish the quantitative relationship between grayscale distribution features and the pore extraction threshold. $\beta_0$ is the intercept term, $\beta_1$ to $\beta_5$ are the regression coefficients corresponding to the five grayscale statistical features, $\tilde{x}_1 to \tilde{x}_5$ represent the standardized grayscale mean, median, mode, standard deviation, and kurtosis, respectively, and ε represents the random error term.

To eliminate the influence of differences in dimensions and numerical ranges among different grayscale statistical features on the estimation of model coefficients, the StandardScaler method was used to standardize each grayscale distribution feature. For the j-th grayscale statistical feature of the i-th core sample, the standardization formula is expressed as:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{8}$$

where $\tilde{x}_{\mathrm{ij}}$ represents the original value of the j-th grayscale statistical feature of the i-th core sample, μj and σj represent the mean and standard deviation of the j-th grayscale statistical feature in the sample set, respectively, and $x_{ij}$ represents the standardized value of the corresponding grayscale statistical feature. The input variables used in the subsequent model were all standardized grayscale statistical features.

In ridge regression, the regression coefficients are estimated by minimizing the residual sum of squares with an L2 regularization term. The objective function can be expressed as:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \left[ \sum_{i=1}^{n} \left( T_{cal,i} - \beta_0 - \sum_{j=1}^{p} \beta_j \tilde{x}_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right] \tag{9}$$

where $\hat{\beta}$ represents the regression coefficient vector finally estimated by the ridge regression model;arg min denotes the process of searching for the set of regression coefficients $\beta$ that minimizes the objective function; $n$ represents the number of core samples;$p$ represents the number of grayscale statistical features. In this study, $p = 5$, corresponding to the grayscale mean, median, mode, standard deviation, and kurtosis. $T_{\mathrm{cal},i}$ represents the calibrated pore extraction threshold of the $i$-th core sample obtained under the constraint of NMR porosity; $\beta_0$ represents the intercept term of the model;$\beta_j$ represents the regression coefficient corresponding to the $j$-th grayscale statistical feature; $\tilde{x}_{\mathrm{ij}}$ represents the standardized value of the $j$-th grayscale statistical feature of the $i$-th core sample;and $\lambda$ represents the hyperparameter controlling the strength of $L_2$ regularization, which corresponds to the alpha value in the model calculation process.

The flowchart of the porosity extraction threshold prediction method for tight sandstone proposed in this study is shown in Fig.4. In the figure, the red rounded rectangles represent the data in the method, and the green rectangles represent the processing steps. The steps for establishing the porosity extraction threshold prediction method for tight sandstone CT data are as follows:

Step 1: Select a tight sandstone core, dry it, and then use a nuclear magnetic resonance relaxation scanner to measure the CPMG echo signal of the core in the dry state. Acquire CT data of the core using a 3D CT scanner. Next, saturate the core with fluid (vacuum pressurization), and measure the CPMG echo signal of the core again. Use the CPMG signal obtained from NMR to invert the T2 distribution spectrum of the core in both dry and saturated states, and calculate the core porosity.

Step 2: Statistically analyze the grayscale distribution curve of the tight sandstone core CT data, calculate the eigenvalues of the grayscale distribution curve, and use StandardScaler to standardize all initial eigenvalues, transforming them into a distribution interval with a mean of 0 and a standard deviation of 1, eliminating the influence of dimensional differences on the prediction method, and obtaining the grayscale feature vector X of the core CT data.

Step 3: The porosity extraction threshold Y of the core CT data is calibrated using porosity measured by nuclear magnetic resonance. Then, the coefficient $\beta$ and error term $\epsilon$ of the grayscale distribution characteristics are calculated using an L2-regularized ridge regression model. A 5-fold cross-validation strategy is employed during the calculation to obtain the optimal regularization parameter $\alpha$ . Finally, a method for predicting the porosity extraction threshold of tight sandstone core CT data is established based on the results.

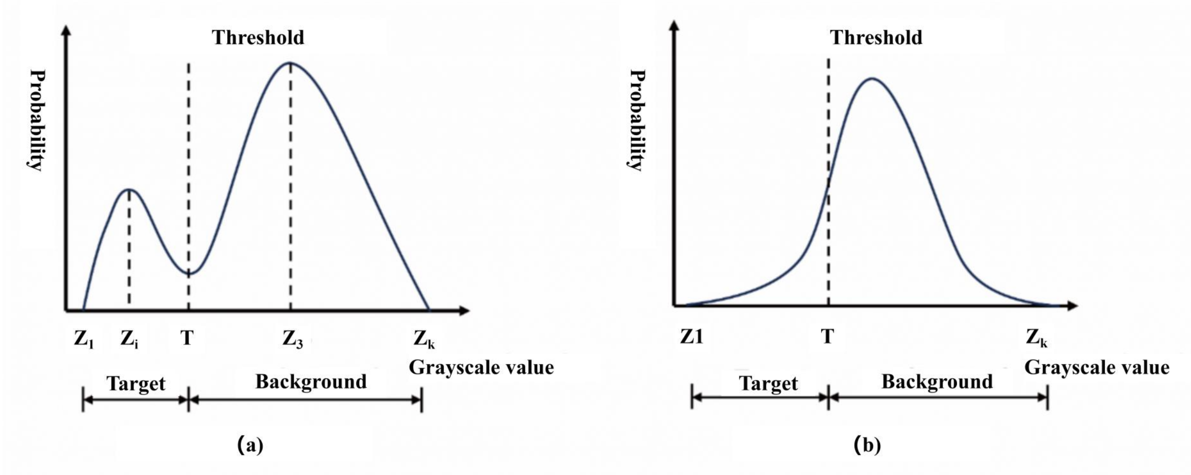


**Fig. 3** Schematic diagram comparing grayscale histograms of CT images of conventional rock samples and tight sandstone

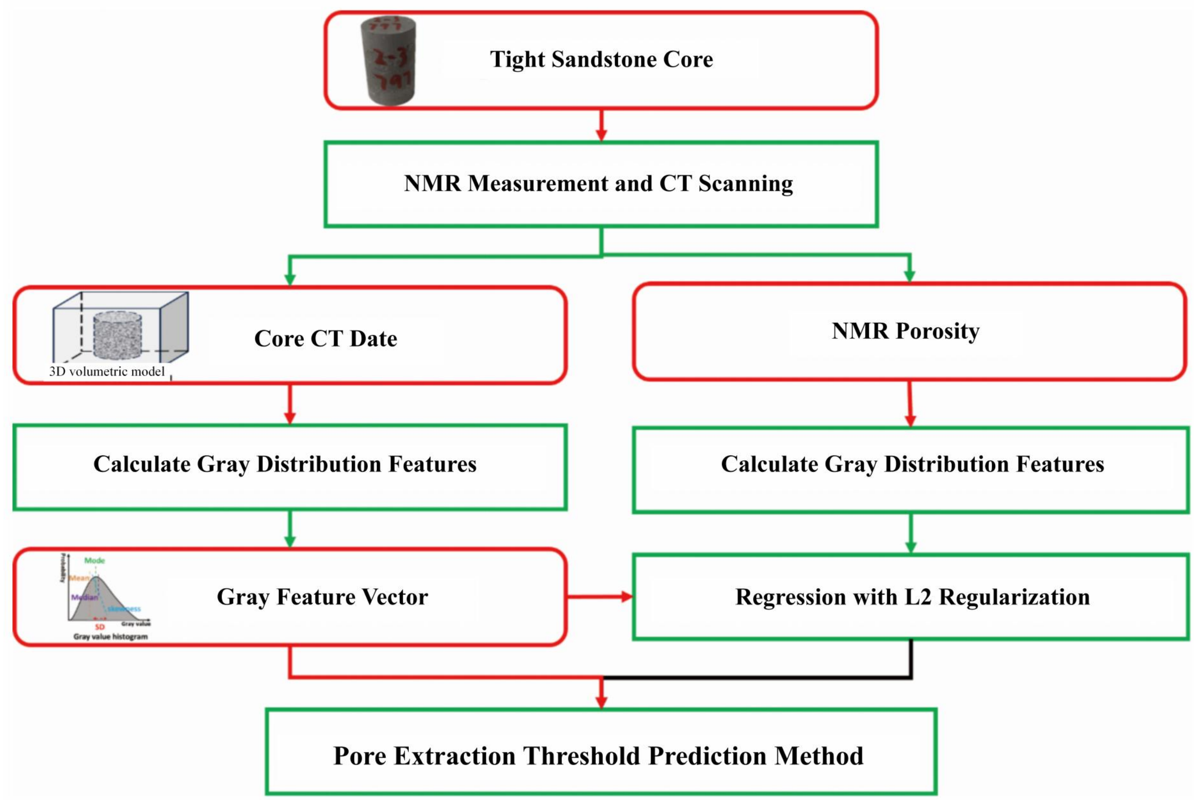


**Fig. 4** Flowchart of the construction and validation of a threshold prediction model based on ridge regression

## 3 Experiments and results

### 3.1 Experimental data and acquisition parameters

Eleven tight sandstone cores were selected for experiments in this study (Fig.5). Both nuclear magnetic resonance (NMR) scanning and CT scanning were performed on the cores. The NMR acquisition parameters included a magnetic field frequency of 21 MHz, a measurement temperature of 32 °C, waiting time of 4000 ms, an echo interval of 0.08 ms, 10000 echoes, and 32 repeated scans to ensure a signal-to-noise ratio exceeding 200. The CT scanning was conducted in spiral mode with a voltage of 140 kV, a current of 200 $\mu$A, an exposure time of 1.5 s per frame, an image merging number of 2, a frame rate of 720 FPS, and a spatial resolution of 9.94 $\mu$m.

The tight sandstone cores used in the experiment were from reservoirs at different depths within the same block, including both homogeneous and heterogeneous reservoir samples. All cores had a porosity of less than 6%, representing typical tight sandstone reservoir core samples and serving to verify the accu-


**Fig. 5** Tight sandstone core sample

racy of the CT porosity extraction threshold prediction method for tight sandstone reservoir cores.

The tight sandstone cores used in the experiment were from reservoirs at different depths within the same block. The apparent porosity measured by nuclear magnetic resonance ranged from 0.88% to 5.55%, including both relatively homogeneous and highly heterogeneous samples. All samples are typical tight sandstone reservoir core samples, effectively validating the accuracy of the CT porosity extraction threshold prediction method for tight sandstone reservoir cores. The basic parameters of the sample are shown in the table below.

During the experiment, core nuclear magnetic resonance testing and three-dimensional core CT scanning experiments were conducted. Nuclear magnetic resonance (NMR) testing was performed using a MesoMR23-060H-I NMR core analyzer (Suzhou Newmai Analytical Technology Co., Ltd.), with a magnetic field frequency of 21 MHz and a measurement temperature of 32 °C. Specific acquisition parameters were set as follows: waiting time 4000 ms, echo interval 0.08 ms, 10000 echoes, and 32 repeated scans to ensure a signal-to-noise ratio (SNR) higher than 200. NMR $T_2$ spectra were measured for the same batch of core samples under both saturated and dried conditions. The $T_2$ distribution spectrum of each sample was obtained by inverting the acquired signals. Subsequently, the $T_2$ spectrum signals were integrated, and the apparent porosity of the core samples was calculated based on the calibration results of standard samples. Test results showed that the apparent porosity of each sample ranged from 0.88% to 5.55%.

Three-dimensional core CT scanning experiments were performed using a nanoVoxel-3000 high-resolution X-ray CT scanning system in spiral CT mode. The scanning parameters were set as follows: scanning voltage of 140 kV, current of 200 $\mu$A, single-frame exposure time of 1.5 s, image merging number of 2, scan frame count of 720, and source-to-detector distance (SDD) of 1016.22 mm. After scanning, the acquired projection data were reconstructed in three dimensions using Avizo software to generate a digital core model for subsequent analysis. The reconstructed three-dimensional grayscale image had a voxel resolution of 9.94 $\mu$m, and the CT data were stored in 16-bit unsigned integer format with grayscale values ranging from 0 to 65535.

To reduce the impact of core boundary effects, clamping devices, and reconstruction artifacts on grayscale statistical results, an effective core area was selected as the region of interest (ROI) within each core sample. Grayscale statistical feature extraction, porosity extraction threshold calibration, and porosity calculation were all performed within a unified ROI. This processing method effectively reduces the interference of edge noise and non-rock areas on statistical results, thereby improving the comparability of grayscale feature analysis and threshold prediction results across different samples.

## 3.2 Experimental results

Following the establishment process of the porosity extraction threshold prediction method for core CT data of tight sandstone reservoirs proposed in this study, the three-dimensional CT data and pore structure of the sample were extracted during the experiment (Fig.6). The porosity extraction threshold of the core CT data was obtained based on porosity calibration using nuclear magnetic resonance (NMR) measurements, and the specific values are shown in Table 3. This process constrains the porosity extraction process of the core CT data, avoiding the subjectivity of the porosity extraction threshold caused by human factors. As can be seen from the table, the porosity of

**Tab. 2** Core Sample Parameters

| Core Sample Number | Length (cm) | Diameter (cm) | Volume (ml) | Dry weight (g) |
|---|---|---|---|---|
| 1 | 4.97 | 2.44 | 23.20 | 62.00 |
| 2 | 4.33 | 2.44 | 20.33 | 52.31 |
| 3 | 4.96 | 2.44 | 23.23 | 62.12 |
| 4 | 4.97 | 2.44 | 23.24 | 61.67 |
| 5 | 4.15 | 2.45 | 19.61 | 51.81 |
| 6 | 4.72 | 2.45 | 22.20 | 58.62 |
| 7 | 4.38 | 2.43 | 20.30 | 55.77 |
| 8 | 4.76 | 2.44 | 22.19 | 59.02 |
| 9 | 4.67 | 2.44 | 21.81 | 57.71 |
| 10 | 3.41 | 2.38 | 15.26 | 39.43 |
| 11 | 5.14 | 2.39 | 23.00 | 59.37 |


**Fig. 6** Three-dimensional CT data and pore structure of tight sandstone core samples

the CT data obtained using these thresholds is highly consistent with the porosity obtained by NMR measurements.

To establish a predictive model between the gray-scale distribution characteristics of CT data from tight sandstone cores and the porosity extraction threshold, the characteristic values of the gray-scale distribution curves of CT data from 11 core samples were calculated. Fig.7 shows the gray-scale distribution information of the core samples, and Table 4 shows the results of the

**Tab. 3** Pore extraction threshold calibrated by NMR measurement of tight sandstone core samples

| Core number | Nuclear magnetic resonance porosity (%) | Pore extraction threshold | CT porosity (%) |
|---|---|---|---|
| 1 | 0.88 | 36100 | 0.87 |
| 2 | 5.48 | 41200 | 5.42 |
| 3 | 3.57 | 36750 | 3.57 |
| 4 | 1.66 | 29550 | 1.69 |
| 5 | 1.40 | 28750 | 1.38 |
| 6 | 2.18 | 33250 | 2.21 |
| 7 | 1.36 | 32850 | 1.34 |
| 8 | 2.23 | 38500 | 2.23 |
| 9 | 2.91 | 29600 | 2.95 |
| 10 | 3.40 | 34900 | 3.41 |
| 11 | 5.55 | 38350 | 5.54 |

gray-scale distribution characteristic values of the core samples. As can be seen from Fig.7, the gray-scale distribution curves of the CT data from tight sandstone reservoir cores all exhibit a single-peak distribution, and the peak value, position, and width of the distribution curves show significant differences. Table 4 shows that the gray-scale distribution characteristic values of CT data from different tight sandstone cores are also different, and there is no obvious pattern. This makes it significant to construct a multivariate predictive model between the gray-scale distribution characteristics of tight sandstone cores and the porosity extraction threshold.

Based on the establishment process of the porosity extraction threshold prediction method for CT data of tight sandstone cores proposed in this study, the multiple regression coefficients between the core grayscale distribution characteristics and the porosity extraction threshold were calculated using an L2-regularized ridge regression model (Table 5). The porosity extraction threshold and porosity of the tight sandstone core CT data established using these coefficients are shown in Table 6. Experiments revealed a strong correlation between the porosity extraction threshold predicted by the proposed method and the threshold determined under NMR constraints (Fig.8), with a correlation coefficient exceeding 0.99. Furthermore, the error between the predicted porosity extraction threshold and the threshold calibrated by NMR measurement is very small. This indicates that the prediction accuracy of the porosity extraction threshold for tight sandstone core CT data proposed in this study is very high. The porosity obtained from the core CT data using the predicted porosity extraction threshold is highly consistent with the porosity measured by NMR, with a correlation coefficient reaching 0.89. Moreover, the absolute error between the predicted core porosity and the porosity measured by NMR is very small.

## 4 Discussion

### 4.1 Relationship between grayscale distribution characteristics and pore extraction threshold

This study calculated the mean, median, mode, standard deviation, and kurtosis of the grayscale distribution curves of 11 tight sandstone core samples, and calculated the correlation between these grayscale distribution characteristics and the pore extraction threshold (Fig. 9). The results showed that there was a strong positive correlation between the gray mean, median, and mode and the pore extraction threshold, with correlation coefficients all greater than 0.94. The correlation coefficient between kurtosis and the pore extraction threshold was 0.82, and the correlation coefficient between standard deviation and the pore extraction threshold was 0.45.

From a physical perspective, the mean, median, and mode of grayscale reflect the concentration level of the component density of the core sample. If the tight sandstone reservoir is located at a greater depth, the rock skeleton density is higher, the corresponding gray value is higher, and the gray value distribution is more concentrated. If the tight sandstone reservoir has a high degree of porosity development, the proportion of low gray voxels will increase, the gray value concentration trend will shift in the negative direction, and the porosity extraction threshold will decrease. Therefore, these three characteristics are strongly positively correlated with the porosity extraction threshold of tight sandstone cores, reflecting the controlling effect of the overall compactness of the core on the location of the porosity extraction threshold.

Kurtosis reflects the steepness of the gray-scale distribution curve. The grayscale distribution of tight sandstone usually exhibits a single-peak shape. The higher the kurtosis value, the more concentrated the grayscale value is in a certain range, and the more homogeneous the lithological components are. The lower the kurtosis value, the more gradual the grayscale distribution, indicating the presence of more transitional components (such as microcracks, transitional pores, high-density minerals, etc.) in the core. The presence of these components will affect the clarity of the boundary between pores and the framework, and thus affect the position of the pore extraction threshold.

Standard deviation characterizes the dispersion of grayscale distribution, reflecting the complexity and heterogeneity of core components (including rock skeleton, pores, clay minerals, high-density minerals, etc.). Tight sandstone reservoirs

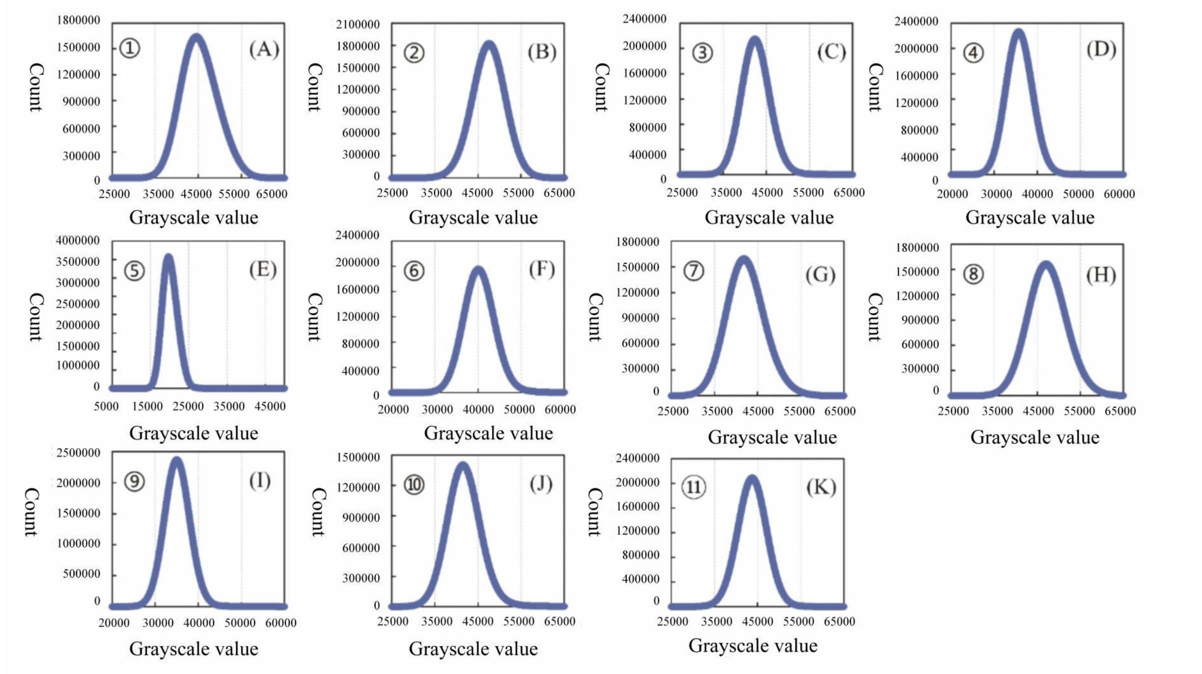


**Fig. 7** Gray-scale distribution curve of CT data of tight sandstone core samples

**Tab. 4** Gray-scale distribution characteristics of tight sandstone core samples

| Sample number | Mean | Median | Mode | Standard deviation | Kurtosis |
|---|---|---|---|---|---|
| 1 | 45316.17 | 45099 | 44602 | 4245.26 | -0.08 |
| 2 | 47485.36 | 47517 | 47549 | 3900.10 | 0.19 |
| 3 | 42562.86 | 42458 | 42294 | 3403.66 | 1.64 |
| 4 | 35901.08 | 35808 | 35727 | 3200.83 | 2.97 |
| 5 | 35666.07 | 35558 | 35357 | 3385.15 | 2.87 |
| 6 | 40347.09 | 40231 | 40089 | 3737.09 | 1.50 |
| 7 | 42296.05 | 42120 | 41872 | 4517.95 | 0.27 |
| 8 | 47325.41 | 47224 | 47014 | 4523.30 | 0.10 |
| 9 | 35152.43 | 35085 | 35048 | 3099.83 | 3.89 |
| 10 | 41807.32 | 41669 | 41489 | 4006.40 | 1.21 |
| 11 | 43733.51 | 43732 | 43810 | 3389.51 | 0.28 |

**Tab. 5** Multiple regression coefficients of pore extraction threshold prediction method

| Model Term | Mean Coefficient | Median Coefficient | Mode Coefficient | Standard Deviation Coefficient | Kurtosis Coefficient | Residual |
|---|---|---|---|---|---|---|
| Value | -523.21 | 429.06 | 5040.78 | -1361.95 | 287.19 | 34527.27 |

typically have diverse mineral compositions and complex pore structures. The larger the standard deviation of the grayscale distribution, the more significant the grayscale differences between different components in the core, the clearer the boundary between pores and the framework, and the easier it is to determine the pore extraction threshold. Conversely, the smaller the standard deviation, the more concentrated the grayscale distribution, the more blurred the grayscale transition between pores and the framework, and the more difficult it is to select the threshold. Although the correlation coefficient between the pore extraction threshold and the standard deviation of the sample in this experiment was only 0.45, the standard deviation characteristic can reflect the heterogeneity of the tight sandstone core and help to accurately predict the pore extraction threshold of the tight sandstone core.

Experiments revealed that the mean, median, mode, kurtosis, and standard deviation of the gray-scale distribution curves of CT data from tight sandstone cores characterized the physi-

**Tab. 6** Pore extraction threshold prediction results information

| Core number | Nuclear magnetic resonance porosity (%) | Calibrated threshold | Predicted threshold | Relative error of threshold | Predicted CT porosity (%) | Absolute error of porosity |
|---|---|---|---|---|---|---|
| 1 | 0.88 | 36100 | 36651 | 1.53% | 1.32 | 0.44% |
| 2 | 5.48 | 41200 | 41156 | 0.11% | 5.31 | 0.17% |
| 3 | 3.57 | 36750 | 36698 | 0.14% | 3.44 | 0.13% |
| 4 | 1.66 | 29550 | 29888 | 1.14% | 2.26 | 0.60% |
| 5 | 1.40 | 28750 | 28914 | 0.57% | 1.59 | 0.19% |
| 6 | 2.18 | 33250 | 33162 | 0.27% | 2.07 | 0.11% |
| 7 | 1.36 | 32850 | 32787 | 0.19% | 1.29 | 0.07% |
| 8 | 2.23 | 38500 | 38754 | 0.66% | 2.29 | 0.06% |
| 9 | 2.91 | 29600 | 29582 | 0.06% | 2.91 | 0.00% |
| 10 | 3.40 | 34900 | 33980 | 2.64% | 1.91 | 1.49% |
| 11 | 5.55 | 38350 | 38229 | 0.31% | 5.17 | 0.38% |

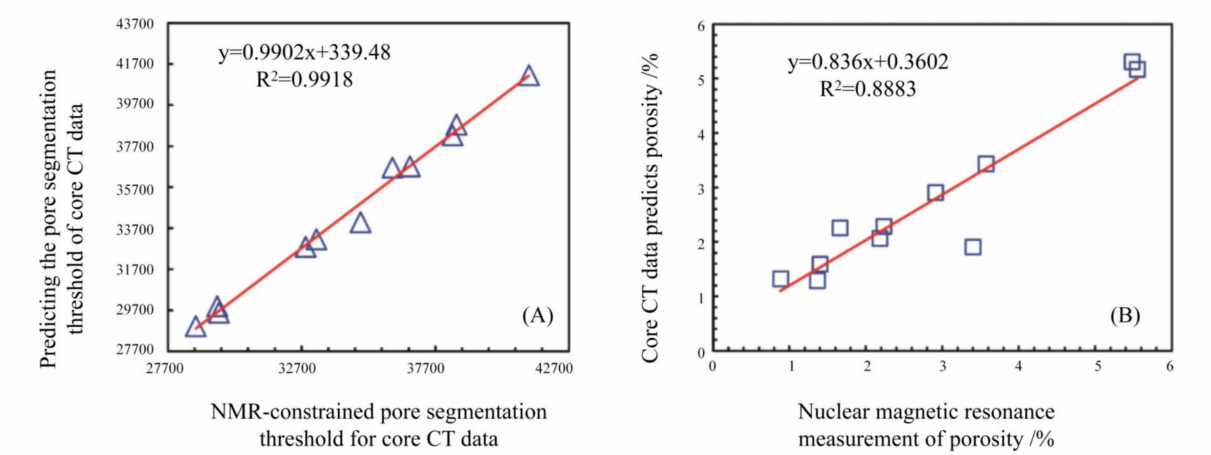


**Fig. 8** Predicted results of porosity extraction threshold (A) and porosity results (B) from core CT data

cal properties of the cores from different dimensions, such as the overall density level, gray-scale distribution pattern, and component heterogeneity, and showed a significant statistical correlation with the pore extraction threshold. These features together constitute the input variables of the prediction model, providing a theoretical and physical basis for establishing a multiple regression prediction model.

## 4.2 Model performance

This study evaluated the performance of the porosity extraction threshold prediction method for CT data of tight sandstone cores from three aspects: prediction accuracy, generalization ability, and stability. The results show that the method proposed in this study has good performance in all performance indicators.

Regarding prediction accuracy, the proposed method achieved a training set determination coefficient $R^2$ of 0.984, indicating strong fitting ability. The model's mean absolute percentage error (MAPE) was 0.69%, and the mean absolute error (MAE) was 237.61. Considering the large numerical range of core CT data and a pore extraction threshold of around 30,000, the absolute error rate of this prediction result is less than 1%.

In terms of generalization ability, the cross-validation determination coefficient (CV-$R^2$) obtained by the 5-fold cross-validation method in the experiment was 0.858, and the cross-validation root mean square error (CV-RMSE) was 383.96. This shows that the method proposed in this study maintains stable prediction performance on unknown data.

From a stability perspective, the method proposed in this study demonstrates good stability. The Durbin-Watson statistic for the predicted results is 1.606, indicating that there is no autocorrelation problem in the residuals.

## 4.3 Impact of CT scan resolution on threshold prediction method

In this study, the spatial resolution of the core CT scan was 9.94 μm. This resolution can identify some micron-sized pores, larger pore throats, and microcracks, but it is difficult to directly distinguish the widely developed submicron and nano-sized pores in tight sandstone. Therefore, the pore extraction threshold prediction method proposed in this paper is mainly applicable to the extraction of identifiable pore space under the current CT scan resolution, and cannot fully characterize the

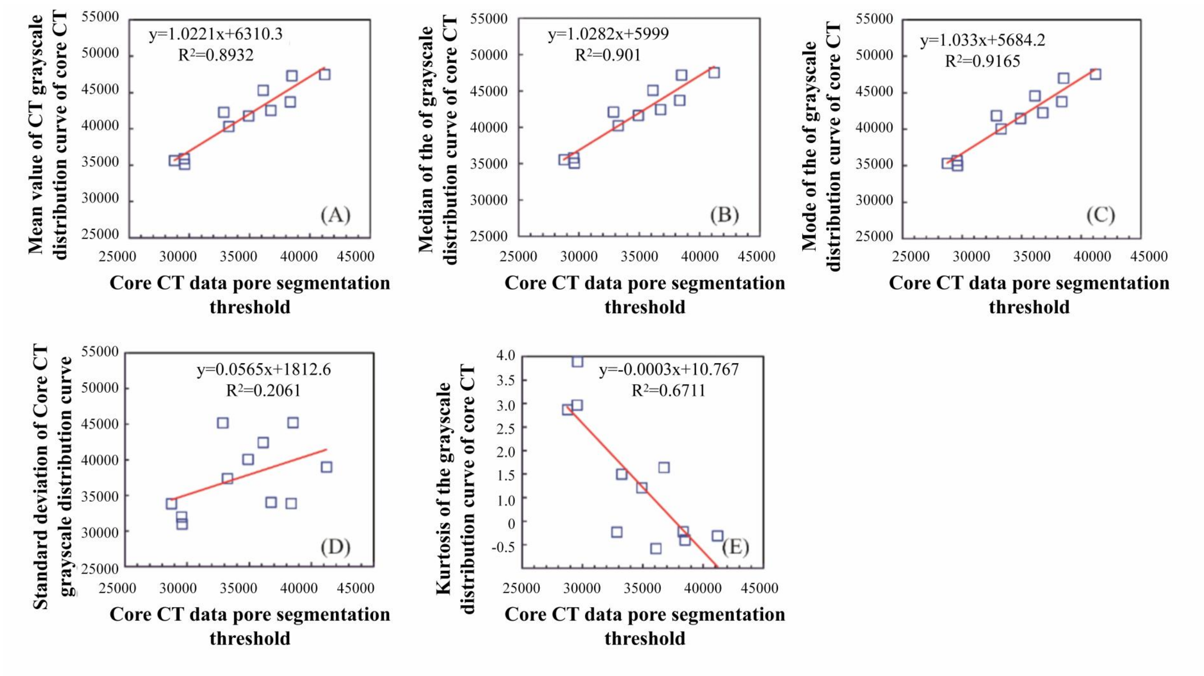


**Fig. 9** Correlation between gray level distribution characteristics and pore extraction threshold

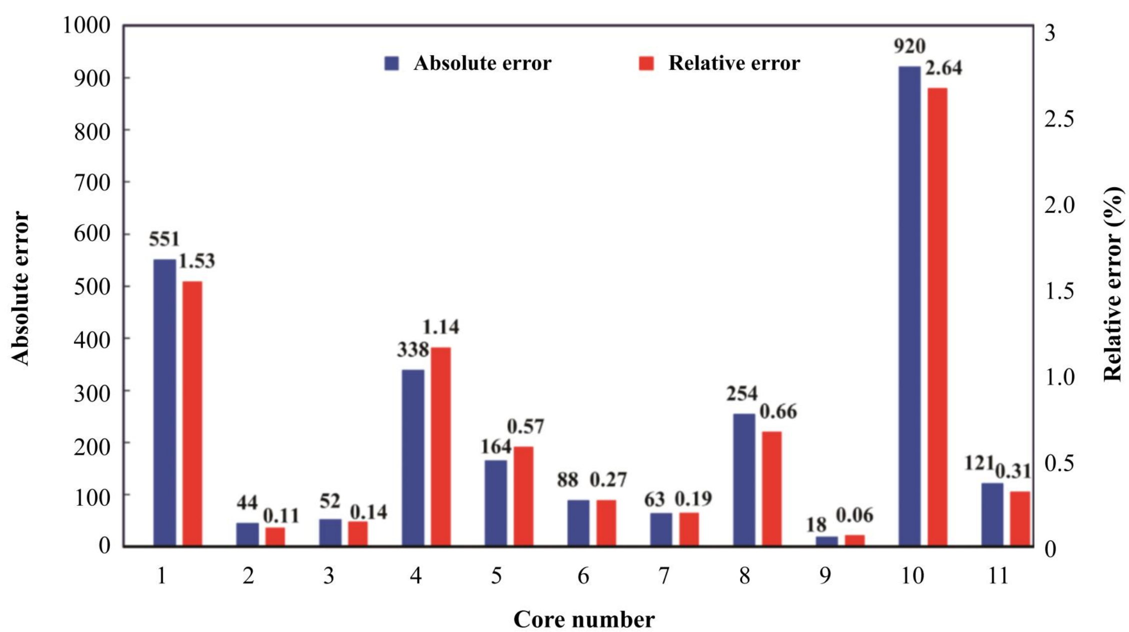


**Fig. 10** Distribution of pore extraction threshold error in CT data of tight sandstone core

full-scale pore system of tight sandstone. For micropores below the CT resolution, their volume fraction may be reflected by NMR signals, but cannot be directly displayed in CT images, which will cause a scale difference between NMR porosity and CT-identifiable porosity. Future research can combine multi-scale data such as micron-CT, nano-CT, scanning electron microscopy, mercury intrusion porosimetry, and NMR $T_2$ spectroscopy to establish a graded threshold prediction method under different resolution conditions, so as to improve the completeness of the multi-scale pore structure characterization of tight sandstone.

## 5 Conclusions

This study addresses the failure of traditional threshold-based pore extraction methods in CT data of tight sandstone cores by proposing a pore extraction threshold prediction method based on gray-level distribution characteristics. This method uses porosity obtained from NMR spectroscopy as a constraint to calibrate the CT pore extraction threshold. It extracts statistical features such as mean, median, mode, standard deviation, and kurtosis from the gray-level distribution curve of the core CT data. A ridge regression model is used to establish a predictive relationship between the gray-level distribution characteristics and the pore extraction threshold, thus achieving the prediction of the pore extraction threshold in CT data of tight sandstone cores. This method does not rely on the bimodal characteristics of the gray-level distribution, reducing the subjective error caused by manual threshold selection, and providing a stable and repeatable technique for pore extraction from CT data of tight sandstone cores.

## Data Availability

Data cannot be shared publicly because of confidentiality agreements. Data are available from the corresponding author upon reasonable request.

## Conflict of interest

The authors declare no competing interest.



## References

Arunpandian M, Arunprasath T, Vishnuvarthanan G, et al. 2018. Soil porosity analysis using combined maximum entropy and class variance thresholding. In: *Microelectronics, Electromagnetics and Telecommunications: Proceedings of the Fourth ICMEET 2018*. Singapore: Springer, 641–650. doi:10.1007/978-981-13-1906-8_65.

Chen CB, Zhao ZY, Fu L, et al. 2021. Interstitial matter and its impact on reservoir development in Chang 6 deepwater tight sandstone in Huaqing area, Ordos Basin. *Oil & Gas Geology*, **42**: 1098–1111. doi:10.11743/ogg20210508.

Deng YW, Yin SL, Li BK, et al. 2024. Fine characterization and evaluation of pore structure in tight sandstone reservoirs based on digital core. *Geological Review*, **70**: 1377–1390. doi:10.16509/j.georeview.2024.01.011.

Desbois G, Urai JL, Hemes S, et al. 2016. Multi-scale analysis of porosity in diagenetically altered reservoir sandstone from the Permian Rotliegend, Germany. *Journal of Petroleum Science and Engineering*, **140**: 128–148. doi:10.1016/j.petrol.2016.01.019.

Feng W, Chen XW, Zhao HF, et al. 2025. Computed tomography and image processing based multi-scale fracture extraction method for core samples. *Petroleum Geology & Experiment*, **47**: 882–894. doi:10.11781/sysydz2025040882.

Hao CG, Guo XY, Deng CB, et al. 2023. Precise identification and threshold inversion of pores and fissures in CT digital coal rock based on Bi-PTI model. *Journal of China Coal Society*, **48**: 1516–1526. doi:10.13225/j.cnki.jccs.2022.0602.

Hayatdavoudi M, Niri ME, Kalhor A. 2025. Comparative analysis of sandstone microtomographic image segmentation using advanced convolutional neural networks with pixelwise and physical accuracy evaluation. *Scientific Reports*, **15**: 22164. doi:10.1038/s41598-025-07211-2.

He TP, Zhou YQ, Zhang ZW, et al. 2025. Study on diagenesis and evolution of porosity in tight sandstone reservoirs. *Carbonates and Evaporites*, **40**: 132. doi:10.1007/s13146-025-01167-0.

Hu R, Han J, Wang C, et al. 2025. Multiscale pore/throat characterization in tight sandstone formation with multi-threshold image segmentation algorithm. *Frontiers in Earth Science*, **13**: 1576484. doi:10.3389/feart.2025.1576484.

Juhos A. 2022. CNN-based regularisation for CT image reconstructions. *arXiv preprint arXiv:2201.09132*. doi:10.48550/arXiv.2201.09132.

Kenyon WE, Day PI, Straley C, et al. 1988. A three-part study of NMR longitudinal relaxation properties of water-saturated sandstones. *SPE Formation Evaluation*, **3**: 622–636. doi:10.2118/15643-PA.

Li YY, Cui HQ, Zhang P, et al. 2020. Three-dimensional visualization and quantitative characterization of coal fracture dynamic evolution under uniaxial and triaxial compression based on $\mu$CT scanning. *Fuel*, **262**: 116568. doi:10.1016/j.fuel.2019.116568.

Liu HF, Ma CH, Zhu CQ. 2022. X-ray micro CT based characterization of pore-throat network for marine carbonates from South China Sea. *Applied Sciences*, **12**: 2611. doi:10.3390/app12052611.

Liu JF, Lin YJ, Wang ZP, et al. 2025. A Laplacian-based grayscale image segmentation algorithm for multi-modal pore characterization of geomaterials. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, **11**: 120. doi:10.1007/s40948-025-01004-7.

Pearson K. 1893. Asymmetrical frequency curves. *Nature*, **48**: 615–616. doi:10.1038/048615a0.

Pearson K. 1894. Contributions to the mathematical theory of evolution. *Proceedings of the Royal Society of London*, **54**: 329–333. doi:10.1098/rspl.1893.0079.

Pearson K. 1905. Skew variation, a rejoinder. *Biometrika*, **4**: 169–212.

Schlüter S, Sheppard A, Brown K, et al. 2014. Image processing of multiphase images obtained via X-ray microtomography: a review. *Water Resources Research*, **50**: 3615–3639. doi:10.1002/2014WR015256.

Wang H, Guo R, Dalton LE, et al. 2024. Comparative assessment of U-Net-based deep learning models for segmenting microfractures and pore spaces in digital rocks. *SPE Journal*, **29**: 5779–5791. doi:10.2118/215117-PA.

Wang H, Yang X, Zhou C, et al. 2025. Constructions of multi-scale 3D digital rocks by associated image segmentation method. *Frontiers in Earth Science*, **12**: 1518561. doi:10.3389/feart.2024.1518561.

Wang M, Song XH, Wang ZG, et al. 2024. Sandstone pore image segmentation based on kernel clustering. *Geophysical Prospecting for Petroleum*, **63**: 1051–1060. doi:10.12431/issn.1000-1441.2024.63.05.014.

Wang W, Chen CB, Xu S, et al. 2022. Fractal characteristics and its controlling factors of pore-throat with different scales in tight sandstones of the Yanchang Formation in the Ordos Basin. *Petroleum Geology & Experiment*, **44**: 33–40. doi:10.11781/sysydz202201033.

Wang XG, Huan JL, Peng XD, et al. 2023. Pore structure and seepage mechanism of tight sandstone reservoirs based on digital core. *Petroleum Geology and Recovery Efficiency*, **29**: 22–30. doi:10.13673/j.cnki.cn37-1359/te.202109036.

Wildenschild D, Sheppard AP. 2013. X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Advances in Water Resources*, **51**: 217–246.

doi:10.1016/j.advwatres.2012.07.018.

Xue DY, Zhang LG, Liu ZY, et al. 2025. Predicting porosity in tight sandstone reservoirs based on logging while drilling engineering parameters. *Scientific Reports*, **15**: 14241. doi:10.1038/s41598-025-98311-6.

Zhang YD, Gao GH, Liu ZP, et al. 2021. Differential characteristics of fluid occurrence in tight sandstone reservoirs: a case study of Triassic Yanchang Formation in Ordos Basin. *Petroleum Geology & Experiment*, **43**: 1024–1030. doi:10.11781/sysydz2021061024.