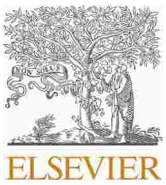
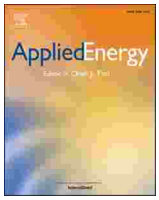
Applied Energy 333 (2023) 120604



Contents lists available at ScienceDirect

Applied Energy

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



An advanced prediction model of shale oil production profile based on source-reservoir assemblages and artificial neural networks

Yazhou Liu [a,b], Jianhui Zeng [a,b,*], Juncheng Qiao [a,b,*], Guangqing Yang [a,b], Shu'ning Liu [a,b], Weifu Cao [c]

[a] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, PR China
[b] College of Geosciences, China University of Petroleum (Beijing), Beijing 102249, PR China
[c] Exploration and Development Research Institute, PetroChina Daqing Oilfield Company, Daqing 163712, PR China

HIGHLIGHTS

- Propose a universal multidisciplinary data-driven workflow to predict shale oil production through ANN.
- Comprehensive quantitative evaluation of geological factors in the source-reservoir assemblages.
- The geometric and harmonic averaging algorithm strategies are recommended to characterize reservoir heterogeneity.
- Reservoir storage capacity has the most significant contribution to shale oil production.

ARTICLE INFO

Keywords:
Sweet spots
Production evaluation
Machine learning
Quantitative evaluation
Shale oil

ABSTRACT

Over the past decade, hydrocarbon production from shale oil reservoirs has become increasingly common, and successful shale oil exploration and development depends significantly on the accurate evaluation of the sweet spots. However, different scholars have established different evaluation standards for sweet spots under different geological settings, and it is difficult for these standards to form a universal evaluation standard. The sweet spots should be synonymous with the overall combination of geological, engineering and economic sweet spots. The shale oil production evaluation would be a valid indicator due to the comprehensive combination of the above three perspectives. This paper demonstrates a multidisciplinary data-driven workflow to predict shale oil production through machine learning and quantitative evaluation of geological variables. 48 test sections from 30 exploratory wells in the Lucaogou Formation of the Jimusaer Sag are taken as an example for application demonstration. The proposed 13 geological variables based on source-reservoir assemblage types, source rock quality, reservoir quality, migration dynamics, and conduit conditions realize a systematic and comprehensive characterization of hydrocarbon generation, storage, dynamics, and flow stimulation. Based on the different averaging algorithms in the ANN model, good agreement has been observed between predicted and simulated data for training ($R > 0.95$) and validation ($R > 0.87$). Moreover, the geometric and harmonic averaging algorithms are preferentially recommended to characterize reservoir heterogeneity. In unconventional reservoirs, there is no single attribute that can be used to predict success or failure. The training results of the advanced prediction model are better than the other five single reservoir characterization models. On the well J174 dataset, the sweet spot predicted by the model matches well with the oil test results. The increase in liquid hydrocarbon content, mud gas content, TOC and normal faults percentage has positive effects on shale oil production, while the increase in reverse faults percentage has negative effects on shale oil production. This research provides ideas for intelligent prediction of the distribution of sweet spots in unconventional resources, and is also important for the development of intelligent hydrocarbon exploration technology.

* Corresponding authors.
E-mail addresses: zengjh@cup.edu.cn (J. Zeng), Juncheng.Qiao@cup.edu.cn (J. Qiao).

https://doi.org/10.1016/j.apenergy.2022.120604
Received 25 July 2022; Received in revised form 9 December 2022; Accepted 28 December 2022

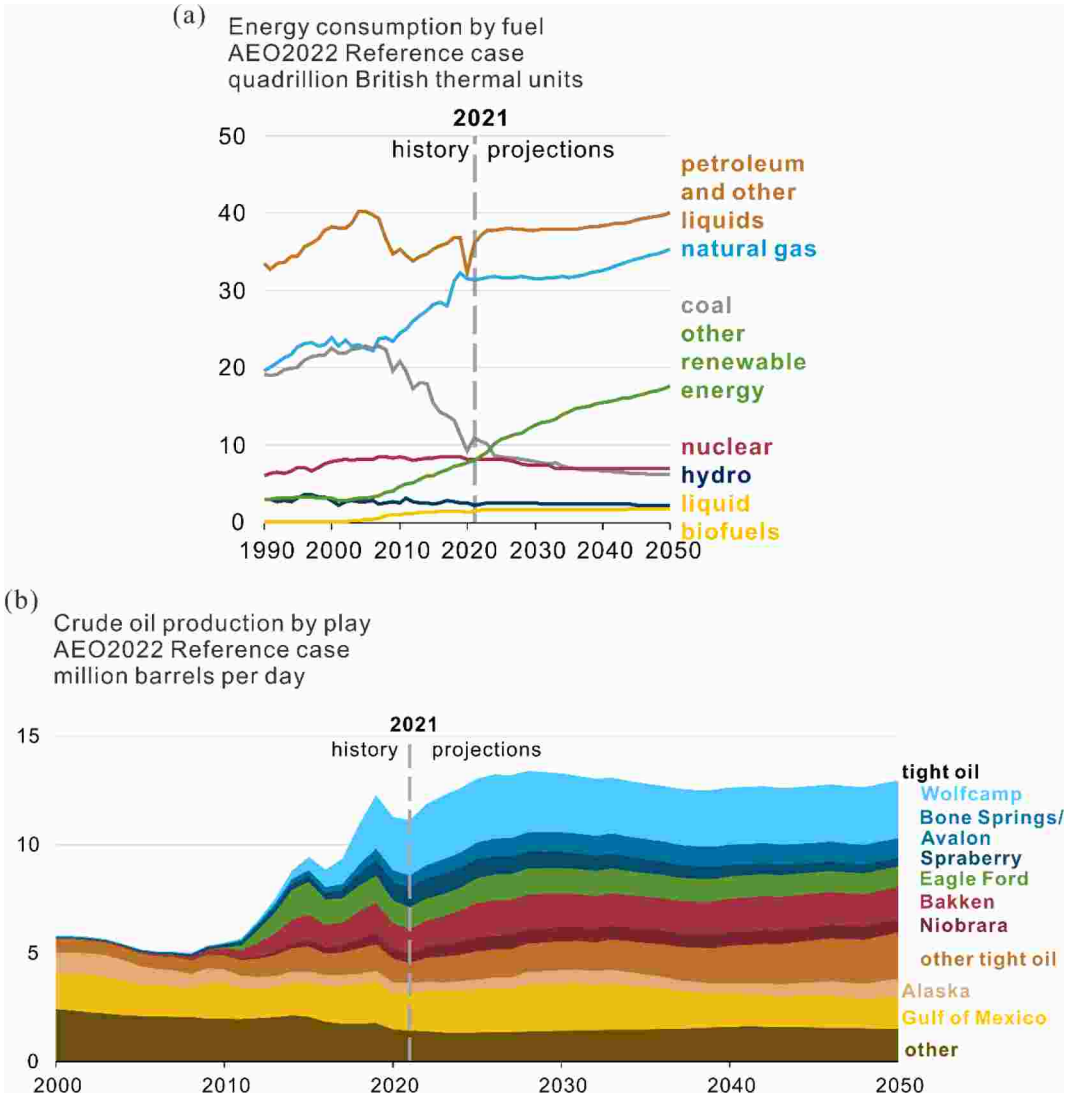


Fig. 1. (a) Energy consumption by fuel from 1990 to 2050. (b) Sources of U.S. crude oil production from 2000 to 2050 [4].

1. Introduction

In the 1800s, coal was the primary fuel source supporting human development, but in the 1900s, petroleum and natural gas significantly replaced coal [1]. Since 1990, petroleum and other liquid resources have become major energy consumption, providing for the development of science, technology, medicine, etc (Fig. 1a). About a decade ago, "petroleum and gas from shale" would not have been considered a valid reservoir. Since 2010, multistage hydraulic fracturing and horizontal wells have made it possible to produce petroleum and natural gas from shale in an economical way [2–3]. The U.S. Energy Information Administration (EIA) estimates that U.S. crude oil production in 2021 from tight (shale) oil resources reaches about 2.64 billion barrels (or about 7.22 million barrels per day) (Fig. 1b) [4]. This is equivalent to about 65% of total U.S. crude oil production in 2021. Tight (shale) oil is a type of oil that occurs in low-permeability shale, sandstone and limestone deposits. Outside the United States, four countries account for more than half of the identified shale oil resources, including Russia, China, Argentina and Libya [5]. Compared with the 41 countries assessed, China's shale oil resources are ranked third after the United States, with approximately 32 billion barrels. Drawing on the successful experience of shale oil exploration and development in North America, China has made some significant achievements in the "13th Five-Year Plan" period (2016–2020). For example, the cumulative production of shale oil in the Dagang oilfield exceeds 100,000 tons; the shale oil resource in Jimusaer Sag is 1.12 billion tons and aims to achieve the target of producing 500,000 tons per year in 2022. At present, the China Energy Administration has included the strengthening of shale oil exploration and development in the "14th Five-Year Plan" (2021–2025). Although new technologies have driven down the cost of extraction, extracting hydrocarbons from shale oil reservoirs remains highly risky. Currently, horizontal well drilling and fracturing techniques can only produce less than 10% of shale oil [6]. Thus, the prediction of shale oil production will help to screen the sweet spots for the commercial development of shale oil.

In unconventional resources, production performance is extremely dependent on accurately targeting multi-stage hydraulic fracturing treatments in sweet spots with quality rock properties and good production potential [7–9]. However, the identification of sweet spots faces two primary challenges, namely reservoir characterization and reservoir heterogeneity [10]. Over the recent years, parameters based on geology, geochemistry, geophysics, reservoir engineering and geomechanics have been presented for reservoir characterization and reservoir heterogeneity considering different geological factors (Table 1). Many available projects and studies have focused on identifying sweet spots based on the adsorption capacity and mobility of shale oil. Many of these studies use S_1/TOC > 100 mg HC/g TOC to screen prospective development zones [11–13]. Although this method is a classic approach, it

Table 1
Summary of sweet spot evaluation methods in unconventional resources.

Author(s)	Method	Disadvantage	Geology		Geochemistry		Petrophysics			Reservoir engineering			Geomechanics	
			Sequence stratigraphy	Sedimentary analysis	Source rock	Shale oil	Porosity	Oil saturation	Oil content	Pore pressure	AP&GOR	Production data	Brittleness	Stress regime
Jarvie [12]	Oil saturation index	A large number of core data; Not applicable to formations with low S_1 and TOC; Empirical threshold values			√				√			√		
Hu et al. [11]; Li et al. [13]	Graded evaluation method	Empirical threshold values; A large number of core data			√				√			√		
Liu et al. [14]	Oil content evaluation index	Human factors affecting oil-bearing area description			√		√	√	√		√	√		
Sharma and Sircar [15]	Multi-attribute analysis and shale potential	Lack of optimizing key parameters; Simple multi-attribute stacking			√			√	√			√	√	√
Zhao et al. [16]	Multi-attribute stacking	Subjective selection of variable threshold values	√	√	√		√	√	√			√	√	
Cudjoe et al. [17]	Quality index reservoir properties	Bias to low-value variables contribution			√	√	√	√				√	√	√
Eid et al. [18]	Productivity and fracture index	Lack of optimizing key parameters			√		√	√				√	√	√
Hou et al. [19]	Sweet spot index	Bias to low-value variables contribution			√		√	√		√		√	√	
Licitra et al. [20]	Analysis of highly productive zones	Lack of systematic quantitative geological evaluation models; too many parameters	√	√	√		√			√	√	√	√	√
Ter Heege et al. [21]	Key performance indicators	Too simple; lack of validation with production data			√		√		√				√	
This study	Source-reservoir assemblages and artificial neural networks		√	√	√	√	√	√	√	√	√	√	√	√

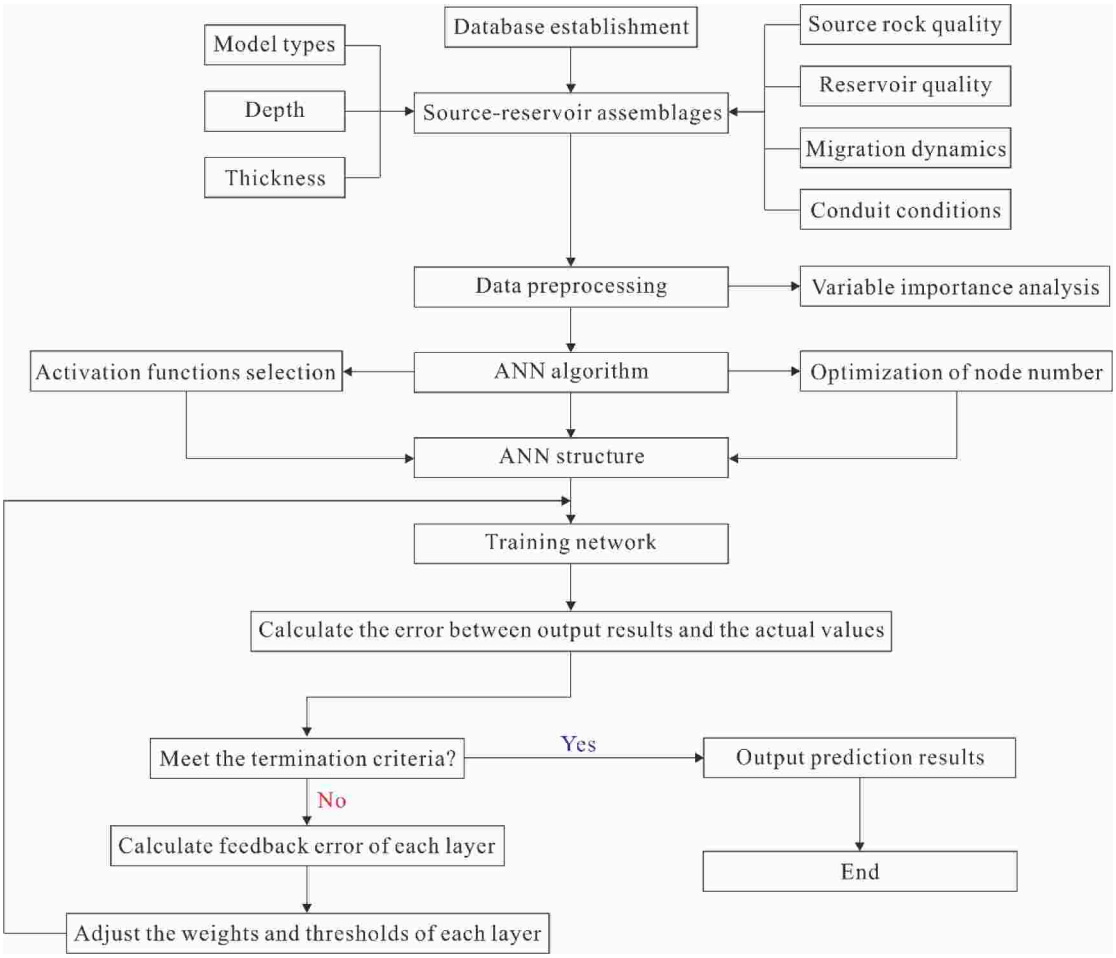


Fig. 2. Flowchart for building shale oil production profile prediction model based on ANN algorithm.

still has many disadvantages. Firstly, a large amount of core data needs to be obtained. Secondly, the threshold value (100 mg HC/g TOC) is an empirical value, which is based on the statistical analysis of marine shale oil production data. This value is not necessarily applicable to the sweet spot evaluation of lacustrine shales.

Other studies have established best practices or general methods for identifying sweet spots [14–16]. Many previous studies have not identified specific quantitative metrics, or only semi-quantitative metrics. Some screening criteria contain a mixture of objective threshold indicators and subjective descriptive standards. The screening criteria and indicators of sweet spots are not consistent in different geological settings. Thus, in the case of subjective criteria, this may cause vagueness and make it difficult to determine the final desired development zones. While these studies help to identify potentially suitable prospective plays from the basin scales, more detailed methods are needed to screen specific development zones from the sand body scales.

Moreover, many factors can affect shale oil production or enhanced oil recovery (EOR), ranging from source-reservoir assemblage types, source rock quality, reservoir quality, migration dynamics to conduit conditions [7]. Previous studies of sweet spot evaluation criteria reflect the broad expertise of researchers and are not comprehensive in presenting holistic criteria that address all factors [17–21]. Furthermore, these evaluation parameters are excessive and lack systematic and comprehensive. Meanwhile, some evaluation parameters are interrelated such as oil content, oil saturation and porosity, resulting in unclear control factors of shale oil sweet spots. Shale is highly heterogeneous vertically and laterally over very short distances. Due to the heterogeneity of unconventional reservoirs, it is often necessary to gather a large amount of data to characterize this heterogeneity to optimize production enhancement designs and completion practices. Besides, each shale play has a unique set of geological settings and conditions that control production performance. In these plays, there is no single attribute that can be used to predict success or failure. Many previous studies have focused more on specific plays or specific projects rather than developing a common set of sweet spot evaluation methods that can be applied to any area of interest [10–21] (Table 1).

As a result, there is a need for a systematic and optimized sweet spot evaluation workflow. Unconventional resource development requires the input of many disciplines and the application of the right technologies. How can all of this be tied together and optimized? In this paper, a data-driven machine learning approach is presented and an advanced prediction model based on source-reservoir assemblages and BP neural networks is developed for predicting shale oil production. Our methodology is applicable to different regions and projects and presents a comprehensive approach to quantitative evaluation addressing all geological factors. Compared with previous studies, the innovations of this study involve the following perspectives: (1) Optimizing comprehensive quantitative characterization of reservoir heterogeneity in terms of source-reservoir assemblage types, source rock quality, reservoir quality, migration dynamics, and conduit conditions, including data from geology, geophysics, geochemistry, geomechanics, and reservoir engineering; (2) Establishing a multidisciplinary data-driven evaluation workflow to meet the needs of each stage of development; (3) Providing ideas for building digital oil fields through machine learning and geological evaluation. The purpose of this paper is to demonstrate shale oil production predictions to reduce the risk of shale oil exploration and

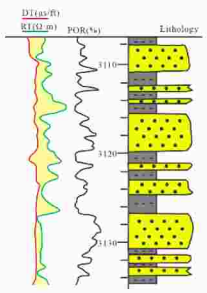
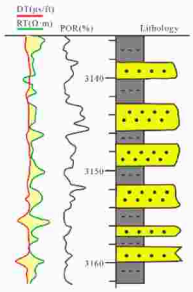
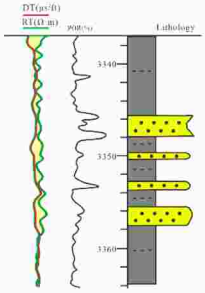
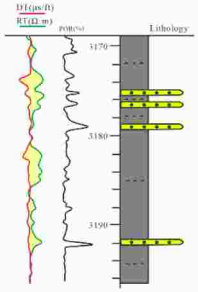
Model types	"Reservoir-based sandwiched" model	"Highly interbeded" model	"Source-based sandwiched" model	"Massive source" model
Source-reservoir assemblages schematic diagrams				
Reservoirs thickness ratio	>0.6	0.6~0.4	0.4~0.1	0.1>
Expulsion efficiency	Excellent	Good	Fair	Poor
Secondary migration	High potential	Medium potential	Low potential	Little potential
Shale plays	Poor	Fair	Good	Excellent
Conventional resources	High potential	Medium potential	Low potential	Little potential
Sedimentary environmrnt	Lakeshore	Shallow lake	Semi-deep lake	Deep &semi-deep lake

Fig. 3. Schematic diagram of source-reservoir assemblage models.

development and to avoid adopting a trial-and-error approach to blindly drill numerous wells and hydraulic fracturing.

2. Methods

2.1. Data establishment

The data presented in this paper were collected from the Research Institute of Experimental Testing of PetroChina Xinjiang Oilfield Company and the State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing). Relevant matched data were gathered for 48 test sections from 30 exploratory wells, where these data consisted of different types of data from geology, geophysics, geochemistry, geomechanics and reservoir engineering. Furthermore, these data could realize a comprehensive characterization of source rock quality, reservoir quality, migration dynamics and conduit conditions from qualitative to quantitative characterization, thereby more accurately predicting shale oil production in the target section.

2.2. Workflow and steps

Fig. 2 illustrates the process of how to construct an advanced shale oil production profile prediction model. This process is mainly divided into four steps: (1) database establishment; (2) develop source-reservoir assemblages assessment; (3) data preprocessing and variable importance analysis; (4) develop and evaluate the prediction models for shale oil production profile based on ANN algorithm (Fig. 2). Among them, the assessment of source-reservoir assemblages covers the model types, depth, thickness, source rock quality, reservoir quality, migration dynamics and conduit conditions in the test sections.

2.2.1. Source-reservoir assemblage models

Shale oil systems are self-contained source-reservoir systems in which oil generated from organic-rich mudstones is stored in organic-rich mudstone intervals or migrates to juxtaposed, continuous organic-lean intervals [21–24]. Therefore, in shale oil systems, the spatial interconfiguration between source rocks and reservoirs in terms of lithology, contact patterns, thickness differences, and conduit types (e.g., pore-type media and fracture-type media), namely the source-reservoir assemblages, has a significant influence on the generation, migration, and accumulation of shale oil. Examination of shale oil prospects suggests four primary models of source-reservoir assemblages (Fig. 3). The first model is presented as a "massive source" model, where the ratio of reservoir thickness to formation thickness is less than 0.1. In this system, a large amount of hydrocarbons generated from source rocks are retained within source rocks, and the overall hydrocarbon expulsion efficiency is poor. Due to the low potential for secondary migration here, this will result in excellent shale plays and limited conventional resources. This model is usually developed in deep and semi-deep lake environments. The second model is when the reservoirs are sandwiched between organic-rich source rocks, where the ratio of reservoir thickness to formation thickness varies from 0.1 to 0.4. In this system, the hydrocarbon expulsion efficiency is higher than that of the first model, and the upper and lower source rocks can contribute hydrocarbons to the same reservoir, but it is still fair. Depending on reservoir quality, secondary migration may be low potential, which would result in good shale plays and low potential conventional resources. The model is typically developed in a semi-deep lake environment. The third model is source rock and reservoir interbedding, where the ratio of reservoir thickness to formation thickness ranges from 0.4 to 0.6. This system leads to good hydrocarbon expulsion efficiency. Secondary migration with medium potential could occur in this system and it could result in fair shale plays and medium potential conventional resources. The model is frequently found in shallow lake environments. The fourth model is when the source rocks are sandwiched between reservoirs, where the ratio of reservoir thickness to formation thickness is greater than 0.6. The hydrocarbon expulsion efficiency in such a system is the highest among all source-reservoir assemblage models. Depending on the source rock quality, secondary migration occurs with high potential and the maturity of the crude oil in the reservoir may be higher than the in situ source rock. This model is typically observed in lakeshore environments. Each source-reservoir assemblage model is closely related to

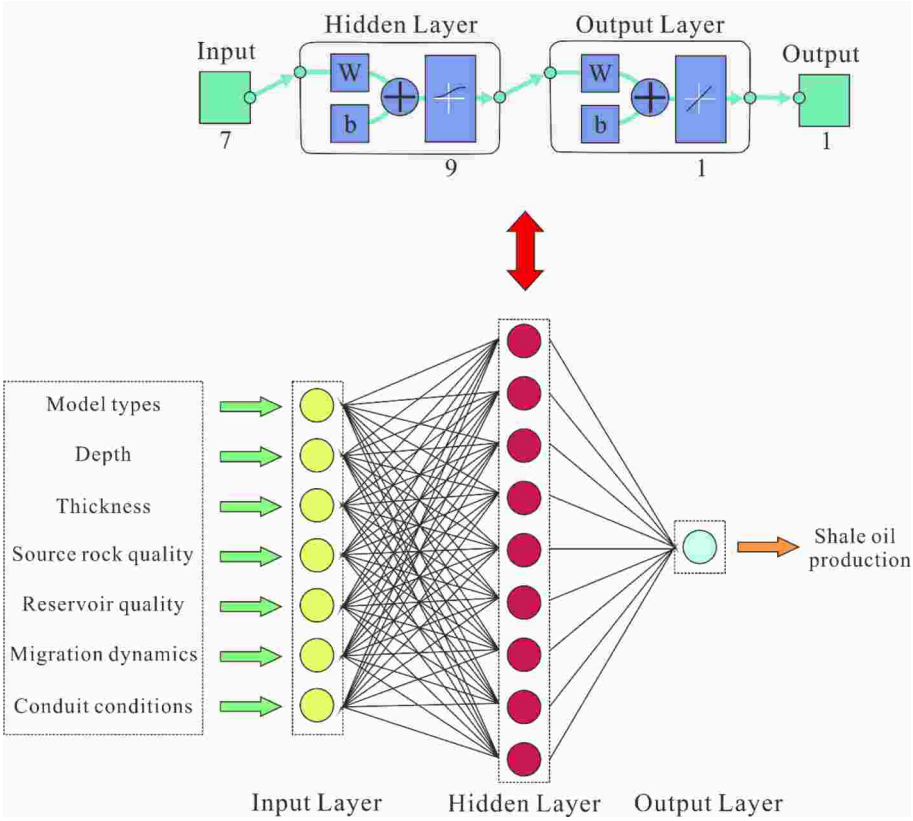


Fig. 4. Single-layer neural network structure.

changes in the depositional environment. Moreover, different source-reservoir assemblages require different drill and development techniques. In all models, the ratio of reservoir thickness to formation thickness is obtained based on the isochronous stratigraphic framework.

The isochronous stratigraphic framework is established based on the theory of high-resolution sequence stratigraphy (HRSS). Layer correlation is a fundamental task in oil and gas exploration and development [25–27]. Cross [28] establishes a new method, HRSS, and applies it to layer correlation. Based on outcrops, cores, logging and seismic sections, HRSS divides the strata into precise stratigraphic sequences, and converts 1D drilling or logging information into 3D stratigraphic stacking relationships, thereby establishing genetically linked layer correlation at different scales. Over the years, many scholars have made a lot of efforts on the theory and application of HRSS [29–34]. To establish isochronous stratigraphic frameworks in HRSS, important markers for identification include unconformity interfaces, flooding surfaces, and event-specific depositional surfaces [35–37]. Based on base-level cycles, sequence units are further divided into long-term, middle-term and short-term cycles. A middle-term base-level cycle is formed during a period of modest water depth variation with genetically related stratigraphic stacks. During a period of small water depth variation, genetically-related stratigraphic stacks with similar lithology and lithofacies form a short-term base-level cycle. In this study, the establishment of the isochronous stratigraphic framework is based on middle-term and short-term base-level cycles. Taking the short-term base-level cycle as the formation unit, the ratio of the reservoir thickness to the formation thickness of each formation unit is calculated. Based on this, the source-reservoir assemblage model of the test section is determined. If the test section is located in more than one formation unit, the composite naming principle is employed.

2.2.2. *Reservoir heterogeneity characterization*

The test sections are often constituted by source rocks and reservoirs. Moreover, even within source rocks or reservoirs, there is substantial heterogeneity in the values of target parameters corresponding to different depths. Owing to its highly variable nature, some form of target parameter averaging is usually required. The question is which average is more reliable? The scale-up of these target parameters is becoming more problematic due to the complexity and uncertainty of the geological body itself at different scales [38–41]. Here, four common simple averaging methods are adopted and their effectiveness is compared, namely (1) harmonic averaging, (2) geometric averaging, (3) arithmetic averaging, and (4) quadratic averaging. The weights of the target parameters corresponding to each depth are the same. These averages are calculated using equations (1)-(4), respectively. Ultimately, these averages have the following relationship, i.e., $x_h \le x_g \le x_a \le x_q$.

$$x_h = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right]^{-1} \tag{1}$$

$$x_g = \left[\prod_{i=1}^{n} x_i\right]^{\frac{1}{n}} \tag{2}$$

$$x_a = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

$$x_q = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right]^{\frac{1}{2}} \tag{4}$$

where, x_h is the harmonic mean. x_g is the geometric mean. x_a is the arithmetic mean. x_q is the quadratic mean. x_i is the value of the target parameter at the i-th depth point.

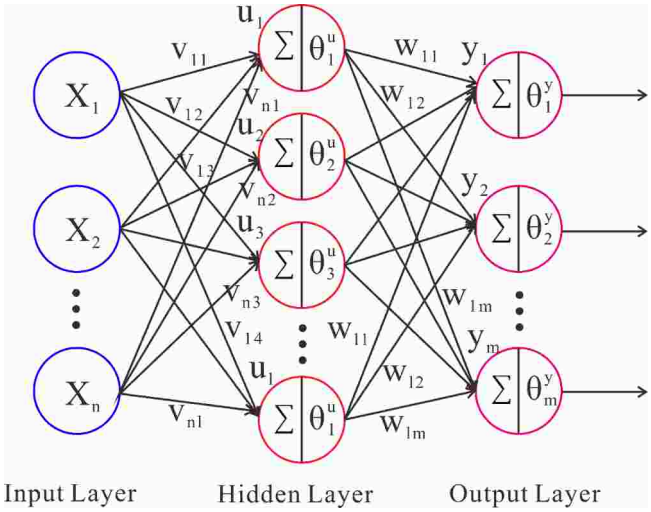


Fig. 5. The workflow of the BP neural network algorithm.

2.2.3. Data preprocessing

Considering that the physical meaning and scope of each target parameter are quite different, the input data of the target parameters in this model are normalized in advance. Normalization is the process of scaling down or up the raw data to the desired scope, which can then be applied to further stages. In particular, normalization plays an important role in areas such as soft computing and cloud computing [42–44]. There are many normalization techniques, such as Min-Max normalization, Z-score normalization, and Decimal scaling normalization. Here, we use the Min-Max normalization technique to normalize the target parameters [45–47]. This expression is as follows:

$$x_{nor} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

where, x is the absolute value of the target parameter. $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the target parameter dataset, respectively.

2.2.4. Variable importance analysis

The determination of the main control factors is not only the basis for building prediction models, but will also have a profound influence on the decision-making of shale oil exploration and development. With the advancement of mathematical algorithms, there are many methods to determine the importance of variables [48–55]. In this paper, based on the target parameters we established, we utilize the grey correlation coefficient, Pearson correlation coefficient, Spearman correlation coefficient and Kendall correlation coefficient to rank and analyze the influencing factors of shale oil production profiles, respectively.

2.3. Artificial neural network (ANN)

2.3.1. ANN structure

In this study, MATLAB R2018a is employed to predict shale oil production. For the structure of ANN, it covers the input layer, the hidden layer, the number of hidden layer nodes, and the output layer. In this work, model types, depth, thickness, source rock quality, reservoir quality, migration dynamics and conduit conditions are used as input layers for the neural network. Of these, the quantitative evaluation parameters established for the latter four are discussed below. The shale oil production from a single well is used as the output layer. In the subsequent work, the number of hidden layers is determined and the number of hidden layer nodes is optimized. The structure of the single-layer network framework model is shown in Fig. 4 (taking the number of 9 hidden layer nodes as an example).

2.3.2. BP neural network algorithm

ANN prediction is a nonlinear transformation system with the ability to organize and learn. It can be trained and validated spontaneously by learning data samples eventually making the output value close to the desired output value [56–58]. BP (backpropagation) neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986 [59]. It is a multilayer feedforward neural network trained according to the error back propagation algorithm and is one of the most widely applied neural network models. BP neural network has the ability of arbitrarily complex pattern classification and excellent multi-dimensional function mapping, which solves some problems that cannot be solved by simple perceptrons. In essence, the BP algorithm is to use the gradient descent method to calculate the minimum value of the squared network errors [60–61]. Before the BP neural network can make predictions, the data set must be trained. Fig. 5 illustrates the workflow of the BP neural network algorithm, which follows the following steps [62–64]:

Assume that a neural network has n input nodes, l hidden layer nodes and m output nodes. The threshold of the k-th node in the output layer is expressed by θ_k^y, and the threshold of the j-th node in the hidden layer is expressed by θ_j^u. The weight between the i-th node in the input layer and the j-th node in the hidden layer is represented by V_{ij}, and the weight between the j-th node in the hidden layer and the k-th node in the output layer is represented by W_{jk} (Fig. 5). A certain learning rate and node activation function are also given.

Using the input vector X, the weight V_{ij} between the input layer and the hidden layer and the hidden layer threshold θ_j^u, as well as the hidden layer activation function, the hidden layer output u is computed.

$$u_j = f\left(\sum_{i=1}^{n} v_{ij}x_i + \theta_j^u\right) \quad j = 1, 2, \ldots, l \tag{6}$$

$$f(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

where, l is the number of nodes in the hidden layer, and f is the activation function of the hidden layer.

Utilizing the hidden layer output u, the weight w_{jk} between the hidden layer and the output layer and the output layer threshold θ_k^y, as well as the output layer activation function, the output layer output y is calculated.

$$y_k = g\left(\sum_{j=1}^{l} w_{jk}u_j + \theta_k^y\right) \quad k = 1, 2, \ldots, m \tag{8}$$

$$g(x) = x \tag{9}$$

where, m is the number of nodes in the output layer, and g is the activation function of the output layer.

Based on the network prediction output y and the actual value y_{actual}, the error J is evaluated.

$$J_k = (y_k - y_{k,actual})^2 \quad k = 1, 2, \ldots, m \tag{10}$$

If the error does not meet the desired value, the individual weights and thresholds need to be adjusted to make the network prediction closer to the actual value. In this study, the gradient descent method is employed to update the weights and thresholds. The updated weights are calculated as follows:

$$\dot{v_{ij}} = v_{ij} - \eta\frac{\partial J_k}{\partial v_{ij}} \tag{11}$$

$$\frac{\partial J_k}{\partial v_{ij}} = 2\left(y_k - y_{k,actual}\right)w_{jk}u_j\left(1 - u_j\right)x_i \tag{12}$$

$$\dot{w_{jk}} = w_{jk} - \eta\frac{\partial J_k}{\partial w_{jk}} \tag{13}$$

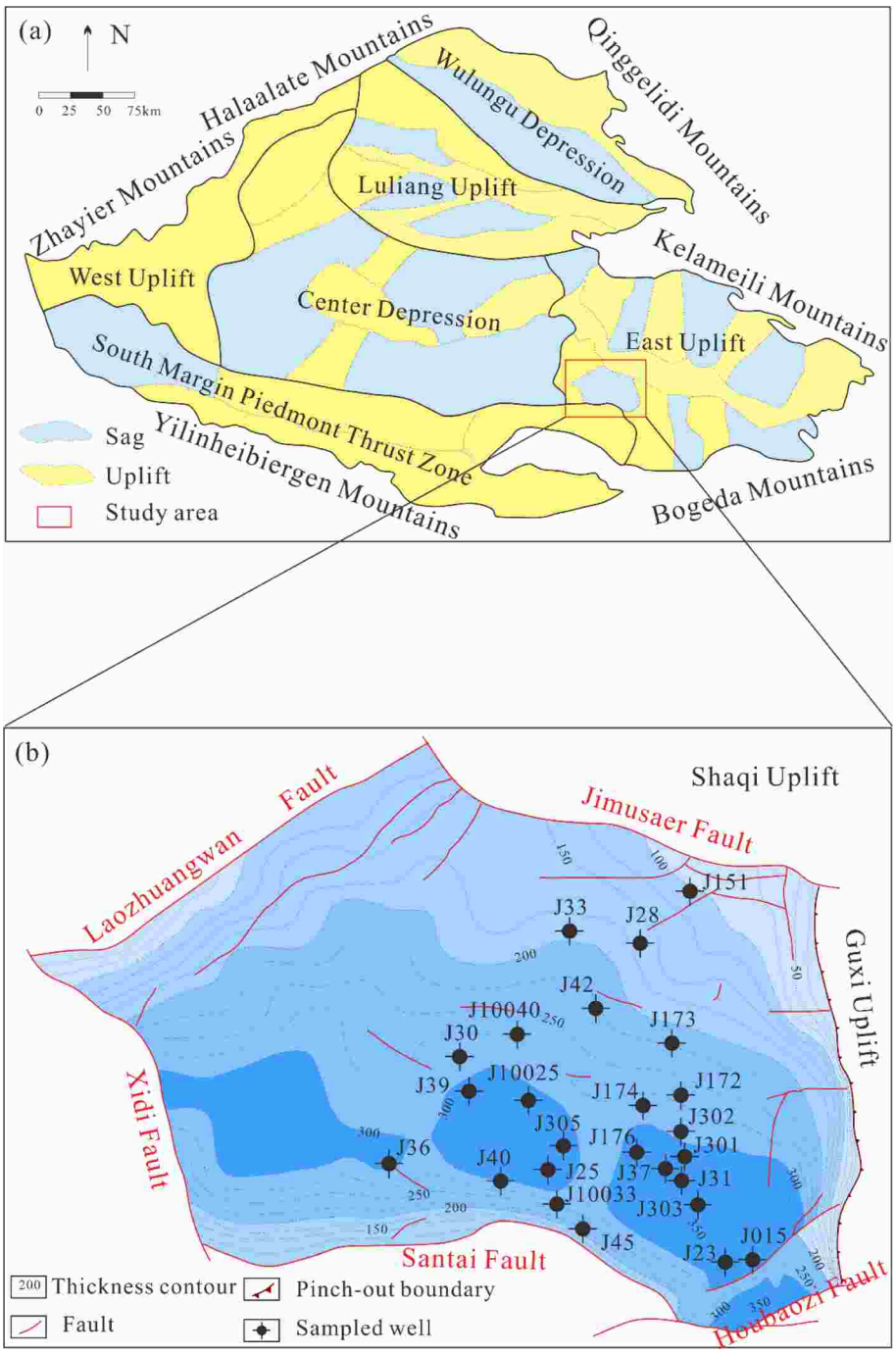


Fig. 6. Distribution of well locations with test data in the Jimusaer Sag. (a) Geographical Location of the Jimusaer Sag in the Junggar Basin; (b) Thickness contour map of the Lucaogou formation in the study area (modified from Liu et al. [14,71]).

$$\frac{\partial J_k}{\partial w_{jk}} = 2(y_k - y_{k,actual})u_j \tag{14}$$

where, $v_{ij}^{'}$ and $w_{jk}^{'}$ are the updated weights, and η is the learning rate. The updated thresholds are calculated as follows:

$$\theta_j^{u'} = \theta_j^{u} - \eta \frac{\partial J_k}{\partial \theta_j^{u}} \tag{15}$$

$$\frac{\partial J_k}{\partial \theta_j^{u}} = 2\left(y_k - y_{k,actual}\right) w_{jk} u_j \left(1 - u_j\right) \tag{16}$$

$$\theta_k^{y'} = \theta_k^{y} - \eta \frac{\partial J_k}{\partial \theta_k^{y}} \tag{17}$$

$$\frac{\partial J_k}{\partial \theta_k^{y}} = 2\left(y_k - y_{k,actual}\right) \tag{18}$$

where, $\theta_j^{u'}$ and $\theta_k^{y'}$ are the updated thresholds.

Determine if the error has reached the desired value or the number of iterations of the algorithm has met the maximum value. If not, keep repeating the above steps.

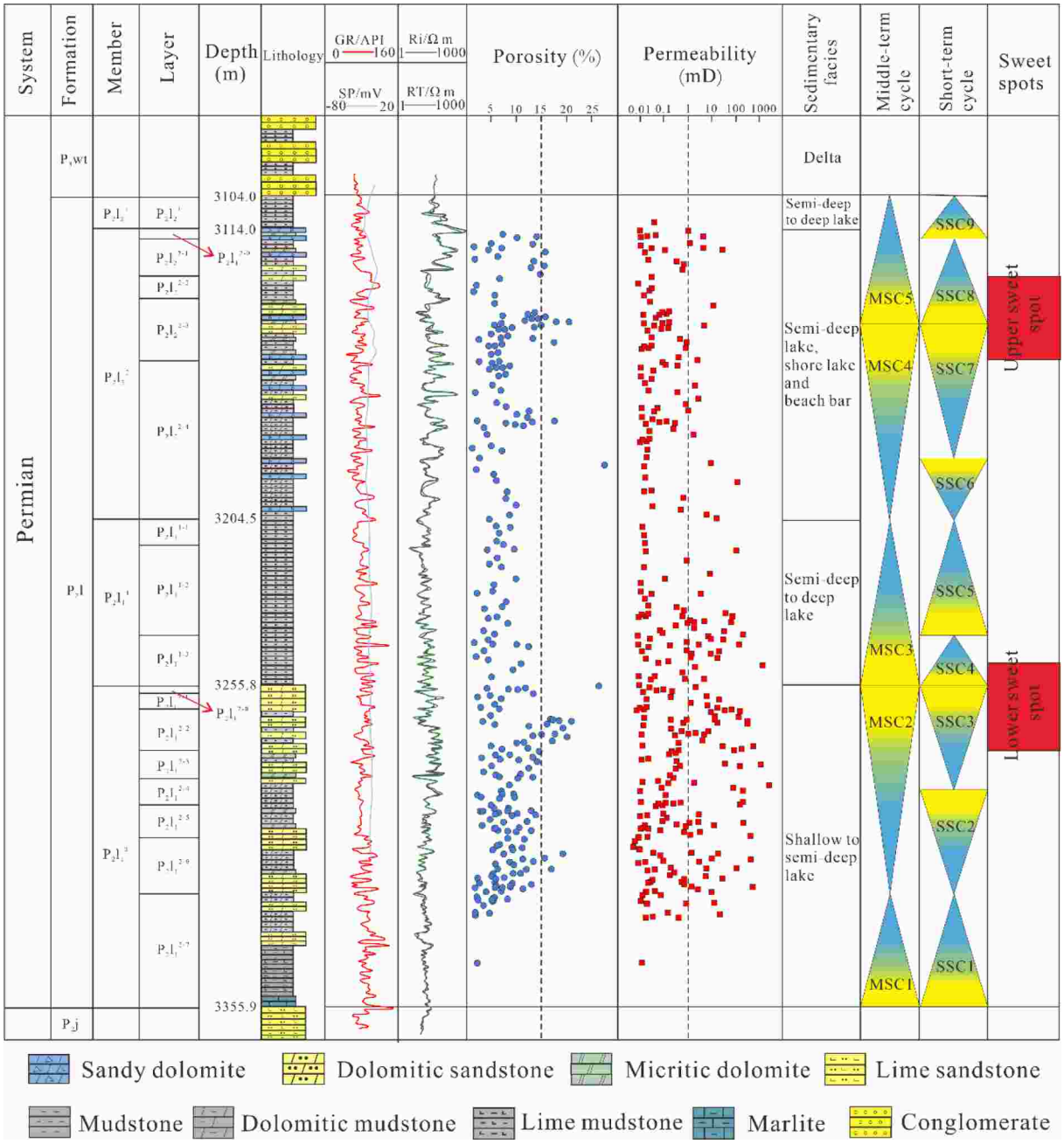


Fig. 7. The medium-term and short-term cycle sequences classification of the Lucaogou Formation in the Jimusaer Sag.

2.3.3. *Optimization of the node number in the hidden layer*

BP neural networks can incorporate multiple hidden layers. Too many hidden layers may result in overfitting, increasing the training difficulty, and possibly even causing the model to fail to converge. In general, one hidden layer can approach most nonlinear continuous functions with any precision [56]. Therefore, a neural network structure with one hidden layer is adopted in this model. Similarly, it is quite essential to choose the proper number of nodes in the hidden layer. Multiple nodes can reduce the error between the predicted and actual values, but too many nodes can lead to overfitting and too much training time. Therefore, it is necessary to optimize the node number in the hidden layer. Usually, the node number of the hidden layer is calculated by the following equation [56,65–66].

$$l = \sqrt{n+m} + b \tag{19}$$

where, l is the number of nodes in the hidden layer, n is the number of nodes in the input layer, and m is the number of nodes in the output layer. b is a constant from 1 to 10.

2.3.4. *Activation function*

The activation function, which is the function that runs on the nodes of the artificial neural network, is responsible for mapping the inputs of the nodes to the outputs [67–68]. If there is no activation function, then the network can only express linear mappings. At this time, even if there are more hidden layers, the entire network is equivalent to a single-layer neural network. If the activation function introduces nonlinear functions to the nodes, this will allow the neural network to arbitrarily approximate any nonlinear function. Therefore, this neural network can be applied to a wide range of nonlinear models. Currently, there are many different types of activation functions, such as Sigmoid, Tanh, Soft-max, ReLU, and Max-out functions. The input value of the Sigmoid function can be any real number, which can satisfy the large-scale and multi-

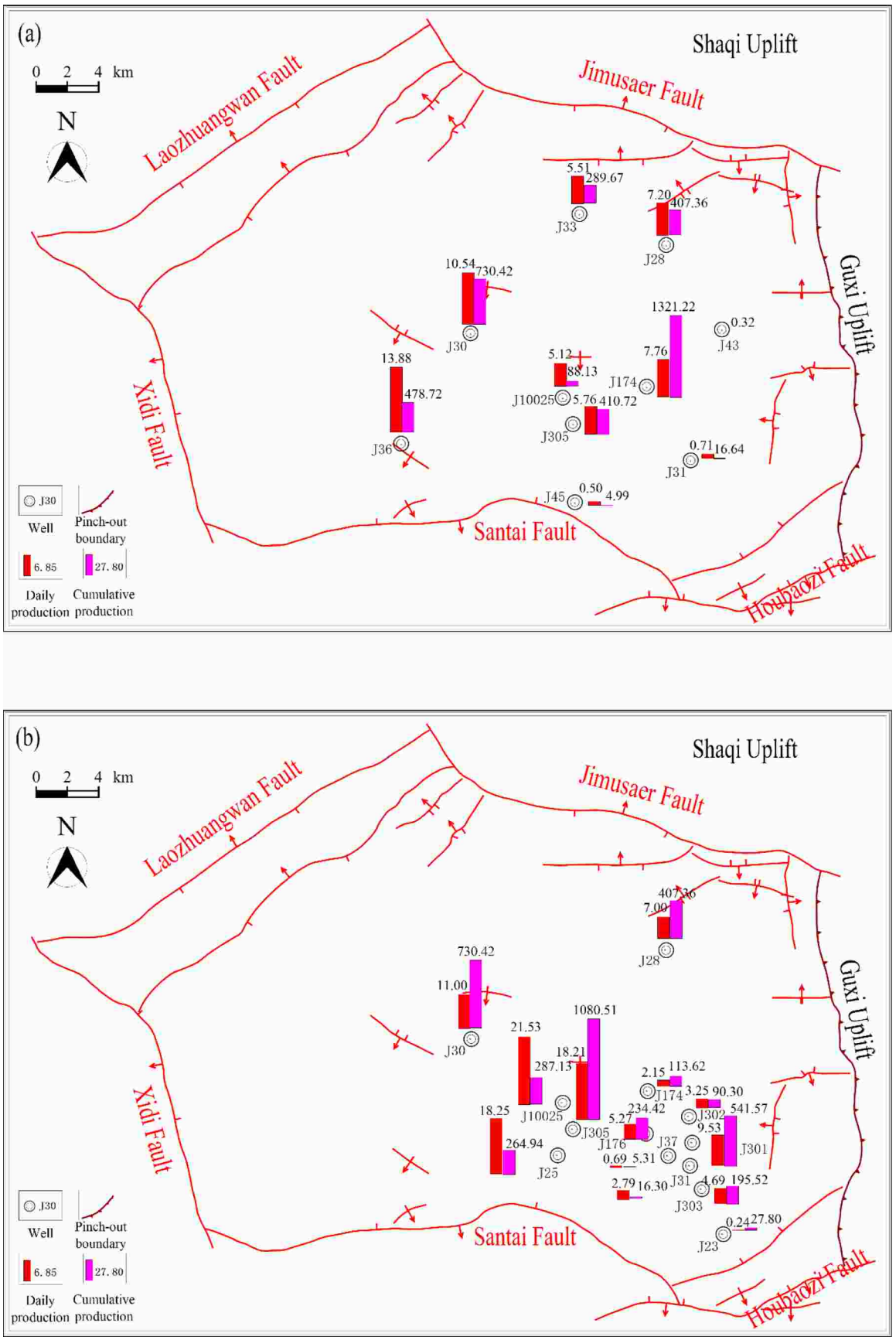


Fig. 8. Single-well production distribution of the Lucaogou Formation in the Jimusaer Sag. (a) The P_2l_1 interval; (b) The P_2l_2 interval.

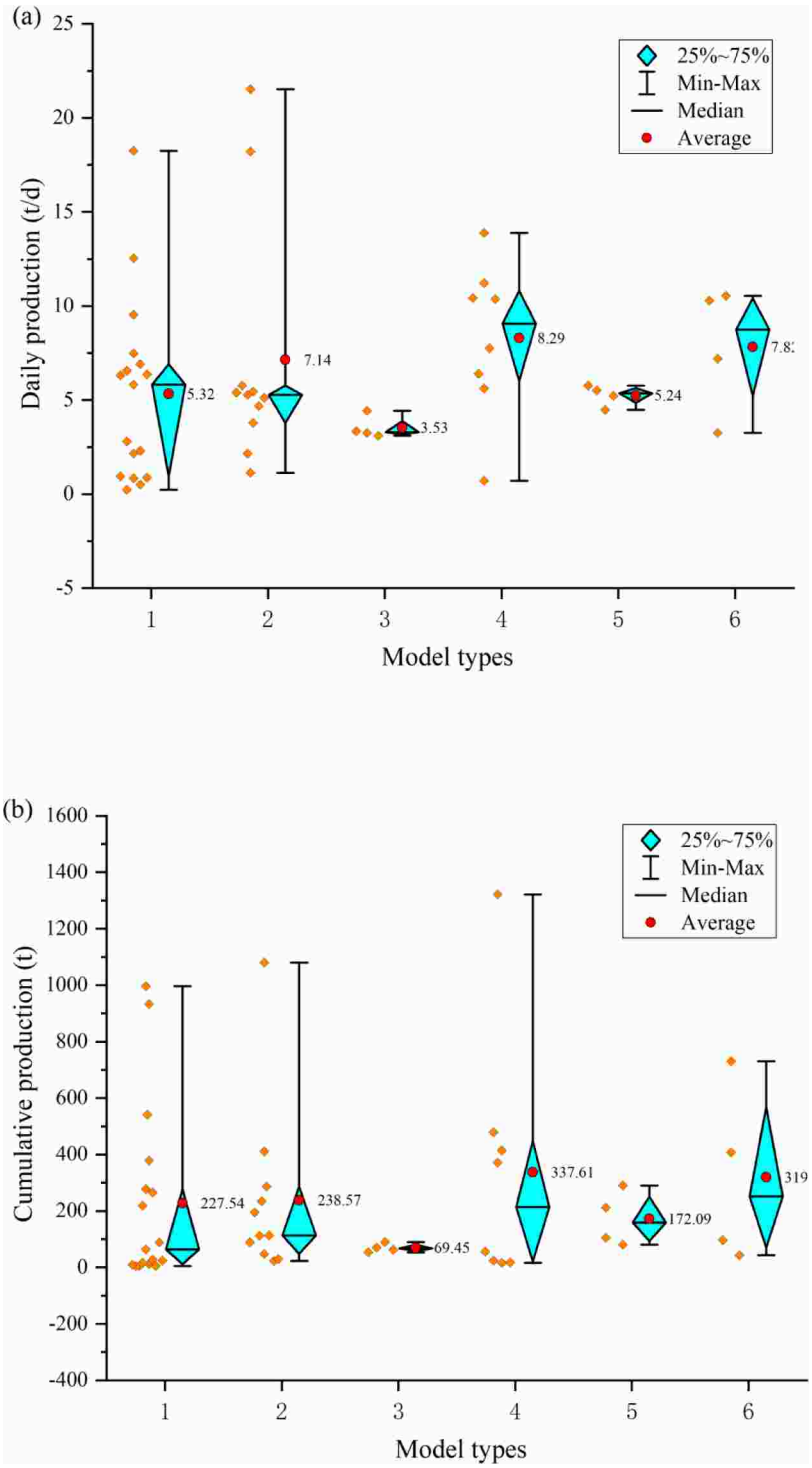


Fig. 9. Relationship between different source-reservoir assemblages and production in Lucaogou Formation, Jimusaer Sag. (a) Source-reservoir assemblage model types versus daily production; (b) Source-reservoir assemblage model types versus cumulative production. Key: 1 = "Reservoir-based sandwiched" model; 2 = "Highly interbedded" model; 3 = "Source-based sandwiched" model; 4 = "Reservoir-based sandwiched" & "Highly interbedded" model; 5 = "Highly interbedded" & "Source-based sandwiched" model; 6 = "Source-based sandwiched" & "Massive source" model.

parameter requirements of this model. The differentiability can significantly minimize the computational effort during the training process [56,69]. Therefore, the Sigmoid function is selected at the hidden layer nodes of the model in this study. The linear activation function can maintain the range of output values from the previous layer and can also produce arbitrary predicted values. Therefore, the linear activation function is chosen for the output layer nodes of the model [56]. The final BP neural network framework is built, as presented in Fig. 4.

3. Geological setting

3.1. Geographical location and sequence stratigraphy

The Junggar Basin in China is a typical intracontinental superimposed basin located at the intersection of the Kazakhstan Plate, the Siberian Plate, and the Tarim Plate. The Jimusaer sag is situated in the southeastern region of the Junggar Basin, with an area of about 1300 km^2 (Fig. 6a) [14,70–71]. The sag is adjacent to the Jimusaer Fault to the north, the Laozhuangwan Fault and the Xidi Fault to the west, and the Santai Fault to the south, while to the east it shows a gradual uplift

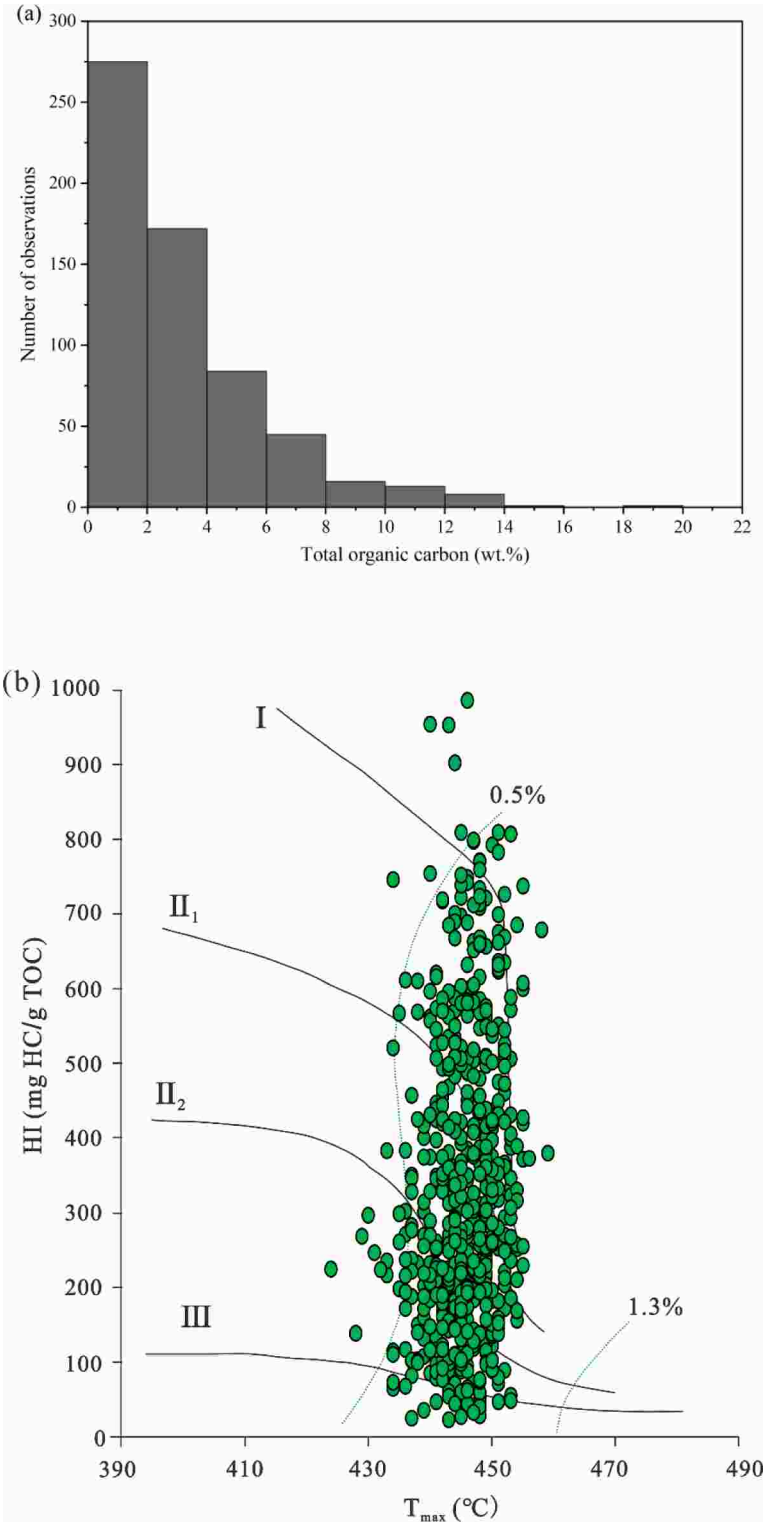


Fig. 10. Geochemical summary of source rocks from the Lucaogou Formation. (a) Histogram of total organic carbon content; (b) The relationships between hydrogen index (HI) and T_{max}; (c) The relationships between production index ($S_1/S_1 + S_2$) and T_{max}; (d) Histogram of the S_1/TOC*100 ratios.

slope that eventually transitions to the Guxi Uplift (Fig. 6b). The sag is a dustpan-shaped sag with high in the east and south, but low in the west and north. Although the sag has undergone several tectonic events, the sag is tectonically stable with formation dips between 3° and 5° in the main part, and large faults are not developed [72–73].

The Lucaogou Formation of the Jimusaer Sag is a typical saline lacustrine-deltaic sedimentary sequence. Affected by mechanical deposition, chemical deposition and biological deposition, the Lucaogou Formation is a fine-grained mixed sedimentary succession, including mudstones, siltstones, sandstones and carbonates [74–76]. This formation can be further subdivided into $P_2l_1^2$, $P_2l_1^1$, $P_2l_2^2$ and $P_2l_2^1$ intervals from the bottom to the top based on third-order cycles and lithological

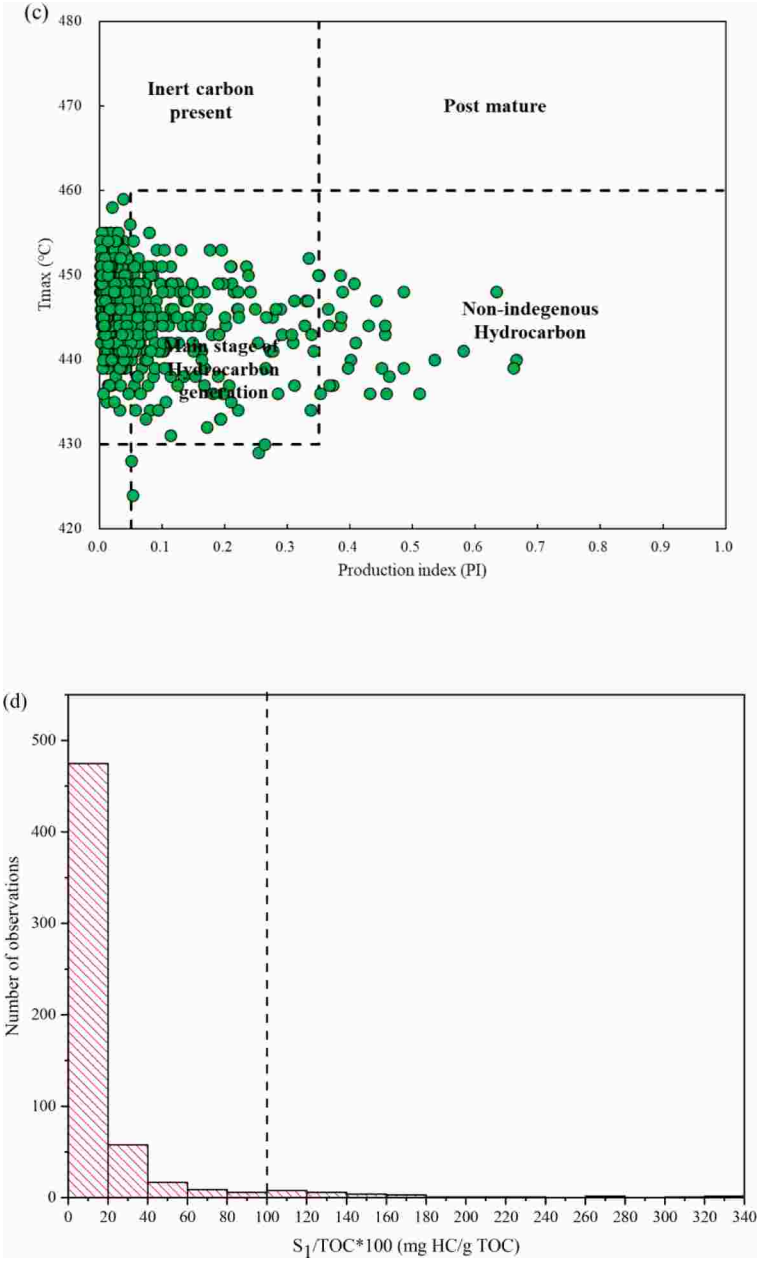


Fig. 10. (*continued*).

variations. There are two sweet spots in the Lucaogou Formation, the upper and lower ones (Fig. 7). The "sweet spot" here refers to the reservoir with porosity greater than 5%, oil saturation greater than 45%, and industrial oil flow under the existing development method [77]. The lower sweet spot is mainly located in the $P_2l_1^2$ interval and is dominated by clastic sedimentation, consisting primarily of siltstone, muddy siltstone, dolomitic siltstone and mudstone. It is distributed throughout the sag, with a thickness ranging from 18 to 68 m and an average value of 42.8 m. The sedimentary facies is dominated by the semi-deep to the deep lake and delta-front facies. The upper sweet spot is mainly located in the $P_2l_2^2$ interval and is dominated by carbonate sedimentation, including micritic dolomite, lime mudstone, sandy dolomite, and dolomitic mudstone. It is confined at the eastern slope of the sag, with a thickness ranging from 13 to 43 m and an average value of 33.0 m. The sedimentary facies is dominated by the semi-deep lake, shore lake and beach bar facies. Based on the base level cycles and sedimentological response characteristics, the Lucaogou Formation can be divided into five medium-term cycles (MSC1-MSC5) and nine short-term cycles (SSC1-SSC9) [78–79].

3.2. Single well production difference

At present, Xinjiang Oilfield Company adopts large-scale hydraulic fracturing to increase production for the oil resources in the Permian Lucaogou Formation, which uses swabbing-based production techniques for sustainable exploration and development. From the data collected in October 2020, the daily oil production of a single well in the Lucaogou Formation showed a trend of gradually increasing from east to west. The daily oil production in the P_2l_1 interval ranged from 0.50 to 13.88 t/d, with an average value of 6.08 t/d (Fig. 8a); while the daily oil production in the P_2l_2 interval varied from 0.24 to 21.53 t/d, with an average value of 6.24 t/d (Fig. 8b). The cumulative oil production of single wells in the Lucaogou Formation also presented the same trend. The cumulative oil production in the P_2l_1 interval spanned from 4.99 to 1321.22 t, with an average value of 213.01 t (Fig. 8a), while the cumulative oil production in the P_2l_2 interval was between 5.31 and

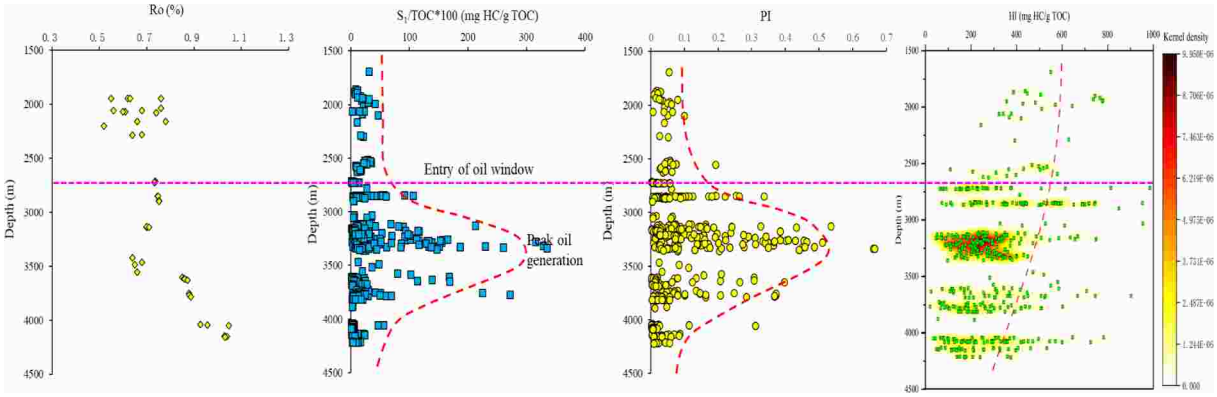


Fig. 11. The natural maturity shale sequence section of the Lucaogou Formation (TOC range 1–5 wt%).

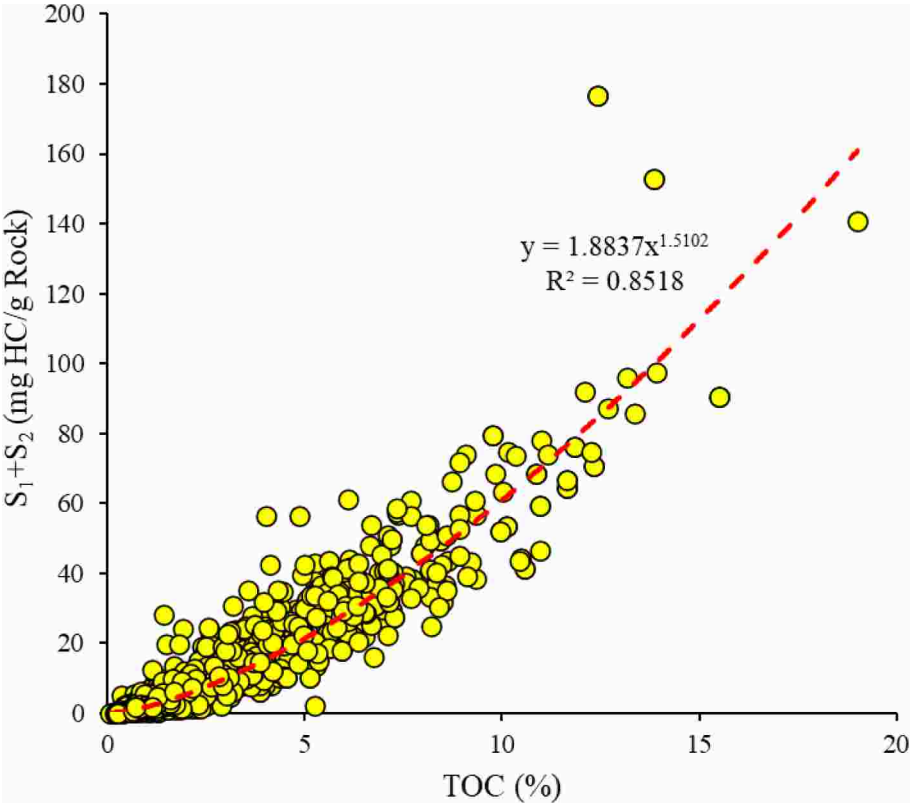


Fig. 12. The relationships between TOC and genetic potential ($S_1 + S_2$).

1080.51 t, with an average value of 232.27 t (Fig. 8b). The variation trend of the production is largely related to the hydrocarbon generation amount of the source rocks, and the burial depth gradually increases from east to west in the tectonic framework. At present, the daily oil production and cumulative oil production of a single well in the P_2l_2 interval are higher than those of the P_2l_1 interval. However, the exploration area of the P_2l_1 interval is more extensive and fewer wells have been drilled in the P_2l_1 interval, so the P_2l_1 interval has more promising exploration prospects.

4. Results and modeling process

4.1. Source-reservoir assemblages assessment

Taking the short-term cycle as the formation unit for source-reservoir assemblage classification, it can be found that there are multiple different source-reservoir assemblages in the test sections (Fig. 9). For the single source-reservoir assemblage types, the highest daily shale oil production appears in the "Highly interbedded" model (average 7.14 t/d), compared to the "Reservoir-based sandwiched" model (average 5.32 t/d) and the "Source-based sandwiched" model (average 3.53 t/d) (Fig. 9a). This trend can also be observed in the composite source reservoir assemblage types. The average daily shale oil production for the "Reservoir-based sandwiched" & "Highly interbedded" model is 8.29 t/d, compared with 7.82 t/d for the "Source-based sandwiched" & "Massive source" model and 5.24 t/d for the "Highly interbedded" & "Source-based sandwiched" model. The variations in cumulative shale oil production follow the variations in daily shale oil production, showing high average production for the "Highly interbedded" model and the "Reservoir-based sandwiched" & "Highly interbedded" model. For the single source-reservoir assemblage types, the average of cumulative shale oil production changes from 227.54 t in the "Reservoir-

Table 2
Reservoir physical property data of Lucaogou Formation in Jimusaer Sag.

Sweet spots	Porosity (%)			Permeability (mD)		
	Maximum	Minimum	Average	Maximum	Minimum	Average
Upper sweet spot	2.00	25.50	11.38	0.010	36.300	0.106
Lower sweet spot	2.00	27.40	9.66	0.010	44.900	0.041

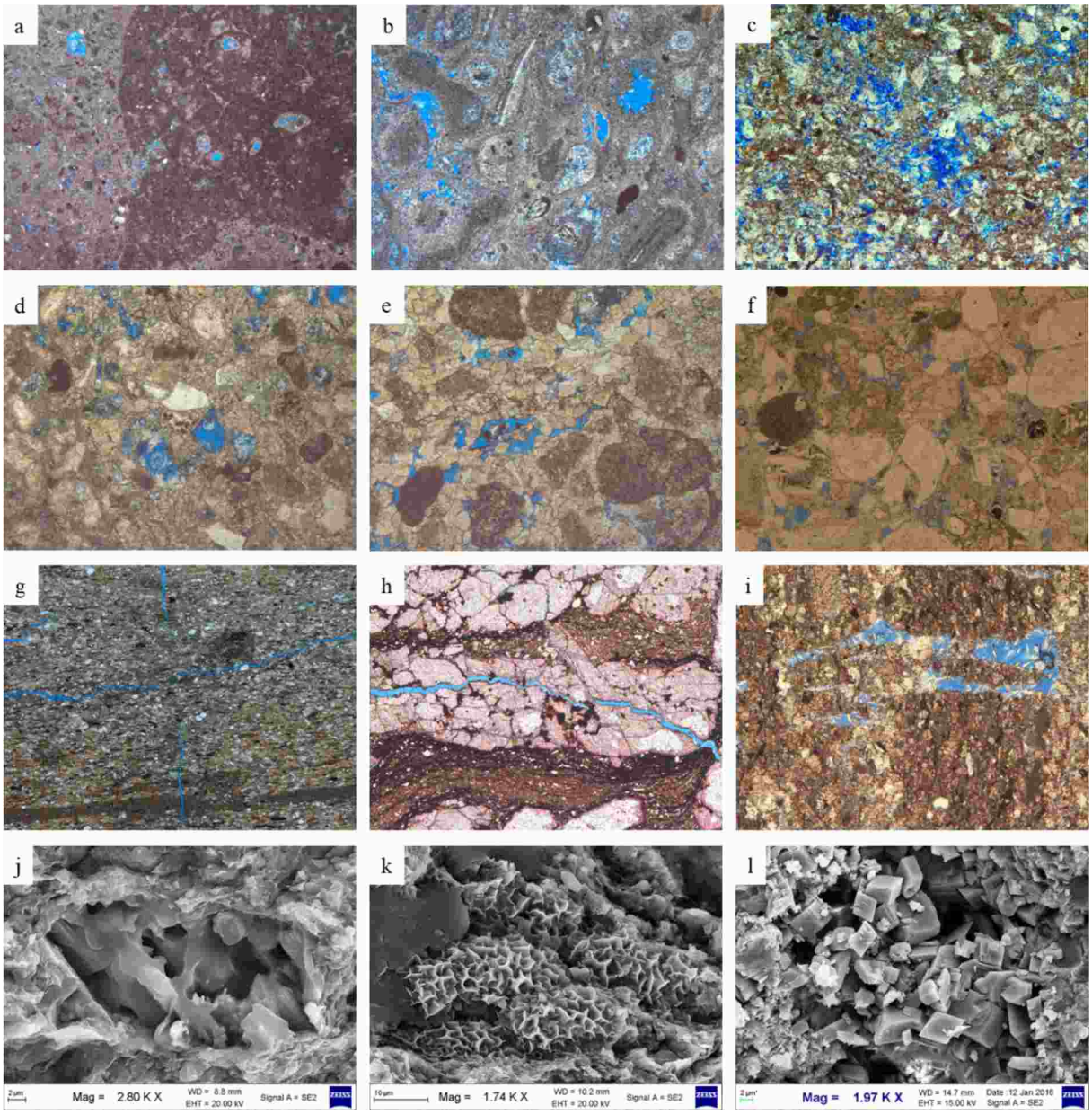


Fig. 13. Reservoir space characteristics of Lucaogou Formation in Jimusaer Sag (these pictures originate from the internal data of Xinjiang Oilfield). (a) J174, 3067.3 m, intragranular dissolved pores; (b) J174, 3180.0 m, intragranular and intergranular dissolved pores; (c) J174, 3283.7 m, intergranular dissolved pores; (d) J30, 4043.4 m, residual intergranular and intragranular dissolved pores; (e) J30, 4052.6 m, intergranular and intercrystalline pores; (f) J302, 2856.3 m, Intergranular and intragranular dissolved pores; (g) J10025, 3571.1 m, fractures; (h) J251, 3759.6 m, fractures; (i) J251, 3754.5 m, fractures and dissolved pores; (j) J10025, 3462.2 m, intragranular dissolved pores in feldspar; (k) J10025, 3491.8 m, honeycomb-like montmorillonite-illite mixed layer, intercrystalline pores; (l) J301, 2760.5 m, diamond-shaped granular dolomite grains and intercrystalline pores.

based sandwiched" model to 238.57 t in the "Highly interbedded" model, and then decreases sharply to 69.45 t in the "Source-based sandwiched" model (Fig. 9b). For the composite source-reservoir assemblage types, the maximum average cumulative shale oil production for the "Reservoir-based sandwiched" & "Highly interbedded" model is 337.61 t, followed by 319.38 t for the "Source-based sandwiched" & "Massive source" model, and then by 172.09 t for the "Highly interbedded" & "Source-based sandwiched" model. Based on this, we can conclude that whether it is a single source-reservoir assemblage or a composite source-reservoir assemblage, as the proportion of source

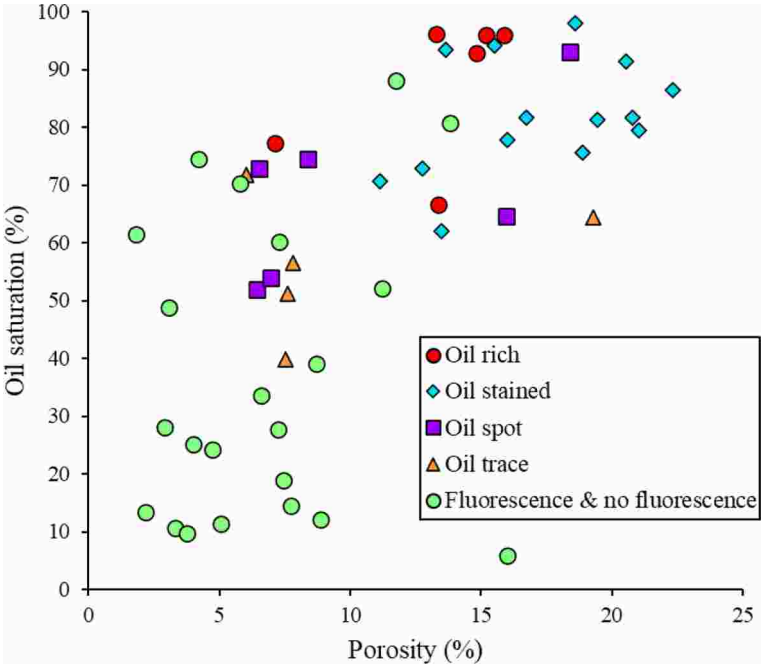


Fig. 14. The relationships between hydrocarbon shows, porosity, and oil saturation (modified from Wang et al. [125]).

rocks increases, the minimum daily production and cumulative production of shale oil increase. In terms of single source-reservoir assemblages, the "Reservoir-based sandwiched" model and the "Highly interbedded" model have the highest production. As far as the composite source-reservoir assemblage is concerned, the "Reservoir-based sandwiched" & "Highly interbedded" model have the highest production. The production of composite source-reservoir assemblages is higher than that of single source-reservoir assemblages.

4.2. Source rock quality assessment

4.2.1. Qualitative assessment

The source rocks of the Lucaogou Formation consist of dark shale, mudstone, dolomitic mudstone and lime mudstone, with abundant organic matter. The organic carbon content ranges from 0.11 to 19.01 wt%, with an average of 3.15 wt% (Fig. 10a). Approximately 55% of the samples exceed 2.0 wt%. The organic matter is a mixture of Type II_1 and II_2 kerogen (Fig. 10b). The hydrogen index varies between 23 mg HC/g TOC and 987 mg HC/g TOC with an average of 329 mg HC/g TOC. The identified type I and II kerogens indicate that the lake basin has an efficient mechanism for capturing terrestrial material, which is introduced into the main lake body and mixed with terrestrial organic matter. Actively generated and retained hydrocarbons in the Lucaogou Formation shales could be observed through elevated S_1 values. This is reflected in the anomalous production index ($S_1/S_1 + S_2$) versus T_{max} values (Fig. 10c) and the elevated S_1/TOC ratios (Fig. 10d). Interestingly, many of the samples fall within the inert carbon present area (Fig. 10c), which may be attributed to the fact that the total S_2 peak contains a small fraction of higher molecular free hydrocarbons (heavy oil residues) [80–84]. Establishing hydrocarbon generation curves for the natural maturity shale sequences requires that the samples ideally have the same organic matter composition, as the hydrocarbon generation curves depend significantly on the types of kerogen [85–87]. However, the dataset shows a large variation in organic matter. Therefore, the dataset is filtered and only the range from 1% to 5% TOC is used to build the hydrocarbon generation curves, as it is expected that narrowing the TOC range would narrow the organic matter types. Subsequently, trend lines for vitrinite reflectance, S_1/TOC, production index, and hydrogen index are constructed (Fig. 11). Fig. 11 suggests that the top of the oil window is at about 2800 m and the peak oil generation is at around 3400 m.

4.2.2. Quantitative assessment

Source rocks contain large amounts of organic matter, usually quantified as total organic carbon content "TOC", and are able to produce hydrocarbons during thermal maturation [87–89]. Therefore, the hydrocarbons produced depend on TOC, organic matter type and maturity. As shown in Fig. 12, TOC has a good exponential function relationship with genetic potential ($S_1 + S_2$), and the genetic potential increases with increasing TOC. Moreover, since the Lucaogou Formation has not undergone erosion events, the burial depth can be used to evaluate the source rock maturity. Direct TOC measurement of samples in the laboratory is accurate, but costly and time-consuming. There are many methods to realize TOC prediction [90–93], among which the ΔlogR method is the most widely applied one. In this method, the separation between porosity logging curves (e.g., sonic transit time curves) and deep resistivity logging curves is considered indicative of TOC. this paper uses the ΔlogR method, as shown in Eqs. (20) and (21).

$$\Delta\log R = \log_{10}(R_t/R_{baseline}) + 0.02 \times (\Delta t - \Delta t_{baseline}) \quad (20)$$

$$TOC = \Delta\log R \times 10^{(2.297 - 0.1688 \times LOM)} \quad (21)$$

where, ΔlogR is the separation between the resistivity log and porosity log. R_t and Δt are the resistivity (Ω•m) and sonic transit time (μs/ft) of the target formation, respectively. $R_{baseline}$ and $\Delta t_{baseline}$ are the resistivity (Ω•m) and sonic transit time (μs/ft) of the non-source rock intervals (organic-lean source rock), respectively. LOM is the level of organic maturity.

4.3. Reservoir quality assessment

4.3.1. Qualitative assessment

Based on the measured porosity and permeability of the cores, it can be seen that the porosity of the upper sweet spot ranges from 2.00 to 25.50% with an average value of 11.38%, and the permeability is between 0.010 and 36.300 mD with an average value of 0.106 mD. The porosity of the lower sweet spot varies from 2.00 to 27.40% with an average value of 9.66%, and the permeability spans from 0.010 to

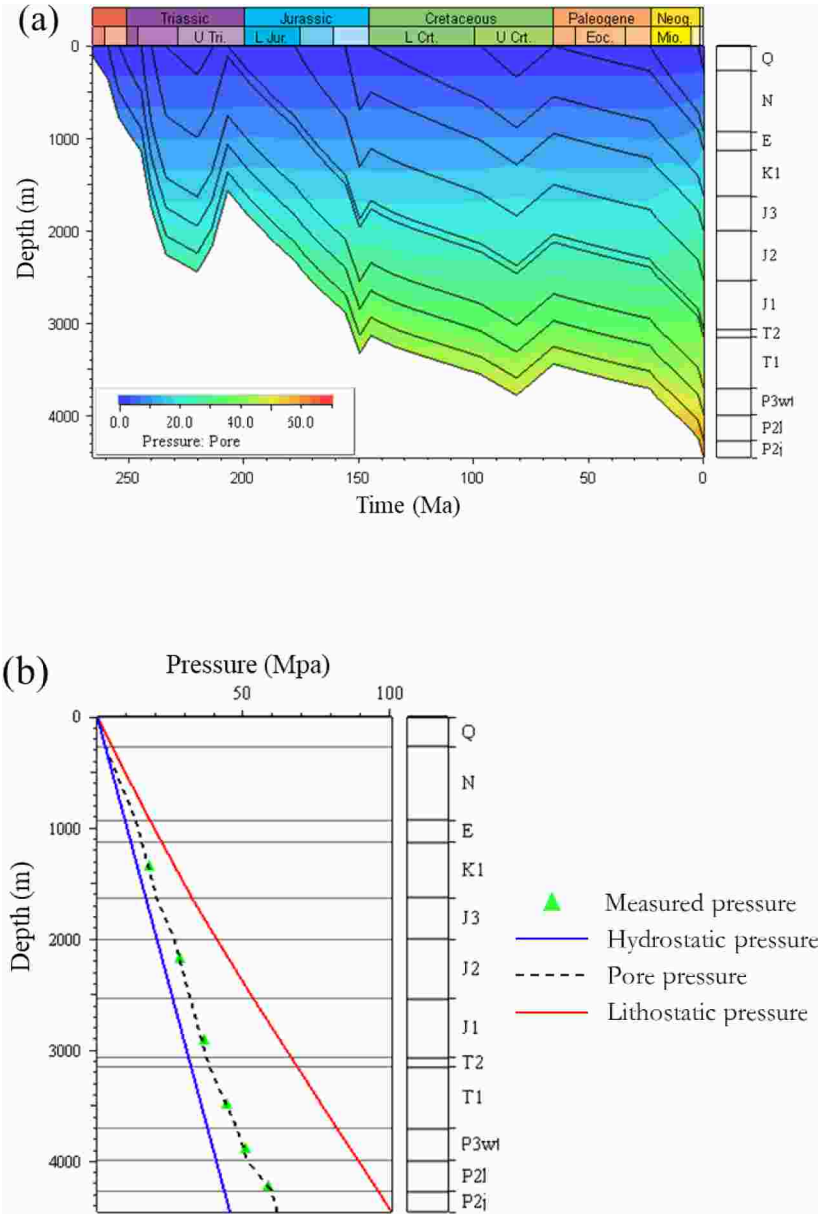


Fig. 15. Basin Model of well J30 in Jimusaer Sag. (a) The superimposed map of burial history and pore pressure; (b) The relationship between measured pressure, hydrostatic pressure, pore pressure, lithostatic pressure and depth; (c) The superimposed map of burial history and excess pressure; (d) The excess pressure evolution history of the Lucaogou Formation.

44.900 mD, with an average value of 0.041 mD (Table 2). In summary, the reservoir of the Lucaogou Formation has the characteristics of medium to low porosity and extra-low permeability. A variety of reservoir space types are observed by thin section and scanning electron microscopy (SEM) (Fig. 13). Among them, intergranular dissolved pores, intragranular dissolved pores and intercrystalline pores are dominant, followed by the residual intergranular pores. Intergranular and intragranular dissolved pores are the main types of reservoir spaces in the Lucaogou Formation and usually occur in unstable components such as feldspar, carbonates, rock debris and biotite cavities, which eventually appear as honeycomb or irregular harbor shapes (Fig. 13a-f and j). Intercrystalline pores refer usually to the pores between crystals formed during precipitation or recrystallization of mineral crystals (dolomite, clay minerals and pyrite). Intercrystalline pores in the Lucaogou Formation reservoir are mostly developed between dolomite and clay minerals (Fig. 13e and k-l). The residual intergranular pores are the native pores remaining after the mechanical compaction and cementation of the rocks, and show a scattered distribution (Fig. 13d). The fractures in the Lucaogou Formation are divided into diagenetic and tectonic fractures, which have limited and localized reservoir space, but can greatly improve the permeability of the extra-low permeability reservoir (Fig. 13g-i). Integrating with the hydrocarbon shows, it can be found that as the hydrocarbon shows better, the porosity and oil saturation are also higher (Fig. 14). Interestingly, low porosity does not imply low oil saturation. Some samples with low porosity have high oil saturation.

4.3.2. *Quantitative assessment*

There are many indicators to characterize the reservoir storage capacity (oil content), such as oil saturation (S_o), S_1 and chloroform extracts. These indicators have different geological meanings. The S_o refers to the liquid hydrocarbon content per unit pore volume [14]. The S_1 and

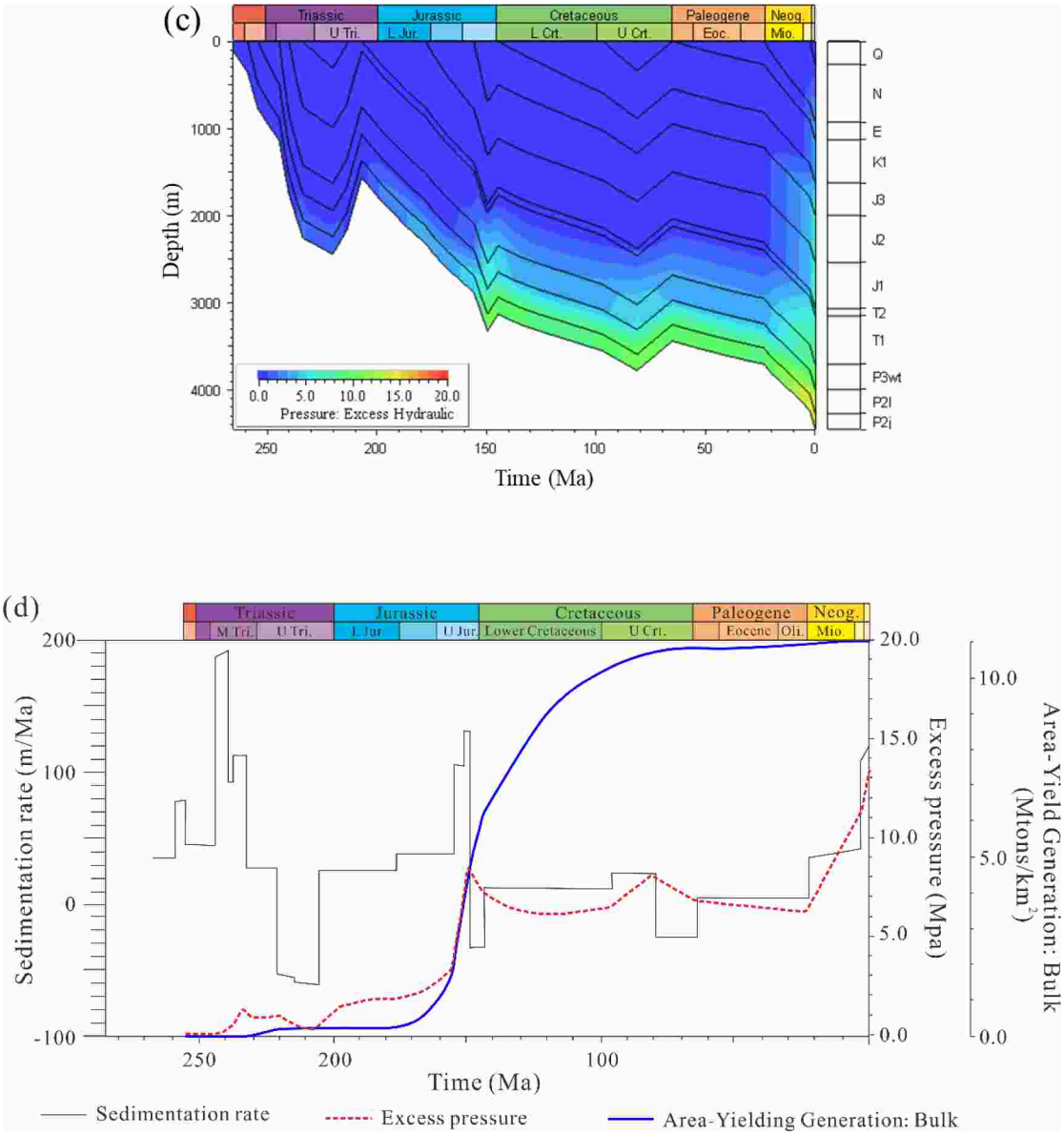


Fig. 15. (*continued*).

chloroform extracts represent the liquid hydrocarbon content per unit mass of rock [14]. As mentioned above, if an interval has very little porosity and high S_o, such an interval is not very valuable for development. Therefore, oil saturation is highly dependent on porosity. Measuring S_1 and chloroform extracts from samples directly in the laboratory is accurate, but expensive and time-consuming. Furthermore, light hydrocarbons are quickly lost during sample storage and experiments due to their low boiling point. To solve the above challenges, Liu et al. [14] build a novel method to characterize the oil content in shale oil systems based on geophysical logging data, and eventually realize the continuous evaluation of the target formation. Among them, mud gas logging is utilized to assess the light hydrocarbon content under actual geological conditions. With the advance in flame ionization detectors, mud gas logging systems can provide reliable light hydrocarbon analysis. And the equation for liquid hydrocarbons content is as follows:

$$LI = \frac{m_{HC}}{m_b} \times 100\%$$

$$= \frac{\rho_{HC} \times V_{HC}}{\rho_b \times V_b} \times 100\%$$

$$= \frac{V_{HC}}{V_{Por}} \times \frac{V_{Por}}{V_b} \times \frac{\rho_{HC}}{\rho_b} \times 100\%$$

$$= S_o \times \phi \times \frac{\rho_{HC}}{\rho_b} \times 100\% \quad (22)$$

where, *LI* is the liquid hydrocarbon content at a certain depth, %. m_{HC} and m_b are the masses of liquid hydrocarbon and rock, g. ρ_{HC} and ρ_b are the densities of liquid hydrocarbon and rock, g/cm^3. V_{HC}, V_{por} and V_b are the volumes of liquid hydrocarbon, pore and rock, cm^3. S_o is the oil saturation, %. φ is the porosity, %.

4.4. Migration dynamics conditions

4.4.1. Paleo-pressure evolution

Basin simulation is now a well-established technique to build burial history, thermal history, hydrocarbon generation history, hydrocarbon expulsion history, and paleo-pressure evolution. Integrating the geological conditions of typical wells in the Jimusaer Sag [94–95], PetroMod software is utilized to recover the pressure evolution history of the Lucaogou Formation. The results from the pressure simulation of

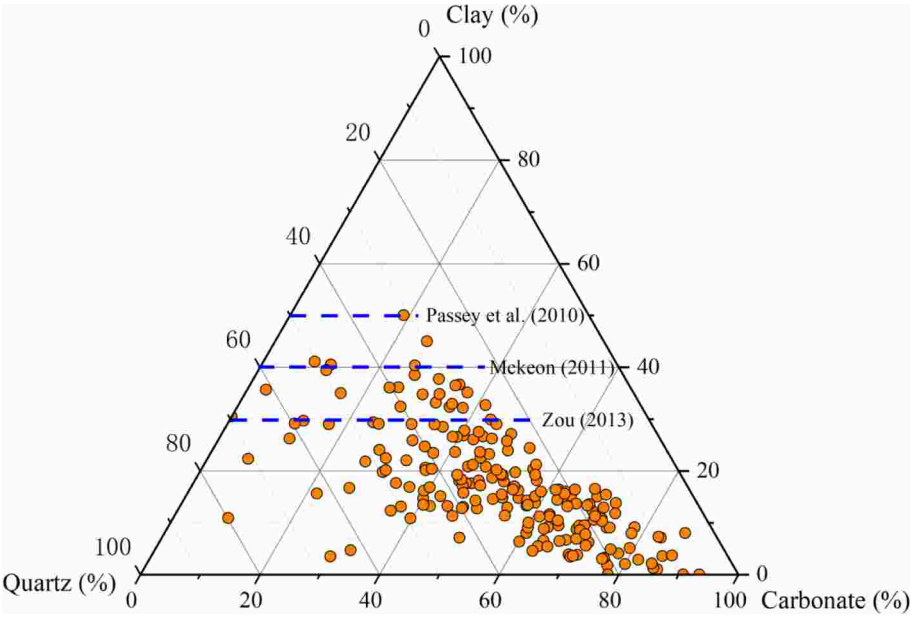


Fig. 16. Ternary diagram of relative abundance of quartz, carbonate minerals and clays in shales of the Lucaogou Formation, Jimusaer Sag (modified from Katz et al. [126]).

Table 3
The published relationships/models used to estimate BI based on mineralogy (modified from Gogoi et al. [127]).

Parameter	Description	Reference
$BI_1 = \frac{w_{Qtz}}{w_T}$	w_{Qtz} is the weight of the quartz fraction. w_T is the total weight of the mineral.	Jarvie et al. [3]
$BI_2 = \frac{w_{Qtz} + w_{Dol}}{w_T}$	w_{Dol} is the weight of the dolomite fraction.	Wang and Gale [128]
$BI_3 = \frac{w_{Qtz} + w_F + w_M + w_{Cal} + w_{Dol}}{w_T}$	w_F, w_M and w_{Cal} are the weight fraction of feldspar, mica and calcite.	Jin et al. [129]
$BI_4 = \frac{w_{Qtz} + +w_{Cal} + w_{Al}}{w_{Qtz} + +w_{Cal} + w_{Al} + w_{Mus} + w_{Cl}}$	w_{Al}, w_{Mus} and w_{Cl} are the weight fraction of albite, muscovite and clay.	Gholami et al. [130]; Glorioso et al. [131]; Guo et al. [132]; Kia et al. [133]; Lai et al. [134]

well J30 suggest that the pore pressure increases with the increase of burial depth (Fig. 15a), and different levels of excess pressure are developed in different formations (Fig. 15b and c). Together with the evolution history of the excess pressure in the Lucaogou Formation (Fig. 15d), the evolution of the excess pressure can be divided into three complete pressurization-release cycles as well as one pressurization cycle. From the deposition of the Lucaogou Formation to the Late Triassic, the Lucaogou Formation developed weak overpressure with an excess pressure of less than 5 MPa, followed by the first pressure release by tectonic uplift under the influence of the Indochina movement; Subsequently, the Lucaogou Formation continued to subside, and the Lucaogou Formation began to generate a large amount of hydrocarbons in the Middle Jurassic, besides being influenced by the under-compaction, and the excess pressure reached nearly 10 MPa in the Late Jurassic, followed by the second pressure release under the influence of the Late Jurassic-Early Cretaceous 2nd tectonic uplift; During the Cretaceous period, the sedimentation rate of Lucaogou Formation was low, and the overpressure caused by hydrocarbon generation was not obvious. The excess pressure of the Lucaogou Formation reached 9 Mpa in the Late Cretaceous, and then experienced the third pressure release due to the tectonic uplift; Since the Paleogene, with the continuous increase of the burial depth of the Lucaogou Formation, especially in the Neogene, the excess pressure of the Lucaogou Formation has gradually increased, and now the excess pressure has reached the maximum value (14.7Mpa). Thus, the excess pressure is now in a process of pressurization, and the accurate prediction of the excess pressure is beneficial to determine the dynamics of hydrocarbon migration.

4.4.2. *Estimation of excess pressure and pressure coefficient*

Pore pressure is the pressure of fluid trapped in the pore spaces of a rock [96]. An accurate understanding of pore pressure has significant implications for safe drilling and fluid migration. Geologically, porosity is exponentially decreasing with depth in normally compacted sedimentary sequences [97]. Hence, deviations from the normal compaction trend (NCT) may indicate the presence of abnormal pressures. Pore pressure is predicted in shale formations due to the significant variation of petrophysical properties of shales with depth. Furthermore, pore pressure prediction in shale formations will provide an early warning of abnormal pore pressures in the underlying reservoirs before drilling into them. Eaton [98] proposed pore pressure prediction models based on resistivity and sonic transit time logging (Eqs. (24) and (25), respectively). Eaton's model is one of the most widely applied pore pressure prediction methods, calibrated with measured pore pressure data.

For resistivity and sonic transit time logging, the equations for estimating pore pressure are as follows:

$$S_v = \int_0^H \rho_b(H) g dH \tag{23}$$

where, S_v is the vertical stress, Mpa. $\rho_b(H)$ is the rock density at depth (H), g/cm^3. g is the gravitational acceleration, assumed as 9.8 m/s^2.

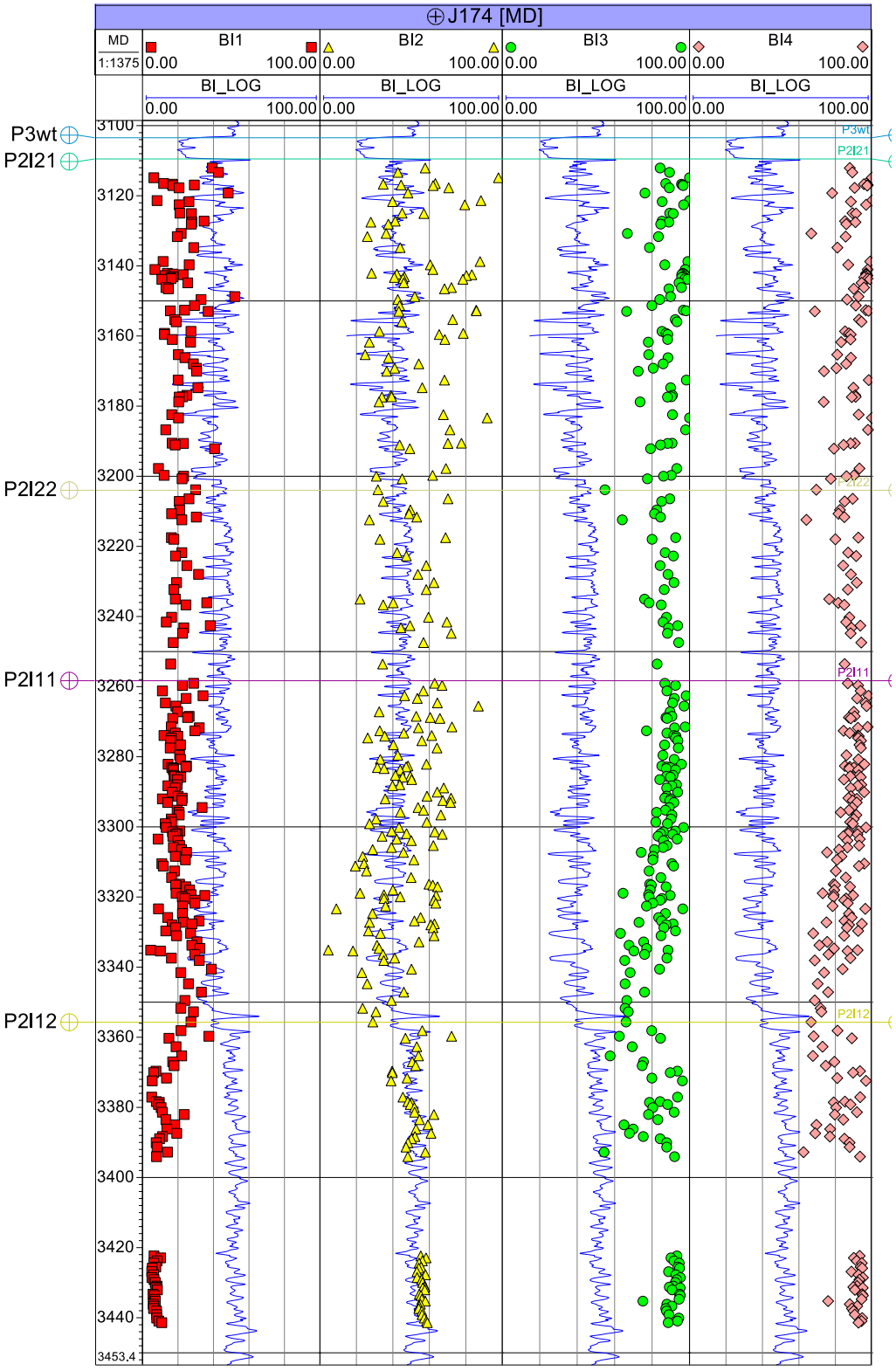


Fig. 17. The superimposed map of the BI (BI_LOG) based on Young's modulus and Poisson's ratio, and the BI (BI1, BI2, BI3 and BI4) obtained using mineralogy. The calculation formulas of BI1, BI2, BI3 and BI4 are shown in Table 3.

$$PP = S_v - (S_v - P_{hyd}) * \left(\frac{R_t}{R_{NCT}}\right)^{1.2} \quad (24)$$

$$PP = S_v - (S_v - P_{hyd}) * \left(\frac{\Delta t}{\Delta t_{NCT}}\right)^{3} \quad (25)$$

where, PP is the formation pore pressure, Mpa. P_{hyd} is the hydrostatic pressure, MPa. R_{NCT} and Δt_{NCT} are the formation true resistivity and the sonic transit time in the shales calculated from the normal compaction trend (NCT), Ω•m, μs/ft, respectively. R_t is the formation resistivity at a certain depth, Ω•m. Δt is the sonic transit time at a certain depth, μs/ft.

To better assess the degree of overpressure, this article calculates the

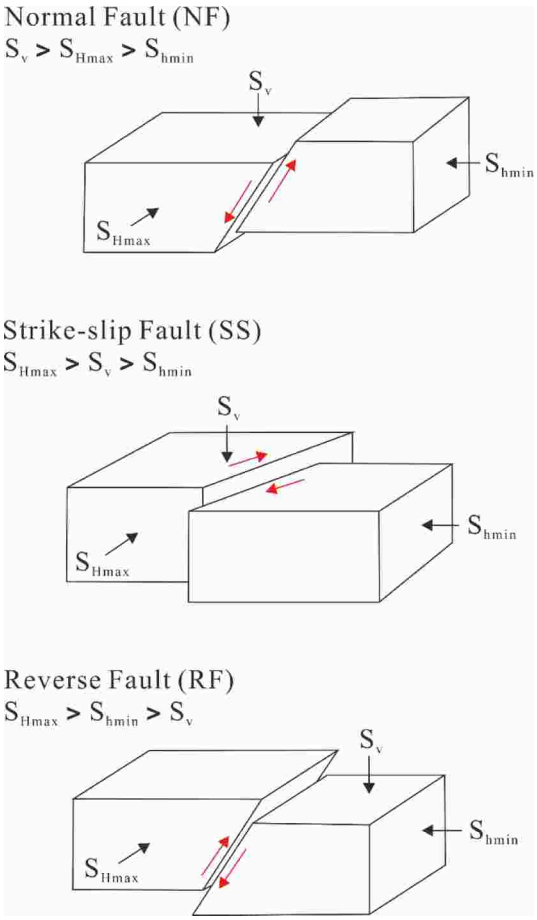


Fig. 18. Anderson fault types and their stress regimes.

excess pressure and pressure coefficient, as shown in Eqs. (26) and (27).

$$EP = PP - P_{hyd} \tag{26}$$

$$PC = PP/P_{hyd} \tag{27}$$

where, *EP* is the excess pressure, Mpa. *PC* is the pressure coefficient.

4.5. Conduit conditions

4.5.1. Brittleness index calculation

Due to the low permeability characteristics of tight rock plays, the most effective method to obtain commercial production rates is artificial fracturing. Therefore, the brittleness or fracture potential of the rock is a very critical prerequisite. Passey et al. [99] considered that sufficient brittleness in the producing fields is obtained at clay contents below 50 wt%. Other scholars have different views on clay mineral thresholds. Mckeon [100] suggested a clay content below 40 wt%, while Zou [101] suggested a clay content below 30 wt%. The clay content of the Lucaogou Formation shales is highly variable (Fig. 16). The clay content of most of these samples is less than 40%, which is favorable for large-scale artificial fracturing. High brittleness formations are generally considered targets for hydraulic fracturing, and therefore brittleness prediction needs to be realized for single wells. In general, Young's modulus (E) and Poisson's ratio (υ) are the most popular parameters to characterize the brittleness of rocks. Rickman et al. [102] proposed equations for the shale brittleness index based on these parameters, which can be expressed as:

$$E = \rho_b V_S^2 \frac{3V_P^2 - 4V_S^2}{V_P^2 - V_S^2} \tag{28}$$

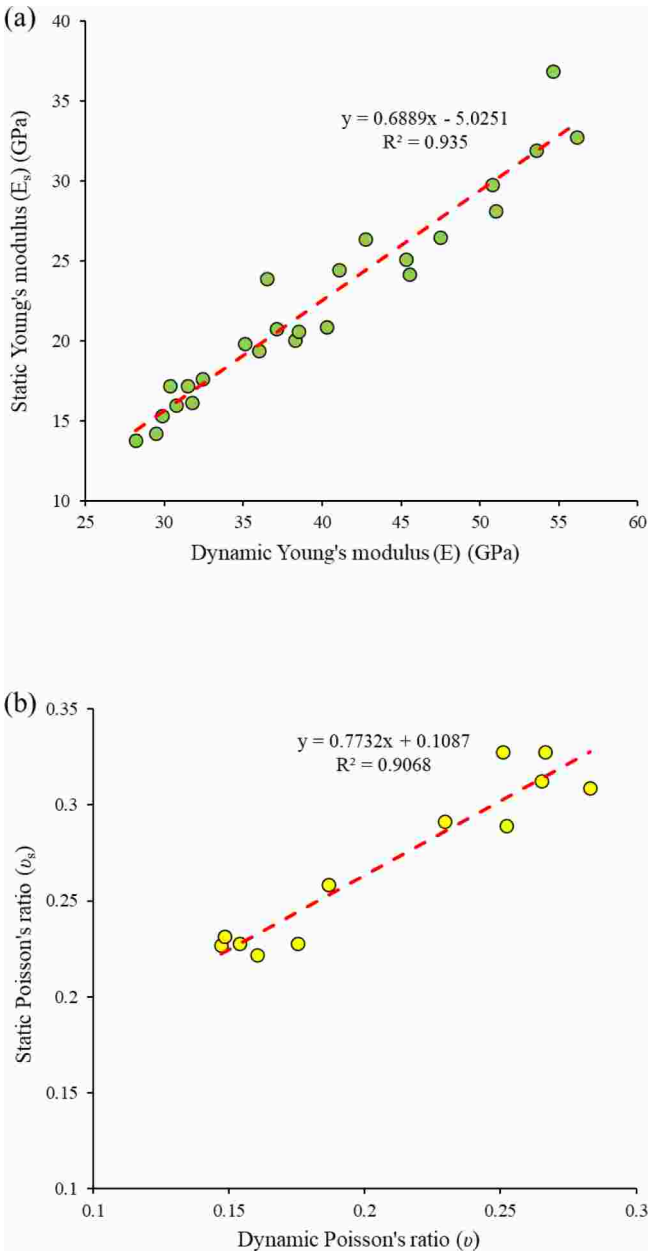


Fig. 19. The relationship between dynamic elastic moduli parameters and static elastic moduli parameters. (a) Dynamic Young's modulus (E) versus Static Young's modulus (E_s); (b) Dynamic Poisson's ratio (υ) versus Static Poisson's ratio (υ_s).

$$\upsilon = \frac{V_P^2 - 2V_S^2}{2V_P^2 - 2V_S^2} \tag{29}$$

$$E_{BI} = \frac{E - E_{min}}{E_{max} - E_{min}} \tag{30}$$

$$\upsilon_{BI} = \frac{\upsilon - \upsilon_{min}}{\upsilon_{max} - \upsilon_{min}} \tag{31}$$

$$BI = \frac{E_{BI} + \upsilon_{BI}}{2} \tag{32}$$

where, V_P and V_S are the compressional and shear sonic wave velocities, m/s, respectively. ρ_b is the rock density, g/cm^3. E_{max} and E_{min} are the maximum and minimum values of Young's modulus of the rocks, GPa, respectively. υ_{max} and υ_{min} are the maximum and minimum values

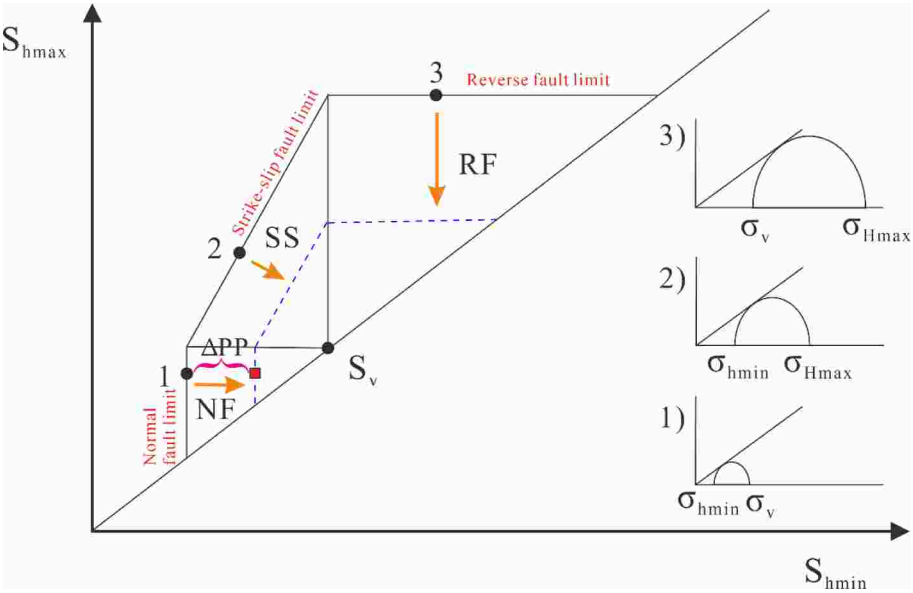


Fig. 20. Schematic diagram of stress polygon constrained rock mass (modified from Zoback et al. [113]).

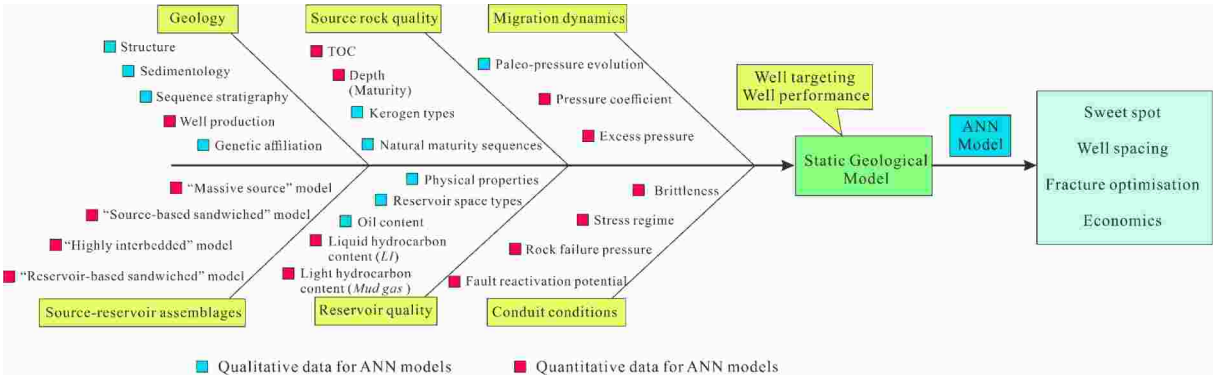


Fig. 21. Fishbone diagram for well targets assessment.

of Poisson's ratio of the rocks, dimensionless. BI is the brittleness index of the rocks.

Generally, the BI calculated based on Young's modulus and Poisson's ratio is different from the BI obtained using mineralogy. Various scholars have proposed different mineral assemblages under the brittleness and ductility classifications to calculate the BI, and these methods are summarized in Table 3. Fig. 17 presents the mineralogy-derived BI (BI1, BI2, BI3 and BI4) values are superimposed on the logging-derived BI (BI_LOG) data in well J174. It can be observed that BI_LOG has a good match with BI2 (Fig. 17). It is noticed that the mineralogy-derived BI values tend to be higher than the BI_LOG values. This is due to the fact that in-situ formation conditions and fluid properties are not accounted for in the mineralogy-derived BI calculations. Instead, the BI based on Young's modulus and Poisson's ratio fully considers the mineral composition, pore, fluid properties and in-situ formation conditions. In summary, the BI calculated based on Young's modulus and Poisson's ratio are good predictors of the weight percent of quartz and dolomite in the total minerals.

4.5.2. Rock fracture pressure

Fracture pressure is the pressure required to induce fractures in the formation. In hydrocarbon exploration and development activities, obtaining accurate fracture pressure information is critical to maintaining control of the wellbore at all times. Accurate assessment of fracture pressure is a critical parameter for safe drilling, fluid design, casing emplacement, enhanced wellbore stability and hydraulic fracturing optimization [103–106]. The Eaton fracture method is widely used worldwide for the estimation of fracture pressures [107]. Based on Poisson's ratio, the equation is as follows:

$$P_{FP} = \frac{\upsilon}{1-\upsilon}(S_v - PP) + PP \tag{33}$$

where, P_{FP} is the fracture pressure, Mpa. υ is the Poisson's ratio, dimensionless. S_v is the vertical stress, Mpa. PP is the formation pore pressure, Mpa.

To better evaluate the ease of rock fracture, the difference between the fracture pressure and the pore pressure is calculated in this paper, as shown in Eq. (34).

$$P_{FP\_PP} = P_{FP} - PP \tag{34}$$

where, P_{FP_PP} is the difference between the fracture pressure and the pore pressure, Mpa.

4.5.3. Stress regime

Geostress is the internal stress within the Earth's crust and is strongly correlated with gravitational and tectonic stresses [108]. Typically, the viability of unconventional resource production depends heavily on artificial fracturing. The ability of these fractures to transmit fluids relies in part on the stress regimes. Moreover, the understanding of the stress regimes allows for better exploration and development of

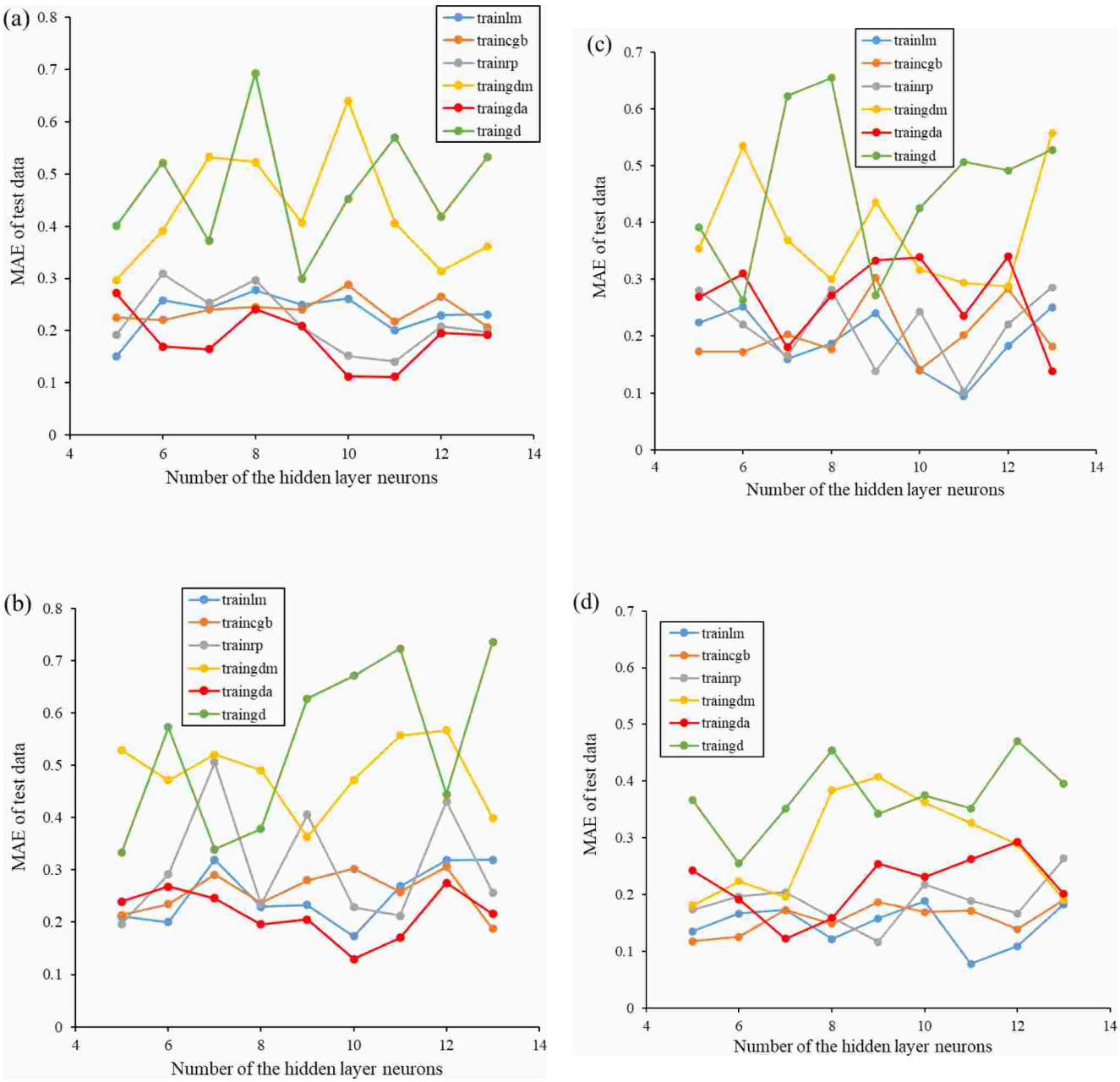


Fig. 22. MAE of shale oil productivity predictions for test data with different training functions and different averaging algorithm models. (a) The quadratic average algorithm model; (b) The arithmetic average algorithm model; (c) The geometric average algorithm model; (d) The harmonic average algorithm model.

Table 4
Training results for different hidden layer nodes in the different average algorithm models.

Different average algorithms	Number of hidden layer nodes	5	6	7	8	9	10	11	12	13
Quadratic average algorithm	MAE	0.2716	0.1697	0.1644	0.2407	0.2091	0.1123	0.1116	0.1952	0.1920
	MSE	0.0941	0.0456	0.0468	0.0808	0.0673	0.0182	0.0211	0.0476	0.0817
	RMSE	0.3067	0.2135	0.2162	0.2842	0.2594	0.1349	0.1453	0.2182	0.2858
Arithmetic average algorithm	MAE	0.2394	0.2679	0.2458	0.1959	0.2056	0.1296	0.1702	0.2755	0.2159
	MSE	0.1188	0.0834	0.0722	0.0435	0.0622	0.0424	0.0451	0.0995	0.0679
	RMSE	0.3447	0.2887	0.2687	0.2087	0.2494	0.2059	0.2123	0.3155	0.2606
Geometric average algorithm	MAE	0.2237	0.2520	0.1594	0.1869	0.2409	0.1400	0.0943	0.1829	0.2504
	MSE	0.0899	0.0902	0.0319	0.0607	0.0866	0.0307	0.0102	0.0511	0.0996
	RMSE	0.2999	0.3003	0.1785	0.2464	0.2943	0.1751	0.1008	0.2261	0.3157
Harmonic average algorithm	MAE	0.1355	0.1668	0.1732	0.1220	0.1580	0.1891	0.0779	0.1093	0.1835
	MSE	0.0270	0.0402	0.0345	0.0199	0.0385	0.0515	0.0087	0.0185	0.0479
	RMSE	0.1643	0.2005	0.1857	0.1410	0.1962	0.2270	0.0933	0.1359	0.2190

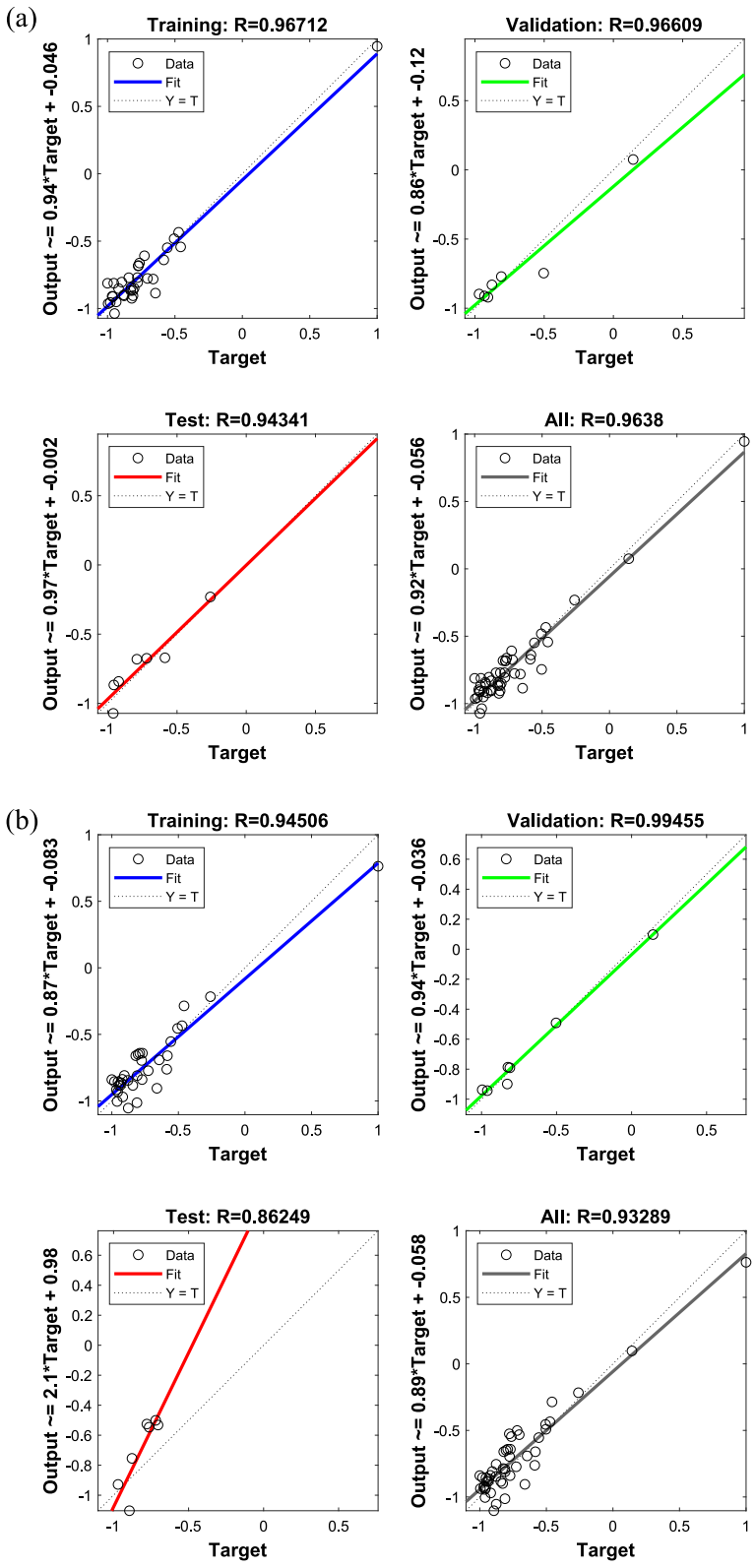


Fig. 23. Correlation coefficients of shale oil productivity under different averaging algorithm models. (a) The quadratic average algorithm model; (b) The arithmetic average algorithm model; (c) The geometric average algorithm model; (d) The harmonic average algorithm model.

unconventional resources, wellbore stability, enhanced oil recovery techniques, and reservoir management [109–113]. Based on Anderson's fault theory [114], three types of stress regimes are determined using the relative magnitudes of the maximum horizontal stress (S_{Hmax}), minimum horizontal stress (S_{hmin}) and vertical stress (S_v): (1) Normal fault stress regime ($S_v > S_{Hmax} > S_{hmin}$); (2) Strike-slip fault stress regime ($S_{Hmax} > S_v > S_{hmin}$); (3) Reverse fault stress regime ($S_{Hmax} > S_{hmin} > S_v$) (Fig. 18). Since the calculation process of S_v has been described above,

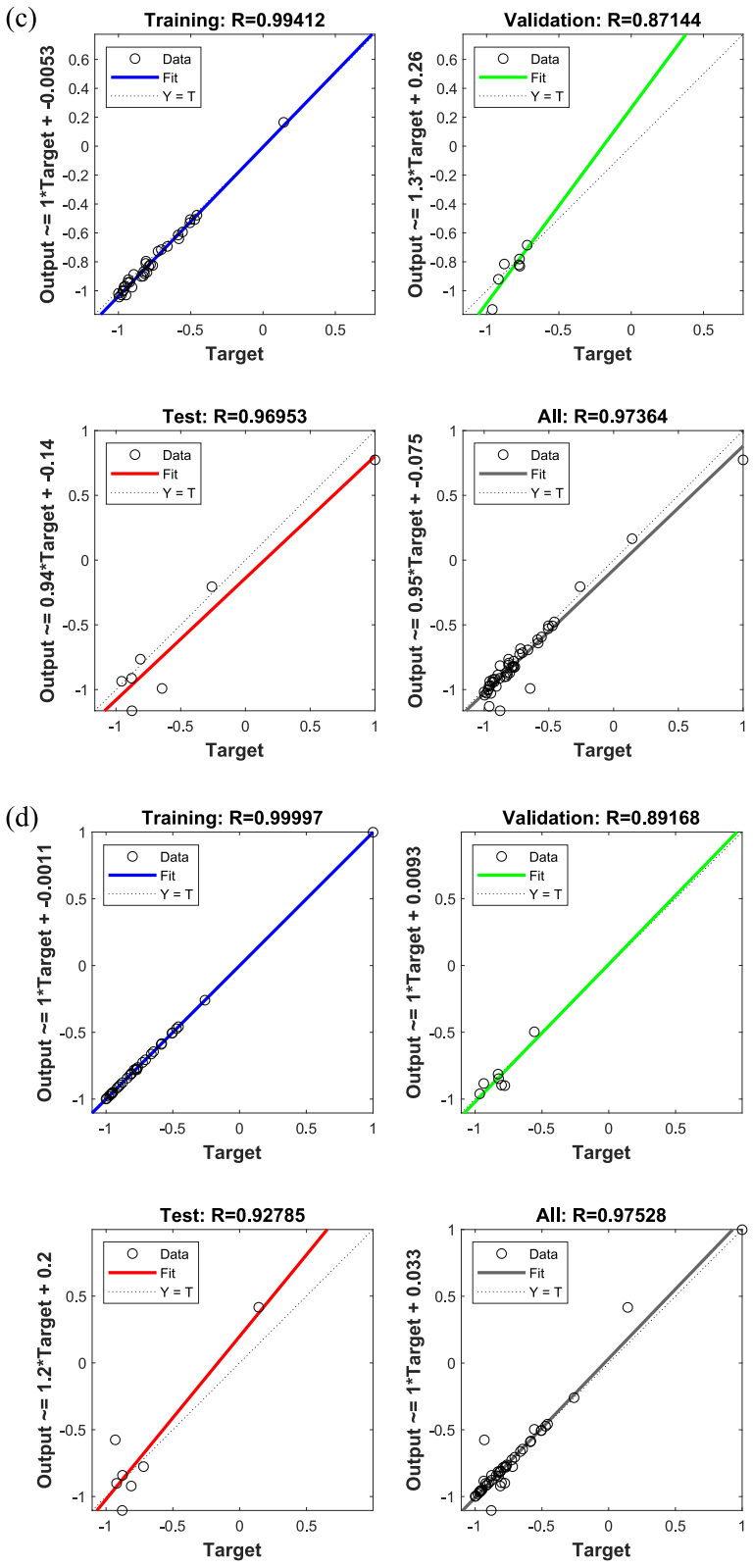


Fig. 23. (*continued*).

the methods and workflows used to calculate S_{Hmax} and S_{hmin} are highlighted in this section. In this paper, we adopt the poroelastic horizontal strain model to determine the S_{hmin} magnitude. This model assumes that the tectonic stresses in the formation are anisotropic in all directions [115–117]. The equation for estimating S_{hmin} using the poroelastic strain model is as follows:

$$S_{hmin} = \frac{\upsilon_s}{1-\upsilon_s}(S_v - \alpha PP) + \alpha PP + \frac{E_s}{1-\upsilon_s^2}\varepsilon_h + \frac{E_s\upsilon_s}{1-\upsilon_s^2}\varepsilon_H \tag{35}$$

$$\varepsilon_h = S_v\frac{\upsilon_s}{E_s}\left(\frac{1}{1-\upsilon_s} - 1\right) \tag{36}$$

Table 5
Correlation coefficients of shale oil productivity under different averaging algorithms.

Different averaging algorithms	Training	Validation	Test	All
Quadratic average algorithm	0.96712	0.96609	0.94341	0.9638
Arithmetic average algorithm	0.94506	0.99455	0.86249	0.93289
Geometric average algorithm	0.99412	0.87144	0.96953	0.97364
Harmonic average algorithm	0.99997	0.89168	0.92785	0.97528

$$\varepsilon_H = S_v \frac{\upsilon_s}{E_s}\left(1 - \frac{\upsilon_s^2}{1-\upsilon_s}\right) \tag{37}$$

$$\alpha = 1 - \exp\left(7.5\tan\frac{\phi\pi}{2}\right) \tag{38}$$

where, υ_s is the static Poisson's ratio, dimensionless. E_s is the static Young's modulus, N/m^2. α is Biot's effective-stress coefficient, dimensionless. ε_h and ε_H are two horizontal strain components along the S_{hmin} and S_{Hmax} directions. φ is the porosity, %.

Young's modulus and Poisson's ratio calculated based on V_p, V_s and ρ_b are dynamic elastic moduli parameters (E and υ). Young's modulus and Poisson's ratio are static elastic moduli parameters based on the studies of core deformation in the laboratory (E_s and υ_s). In this paper, based on the actual core measurement data and logging data (Fig. 19a and b), the relationship equations of dynamic elastic parameters and static elastic parameters are established, which are as follows:

$$E_s = 0.6889E - 5.0251 \tag{39}$$

$$\upsilon_s = 0.7732\upsilon + 0.1087 \tag{40}$$

Similarly, we still use the poroelastic strain model to estimate S_{Hmax} as follows:

$$S_{Hmax} = \frac{\upsilon_s}{1-\upsilon_s}(S_v - \alpha PP) + \alpha PP + \frac{E_s}{1-\upsilon_s^2}\varepsilon_H + \frac{E_s\upsilon_s}{1-\upsilon_s^2}\varepsilon_h \tag{41}$$

The parameters input in Eq. (41) are described above.

4.5.4. Fault reactivation potential

Zoback et al. [106,113] built the stress polygon model by combining the Anderson fault theory with the Mohr-Coulomb failure criterion. It is assumed that the stress regime of the rock mass is limited by the Mohr-Coulomb failure criterion, which means that fault surface sliding occurs when the ratio of the maximum to minimum effective principal stresses exceeds a particular value defined by the fault frictional coefficient (μ). The relevant formula is as follows:

$$\frac{\sigma_1}{\sigma_3} = \frac{S_1 - PP}{S_3 - PP} = \left[\left(\mu^2+1\right)^{0.5} + \mu\right]^2 \tag{42}$$

where, σ_1 and σ_3 are the maximum and minimum effective principal stresses, MPa, respectively. S_1 and S_3 are the maximum and minimum principal stresses, MPa, respectively. PP is the pore pressure, MPa. μ is the fault frictional coefficient, and the average value of μ is assigned to 0.6.

Considering the variability of stress regimes for different fault types, Eq. (42) can be further converted to:

(1) Normal Fault (NF):

$$\frac{S_v - PP}{S_{hmin} - PP} \leqslant \left[\left(\mu^2+1\right)^{0.5} + \mu\right]^2 \tag{43}$$

(2) Strike-slip Fault (SS):

$$\frac{S_{Hmax} - PP}{S_{hmin} - PP} \leqslant \left[\left(\mu^2+1\right)^{0.5} + \mu\right]^2 \tag{44}$$

(3) Reverse Fault (RF):

$$\frac{S_{Hmax} - PP}{S_v - PP} \leqslant \left[\left(\mu^2+1\right)^{0.5} + \mu\right]^2 \tag{45}$$

The parameters input in Eqs. 43–45 are as stated above.

The stress polygon delineates the range of possible stress magnitudes at a specific depth and pore pressure for a specific frictional coefficient, as shown in Fig. 20. In the stress regime outside the stress polygon region, the rock mass will be in an unstable state under natural conditions. Therefore, we use the stress polygon to calculate how much additional pore pressure is required to reactivate the fault. For example, the stress regime of a rock mass at a certain depth is shown as the red box in Fig. 20, which requires additional pore pressure of ΔPP for fault reactivation. In this paper, we calculate the ΔPP of each rock mass to quantitatively characterize the fault reactivation potential.

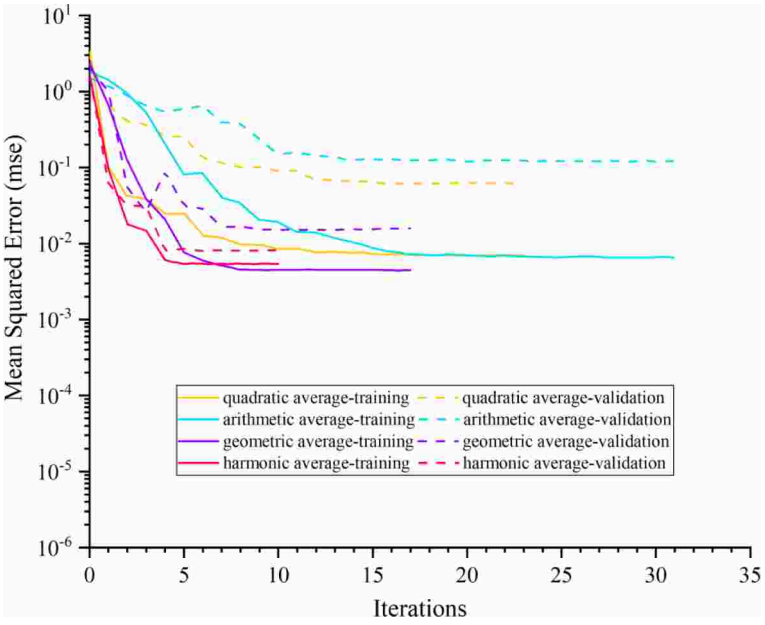


Fig. 24. Training and validation loss curves for four different reservoir heterogeneity characterization strategies.

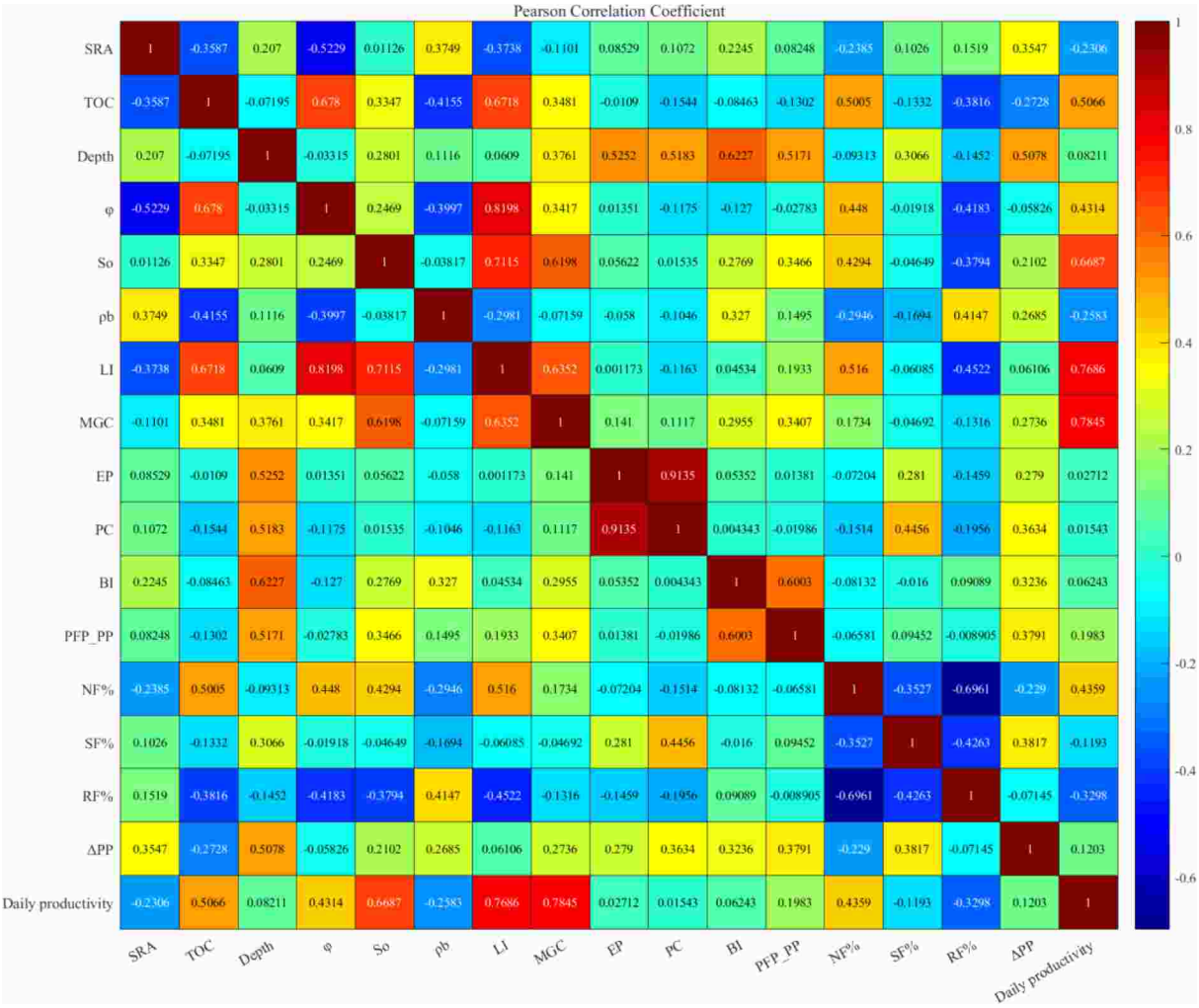


Fig. 25. The Pearson correlation coefficient heatmap showing the relationship between the 17 variables. Key: SRA = source reservoir assemblages; φ = porosity; So = oil saturation; ρ_b = the densities of rock; LI = liquid hydrocarbon content; MGC = mud gas content; EP = excess pressure; PC = pressure coefficient; BI = brittleness index; PFP_PP = rock fracture pressure; NF% = normal faults percentage; SF% = strike-slip faults percentage; RF% = reverse faults percentage; ΔPP = fault reactivation pressure.

4.6. Development of ANN prediction model

In this paper, there are quantitative datasets of 48 test sections as input data for the ANN prediction model. When the BP neural network algorithm is applied for training, the initial data set is randomly divided into training, testing and validation samples in the ratio of 7:1.5:1.5. Fig. 21 shows a fishbone diagram of the model building process. Previous studies have shown that the main factors for shale oil production (sweet spot) prediction are mostly subjective and unsystematic, which may introduce large errors in model building. In this study, these data are systematically qualitative to quantitative evaluation from geology, source-reservoir assemblages, source rock quality, reservoir quality, migration dynamics and conduit conditions. Then, based on the ANN model, a prediction model is established through repeated training. Ultimately, better recommendations are provided for the sweet spot, well spacing, fracture optimization and economics. The maximum number of training times for the target is set to 1000, the minimum gradient is 1×10^{-6}, and the learning rate is 0.001. Model validation is an important step to verify the validity of the model. This paper uses mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE) to calculate the error between the predicted and actual data in the testing dataset. These error calculation expressions are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\text{Actual}_i - \text{Predicted}_i| \tag{46}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i - \text{Predicted}_i)^2 \tag{47}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i - \text{Predicted}_i)^2} \tag{48}$$

where, $Actual_i$ is the i-th actual shale oil production data, t/d•m. $Predicted_i$ is the i-th predicted shale oil production data, t/d•m.

As you know, ANN training is nothing more than a black box training. Therefore, the pretrained models are only useful if they are cross-checked with the simulation models for sensitivity analysis. The k-fold cross-validation technique is one of the most commonly used approaches to evaluate the performance of machine learning models and model selection [118]. In this method, the initial data set is randomly divided into k subsets of equal size. Each subset is made a test set once separately, while the remaining k-1 subsets are used as training sets. The previous process is further repeated k times to ensure that each subset is validated once. Finally, the results of the k validations are averaged and used as the sensitivity analysis of this model [119–120]. The cross-validation effectively utilizes all available data and avoids over- and

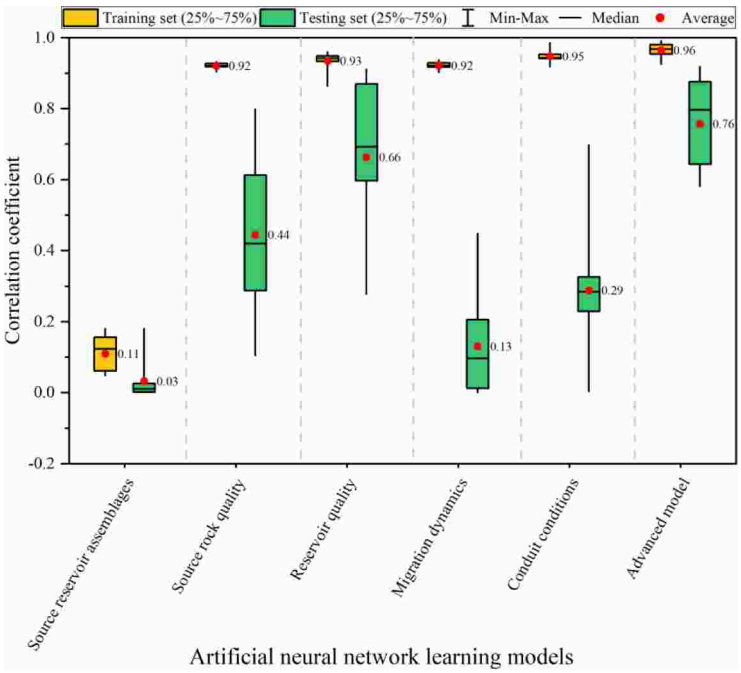


Fig. 26. Box plots of 8 times 8-fold cross-validation results for 6 reservoir characterization element models on the dataset of this study.

under-learning states, and the results obtained are more convincing. The stability and accuracy of the evaluation results of the cross-validation technique are highly dependent on the k-values. To minimize this effect, multiple times k-fold cross-validation tests are employed [121].

5. Discussion

5.1. Comparison of different reservoir heterogeneity models

Considering the thickness difference of the test sections, we divide the shale oil production by the thickness to obtain the shale oil productivity as the output layer. And the input layer includes 13 variables, namely source reservoir assemblages, TOC, depth, liquid hydrocarbon content (*LI*), mud gas content, excess pressure (*EP*), pressure coefficient (*PC*), brittleness index (*BI*), rock fracture pressure (P_{FP_PP}), percentage of normal faults, percentage of strike-slip faults, percentage of reverse faults, and fault reactivation pressure (ΔPP). The learning rule of the BP neural network algorithm is to use the fastest descent technique to continuously change the weights and thresholds of the network by back propagation to reduce the sum of squared errors of the network. For different domains, there are various training learning methods for BP neural networks, including *trainlm*, *traincgb*, *trainrp*, *traingdm*, *traingda* and *traingd*. In this paper, we determine the optimal training function by comparing the MAE of test data under different training functions and different averaging algorithms. As shown in Fig. 22, different average algorithm models have different optimal training functions. Among them, the *traingda* function is the optimal training function for the quadratic and arithmetic average algorithm models (Fig. 22a and b), while the *trainglm* function is the optimal training function for the geometric and harmonic average algorithm models (Fig. 22c and d). Based on the above quantitative characterization of the oil test section, the number of nodes in the hidden layer is between 4 and 14 due to the number of nodes in the input layer being 13 and the number of nodes in the output layer being 1. In this paper, we use MATLAB R2018a to determine the number of optimal hidden layer nodes. MAE, MSE and RMSE are regarded as indicators of prediction accuracy by training different training functions, number of nodes in the hidden layer and averaging algorithms. As shown in Table 4, the optimal number of nodes for the hidden layer is 10 in the quadratic and arithmetic average algorithm models, while the optimal number of nodes for the hidden layer is 11 in the geometric and harmonic average algorithm models.

The correlation coefficient (R) measures the correlation between the predicted and actual values. When R is close to 1, it indicates a strong correlation between them. Shale oil productivity prediction models with different averaging algorithms were trained, and the results are shown in Fig. 23. Table 5 shows the correlation coefficients of the different averaging algorithm models in the training stage. It can be observed that the training results of different average algorithm models are satisfactory with high correlation coefficients. Among the shale oil productivities analyzed, the overall correlation coefficients were 0.9638, 0.93289, 0.97364 and 0.97528 for the quadratic average, arithmetic average, geometric average and harmonic average algorithm models, respectively. The correlation coefficients of the training and validation data for all models exceeded 0.87. As shown in Fig. 23, the black dashed line represents that the predicted and actual values are equal. Based on the distribution of points and correlation coefficients, the geometric average algorithm and the harmonic average algorithm are preferentially recommended as a characterization of reservoir heterogeneity. In addition, the performance of ANN can be assessed by the speed of model convergence and the training and validation loss curves of the model. Fig. 24 illustrates the training and validation loss curves for four different reservoir heterogeneity characterization strategies. Among all strategies, the harmonic and geometric averaging algorithms perform the best: they converge quickly and have the lowest training and validation losses. Therefore, we suggest using either the harmonic average algorithm or the geometric average algorithm strategy to characterize reservoir heterogeneity.

5.2. Performance assessment of different reservoir characterization models

During the development stages of unconventional resources, is our proposed reservoir characterization model better than other well-known reservoir characterization models? The heatmap of correlation coefficients provides a full picture of the data relationships and comprehensively identifies attribute-to-attribute relationships. Fig. 25 presents the heatmap correlation matrix that correlates the 17 variables with each other. The matrix shows that liquid hydrocarbon content (LI) and

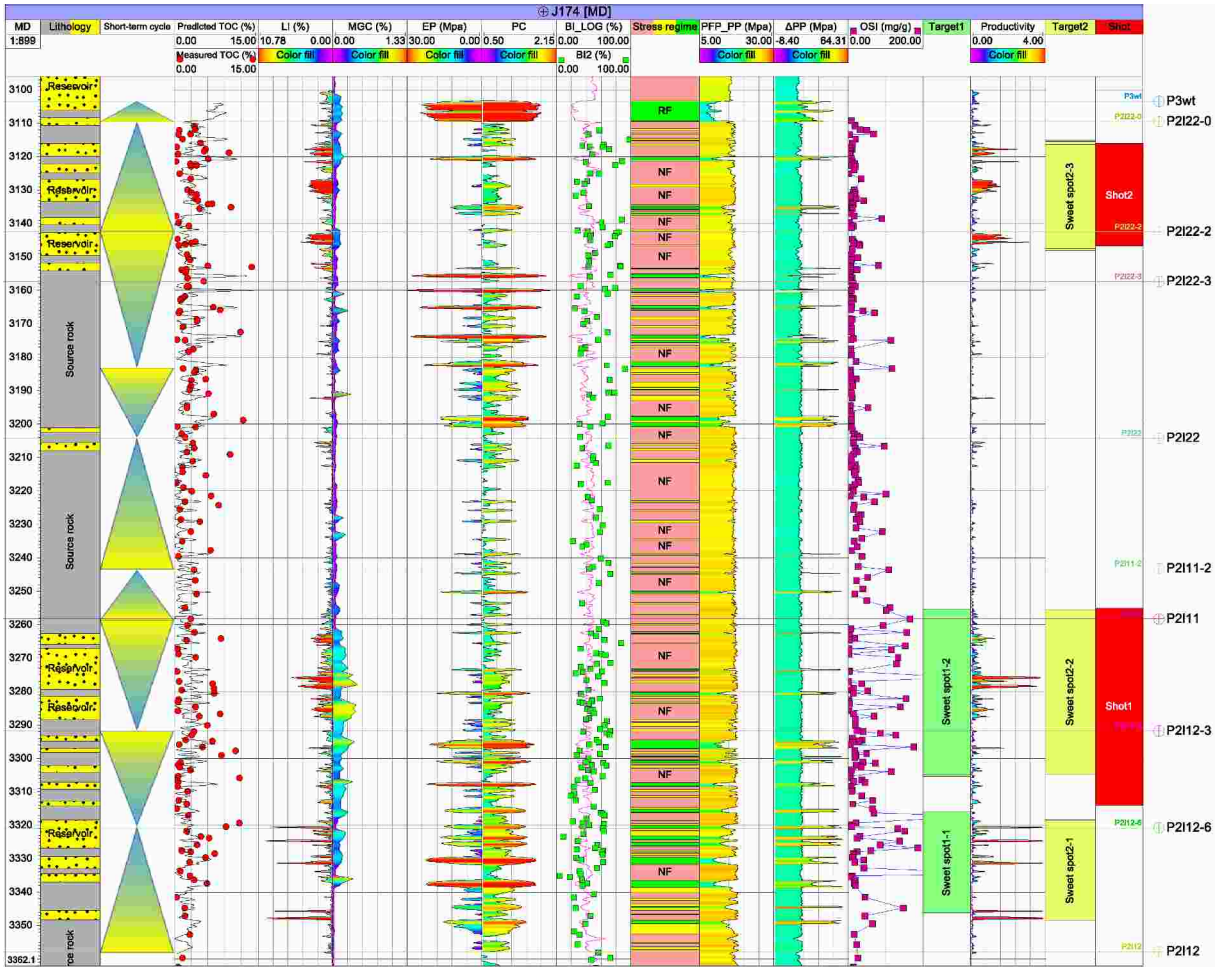


Fig. 27. Prediction of potential sweet spots of the Lucaogou Formation in well J174, Jimusaer Sag. Key: LI = liquid hydrocarbon content; MGC = mud gas content; EP = excess pressure; PC = pressure coefficient; BI = brittleness index; NF = normal faults; SF = strike-slip faults; RF = reverse faults; PFP_PP = rock fracture pressure; ΔPP = fault reactivation pressure; Track target1 = the sweet spots identified by OSI > 100. Track target2 = the sweet spots identified by the method proposed in this study (Track productivity); Track shot = the oil test section.

mud gas content (MGC) are highly correlated with output daily productivity, compared with porosity (φ), oil saturation (So) and rock density (ρ_b); therefore, the quantitative evaluation model we built for reservoir storage capacity is more effective. The matrix also confirms that there is no single variable that determines the quality of unconventional reservoirs, but that many variables are important. Therefore, unconventional reservoir characterization should use an integrated and multidisciplinary approach that includes the study of all important parameters, whenever they are available, taking into account their scales, heterogeneity and relationships. 5 reservoir characterization elements are used individually to predict shale oil productivity. 8 times 8-fold cross-validation and correlation coefficients are employed to evaluate the performance of the shale oil productivity prediction models constructed from 6 ANN inputs. The training and testing results of the 6 prediction models present significant differences. Based on the average and median of the correlation coefficients, the six prediction models are ranked as follows: Advanced model (this study) (average = 0.76, median = 0.80) > Reservoir quality model (average = 0.66, median = 0.69) > Source rock quality model (average = 0.44, median = 0.42) > Conduit conditions model (average = 0.29, median = 0.28) > Migration dynamics model (average = 0.13, median = 0.10) > Source reservoir assemblages model (average = 0.03, median = 0.01) (Fig. 26). The advanced model is better than other single geological factor evaluation models. In addition, evaluation parameters with high correlation coefficients as inputs can make prediction models more accurate (Fig. 25). However, here we place more emphasis on multidisciplinary unconventional characterization approaches that make them applicable to different geological settings.

To further verify the effectiveness of the proposed method to determine the sweet spots, the results obtained by the method are combined with the oil test results and compared with the methods from the state-of-the-art. Among them, Jarvie et al. [12] proposed OSI (S_1/TOC) > 100 as the most classical method to evaluate the potential sweet spot of shale oil. Well J174, which was continuously cored from 3109.2 m to 3440.2 m, is a typical well for studying the continental shale oil systems in China. The test results showed that the Shot1 section produced 7.76 tons of oil per day, while the Shot2 section produced 2.15 tons of oil per day (Fig. 27). Fig. 27 presents the evaluation results for each geological factor, as described above. It can be observed that the entire section of the Lucaogou Formation has hydrocarbon shows, but there are differences in the intensity of the hydrocarbon shows. Track target1 shows the sweet spots identified by OSI > 100, while track target2 presents the sweet spots identified by the shale oil productivity calculated in this study. The results obtained by these two methods are basically the same. In addition, sweet spot2-3 could also be a potential shale oil production interval (Fig. 27). The testing result of this interval is 2.15 t/d, which provides sufficient evidence to test the effectiveness of the model proposed in this paper. Furthermore, the proposed shale oil productivity prediction model not only provides favorable locations for sweet spots, but also provides suggestions for development thickness to meet the

Table 6
Ranking of factors affecting shale oil productivity.

Parameters	Source reservoir assemblages	Source rock quality		Reservoir quality		Migration dynamics		Conduit conditions				Fault reactivation pressure (ΔPP)
		TOC	Depth	Liquid hydrocarbon content (LI)	Mud gas content	Pressure coefficient (PC)	Excess pressure (EP)	Brittleness index (BI)	Rock fracture pressure (P_{FP_PP})	Stress regime		
										Normal faults percentage	Strike-slip faults percentage	Reverse faults percentage	
Grey	0.6716	0.5824	0.5590	0.6956	0.8254	0.7452	0.8077	0.5410	0.4941	0.5865	0.6634	0.6772	0.6162
Pearson	−0.2306	0.5070	0.0821	0.7690	0.7840	0.0154	0.0271	0.0624	0.1983	0.4360	−0.1193	−0.3300	0.1203
Spearman	−0.1404	0.6530	0.0763	0.7970	0.5090	−0.0690	0.0536	0.1480	−0.0040	0.3970	−0.0336	−0.4220	0.1299
Kendall	−0.1167	0.4460	0.0515	0.6290	0.3700	−0.0399	0.0319	0.1038	−0.0044	0.2800	−0.0115	−0.2910	0.0843
Rank	5	2	8	1	1	7	7	7	10	4	9	3	6

lower limit of industrial oil flow. Although sweet spot1-1 or sweet spot2-1 is identified by both methods, the development location of the interval is preferably located in the interval with tension stress regime and high productivity (Fig. 27). This indicates that the new method can be successfully supplemented to predict the sweet spots in unconventional reservoirs.

5.3. Variable importance analysis

The importance analysis of variables affecting shale oil productivity is significant for the next step of shale oil exploration and development. In this paper, the grey correlation coefficient, Pearson correlation coefficient, Spearman correlation coefficient and Kendall correlation coefficient are used to evaluate the correlation between 13 variables and shale oil productivity. Among them, the grey correlation coefficient is suitable for solving complex interrelationships between multiple factors and variables according to the degree of similarity of the change curves [122–123]. The Pearson correlation coefficient is employed to assess the linear relationship of the covariance matrix of the data set [124]. The Spearman correlation coefficient is usually used to assess the nonlinear correlation between data based on nonparametric statistical theory [124]. The Kendall correlation coefficient is another alternative method for assessing nonlinear correlations, counting two variables as consistent or inconsistent pairs based on their rank [124]. When dealing with tied rankings, this method outperforms the Spearman correlation coefficient. In this paper, the correlation analysis reflects only the relationship between the influencing factors and shale oil production, without considering the effects of other variables. The correlation coefficient rankings of the four methods are shown in Table 6. The results show that there are four main controlling factors affecting shale oil productivity, namely reservoir quality, source rock quality, stress regime and source reservoir assemblages, as shown in Table 6. Prior to our work, few scholars had systematically integrated geological and mathematical analysis of the various factors affecting shale oil production. Fig. 28 shows the comprehensive ranking of the correlations of different influencing factors under different correlation coefficient algorithms. Since the method of calculating the correlation coefficient is different, the obtained correlation coefficient is also different. In this paper, the correlation coefficients obtained by the four methods are sorted and then averaged, and finally the comprehensive sorting results of different influencing factors are obtained. It can be found that the liquid hydrocarbon content and mud gas content are tied for the first place in the influencing factors, which can represent the reservoir storage capacity. TOC, reverse faults percentage, normal faults percentage and the source-reservoir assemblages are closely related, ranking 2nd to 5th.

5.4. Uncertainty analysis and recommendations

Quantitative assessment of shale oil production is not accurate science, but it can reduce the risk of shale oil exploration and development. The quantitative evaluation models and methods described above mainly depend on the accuracy of the source-reservoir assemblage models, source rock quality evaluation, reservoir quality evaluation, migration dynamics prediction and conduit condition evaluation. These evaluation conditions include a total of 13 evaluation indicators. Although all elements are analyzed equally in Fig. 21, we believe that different geological settings have different key elements for high shale oil production. Therefore, different methods of obtaining correlation coefficients are used to determine the contribution of each variable. However, there are still some uncertainties in this evaluation system, which arise from the three-dimensional reservoir performance, shale oil mobility and three-dimensional fracture network. Currently, our selection of perforation targets is based on single well production predictions without a clear understanding of the lateral performance of the targets. Ultimately it may affect the economic value of target development. The shale oil system is dominated by in-situ accumulation without obvious

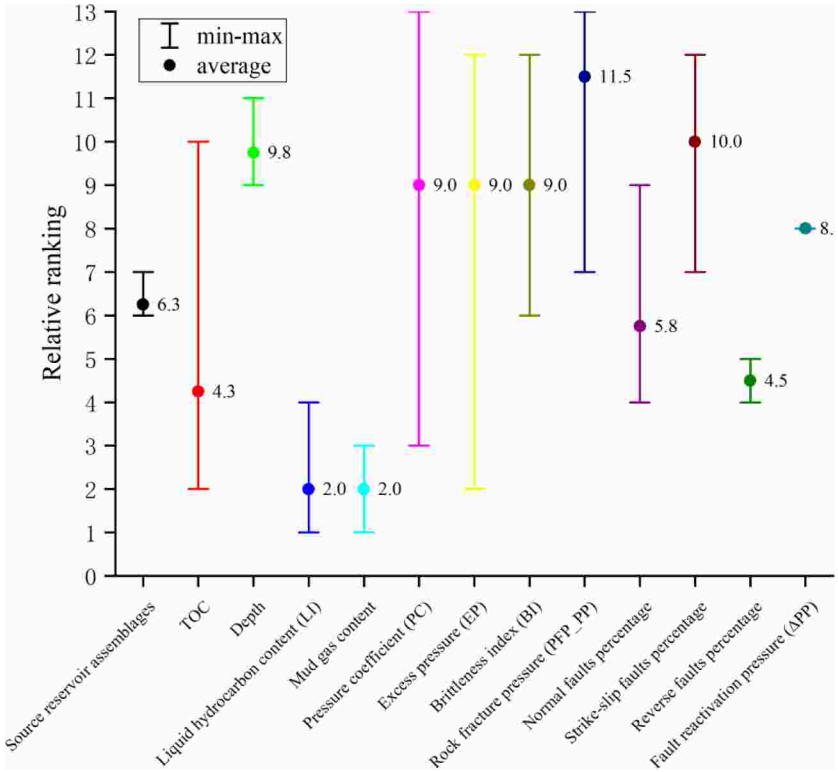


Fig. 28. Comprehensive ranking of different factors affecting shale oil productivity.

fractionation. Therefore, it often has the characteristics of high density and high viscosity. The quality or mobility of shale oil finally affects shale oil production. Furthermore, a large amount of measured data is required to calibrate the predicted data, which is not a problem for mature basins, but there is a large uncertainty for immature exploration plays. However, this disadvantage can be solved by analogy with mature exploration areas with similar geological settings. How to reduce the risk of shale oil exploration and development? Different source-reservoir assemblages have different degrees of hydrocarbon enrichment and economic development values. The analysis of the causes of high shale oil production is the key to further exploration of shale oil. Firstly, different stratigraphic sequences are classified for the target and the corresponding geological framework model is established. Geological, geochemical, geophysical, rock mechanics and reservoir engineering evaluations are conducted on the present test sections to identify the reasons for the hydrocarbon differential accumulation. Finally, well seismic integration is performed to predict key elements and preferably select favorable exploration plays and intervals. Furthermore, it is strongly recommended that integrated geological evaluation and machine learning approaches can further attempt to improve the overall understanding of unconventional resource evaluation and prediction as future work.

6. Conclusions

This thesis proposes a new approach incorporating ANN techniques to evaluate the complex characteristics of shale oil reservoirs with the objective of optimizing the production performance of these resources, while developing new methods and guidelines for the development of shale oil resources. The main conclusions are summarized as follows:

(1) The proposed quantitative evaluation method based on source-reservoir assemblage types, source rock quality, reservoir quality, migration dynamics and conduit conditions can systematically and comprehensively characterize hydrocarbon generation, storage, dynamics and flow stimulation.
(2) The ANN architectures developed based on different averaging algorithms have good agreement between predicted and simulated data for training ($R > 0.95$) and validation ($R > 0.87$) of the models. Based on the correlation coefficients and loss function plots, the geometric and harmonic averaging algorithm strategies are preferentially recommended for characterizing reservoir heterogeneity.
(3) There is no single variable that determines the quality of unconventional reservoirs, but many variables are important. The training results of the advanced prediction model are better than the other five single reservoir characterization models. The sweet spot of shale oil predicted by the model is in good agreement with the test oil results, which is a good complement to the current state-of-the-art methods.
(4) For variable importance analysis, liquid hydrocarbon content and mud gas content are tied for first place, followed by TOC, reverse faults percentage, normal faults percentage and the source-reservoir assemblages. Among them, the increase in liquid hydrocarbon content, mud gas content, TOC and normal faults percentage has positive effects on shale oil production, while the increase in reverse faults percentage has negative effects on shale oil production.

Due to the specificity of shale oil migration and accumulation, it is suggested to better focus on the three-dimensional reservoir performance, shale oil mobility and three-dimensional fracture network in future studies. Furthermore, geological evaluation and machine learning are strongly recommended to understand the evaluation and simulation of unconventional reservoirs.

CRediT authorship contribution statement

Yazhou Liu: Methodology, Validation, Data curation, Software,

Writing-original draft, Writing- review & editing. **Jianhui Zeng:** Conceptualization, Project administration, Supervision, Funding acquisition, Resources. **Juncheng Qiao:** Supervision, Writing- review & editing. **Guangqing Yang:** Methodology, Formal analysis. **Shu'ning Liu:** Software, Visualization. **Weifu Cao:** Data curation, Visualization.

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors sincerely thank the Research Institute of Experimental and Testing, PetroChina Xinjiang Oilfield Company for data collection. This work was financially supported by the National Natural Science Foundation of China (Grant No. 41972147), China and the 2021 AAPG Foundation Grants-in-Aid Program (Roger W. Stoneburner Memorial Grant), United States. I would like to express my sincere thanks to Paul Hackley and Hamed Sanei for giving me the opportunity to study at Aarhus University. Moreover, we would like to thank Prof. Zita Vale and two anonymous reviewers for their excellent comments on the manuscript.

References

[1] Ahmed U, Meehan DN. Unconventional oil and gas resources: exploitation and development. CRC Press; 2016.
[2] Jarvie D, Breyer J. Shale resource systems for oil and gas: Part 1–Shale gas resource systems; Part 2–Shale oil resource systems. Shale reservoirs–Giant resources for the 21st century. AAPG Mem 2011;97:1–31.
[3] Jarvie DM, Hill RJ, Ruble TE, Pollastro RM. Unconventional shale-gas systems: The Mississippian Barnett Shale of north-central Texas as one model for thermogenic shale-gas assessment. AAPG Bull 2007;91(4):475–99.
[4] EIA, 2022. Annual Energy Outlook 2022. https://www.eia.gov/.
[5] EIA, 2016. Annual Energy Outlook 2016. https://www.eia.gov/.
[6] Sheng JJ. Critical review of field EOR projects in shale and tight reservoirs. J Pet Sci Eng 2017;159:654–65.
[7] Syed FI, Dahaghi AK, Muther T. Laboratory to field scale assessment for EOR applicability in tight oil reservoirs. Pet Sci 2022.
[8] Zhang S, Liu H, Liu Y, Wang Y, Wang M, Bao Y, et al. Main controls and geological sweet spot types in Paleogene shale oil rich areas of the Jiyang Depression, Bohai Bay basin, China. Marine Pet Geol 2020;111:576–87.
[9] Zou C, Zhang G, Yang Z, Tao S, Hou L, Zhu R, et al. Geological concepts, characteristics, resource potential and key techniques of unconventional hydrocarbon: On unconventional petroleum geology. Pet Explor Dev 2013;40(4):385–99.
[10] Gharavi A, Hassan M, Gholinezhad J, Ghoochaninejad H, Barati H, Buick J, et al. Application of machine learning techniques for identifying productive zones in unconventional reservoir. Int J Intell Netw 2022;3:87–101.
[11] Hu T, Pang X, Jiang S, Wang Q, Zheng X, Ding X, et al. Oil content evaluation of lacustrine organic-rich shale with strong heterogeneity: A case study of the Middle Permian Lucaogou Formation in Jimusaer Sag, Junggar Basin, NW China. Fuel 2018;221:196–205.
[12] Jarvie DM. Unconventional oil petroleum systems: shales and shale hybrids. Society of Petroleum Engineers Search and Discovery Article 2011:80131.
[13] Li W, Lu S, Xue H, Zhang P, Hu Y. Oil content in argillaceous dolomite from the Jianghan Basin, China: application of new grading evaluation criteria to study shale oil potential. Fuel 2015;143:424–9.
[14] Liu Y, Zeng J, Yang G, Jia W, Liu S, Kong X, et al. An innovative method for the characterization of oil content in lacustrine shale-oil systems: A case study from the Middle Permian Lucaogou Formation in the Jimusaer Sag, Junggar Basin. Marine Pet Geol 2021;130:105112.
[15] Sharma V, Sircar A. Demarcating sweet spots in cambay shale by integrating rock eval pyrolysis, geomechanics and seismic data. Abu Dhabi International Petroleum Exhibition & Conference: OnePetro; 2018.
[16] Zhao X, Zhou L, Pu X, Han W, Jin F, Xiao D, et al. Exploration breakthroughs and geological characteristics of continental shale oil: A case study of the Kongdian Formation in the Cangdong Sag, China. Marine Pet Geol 2019;102:544–56.
[17] Cudjoe S, Vinassa M, Gomes JHB, Barati RG. A comprehensive approach to sweet-spot mapping for hydraulic fracturing and CO2 huff-n-puff injection in Chattanooga shale formation. J Nat Gas Sci Eng 2016;33:1201–18.
[18] Eid M, Abedeen A, Kurri J. Petrophysics of source rock reservoirs: a complete workflow. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition: OnePetro; 2018.
[19] Hou L, Zou C, Yu Z, Luo X, Wu S, Zhao Z, et al. Quantitative assessment of the sweet spot in marine shale oil and gas based on geology, engineering, and economics: A case study from the Eagle Ford Shale, USA. Energ Strat Rev 2021;38:100713.
[20] Licitra D, Lovrincevich E, Vittore F, Quiroga J, Oviedo P, Montoya V, et al. Sweet spots in Vaca Muerta: Integration of subsurface and production data in Loma Campana shale development, Argentina. SPE/AAPG/SEG Unconventional Resources Technology Conference: OnePetro; 2015.
[21] Ter Heege J, Zijp M, Nelskamp S, Douma L, Verreussel R, Ten Veen J, et al. Sweet spot identification in underexplored shales using multidisciplinary reservoir characterization and key performance indicators: Example of the Posidonia Shale Formation in the Netherlands. J Nat Gas Sci Eng 2015;27:558–77.
[22] Badics B, Vető I. Source rocks and petroleum systems in the Hungarian part of the Pannonian Basin: The potential for shale gas and shale oil plays. Mar Pet Geol 2012;31(1):53–69.
[23] Hackley PC, Cardott BJ. Application of organic petrography in North American shale petroleum systems: A review. Int J Coal Geol 2016;163:8–51.
[24] Jarvie DM. Shale resource systems for oil and gas: Part 2—Shale-oil resource systems. 2012.
[25] Chen L. On the Essence of Stratigraphic Diachronism from the Basic Sequence Stratigraphic Model—An example from the Devonian sequence stratigraphic framework of the Dianqiangui basin. Northwest Geol 2008;41(1):50–8.
[26] Peng Y, Chen Y, Liu Y. Benxi Formation—lithohorizon and chronohorizon with diachronism. Global Geol 2003;22:111–8.
[27] Xuemei D, Hongliang W. High-resolution sequence stratigraphy in layer correlation: A case study from Shahejie Formation in Huanghua Depression. J Pet Sci Eng 2016;147:1–6.
[28] Cross TA. Controls on coal distribution in transgressive-regressive cycles. Western Interior, USA: Upper Cretaceous; 1988.
[29] Ayranci K, Harris NB, Dong T. High resolution sequence stratigraphic reconstruction of mud-dominated systems below storm wave base; a case study from the Middle to Upper Devonian Horn River Group, British Columbia. Canada Sedimentary Geol 2018;373:239–53.
[30] Catuneanu O, Zecchin M. High-resolution sequence stratigraphy of clastic shelves II: controls on sequence development. Mar Pet Geol 2013;39(1):26–38.
[31] Dalman R, Weltje GJ, Karamitopoulos P. High-resolution sequence stratigraphy of fluvio–deltaic systems: prospects of system-wide chronostratigraphic correlation. Earth Planet Sci Lett 2015;412:10–7.
[32] Kocurek G. Aeolian system response to external forcing factors—a sequence stratigraphic view of the Saharan region. Quaternary deserts and climatic change: CRC Press; 2020. p. 327–37.
[33] Shang W, Xu S, Mao Z, Li X, Gao G, Li Z, et al. High-resolution sequence stratigraphy in continental lacustrine basin: A case of Eocene Shahejie formation in the Dongying Depression, Bohai Bay Basin. Mar Pet Geol 2022;136:105438.
[34] Zecchin M, Catuneanu O. High-resolution sequence stratigraphy of clastic shelves I: units and bounding surfaces. Mar Pet Geol 2013;39(1):1–25.
[35] Amorosi A, Rossi V, Sarti G, Mattei R. Coalescent valley fills from the late Quaternary record of Tuscany (Italy). Quat Int 2013;288:129–38.
[36] Bourquin S, Rigollet C, Bourges P. High-resolution sequence stratigraphy of an alluvial fan–fan delta environment: stratigraphic and geodynamic implications–an example from the Keuper Chaunoy Sandstones. Paris Basin Sedimentary Geology 1998;121(3–4):207–37.
[37] Partington M, Mitchener B, Milton N, Fraser A. Genetic sequence stratigraphy for the North Sea Late Jurassic and Early Cretaceous: distribution and prediction of Kimmeridgian–Late Ryazanian reservoirs in the North Sea and adjacent areas. Conference Genetic sequence stratigraphy for the North Sea Late Jurassic and Early Cretaceous: distribution and prediction of Kimmeridgian–Late Ryazanian reservoirs in the North Sea and adjacent areas, vol. 4. Geological Society of London, p. 347-70.
[38] Begg S, Carter R, Dranfield P. Assigning effective values to simulator gridblock parameters for heterogeneous reservoirs. SPE Reserv Eng 1989;4(04):455–63.
[39] Li D, Beckner B, Kumar A. A new efficient averaging technique for scaleup of multimillion-cell geologic models. SPE Reserv Eval Eng 2001;4(04):297–307.
[40] Pyrcz MJ, Deutsch CV. Geostatistical reservoir modeling. Oxford University Press; 2014.
[41] Tran TT. Improving variogram reproduction on dense simulation grids. Comput Geosci 1994;20(7–8):1161–8.
[42] Deka PC. Support vector machine applications in the field of hydrology: a review. Appl Soft Comput 2014;19:372–86.
[43] Patro S, Sahu KK. Normalization: A preprocessing stage. arXiv preprint arXiv: 150306462. 2015.
[44] Zhang C, Chen C, Streimikiene D, Balezentis T. Intuitionistic fuzzy MULTIMOORA approach for multi-criteria assessment of the energy storage technologies. Appl Soft Comput 2019;79:410–23.
[45] Al Shalabi L, Shaaban Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. Conference Normalization as a preprocessing engine for data mining and the approach of preference matrix. IEEE, p. 207-14.
[46] Mohamad IB, Usman D. Standardization and its effects on K-means clustering algorithm. Res J Appl Sci Eng Technol 2013;6(17):3299–303.
[47] Saranya C, Manikandan G. A study on normalization techniques for privacy preserving data mining. Int J Eng Technol (IJET) 2013;5(3):2701–4.

[48] Asuero AG, Sayago A, González A. The correlation coefficient: An overview. Crit Rev Anal Chem 2006;36(1):41–59.
[49] Chen M-F, Tzeng G-H. Combining grey relation and TOPSIS concepts for selecting an expatriate host country. Math Comput Model 2004;40(13):1473–90.
[50] Gocic M, Trajkovic S. Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Global Planet Change 2013;100:172–82.
[51] Nathans LL, Oswald FL, Nimon K. Interpreting multiple linear regression: a guidebook of variable importance. Pract Assess Res Eval 2012;17(9):n9.
[52] Obilor EI, Amadi EC. Test for significance of Pearson's correlation coefficient. Int J Innovat Math Statist Energy Policies 2018;6(1):11–23.
[53] Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018;126(5):1763–8.
[54] Taylor R. Interpretation of the correlation coefficient: a basic review. J Diagn Med Sonogr 1990;6(1):35–9.
[55] Yue S, Pilon P, Cavadias G. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. J Hydrol 2002;259(1–4):254–71.
[56] Chen H, Wang Y, Zuo M, Zhang C, Jia N, Liu X, et al. A new prediction model of CO2 diffusion coefficient in crude oil under reservoir conditions based on BP neural network. Energy 2022;239:122286.
[57] Yang H, Jin J, Hou F, He X, Hang Y. An ANN-based method for predicting Zhundong and other Chinese coal slagging potential. Fuel 2021;293:120271.
[58] Deng Y, Zhu M, Xiang D, Cheng X. An analysis for effect of cetane number on exhaust emissions from engine with the neural network. Fuel 2002;81(15):1963–70.
[59] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for. Cognit Sci 1985.
[60] Ahmad F, Isa NAM, Osman MK, Hussain Z. Performance comparison of gradient descent and Genetic Algorithm based Artificial Neural Networks training. Conference Performance comparison of gradient descent and Genetic Algorithm based Artificial Neural Networks training. IEEE, p. 604-9.
[61] Wang J, Wen Y, Gou Y, Ye Z, Chen H. Fractional-order gradient descent learning of BP neural networks with Caputo derivative. Neural Netw 2017;89:19–30.
[62] Cheng J, Wang X, Si T, Zhou F, Zhou J, Cen K. Ignition temperature and activation energy of power coal blends predicted with back-propagation neural network models. Fuel 2016;173:230–8.
[63] Hariharan N, Senthil V, Krishnamoorthi M, Karthic S. Application of artificial neural network and response surface methodology for predicting and optimizing dual-fuel CI engine characteristics using hydrogen and bio fuel with water injection. Fuel 2020;270:117576.
[64] Karkevandi-Talkhooncheh A, Hajirezaie S, Hemmati-Sarapardeh A, Husein MM, Karan K, Sharifi M. Application of adaptive neuro fuzzy interface system optimized with evolutionary algorithms for modeling CO2-crude oil minimum miscibility pressure. Fuel 2017;205:34–45.
[65] Shen H, Wang Z, Gao C, Qin J, Yao F, Xu W. Determining the number of BP neural network hidden layer units. J Tianjin Univ Technol 2008;24(5):13.
[66] Stathakis D. How many hidden layers and nodes? Int J Remote Sens 2009;30(8):2133–47.
[67] de Campos Souza PV, Torres LCB, Guimaraes AJ, Araujo VS, Araujo VJS, Rezende TS. Data density-based clustering for regularized fuzzy neural networks based on nullneurons and robust activation function. Soft Comput 2019;23(23):12475–89.
[68] Warner B, Misra M. Understanding neural networks as statistical tools. Am Stat 1996;50(4):284–93.
[69] McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychol Rev 1981;88(5):375.
[70] Cao Z, Jiang H, Zeng J, Saibi H, Lu T, Xie X, et al. Nanoscale liquid hydrocarbon adsorption on clay minerals: A molecular dynamics simulation of shale oils. Chem Eng J 2021;420:127578.
[71] Liu Y, Zeng J, Jin J, Yang G, Xiang B, Zhou N, et al. Geochemical evaluation of produced petroleum from the Middle Permian Lucaogou reservoirs Junggar Basin and its implication for the unconventional shale oil play. J Pet Sci Eng 2022;211:110202.
[72] Luo X, Wang Z, Zhang L, Yang W, Liu L. Overpressure generation and evolution in a compressional tectonic setting, the southern margin of Junggar Basin, northwestern China. AAPG Bull 2007;91(8):1123–39.
[73] Xiao W, Han C, Yuan C, Sun M, Lin S, Chen H, et al. Middle Cambrian to Permian subduction-related accretionary orogenesis of Northern Xinjiang, NW China: implications for the tectonic evolution of central Asia. J Asian Earth Sci 2008;32(2–4):102–17.
[74] Wang J, Cao Y, Liu K, Gao Y, Qin Z. Fractal characteristics of the pore structures of fine-grained, mixed sedimentary rocks from the Jimsar Sag, Junggar Basin: Implications for lacustrine tight oil accumulations. J Pet Sci Eng 2019;182:106363.
[75] Yang Y, Qiu L, Wan M, Jia X, Cao Y, Lei D, et al. Depositional model for a salinized lacustrine basin: The permian Lucaogou formation, Jimsar sag, Junggar Basin, NW China. J Asian Earth Sci 2019;178:81–95.
[76] Zhang S, Cao Y, Liu K, Jahren J, Xi K, Zhu R, et al. Characterization of lacustrine mixed fine-grained sedimentary rocks using coupled chemostratigraphic-petrographic analysis: A case study from a tight oil reservoir in the Jimusar Sag, Junggar Basin. Marine Pet Geol 2019;99:453–72.
[77] Dong Y, Xu D, Qian G, Wang X, Dai Y. Shale oil "sweet-spot" prediction in Jimusar Sag. Specail Oil & Gas Reservoirs 2020;27(3):54–9.
[78] Huang H, Gao Y, Jones MM, Tao H, Carroll AR, Ibarra DE, et al. Astronomical forcing of Middle Permian terrestrial climate recorded in a large paleolake in northwestern China. Palaeogeogr Palaeoclimatol Palaeoecol 2020;550:109735.
[79] Zhang Y, Ma S, Gao Y, Li Y, Zhang J, Wang L, et al. Characteristics of the High-resolution Sequence Stratigraphy and the Distribution of Tight Oil Reservoirs in the Salt Lake: A Case from the A Region of Lucaogou Formation, Jimsar Sag. Geoscience 2016;30(5):1096.
[80] Hou L, Luo X, Zhao Z, Zhang L. Identification of oil produced from shale and tight reservoirs in the Permian Lucaogou Shale sequence, Jimsar Sag, Junggar Basin, NW China. ACS omega 2021;6(3):2127–42.
[81] Ardakani OH, Sanei H, Ghanizadeh A, Lavoie D, Chen Z, Clarkson CR. Do all fractions of organic matter contribute equally in shale porosity? A case study from Upper Ordovician Utica Shale, southern Quebec, Canada. Marine Pet Geol 2018;92:794–808.
[82] Sanei H, Ardakani OH, Akai T, Akihisa K, Jiang C, Wood JM. Core versus cuttings samples for geochemical and petrophysical analysis of unconventional reservoir rocks. Sci Rep 2020;10(1):1–10.
[83] Sanei H, Wood JM, Ardakani OH, Clarkson CR, Jiang C. Characterization of organic matter fractions in an unconventional tight gas siltstone reservoir. Int J Coal Geol 2015;150:296–305.
[84] Sun J, Dong X, Wang J, Schmitt DR, Xu C, Mohammed T, et al. Measurement of total porosity for gas shales by gas injection porosimetry (GIP) method. Fuel 2016;186:694–707.
[85] Espitalié J, Maxwell J, Chenet Y, Marquis F. Aspects of hydrocarbon migration in the Mesozoic in the Paris Basin as deduced from an organic geochemical survey. Organic Geochemistry In Petroleum Exploration: Elsevier; 1988. p. 467–81.
[86] Justwan H, Dahl B. Quantitative hydrocarbon potential mapping and organofacies study in the Greater Balder Area, Norwegian North Sea. Conference Quantitative hydrocarbon potential mapping and organofacies study in the Greater Balder Area, Norwegian North Sea, vol. 6. Geological Society of London, p. 1317-29.
[87] Welte D, Tissot P. Petroleum formation and occurrence. Springer 1984.
[88] Bordenave ML. Applied petroleum geochemistry. 1993.
[89] Khoshnoodkia M, Mohseni H, Rahmani O, Mohammadi A. TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data. J Pet Sci Eng 2011;78(1):119–30.
[90] Passey Q, Creaney S, Kulla J, Moretti F, Stroud J. A practical model for organic richness from porosity and resistivity logs. AAPG Bull 1990;74(12):1777–94.
[91] Rui J, Zhang H, Ren Q, Yan L, Guo Q, Zhang D. TOC content prediction based on a combined Gaussian process regression model. Mar Pet Geol 2020;118:104429.
[92] Shalaby MR, Malik OA, Lai D, Jumat N, Islam MA. Thermal maturity and TOC prediction using machine learning techniques: case study from the Cretaceous-Paleocene source rock, Taranaki Basin, New Zealand. J Pet Explor Prod Technol 2020;10(6):2175–93.
[93] Zhu L, Zhang C, Zhang C, Zhang Z, Zhou X, Liu W, et al. A new and reliable dual model-and data-driven TOC prediction concept: A TOC logging evaluation method using multiple overlapping methods integrated with semi-supervised deep learning. J Pet Sci Eng 2020;188:106944.
[94] Gong D, Song Y, Peng M, Liu C, Wang R, Wu Wa. The Hydrocarbon Potential of Carboniferous Reservoirs in the Jimsar Sag, Northwest China: Implications for a Giant Volcanic-Petroleum Reserves. Front. Frontiers in Earth Science. 2022;10:879712.
[95] He W, Liu Y, Wang D, Lei D, Liu G, Gao G, et al. Geochemical Characteristics and Process of Hydrocarbon Generation Evolution of the Lucaogou Formation Shale, Jimsar Depression, Junggar Basin. Energies 2022;15(7):2331.
[96] Oloruntobi O, Butt S. Energy-based formation pressure prediction. J Pet Sci Eng 2019;173:955–64.
[97] Czerniak M. RhoVe method: A new empirical pore pressure transform. Mar Pet Geol 2017;86:343–66.
[98] Eaton BA. The equation for geopressure prediction from well logs. Conference The equation for geopressure prediction from well logs. OnePetro.
[99] Passey QR, Bohacs K, Esch WL, Klimentidis R, Sinha S. From oil-prone source rock to gas-producing shale reservoir–geologic and petrophysical characterization of unconventional shale-gas reservoirs. Conference From oil-prone source rock to gas-producing shale reservoir–geologic and petrophysical characterization of unconventional shale-gas reservoirs. OnePetro.
[100] McKeon M. Horizontal fracturing in shale plays. Petroleum Technology Transfer Conference, Taking A Deeper Look At Shales: Geology And Potential Of The Upper Ordovician Utica Shale In The Appalachian Basin, New Philadelphia, OH, June2011.
[101] Zou C. Unconventional petroleum geology. Elsevier; 2017.
[102] Rickman R, Mullen MJ, Petre JE, Grieser WV, Kundert D. A practical use of shale petrophysics for stimulation design optimization: All shale plays are not clones of the Barnett Shale. Conference A practical use of shale petrophysics for stimulation design optimization: All shale plays are not clones of the Barnett Shale. OnePetro.
[103] Radwan A, Abudeif A, Attia M, Elkhawaga MA, Abdelghany WK, Kasem AA. Geopressure evaluation using integrated basin modelling, well-logging and reservoir data analysis in the northern part of the Badri oil field, Gulf of Suez, Egypt. J Afric Earth Sci 2020;162:103743.
[104] Radwan A, Abudeif A, Attia M, Mohammed M. Pore and fracture pressure modeling using direct and indirect methods in Badri Field, Gulf of Suez, Egypt. J Afric Earth Sci 2019;156:133–43.
[105] Zhang J. Pore pressure prediction from well logs: Methods, modifications, and new approaches. Earth Sci Rev 2011;108(1–2):50–63.

[106] Zoback M, Barton C, Brudy M, Castillo D, Finkbeiner T, Grollimund B, et al. Determination of stress orientation and magnitude in deep wells. Int J Rock Mech Min Sci 2003;40(7–8):1049–76.
[107] Eaton BA. Fracture gradient prediction and its application in oilfield operations. J Petrol Tech 1969;21(10):1353–60.
[108] Ju W, Shen J, Qin Y, Meng S, Wu C, Shen Y, et al. In-situ stress state in the Linxing region, eastern Ordos Basin, China: implications for unconventional gas exploration and production. Mar Pet Geol 2017;86:66–78.
[109] Binh NTT, Tokunaga T, Son HP, Van Binh M. Present-day stress and pore pressure fields in the Cuu Long and Nam Con Son Basins, offshore Vietnam. Mar Pet Geol 2007;24(10):607–15.
[110] Finkbeiner T, Zoback M, Flemings P, Stump B. Stress, pore pressure, and dynamically constrained hydrocarbon columns in the South Eugene Island 330 field, northern Gulf of Mexico. AAPG Bull 2001;85(6):1007–31.
[111] Liu R, Liu J, Zhu W, Hao F, Xie Y, Wang Z, et al. In situ stress analysis in the Yinggehai Basin, northwestern South China Sea: Implication for the pore pressure-stress coupling process. Mar Pet Geol 2016;77:341–52.
[112] Tingay MR, Hillis RR, Morley CK, King RC, Swarbrick RE, Damit AR. Present-day stress and neotectonics of Brunei: Implications for petroleum exploration and production. AAPG Bull 2009;93(1):75–100.
[113] Zoback MD. Reservoir geomechanics. Cambridge University Press; 2010.
[114] Anderson EM. The dynamics of faulting and dyke formation with applications to. Britain: Hafner Publishing Company; 1972.
[115] Ganguli SS, Sen S. Investigation of present-day in-situ stresses and pore pressure in the south Cambay Basin, western India: Implications for drilling, reservoir development and fault reactivation. Mar Pet Geol 2020;118:104422.
[116] Javani D, Aadnoy B, Rastegarnia M, Nadimi S, Aghighi MA, Maleki B. Failure criterion effect on solid production prediction and selection of completion solution. J Rock Mech Geotech Eng 2017;9(6):1123–30.
[117] Najibi AR, Ghafoori M, Lashkaripour GR, Asef MR. Reservoir geomechanical modeling: In-situ stress, pore pressure, and mud design. J Pet Sci Eng 2017;151:31–9.
[118] Stone M. Cross-validatory choice and assessment of statistical predictions. J Roy Stat Soc: Ser B (Methodol) 1974;36:111–33.
[119] Bengio Y, Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. Advances in Neural Information Processing Systems. 2003;16.
[120] Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 2009;32:569–75.
[121] Refaeilzadeh P, Tang L, Liu H. Cross-validation Encyclopedia of database systems 2009;5:532–8.
[122] Gao C-l, Li S-c, Wang J, Li L-p, Lin P. The risk assessment of tunnels based on grey correlation and entropy weight method. Geotech Geol Eng 2018;36(3):1621–31.
[123] Lee W-S, Lin Y-C. Evaluating and ranking energy performance of office buildings using Grey relational analysis. Energy 2011;36(5):2551–6.
[124] Li Z, Gao X, Lu D. Correlation analysis and statistical assessment of early hydration characteristics and compressive strength for multi-composite cement paste. Constr Build Mater 2021;310:125260.
[125] Wang L, Ye Y, Qin J, Gao Y, Deng Y, Li Y, et al. Microscopic pore structure characterization and oil-bearing property evaluation of lacustrine shale reservoir: A case study of the Permian Lucaogou Formation in Jimsar Sag. Junggar Basin Oil & Gas geology 2022;43(1):149–60.
[126] Katz B, Gao L, Little J, Zhao YR. Geology still matters–Unconventional petroleum system disappointments and failures. Unconventional Resources 2021;1:18–38.
[127] Gogoi T, Chatterjee R. Multimineral modeling and estimation of brittleness index of Shaly sandstone in Upper Assam and Mizoram areas. India SPE Reservoir Evaluation & Engineering 2020;23(02):708–21.
[128] Wang FP, Gale JFW. Screening criteria for shale-gas systems. Gulf Coast Assoc Geol Soc Trans 2009;59:779–93.
[129] Jin X, Shah SN, Roegiers J-C, Zhang B. Fracability evaluation in shale reservoirs-an integrated petrophysics and geomechanics approach. Conference Fracability evaluation in shale reservoirs-an integrated petrophysics and geomechanics approach. OnePetro.
[130] Gholami R, Rasouli V, Sarmadivaleh M, Minaeian V, Fakhari N. Brittleness of gas shale reservoirs: A case study from the north Perth basin, Australia. J Nat Gas Sci Eng 2016;33:1244–59.
[131] Glorioso JC, Rattia A. Unconventional reservoirs: basic petrophysical concepts for shale gas. Conference Unconventional reservoirs: basic petrophysical concepts for shale gas. European Association of Geoscientists & Engineers, p. cp-285-00058.
[132] Guo T, Zhang S, Ge H, Wang X, Lei X, Xiao B. A new method for evaluation of fracture network formation capacity of rock. Fuel 2015;140:778–87.
[133] Kias E, Maharidge R, Hurt R. Mechanical versus mineralogical brittleness indices across various shale plays. Conference Mechanical versus mineralogical brittleness indices across various shale plays. OnePetro.
[134] Lai J, Wang G, Huang L, Li W, Ran Y, Wang D, et al. Brittleness index estimation in a tight shaly sandstone reservoir using well logs. J Nat Gas Sci Eng 2015;27:1536–45.