processes

MDPI

Review

Research Progress and Prospects of Oil Saturation Evaluation Methods in Shale Oil Reservoirs

Haibo Zhao [1], Jiandong Zheng [1], Zhaoqian Zhang [1], Hao Qin [1], Chunyan Wang [1], Jianhua Zhu [1], Anqi Chen [1], Longchuan Chen [1], Peng Liu [2], Wangwang Yang [2,*] and Xuan Hu [2]

[1] PetroChina Daqing Oilfield Co., Ltd., Daqing 163453, China; zhaohaibo@petrochina.com.cn (H.Z.); zhengjd@petrochina.com.cn (J.Z.); zhangzhaoqian1@petrochina.com.cn (Z.Z.); qinhao@petrochina.com.cn (H.Q.); chyang@petrochina.com.cn (C.W.); zhujianhua@petrochina.com.cn (J.Z.); chenanqi123@petrochina.com.cn (A.C.); alone43@petrochina.com.cn (L.C.)

[2] Research Institute of Petroleum Exploration and Development, Xinjiang Oilfield Company, PetroChina, Karamay 834000, China; liupeng4642@petrochina.com.cn (P.L.); huxuan123@petrochina.com.cn (X.H.)

* Correspondence: ywwwelllogging@163.com

Abstract: The core of the exploration and development of unconventional oil and gas resources, such as shale oil, lies in the effective identification and scale utilization of the sweet spot. Oil saturation is an important parameter in evaluating the sweet spot. Aiming to solve the many current problems of oil saturation logging evaluations of shale oil reservoirs, this study outlines the research progress of oil saturation logging evaluations of shale oil from the three aspects, namely, the electrical method, non-electrical method, and machine learning, through researching the literature and practical applications. At the same time, several typical saturation models are applied to shale oil reservoirs in the Qingshankou Formation of the Gulong Depression, and applicability analyses are conducted. Lastly, the advantages and disadvantages of each oil saturation calculation model are summarized, and suggestions are given for conducting research using each type of method. This study has certain significance in the selection of oil saturation logging evaluation methods for shale oil reservoirs and provides directions for improvement.

Keywords: shale oil; oil saturation; logging evaluation; reservoir quality

check for updates

Citation: Zhao, H.; Zheng, J.; Zhang, Z.; Qin, H.; Wang, C.; Zhu, J.; Chen, A.; Chen, L.; Liu, P.; Yang, W.; et al. Research Progress and Prospects of Oil Saturation Evaluation Methods in Shale Oil Reservoirs. *Processes* **2024**, *12*, 2421. https://doi.org/10.3390/pr12112421

Academic Editor: Youguo Yan

Received: 2 September 2024
Revised: 19 October 2024
Accepted: 23 October 2024
Published: 2 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Shale oil is an essential example of unconventional oil and gas resources, which China is rich in, such as shale oil, with huge resource potential [1–3]. However, this type of reservoir is extremely non-homogeneous, with a complex pore structure [4,5], various oil and water storage modes, and complex mineral types, which lead to a significant error in the calculation of oil saturation, making it challenging to identify an adequate reservoir and restricting the exploration and development processes of this type of oil and gas resource, so it is of great significance to accurately characterize oil saturation for the exploration and development of shale oil [6].

The current methods of calculating the oil saturation in shale oil reservoirs can be summarized into three main categories. The first one is the electrical saturation model developed based on Archie's formula [7], first used by Archie in 1942 to calculate formation water saturation using resistivity logging. The formula is an empirical model developed based on pure sandstone formation, and Archie's formula cannot be applied to shale formations with a high clay content. As a result, many scholars at home and abroad have addressed the limitations of using Archie's formula in shale formations. The presence of mud will lead to additional conductivity in formation water; for this reason, many scholars have found that additional conductivity in clay leads to a decrease in the resistivity curve. In 1950, DeWitte [8] assumed that the mud in shale formations is dispersed and put forward a conductivity model constituting dispersed mud and pore water. In 1954, Poupon et al. [9]

assumed that shale is a combination of various thin interbedded layers of mudstone and sandstone, which led to the development of a simplified model for determining shale water saturation. In 1963, Simandoux [10] proposed a Simandoux equation for the distribution of dispersed mud, which is still one of the more popular models for water saturation. In 1968, Waxman and Smits [11] found that additional conductivity in clay is caused by cation exchange in muddy sandstone and then proposed a W-S saturation evaluation model while considering the cation exchange capacity. In 1971, Poupon and Leveaux et al. [12] carried out field observations in Indonesia and, based on the field development data of the oil field, proposed the Indonesian equation. In 1972, Schlumberger [13] proposed a whole shale saturation model based on the Simandoux degree equation, without considering the form of mud distribution and correcting the mud conductive part. In 1984, Clavier et al. [14] proposed a two-water model based on the theory of bilateration by dividing the formation water into two parts, namely, clay water and free water. Based on this, in 2001, Zhang et al. conducted a study [15]. Formation water could be subdivided into clay water, capillary water, and free water, and a three-water model could be proposed based on the principle of parallel conductivity. In 2016, Kadkhodaie [16] argued that organic matter in shale formations caused an anomalous increase in resistivity and proposed a parallel conductivity model for clay and organic matter. In 2022, Yu [17] argued that conductive minerals in shale formations significantly affect rock conductivity, which led to the proposal of a saturation model that considers conductive minerals. More factors affect resistivity in shale formations, resulting in abnormal low- or high-resistance phenomena. To avoid this influence, many scholars try to evaluate the saturation degree of the non-electric method. Among non-electric logging, nuclear magnetic resonance (NMR) logging is one of the indispensable techniques for assessing the quality of shale reservoirs. In 2021, Liu Yahui et al. [18] chose different starting times for the NMR T2 spectra to calculate the oil-bearing pore space and, thus, calculate the oil-bearing saturation, and the saturation degree was calculated by a core. In 2023, Zeng Jingbo et al. [19] determined the optimal centrifugation time through multi-stage centrifugation 1D NMR experiments, combined NMR and Hg pressure experiments by scaling them together to establish the classification standard of the reservoir pore structure, and established a water saturation model based on the classification standard. Using 2D NMR, in 2021, Yan Weilin et al. conducted a study [20]. The blind source separation technique was used to cluster and partition the T1-T2 spectra of CMR-NG, and the oil saturation of the shale reservoir was calculated based on the partitioning percentage. In 2022, Qin Yingyao et al. conducted a study [21]. The T2-T1 fluid interpretation plate of the Jimushar shale oil reservoir was obtained through 2D NMR experiments, and the fluid saturation could be calculated based on the interpretation plate. In recent years, the artificial intelligence industry has developed rapidly; computer computational performance has become more and more powerful, and deep learning has become a research hotspot in geophysical modeling [22–24]. Shale oil reservoirs are highly inhomogeneous and have complex pore structures, and often, there is a complex linear relationship between the logging curve and oil saturation. The prediction of oil saturation in deep learning can be categorized as a regression problem, often a multi-parameter regression problem. Deep learning has an absolute advantage over nonlinear problems compared to multiple fittings. More common deep learning algorithms mainly include random forest algorithms, fully connected neural networks, convolutional neural networks rec, current neural networks, and other branch algorithms.

This study investigates the literature related to the oil saturation evaluation of shale oil reservoirs using three methods from different fields: the electric method, the non-electric method, and a machine learning algorithm. The methods in each field have unique features, and their principles, technologies, and modeling processes are very different. Regarding the calculation and prediction of oil saturation in shale oil reservoirs, previous researchers have not systematically summarized and organized this method system; this study combs through the representative methods of three major fields and outlines, in detail, the current research status, principles, and applicability of these methods. Finally, the corresponding

application suggestions for these three methods and the outlook for future technology development are given. These can provide comprehensive technical ideas for calculating oil saturation in shale oil reservoirs and lay the foundation for shale oil reservoir evaluation, sweet spot evaluation, and oil test and recovery.

2. Current Status of Shale Oil Reservoir Saturation Logging Evaluation Research

2.1. Electrical Method Saturation Model

Archie's formula is the earliest saturation model proposed for pure sandstone formations and the most widely used saturation model in conventional reservoirs. Through petro-electric experiments, Archie found that the resistivity of water-bearing rocks is always proportional to the resistivity of the contained formation water. The scaling factor is called the formation factor, which is denoted by F. The resistivity of water-bearing rocks is always proportional to the resistivity of the contained formation water. The stratigraphic factor is only related to porosity, cementation, and pore shape, as shown in the following equation. For hydrophilic rocks, the rock resistivity is always proportional to the resistivity of the water-saturated rock. This proportionality coefficient is known as the resistivity increase factor, expressed as I. The resistivity increase factor is related to the oil saturation degree of the rock, and the relationship is given in the following equation.

$$F = \frac{R_0}{R_w} = \frac{a}{\varphi^m} \tag{1}$$

$$I = \frac{R_t}{R_0} = \frac{b}{{S_w}^n} \tag{2}$$

where R_0 is the rock resistivity ohm-m saturated with water, ohm-m. R_w is the formation water resistivity, ohm-m, a is a scaling factor related to lithology, m is the cementation index, and φ is the porosity. R_t is the rock resistivity, ohm-m, b is the coefficient, n is the saturation index, and S_w is the water saturation.

(i) Water saturation model taking into account the overall effect of sediment quality

Clay-rich shale oil reservoirs are the most typical feature. Due to the presence of mud and clay minerals, the electrical conductivity of shale reservoirs is greatly enhanced, leading to the problem of overestimation of water saturation. Many scholars have proposed several saturation models for the influence of mud according to its distribution form.

Dispersed mud is composed of clay particles, fragments, or crystals occupying the matrix pores of the rock skeleton, significantly reducing the effective porosity and permeability. DeWitte assumed that the formation is mainly composed of pore water and dispersed clay to form a conductive network and, thus, proposed a model for calculating the water saturation of dispersed muddy sandstone strata with the following specific formulas.

$$S_W = \frac{1}{1-q}\sqrt{\frac{aR_W}{\varphi^2 R_t} + \frac{q^2}{4}} - \frac{q}{2} \tag{3}$$

$$q = \frac{\varphi_S - \varphi_D}{\varphi_S} \tag{4}$$

where q is acoustic response, φ_S is acoustic porosity, and φ_D is density porosity.

Also, for dispersed clays, there is the Simandoux model (Figure 1), proposed by Simandoux et al. for application with dispersed clays. The model was derived from an experimental study in the Institut Français de Recherche Pétrolière (IFP) laboratory using a reservoir physical model simulating sands and clays. The Simandoux model assumes that there is also a certain amount of oil and gas storage properties in the clay fraction, whereas Archie's formula is used in sandstone. The Simandoux model equates the rocks into sandstone and mudstone based on a volume model and derives a conductivity model of rock resistivity versus mud volume based on the principle of parallel conductivity:

$$\frac{1}{Rt} = \frac{{S_W}^n V_{sh}}{R_{sh}} + \frac{{S_W}^n {\varphi_e}^m}{aR_W} \tag{5}$$

where V_{sh} is the mud content, %. ${\varphi_e}^m$ is the effective porosity, and R_{sh} is the resistivity of the fully cemented mass, ohm-m.

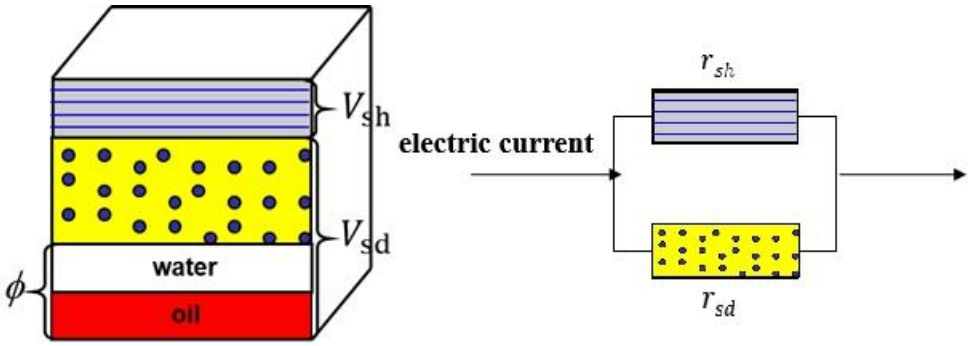


Figure 1. Schematic diagram of Simandoux' conductivity model [25].

In 1954, Poupon et al. described laminated mud as shale derived from the interaction of multiple thin layers of mudstone and sandstone within the vertical resolution of resistivity logging, and they noted that laminated shale does not affect the porosity or permeability of the sandstone itself. However, as the number of laminated shales increases, the same will cause the pore medium to decrease, and eventually, the porosity of the shale reservoir will decrease. As a result, Poupon proposed a simplified model for determining the water saturation model of laminated muddy sandstone formation. The specific formula is shown below:

$$S_w = \sqrt{\frac{aR_W(1 - V_{sh})}{\varphi^m}\left(\frac{R_{sh} - V_{Lam}R_t}{R_t R_{sh}}\right)} \tag{6}$$

where V_{Lam} is the volume of laminated mud in the formation, %. In 1971, Poupon and Leveaux, also for layered shale, proposed the Indonesian equation through field observation and analysis of several oil fields in Indonesia, with the following formula:

$$S_w = \frac{1}{R_t}\left(\frac{\sqrt{aR_W R_{sh}}}{V_{sh}\sqrt{aR_W + \varphi^{m/2}\sqrt{R_{sh}}}}\right) \tag{7}$$

All the parameters of the above equations are given as interpretations in the previous equations.

Schlumberger, to simplify the form of mud distribution in shale, based on Simandoux's experimental study and drawing on Poupon's field experience in the Niger Delta, proposed a model for estimating water saturation in shale, which simplifies the form of mud distribution in shale and can be applied to most shale formations. Since the model is straightforward and practical, many improved models have been developed for further research, and the earliest model is listed in this study as the following equation:

$$S_W = \frac{aR_w(1 - V_{sh})}{R_{sh}}\left[\left(\frac{-V_{sh}}{R_{sh}}\right) + \sqrt{\left(\frac{V_{sh}}{R_{sh}}\right) + \left(\frac{4\varphi^2}{aR_w R_t(1 - V_{sh})}\right)}\right] \tag{8}$$

(ii) Modeling of water saturation taking into account mud cation exchange capacity

The several saturation models described in the previous section are mainly from a macroscopic point of view to categorize the mud in the shale formation into different distribution forms to consider the influence of mud as a whole on the electrical conductivity of the reservoir. With the understanding of mud sandstone, more and more people began to

pay attention to the electrochemical properties of mud itself. Waxman and Smits found the phenomenon that the conductivity of cement-bearing sandstone increases sharply with an increase in the concentration of equilibrium electrolyte solution based on the experiments of Hill and Milburn, and they concluded that the conductivity of mud sandstone consists of the conductivity contributed by the exchange of the free electrolyte and the clay cations in the pore volume of the rock, which is composed of the conductivity contributed by the exchange of free electrolytes and clay cations in the rock. The conductivity contributed by the free electrolyte and clay cation exchange in the rock pore volume consists of the conductivity expressed by the following equation [26,27]:

$$C_o = xC_e + yC_w \tag{9}$$

where C_o is the conductivity of the fully hydrated sandstone, C_e is the conductivity contributed by clay cation exchange, C_w is the conductivity contributed by the equilibrium electrolyte solution, and x,y are constants.

Waxman and Smits concluded that the electrochemical efficiency of clayey sandstones increases with increasing clay content, and the conductivity is related to the clay cation exchange capacity, as confirmed in Hill and Milburn's experiments. The following equation can express the equation for the conductivity of saturated cemented sandstones:

$$C_o = \frac{1}{F^*}(BQ_v + C_w) \tag{10}$$

where F^* is the formation factor, B is the equilibrium cation-equivalent electrical conductance, mL/(Ω·m·mmol), and Q_v is the cation exchange capacity, mmol/mL.

Similarly, Waxman and Smits extended the above conductivity equations to oil- and salt-water-bearing argillaceous sandstones, concluding that when the rock is oil-bearing (sw < 1), the exchange cations bound to the clay are more concentrated in the pore water, and further describing the electrical conductivity of oil-bearing argillaceous sandstones in terms of a similar equation as above (W-S model):

$$C_t = \phi^{m^*} Sw^{n^*}(C_\mathrm{w} + BQ_V/S_W) \tag{11}$$

where C_t is the conductivity of hydrocarbon-bearing mud sandstone, S/m; m^* is the cementation index in the W-S model, and n^* is the saturation index in the W-S model.

In a subsequent study, Clavier pointed out that the W-S model has certain shortcomings, such as the fact that the W-S model does not consider the salt removal effect of adsorbed water on the clay surface. Clavier noted, based on the theory of a diffusion bilayer (as shown in Figure 2), that the water in the muddy sandstone for the clay adsorbed water and free water, so the electrical conductivity of the muddy sandstone is the parallel conductivity of the clay adsorbed water and free water, according to which the two-water model is proposed:

$$C_t = \frac{1}{F^o} {S_W}^n \left[C_W + \frac{aV_Q Q_V}{S_W}(C_{cw} - C_w) \right] \tag{12}$$

where F^O is the stratigraphic factor in the two-water model; a is the coefficient; V_q is the amount of clay water per unit pore volume when Q_v is 1, mL/mmol; and β is the equilibrium cation-equivalent conductivity, (S/m) (mmol/mL).

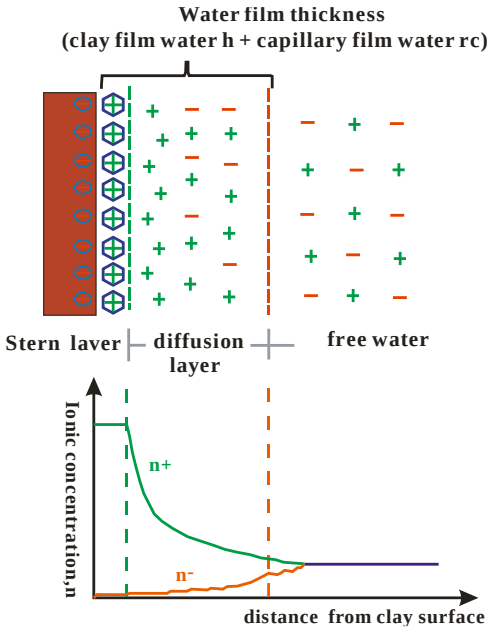


Figure 2. Principle of diffuse double-electric layer [28].

2.2. Non-Electrical Saturation Models

Subject to clay content, shale reservoirs often exhibit low resistance, resulting in lower resistivity curves than the actual formation resistance. The non-electrical saturation interpretation model can directly bypass the resistivity as a factor in calculating the saturation of shale oil reservoirs. Nuclear magnetic resonance (NMR) logging is unaffected by reservoir electrical properties and lithology. NMR is currently an indispensable technology for evaluating the physical properties and oil content of unconventional reservoirs. This section focuses on applying NMR technology to calculate oil saturation in shale oil reservoirs.

(iii) One-dimensional nuclear 3magnetic resonance (NMR) logging saturation interpretation models

Usually, the one-dimensional NMR logging saturation interpretation model is mainly based on the core analytical saturation and core analytical porosity to calculate the oil-bearing porosity of the core. Then, the T2 spectral area truncation method is used to scale the optimal oil-bearing porosity starting time, and we calculate the oil-bearing saturation according to the following formula.

$$S_o = PHI_{OIL}/PHIE \tag{13}$$

where S_o is the oil-bearing saturation, PHI_{OIL} is the oil-bearing porosity, and $PHIE$ is the NMR-calculated effective porosity.

In addition, a more direct way can be used through the analysis of the NMR T2 spectrum. The core saturation is directly used to scale the T2 cutoff value and then determine the oil-bearing volume of the shale reservoir in the T2 spectrum to obtain the oil-bearing saturation, in which the T2 cutoff value is the most critical step. Generally, for the same set of strata in a region, the T2 cutoff value should be close, and different cutoff values can be designed to calculate the oil saturation using the following formula and the T2 cutoff value at the time of minimizing the average absolute error by traversing the method. The following figure shows a schematic diagram for determining the T2 cutoff value/the average absolute error is minimized when T2 = 10 ms, so the T2 cutoff value for oil saturation calculation is 10 ms (Figure 3).

$$So = \frac{\int_{T_{2cutoff}}^{T_{2max}} S(T)dt}{\int_{T_{2min}}^{T_{2max}} S(T)dt} \quad (14)$$

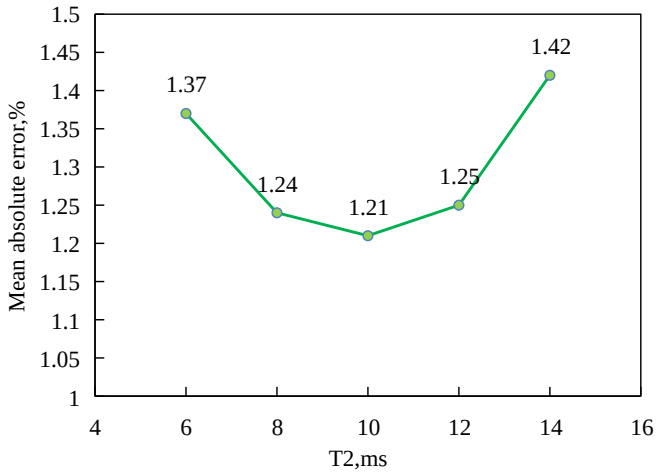


Figure 3. Calculation error diagram for different starting times.

(iv) Two-dimensional nuclear magnetic resonance (NMR) logging saturation interpretation models

For shale oil reservoirs, the oil and water storage state is complicated, and the micro and nano pores of organic matter are developed. Thus, it is difficult to distinguish the oil and water signals using one-dimensional NMR technology. Fluid signals will overlap in the transverse relaxation time, and the oil saturation calculated often has some errors. Two-dimensional NMR technology can obtain the relaxation information of hydrogen nuclei of different components in different shales, including the longitudinal relaxation time T1 and transverse relaxation time T2. Usually, the calculation of oil saturation using two-dimensional NMR technology relies on the T1-T2 two-dimensional NMR technology, in which the main methods can be summarized into two kinds: one is them is the plate method based on the region's experience, and the other is to avoid the time-consuming experiments and interpretation processes. Many researchers use signal processing methods. Venkataramanan et al. [29,30], in 2018, used an unsupervised clustering algorithm to separate the fluid components of the T1-T2 maps by decomposing a complete signal into multiple signals, which is also known as the blind source separation method. Subsequently, the oil saturation was calculated based on the location of individual fluid signals in the region as well as the volume. The plate method is based on regional experience and core NMR experiments. The T1-T2 NMR plates of multiple cores are analyzed to summarize individual fluid signals' relaxation signal characteristics to formulate the region's T1-T2 fluid component identification plate (shown in Figures 4 and 5). Then, the oil saturation degree of the shale reservoir is determined by the ratio of the sum of the signal volumes of free oil and adsorbed oil in the plate to the total signal, with the specific formula as follows:

$$S_O = \frac{F_{\text{absorbent oil}} + F_{\text{free oil}}}{F_t} \quad (15)$$

where $F_{\text{adsorbet oil}}$ is the adsorbed oil signal quantity, $F_{\text{free oil}}$ is the free oil signal quantity, F_t total signal quantity.

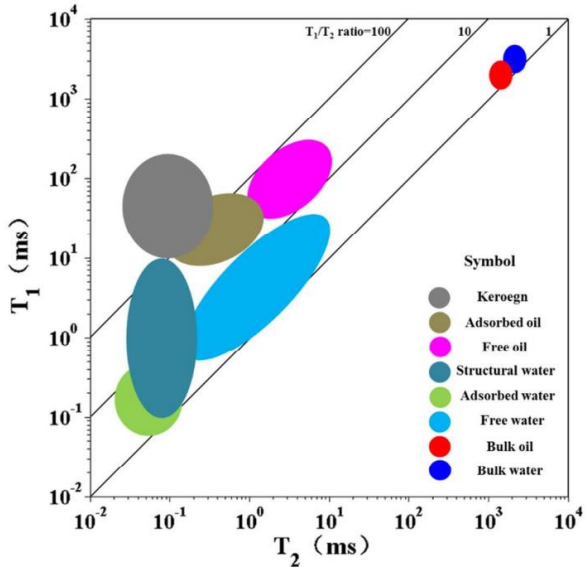


Figure 4. T2-T1 2D NMR fluid identification plate (Li et al., 2018 [31]).

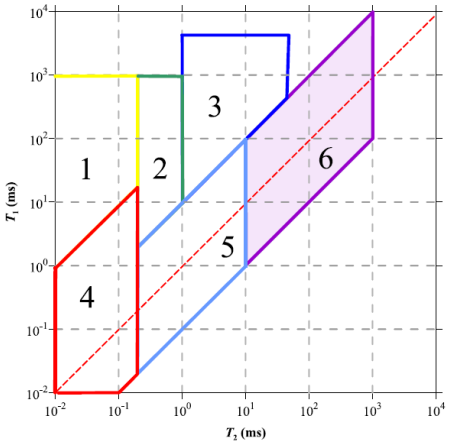
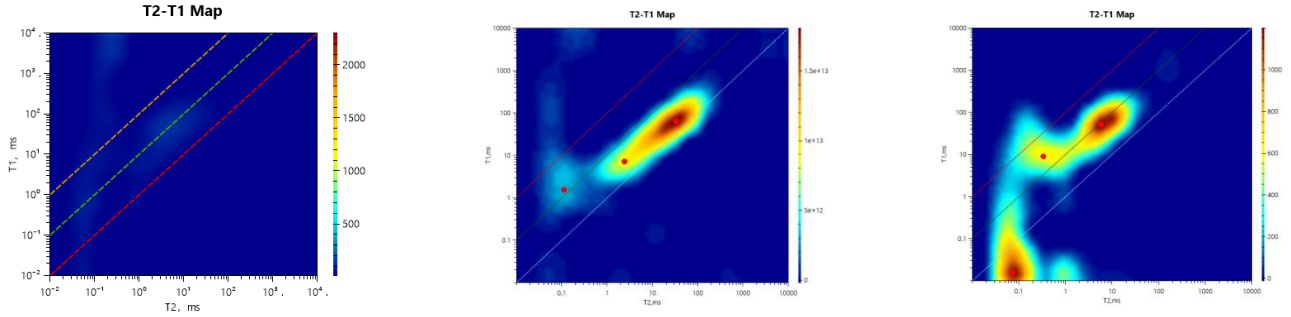


Figure 5. T_2-T_1 2D NMR fluid identification plate (1—kerogen/asphalt; 2—bound oil; 3—mobile oil; 4—bound water; 5—mobile water; 6—mobile water in macropore; Li et al., 2020 [32]).

Jiacheng Zhang [33] and Jiawei Zhang [34] combined a two-dimensional component decomposition method with a multi-state two-dimensional NMR experiment on rock cores. The primary process is to wash the core's original sample with oil and salt, followed by drying. After drying, the dry sample is saturated with water and oil to obtain the T2-T1 spectral distribution state of oil and water, which can roughly determine the location of oil and water signals. Figure 6 shows the T2-T1 spectral map of the core in different states of the Fengcheng Formation in the Mabei area of the Junggar Basin. Based on the above experimental analysis, to determine the fluid components more accurately and quantitatively, it is necessary to use the two-dimensional component decomposition method, which, at present, mainly adopts the image morphology technology, to find the peak point of each fluid component, and then we use Gaussian fitting or K-mean clustering and other algorithms to obtain the boundaries of each fluid component, so that the specific location of each element is obtained (shown in Figure 7). Using the above-mentioned core multi-state 2D NMR experiments, each fluid component is calibrated, and then the oil saturation is obtained according to the ratio of the total adsorbed oil and free oil signals using the above-mentioned formula. Through the 2D NMR experiments, Qin Yingyao summarized the T2-T1 2D NMR fluid identification plates of a few typical land-phase shale oil blocks in China, shown in Figure 8 below.



Drying status Water-saturated state Oil-saturated state

Figure 6. T2-T1 spectra of cores in different states of the Marbei Fengcheng Formation, Junggar Basin.

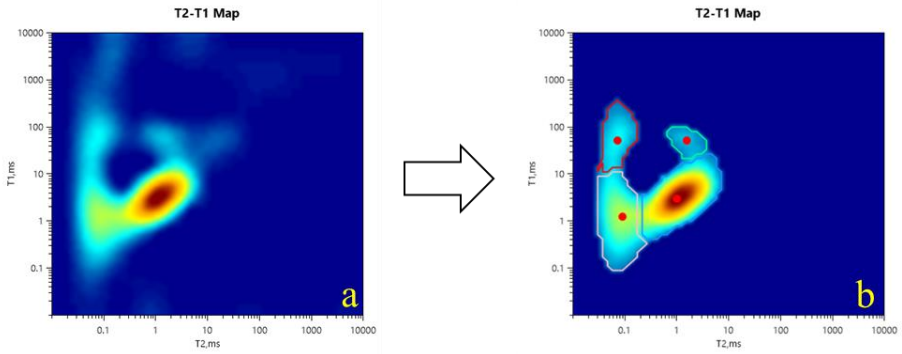


Figure 7. (**a**) Original T_2-T_1 two-dimensional spectrum; (**b**) T_2-T_1 two-dimensional spectrum after boundary division.

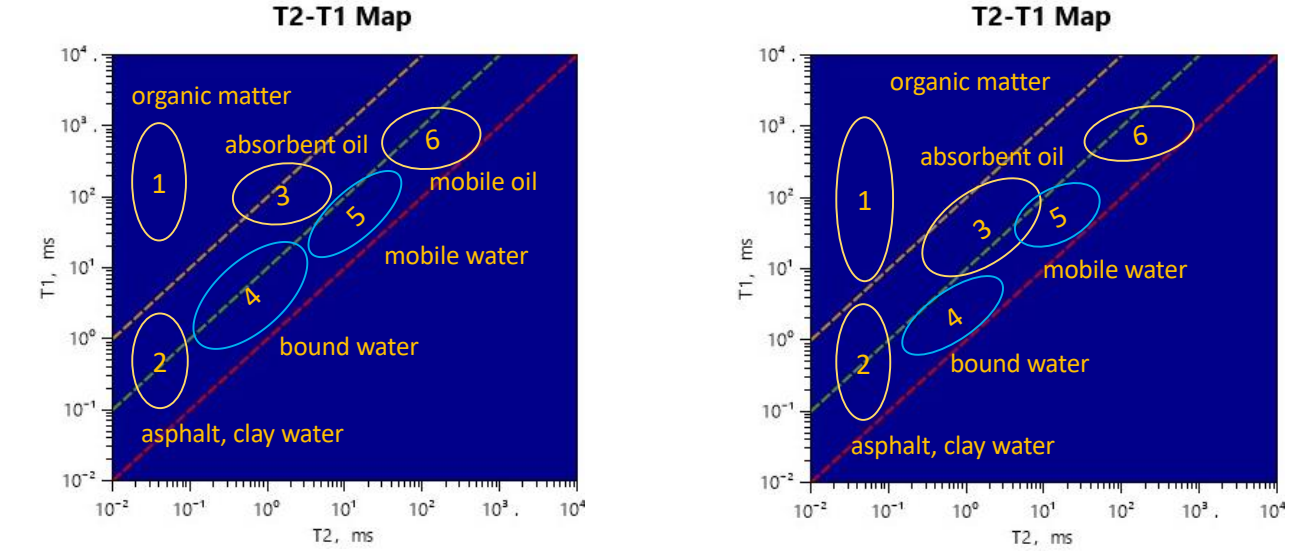


Plate T2-T1 of the Luchaogou Formation, Jimushar Region Plate T2-T1 of the Yanchang Formation, Chang 7 Region

Figure 8. *Cont.*

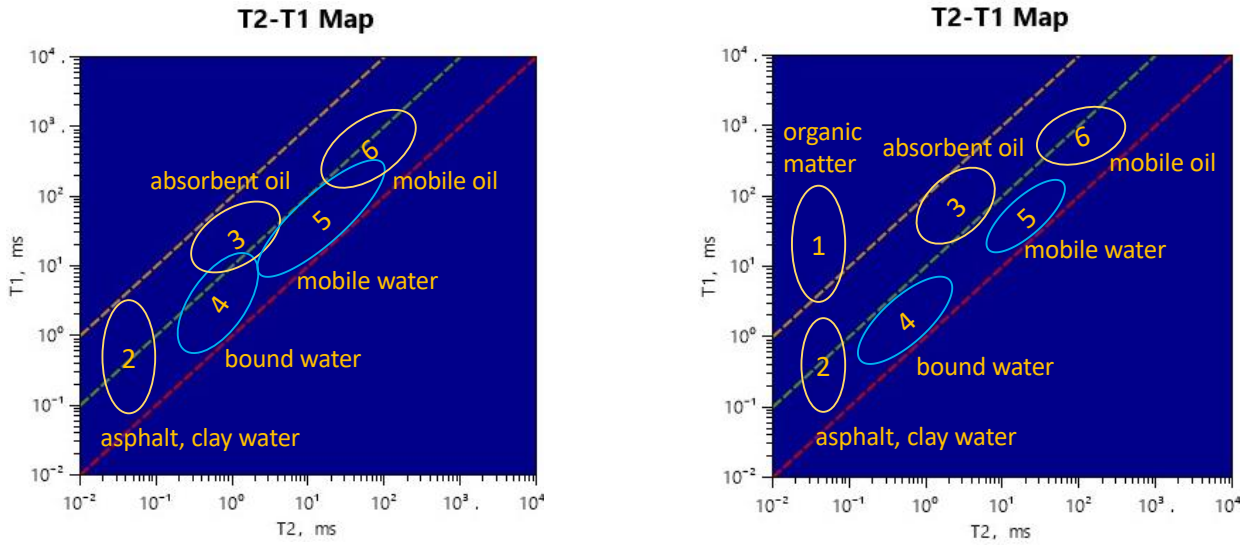


Plate T2-T1 of the Fengcheng Formation, Mabei Region

Plate T2-T1 of the Qingshankou Formation, Gulong Region

Figure 8. Two-dimensional NMR T2-T1 plate of typical shale oil blocks in China [35].

2.3. Machine Learning Models

Saturation prediction based on machine learning algorithms can be viewed as a regression task in mathematical modelling, in which multiple regression fitting based on statistical analysis is the most commonly used technique. With the in-depth exploration and development of unconventional reservoirs such as shale oil, shale oil reservoirs have several prominent features compared with conventional reservoirs. One is the complexity of mineral types and the rapid change in lithological complexity; the second is the complexity of the pore structure and the development of nanoscale pores; and the third is the complexity of wettability and the unclear state of oil and water fugacity. These three characteristics cause the shale oil quantitative characterization accuracy to be low, the logging curve response characteristics and formation oil saturation is no longer a simple linear relationship, and the traditional multiple linear regression model is no longer applicable. Machine learning algorithms are increasingly used in saturation prediction with the gradual improvement in computer hardware and software performance. Machine learning algorithms can be summarized into two kinds. One is shallow learning algorithms, such as random forest [36–38], support vector machine [39], K-nearest neighbor [40], naive Bayes [41], and other algorithms. These algorithms generally apply to fewer coring data, high data quality, and reservoir homogeneity of good reservoirs.

Developing grain layers, rapid lithological changes, and non-homogeneity for shale oil reservoirs often require deep learning algorithms with deeper network structures. The deep learning networks currently used for saturation prediction are mainly deep neural networks, recurrent neural networks, and deep learning networks based on migration learning [42,43]. Figure 9 shows the saturation prediction model based on deep neural networks, generally using the input feature data for conventional logging curves. These curves constitute the input layer; the hidden layer consists of several neurons, which can contain the activation function, pooling layer, and the full link layer, etc., mainly used to make nonlinear changes and process the input data to achieve the purpose of feature extraction. The output layer in the general regression task mainly uses the fully connected layer, which connects with all the neurons in the previous layer and outputs the predicted value. The whole deep neural network operation process can be summarized as follows: the input layer is responsible for accepting the logging curve values and passing them to the hidden layer, which nonlinearly changes the data through the activation function and generates new features, and then the output layer is responsible for outputting the

predicted saturation, and the training process is to take the confined core saturation as a label and, through the error computation, use the gradient descent algorithm to update the weight and bias continually. The optimal parameters save the optimal model, that is, on behalf of the completion of training; after the completion of training, the logging curve input to the network can obtain the saturation prediction value.

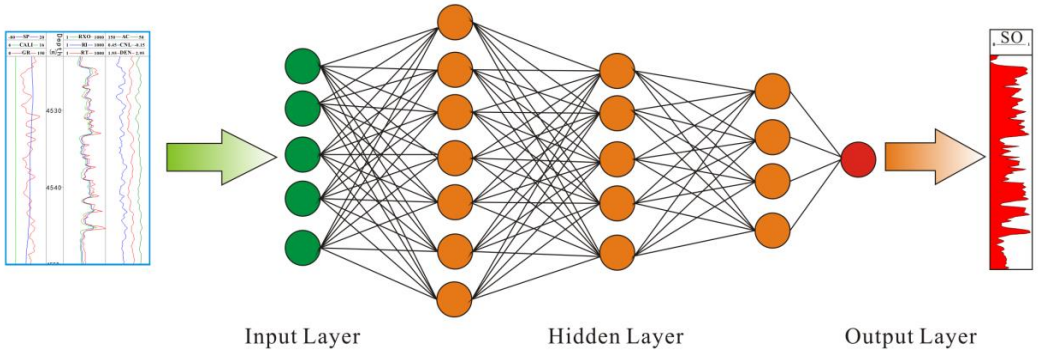


Figure 9. Schematic diagram of deep neural network model saturation prediction.

A recurrent neural network has a strong advantage in the task of processing time-series data, which is usually used in text processing, translation, etc. Saturation, as an important parameter of the reservoir, can reflect the depositional characteristics of acoustic, electric, radioactive, etc., of each geological period to a certain extent, so the saturation prediction can also be viewed as the task of time-series data prediction. Figure 10 shows the recurrent cell structure of the LSTM network at the moment t, where x_t is the input at the moment t; h_{t-1} is the output of the hidden layer for the moment $t-1$; c_{t-1} is the storage unit retained at the moment $t-1$; f_t, i_t, and O_t are the activation vectors of the forgetting gate, the input gate, and the output gate at the moment t, respectively; c'_t is the neuron cell state vector obtained after the tanh layer transformed neuron cell state vector; c_t is the storage unit updated at moment t; h_t is the output of the hidden layer at moment t; and σ is the sigmoid activation function. As shown in Figure 10, firstly, we need to copy four copies of the output of the hidden layer at the moment $t-1$ and the input at the moment t respectively, randomly initialize different weights, and then splice the h_{t-1} with the x_t vector, and, finally, nonlinearly transform by the activation function, sigmoid function, or the tanh function to obtain the oblivious gate activation vector, the input gate activation vector, the output gate activation vector and the transformed neuron cell state vector. Figure 11 exhibits a saturation prediction flowchart of the long- and short-term memory recurrent neural network. The long- and short-term memory recurrent neural is a variant of the recurrent neural network, which avoids the problem of long-range dependence on the recurrent neural network by adding a gating mechanism, which allows the recurrent unit to save useful historical information and forget useless historical information selectively. The figure details the workflow from the original logging curve to the saturation prediction value. It can be seen that before the original logging curve is input to the network, the original logging curve first needs to carry out the pre-data processing, such as the multi-well standardization, the spreading correction, the core homogenization, the data cleaning, and other steps, to ensure the reliability of the data. Secondly, the core analysis saturation and logging curve need to be analyzed for sensitivity to determine the preferred logging curve. Subsequently, the training set and test set are divided proportionally. After standardization or normalization, it is inputted into the network for training. The final saturation prediction is obtained based on the trained model.

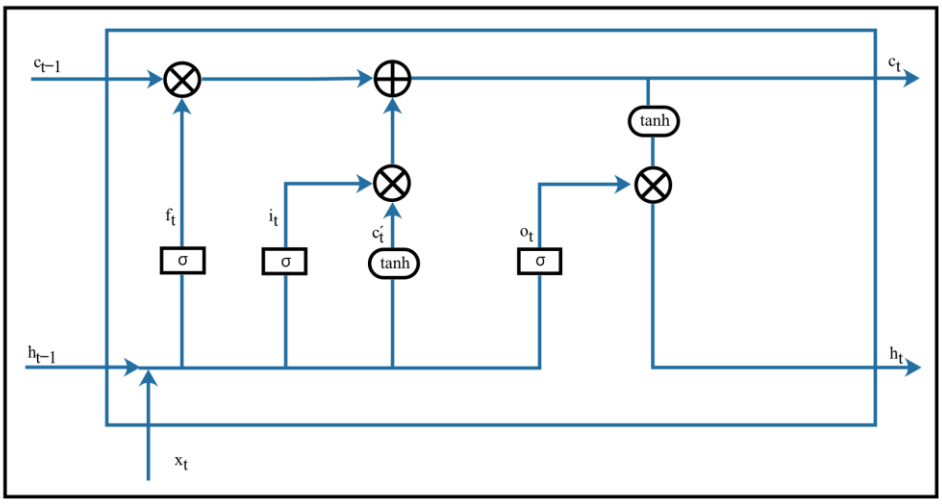


Figure 10. Working principle of LSTM network at t time.

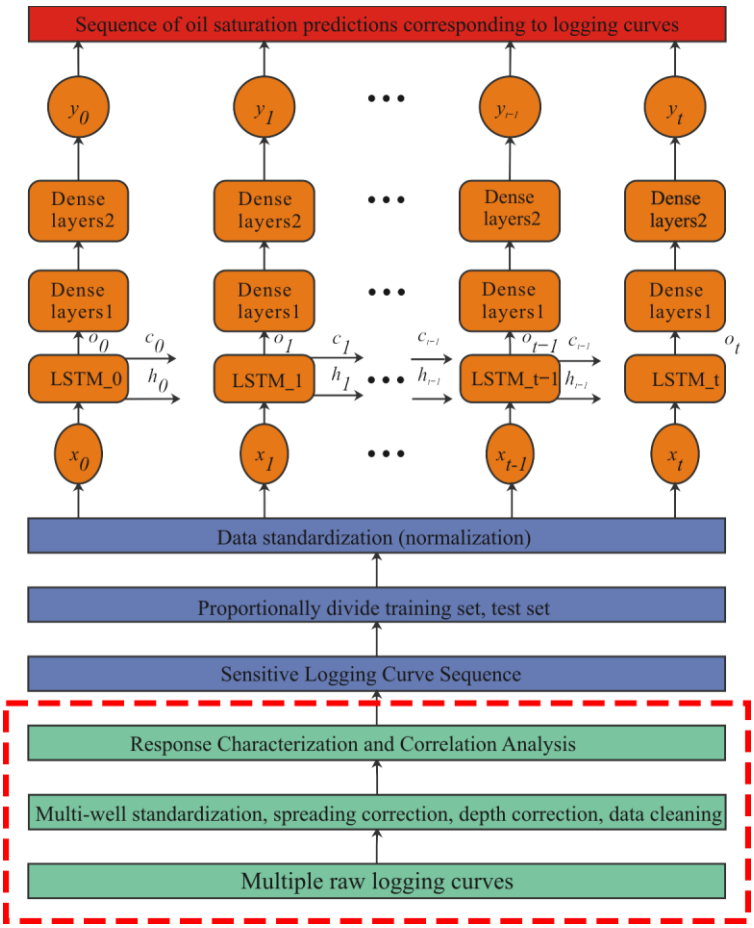


Figure 11. Schematic of the saturation prediction of the neural network for long- and short-term memory cycles [44].

3. Analysis of the Applicability of Typical Saturation Models

The previous subsection described each model's research progress, construction principle, and application mode from three aspects: electric method, nuclear magnetism, and machine learning. In this section, several typical saturations are selected to evaluate the saturation logging of shale oil reservoirs in the Qingshankou Group of Gulong Depression, and the relatively accurate saturation of the confined core is used for comparison to verify

the calculation effect of each model. The reservoir of Qingshankou Formation in Gulong Depression is a typical terrestrial shale oil reservoir with a high clay content (30–40%), small reservoir space, and predominantly micro- and nano-pore spaces. The hydrocarbon source rock is characterized by the development of organic matter and high maturity, mainly dominated by Type I and Type II_I organic matter. Figure 12 shows the comparative results of the saturation calculation of G1 wells in the Qingshankou Formation of the Gulong Depression. In this study, the total shale model (eighth pass), Indonesia model (ninth pass), Simandoux model (tenth pass), and double-water model (eleventh pass) in the saturation model of the electric method are selected.

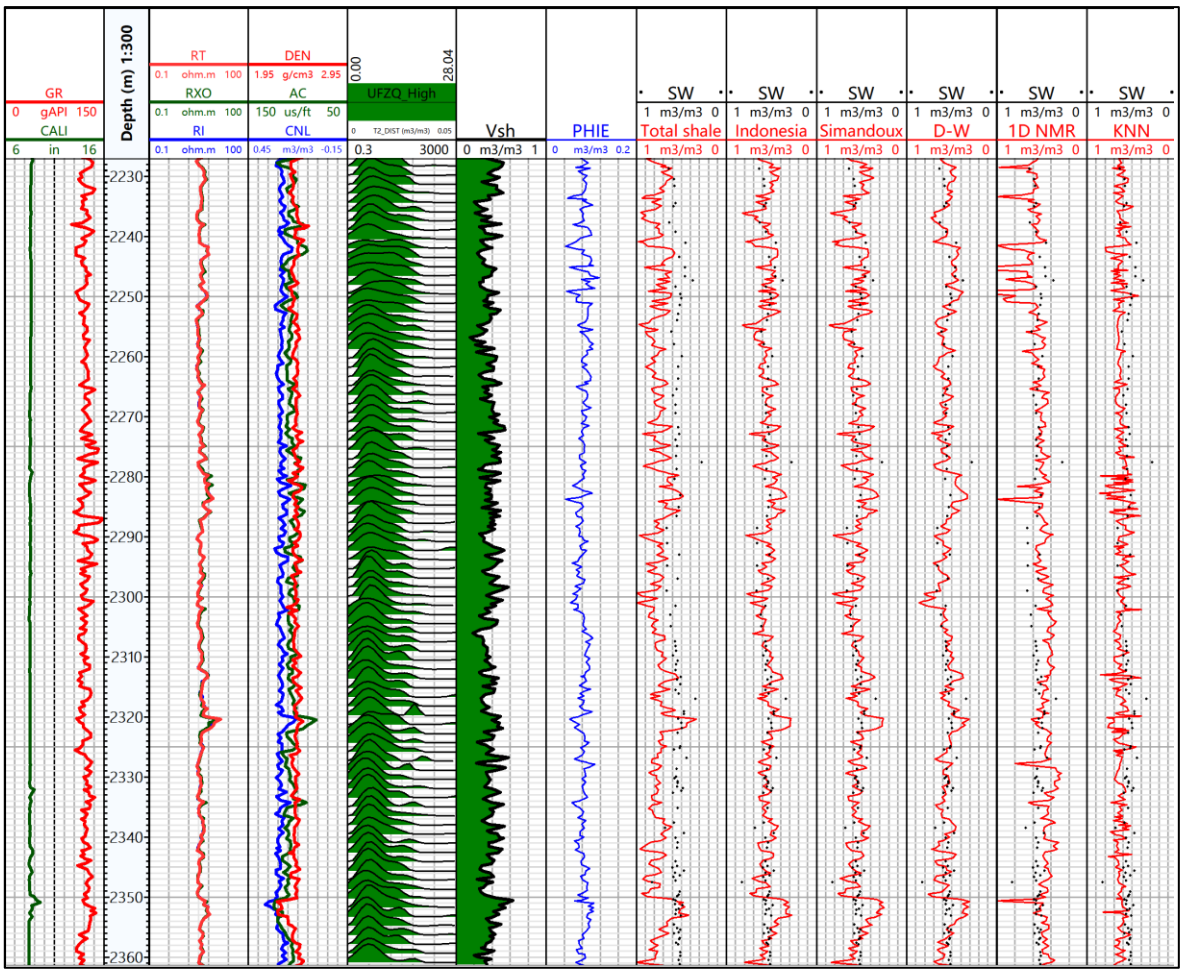


Figure 12. Comparison of saturation calculation of G1 wells in Qingshankou Formation, Gulong Depression.

Among the NMR saturation models, the one-dimensional NMR saturation model (twelfth pass) was chosen in this work. In the machine learning model, the more classical KNN model (the thirteenth course) was selected in this study. The rock electric parameters and stratigraphic used in the electrical model are consistent. The prediction results of the total shale model could be better calculated compared to the other three electrical method models, with an absolute error of 19% and a relative error of 38%, see Table 1. The overall calculation results of the Indonesian model and the Simandoux model are comparable. The difference between the two models is that the Simandoux model tends to be more sensitive in the layers with lower or higher resistivity, calculating higher oil saturation at higher resistivity and predicting lower oil saturation at lower resistivity, which makes the Indonesian model more stable in one sense. Still, the Simandoux model can express the oil content of the stratum more favorably from another perspective—the oil content of the

formation. The two-water model calculates the saturation and core saturation where the depth change trend is basically the same, its calculation absolute error is 7%, and relative error is 12%, so there is, relatively speaking, higher accuracy. However, its model requires more parameters, because the lack of cation exchange experimental data for the block is difficult to apply. The one-dimensional NMR model is calculated by focusing on the core well data, and the cutoff value of oil-bearing porosity determined by analysis is 6ms. Overall, the error is relatively large, mainly because the 1D NMR T2 spectrum is greatly affected by clay minerals and pore structure, and the cutoff values of different layers in different planar positions have certain changes, so it is necessary to carry out the modelling work of the cutoff value of oil-bearing porosity. The last KNN model calculation results show a relatively large error and do not have a good response trend with the analyzed oil saturation. This is mainly because the research area has less available training data, which cannot meet the requirements of the deep network model application. The relatively shallow KNN model has a limited ability to express nonlinear mapping.

Table 1. Comparison of errors in calculating oil saturation by models.

Model	Total Shale	Indonesia	Simandoux	D-W	1D NMR	KNN
absolute error	18.6%	8.3%	8.6%	7.3%	10.8%	9.2%
relative error	35.1%	13.1%	14.2%	11.5%	19.6%	17.4%

4. Conclusions and Outlook

In this study, the research progress of oil saturation in shale oil reservoirs is elaborated on in three aspects: the electric method model, the nuclear magnetic model, and the machine learning model. It is further subdivided and summarized in these three aspects. Finally, several typical saturation models are selected for calculating oil saturation in shale oil reservoirs of the Qingshankou Formation in the Gulong Depression and are compared and analyzed. The following conclusions are drawn after further analysis, summarization, and sorting.

(1) Electrical modelling mainly focuses on mud's overall and microscopic effects, such as mud content, mud resistivity, clay cation exchange capacity, and other factors. For shale oil and water storage state and wettability, whether the large and small pores contain oil and wettability characteristics of different large and small pore walls still need to be considered. In the future, an in-depth analysis of the oil–water storage state and wettability of shale oil reservoirs can be carried out, and an electrical conductivity model considering the characteristics of the oil–water storage state and wettability can be established.

(2) For the NMR oil saturation model, the one-dimensional NMR T2 spectrum often makes it difficult to distinguish between the signals of oil and water and the signals of oil and water overlapping in the transverse relaxation time. Meanwhile, due to the changes in clay minerals and the formation of the pore structure, the one-dimensional NMR model based on a fixed cutoff value does not have a strong generalization, and it is challenging to meet the requirements for saturation calculation accuracy. For the two-dimensional NMR model, the T2-T1 mapping currently relies on the core two-dimensional NMR experimental calibration to establish the fluid identification plate and then calculate the saturation degree. In the future, it is necessary to combine semantic segmentation or instance segmentation in computer vision with 2D NMR mapping to accurately recognize oil, gas, and water signals and their signal quantities and then calculate the saturation degree.

(3) For machine learning models, traditional shallow algorithms are less affected by the sample size, but the prediction accuracy is not enough, and the deep learning models have strong nonlinear expression ability; however, they are greatly affected by the number of samples, and there are often fewer coring data, which cannot meet the deep learning modelling needs. In the future, some optimization algorithms and

deep learning networks can be fused, or different neural networks can be fused with each other to mine as much useful information as possible in the case of limited data samples.

Funding: This research was funded by [China National Petroleum Corporation (CNPC) major scientific and technological special project on "Onshore Shale Oil Scale Reserve Increase and Production, Exploration and Development Technology Research"] grant number [2023ZZ15].

Conflicts of Interest: Authors Haibo Zhao, Jiandong Zheng, Zhaoqian Zhang, Hao Qin, Chunyan Wang, Jianhua Zhu, Anqi Chen, Longchuan Chen were employed by the company PetroChina Daqing Oilfield Co., Ltd. Authors Peng Liu, Wangwang Yang and Xuan Hu were employed by the company Xinjiang Oilfield Company, PetroChina. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The companies had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, N.; Feng, Z.; Wu, H.; Tian, H.; Liu, P.; Liu, Y.; Liu, Z.; Wang, K.; Xu, B. New advances in methods and technologies for well logging of continental shale oil in China. *Acta Pet. Sin.* **2023**, *44*, 28–44.
2. Zou, C.; Yang, Z.; Zhang, G.; Hou, L.; Zhu, R.; Tao, S.; Yuan, X.; Dong, D.; Wang, Y.; Guo, Q.; et al. Conventional and unconventional Petroleum "orderly accumulation": Concept and practical significance. *Pet. Explor. Dev.* **2014**, *41*, 14–26. [CrossRef]
3. Suyun, H.; Wenzhi, Z.; Lianhua, H.; Zhi, Y.; Rukai, Z.; Songtao, W.; Bin, B.; Xu, J. Development potential and technical strategy of continental shale oil in China. *Pet. Explor. Dev.* **2020**, *47*, 819–828.
4. Li, Q.; Wang, Y.; Wang, F.; Wu, J.; Usman Tahir, M.; Li, Q.; Yuan, L.; Liu, Z. Effect of thickener and reservoir parameters on the filtration property of CO2 fracturing fluid. *Energy Sources* **2020**, *42*, 1705–1715. [CrossRef]
5. Li, Q.; Cheng, Y.; Li, Q.; Zhang, C.; Ansari, U.; Song, B. Establishment and evaluation of strength criterion for clayey silt hydrate-bearing sediments. Energy sources, Part A. *Recovery Util. Environ. Eff.* **2018**, *40*, 742–750.
6. Ning, L.I.; Weilin, Y.A.N.; Hongliang, W.U. Current situation, Problems and countermeasures of the well-logging evaluation technology for Gulong shale oil. *Pet. Geol. Oilfield Dev. Daqing* **2020**, *39*, 117–128.
7. Archie, G.E. The electrical resistivity log as an aid in determining some reservoir characteristics. *Trans. Aime* **1942**, *146*, 54–62. [CrossRef]
8. DeWitte, L. Relations between resistivities and fluid contents of porous rocks. *Oil Gas. J.* **1950**, *49*, 120–134.
9. Poupon, A.; Loy, M.E.; Tixier, M.P. A Contribution to electrical log interpretation in shaly sands. *J. Petrol. Technol.* **1954**, *6*, 27–34. [CrossRef]
10. Simandoux, P. Mesuresd ielectriques en milieu poreux, application a mesure des saturations en eau, Etude du Comportment des massifs Argileux. Supplementary Issue. *Rev. Dei'institut Fr. Du.Petrol.* **1963**, 193–215.
11. Waxman, M.; Smits, L. Electrical Conductivities in Oil-Bearing Shaly Sands. *Soc. Pet. Eng. J.* **1968**, *8*, 107–122. [CrossRef]
12. Poupon, A.; Leveaux, J. Evaluation of water saturation in shaly formations. In Proceedings of the The SPWLA 12th Annual Logging Symposium, Dallas, TX, USA, 2–5 May 1971; *SPWLA*-1971-O.
13. Schlumberger. *Log Interpretation*; Schlumberger: New York, NY, USA, 1972; Volume 1—Principles, Chapter 16.
14. Clavier, C.; Coates, G.; Dumanoir, J. Theoretical and Experimental Bases for the Dual-Water Model for Interpretation of Shaly Sands. *Soc. Pet. Eng. J.* **1984**, *24*, 153–168. [CrossRef]
15. Zhang, L.H.; Pan, B.Z.; Li, Z.B.; Mo, X.W.; Xia, Z.L.; Xu, W.L. New three-water conduction model and its application in evaluation of low porosity and permeability reservoir. *OGP* **2010**, *45*, 431–435.
16. Kadkhodaie, A.; Rezaee, R. A new correlation for water saturation calculation in gas shale reservoirs based on compensation of kerogen-clay conductivity. *J. Pet. Sci. Eng.* **2016**, *146*, 932–939. [CrossRef]
17. Yu, C.; Zhang, C.; Zhang, Z.; Wang, K.; Shi, W.; Zhou, X.; Liu, W. Analysis of low- resistivity genesis and saturation calculation of marine-continental transitional shale: A case study of the Longtan Formation in A Area form southeast of Sichuan Basin. *J. Northeast. Pet. Univ.* **2022**, *46*, 55–68.
18. Liu, Y.H.; Wang, C.Z.; Liu, Z.H.; Wang, H.; Liu, Y.M. A logging method for evaluating oil-bearing property of Jimsar shale oil: Case study of Sag in Junggar Basin. *Nat. Gas Geosci.* **2021**, *32*, 1084–1091.
19. Zeng, J.; Zheng, Q.; Xie, X.; Yang, F.; Cheng, L.; Xia, Y. New method of saturation calculation based on NMR experiment and application. *Prog. Geophys.* **2023**, *38*, 1228–1237.
20. Yan, W.; Zhang, Z.; Chen, L.; Zhao, Z.; Wang, W. New evaluating method of oil saturation in Gulong Shale based on NMR technique. *Pet. Geol. Oilfield Dev. Daqing.* **2021**, *40*, 78–86.
21. Qin, Y.; Zhang, G.; Luo, C.; Zhang, J.; Wang, Z. Two-dimensional NMR characteristics of Jimsar shale oil reservoir. *J. Cent. South. Univ.* **2022**, *53*, 3387–3400.

22. Li, N.; Xu, B.; Wu, H.; Feng, Z.; Li, Y.; Wang, K.; Liu, P. Application status and prospects of artificial intelligence in well logging and formation evaluation. *Acta Pet. Sin.* **2021**, *42*, 508–522.
23. Nikravesh, M.; Aminzadeh, F. Past, present and future intelligent reservoir characterization trends. *J. Pet. Sci. Eng.* **2001**, *31*, 67–79. [CrossRef]
24. Alkinani, H.H.; Al-Hameedi, A.T.; Dunn-norman, S. Applications of artificial neural networks in the petroleum industry: A review. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 18–21 March 2019; SPE: Richardson, TX, USA, 2019; 195072.
25. Meng, L.; Shi, X.; Yang, W.; Zhang, C.; Xu, W. Study on the Evaluation Method of the Water Saturation Logging of a Low-Resistance Oil Reservoir in the Guantao Formation, Bohai Basin. *Processes* **2023**, *11*, 2890. [CrossRef]
26. Hill, H.J.; Klein, G.E.; Shirley, O.J.; Thomas, E.C.; Waxman, W.H. Bound water in shaly sands: Its relation to Qv and other formation properties. *Log. Anal.* **1979**, *20*, 3–19.
27. Wang, L.; Li, Y.; Ren, L.; Song, L.; Su, J. Comparison of W-S Model and Dual Water Model in Describing the Additional Conductivity of Argillaceous Sandstone. *Well Logging Technol.* **2018**, *42*, 622–628.
28. Cui, Y.; Ming, J.; Shi, X.; Yang, W.; Zhang, Z.; Zhang, C. A New Method for Calculating Reservoir Core-Bound Water Saturation Using the Cast Thin Section. *Processes* **2023**, *11*, 1397. [CrossRef]
29. Jiang, T.; Phillips, C.; Bonnie, R.J.; Correa, T.S.; Krueger, M.C.; Kelly, S.A.; Wasson, M.S. Integrated Reservoir Characterization Using Unsupervised Learning on Nuclear Magnetic Resonance (NMR) T1-T2 Logs. *Petrophysics–SPWLA J. Form. Eval. Reserv. Descr.* **2022**, *63*, 277–289. [CrossRef]
30. Venkataramanan, L.; Evirgen, N.; Allen, D.F.; Mutina, A.; Cai, Q.; Johnson, A.C.; Green, A.Y.; Jiang, T. An Unsupervised Learning Algorithm to Compute Fluid Volumes From NMR T1-T2 Logs in Unconventional Reservoirs. *Petrophysics–SPWLA J. Form. Eval. Reserv. Descr.* **2018**, *59*, 617–632. [CrossRef]
31. Li, J.; Huang, W.; Lu, S.; Wang, M.; Chen, G.; Tian, W.; Guo, Z. Nuclear Magnetic Resonance T1–T2 Map Division Method for Hydrogen-Bearing Components in Continental Shale. *Energy Fuels* **2018**, *32*, 9043–9054. [CrossRef]
32. Li, J.; Jiang, C.; Wang, M.; Lu, S.; Chen, Z.; Chen, G.; Li, J.; Li, Z.; Lu, S. Adsorbed and free hydrocarbons in unconventional shale reservoir: A new insight from NMR T1-T2 maps. *Mar. Pet. Geol.* **2020**, *116*, 104311. [CrossRef]
33. Zhang, J. *Research and Application of Nuclear Magnetic Resonance Data Processing Methods for Shale Oil Reservoirs*; Yangtze University: Jingzhou, China, 2023.
34. Zhang, J.; Liao, G.; Xiao, L.; Luo, S. Quantitative Evaluation for Fluid Components on 2D NMR Spectrum Using Image Boundary Tracking And Modified GMM Clustering Method. In *SPWLA Annual Logging Symposium*; SPWLA: Houston, TX, USA, 2024; *SPWLA*-2024-0090.
35. Qin, Y. *NMR Response Mechanism and Evaluation Method for Shale Oil Reservoirs*; Yangtze University: Hubei, China, 2023.
36. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Huang, L.S.; Yan, J.P.; Guo, W.; Zheng, M.J.; Zhong, G.H.; Huang, Y. Evaluation of Low Resistivity Shale Gas Reservoir Saturation Based on Random Forest Regression Method. *Well Logging Technol.* **2023**, *47*, 22–28.
39. Zhou, Z. A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **2018**, *5*, 44–53. [CrossRef]
40. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1953**, *13*, 21–27. [CrossRef]
41. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Statal Assoc.* **2008**, *103*, 681–686. [CrossRef]
42. Shao, R.; Xiao, L.; Liao, G.; Shi, Y.; Zhou, J.; Li, G.; Hou, X. Multitask Learning based reservoir parameters prediction with geophysical logs. *Chin. J. Geophys.* **2022**, *65*, 1883–1895.
43. Shao, R.; Xiao, L.; Liao, G.; Zhou, J.; Li, G. A reservoir parameters prediction method for geophysical logs based on transfer learning. *Chin. J. Geophys.* **2022**, *65*, 796–808.
44. Yang, W.; Zhang, C.; Yang, M.; Zhang, Y.; Wang, M.; Sun, K. Permeability logging evaluation of carbonate reservoirs in Oilfield H of Iraq in the Middle East based on long short-term memory recurrent neural network. *Pet. Geol. Oilfield Dev. Daqing* **2022**, *41*, 126–133.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.