

## Research Article

# Optimization of Logging Interpretation Parameters Based on Committee Machine

Ziyi Na <sup>1</sup>, Lixia Dang <sup>1</sup>, Mingyang Chang <sup>2</sup>, and Rui Deng <sup>1</sup>

<sup>1</sup>Key Laboratory of Oil and Gas Resources and Exploration Technology of Ministry of Education, Yangtze University, Wuhan 430100, Hubei, China

<sup>2</sup>Shandong Coalfield Geological Survey and Research Institute, Jinan 250104, Shandong, China

Correspondence should be addressed to Rui Deng; dengrui@yangtzeu.edu.cn

Received 22 June 2022; Revised 7 August 2022; Accepted 15 October 2022; Published 17 November 2022

Academic Editor: Taoreed Owolabi

Copyright © 2022 Ziyi Na et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problems of low precision and low applicability of the selection method of macroscopic capture cross-section parameters of rock skeleton, formation water, oil and gas, and mud in the volume model of remaining oil saturation logging, this study proposes a method for optimizing the capture section parameters based on the committee machine regression model. To begin, we select well sections from well logging data in a reasonable manner, allocate well sections with different component parameters in the same proportion, and construct sample datasets. The basic experts are then selected as multiple regression models, particle swarm optimization, and robust regression methods to train and learn the input parameters. By combining multiple experts, the regression committee machine improves the overall performance of the intelligent model. Finally, the genetic algorithm is used as a combiner to determine the contribution of each basic expert network in the final output, and the optimized parameters are obtained, which are then fed into the volume model to calculate the remaining oil saturation of the newly developed production wells, guiding the perforation and development work. The model is used to evaluate the remaining oil in the X oilfield, and the calculated water saturation matches the oil test results, proving the model's accuracy and availability. The use of real-world data demonstrates that this method can effectively characterize the four parameter values in the volume model and provide reliable geophysical technical support for the evaluation of remaining oil.

## 1. Introduction

Most of China's oilfields have been developed by water injection and have gradually entered the later stage of development, the evaluation of residual oil saturation is becoming more and more difficult. Residual oil logging evaluation is currently the main means of macroscopic evaluation of residual oil. In the process of oil and gas reservoir development, the geophysical characteristics will change accordingly, and the study of changes in logging response during the development process can guide the residual oil tapping. The methods of identifying and evaluating the water flooded layer can be mainly divided into three types: one is the laboratory geological analysis technology; the second is the reservoir geochemical technology; the third is the geophysical technology [1], among which the

neutron lifetime logging is an important means of identifying and evaluating water flooded formations in geophysical logging. Due to the high sensitivity of neutron lifetime logging technology in water flooded layers with high salinity, this technology is usually used in oil fields to qualitatively identify and quantitatively evaluate water flooded layers, mainly by measuring the macro capture section of the formation and calculating the remaining oil saturation of the reservoir [2].

Based on the petrophysical volume model, the relationship between the capture section and water saturation can be established [3]. The rock volume model refers to dividing the rock volume into different parts according to the composition of the reservoir and its physical properties, and regards the logging results as the sum of the contributions of each part. The formation porosity and shale

content required to obtain water saturation by using the volume model can be obtained from the open hole logging data [4], and the reliability is high. The porosity and shale content of the formation required for water saturation can be obtained with high reliability from open hole logging data; however, the selection of macro capture cross sections, such as rock skeleton, formation water, oil and gas, and mud, is based on empirical values, which do not accurately reflect formation characteristics and affect residual oil saturation calculation accuracy.

Artificial neural networks and geophysical logging have a growing number of interdisciplinary applications as big data and machine learning technology become more advanced. Predecessors have conducted a great deal of exploratory research and made some progress on the topic of calculating reservoir physical parameters. Liu et al. used the hilchie index when calculating the formation mud content [5]. Kang et al. used the principle of normalized nonlinear least square method to optimize the parameters of regional formation components [6]. An et al. established LSTM cyclic neural network method to predict mud content and porosity [7]. Zhai and Dong used the density regression method to calculate the density value of the skeleton [8].

The majority of the currently used machine learning applications for parameter optimization of volume models only take into account one model, and frequently only one of the four parameters is used as the research subject without explaining the entire set. This study applies the committee machine model in light of this issue. By combining multiple experts, the regression committee machine enhances the performance of the intelligent model overall. It selects the multiple regression model, particle swarm optimization algorithm, and robust regression model as the expert network to optimize the four parameters of rock skeleton, formation water, oil and gas, and mud macrocapture cross section in the volume model, and determines the weight of each expert through genetic algorithm. Finally, the optimized four capture cross-section values are output and brought into the water saturation calculation formula to calculate the remaining oil saturation of the newly developed production wells, so as to improve the accuracy of interpretation of water flooded zones in this area.

## 2. The Principle of Parameter Optimization

According to the petrophysical volume model, the reservoir can be regarded as a simple structure composed of mud, framework, and pores. The framework often includes different lithologic components, and the pores contain oil, gas, water, and other fluids [9], as shown in Figure 1.

Taking the capture cross-section value measured by neutron logging instrument as the comprehensive response of macrocapture cross sections of skeleton, mud, water, and hydrocarbon, the calculation formula of water saturation calculated by volume model method is as follows:

$$S_w = \frac{(\Sigma_{log} - \Sigma_{ma}) - \phi(\Sigma_h - \Sigma_{ma})}{\phi(\Sigma_w - \Sigma_h)} - \frac{V_{sh}(\Sigma_{sh} - \Sigma_{ma})}{\phi(\Sigma_w - \Sigma_h)}, \quad (1)$$

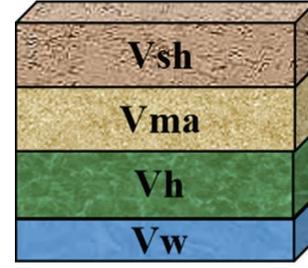


FIGURE 1: Schematic diagram of the volume model.

where  $\Sigma_{log}$  is the macroscopic capture section of the formation,  $\Sigma_w$  is the macroscopic capture section of water,  $\Sigma_{ma}$  is the macroscopic capture section of all dry solids (framework, silt, and dry clay colloid),  $\Sigma_{sh}$  is the macroscopic capture section of hydrocarbons,  $\Sigma_h$  is the macroscopic capture section of the formation mud,  $V_{sh}$  is the shale content of the formation,  $\phi$  is the total porosity of the formation (including shale irreducible water), and  $S_w$  is the water saturation.

Apart from the precise macrocapture cross section, the formula requires the determination of six parameters, including  $V_{sh}$ ,  $\Sigma_{ma}$ ,  $\Sigma_{sh}$ ,  $\Sigma_h$ ,  $\Sigma_w$ , and  $\phi$ . Formation porosity and shale content can be obtained from open hole well logging data with high reliability. In previous studies, the combination of empirical values and theoretical values is usually used to determine the four parameters:  $\Sigma_{ma}$ ,  $\Sigma_{sh}$ ,  $\Sigma_h$ , and  $\Sigma_w$ , but the thermal neutron capture of each component in the formation is not constant [10], and it varies depending on salinity and hydrogen level in the formation. The parameters chosen based on empirical values are entered into the model, and the resulting water saturation does not correctly reflect the formation's genuine state [11]. In this instance, the parameters of each component skeleton can be inversed according to the actual measured values utilizing the parameter optimization approach, allowing the selected parameters to reflect the formation's actual conditions.

The optimization method of the capture cross-section parameters of rock components is to take the parameters of  $\Sigma_{ma}$ ,  $\Sigma_{sh}$ ,  $\Sigma_h$ , and  $\Sigma_w$  as unknown quantities, and use logging data to obtain porosity, shale content, and the total capture cross-section values in the formation as known quantities to construct a linear system of equations::

$$\begin{aligned} \Sigma_z &= V_{sh} \Sigma_{sh} + V_{ma} \Sigma_{ma} + V_w \Sigma_w + V_o \Sigma_o, \\ s \cdot t \cdot V_{sh} + V_{ma} + V_w + V_o &= 1, \end{aligned} \quad (2)$$

where  $\Sigma_z$  is the total trapping section of the formation obtained by measurement,  $\Sigma_{sh}$ ,  $\Sigma_{ma}$ ,  $\Sigma_w$ , and  $\Sigma_o$  are the parameters to be determined, and the constraint condition is that the sum of the volume content of each component is 1.

## 3. Committee Machine Model

Nilsson proposed the concept of committee machine in 1965, which is actually a two-layer network, but only needs to adjust the weights of one-layer network during training. The committee machine is usually divided into a neural network

ensemble, which is made up of several expert machine with independent prediction or classification functions. The local results obtained by each expert machine are combined and decision-making is made, and then the global decision-making model of the entire committee machine system is obtained through some synthetic mechanism.

Different networks in the committee machine are different experts. The committee machine requires the experts to have necessary differences. The performance of the committee machine can be improved by integrating different experts through the committee machine. Expert differences can be obtained by varying the initial training settings and training sets of experts, or by modifying the network differences [12].

Haykin divides committee machine into two categories [13], one is static structure and the other is dynamic structure. The integrated decision-making method of committee machine output to each expert does not involve the committee machine's input in the static structure; instead, the committee machine's input is used to determine how to combine the partial solutions of each expert's output. The committee machine model presented in this research is classified as a static structure, and the model flow chart is shown in Figure 2. The multiple linear regression model, the particle swarm optimization algorithm, and the robust regression model are the three different networks selected in the field of experts. The three methods of network training are different, which meet the different conditions of the experts, and finally output the results through the combiner.

### 3.1. Expert Selection

**3.1.1. Linear Regression Model.** Linear regression is one of the most well-known modeling techniques and one of the first-choice techniques people use when learning predictive models. In this technique, the dependent variable is continuous, the independent variable can be either continuous or discrete [14], and the regression line is linear in nature.

The condition that the multiple linear regression model needs to meet is that there is a linear relationship between the explanatory variable  $Y$  and multiple explanatory variables  $x_1, x_2, \dots, x_k$  [15], which is a multiple linear function of the explanatory variables, namely,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \mu, \quad (3)$$

where  $\beta_0, \beta_1 \dots \beta_k$  are  $k+1$  unknown parameters,  $\beta_0$  is known as regression constants, in which  $\beta_0, \beta_1 \dots \beta_k$  are regression coefficients,  $y$  is called as explained variables, and  $x_1, x_2, \dots, x_k$  are  $k$  explanatory variables, in which  $\mu$  known as random errors.

Corresponding to the model in parametric inversion,

$$\sum_z = V_{sh} \sum_{sh} + V_{ma} \sum_{ma} + V_w \sum_w + V_h \sum_h. \quad (4)$$

In this case,  $k = 4$ , and the constant term  $\beta_0$  is 0,  $\beta_1$  for  $V_{sh}$ ,  $\beta_2$  for  $V_{ma}$ ,  $\beta_3$  for  $V_w$ ,  $\beta_4$  for  $V_h$ ,  $X_1$  for  $\sum_{sh}$ ,  $X_2$  for  $\sum_{ma}$ ,  $X_3$  for  $\sum_w$ ,  $X_4$  for  $\sum_h$ , and  $Y$  for  $\sum_z$ , to satisfy the conditions for a multiple regression model.

The use of multiple linear regression models requires a goodness of fit test [16]; that is, for formula (3), the regression equation is obtained as follows:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k. \quad (5)$$

For the evaluation of the fitting effect of the regression model, it is usually selected  $\bar{R}^2$  as the test index [17], and its value is between 0 and 1. The larger the value, the better the fitting effect of the equation. The calculation formula is as follows:

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{SSE/n - p - 1}{SST/n - 1} \\ &= 1 - \frac{\sum (Y_i - \hat{Y})^2 / n - p - 1}{\sum (Y_i - \bar{Y})^2 / n - 1}, \end{aligned} \quad (6)$$

where  $n - p - 1$  and  $N - 1$  are the degrees of freedom of SSE and SST.

**3.1.2. Particle Swarm Optimization Algorithm Model.** The particle swarm optimization algorithm model is derived from simulations of bird migration and aggregation in the foraging process [18], and the birds achieve the optimal goal through a cooperative group, by following the optimal value of the current global optimization search, with fewer parameters and faster convergence [19].

The mathematical description of the particle swarm optimization algorithm model is that, for a population, given an  $N$ -dimensional search space, a total number of  $M$  particles are generated in the process of its movement, and the population is set as  $X = (x_1, x_2, \dots, x_m)$ , the corresponding position vector of each particle is  $\text{part}x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ , velocity vector  $\text{part}v_k = (x_{i1}, x_{i2}, \dots, x_{im})$ ,  $i = 1, 2, \dots, M$ , position vector represents the potential solution of the objective function, and the velocity vector can control the size of the step size when searching for the solution. The optimal position of each particle in the pursuit of its own history is set as  $\text{poi} = (p_{k1}, p_{k2}, \dots, p_{km})$ ; the optimal solution found by the pursuit group is set as  $\text{goi} = (g_1, g_1, \dots, g_n)$ . After  $n+1$  iterations, the change of the  $K$ th dimension velocity and position can be expressed by the following formula:

$$\begin{aligned} vi_k^{i+1} &= W \times V_{ik}^i + c_1 \times \text{rand1} \times (pb_{ik}^i - P_{ik}^i) \\ &\quad + c_2 \times \text{rand2} \times (gb_k^i - P_{ik}^i), \\ P_{ik}^{i+1} &= P_{ik}^i + V_{ik}^{i+1}, \end{aligned} \quad (7)$$

where  $i$  represents the  $i$ th particle,  $k$  represents the dimension of the current solution,  $t$  is the number of iterations,  $c_1$  and  $c_2$  are the learning factors,  $\text{rand1}$  and  $\text{rand2}$  represent random numbers, and  $W$  represents the weight.

Given different learning factors, the particles can have different learning abilities, and the learning factors can follow a certain change rule, which plays an important role in regulating the learning method of the particles [20]. For example, making the parameters decrease linearly from the

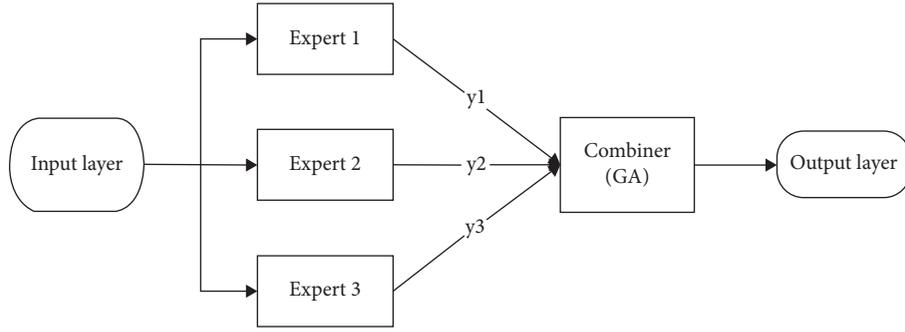


FIGURE 2: Schematic diagram of the committee machine model.

maximum value indicates that the particles are less dependent on the population in the later stage of the iteration. By making the two learning factors change asynchronously, particles can take different optimization forms in the early and late optimization process [21].

The weight plays an important role in particle solution finding [22]. By changing the size of the weight, you can decide whether to let it solve in the original direction. The larger the weight value, the faster the particle moves in the original direction, and the smaller the correction amount of the speed, so that it has a smaller convergence speed. When the weight value decreases, it means to increase the amount of correction to the particle direction, so that it can find a solution in the new direction.

**3.1.3. Robust Regression.** At present, most of the regression models for regression analysis of logging data are implemented by the least squares method. However, during the logging process, due to the influence of the logging instrument and accuracy, there will be abnormal points in the logging data, and these abnormal points will affect the regression model. There is a great influence, and the regression model obtained using the least squares method is quite different from the actual situation.

Robust regression is a method used for regression when the least squares method has abnormal points in the data samples [23], which can detect abnormal points or sample values that have the greatest impact on the model. The most commonly used in robust regression is the maximum likelihood  $M$  estimation [24], and the general model of  $M$  estimation is as follows:

$$Y_i = \sum_{j=1}^p x_{ij}\beta_j + e_i, \quad i = 1, \dots, n, \quad (8)$$

where  $\beta_1, \dots, \beta_p$  is the regression coefficient,  $e_1, \dots, e_i$  is independent and identically distributed, and the variance is 0. The purpose of the least square method is to obtain the value of the regression coefficient that minimizes the value of the objective function. The idea of using the  $M$  estimation of robust regression is to iteratively weight the coefficients of the least squares method [25], and then combine the value of the regression residuals to determine the weight of each point, so as to make the function robust. The optimized objective function is as follows:

$$\sum_{i=1}^n w_i \left( Y_i - \sum_{j=1}^p x_{ij}\beta_j \right)^2 = \min. \quad (9)$$

$M$  estimation reduces the effect of outliers by controlling the weight of outliers. By giving smaller weights to points with large residual values and larger weights to points with small residuals, the purpose of screening data is achieved and the minimum value is established. The optimal solution is obtained by quadratic estimation, and the weight coefficients are improved by iterative iterations. There are many ways to construct weights. In this experiment, the Biweight method [26] provided by DPS is adopted:

$$w_i = \begin{cases} 0, & \left| \frac{u_i}{c} \right| > 1, \\ \left( 1 - \left| \frac{u_i}{c} \right|^2 \right)^2, & \left| \frac{u_i}{c} \right| \leq 1, \end{cases} \quad (10)$$

where  $u_i = e_i/s$  is the residual index and  $s$  is the residual scale.

**3.2. Combiners.** The simple average method or the weighted average method is often chosen as the combination strategy in the regression committee model. Based on the weighted average method, the final output of the committee machine is obtained by weighted sum of the output of each expert. The weight can be selected by experience method, artificial method, or by optimization algorithm. In this paper, genetic algorithm is used as a weight calculator to search the global optimal solution randomly.

Under the condition of producing a certain number of individuals first [27], the genetic algorithm produces more individuals through breeding, crossing and variation, and screening these individuals according to certain conditions, and in the screening process, keep the number of individuals at a constant level and repeat the selection process to find the optimal solution [28]. The specific operation steps encode the actual problem as a chromosome, randomly generate a certain number of individuals, that is, a population, calculate the fitness value of each individual, and judge whether the initial solution is the optimal solution through a given termination condition, and then choose whether to output. If the result is not its optimal solution, a new population is

generated through the genetic operator, and the above operations are repeated. Real-coded genetic algorithms (GAs) are used by Roy et al. to train ANNs, and they produce effective results. [29].

The optimization objective function designed by genetic algorithm is as follows:

$$E = \sum_{j=1}^N -\ln \sum_{i=1}^3 w_i \exp\left(-\frac{1}{2}\|Y_{\text{hat}}^j - Y_0^j\|^2\right). \quad (11)$$

The restrictions are as follows:

$$\sum_{i=1}^3 w_i = 1, \quad 0 \leq w_i \leq 1, \quad (12)$$

where  $w_1$ ,  $w_2$ , and  $w_3$  are the weights of multiple linear regression, robust regression, and particle swarm optimization, respectively,  $Y_{\text{hat}}$  is the prediction value,  $Y_0$  is the target value, and  $N$  is the quantity of training data.

#### 4. Interpretation of Residual Oil Saturation of the Committee Machine Model

X oilfield has gradually entered the aquatic production period of medium and high water content. The mineralization of distinct formations in the vertical direction is complex and variable due to the uneven advancement and alternating mixed injection of clean water and sewage induced by plane water injection. The saturation logging interpretation settings are chosen using a traditional statistical method that does not adjust to the present scenario. Early open hole interpretation and logging can be used to classify wells with comparable lithology and water injection distribution as one type of interpretation parameter, avoiding unpredictability in interpretation. In study area A of X oilfield, we select the wells with the same lithology and formation, as well as the horizon of interpretation results and perforation verification. Finally, 998 logging data groups from 30 horizons were obtained. The committee machine model's training target horizon types include oil layer, oil-water layer, water layer, and dry layer. The interpretation parameters with popularization and application value in the block are obtained through the unified selection of the component parameters of different reservoir types. The committee machine is used to select the interpretation parameters of block A of the X oilfield, output the parameter values with high accuracy, and then feed them into the volume model to calculate the remaining oil saturation of the newly developed production wells, thereby guiding perforation and development work.

**4.1. Data Processing.** First, we perform correlation analysis on  $V_{sh}$ ,  $V_{ma}$ ,  $V_w$ , and  $V_h$  to avoid the problem of multicollinearity and calculate the correlation coefficient between the four variables. The correlation coefficient is a statistical indicator of the closeness of the relationship between the response variables. The value range of the correlation coefficient is between  $-1$  and  $1$ .  $1$  means that the two variables

are perfectly linearly correlated,  $-1$  means that the two variables are completely negatively correlated, and  $0$  means that the two variables are not correlated. The closer the data is to  $0$ , the weaker the correlation is.

The calculation results are shown in Figure 3. The absolute value of the correlation coefficient between the four variables does not exceed  $0.7$ . It can be considered that the correlation between the variables is not strong and multicollinearity processing is not required.

**4.2. Parameter Optimization.** Based on the above regression, the committee machine built the parameter optimization model of the rock component capture section in well XII in the X oilfield, input the test set's logging data into the committee machine model, first output three groups of parameter values through the calculation of three experts, and then optimize the combination weight of the three groups of parameter values via the genetic algorithm combiner to obtain the committee machine out. Finally, the results are entered into the verification set to validate the model's effect.

The performance evaluation indicators of the committee's machine model are shown in Table 1. The average absolute error (MAE) and the adjusted goodness of fit ( $\bar{R}^2$ ) are selected as the two indicators of model performance evaluation. The MAE is a measure of the error between pairs of observations expressing the same phenomenon. Outliers have little influence when using the MAE loss function, and the fitting straight line can better represent the distribution of normal data. The adjusted goodness of fit is used to assess the overall fitness of the regression equation and to express the overall relationship between the dependent and independent variables. In the training set, the mean absolute error (MAE) between the formation capture cross-section value calculated by the parameters output by the multiple linear regression expert and the actual formation capture cross-section value is  $1.0171$ , the error is the largest, the MAE of the committee machine is  $1.0100$ , and the error is the smallest. The MAE of the robust regression expert and the particle swarm optimization expert are  $1.0118$  and  $1.0161$ , respectively. We verify that the MAE value of the centralized committee machine is  $1.0529$ , with the minimum error. The adjusted goodness of fit is the highest value of the committee machine in both the training and validation sets, indicating that it performs better than a single expert network.

It can be seen from Figure 4 that the data point distribution of the committee's machine model and the actual data is the most uniform and the convergence degree is the best with respect to the 45 degree line distribution.

Using the committee machine model constructed above, the parameters of each component capture section of well X in oilfield are optimized. The parameter values of  $\sum_{sh}$ ,  $\sum_{ma}$ ,  $\sum_w$ , and  $\sum_o$  obtained by the committee machine are  $33.1747c.u.$ ,  $11.4581c.u.$ ,  $50.3596c.u.$ , and  $10.3007c.u.$ , respectively. As shown in Table 2, the capture cross-section values calculated by the committee machine model itself and the three expert systems included in it are within the theoretical range.

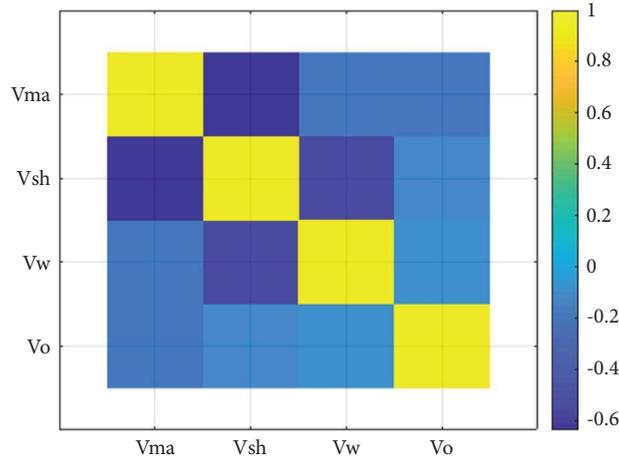


FIGURE 3: Correlation coefficient analysis diagram.

TABLE 1: Committee machine model performance evaluation index.

Training set	MAE	$\bar{R}^2$	Validation set	MAE	$\bar{R}^2$
Multiple regression	1.0171	0.6820	Multiple regression	1.0598	0.6328
Robust regression	1.0118	0.6632	Robust regression	1.0540	0.6262
Particle swarm optimization	1.0161	0.6123	Particle swarm optimization	1.0555	0.5119
Committee machine	1.0100	0.6825	Committee machine	1.0529	0.6377

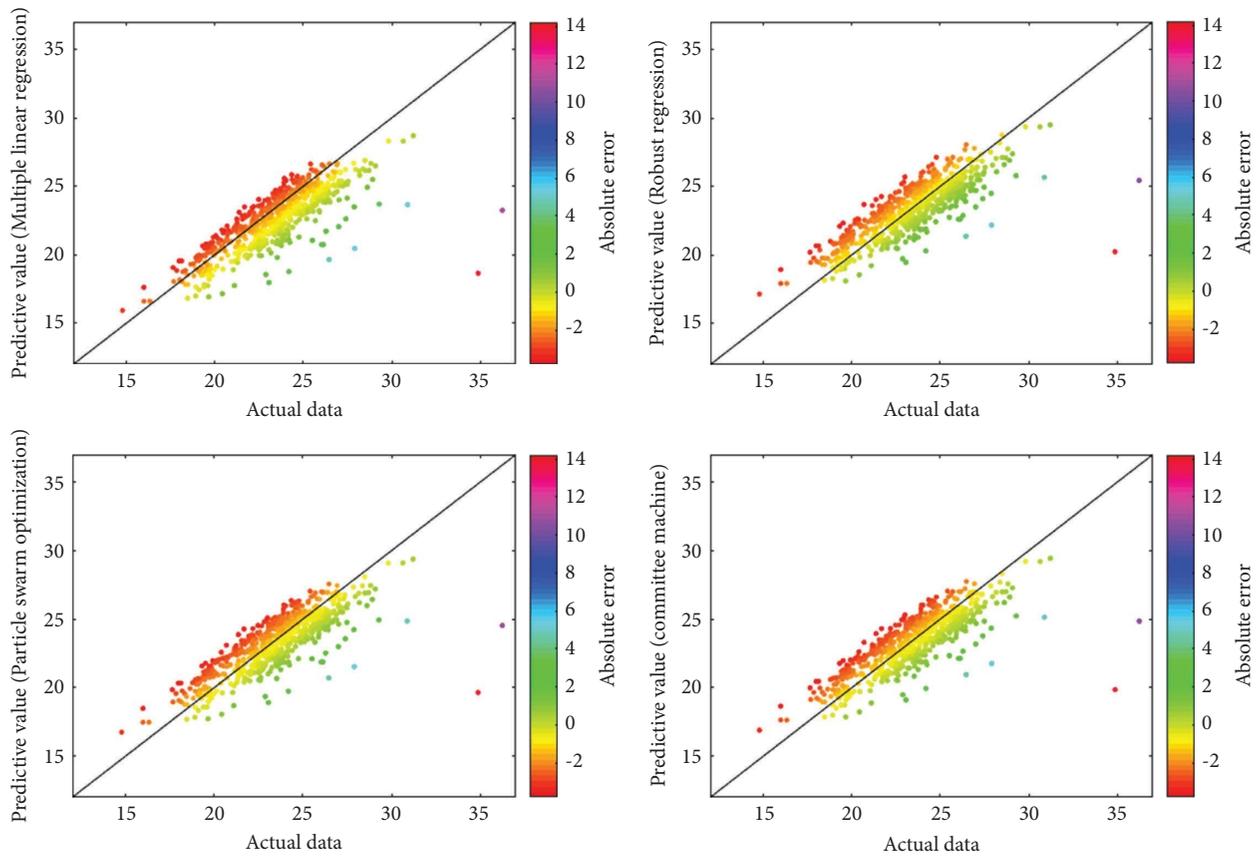


FIGURE 4: Correlation coefficient analysis diagram.

TABLE 2: Committee machine model output parameter results.

	$\sum_{sh}(c.u.)$	$\sum_{ma}(c.u.)$	$\sum_w(c.u.)$	$\sum_o(c.u.)$
Particle swarm optimization	10.9326	30.8733	52.632	8.9641
Multiple regression	11.4929	34.1097	48.8058	8.9307
Robust regression	11.4262	32.6840	51.0242	10.9495
Committee machine	11.4581	33.1747	50.3596	10.3007

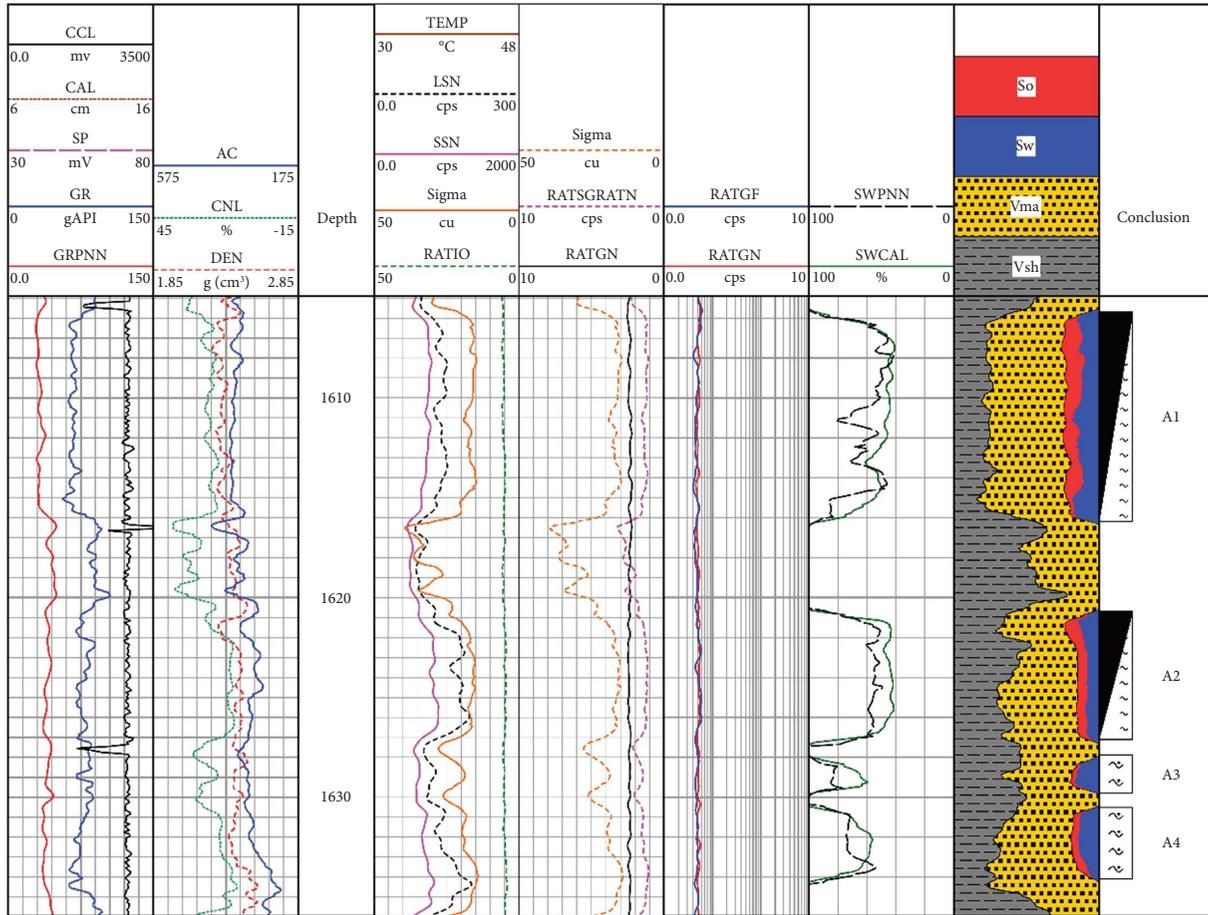


FIGURE 5: Interpretation results of well XII.

4.3. *Residual Oil Saturation Interpretation.* The four capture section parameters output by the committee machine are applied to equation (1) to obtain the water saturation value after parameter optimization, and the following interpretation results of well XII can be obtained. In Figure 5, SWPNN is the water saturation curve calculated by using the empirical value of the previous application area, and SWCAL is the water saturation curve calculated by using the interpretation parameters output by the above committee machine. It is obvious that the water saturation curve calculated after the optimized capture section parameters is obviously smooth and consistent with the shape corresponding to other curves. The correlation shows that the accuracy of this interpretation has been improved. It can be seen from the interpretation results that the remaining oil saturation of layers A1 and A2 is about 60%, which has development and production potential. The remaining saturation of layers A3 and A4 is lower than 40%, and the

porosity is smaller than that of layers A1 and A2. The degree of water flooded layer is high. It is basically an oil-bearing water layer with low exploitation potential.

After the parameter optimization interpretation, the study area conducted perforation analysis on the potential A1 and A2 layers of the well according to the interpretation results and verified the interpretation profile of the oil production rate of the interpreted layer. The interpretation conclusions are shown in Table 3.

The results show that the accuracy of the committee machine output parameters in the calculation of residual oil saturation has improved, and in terms of water production rate, the value calculated after parameter optimization is significantly closer to the value of water production rate, indicating that the regression method of the capture section parameter value in the block through the committee machine model is reliable and can be popularized and applied.

TABLE 3: Calculation of oil production rate.

Horizon	SWPNN (%)	SWCAL (%)	Perforation water production rate (%)	Interpretation conclusion
A1	61.6	42.8	48.5	Low water flooded layer
A2	58.7	41.2	46.9	Low water flooded layer

## 5. Conclusion

Aiming at the problem that the capture cross-section value of each component of the formation cannot be accurately determined in the process of obtaining water saturation from the volume model, a method for parameter inversion based on the committee machine method is proposed. A static committee machine is constructed by combining multiple experts using genetic algorithm. This method is used to explain and analyze the water flooding situation of some wells in X oilfield and to verify the results of actual measures. It is confirmed that the method has high accuracy in practical application.

- (1) The three experts selected by the committee machine are different from each other. The combination of multiple experts can avoid the problem of unreliable prediction results of a single algorithm. The goodness of fit of the results output by the committee machine reaches 0.6825, which is close to the actual situation of the formation. Using a genetic algorithm as a combiner can better express the differences between experts than a simple average method, resulting in a better final output effect.
- (2) The macrocapture cross section after parameter optimization through the committee machine model is more scientific and accurate than that selected according to the theoretical empirical value. The water saturation obtained in the volume model is closer to the perforation water production rate, which provides a new idea for the quantitative interpretation of residual oil saturation logging and has guiding significance for the selection of the next key production target layer in the oilfield.
- (3) In residual oil saturation logging, the optimization of macrocapture cross-section parameters of rock skeleton, formation water, oil and gas, and mud in the volume model is essentially a regression problem. The model's application effect is directly related to data quality. The fitting effect of the model output can be improved by increasing the number of actual logging data and enriching the sample types. Furthermore, how to choose the basic expert network with the best performance to combine and improve the model's generalization ability requires more research.

## Data Availability

The data that support the findings of this study can be obtained from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Ziyi Na conceptualized the study, curated the data, developed the methodology, investigated the study, validated the study, and wrote the original draft. Lixia Dang and Rui Deng supervised the study, carried out formal analysis and funding acquisition, collected the resources, and wrote, reviewed, and edited. Mingyang Chang supervised, carried out formal analysis, and collected the resources.

## Acknowledgments

The research was funded by the Major National Science and Technology Projects of China "Multidimensional and high precision imaging logging series" (no. 2017ZX05019001) and the Key Project of Science and Technology Research Program of Hubei Provincial Department of Education (Grant no. D20191302).

## References

- [1] H. Shunyi, X. Nan, P. Hongboand, and Z. Chengxue, "Identification and evaluation of water flooded zones with logging data and its application," *Offshore Oil*, vol. 30, no. 2, pp. 91–95, 2010.
- [2] D. Parasnis, *Mining Geophysics*, Elsevier, Amsterdam, Netherlands, 2013.
- [3] F. Zhang, J. P. Xu, L. M. Hu, C. H. Xiu, and J. M. Sun, "Monte Carlo simulation for the pulsed neutron-neutron logging method," *Chinese Journal of Geophysics*, vol. 50, no. 6, pp. 1654–1661, 2007.
- [4] H. Yang, S. Li, and X. Liu, "Characteristics and resource prospects of tight oil in Ordos Basin, China," *Petroleum Research*, vol. 1, no. 1, pp. 27–38, 2016.
- [5] W. Liu, "XSong et al., "Calculation of formation mud content using natural gamma logging data," *Uranium Geology*, vol. 33, no. 1, pp. 45–48, 2017.
- [6] X. Kang, D. Wu, X. Zhao, Z. He, and B. Lu, "Macroscopic capture cross section of thermal neutrons in each component of strata  $\Sigma$  Optimization determination of values," *Logging technology*, vol. 42, no. 06, pp. 658–661, 2018.
- [7] P. An, D. Cao, B. Zhao, X. Yang, and M. Zhang, "Research on prediction method of reservoir physical parameters based on LSTM cyclic neural network," *Progress in Geophysics*, vol. 34, no. 05, pp. 1849–1858, 2019.
- [8] L. Zhai and W. Dong, "Determination of rock skeleton density based on mineral composition method and density regression method and its application in well X Area," *China Petroleum and chemical industry standard and quality*, vol. 42, no. 07, pp. 103–105, 2022.
- [9] F. Zhang, D. Jiang, R. Du, and J. Liu, "Research on TOC inversion method based on shale linear petrophysical model," *Progress in Geophysics*, vol. 36, no. 167, pp. 1154–1165, 2021.
- [10] R. Deng, M. Li, L. Song, R. Yang, and J. Xie, "Research on calculation method of steam Absorption in steam injection thermal recovery technology," *Fresenius Environmental Bulletin*, vol. 30, no. 05, pp. 5362–5369, 2021.

- [11] X. Zhu, S. Shan, Z. Yin, and X. Kong, "Calculation method of remaining oil saturation in water-flooded layer mixed with PNN logging and decontamination," *Geophysical and Geochemical Exploration*, vol. 45, no. 03, pp. 679–685, 2021.
- [12] J. Jin, D. Liang, Y. Bao, and G. Huang, "A committee machine with two layer expert nets," *Kybernetes*, vol. 39, no. 6, pp. 961–972, 2010.
- [13] S. Haykin and, "A comprehensive foundation," *Neural Networks*, vol. 200441 pages, 2004.
- [14] N. R. D. H. Smith, *Applied Regression Analysis*, John Wiley & Sons, Hoboken New Jersey, 1998.
- [15] J. F. Leng, X. Gao and, and J. P. Zhu, "Application of multivariate linear regression statistical prediction model," *Statistics & Decisions*, vol. 7, pp. 82–85, 2016.
- [16] N. C. Giri, *Multivariate Statistical Analysis: Revised and Expanded*, CRC Press, Boca Raton Florida, 2003.
- [17] D. George and P. Mallery, *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*, Routledge, Milton Park Abingdon-on-Thames Oxfordshire England UK, 2019.
- [18] P. Roy, G. S. Mahapatra, and K. N. Dey, "An efficient particle swarm optimization-based neural network approach for software reliability assessment," *International Journal of Reliability, Quality and Safety Engineering*, vol. 24, no. 04, Article ID 1750019, 2017.
- [19] M. Shokooch-Saremi and R. Magnusson, "Particle swarm optimization and its application to the design of diffraction grating filters," *Optics Letters*, vol. 32, no. 8, pp. 894–896, 2007.
- [20] J. Liu, Q. Shi, R. Han, and J. Yang, "A hybrid GA-PSO-CNN model for ultra-short-term wind power forecasting," *Energies*, vol. 14, no. 20, p. 6500, 2021.
- [21] P. Roy, G. S. Mahapatra, and K. N. Dey, "Forecasting of software reliability using neighborhood fuzzy particle swarm optimization based novel neural network," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1365–1383, 2019.
- [22] F. Wang, J. Li, Z. Li et al., "Design of model predictive control weighting factors for PMSM using Gaussian distribution-based particle swarm optimization," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 11, pp. 10935–10946, 2022.
- [23] R. Andersen, *Modern Methods for Robust Regression*, Sage, California, 2008.
- [24] F. Fu, Z. Zeng, and X. Liu, "Difference-based M-estimator of generalized semiparametric model with NSD errors," *Journal Inequalities Applications*, vol. 2019, no. 1, pp. 61–15, 2019.
- [25] H. Ertaş and S. K. H. Güler, "Robust Liu-type estimator for regression based on M-estimator," *Communications in Statistics - Simulation and Computation*, vol. 46, no. 5, pp. 3907–3932, 2017.
- [26] Q. Y. Tang and C. X. Zhang, "Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research," *Insect Science*, vol. 20, no. 2, pp. 254–260, 2013.
- [27] G. S. Mahapatra, B. Maneckshaw, and K. Barker, "Multi-objective reliability redundancy allocation using MOPSO under hesitant fuzziness," *Expert Systems with Applications*, vol. 198, Article ID 116696, 2022.
- [28] F. Xu, L. Y. Mo, H. Chen, and J. M. Zhu, "Genetic algorithm to optimize the design of high temperature protective clothing based on BP neural network," *Frontiers in Physics*, vol. 9, Article ID 600564, 2021.
- [29] P. Roy, G. S. Mahapatra, P. Rani, S. Pandey, and K. Dey, "Robust feedforward and recurrent neural network based dynamic weighted combination models for software reliability prediction," *Applied Soft Computing*, vol. 22, pp. 629–637, 2014.