IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 June 2024, accepted 13 July 2024, date of publication 19 July 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3431015

RESEARCH ARTICLE

Research on Correction of Temperature and Mineralization and Prediction of Water Holdup Based on Machine Learning

XIAOMEI DAI[1], YONG WEI[1], RUYI GAN[1], BAOJUN WEI[2], YALIN XIANG[3], AND PING LIU[2]

[1]School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou 434023, China
[2]Changqing Branch of China Petroleum Group Logging Company Ltd., Xi'an 710077, China
[3]Southwest Branch of China Petroleum Group Logging Company Ltd., Chongqing 401120, China

Corresponding author: Yong Wei (weiyong@yangtzeu.edu.cn)

This work was supported in part by the Scientific Research and Technology Development Project of China National Petroleum Corporation (Research on Key Technologies of Logging Acquisition, Processing, and Interpretation) under Grant 2021 DJ4004, and in part by Hubei Provincial Natural Science Foundation of China (Research on a New Type of Detector to Improve the Accuracy of Water Holdup Detection) under Grant 2022 CFB187.

ABSTRACT Accurate detection of water holdup in oil-water two-phase flow is crucial for optimizing production and improving crude oil recovery. The transmission lines method is currently one of the few effective methods to measure the water holdup of oil-water two-phase flow. However, variations in temperature and mineralization will alter the dielectric constant and conductivity of the oil-water mixture respectively, posing challenges for precise water holdup measurement. The complex nonlinear relationship between these factors limits the prediction range and accuracy of widely used models, such as the BP neural network and Support Vector Machine (SVM). In order to overcome these issues, this paper establishes a multi-sensor oil-water two-phase flow indoor experiment system and studies the complex relationship between the phase shift of sensor signal and influencing factors. On this basis, this paper proposes a combined water holdup prediction model (BO-XGBoost) of Bayesian optimization (BO) algorithm and extreme gradient boosting (XGBoost). The results demonstrate that the XGBoost model outperforms traditional BP neural network and SVM in predicting water holdup across the full range of 0%-100%. The average absolute error of the BO-XGBoost model is only 1.50%. The above research achieves a full-range, high-precision water holdup prediction, providing a new solution for oilfield development and possessing practical engineering significance.

INDEX TERMS BO-XGBoost, XGBoost, BP, SVM, transmission lines method, water holdup.

I. INTRODUCTION

The water holdup of oil-water two-phase flow is a key parameter that reflects the production status of oil fields. It provides direct basis for evaluating the productivity of oil wells, predicting the development life of oil wells, and optimizing the extraction and transportation process of crude oil [1], [2]. Real-time accurate measurement of the water holdup of each layer of the oil well can provide a reference for effectively plugging the layers with excessive water holdup during injection and production, and avoid the ''empty pumping'' of oil production machines. It not only improves the extraction efficiency of oil wells, but also avoids the waste of human and material resources, which is of great significance to the development of oil fields.

At present, the common methods for measuring water holdup mainly include capacitance method [3], conductance method [4], acoustic wave method [5], radio frequency method [6] and transmission lines method [7]. The capacitance method uses the proportional relationship between the capacitance and the dielectric constant of the oil-water mixture for detection, but in high water holdup, the water

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Abdur Razzaque.

phase connection reduces the resolution. It is generally suitable for the water holdup less than 50%. The conductance method uses the different conductance of the oil-water mixture to detect, which is generally applicable to the water holdup higher than 50%. It loses the resolution that oil forms continuous phase under low water holdup. The acoustic wave method utilizes the lost energy during the transmission process caused by the scattering and reflection of sound waves at the oil-water interface. However, it has poor resolution for dispersed oil slugs in water, and is fit for oil-water two-phase flow with low mixing speed and high water holdup. The radio frequency method relies on the attenuated variation of the amplitude of the radio frequency signal caused by the difference in the oil-water mixture medium. But it is easily affected by the uneven distribution of the oil-water mixture and measures difficultly under low water holdup, there is also the risk of radiation leakage. In comparison, the fundamental principle of the transmission lines method is to use the phase shift of initial and terminal signals of the transmission lines that has a monotonic relationship with the dielectric constant of the fluid detected to measure the water holdup [7], [8]. Both theoretical analysis and experiments have shown that this method is valid at the water holdup from 0% to 100%. It can be well applied to monitoring and evaluating the production wells.

However, the actual application has shown that the dielectric constant of downhole mixed fluid decreases with the rising temperature, and the conductive current between transmission lines increases with the rising mineralization. Accordingly, it is still influenced by the temperature and mineralization of the downhole fluid, thus causing measurement error. Although methods such as increasing the emission frequency of signal and covering the insulation layer have been adopted to reduce the impact of mineralization on measurement, the measurement error caused by the above factors cannot be eliminated completely. Therefore, it is extremely important to choose a method that can correct the affection of temperature and mineralization.

Due to the complexity of downhole situation, there is a non-linear relationship between the influencing factors and the measurement results of the water holdup [9]. Therefore, it is generally difficult to establish a suitable mathematical model for the problem of multiple factors affecting the measurement of downhole water holdup [10], [11], [12]. Domestic and foreign scholars have also done extensive research work. Zhou et al. [13] used a multi-parameter coupled BP neural network model to predict the water holdup and achieved good test results in the range of water holdup from 60% to 100%. However, the types and number of test samples are too few to be general. In view of the problems that BP cannot seek the global optimal solution and has slow convergence speed. Chang and Chen [14] proposed the BP neural network model improved by heuristic improvement and numerical optimization method to predict the water holdup of crude oil, which enhanced the convergence speed and prediction accuracy. But there is still a large prediction error at low water holdup. On this basis, Zhang and Xia [15] established a soft sensor model based on RS-SVM classifier and GA-NN model to predict the water holdup, which was capable of segmented prediction of water holdup. However, its generalization ability is poor owing to the small number of model samples.

In order to overcome the certain limitations of traditional prediction models in prediction range and accuracy, this paper uses an integrated learning algorithm XGBoost to correct the effects of temperature and mineralization. Compared with traditional BP neural network and SVM, the XGBoost model has the advantages of higher prediction accuracy and wider prediction range. The other parts of the paper are organized as follows. Section II theoretically analyzes the influence of electrical parameters of oil-water mixture. Section III introduces the water holdup experimental device based on the transmission lines method and studies the effects of temperature and mineralization on the measurement results. Section IV constructs water holdup prediction models of machine learning algorithms, then compares and analyzes their prediction performance. Finally, the conclusions are given in Section V.

II. EFFECT OF TEMPERATURE AND MINERALIZATION ON THE ELECTRICAL PARAMETERS OF OIL-WATER MIXTURE

A. EFFECT OF TEMPERATURE ON THE DIELECTRIC CONSTANT OF OIL-WATER MIXTURE

In the downhole environment, the dielectric constant of the oil-water mixture varies with the temperature. The relative permittivity of oil is slightly affected by temperature, generally between 2.2 and 2.5, while the relative permittivity of water is greatly affected by the temperature. Table 1 reflects the relationship between them [16].

TABLE 1. Relative permittivity of water at different temperatures.

t(℃)	10	20	30	40	60	80	100
ε_w	83.8	80	76	73	66.8	61.0	55.7

The quadratic relation fitted to the relative permittivity ε_w of water and the temperature t is shown as equation (1).

$$\varepsilon_w = 91.1238 - 0.5077t + 0.0015t^2 \tag{1}$$

When the water holdup of the oil-water mixture varies, its dielectric constant also varies obviously. According to the distribution state of oil and water, its equivalent dielectric constant is:

$$\varepsilon_m = [\sqrt{\varepsilon_o}(1 - Y_w) + (Y_w \sqrt{\varepsilon_w})]^2 \tag{2}$$

In equation (2), ε_m is the equivalent dielectric constant of the oil-water mixture, ε_o is the relative permittivity of crude oil, Y_w is the water holdup of the oil-water mixture, ε_w is the relative permittivity of the water.

By substituting equation (1) into equation (2), the relationship between the equivalent dielectric constant and water

TABLE 2. Conductivity data of oil-water mixed samples at different degrees of mineralization.

Mineralization (mg/L)		0	1000	2000	5000	8000	10000	50000	80000
Conductivity of oil-water mixed samples (ms/cm)	100%	0.19	1.18	2.12	4.66	6.80	8.10	23.00	28.70
	80%	0.14	0.87	1.41	2.85	4.36	5.63	11.71	24.09
	60%	0.07	0.63	1.02	2.62	4.01	4.93	17.05	23.70

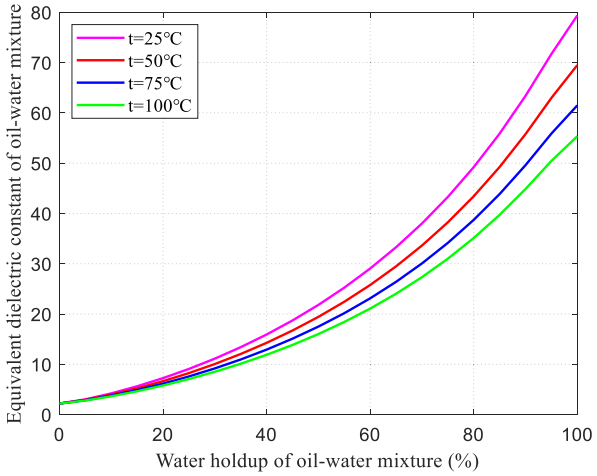

FIGURE 1. Relationship between equivalent dielectric constant and water holdup of oil-water mixture at different temperatures.

holdup of the oil-water mixture at different temperatures can be obtained. The numerical simulation is shown in Fig. 1.

It can be seen from Fig. 1 that the equivalent dielectric constant of the oil-water mixture increases monotonically with the rising water holdup. The lower the temperature, the higher the equivalent dielectric constant.

B. EFFECT OF MINERALIZATION ON THE CONDUCTIVITY OF OIL-WATER MIXTURE

In order to simulate the downhole two-phase flow environment, diesel oil and water are used to prepare the experimental samples with water holdup of 60%, 80% and 100%, respectively. The mixture is stirred fully by using a blender, and a certain mass of NaCl is added to each sample with different mineralization. The conductivity of the mixture is measured by a conductivity meter. The experimental results are shown in Table 2.

According to the experimental data in Table 2, the fitted curves of conductivity and mineralization under different water holdup samples are drawn as shown in Fig. 2. It shows the quadratic fitted function and curve of conductivity and mineralization, indicating that they have a parabolic relationship in the range of 0-8 $\times$ 10^4mg/L mineralization. Fig. 2 (b) is an enlarged view of the green circle in Fig. 2 (a), which shows that they have a good linear relationship in the range of 0-1 $\times$ 10^4mg/L mineralization. Therefore, the conductivity and mineralization of the oil-water mixture show a monotonic increasing relationship. The greater the mineralization, the higher the conductivity.

C. NUMERICAL SIMULATION OF MULTIPLE FACTORS AFFECTING THE MEASUREMENT OF WATER HOLDUP

The above theory analyses the significant effect of temperature and mineralization on the electrical parameters of oil-water mixture. When measuring water holdup using the transmission lines method, changes in the equivalent dielectric constant and conductivity of the oil-water mixture combine to affect the phase shift of the transmission lines signal [16]. In order to further analyze their common influence, the numerical simulation result is shown in Fig. 3.

When the dielectric constant of the oil-water mixture is certain, the higher the conductivity, the greater the phase shift. When the conductivity of the oil-water mixture is fixed, the larger the dielectric constant, the greater the phase shift. As a result, when measuring the water holdup of oil-water mixture, it is definitely affected by temperature and mineralization, then leading to measurement error.

III. EXPERIMENTAL STUDY ON THE EFFECT OF TEMPERATURE AND MINERALIZATION ON THE MEASUREMENT OF WATER HOLDUP

A. EXPERIMENTAL DEVICE

In order to test the effect of temperature and mineralization on the phase shift of the transmission lines signal, a measuring device as shown in Fig. 4 is designed to study the influence of different temperatures and mineralization on the measurement of water holdup. The device consists of a mixing container, a measuring container, sensors, a blender, temperature controller, a water pump, heating rods, etc. The mixing container is filled with mixed liquid of a certain proportion of diesel and water, as well as a certain quality of NaCl. The temperature controller can achieve real-time temperature display and control, and simulate the oil-water two-phase flow environment at various temperatures, mineralization and downhole water holdup.

The experimental process is as follows. First, prepare the oil-water mixture sample according to the water holdup value, pour it into the mixing container, open the blender to stir the mixture fully, and then pump it into the measuring container through a water pump. Last, set the preheating temperature of the temperature controller, wait for the temperature to reach the set value, and record the sensor count value after the mixture is sufficiently stable.

B. EXPERIMENTAL DATA ANALYSIS

In order to test the count values of the two sensors at different samples, the experiment is carried out at the experimental

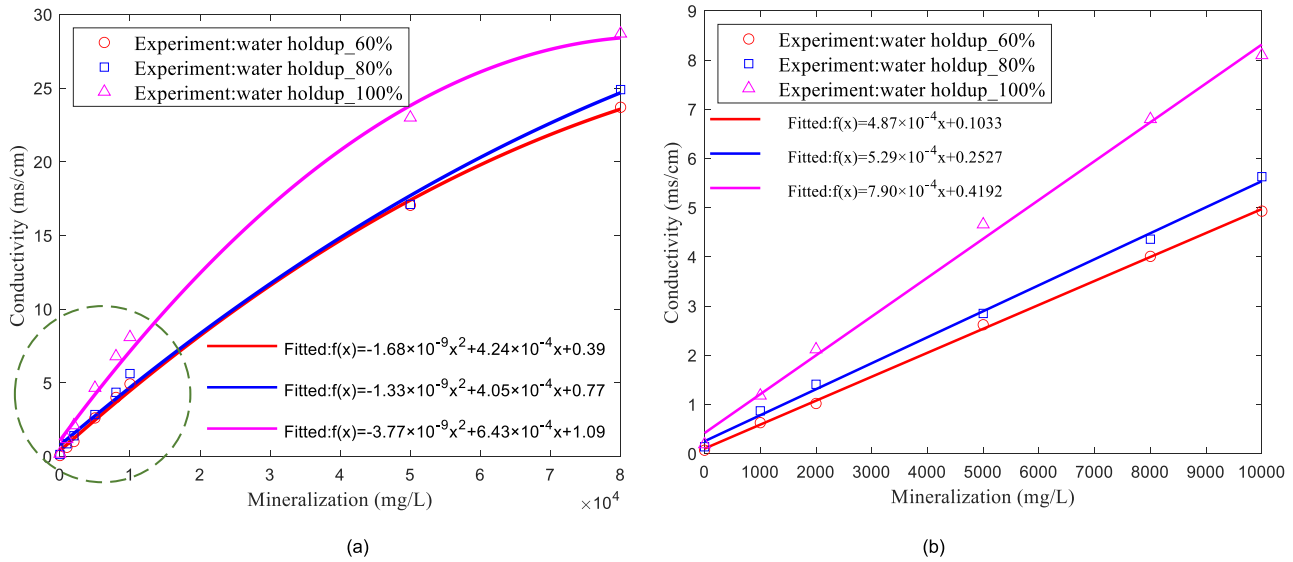

FIGURE 2. Relationship between conductivity and mineralization under different water holdup. (a) 0-8 × 10^4mg/L and (b) 0-1 × 10^4mg/L.

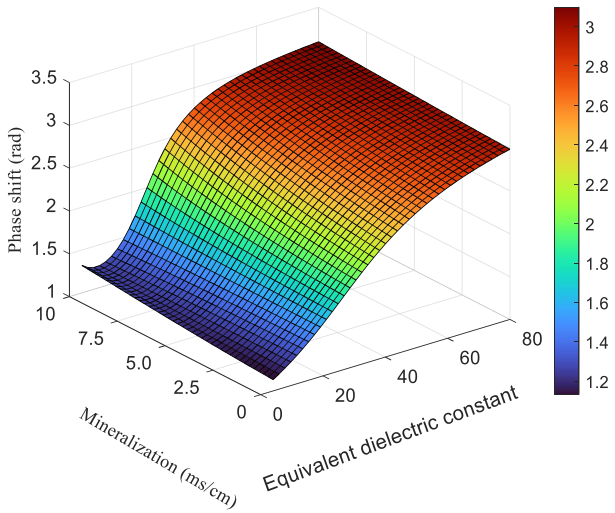

FIGURE 3. Numerical simulation of the effect of equivalent dielectric constant and mineralization on phase shift.

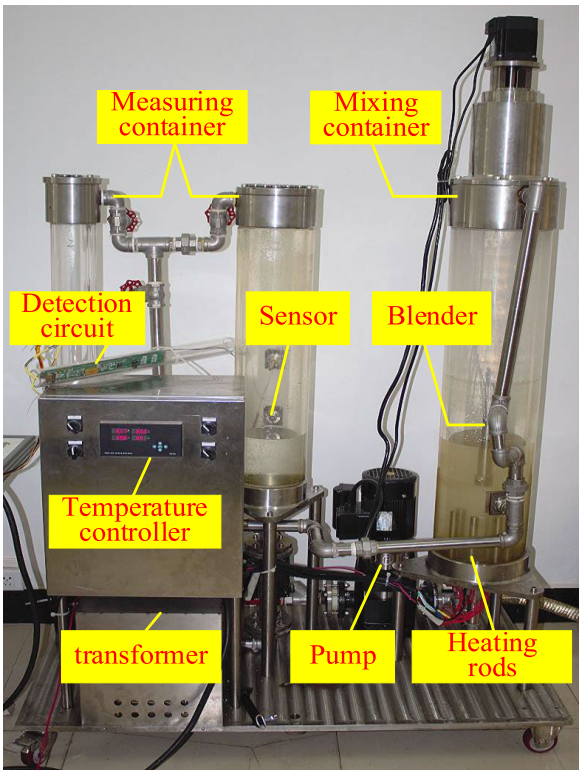

FIGURE 4. Experimental device.

samples with temperature of 25°C, water holdup of 0% to 100%, an interval of 10%, and 0 mg/L mineralization. At the same time, in order to make the count values of the two sensors comparable, the phase shift count value N_x is normalized to obtain the phase shift count percentage (CP) N_{CP} as equation (3).

$$N_{CP} = \frac{N_x - N_o}{N_w - N_o} \times 100\% \tag{3}$$

where N_o is the count value under air, N_w is the count value under full water.

Fig. 5 shows the N_{CP} curves of the two sensors at 0mg/L mineralization. The curve shows that: (1) Within the range of full oil to full water, the N_{CP} of the sensor shows a monotonic increasing relationship with the water holdup. (2) The first and second N_{CP} of the same sensor are basically consistent with good repeatability. (3) Due to the immiscibility of oil-water medium, there is a phenomenon of uneven mixing, which causes the N_{CP} of sensor to fluctuate within a certain range. So the strip curve reasonably describes the fluctuation in the counting. (4) The sensitivity of the two sensors is equivalent at the low, medium, and high water holdup, and their consistency is pretty well.

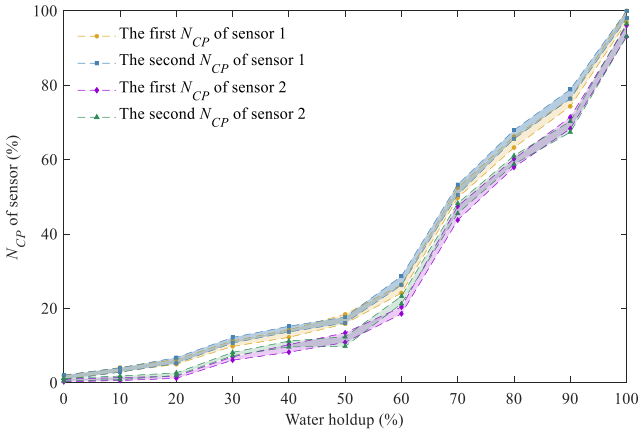

FIGURE 5. N_{CP} curves of two sensors at 0mg/L mineralization.

In order to obtain test samples of various temperatures, mineralization and water holdup, the temperatures of 25°C, 45°C, and 65°C are selected during the experiment. Nine mineralization degrees are selected, including 0mg/L, 1000mg/L, 2000mg/L, 5000mg/L, 8000mg/L, 10000mg/L, 20000mg/L, 50000mg/L, and 80000mg/L. A total of 11 water holdup values are selected, ranging from 0% to 100%, with an interval of 10%. A total of 3 × 9×11=297 groups of experimental samples are produced under the above conditions. Because of the limitation of experimental conditions, this experiment is unable to simulate the high temperature underground except. It basically covers the variable range of mineralization and water holdup in actual well condition (because water exceeding 100°C will boil under normal surface pressure condition and diesel vapor has the risk of explosion). Fig. 6 shows the N_{CP} curves of the sensor at three temperatures.

As can be seen from Fig. 6, (1) Within the range of 0%-40% water holdup, as the mineralization increases, the N_{CP} of the sensor hardly changes. This phenomenon is related to the oil-water mixture presenting a state of ''water-in-oil''. The conductive ions in the bubbles are separated by oil, and the conductive current cannot be formed between the transmission lines, which has little impact on the count value of sensor. (2) Within the range of 40%-90% water holdup, some N_{CP} of sensor increase monotonically as the mineralization increases, while others show a trend of first increasing and then decreasing with the rising mineralization. This is related to the combined effect of conductive current and oil-water slippage. When conductive current dominates, the N_{CP} increases with the rising mineralization, and when slip phenomenon dominates, the N_{CP} decreases with the rising mineralization [17], [18]. (3) The N_{CP} of the sample with 100% water holdup increases monotonically as the mineralization increases. Under the same mineralization condition, the N_{CP} decreases monotonically with rising temperature, which is compatible with the theoretical analysis in Section II.

In order to further illustrate the impact of temperature on the measurement results, the count results at each temperature and water holdup are averaged. For the convenience of comparison, the measurement results of mineralization at four typical working conditions of 0 mg/L, 5000 mg/L, 20000 mg/L, and 80000 mg/L are selected as shown in Fig. 7.

As can be seen from Fig. 7, (1) The N_{CP} is basically not affected by temperature within the range of 0% to 30% water holdup. This is because the proportion of oil in the mixture is large, and the equivalent dielectric constant of the oil-water mixture is not sensitive to temperature. (2) When the water holdup ranges from 40% to 100%, the N_{CP} shows a gradual decline trend with the temperature rising. This is mainly related to the sensitivity of the relative permittivity of oil and water to temperature. Because the relative permittivity of water decreases with the rising temperature, so the equivalent dielectric constant of oil-water mixture is sensitive to temperature.

From the above experimental results, it is clear that the measurement of water holdup by the transmission lines method must be affected by temperature and mineralization of the oil-water mixture. However, we can build the right prediction model to deal with the complex non-linear influence relationship. In the actual downhole environment, the temperature and mineralization of oil-water two-phase flow can be measured on-line. By inputting the temperature, mineralization, and count value of sensor into the prediction model, the impact of multiple factors on measurement of the water holdup can be reduced, and water holdup can be predicted.

IV. THE WATER HOLDUP PREDICTION MODEL BASED ON MACHINE LEARNING

A. ESTABLISHMENT OF PREDICTION MODEL

The input data of the model are temperature, mineralization, and count value of sensor, and the output is the predicted value of water holdup as shown in Fig. 8. The construction process of the prediction model is divided into the selection of data sets, data processing, model construction, and the evaluation and analysis of prediction result [19], [20]. The training sets select the above 297 sets of experimental samples. The test sets select temperatures of 35°C and 55°C, water holdup ranging from 5% to 95%, with an interval of 5%. The mineralization levels are 1500mg/L, 3500mg/L, 6500mg/L, 9000mg/L, 15000mg/L, 25000mg/L, 35000mg/L, 45000mg/L, 55000mg/L, with a total of 180 groups of combined samples. For BP and SVM, data sets need to be normalized to avoid certain input variables that are too large to affect training. However, XGBoost does not need to be normalized because it only considers the distribution of each variable and the conditional probabilities between them.

B. BP

BP is a multi-layer feedforward neural network that learns based on error backpropagation [21], [22], [23]. On the premise of satisfying the model training accuracy, this paper

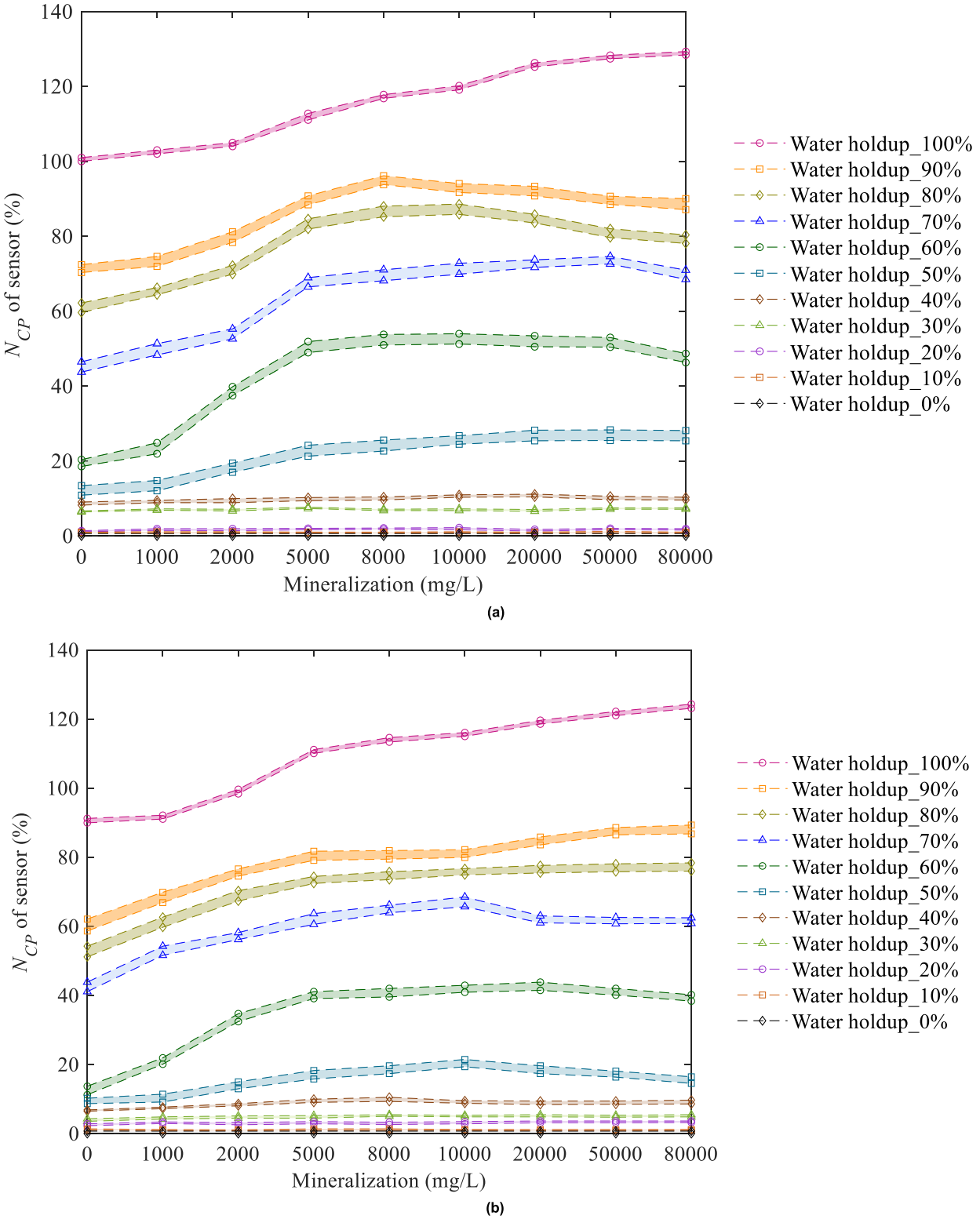

FIGURE 6. N_{CP} of the sensor at three temperatures. (a) 25°, (b) 45°.

adopts a BP structure with two hidden layers. Using the node trial-and-error method, the number of nodes in the two hidden layers is 8 and 4 respectively. The learning rate is set to 0.01, the target error is set to 1e-4, and the iterations is set to 5000. Use three training functions of the MATLAB toolbox: ''Resilient backpropagation algorithm (trainrp)'', ''Bayes rule training algorithm (trainbr)'', and ''Gradient descent with momentum and adaptive learning rate

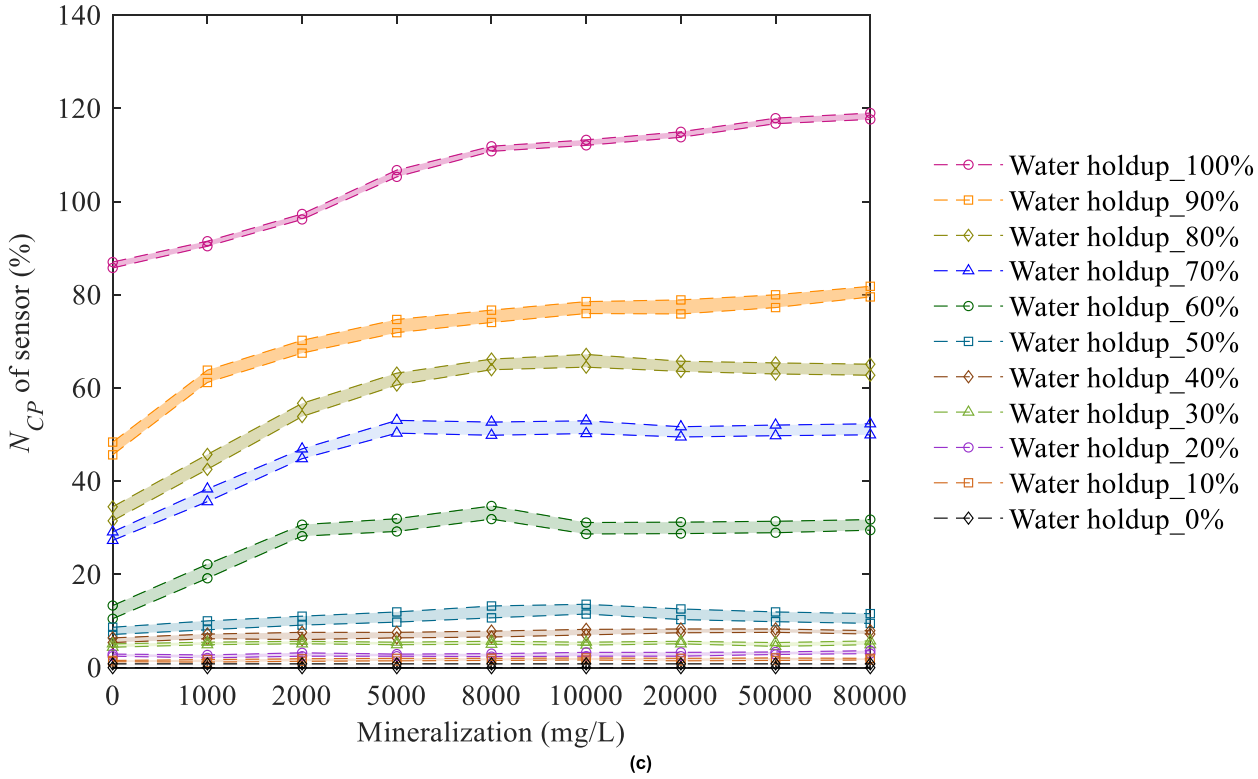

FIGURE 6. *(Continued.)* N_{CP} of the sensor at three temperatures. (c) 65°.

backpropagation (traingdx)'' for training respectively. After experiment, it is found that the ''trainrp'' algorithm can make the model converge with the least iterations and the best training effect. Therefore, ''trainrp'' is chosen as the training algorithm for the BP network.

C. SVM

The principle of SVM is to find a classification decision hyperplane that maximizes the distance between samples of different categories and the hyperplane [24], [25]. The Lagrange multiplier is introduced according to the principle of empirical risk minimization, and the decision function is shown as in equation (4):

$$f(x) = \sum\nolimits_{i,j=1}^{n} \alpha_i y_i K(x_i, x_j) + b \tag{4}$$

where y_i is the actual value of the i-th sample, α_i is Lagrangian operator, $K(x_i,x_j)$ is a kernel function, x_i is the i-th sample, and b is the bias.

The core parameters of SVM are ''the penalty coefficient (C)'' and ''the kernel function coefficient (*gamma*)''. The common method at present is the cross-validation grid search method [26], [27]. First, use the training sets as the data sets for selection of parameters, set the algorithm type to ''support vector regression machine (SVR)'', then set kernel = ''rbf''. The selection range of ''C'' and ''*gamma*'' is 10^{-10} to 10^{10}. After the model is trained, the optimal parameter $C = 90.51$ and $gamma = 0.71$ is obtained.

D. XGBOOST

The principle of the XGBoost algorithm is to learn new weak learners based on decision trees, and then combine them into a strong learner [28], [29].

In the XGBoost model, the objective function is shown as in equation (5):

$$obj^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)}) + \Omega(f_k) \tag{5}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{6}$$

In equation (5), $\hat{y}_i^{(t-1)}$is the predicted value of the i-th sample after the (t-1)-th iteration, and $l(y_i, \hat{y}_i^{(t-1)})$ is the training error, $\Omega(f_k)$ is the regularization penalty term. In equation (6), T is the number of leaf nodes, γ and λ are adjustable penalty term parameters, ω_j is the output value of the j-th leaf node.

E. BO-XGBoost

Bayesian optimization algorithm (BOA) is a black-box optimization algorithm [30], [31]. The BOA framework consists of an agent model of the objective function based on the historical evaluation results of the objective function, and an acquisition function constructed using the posterior information of the agent model [32], [33]. The gaussian process is used as the agent model here, it consists of a mean

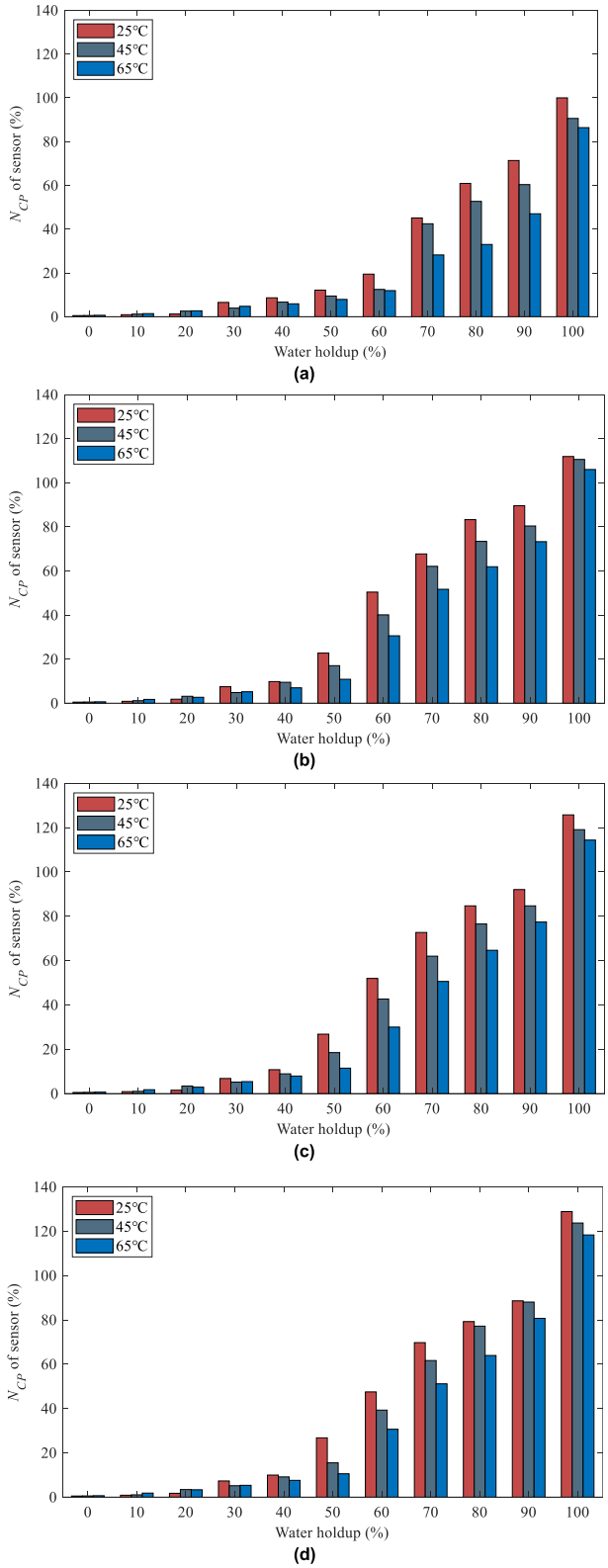

FIGURE 7. Measurement results of water holdup at various temperatures and mineralization. (a) 0mg/L, (b) 5000mg/L, (c) 20000 mg/L, and (d) 80000mg/L.

function and a covariance function:

$$f(x) \sim GP[m(x), k(x, x^t)] \tag{7}$$

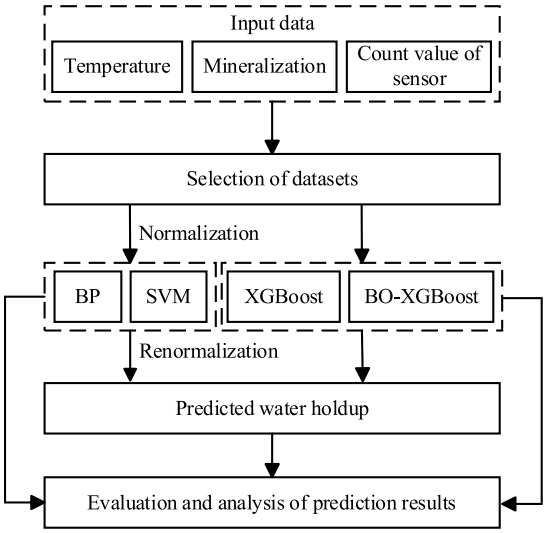

FIGURE 8. Flowchart of water holdup prediction model based on machine learning.

where $m(x) = E[f(x)]$ is the mean function, $k(x, x^t) = E[(f(x)\text{-}m(x))(f(x^t)\text{-}m(x^t))]$ is the covariance function.

The (PI) function is chosen for this model as shown in equation (8), and the main idea is to allow the posterior distribution sampling points to maximize the probability of boosting the value.

$$f(x; D_{1:t}) = p[f(x) \geq f(x^+) + \xi] = \Phi(\frac{\mu(x) - f(x^+) - \xi}{\sigma(x)}) \tag{8}$$

where $f(x^+)$ is the optimal function value in the current tested hyperparameter combination, $\mu(x)$ and $\sigma(x)$ are the mean and variance of the agent model, respectively, $\hat{\Phi}$ is the cumulative density function of a normal distribution, and ξ is a balancing parameter.

The bayesian optimized XGBoost algorithm mainly includes 6 steps.

Step 1: Use the mean squared error of the training sets after 5-fold cross-validation on the XGBoost model as the objective function. Set the main parameters of the model as the array to be optimized and initialize the parameters optimization range, and randomly generate initial point parameters sets and find the function value of each parameters set.

Step 2: Build a gaussian agent model based on the parameters sets of the current initial points and the function value of the parameters sets.

Step 3: Maximize the acquisition PI function to select the next most potential hyperparameter combination x_{t+1}.

Step 4: Evaluate the value of the objective function $f(x_{t+1})$ based on the selected hyperparameter combination x_{t+1}.

Step 5: Add the newly obtained input hyperparameter combination - observation value pair $\{x_{t+1}, f(x_{t+1})\}$ to the historical observation data $D_{1:t}$, and update the Gaussian agent model to prepare for the next iteration.

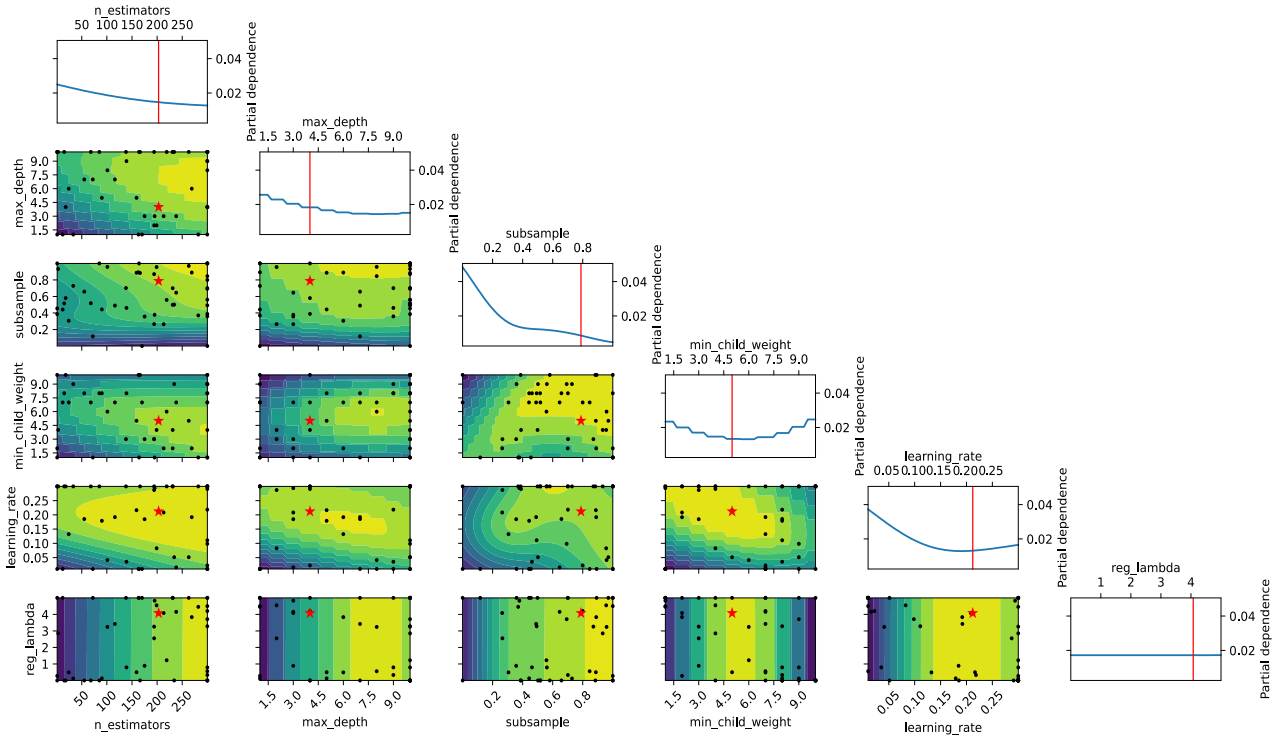

FIGURE 9. Bayesian search hyperparameters of XGBoost.

Step 6: Judge whether to meet the preset error condition of the model, if not, continue to iterate. If the condition is met, output the best hyperparameter combinations to complete the construction of the model.

The bayesian optimized hyperparameters of XGBoost model are shown in Fig. 9. The selection of model hyperparameters is determined based on the mean squared error. When the mean squared error reaches the minimum, the optimal parameters can be found. The coordinates of each subplot in the figure are the range of each parameter selection. The black dots indicate the positions of the sampling points during the bayesian optimized process. The red pentagrams are the optimal points of the parameters.

Compare the bayesian optimized hyperparameters in Fig. 9 with the unoptimized default hyperparameters, the results are shown in Table 3.

F. EVALUATION AND ANALYSIS OF PREDICTION RESULTS

By comparing and analyzing the prediction effects of these four models on the test sets, we can find an effective algorithm to correct the impact of multiple factors on the measurement of water holdup. Fig. 10 shows the fitted effect between the predicted results of the four models and the actual water holdup. The closer the scatter in the figure is to the ideal fitted straight line, the closer it is to the actual water holdup. It can be seen that the prediction effect of BO-XGBoost is the best, followed by XGBoost, and the other two models have poor fitted performance at high water holdup.

In order to make further efforts to test the prediction effect of the BO-XGBoost model, 90 water holdup samples at

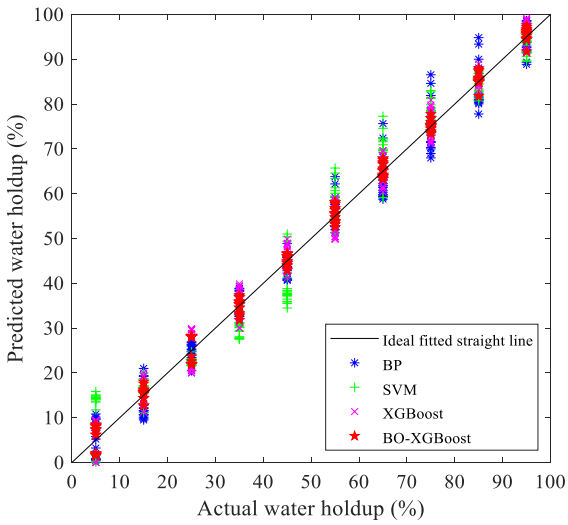

FIGURE 10. Fitted effects of prediction results and actual water holdup.

temperatures of 35°C and 55°C are selected, respectively. The prediction results are shown in Fig. 11. It can be seen that the prediction results have a good performance at both temperature conditions. Fig. 12 reflects the prediction error of the BO-XGBoost model. The error of most samples is within ±2%, the error of a few samples is slightly larger, but still within ±5%, which has a good prediction effect.

In order to more effectively evaluate the prediction performance of the four models, the fitted coefficient (R^2), the root mean square error (*RMSE*), the maximum absolute

TABLE 3. Parameter comparison results.

Hyperparameters	Illustration	BO-XGBoost	XGBoost
n_estimators	number of decision tree	203.0	100.0
max_depth	maximum depth of decision tree	4.0	6.0
subsample	the proportion of randomly selected samples	0.8	1.0
min_child_weight	the minimum number of leaf node samples	5.0	1.0
learning_rate	learning rate	0.21	0.3
reg_lambda	regular term penalty cofficient	4.10	1.0

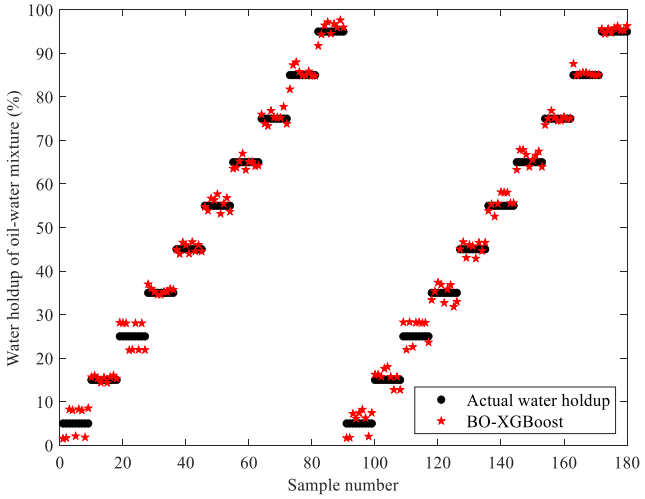

FIGURE 11. Prediction results of BO-XGBoost model.

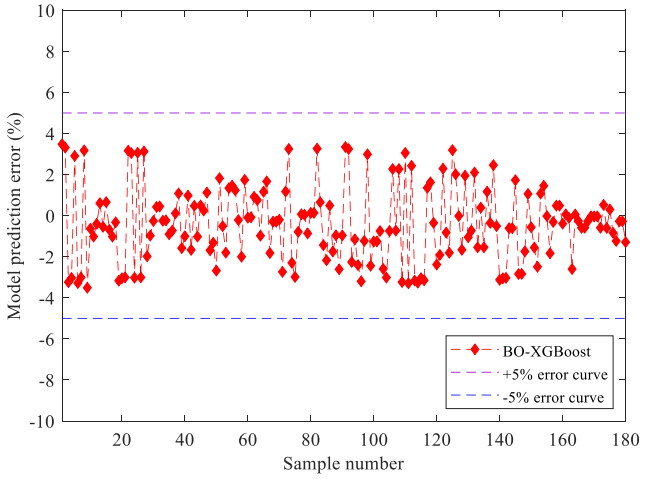

FIGURE 12. Prediction error of BO-XGBoost model.

error ($AE_{\max}$), and the mean absolute error (MAE) are used as evaluation indicators. The calculation formula of the evaluation index is as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{10}$$

$$AE_{\max} = \max\left|y_i - \hat{y}_i\right| \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value of the sample, and $\bar{y}$ is the average value of y_i.

TABLE 4. Comparison of water holdup prediction performance of four models.

Algorithm	R^2	$RMSE$	$AE_{\max}$	MAE
BP	0.9825	0.037	11.53%	3.16%
SVM	0.9677	0.052	12.27%	4.26%
XGBoost	0.9859	0.035	5.87%	3.11%
BO-XGBoost	0.9958	0.019	3.51%	1.50%

According to Table 4, it can be seen that: (1) Comparing R^2 values, BO-XGBoost has the best fitted effect. (2) Comparing $RMSE$ values and $AE_{\max}$ values, BO-XGBoost<XGBoost<BP<SVM, BO-XGBoost has the best stability and accuracy in prediction results. (3) From the perspective of MAE, the MAE of BO-XGBoost remains the smallest. In summary, BO-XGBoost has the best generalization ability and prediction effect, which effectively corrects the influence of temperature and mineralization on the measurement of water holdup.

V. SUMMARY

To correct for the complex nonlinear effects of temperature and mineralization on the measurement of water holdup in oil-water two-phase flow using the transmission lines method. This paper uses a combination of theoretical analysis, numerical simulation analysis and indoor experiments to study the influence of multiple factors on the phase shift of transmission lines signal. BY using 297 pieces of experimental data, prediction models based on BP, SVM, XGBoost, and BO-XGBoost are developed, and 180 sets of test samples are used to verify the prediction effect of water holdup. The results show:

(1) XGBoost outperforms traditional BP neural network and SVM in water holdup prediction, with the advantages of wider prediction range and higher accuracy.

(2) The average absolute error of the XGBoost model after Bayesian optimization is only 1.50%, which provides a solution for reliable prediction of the water holdup of oil-water two-phase flow and has practical engineering significance.

REFERENCES

[1] L. Han, M. Chen, X. Liu, and C. Fu, "New measurement method of oil-water two-phase flow with high water holdup and low rate by phase state regulation," *Meas. Sci. Rev.*, vol. 23, no. 6, pp. 268–274, Dec. 2023, doi: 10.2478/msr-2023-0034.

[2] Y. Lv, S. Chen, G. Lv, and L. He, "Oil-water two-phase flow with three different crude oils: Flow structure, droplet size and viscosity," *Energies*, vol. 17, no. 7, p. 1573, Mar. 2024, doi: 10.3390/en17071573.

[3] H. Zhang, L. Zhai, C. Yan, H. Wang, and N. Jin, "Capacitive phase shift detection for measuring water holdup in horizontal oil–water two-phase flow," *Sensors*, vol. 18, no. 7, p. 2234, Jul. 2018, doi: 10.3390/s18072234.

[4] Y. F. Han, N. D. Jin, Y. Y. Ren, T. H. Han, and Z. H. Huang, "Measurement of local oil holdup for oil-in-water emulsion flows using multiple mini-conductance probes," *Measurement*, vol. 121, pp. 6–18, Jun. 2018, doi: 10.1016/j.measurement.2018.02.025.

[5] A. Zhao, Y.-F. Han, Y.-Y. Ren, L.-S. Zhai, and N.-D. In, "Ultrasonic method for measuring water holdup of low velocity and high-water-cut oil-water two-phase flow," *Appl. Geophys.*, vol. 13, no. 1, pp. 179–193, Apr. 2016, doi: 10.1007/s11770-016-0547-z.

[6] M. Tayyab, M. S. Sharawi, and A. Al-Sarkhi, "A radio frequency sensor array for dielectric constant estimation of multiphase oil flow in pipelines," *IEEE Sensors J.*, vol. 17, no. 18, pp. 5900–5907, Sep. 2017, doi: 10.1109/JSEN.2017.2732164.

[7] Y. Wei, H. Yu, Q. Chen, G. Liu, and J. Chen, "Measurement of water holdup in oil-water two-phase flows using coplanar microstrip transmission lines sensor," *IEEE Sensors J.*, vol. 19, no. 23, pp. 11289–11300, Dec. 2019, doi: 10.1109/JSEN.2019.2935022.

[8] J. Ma, N.-D. Jin, D.-Y. Wang, D.-Y. Liu, and W.-X. Liu, "Measurement of water holdup in vertical upward high water-cut oil-in-water flows using a high frequency sensor," *Sens. Actuators A, Phys.*, vol. 289, pp. 165–179, Apr. 2019, doi: 10.1016/j.sna.2019.02.030.

[9] W. Liu, Z. Chen, Y. Hu, and L. Xu, "A systematic machine learning method for reservoir identification and production prediction," *Petroleum Sci.*, vol. 20, no. 1, pp. 295–308, Feb. 2023, doi: 10.1016/j.petsci.2022.09.002.

[10] F. Harrou, T. Cheng, Y. Sun, T. Leiknes, and N. Ghaffour, "A data-driven soft sensor to forecast energy consumption in wastewater treatment plants: A case study," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4908–4917, Feb. 2021, doi: 10.1109/JSEN.2020.3030584.

[11] P. Yan, B. Shen, and Y. Wang, "Soft sensor for VFA concentration in anaerobic digestion process for treating kitchen waste based on DSTHELM," *IEEE Access*, vol. 8, pp. 223618–223625, 2020, doi: 10.1109/ACCESS.2020.3042512.

[12] L. Hua, C. Zhang, W. Sun, Y. Li, J. Xiong, and M. S. Nazir, "An evolutionary deep learning soft sensor model based on random forest feature selection technique for penicillin fermentation process," *ISA Trans.*, vol. 136, pp. 139–151, May 2023, doi: 10.1016/j.isatra.2022.10.044.

[13] Y. M. Zhou, S. W. Wang, and L. Lin, "An application of BP neural network model to predict the moisture content of crude oil," *Adv. Mater. Res.*, vols. 524–527, pp. 1327–1330, May 2012, doi: 10.4028/www.scientific.net/amr.524-527.1327.

[14] L. Chang and D. Chen, "Research on outside pipe measurement of crude oil water content based on improved BP neural network," *Instrum. Tech. Sens.*, vol. 53, no. 9, pp. 123–126, 2016.

[15] D. Zhang and B. Xia, "Soft measurement of water content in oil-water two-phase flow based on RS-SVM classifier and GA-NN predictor," *Meas. Sci. Rev.*, vol. 14, no. 4, pp. 219–226, Aug. 2014, doi: 10.2478/msr-2014-0030.

[16] Y. Wei, H. Yu, Q. Chen, G. Liu, C. Qi, and J. Chen, "A novel conical spiral transmission line sensor-array water holdup detection tool achieving full scale and low error measurement," *Sensors*, vol. 19, no. 19, p. 4140, Sep. 2019, doi: 10.3390/s19194140.

[17] Q. Wang, X. Fu, F. J. Duan, J. J. Jiang, and T. Y. Li, "Study on simplified compensation model of temperature-salinity for water content measurement of crude oil," *Instrum. Tech. Sens.*, vol. 53, no. 1, pp. 121–126, 2022.

[18] M. Derakhshi, M. Chahardowli, and M. Simjoo, "Mathematical modeling of in-depth gel treatment for water control in a heterogeneous oil reservoir: Gel kinetics coupled by flow transport model," *Petroleum Sci. Technol.*, vol. 39, nos. 7–8, pp. 270–287, Apr. 2021, doi: 10.1080/10916466.2021.1892756.

[19] A. F. Ibrahim, R. Al-Dhaif, S. Elkatatny, and D. A. Shehri, "Applications of artificial intelligence to predict oil rate for high gas–oil ratio and water-cut wells," *ACS Omega*, vol. 6, no. 30, pp. 19484–19493, Aug. 2021, doi: 10.1021/acsomega.1c01676.

[20] C. Huang and Y. Li, "A short-term prediction method for PV power generation based on SVM weather classification and PSO-BP neural network," in *Proc. IEEE 2nd Int. Power Electron. Appl. Symp. (PEAS)*, Nov. 2023, pp. 2544–2549, doi: 10.1109/peas58692.2023.10394800.

[21] S. Shi, J. Liu, H. Hu, and H. Zhou, "A research on a GA-BP neural network based model for predicting patterns of oil-water two-phase flow in horizontal wells," *Geoenergy Sci. Eng.*, vol. 230, Nov. 2023, Art. no. 212151, doi: 10.1016/j.geoen.2023.212151.

[22] J. Tian and J. Shi, "A high-accuracy and fast retrieval method of atmospheric parameters based on genetic-BP," *IEEE Access*, vol. 10, pp. 19458–19468, 2022, doi: 10.1109/ACCESS.2022.3151868.

[23] Y. Gan, B. Meng, Y. Chen, and F. Sun, "An intelligent measurement method of the resonant frequency of ultrasonic scalpel transducers based on PSO-BP neural network," *Measurement*, vol. 190, Feb. 2022, Art. no. 110680, doi: 10.1016/j.measurement.2021.110680.

[24] Z. Zhao, J. Yuan, and L. Chen, "Research on air traffic flow management delay distribution prediction based on IV value and PSO-SVM," *IEEE Access*, vol. 11, pp. 84035–84047, 2023, doi: 10.1109/ACCESS.2023.3300373.

[25] N. D. Pah, V. Indrawati, and D. K. Kumar, "Voice-based SVM model reliability for identifying Parkinson's disease," *IEEE Access*, vol. 11, pp. 144296–144305, 2023, doi: 10.1109/ACCESS.2023.3344464.

[26] M. Wang, H. Song, M. Li, and C. Wu, "Prediction of split-phase flow of low-velocity oil-water two-phase flow based on PLS-SVR algorithm," *J. Petroleum Sci. Eng.*, vol. 212, May 2022, Art. no. 110257, doi: 10.1016/j.petrol.2022.110257.

[27] L. Han, Y. Zhang, X. Wang, and H. Yang, "Current harmonics minimization of model-free predictive current control for PWM rectifiers based on hybrid SVM," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 12, no. 1, pp. 486–495, Feb. 2024, doi: 10.1109/JESTPE.2023.3334308.

[28] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration," *Eng. Comput.*, vol. 38, no. S5, pp. 4145–4162, Dec. 2022, doi: 10.1007/s00366-021-01393-9.

[29] J. Wang, W. Niu, and Y. Yang, "Wind turbine output power prediction by a segmented multivariate polynomial-XGBoost model," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, vol. 46, no. 1, pp. 505–521, Nov. 2023, doi: 10.1080/15567036.2023.2284840.

[30] J. Su, Y. Wang, X. Niu, S. Sha, and J. Yu, "Prediction of ground surface settlement by shield tunneling using XGBoost and Bayesian optimization," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105020, doi: 10.1016/j.engappai.2022.105020.

[31] X. Wang, P. Pan, and J. Li, "Real-time measurement on dynamic temperature variation of asphalt pavement using machine learning," *Measurement*, vol. 207, Feb. 2023, Art. no. 112413, doi: 10.1016/j.measurement.2022.112413.

[32] R. Shi, X. Xu, J. Li, and Y. Li, "Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization," *Appl. Soft Comput.*, vol. 109, Sep. 2021, Art. no. 107538, doi: 10.1016/j.asoc.2021.107538.

[33] C. Zhang and X. Zhou, "Forecasting credit risk of SMEs in supply chain finance using Bayesian optimization and XGBoost," *Math. Problems Eng.*, vol. 2023, pp. 1–14, Dec. 2023, doi: 10.1155/2023/5609996.

XIAOMEI DAI was born in Shiyan, China, in 2000. She received the bachelor's degree in electronic information engineering from Yangtze University, in 2022, where she is currently pursuing the master's degree, with a focus on signal and information processing. She is good at deeply integrating artificial intelligence technology with the oil field, and improving the efficiency and accuracy of oil exploration, mining, production, and management through algorithm optimization and data analysis.

YONG WEI received the B.S. degree in electrical and information engineering from Jianghan Petroleum Institute, China, in 2003, and the M.S. degree in signal and information processing and the Ph.D. degree in geodetection and information technology from Yangtze University, Jingzhou, China, in 2006 and 2016, respectively. From October 2018 to October 2019, he was a Visiting Scholar with the University of Houston, Houston, TX, USA. He is currently a Professor and the Vice President of the School of Electronic Information, Yangtze University. His research interests include new methods and instruments for acoustic and electric well logging.

RUYI GAN received the master's degree in engineering from Chongqing University of Posts and Telecommunications, in June 2017, and the Ph.D. degree in engineering from Chongqing University, in June 2023, with a focus on instrument science and technology. In July 2023, he joined the Automation Department, School of Electronic Information and Electrical Engineering, Yangtze University. His research interest includes underground dynamic liquid level detection.

BAOJUN WEI received the degree in control science and engineering from China University of Petroleum, Beijing, in 2011. He is currently with the Changqing Branch of CNPC Well Logging Company Ltd. He is mainly engaged in production logging-related technology research, with systematic professional knowledge and rich field practical experience, proficient in various logging technologies and methods, focusing on the pain points and difficulties of party A to carry out process technology research, to solve several first-line production problems. He is a Registered Blasting Engineer and a Registered Safety Engineer.

YALIN XIANG received the master's degree in signal and information processing from Yangtze University, in 2004. Currently, he is with Southwest Branch of CNPC Well Logging Company Ltd. He is mainly engaged in on-site data collection work and has the ability to perform open hole logging, casing logging, and through bit storage logging operations. He has long been responsible for shale gas and tight gas logging construction tasks in Sichuan, Chongqing Region. He is proficient in various logging techniques and carries out process technology breakthroughs around the pain points and difficulties of party A, solving multiple frontline production problems.

PING LIU was born in Huangdao, Qingdao, Shandong, China, in 1988. He received the bachelor's degree in geological engineering from Chang'an University, in 2011, and the master's degree in geology from China University of Petroleum, Beijing, in 2014. He has authored one book, more than ten articles, and six inventions. He received the Excellent Paper Award at the 22nd Annual Well Testing Conference of the Chinese Petroleum Society. He is a Reviewer of the International Conference on Oil and Gas Field Exploration and Development (IEFDC).