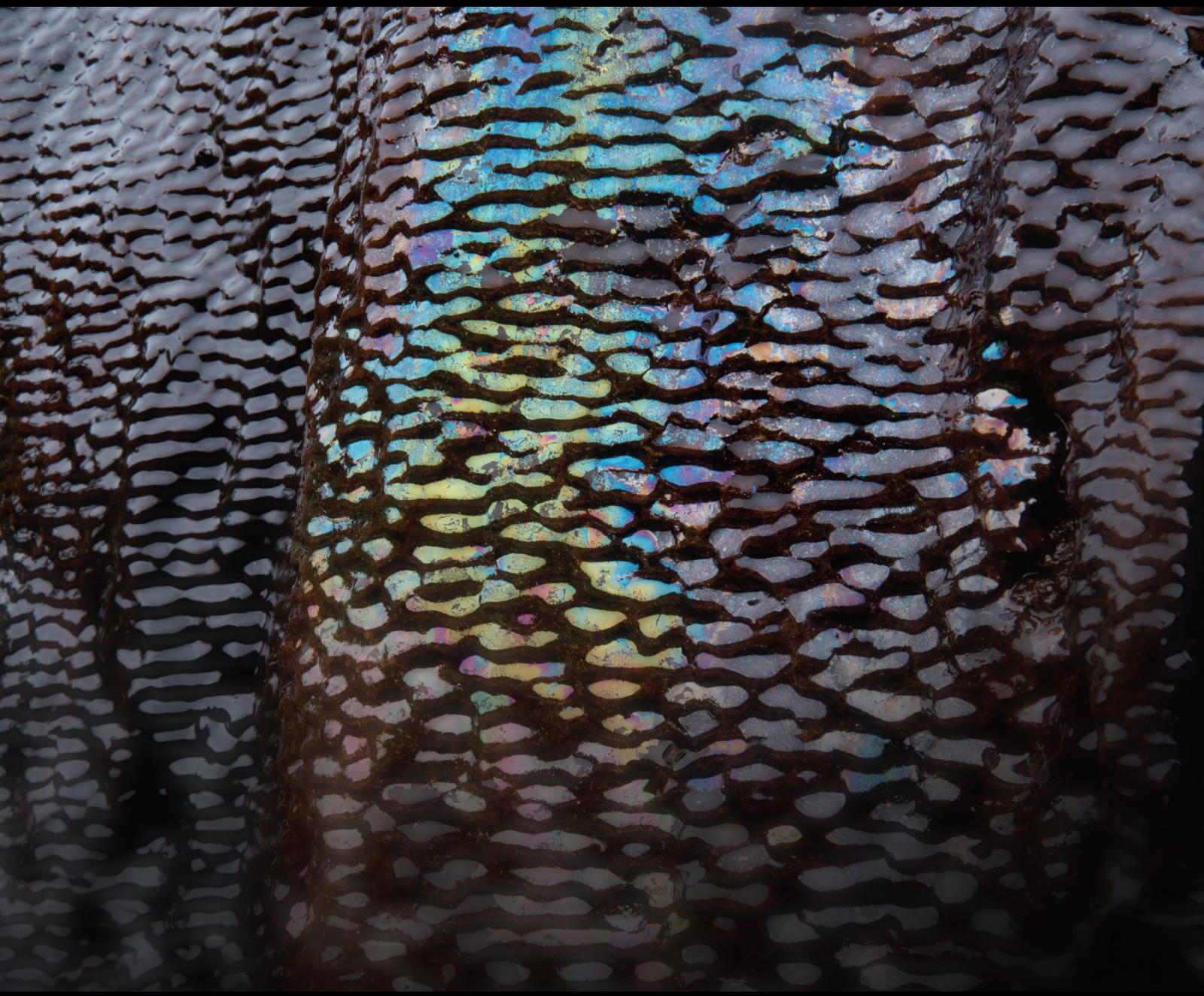
Geofluids

Advanced Data-Driven Modeling of Geological Processes

Lead Guest Editor: Bicheng Yan
Guest Editors: Keliu Wu, Jose Antonio Torres, Zhenzhen Wang, and Yating Wang



Advanced Data-Driven Modeling of Geological Processes

Geofluids

Advanced Data-Driven Modeling of Geological Processes

Lead Guest Editor: Bicheng Yan
Guest Editors: Keliu Wu, Jose Antonio Torres, Zhenzhen Wang, and Yating Wang

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
Geofluids
Volume 2023, Article ID 5467956, 16 pages
https://doi.org/10.1155/2023/5467956

Research Article

CNN-LSTM Model Optimized by Bayesian Optimization for Predicting Single-Well Production in Water Flooding Reservoir

Lei Zhang[1,2], **Hongen Dou**[1,2], **Kun Zhang**[3], **Ruijie Huang**[1], **Xia Lin**[1,2], **Shuhong Wu**[1,2], **Rui Zhang**[4], **Chenjun Zhang**[1] and **Shaojing Zheng**[1]

[1]*Research Institute of Petroleum Exploration and Development, Beijing 100083, China*
[2]*Artificial Intelligence Technology R&D Center for Exploration and Development, CNPC, Beijing 100083, China*
[3]*The Twelfth Oil Production Plant, Changqing Oilfield Company, Qingyang 745099, China*
[4]*Changqing Engineering Design Co., Ltd., PCOC, Xi'an 710021, China*

Correspondence should be addressed to Hongen Dou; douhongen@petrochina.com.cn
and Ruijie Huang; huangruijie@petrochina.com.cn

Received 24 March 2022; Revised 17 November 2022; Accepted 18 March 2023; Published 22 May 2023

Academic Editor: Yating Wang

Geared toward the problems of predicting the unsteadily changing single oil well production in water flooding reservoir, a machine learning model based on CNN (convolutional neural network) and LSTM (long short-term memory) is established which realizes precise predictions of monthly single-well production. This study is considering more than 60 dynamic and static factors that affect the changes of oil well production, introduce water injection parameters into data set, select 11 main control factors, and then, build a CNN-LSTM model optimized by Bayesian optimization. The effectiveness of the proposed model is verified in a realistic reservoir. The experiment results show that the prediction accuracy of the proposed model is over 90%, which suggests a penitential application in an extensive range of applications. Production forecasting by the developed model is simple, efficient, and accurate, which can provide a guidance for the dynamic analysis of a water flooding reservoir, and work as a good reference of the development and production of other types of reservoirs.

1. Introduction

Oil production prediction runs through the entire development course of a water-driven oilfield; it is the foundation of the well stimulation and plays an important role in making investment decisions. For a long time, the main production prediction methods are reservoir numerical simulation, mechanism model, and decline curve analysis, which have their respective advantages and disadvantages. Since the 1990s, fuzzy comprehensive evaluation (FE), BP neural network (BPNN), grey model (GM), and other methods have been applied to oilfield production prediction [1, 2]. Recently, with the maturity and wide application of big data, artificial intelligence, and other technical theories and methods, more and more intelligent algorithms have been introduced into the petroleum industry, providing a new way to solve complex engineering problems [3, 4].

As a nonlinear fitting method, machine learning can learn rules from data and make a prediction. In the past 20 years, there has been extensive research on production prediction based on machine learning. Common methods include random forest (RF), support vector machine (SVM), fuzzy comprehensive evaluation (FE), artificial neural network (ANN), and autoregressive integrated moving average model (ARIMA) [5–14]. These classical machine learning methods have been fully applied in the field of petroleum industry and show strong vitality.

Y. Duana et al. use RTS (Rauch Tung Striebel) to smooth the gas production series, and then, an ARIMA model was established to predict the gas production [6]. J. Gu et al. proposed an oil well production prediction model which combines ARIMA and Kalman filter to eliminate the influence of nonsynchronicity and hysteresis [7]. Viet Nguyen-Le et al. propose 3 ANN models to predict the

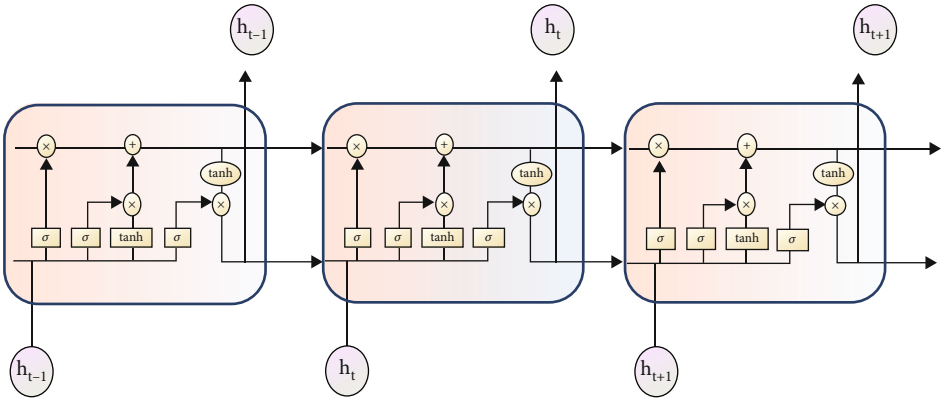


Figure 1: The basic structure of LSTM cell. In each cell unit, long-term information and short-term information are conveyed, respectively.

parameters in Arps's hyperbolic decline and then reconstructs the production profile [13]. Y. Zhu et al. established an "ε-SVM" production prediction model based on sequential minimal optimization (SMO) algorithm using production records and bottom hole pressure data [14].

Since 2015, the recurrent neural networks represented by long short-term memory (LSTM) and gated recurrent unit (GRU) become a new hot spot in the field of production prediction [15–22].

L. Zhang et al. proposed a GRU-FNN model to predict single-well production in water drive oilfield [23]. X. Song et al. adopted particle swarm optimization (PSO) to optimize the basic structure of LSTM model and achieved better model performance [24]. Weiss applied fuzzy ranking and neural network to establish correlations to predict oil production [25]. H. Wang et al. established a production prediction model for high-water-cut oil fields considering both timing and engineering factors based on LSTM and increased the input features to 17 items, with a relative error of only 1% [26]. Sagheer et al. compared the application effects of simple RNN and LSTM in oil and gas production predictions, verified that LSTM had better performance than simple RNN [27].

All the predictions made in the above studies are pretty accurate, but there is still room for further improvements. First, most of the works are aimed at predicting the total production of a reservoir or an oilfield, few research focused on the single-well production prediction. In fact, accurately predicting the production of single well is not easy because the fluctuation of single-well production is more irregular and more drastic than that of the block/reservoir production. Also, there are too many factors affecting single-well production, and we can hardly analyze them clearly. Second, in most studies, the production series is divided into 3 segments based on time, the earlier one for training, the middle one for validating, and the latest one for testing. However, as we all know, the production changes of oilfield/oil well have significant stages, and the early stage, middle stage, and late stage have different dominant control factors or rules of change. Learning the rules in early stage and predicting the production changes in late stage are bound to cause errors. Finally, there are few studies that consider both dynamic and static characteristics in time series prediction. Based on these considerations, to achieve accurate single-well production prediction in water flooding oilfield, the data set was built considering over 60 geological and development factors that affect oil well production. The influence of the water injection on oil production is also quantified and added to the data set as a feature. The missing data are filled in accordance with the industry knowledge and the distribution pattern of data. Secondly, the feature correlation analyses tailored to time series prediction are carried out, through which 11 dominant control factors of oil well production are selected to build the samples. The production prediction model is established based on a one-dimensional convolution neural network and a long short-term memory neural network, and the model hyperparameters are optimized by Bayesian optimization. Finally, an example numerical test is carried out in a practical reservoir. Results show that the CNN-LSTM model has better performance compared with the Bi-LSTM model, Attention-LSTM model, or any other models. Production forecasting by the developed model is simple, efficient, and accurate, which can provide a guidance for the dynamic analysis of a water flooding reservoir and work as a good reference of the development and production of other types of reservoirs.

2. Methodology

2.1. LSTM and Conv1D. LSTM (long short-term memory) can be the most successful recurrent neural network (RNN) nowadays. Its unique design of gate structure and cell state makes it possible to capture long-term dependencies, which is also the key to its great success. As shown in Figure 1, different from the other simple RNN units, there are three gates in the LSTM cell unit: forgetting gate, input gate, and output gate. In each time step, LSTM cells will partially pass the information to the next step and retain some information in the cell as the cell state, which is the key for LSTM to achieve long-term memory. LSTM neural network performs well in the field of time series prediction [28].

Conv1D (1-dimensional convolutional neural network) is famous in the areas of image recognition, but in recent years, people find that 1-dimensional CNN (Conv1D) can well accomplish the task of time series prediction and even outperforms LSTM and other time series prediction models

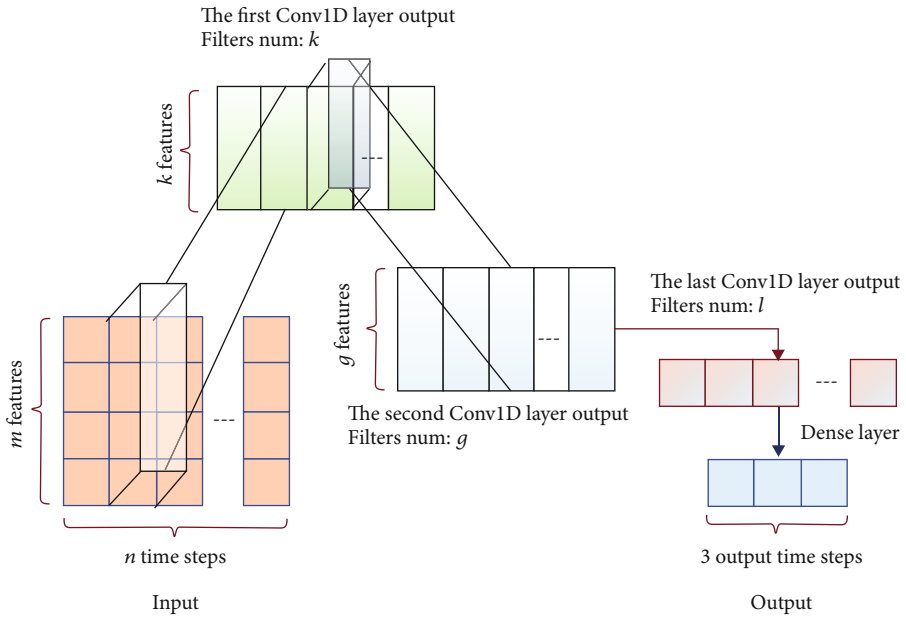


Figure 2: Diagram of Conv1Ds. For the inputs sized $m \times n$ and a Conv1D layer with k filters, the outputs' size is $k \times n$. Using multiple Conv1Ds, the final output size is $1 \times n$. The model illustrated in this figure wants to predict production for the next three months, that is, it wants three outputs, which we achieve using a simple dense layer containing three neurons.

Table 1: Initial data set. The italicized items are category features which need to be one-hot encoding.

Well basic data				Reservoir data of well		
Name	*X*_coordinate	*Y*_coordinate	Depth			
Unit	*Layer*	Start date	Flooding date	Position	Top depth (MD)	Bottom depth (MD)
Inject/production	*Well type*	Initial temperature	Initial pressure	Top depth (TVD)	Bottom depth (TVD)	Thickness of sand bodies
Initial pressure coefficient	Saturation pressure	Initial solution gas-oil ratio	Viscosity	Effective thickness	Porosity	Permeability
Monthly producing well production data						
Time	*Producing scheme*	Production days	Nozzle	Pump diameter	Pump depth	Stoke
Jig frequency	Monthly liquid	Monthly oil	Monthly water	Monthly gas	Gas-oil ratio	Water cut
Cumulative oil	Cumulative water	Cumulative liquid	Recovery	Static pressure	FBHP (bottom hole flowing pressure)	Working fluid level
Monthly water-injecting well injection data						
Injection methods	Injecting days	Pumping pressure	Casing pressure	Oil pressure	Injecting volume monthly	Cumulative injecting volume

in certain cases. When Conv1D is used for time series prediction, the inputs, n temporal steps with m features, can be viewed as an image of size of $n \times m$, and Conv1D can extract a new image from the input data by scanning and calculating in one direction. Changing the number of filters and the size of convolutional kernels, we can easily control the size of outputs (see Figure 2). It should be noted that the convolution kernel of Conv1D is also 2-dimensional, but it can only slide windows in one direction, and that is the difference between Conv1D and 2-dimensional convolutional neural network.

2.2. Spearman Correlation Coefficient. Spearman correlation coefficient, also known as the rank correlation coefficient, can measure the degree of nonlinear correlation between two features and is a method to analyze the correlation between two variables. Spearman correlation coefficient is calculated as in

$$\rho = \frac{\sum_{i=1}^{n}\left(x_j - \bar{x}\right)\left(y_j - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_j - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_j - \bar{y}\right)^2}}. \quad (1)$$

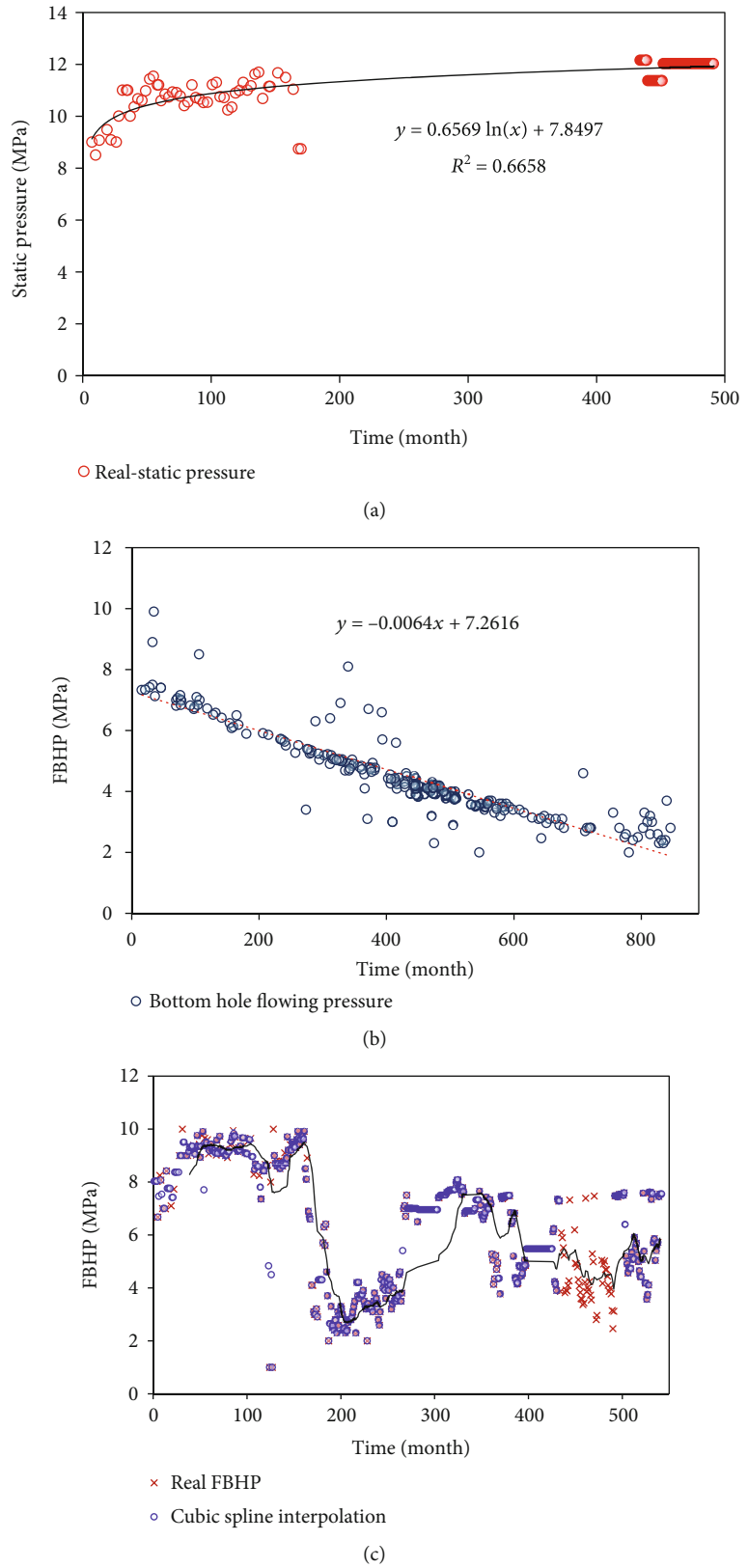


Figure 3: Static pressure, working fluid level and FBHP in well A. (a) The relationship between time and static pressure. (b) Comparison between real values and predicted values of FBHP. (c) The relationship between working fluid level and FBHP.

$\rho \in [-1, 1]$. When the ρ is negative, it means that the two features have a negative correlation, one feature increases means the other decreases. Similarly, if ρ is positive, the two features are positively associated. The larger the absolute value is, the stronger the correlation between two features is.

2.3. Bayesian Optimization. Acclaimed as one of the best hyperparameter tuning algorithm at present stage, Bayesian optimization is a general gradient-free global optimization strategy, which can identify a good set of hyperparameters with few iterations. It is suitable in two cases: (1) the objective function is extremely complex and time-consuming to evaluate; (2) the target function is difficult to differentiate in respect to the independent variable.

In each iteration, Bayesian optimization decomposes the optimization problem into multiple small optimization problems. It samples the original function curve by a certain method and builds an alternative curve by fitting these points. Bayesian optimization builds a model to describe the parameter-distribution of the objective function using the Gaussian process model and then solves the minimum value of the alternative model and uses this minimum value as the optimal solution of the original function in this iteration. This process is called surrogate optimization. By gradually increasing the number of sampling points, the model will gradually approach the original objective function, and an optimal combination of hyperparameters will be obtained.

In this paper, Bayesian optimization is used to optimize network model hyperparameters, including network depth and width, initial learning rate, and network activation functions. For all hyperparameters to be optimized, a maximum value and a minimum value are first specified for them, respectively. Based on the Bayesian optimization method in the package *Keras_tuner*, appropriate maximum iteration and optimization objective are set and then executed to obtain the optimal combination of hyperparameters.

3. Data Processing and Feature Engineering

3.1. Sources of Data. Data used in this paper were collected from 426 oil wells and 94 water wells in a Chinese reservoir named A with an average production time of 406 months. The oldest wells in the data set can trace back to 1956. During more than 60 years of development, the reservoir has experienced a variety of production methods such as pumping production, water flooding production, and fracturing production. At present, the reservoir has entered the ultrahigh water cut stage, with an average water cut of more than 95%.

The initial data include monthly production data of 426 oil wells, monthly injection data of 94 water-injecting wells, and single-well reservoir information. The specific items are shown in Table 1.

3.2. Data Preprocessing

3.2.1. Filling. In this paper, most of the missing data are bottom hole flowing pressure, reservoir static pressure, or dynamic liquid level height. According to their own characteristics, we use different filling methods, taking well A as an example to illustrate.

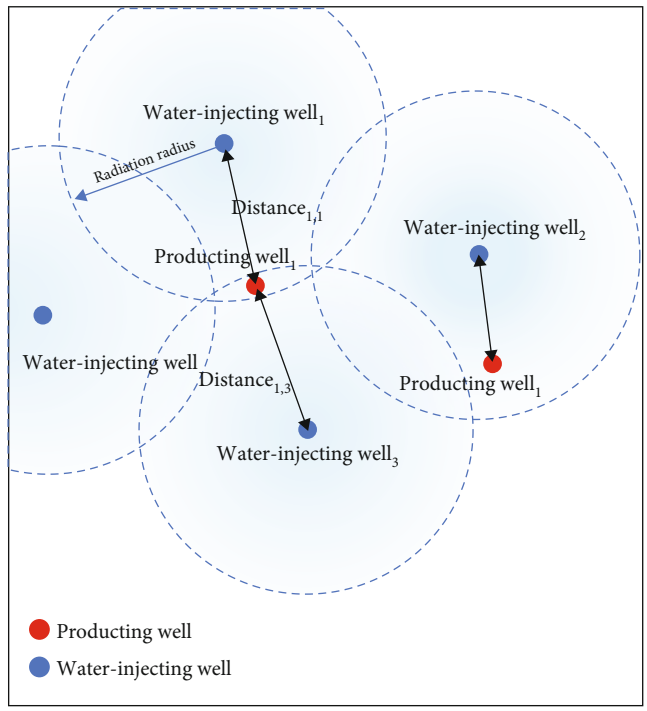


Figure 4: Diagram of injection-production relationship. Producing wells always be influenced by several water-injecting wells.

(1) Reservoir static pressure

According to the oil field development experience, the change of static pressure during the oilfield development process will not be significant and may have an obvious trend; therefore, the regression fitting method can be useful. Through the regression analysis using existing data, the relationship between the static pressure and time can be obtained, and the missing static pressure data will be calculated, as shown in Figure 3(a).

(2) Bottom hole flowing pressure (FBHP)

Figure 3(b) shows that changes of bottom hole flowing pressure during a well's production life has stages, and the range of FBHP in different stages varies greatly, which limits the application of fitting regression method. In this case, we use the cubic spline interpolation, and the filling results are shown in Figure 3(b), which is basically consistent with the actual value distribution.

(3) Working fluid level

As the working fluid level and FBHP were highly correlated (see Figure 3(c)), we can easily obtain a linear relationship between the two features, and one can be calculated from another. If either the working fluid level height or bottom hole flowing pressure is available, cubic spline interpolation is used.

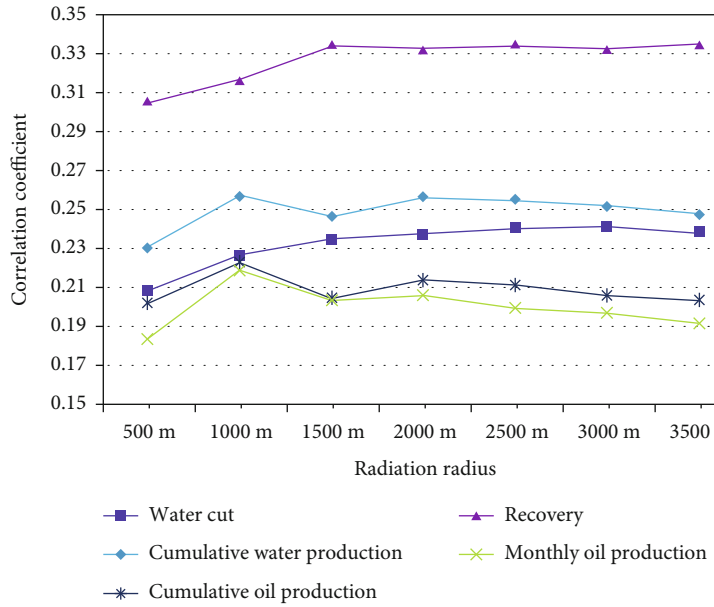


Figure 5: The correlation between producing well injected volume and production features. Seven correlation coefficients are used to compare; they are ρ (radiation radius, water cut), ρ (radiation radius, cumulative oil production), ρ (radiation radius, cumulative water production), ρ (radiation radius, monthly oil production), and ρ (radiation radius, recovery).

3.2.2. Encoding. Since a large part of static features in the data set are category features, the one-hot encoding process for category features should be carried out to concordance numerical features and category features. The features that are treated with this operation include units, layers, well types, and producing schemes.

3.2.3. Measurement of Injection-Production Relationship. A major factor that affects the production of oil well in water-driven reservoir is waterflooding measure. In this paper, we propose the parameter injection_i^t, producing well i injected volume in month t, and add it into the data set as a feature "inject."

As shown in Figure 4, an oil/producing well is often controlled by n water/water-injecting wells. We assume that a water-injecting well has a certain radiation radius, and in the radiation radius around a water well, the oil well within the radius is affected by that water-injecting well, and the smaller the distance is, the greater the impact of inject well on producing well is, and the larger producing well injected volume is. The total producing well injected volume equals the superposition of all "affected water-injecting wells."

According to this hypothesis, we can quantify the producing well injected volume, as shown in

$$\text{distance}_{i,j} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}, \tag{2}$$

$$\text{injection}_i^t = \sum_{j=1}^{m} \frac{\alpha_j^t}{\text{distance}_{i,j}}. \tag{3}$$

Here, i denotes the producing well i, j denotes the water-injecting well j, m denotes the number of water-injecting wells who influences producing well i. And the distance between producing well i and water-injecting well j is $\text{distance}_{i,j}$. In tth month, α_j^t units of water were injected to reservoir by water-injecting well j.

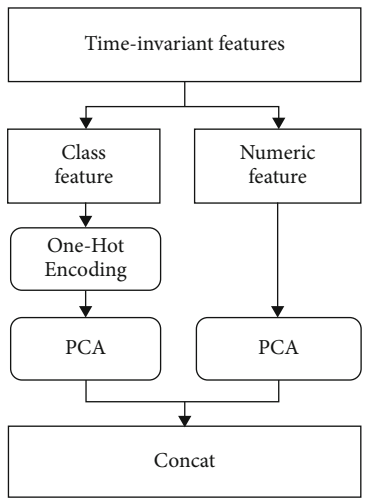


Figure 6: The data processing working flow. Static features and dynamic features need to be treated separately, also the discrete features and continuous features do.

The radiation radius is determined by correlation analysis: the closer the radiation radius to the actual radiation radius, the higher the correlation between producing well injected volume and the well production. By establishing the relationship between radiation radius and correlation coefficient, we can find the optimal radiation radius. As shown in Figure 5, 1000 m is the optimal radiation radius. Under this radius, producing well injected volume in a

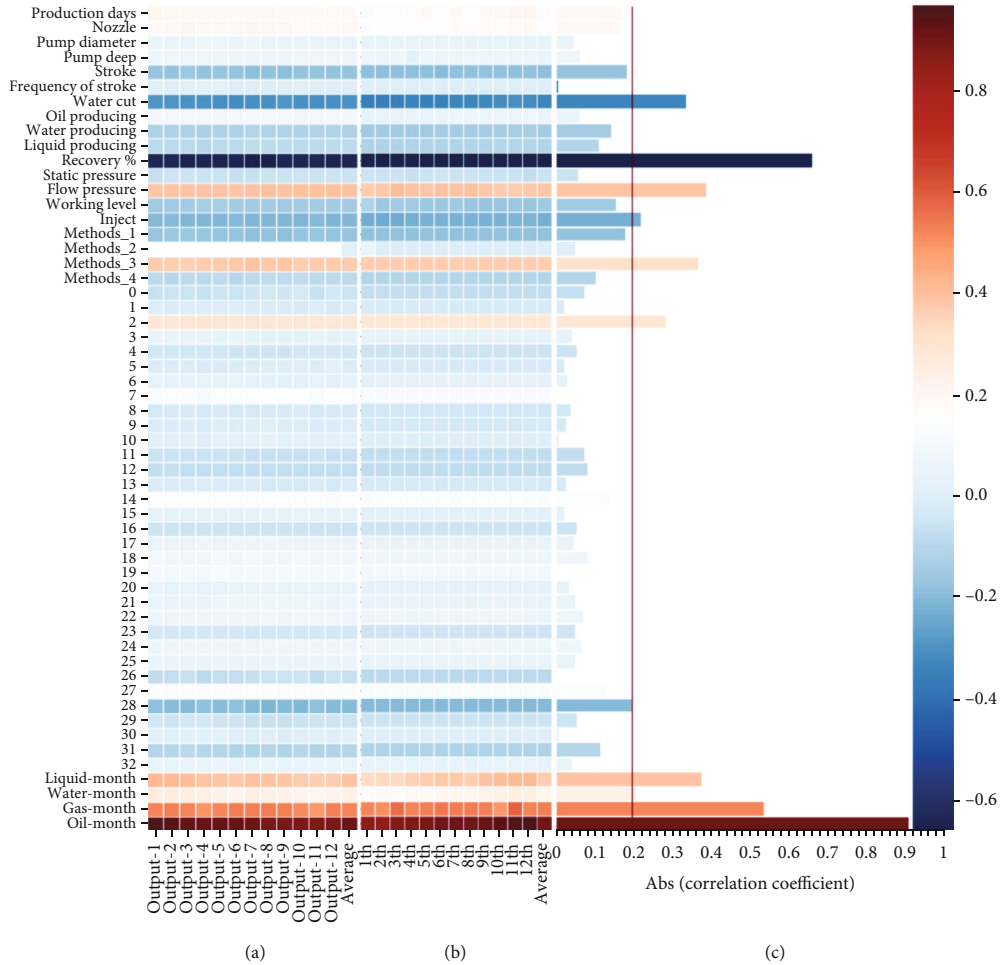


Figure 7: Heatmap of correlation coefficient between inputs and outputs. (a) The average of 12 correlation coefficients between the production in the next nth month and the features in the last 12 months. (b) The correlation coefficient between the production in the next 1th month and the features in the last mth months. (c) The average of 12 correlation coefficients between the production in the next 1th month and the features in the last 12 months.

certain time and certain producing well is calculated and added to the data set as the feature "inject."

3.3. Feature Engineering

3.3.1. Feature Dimension Reduction. For better prediction accuracy, feature compression is needed to improve the quality of features and limit the number of features. In the process of feature compression, static features and dynamic features need to be treated separately, so are the discrete features and continuous features. Otherwise, the static features representing the characteristics of individual wells will submerge in the continuously changing dynamic features. Similarly, the highly sparse 0-1 features created by one-hot encoding will also submerge in the continuously changing numerical features. In the data processing working flow of this paper (Figure 6), we first separate the dynamic features (features that change over time), as shown in Table 1. There are in total 11 dynamic features without dimension reduction. Secondly, we separate the 0-1 features from the static features for principal component analysis (PCA) dimensionality reduction (with a confidence of 95%); then,

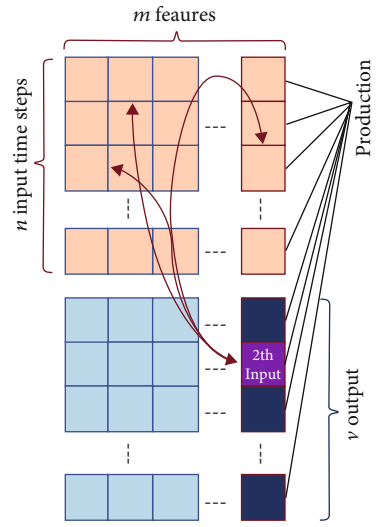


Figure 8: Correlation analysis method in this paper. The correlation is between inputs' features and outputs, not simply between the different features.

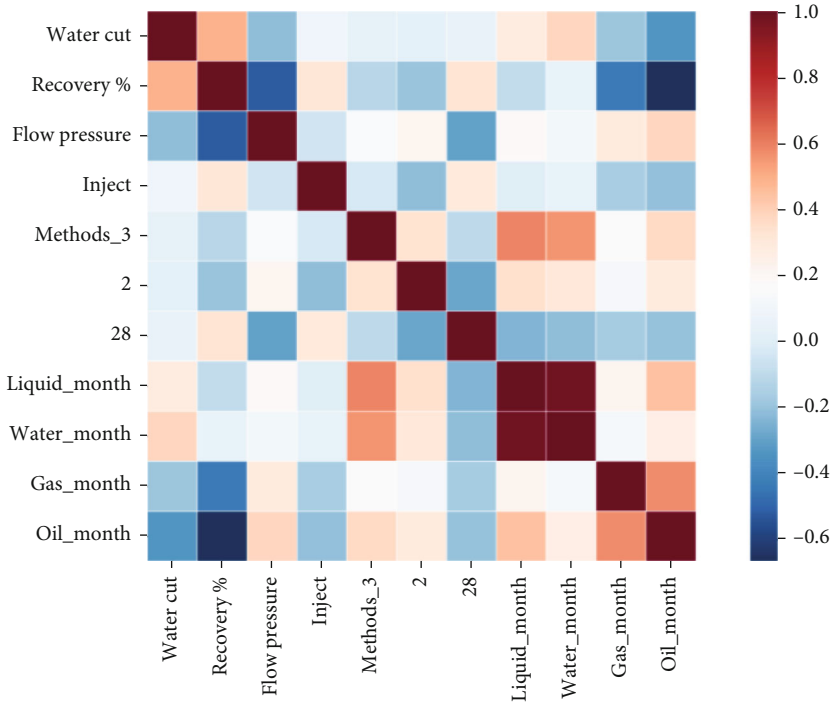


Figure 9: Heatmap of correlation coefficients between 11 features.

Table 2: Model evaluation metrics.

MAE	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{N}\lvert y_i - \widehat{y}_i\rvert$
MAPE	$\mathrm{MAPE} = \left(\frac{1}{n}\sum_{i=1}^{N}\frac{\lvert y_i - \widehat{y}_i\rvert}{y_i}\right) \times 100$
RMSE	$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\lvert y_i - \widehat{y}_i\rvert}{N}}$
R^2	$R^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2 - \sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Accuracy	$\mathrm{accurate} = 1 - \left(\frac{\sum_{i=1}^{N}((\lvert y_i - \widehat{y}_i\rvert)/y_i)}{N}\right)$

y_i: the measured value at time i; $\widehat{y}_i$: the predicted value at time i; $\bar{y}$: the mean of y_i; N: sample size.

the remaining numerical features are compressed by another PCA model (with a confidence of 95%), and two sets of compressed features are spliced to form the features numbered no. 0-no. 32 (see Figure 7) finally.

3.3.2. Feature Selection. In previous studies, the analyses of correlation with time series are generally carried out in the same time step in terms of production prediction; this correlation is the relationship between the current month's output and the current month's flowing pressure (or other features). This correlation can reflect the relationship between targets and variables and between variables and variables. However, there are still some loose places: the data used in prediction is not the data of the current month, but the data of the previous n months, which means that if we want to get a more accurate correlation coefficient, the calculation should also be carried out between the target month's output and the input data used to predict it (see Figure 8).

The study takes the input length of 12 and the output length of 12 (using the previous 12 months' data to predict the next 12 months' production) as an example to illustrate the problem.

Figure 7 shows the visualization results of Spearman correlation coefficient between the inputs and outputs. Figure 7(a) shows the average of 12 correlation coefficients between the production in the next nth month and the feature^k in the last m months:

$$\rho\left(\mathrm{output}_n, \mathrm{feature}^k\right) = \frac{\sum_{m=1}^{12}\rho\left(\mathrm{output}_n, \mathrm{fature}_m^k\right)}{12}. \quad (4)$$

Here, $\rho(\mathrm{output}_n, \mathrm{feature}^k)$ denotes the Spearman correlation coefficient between output$_n$ and feature^k; output$_n$ denotes the production in the nth month; feature^k denotes the kth feature; fature$_m^k$ denotes the feature^k in the last m months.

Figure 7(b) shows the correlation coefficient between the production in the next 1th month and the features in the last mth months. When a certain feature is fixed, from left to right, the color of the grids in Figure 7(a) gradually becomes lighter, and the color of the grids in Figure 7(b) gradually becomes darker, which means that the larger the time interval between the inputs and outputs, the lower the correlation between them, and the greater the difficulty of prediction.

Table 3: Model structure (CNN-LSTM).

Layer (type)	Hyperparameters		
	Filter	Kernel size	Dilation rate
Conv1D_1	128	1	1
Conv1D_2	241	1	1
	Unit		Activation
LSTM_1	156		Softsign
LSTM_2	136		Softsign
		Dropout rate	
Dropout		0.2	
	Unit		Activation
LSTM_3	57		None
Dense	Prediction time steps		None

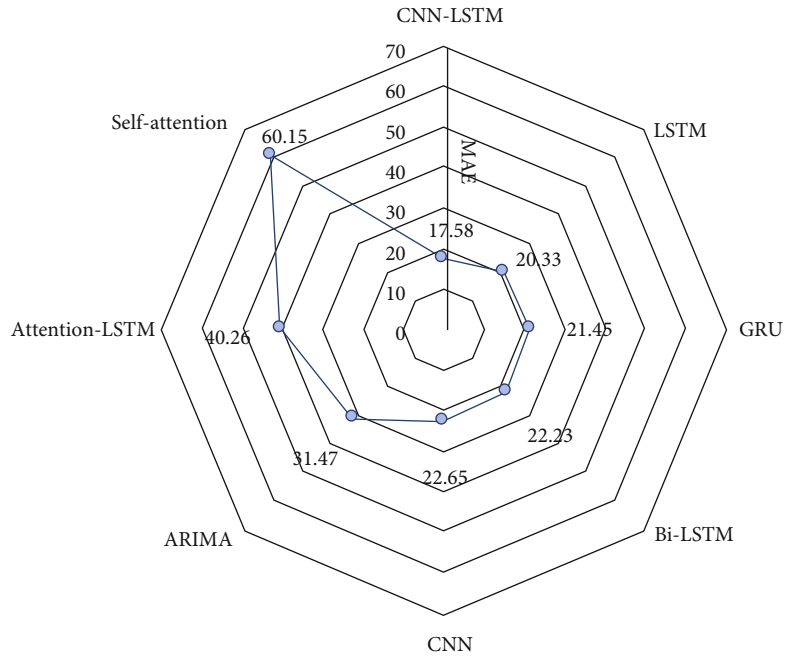


Figure 10: The predicting performance of different models.

Observing row by row, we can easily find that dynamic features have stronger correlation with future productions than static features, which can be easily explained: the correlation coefficient depends to a certain extent on the relative variation of the value of features in the sample. Compared with the production which remains changed over time, static features only vary from well to well, so the overall correlation is weak. We also note that the forecast target (production) has a strong autocorrelation—future production is far more correlated with past production than any other characteristic (up to 0.9), which means that the forecast value will be largely determined by the past production data.

To reduce the degree of difficulty in model fitting, accelerate the convergence, and eliminate invalid features, features are selected, as shown in Figure 7(c): with ±0.2 as the threshold, the features of correlation (accumulative correlation score over all outputs) ≥ 0.2 or ≤ -0.2 will be retained, and 11 features are eventually entered into the data set. The correlation between them is shown in Figure 9. We can find that the correlation between monthly water production and monthly liquid production is much greater than that between monthly oil production and monthly liquid production, which confirms that the block has entered the ultrahigh water cut stage, and most of the produced liquid is water.

3.4. Sample Generation. Before generating the input samples, we need to standardize the data, so that different features have the same scale, and they will have a fair chance to be learned by the model. As mentioned in the introduction, in order to enable the model to learn the rule of production variation at each production stage, we divided the data set by wells: 341 wells constitute the training set for model training, 43 wells constitute the validation set for hyperparameter optimization, and 42 wells constitute the test set for model

Table 4: Performance of different models (84-3).

Model	MAE	MAPE	R^2	RMSE	Accurate	Training time/min
CNN-LSTM	17.58	6.54	0.99	37.26	0.93	6.33
LSTM	20.33	7.02	0.99	44.36	0.9	5.18
GRU	21.45	7.15	0.99	47.14	0.9	3.04
Bi-LSTM	22.23	7.69	0.99	48.12	0.88	6.44
CNN	22.65	7.73	0.99	48.87	0.88	3.16
ARIMA	31.47	15.68	0.98	37.17	0.83	2.11
Attention-LSTM	40.26	18.26	0.97	60.68	0.82	8.26
Self-attention	60.15	25.36	0.96	80.67	0.66	5.01

84-3: using the data from the previous 84 months to predict oil production of the next three months.

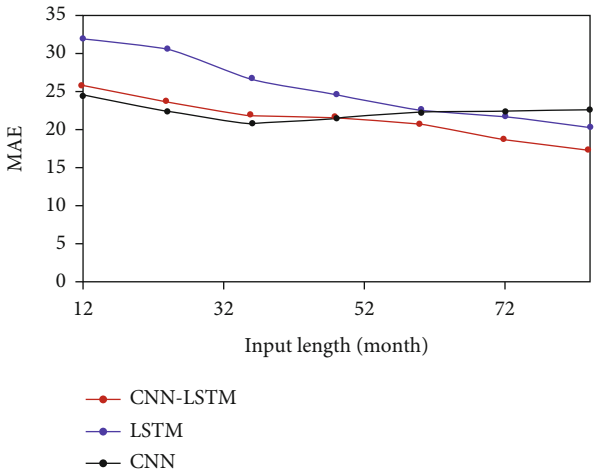


Figure 11: Different input length vs. model prediction error.

testing. The samples were manufactured by sliding time window and rearrange order to achieve better training effect.

4. Model Design and Evaluation

4.1. Evaluation Metrics. In this paper, in order to evaluate the prediction performance of the model comprehensively, five evaluation metrics are used to evaluate the model, as shown in Table 2. It is worth noting that, to reduce the uncertainty caused by the potential randomness of the model, the model evaluation results in this paper are all from the average of 3 experiments under the same setting.

4.2. Model Structure Design. After extensive investigations and comparative experiments, LSTM and CNN are selected to build a hybrid model. CNN is stacked in the first place, its excellent ability of feature extraction enables the model to extract as much hidden knowledge as possible. After the CNN layers, we stacked the LSTM layers in the hope that the model could learn the changes of timing sequence better. In addition, the layer normalization is used between the LSTM layers, which prevents possible gradient extinction and gradient explosion.

In this paper, Bayesian optimization is used to optimize the hyperparameters and the structure of the model. Since the limited space, only the structure and hyperparameter optimization results of the optimal model (CNN-LSTM) are presented here, as shown in Table 3.

4.3. Comparative Model. There were 8 models built for the optimal model selection, as shown below:

CNN-LSTM: Conv1D for feature extraction. LSTM for timing sequence capture [29, 30].

LSTM: classical LSTM model.

Bi-LSTM: the normal LSTM model can only learn the information from front to back, but cannot catch the information from back to front. Bidirectional LSTM is an improvement in this aspect. It combines forward LSTM and backward LSTM to capture forward and backward information at the same time [31, 32].

GRU: GRU is one of the varieties of LSTM and maybe the most successful one. It simplifies the three gate structures of LSTM cells into two and can achieve almost the same prediction accuracy as LSTM while greatly accelerating convergence [23].

CNN: one-dimensional convolutional neural network model for time series prediction [33].

Attention-LSTM: the LSTM model supplemented with an attention mechanism in hope that the addition of attention mechanism can help the model better capture the critical time steps which may contain the key information about production changes. This paper adopts Luong Attention mechanism and "General" score function. To focus attention on multiple time steps instead of one, we modify the weight activation function to "sigmoid" [34–36].

Self-Attention: do not use RNNs or CNNs, and the multihead self-attention mechanism is used to realize the prediction of time series. The structure of the model refers to BERT [37], but word embedding and location embedding are dropped, and the activation function adopts "sigmoid" [38].

ARIMA: one of the most common time series prediction models.

It is worth noting that the hyperparameters of all the above models are also the most combined ones obtained by Bayesian optimization.

4.4. The Training Set. The input of the neural network is selected features as well as production records in a certain

Table 5: Comparison of prediction accuracy of different models under different feature selection strategies (84-3).

Models (84-3)		MAE	MAE Max	MAE Min	MAPE/% MAPE	MAPE/% Max	MAPE/% Min	Accurate/%	R^2
A	CNN-LSTM	17.60	123.1	0.24	6.55	34.2	0.04	93.30	0.99
	CNN	22.66	125.31	0.21	7.19	38.1	0.03	88.37	0.99
	LSTM	23.55	120.13	0.19	7.32	35.1	0.03	86.11	0.99
	GRU	22.45	121.45	0.23	7.31	37.21	0.04	85.34	0.99
B	CNN-LSTM	17.58	120.03	0.23	6.54	35.4	0.03	93.23	0.99
	CNN	22.65	125.47	0.22	7.73	37.9	0.03	88.36	0.99
	LSTM	20.33	120.15	0.2	7.02	34.1	0.03	90.01	0.99
	GRU	21.45	121.63	0.24	7.15	34.2	0.04	90.24	0.99

A: use all features; B: use selected features.

duration, and the model is trained in a "supervised" fashion with future well production being the outputs. In this case, the inputs at each step of the model include historic production/features data for 82 consecutive months; the output is the production in the next 3 months.

In the training process, the model based on the encoder-decoder uses the "Teacher Forcing" hybrid training strategy [35], in which 60% of the input data of each time step in the decoder stage is real data, and the other 40% is the predicted value of the previous model output. This hybrid strategy can prevent overfitting caused by rapid convergence while ensuring the prediction accuracy of the model. Through the experiment, "*Adagrad*" was chosen as the model optimizer. The size of batch and the initial learning rate were set at 36 and 0.05, respectively. Callback function *ReduceLROnPlateau* was used to adapt the learning rate. The maximum epoch is 150; training process is controlled by the callback function *EarlyStop*. And model loss is MAE of predicted production.

5. Results and Discussion

5.1. Discussion of Different Models. Among the 8 models proposed, CNN-LSTM model achieved the best performance, and the predicting performance of other models are shown in Figure 10 and Table 4.

Results show that the prediction precision of attention mechanism model (including self-attention mechanism) is the worst. In many papers, attention mechanism improves the score of time series prediction [32, 34–36]. However, it is not the case in this paper. In fact, the production of a well in a certain month does not heavily depend on one or more certain previous months, so adding weights to different time steps is not very helpful in forecasting. Besides, the addition of attention mechanism greatly aggravates the training burden of the model, which is another possible reason for the low prediction accuracy of the model.

Bi-LSTM model also failed to reach the expected score, and its performance was slightly worse than that of classical LSTM model. Unlike the semantic recognition task such as machine translation, changing the order of oil well production series makes no significant difference in predicting production, and the use of Bi-LSTM brings more complex model structure undoubtedly, making model training more difficult.

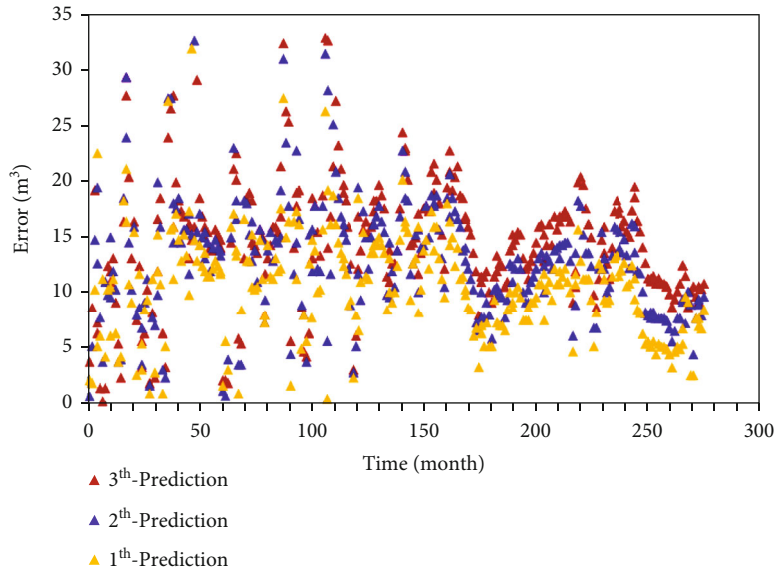
GRU and CNN models gain similar results, with LSTM model having slightly higher accuracy than LSTM, but it should be noted that the training time of GRU and CNN was almost half of LSTM. GRU or CNN models may be more appropriate in some less-desirable cases.

5.2. Discussion of Model Input Length. In this paper, CNN, LSTM, and CNN-LSTM model were used to conduct contrast experiments, respectively, and to verify the influence of different input length on model prediction error, as shown in Figure 11.

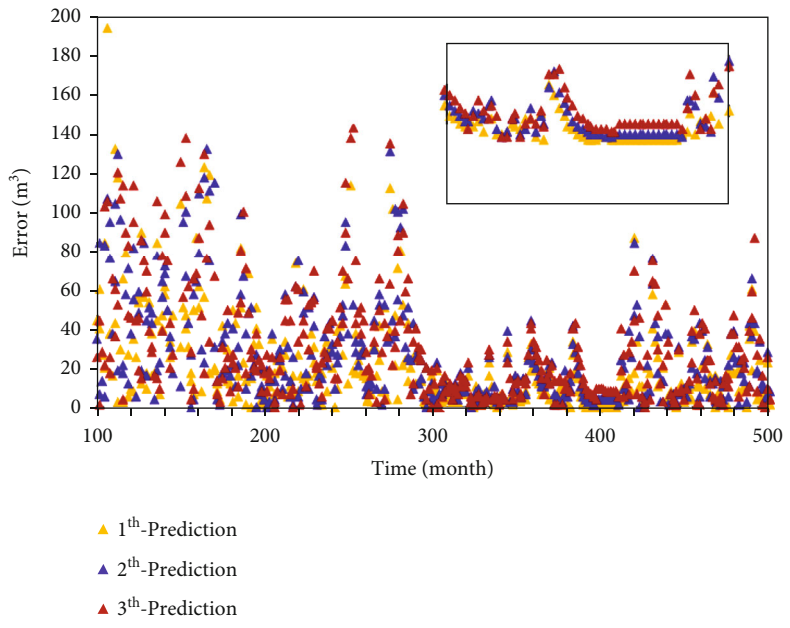
It is easy to see that the performance of CNN model is better than that of LSTM and CNN-LSTM when the input length is short. However, with the increase of the input length, the prediction error of CNN model also increases significantly, while that of LSTM and CNN-LSTM is decreasing. Compared with CNN and LSTM, CNN-LSTM model seems to inherit advantages from both, better than LSTM in short case and better than the other two in long case. Considering data characteristics, case requirements, and model performance, the following example uses 84 months' data to predict monthly well production over the next three months.

5.3. Discussion of Feature Selection. To verify the effectiveness of feature selection, two feature selection plan were used for feature selection experiment. Plan A: input all features; plan B: input select features. The prediction accuracy of each model using different features is shown in Table 5.

It is found that features and algorithms have cross influence on the prediction accuracy of the model. The RNN model has strong sensitivity to the number of features, and the selection of the number of features can reduce its prediction error. The model with convolutional layer is insensitive to the changes in the number of features. It can be seen from Table 5 that there is not much difference in model prediction accuracy before and after feature selection. The analysis shows that the RNN (LSTM/GRU) model is more inclined to capture the connection of samples in time series, while the convolution structure model is more inclined to analyze and extract high-dimensional features. Therefore, prior feature selection for the LSTM model means that part of the

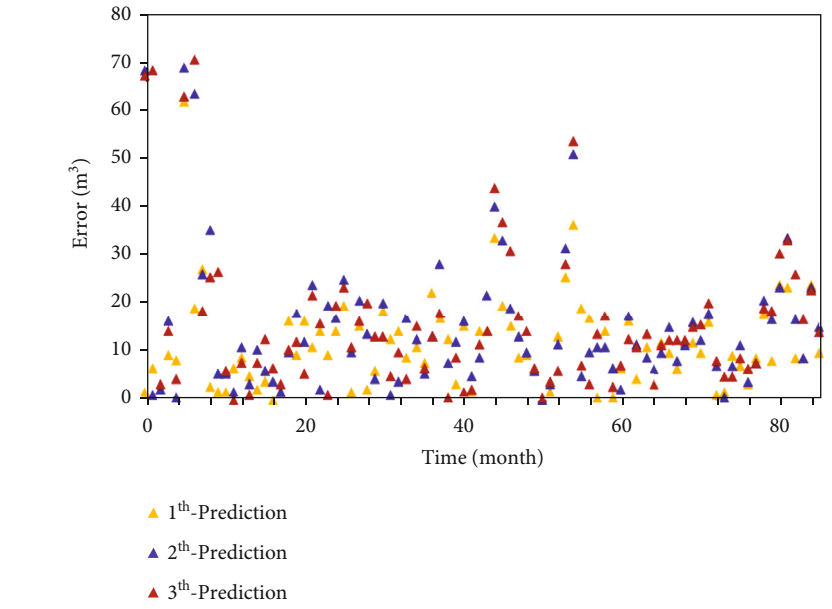


(a)

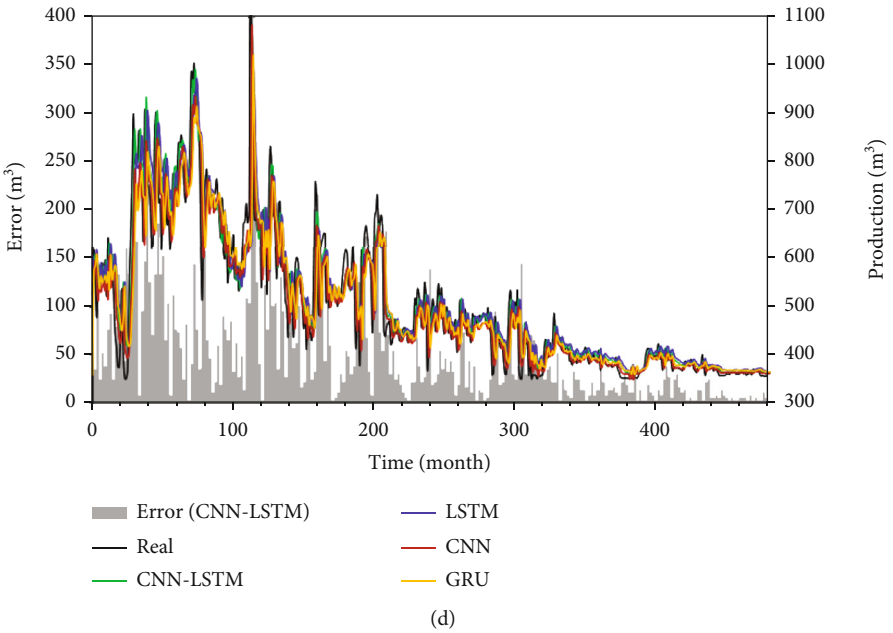


(b)

Figure 12: Continued.



(c)



(d)

Figure 12: Continued.

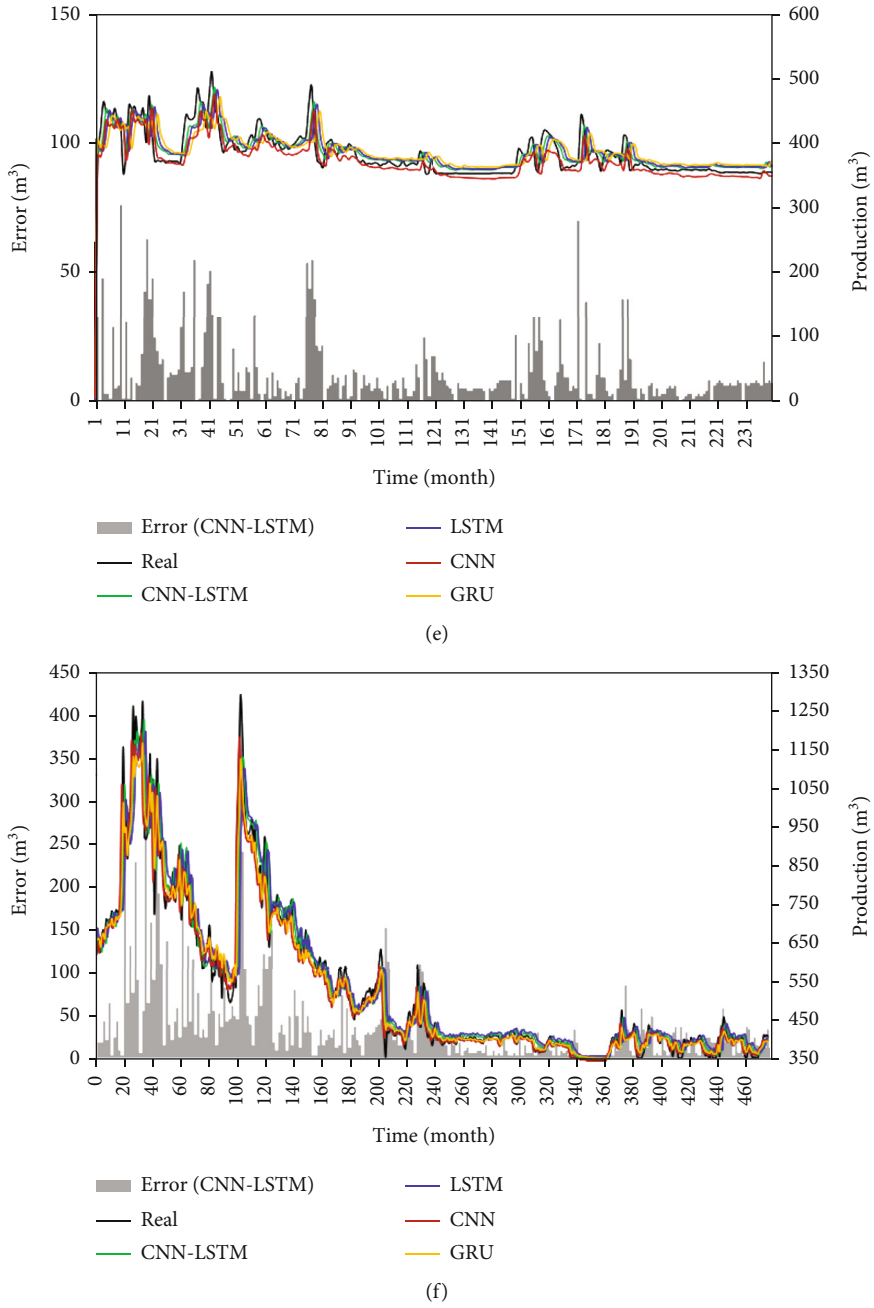


Figure 12: The prediction results of 6 wells in reservoir A. (a–c) The error distribution of the 3 predicted production of well A\B\C by CNN-LSTM model. "1th-prediction" denotes the next month's predicted production. Similarly, "2th-prediction" denotes the predicted production of the month after next. And "3th-prediction" means the predicted production of the third month in the future. (e–f) The comparison of real production curves and the 1th predicted production curves of well E/F/G by CNN-LSTM model, LSTM model, GRU model, and CNN model. The absolute error between real production and production predicted by CNN-LSTM is also shown.

feature extraction work has been completed before model training, which reduces the difficulty of LSTM training, so the LSTM model has higher prediction accuracy for data sets with fewer features. CNN itself can well realize feature extraction, and feature selection only accelerates its training speed, but has little impact on the final prediction accuracy.

5.4. Case Verification. The proposed models are used to predict 43 wells' production in the test set, and the results of 6 wells were selected randomly. The results are shown in Figure 12.

Figures 12(a)–12(c) compare the prediction error of the three outputs of the model with wells B, C, and D as examples. Obviously, the 1th outputs have the minimum error consistent with the correlation analysis.

Figures 12(d)–12(f) compared different models' prediction results of wells E, F, and G. Similar to the results shown in Table 4, CNN-LSTM model's prediction is closer to the

real tendency, but other models also can fit the curve closely. Another obvious phenomenon is that the predictions at earlier time stages are worse compared to latter ones; in other words, the error at high production is greater than the error at low production. This is because production changes are more dramatic in the early stages of the well development.

Extensive experiments show that proposed scheme significantly improves production prediction accuracy and enhances predict efficiency.

6. Conclusion

This paper proposes a machine learning model for predicting single-well production in water flooding reservoir. The specific conclusions are as follows:

(1) More than 60 factors of geology and development that affect the changes of oil well production was comprehensively considered to build the data set. Data filling and feature extraction were carried out, respectively, according to the characteristics of data. Features were analyzed and selected from the perspective of time sequence. The data set is divided by wells to make the sample distribution more practical

(2) Eight models with good performance in the time series prediction are constructed. By comparison, CNN-LSTM model gains the best score, while the improvement of attention mechanism and Bi-LSTM model is limited. It also illustrates that complex models that are doing well in other tasks may not be suited to the well production prediction

(3) The Bayesian optimization is used to optimize the hyperparameter of the models, which can greatly improve the efficiency of hyperparameter optimization and improve the prediction accuracy of the models

(4) An experiment case is carried out in reservoir A, and the results prove that the model proposed in this paper can accomplish the prediction task of single-well production successfully and provides a good reference and guidance for development and production in water flooding reservoir

Data Availability

The manuscript is a self-contained data article; the entire data used to support the findings of this study are included within the article. If any additional information is required, this is available from the corresponding author upon request to zhanglesley@foxmail.com.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank the support from the National Natural Science Foundation of China (grant no. 51974357).

References

[1] J. Huang and J. Zhang, "Overview of oil and gas production forecasting by machine learning," *Petroleum Reservoir Evaluation and Development*, vol. 11, pp. 613–620, 2021.

[2] K. Wang, B. Jiang, H. Li et al., "Rapid and accurate evaluation of reserves in different types of shale-gas wells: production-decline analysis," *International Journal of Coal Geology*, vol. 218, p. 103359, 2020.

[3] G. Luo, L. Xiao, and Y. ShiI, "Machine learning for reservoir fluid identification with logs," *Petroleum Science Bulletin*, vol. 1, pp. 24–33, 2022.

[4] G. Liao, Y. Li, and L. Xiao, "Prediction of microscopic pore structure of tight reservoirs using convolutional neural network model," *Petroleum Science Bulletin*, vol. 1, pp. 26–38, 2020.

[5] R. R. A. Mendes, A. P. Paiva, R. S. Peruchi, P. P. Balestrassi, R. C. Leme, and M. B. Silva, "Multiobjective portfolio optimization of ARMA-GARCH time series based on experimental designs," *Computers and Operations Research*, vol. 66, pp. 434–444, 2016.

[6] Y. Duana, H. Wang, and M. Wei, "Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction," *Petroleum*, vol. 8, no. 2, pp. 270–277, 2021.

[7] J. Gu, G. Sui, and Z. Li, "Oil well production forecasting method based on ARIMA-Kalman filter data mining model," *Journal of Shenzhen University Science and Engineering*, vol. 35, no. 6, pp. 575–581, 2018.

[8] Z. Hou, K. Song, and G. Ma, "Neural units with higher-order synaptic operations for robotic image processing applications," *Soft Computing*, vol. 11, no. 3, pp. 221–228, 2007.

[9] X. Song, Y. Liu, and J. Ma, "Productivity forecast based on support vector machine optimized by grey wolf optimizer," *Lithologic Reservoirs*, vol. 32, no. 2, pp. 134–140, 2020.

[10] L. Gao, C. Tan, and X. Cheng, "Oilfield output forecast based on the chaotic sequence time forecasting method," *Peak Data Science*, vol. 2, p. 5, 2016.

[11] A. D. Morozov, D. O. Popko, and V. M. Duplyakov, *Data-Driven Model for Hydraulic Fracturing Design Optimization: Focus on Building Digital Database and Production Forecast*, 2019.

[12] R. Zhang and H. Jia, "Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs," *Petroleum Exploration and Development*, vol. 48, no. 1, pp. 201–211, 2021.

[13] V. Nguyen-Le and H. Shin, "Artificial neural network prediction models for Montney shale gas production profile based on reservoir and fracture network parameters," *Energy*, vol. 244, article 123150, 2022.

[14] Y. Zhu, Y. Xian, and Q. Li, "Shale gas productivity forecast based on big data," *Well Testing*, vol. 28, no. 1, pp. 1–6, 2019.

[15] G. Hui, S. Chen, Y. He, H. Wang, and F. Gu, "Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational

factors," *Journal of Natural Gas Science and Engineering*, vol. 94, p. 104045, 2021.

[16] J. Zhang, Z. Yan, and X. Zhang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas," *Journal of Hydrology*, vol. 561, pp. 918–929, 2018.

[17] W. Liu, W. D. Liu, and J. Gu, "Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network," *Journal of Petroleum Science and Engineering*, vol. 189, article 107013, 2020.

[18] Y. Ren, J. Gu, and W. Cui, "Oilfield production prediction model based on improved fruit fly algorithm and long-short term memory neural network," *Science Technology and Engineering*, vol. 20, no. 18, pp. 7–21, 2020.

[19] Q. Zhang, Z. Chen, and Y. Zeng, "Data-driven approaches for time series prediction of daily production in the Sulige tight gas field, China," *Artificial Intelligence in Geosciences*, vol. 2, pp. 165–170, 2021.

[20] Z. Guo, J. Zhao, and Z. You, "Prediction of coalbed methane production based on deep learning," *Energy*, vol. 230, article 1208471, 2021.

[21] N. C. Chakra, K. Y. Song, and M. M. Gupta, "An innovative neural forecast of cumulative oil production from a petroleum reservoir employing higher-order neural networks (HONNs)," *Journal of Petroleum Science & Engineering*, vol. 106, no. 6, pp. 18–33, 2013.

[22] R. Huang, C. Wei, B. Wang et al., "Well performance prediction based on long short-term memory (LSTM) neural network," *Journal of Petroleum Science and Engineering*, vol. 208, article 109686, 2022.

[23] L. Zhang, H. Dou, and H. Wang, "Neural network optimized by genetic algorithm for predicting single well production in high water cut reservoir," in *2021 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field (ICMSP)*, pp. 297–306, Xi'an China, 2021.

[24] X. Song, Y. Liu, L. Xue et al., "Time-series well performance prediction based on long short-term memory (LSTM) neural network model," *Journal of Petroleum Science and Engineering*, vol. 186, article 106682, 2020.

[25] W. W. Weiss, R. S. Balch, and B. A. Stubbs, *How Artificial Intelligence Methods Can Forecast Oil Production*, no. article 75143, 2002SPE, 2002.

[26] H. Wang, L. Mu, and F. Shi, "Production prediction at ultra-high water cut stage via recurrent neural network," *Petroleum Exploration and Development*, vol. 47, no. 5, pp. 1009–1015, 2020.

[27] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, 2019.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] A. Agga, A. Abboua, and M. Labbadi, "CNN-LSTM: an efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, vol. 208, article 107908, 2022.

[30] A. Agga, A. Abbou, M. Labbadi, and Y. el Houm, "Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models," *Renewable Energy*, vol. 177, pp. 101–112, 2021.

[31] X. Huang, Q. Li, and Y. Tai, "Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM," *Energy*, vol. 246, article 1234031, 2022.

[32] N. Mughees, S. AliMohsin, and A. Mughees, "Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting," *Expert Systems with Applications*, vol. 175, article 114844, 2021.

[33] Z. Yuan, H. Huang, Y. Jiang, and J. Li, "Hybrid deep neural networks for reservoir production prediction," *Journal of Petroleum Science and Engineering*, vol. 197, article 108111, 2021.

[34] Y. Liu, L. Shan, and D. Yu, "An echo state network with attention mechanism for production prediction in reservoirs," *Journal of Petroleum Science and Engineering*, vol. 209, article 109920, 2022.

[35] M. T. Luong, H. Pha, and C. D. Manning, *Effective Approaches to Attention-Based Neural Machine Translation*, Computer ence, 2015.

[36] J. Qin, Y. Zhang, and S. Fan, "Multi-task short-term reactive and active load forecasting method based on attention-LSTM model," *International Journal of Electrical Power & Energy Systems*, vol. 135, article 108111, 2021.

[37] J. Devlin, M. W. Chang, and K. Lee, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2018.

[38] A. Vaswani, N. Shazeer, and N. Parmar, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

Hindawi
Geofluids
Volume 2022, Article ID 9974157, 14 pages
https://doi.org/10.1155/2022/9974157

Research Article

Prediction of Shear Wave Velocity Based on a Hybrid Network of Two-Dimensional Convolutional Neural Network and Gated Recurrent Unit

Tengfei Chen,[1] **Gang Gao**,[1] **Peng Wang**,[1] **Bin Zhao**,[1] **Yonggen Li**,[2] **and Zhixian Gui**[1]

[1]*Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education, Wuhan 430100, China*
[2]*Research Institute of Petroleum Exploration and Development, CNPC, Beijing 100083, China*

Correspondence should be addressed to Gang Gao; dragon_china316@163.com

Received 9 March 2022; Accepted 7 May 2022; Published 2 June 2022

Academic Editor: Zhenzhen Wang

Compressional and shear wave velocities (Vp and Vs, respectively) are important elastic parameters to predict reservoir parameters, such as lithology and hydrocarbons. Due to acquisition technologies and economy, the shear wave velocity is generally lacking. Over the last few years, some researchers proposed deep learning algorithms to predict the shear wave velocity using conventional logging data. However, these algorithms focus either on spatial feature extraction for different physical properties of rocks or on sequential feature extraction in the depth direction of rocks. Only focusing on feature extraction in a direction of rocks might lead to a decrease in prediction accuracy. Therefore, we propose a hybrid network of a two-dimensional convolutional neural network and the gated recurrent unit (2DCNN-GRU), which can establish more complex nonlinear relationships between the input and output data based on the spatial features extracted by 2DCNN and the sequential features extracted by GRU. In this study, two cases are used to validate the reliability and prediction accuracy of the proposed network. Comparing the prediction results of 2DCNN, GRU, and the proposed network, the proposed network shows better performance. Meanwhile, for improving the prediction accuracy of the proposed network, the relationship is analyzed between the prediction accuracy of the proposed network and the length of the input sample.

1. Introduction

Compressional and shear wave velocity (Vp and Vs, respectively) are very important parameters in hydrocarbon fields for characterizing and evaluating reservoir, identification of the pore types, and estimation of the dynamic properties of rocks [1–4]. Due to various reasons, shear wave velocity is generally lacking. Therefore, it is necessary to study a shear wave prediction method with high prediction accuracy and strong generalization ability to improve the reservoir prediction accuracy.

Currently, empirical regression methods, rock physics methods, and machine learning methods are the main methods for shear wave velocity prediction. Since empirical methods are the fastest and easiest to apply, linear or nonlinear empirical relationships between compression and shear wave velocities have been proposed by various researchers [5–12]; however, they are constrained by site-specific and the rock type.

A variety of methods for predicting the shear wave velocity on the basis of rock physics have been proposed. These rock physics models focused on the modeling of the modulus of the rock matrix, dry rock, and saturated rock of the equivalent medium. In particular, for the modeling of the dry rock modulus of an equivalent medium, complex pore shapes were the research focus. Jørstad et al. [13] used both DEM and self-consistent approximation (SCA) for the shear wave velocity prediction in sandstones and concluded that the effective-medium theories were more accurate by comparing the results with those predicted from empirical regression methods. Xu and White [14] proposed a hybrid

approach to predict the shear wave velocity based on a shaly sandstone formation using a combination of the Kuster and Toksöz (KT) model [15] and the differential effective medium (DEM) model [16]. Based on the widely used Xu–White model and Gassmann's equations [17], Bai et al. [18] analyzed the influence of errors of input parameters of rock matrix, fluid inclusions, porosity, and aspect ratio (AR) on the prediction accuracy of shear wave velocity. Bai et al. [19] illustrated that a variable aspect ratio method in the Xu–White model was significantly improved. Liu et al. [20] proposed a differential Kuster–Toksöz (DKT) model to predict shear wave velocity and focused on the process in which the porosity with certain geometric shapes is gradually increased from zero to its final value to overcome a diluted concentrated pore of the KT model. Yang et al. [21] developed a revised Xu-White model and improved the estimated shear wave velocity for a calciferous sandy shale formation by considering the effect of the volume fraction of limestone. In the past ten years, with the development of unconventional oil and gas, rock physics models of complex reservoirs have been developed rapidly. Xu and Payne [22] extended the Xu-White model, originally designed for clastic rocks, to carbonate rocks and proposed a carbonate rock physics model with complex pore types. Zhang [23] established an anisotropic rock physics model to predict shear wave velocity, which was suitable for rocks with high-angle fractures. Based on the dual pore theory, an anisotropic rock physics model of tight oil sandstone was proposed, and the influence of clay content and type and pore connectivity and type on it was systematically studied by Huang et al. [24]. Assuming that shale is a laterally isotropic medium, Gui et al. [25] proposed a shear wave velocity prediction method that considered the microscopic characteristics of the rock. Liu et al. [26] proposed a method for predicting shear wave velocity suitable for organic-rich rocks. The accuracy of these methods for predicting shear wave velocity depended on the accurate calculation of reservoir geophysical parameters such as porosity, pore type, pore shape, mineral composition, and water saturation. However, these high-precision parameters are difficult to obtain, which increases complexity and indeterminacy of rock physics.

With the rapid development of software and hardware technology, some researchers use machine learning algorithms to predict shear wave velocities using logging data [27–33]. Deep learning developed from artificial neural network algorithms is a research hotspot in academic and industrial circles. Comparing with traditional shallow learning, deep learning improves the accuracy of prediction or classification by constructing many hidden layer machine models with complex function approximation and layer-by-layer feature transformation. The convolutional neural network (CNN) with spatial feature capture has achieved good results in different geophysical fields including interpretation of reservoir parameters from logging data [34, 35], seismic interpretation [36–38], and seismic inversion [39–42]. Based on the characteristics of logging data based on long-term dependencies, a long-short-term memory (LSTM) network was proposed to predict the shear wave velocity and its application in the identification of geophysical parameters of complex reservoirs [43–46]. Comparing with LSTM which required a long training time, the gated recurrent unit (GRU) has the characteristics of faster speed and basically unchanged accuracy by simplifying the internal structure of LSTM [47]. Sun and Liu [48] proposed a GRU-based shear wave velocity prediction method. The above applications show that deep learning models have been successfully applied in the field of geophysics and are in rapid development.

Predicting shear wave velocity is essentially a typical regression problem in data processing. Compared with the empirical and rock physics methods, deep learning is better at handling regression problems by building many hidden-layer machine models with complex function approximations and layer-by-layer feature transformations. In order to fully mine the sequential features in the depth direction of rocks and spatial features of different physical properties of rocks, a hybrid network of two-dimensional convolutional neural network and gated recurrent unit (2DCCN-GRU) was constructed to predict the shear wave velocity using conventional logging data. This network takes full advantage of the powerful spatial features extracted by 2DCNN and the sequential features extracted by GRU. The process of predicting Vs using the 2DCNN-GRU hybrid network included data normalization, generating sample datasets, and constructing the 2DCNN-GRU hybrid network and its training and prediction. The Vs prediction of the two cases confirmed that the 2DCNN-GRU hybrid network was an accurate and reliable method of Vs prediction.

2. Methodology

2.1. Convolutional Neural Network (CNN). CNN, which is a feed-forward artificial neural network, is widely used in the field of vision and image. With the rapid development of deep learning, it has been proven to successfully solve various geological problems, such as fault recognition, reservoir prediction, lithofacies classification, and geological parameter inversion [39, 40]. CNN typically consists of the convolutional layers, the pooling layers, and the fully connected layers (Figure 1). In the convolutional layers, the data of the input layer is convolved with the convolutional kernels of the convolutional layers, which can mine the local features between the data. Its weight sharing feature greatly reduces the complexity of the network. The nonlinear relationship of the data is added through the activation function, usually the rectified linear function ("ReLU") to avoid overfitting. When the data is passed into the convolution layer, the output features can be expressed as

$$y_j^l = \sigma\left(\sum\left(a_i^{l-1} w_{ij}^l\right) + b_j^l\right), \tag{1}$$

where y_j^l represents the j-th feature map of the l-th layer, a_i^{l-1} represents the i-th feature map of the previous layer, w_{ij}^l denotes the weight matrix of the l-th layer, b_j^l represents the corresponding bias term, and σ represents the activation function.

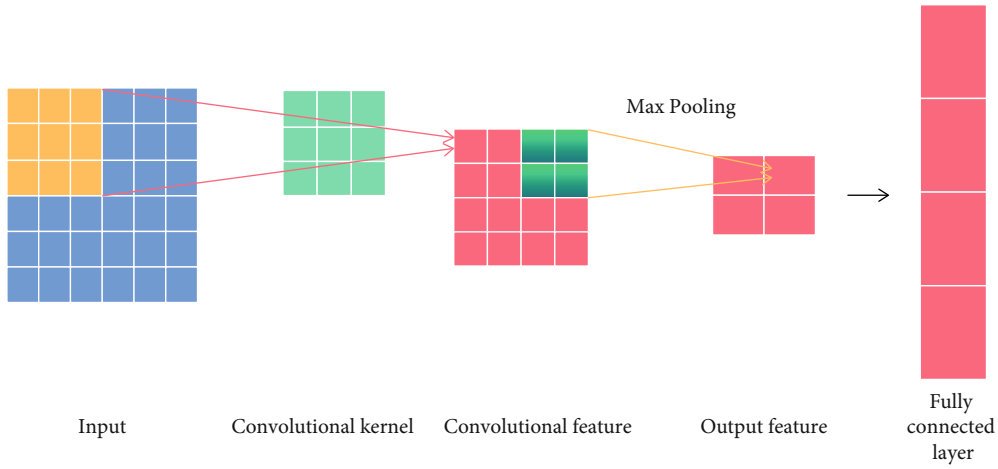


Figure 1: The structure of CNN and its unrolled network.

Considering the intrinsic relationship between shear wave velocity and various logging data, inspired by the extremely high feature extraction ability of a 2D convolutional neural network, the 2D convolution is used to extract more high-dimensional information and preserve topology as well as type and depth of log data.

2.2. Gated Recurrent Unit (GRU). A recurrent neural network (RNN) [49] is very effective for mining data with sequence characteristics. The hidden unit of the RNN with long-term sequence storage contains a loop that can combine the output at the current moment with the input at the next moment as the input at the next moment. Therefore, the RNN is particularly suitable for processing logging data that varies with sedimentary facies in the depth direction. However, due to the relatively simple structure of RNN, problems such as gradient disappearance or explosion are prone to occur in practical applications [50], and it can only hold memory functions for short-term data. In response to the above problems, the RNN variants LSTM [51] and GRU [52] were proposed. LSTM sets three-unit gates (forget gate, input gate, and output gate) to update the input data and obtain the ability of long-term memory data. However, the hidden unit of LSTM has many parameters, a complex structure, and a long training time. Compared with the LSTM network, the reset gate and update gate of GRU can reduce the network training parameters, shorten the training time, and improve the generalization ability of the network under the premise of ensuring the prediction accuracy [53] (Figure 2).

The structure of GRU combined the reset gate (r_t), the update gate (z_t), the output of the hidden state at $t-1$ (h_{t-1}), the output of the hidden state at t (h_t), and the input at t (x_t); they can be expressed as

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r), \tag{2}$$

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z), \tag{3}$$

$$\tilde{h}_t = \tanh\left(W_{\tilde{h}}[h_{t-1} \circ r_t, x_t] + b_{\tilde{h}}\right), \tag{4}$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t, \tag{5}$$

where $W_r, W_z, W_{\tilde{h}}$ and $b_r, b_z, b_{\tilde{h}}$ are the weights and biases, respectively, which are learned, "σ" is the logistic function sigmoid, $\tilde{h}_t$ is the new hidden state at t, "$\circ$" represents the dot product, and "[ ]" represents that two vectors are connected. The reset gate controls how much information from the previous state is retained. On the other hand, the update gate is contrary to the function of the reset gate [52].

2.3. Building a Hybrid Network of 2DCNN-GRU. The shear wave velocity changing with time has a certain periodicity and has a nonlinear relationship with various factors such as density, porosity, Vp, and resistivity. Therefore, a 2DCNN-GRU hybrid network was proposed in this study to solve the problem of lack of shear wave velocity. The structure of 2DCNN uses the convolution kernels to fully excavate the high-dimensional features of different logging data, while the series data of time and depth cannot be accurately excavated. The structure of GRU has a strong ability to capture features in sequence data, while it is easy to introduce noise and lose some features during the calculation process [54], which is difficult to express the spatial features of the data and ultimately leads to deviations in the prediction results. To make up for the shortcomings of a separate network, the 2DCNN and the GRU are combined to make full use of the spatial convolution characteristics of 2DCNN and the sensitivity of GRU to sequence data to establish a nonlinear relationship between input and output. The structure of the 2DCNN-GRU hybrid network (Figure 3) and the flow chart of the shear wave velocity prediction (Figure 4) are as follows.

It can be seen clearly from Figure 3 that the first part of the 2DCNN-GRU is the CNN, which convolves with the input logging data through the convolution kernels to obtain the spatial characteristics and uses padding to fill it which can keep the size of the input sample unchanged after convolution. The second part of 2DCNN-GRU is the

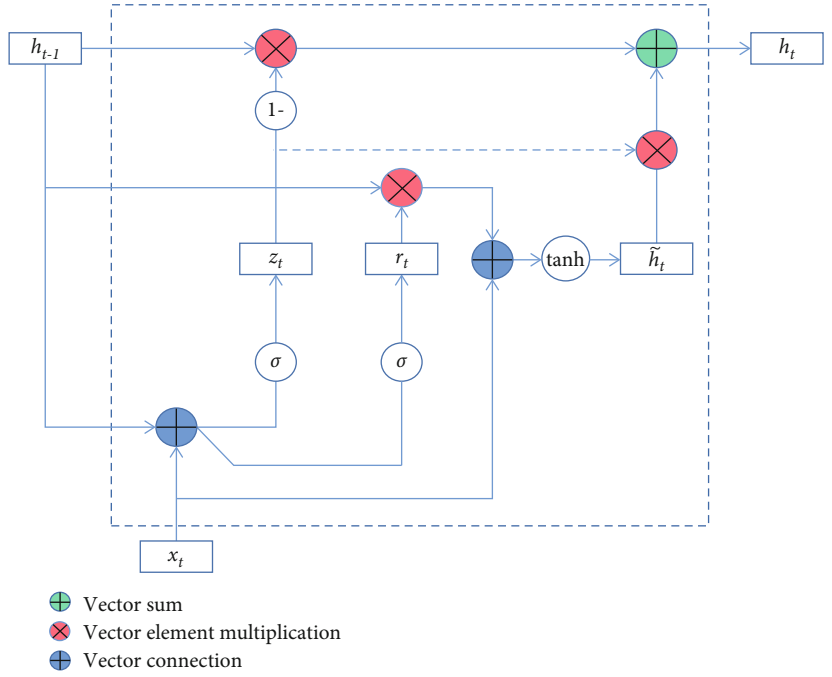


Figure 2: The internal structure of GRU.

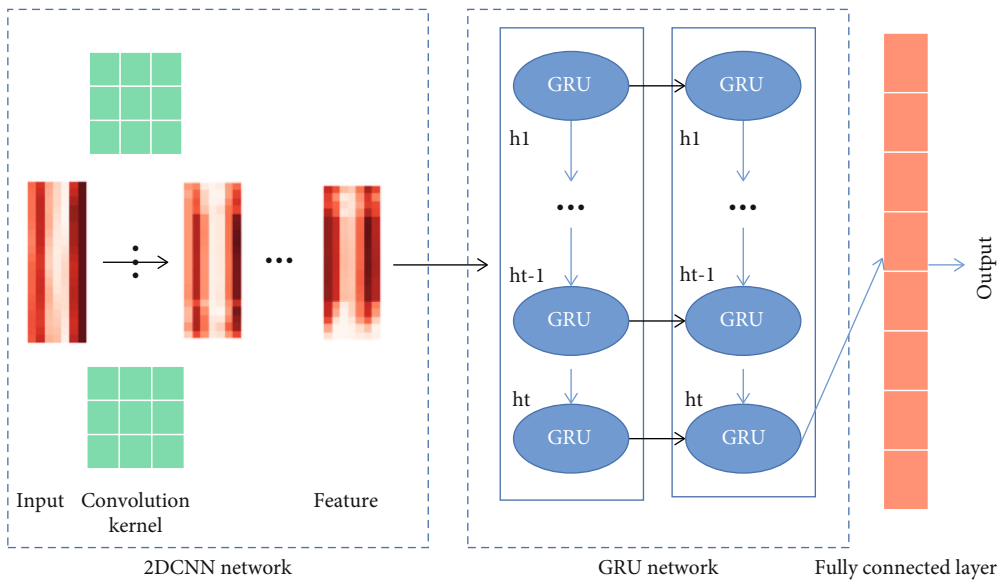


Figure 3: The structure of the 2DCNN-GRU hybrid network. A 2DCNN-GRU network is mainly composed of an input layer, a CNN Network, a GRU Network, and a fully connected layer.

GRU, which uses the spatial features extracted by the first part as the input of this layer. In particular, the first layer of GRU adopts the method of returning intermediate values. Both of the networks use activation functions to increase the nonlinearity of the network and use dropout to prevent overfitting and increase the generalization ability of the network. Finally, these features are taken into the fully connected layer to obtain the prediction of the shear wave velocity.

3. Prediction Framework Based on 2DCNN-GRU

Figure 4 shows the shear wave velocity prediction framework based on the 2DCNN-GRU hybrid network, and the specific process includes the following 4 parts.

3.1. Feature Selection. Deep learning networks are often used to deal with classification and regression problems.

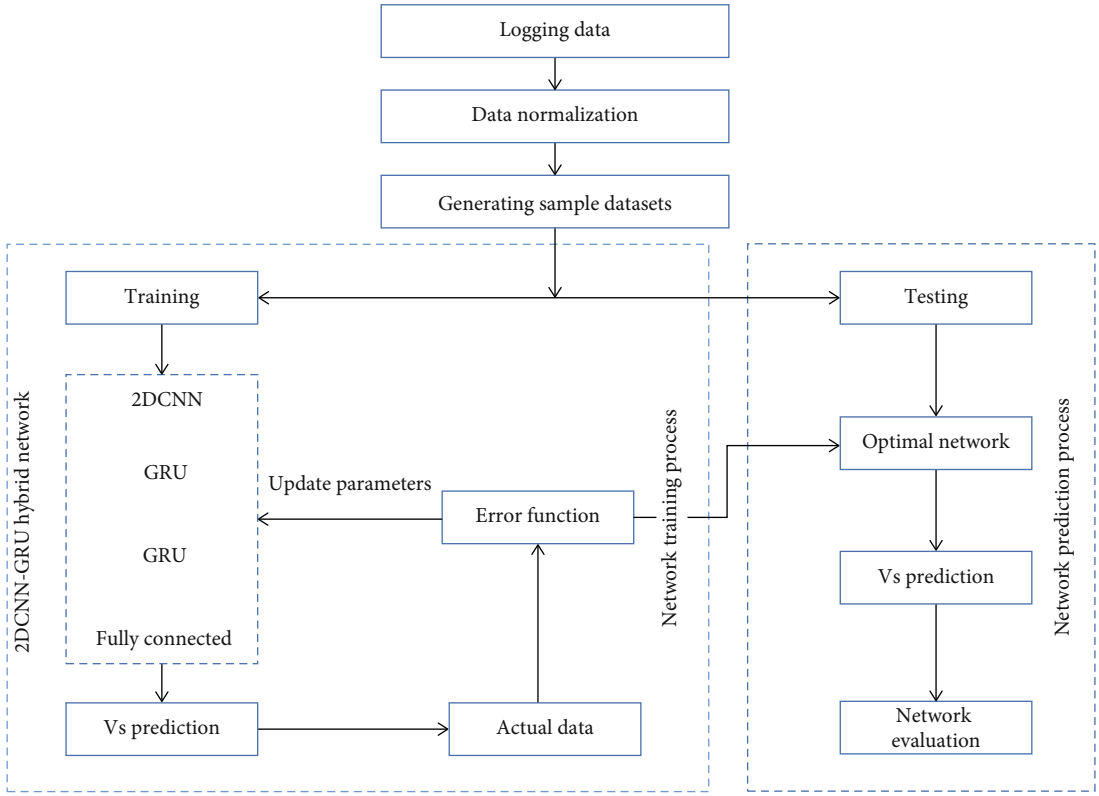


Figure 4: The VS prediction framework of the 2DCNN-GRU hybrid network.

Predicting shear wave velocity using conventional logging data is a typical regression problem. The assumptions that deal with regression problems often require correlations between input and output data. The correlation coefficient between logging data and shear wave velocity is shown in the cross-plot (Figure 5). The correlation coefficients between shear wave velocity (Vs) and compression wave velocity (Vp), neutron porosity (CNL), gamma (Gr), shale volume (Sh), density (RHOB), and water saturation (Sw) are, respectively, 0.791, 0.576, 0.324, 0.300, 0.004, and 0.003. In these selected logs, the correlation between density, Sw, and shear wave velocity is small. In addition, the correlations between other logging data and shear wave velocity are all above 0.3, which satisfies the assumption that deep learning deals with regression problems.

3.2. Data Normalization. Since there are different degrees of differences between different logging data, it is necessary to normalize the logging data to speed up the training process, which can reduce the impact on the network accuracy [55]. The logging data have mapped the range of [0, 1] with the MinMaxScaler normalization method. The normalization formula can be expressed as

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{6}$$

where $X_{\min}$ and $X_{\max}$ are the minimum and maximum of a sequence X, respectively, and Y represents the result of normalization.

3.3. Generating Sample Datasets. The recurrent neural network has various network structures in dealing with time series problems, such as one-to-one, one-to-many, many-to-one, and many-to-many. Due to the depositional law of the subsurface in the depth direction, a many-to-one structure is adopted in the process of the prediction framework based on 2DCNN-GRU (Figure 6).

3.4. Network Training and Evaluation. To speed up the network training, the loss function mean square error (MSE) was used to calculate the gap between the predicted values and the true values in this study; at the same time, the Adaptive Moment Estimation [56] was used to back-propagate to update the weight parameters. The prediction performance of the network was evaluated by mean absolute error (MAE) and correlation coefficient (R^2) in this study, which can be expressed as

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \tilde{y}_{ii})|, \tag{7}$$

$$R^2 = \frac{\sum_{i=1}^{m}(\tilde{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2}, \tag{8}$$

where m represents the number of samples, y_i represents the real value, $\tilde{y}_i$ represents the predicted value, and $\bar{y}_i$ represents the mean of samples.

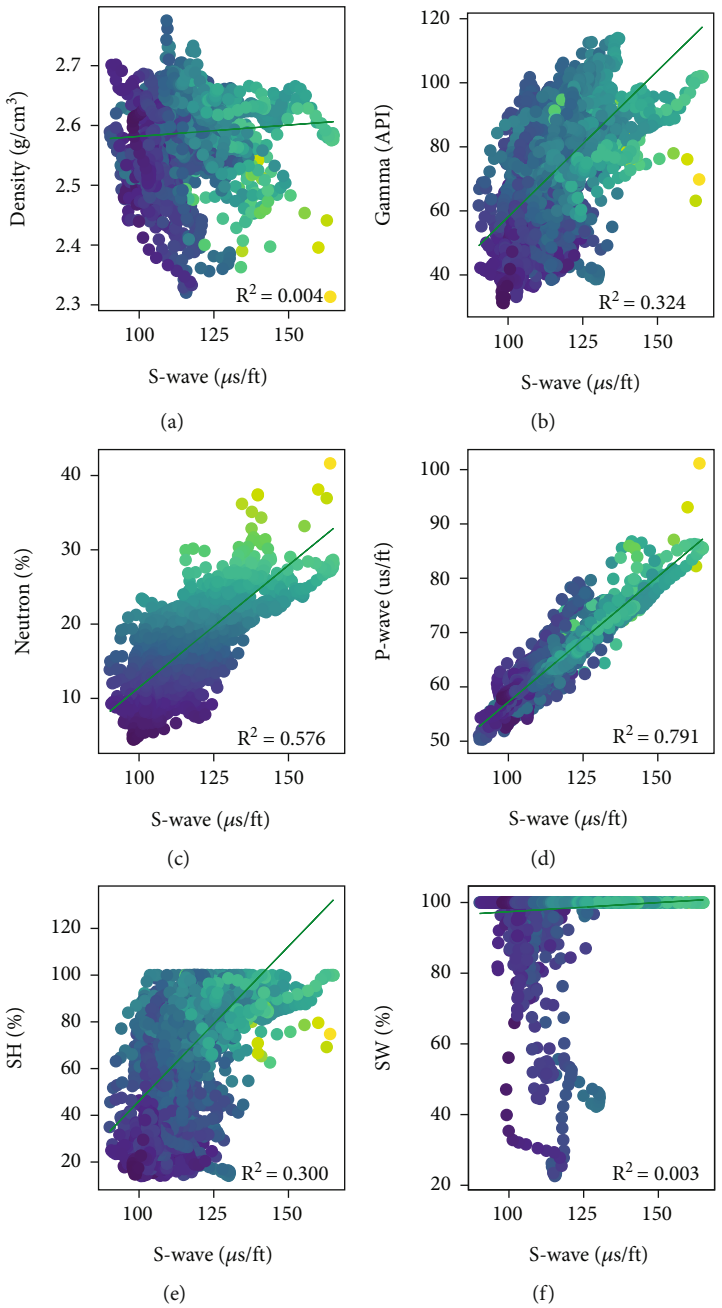


Figure 5: The cross-plot of conventional logging data and shear wave velocity. (a) RHOB–Vs; (b) Gr–Vs; (c) CNL–Vs; (d) Vp–Vs; (e) Sh–Vs; (f) Sw–Vs. *R* is the correlation coefficient between the two variables.

4. Testing and Analysis

The logging data used in our study were derived from the Tarim Basin (Figure 7). The target layer is buried at a depth of about 5500 m, mainly composed of medium-fine sandstone, and the reservoir porosity is less than 10%, which are typical characteristics of deep tight sandstone. In order to verify the prediction accuracy of the 2DCNN-GRU hybrid network proposed and optimize its parameters, the network is trained with the logging data from 8 wells in a certain area, and tested with another 2 wells to verify its accuracy and generalization, and two cases are adopted using the optimization algorithm Adaptive Moment Estimation (Adam), the loss function Mean Squared Error (MSE), and Dropout to avoid overfitting of the network. In case I, the results of the 2DCNN, GRU, and 2DCNN-GRU hybrid network were analyzed and compared to verify the prediction accuracy of the 2DCNN-GRU hybrid network. In case II, the influence of sample length on the prediction accuracy of the 2DCNN-GRU hybrid network was analyzed.

4.1. Case I. Predicting shear wave velocity based on deep learning is essentially a sequence prediction problem. Fully considering the spatial and sequential features of the logging

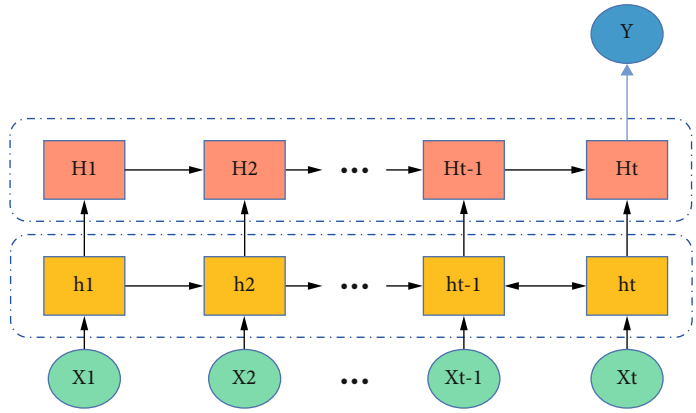


Figure 6: The structure of recurrent neural network (many-to-one).

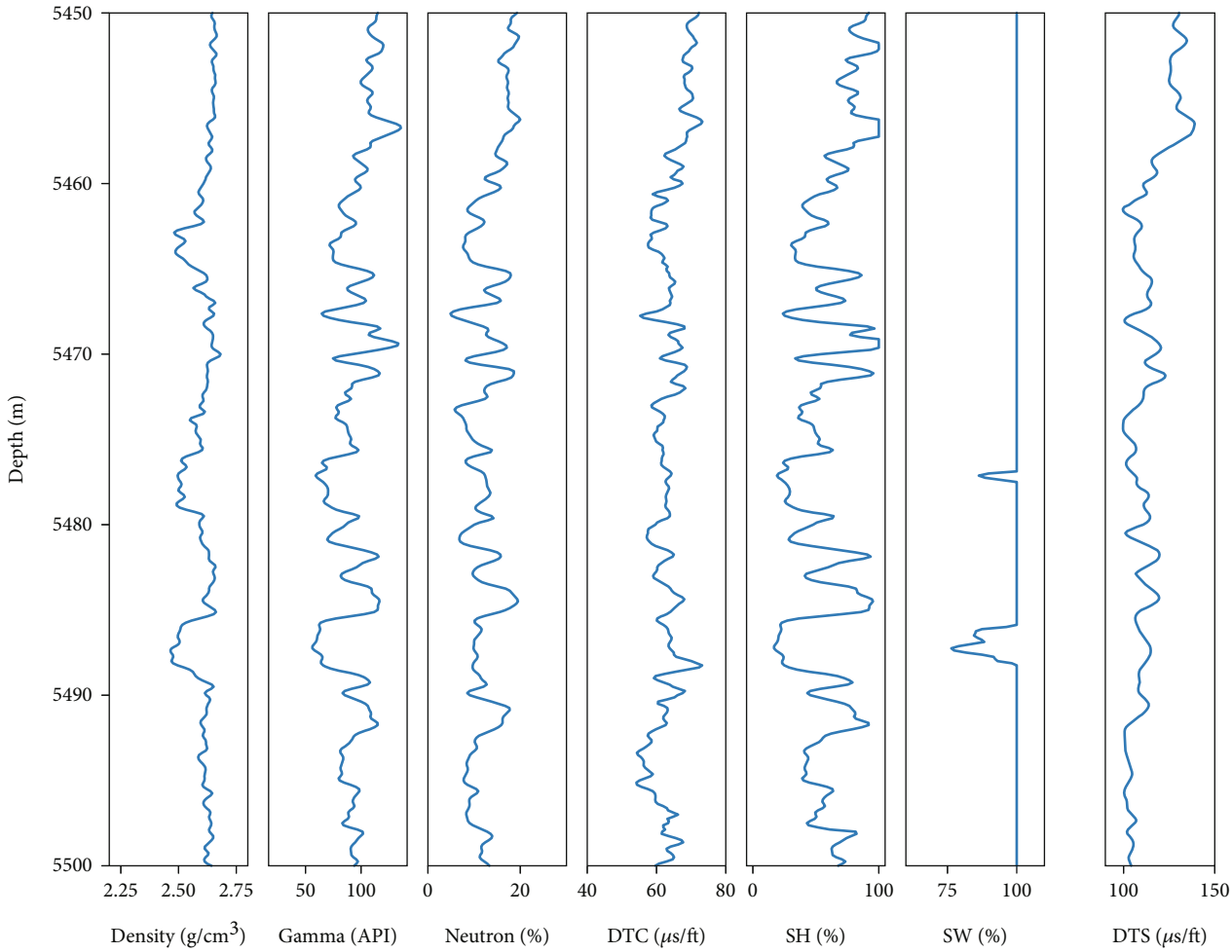


Figure 7: The visualization of logging data.

data, the 2DCNN-GRU hybrid network was established to predict the shear wave velocity, and its results were compared with those of the separate 2DCNN and GRU. The structures of the 2DCNN-GRU, 2CNN, and GRU networks are listed in Table 1.

Figure 8 shows the loss errors of 2DCNN-GRU, 2DCNN, and GRU networks. After a period of training, the loss values of all networks reach the minimum value and remain the same. It can be seen that the 2DCNN-GRU hybrid network has the lowest loss error in that the

Table 1: The structures of the 2DCNN-GRU, GRU, and 2DCNN.

Network name	Architecture	Kernel size of 2DConv layers	Number of conv filters and GRU units	Trainable parameters	Other parameters
2DCNN-GRU	Conv+GRU+GRU +dense	$(15 * 3)$	16 + 32 + 64	32097	Padding = "same" Return_ sequences = true
2DCNN	Conv+conv+dense	$(15 * 3) + (15 * 3)$	16 + 32	30529	Padding = "same"
GRU	GRU+GRU+dense	0	32 + 64	22721	Return_ sequences = true

Conv: 2D convolution+ReLU activation.

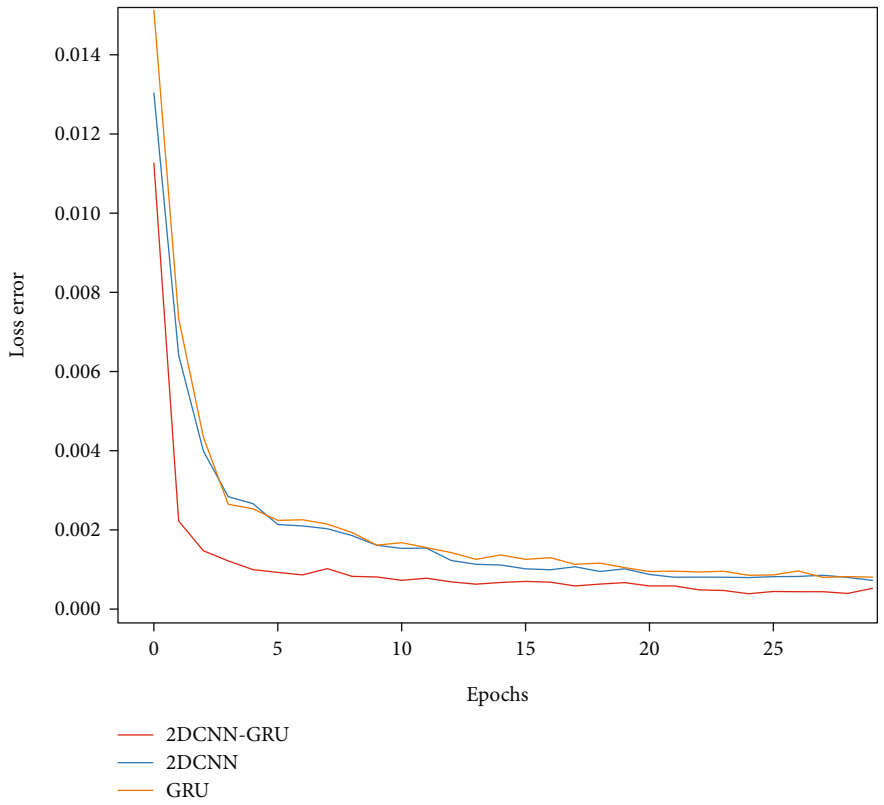


Figure 8: Loss error curves of 2DCNN-GRU, 2DCNN, and GRU networks.

shear wave velocity prediction values are closer to the true values than the other two networks. The logging data are convolved with the convolution kernels to extract the high-dimensional spatial features of the logging data. However, the logging data has time-series features in the depth direction, so the extracted spatial features are inputted into the GRU for time-series feature extraction, which can combine the spatial and time-series features of the logging data to predict shear wave velocity.

To compare the difference between the prediction performances of the three networks, the 2DCNN-GRU, 2DCNN, and GRU perform shear wave velocity prediction on the same test set and the experimental results are shown in Figure 9. Although the predicted values of the three networks are generally similar in trend to the true values, the predicted values of the 2DCNN-GRU hybrid network are closer to the true values than those of the other two networks at 5570-5590 m.

As can be seen from Figure 10, the prediction of a single 2DCNN or GRU at this stage is always slightly higher than the true values, but the prediction effect of the 2DCNN-GRU that integrates spatiotemporal features has been greatly improved. That is to say, combining with the spatiotemporal features of the logging data can better predict shear wave velocity.

In order to analyze the prediction results of the three networks more precisely, the mean absolute error (MAE) and correlation coefficient (R^2) were used to quantitatively evaluate the prediction accuracy of the three networks

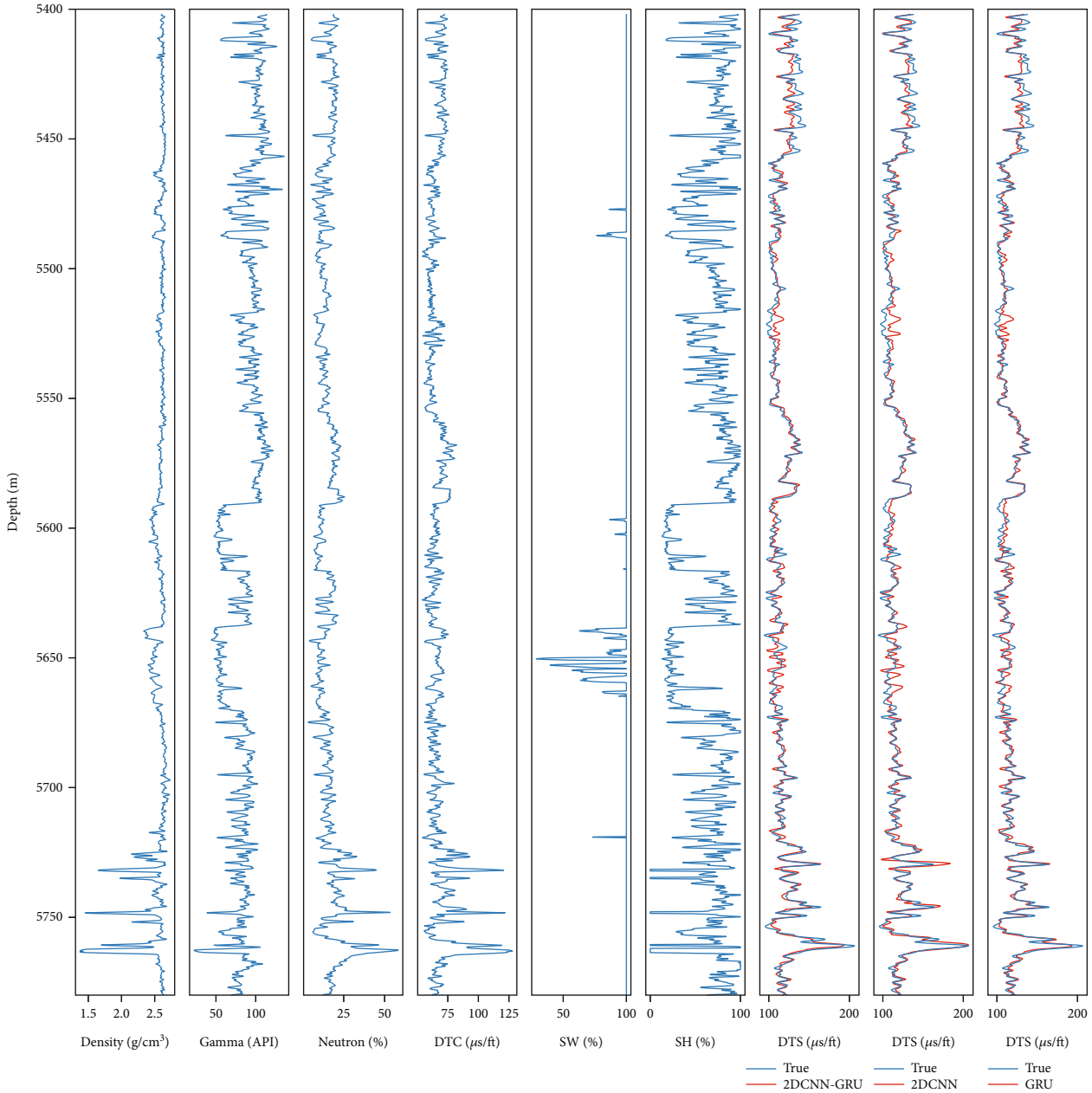


Figure 9: Prediction results of three networks in case I. The last three subgraphs indicate the final prediction results. The blue line represents the true values and the red line represents different prediction results based on three networks. The left one is the prediction result from 2DCNN-GRU, the middle one is that from 2DCNN, and the right one is that from GRU. To analyze the prediction performance of the three networks more clearly, the differences between the prediction of the three networks were compared in the form of local amplification (Figure 10).

(Figure 11). The correlation coefficient between the data predicted by the 2DCNN-GRU hybrid network and the true values was higher than that by the other two networks and was as high as 0.866. Moreover, the MAE of 2DCNN-GRU was lower than that of 2DCNN and GRU and was as low as 0.0165. Compared with the 2DCNN-GRU hybrid network, the reason for the low prediction accuracy of 2DCNN and GRU is that both of them only predict shear waves from single spatial or temporal features of logging data. Therefore, the 2DCNN-GRU hybrid network that comprehensively considers spatiotemporal features improves the accuracy of shear wave velocity prediction.

4.2. Case II. Due to the depositional law of the subsurface in the vertical direction, there is a certain correlation between the sequence sampling points, which indicates that the length of the input sample plays an important role in the prediction of shear wave velocity by deep learning. In order to analyze

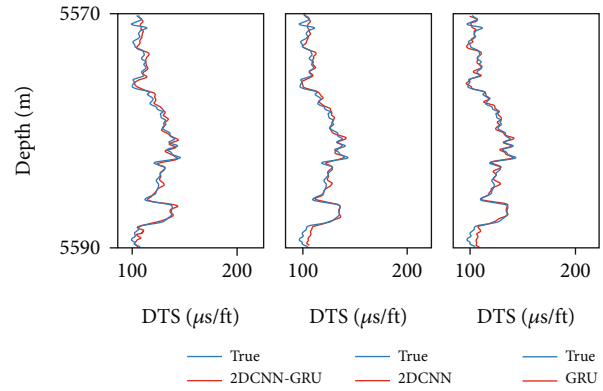


Figure 10: The local amplification comparison of 2DCNN-GRU, 2DCNN, and GRU.

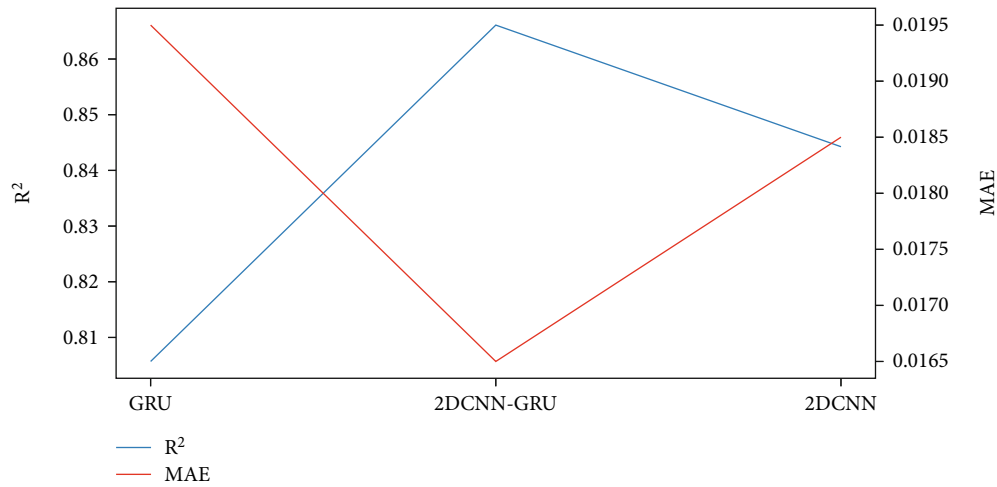


Figure 11: The evaluation metrics of 2DCNN-GRU, 2DCNN, and GRU networks.

Table 2: The structures of the 2DCNN-GRU hybrid network in 9 experiments.

Network name	Architecture	The length of the convolution kernel and sample	Number of conv filters and GRU units	Trainable parameters	Other parameters
2DCNN-GRU	Conv+GRU +GRU	3	32 + 32 + 64	40897	Padding = "same" Return_ sequences = true
		10	32 + 32 + 64	41569	
		15	32 + 32 + 64	42049	
		20	32 + 32 + 64	42529	
		25	32 + 32 + 64	43009	
		30	32 + 32 + 64	43489	
		35	32 + 32 + 64	43969	
		40	32 + 32 + 64	44449	
		65	32 + 32 + 64	46849	

Conv: 2D convolution+ReLU activation+dropout.

the optimal sample length, the vertical length of the convolution kernel in the CNN was consistent with the input sample length, and the horizontal length was set to 3. A total of 9 experiments were performed with sample lengths set to 3, 10, 15, 20, 25, 30, 35, 40, and 65. The structure of the 2DCNN-GRU hybrid network is shown in Table 2.

Prediction results of the 2DCNN-GRU hybrid network are shown in Figure 12. It can be seen that under different

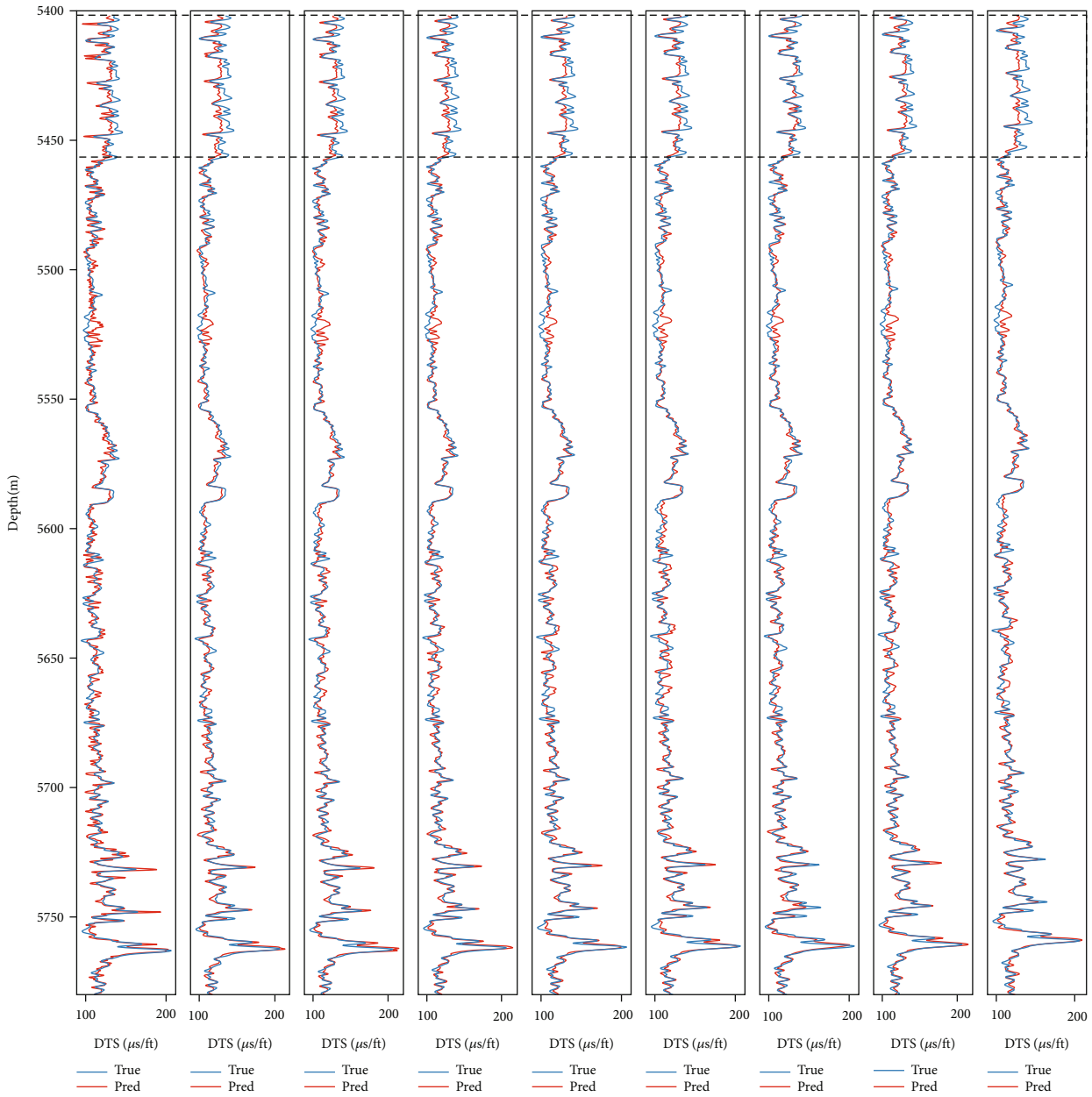


Figure 12: Nine experimental results based on the 2DCNN-GRU hybrid networks. The vertical lengths of the convolution kernels from left to right were set to 3, 10, 15, 20, 25, 30, 35, 40, and 65, respectively.

sample lengths, the prediction effects of 2DCN-GRU are different. With the continuous increase of the sample length, the prediction effect of the network becomes better first and then worse. When the input sample length is 35, the prediction effect of the network is the best. Compared with other sample lengths, the predicted values in the dotted box are the closest to the true values, but with the continuous increase of the sample length, the prediction effect of the network changes gets worse. This is because the convolution kernels of 2DCNN can only extract local features but cannot obtain global information as the sample length increases; at the same time, the GRU cannot effectively associate the input at the current moment with the historical data, which can make the prediction effect much worse than before.

Moreover, all experimental results had been evaluated by the correlation coefficient (R^2) and mean absolute error (MAE) (Figure 13). It can be seen clearly that the correlation coefficient (R^2) first increased and then decreased with the input sample length, while the trend of MAE was just the opposite during the whole testing process. When the length of the input sample reached 35, the correlation coefficient (R^2) reached the highest value at 0.877; at the same time, the MAE reached the lowest value at 0.0160, which indicates that the predicted values were closest to the true values. At the same time, the correlation between the predicted values

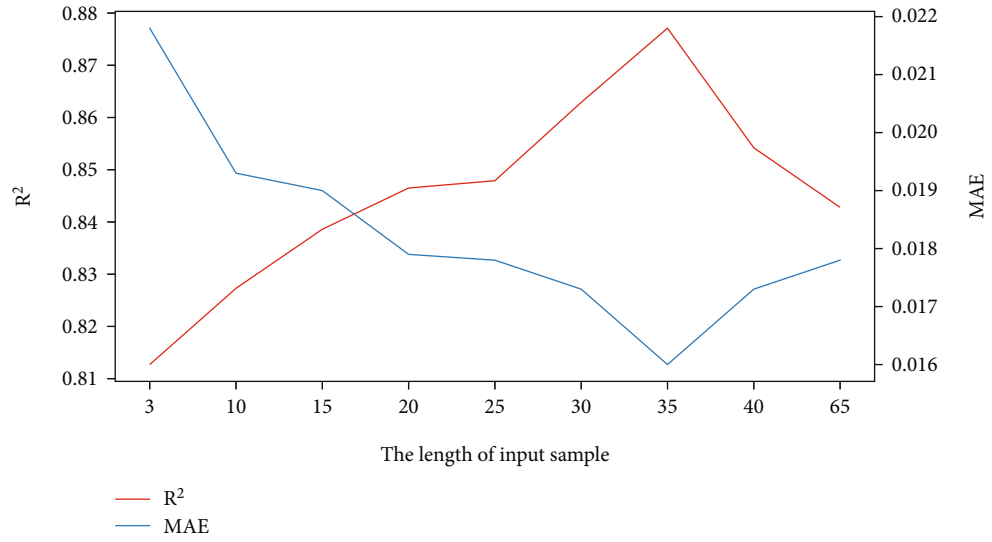


Figure 13: R^2 and MAE comparison of nine experimental results based on 2DCNN-GRU hybrid network.

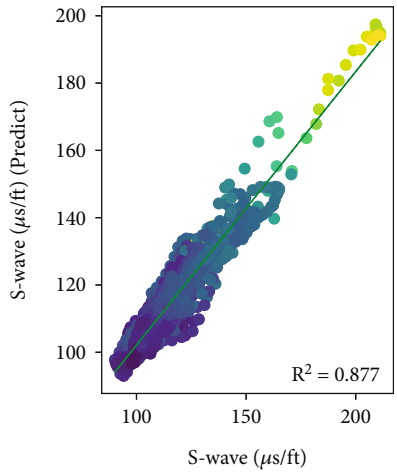


Figure 14: The correlation between shear wave velocity and predicted.

and the true values was analyzed in the form of a cross-plot (Figure 14), and the correlation reached 0.877. To sum up, the prediction accuracy of the 2DCNN-GRU hybrid network is affected by the length of the input samples. When the input sample length and convolution kernel length are 35, the prediction effect of the network is the best.

5. Conclusion

Considering the sequential features in the depth direction of rocks and spatial features of different physical properties of rocks, a new network 2DCNN-GRU hybrid network was proposed in this study, which can extract the spatial features of logging data from the 2DCNN and input them into the GRU to extract the temporal features, fully considering the temporal and spatial features of the logging data to predict Vs. In the case of I, the correlation coefficient, mean absolute error, and loss function of the evaluation parameters of 2DCNN-GRU were better than those of the separate 2DCNN and GRU, reaching 86.6, 0.165, and 5.2375*e*-04, respectively; comparing the prediction results of 2DCNN, GRU, and 2DCNN-GRU, the prediction effect of 2DCNN-GRU is better than that of 2DCNN and GRU alone. In the case of II, the prediction accuracy of 2DCNN-GRU was affected by the input sample length. The prediction accuracy of 2DCNN-GRU first increased and then decreased with the input sample length. When the sample length was 35, the prediction accuracy of the network reached the highest. The experimental results show that the newly proposed 2DCNN-GRU hybrid network outperforms other networks in prediction performance. In addition, the 2DCNN-GRU hybrid network proposed in this study was a supervised machine learning whose prediction accuracy was dependent on training sample accuracy.

Data Availability

The logging data used to support the findings of this study have not been made available because the data are shown in the article graphs.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions

Each author has contributed to the present paper. Tengfei Chen was responsible for drafting the article, programming, and performing the experiments; Gang Gao was responsible for conceiving the method, directing the experiments, and revising the article; Peng Wang was responsible for analysis of the data; Yonggen Li was responsible for interpretation

of the data; Bin Zhao supervised the experiments; and Zhixian Gui was responsible for revising the article.

Acknowledgments

This work is jointly supported by the State Key Program of the National Natural Science Foundation of China (Grant No. 42030805) and Scientific Research & Technology Development Project of the China National Petroleum Corporation (Grant No. 2021DJ3704).

References

[1] R. Saltzer, C. Finn, and O. Burtz, "Predicting Vshale and porosity using cascaded seismic and rock physics inversion," *The Leading Edge*, vol. 24, no. 7, pp. 732–736, 2005.

[2] K. Spikes, T. Mukerji, J. Dvorkin, and G. Mavko, "Probabilistic seismic inversion based on rock-physics models," *Geophysics*, vol. 72, no. 5, pp. R87–R97, 2007.

[3] R. K. Sharma and S. Chopra, "Determination of lithology and brittleness of rocks with a new attribute," *The Leading Edge*, vol. 34, no. 5, pp. 554–564, 2015.

[4] G. Gao, Y. H. Yang, B. Zhao, H. L. Duan, X. Y. Wang, and Y. Z. Wei, "A method for establishing and directly extracting sensitive identification factors of unconsolidated sandstones," *Oil Geophysical Prospecting*, vol. 54, no. 6, pp. 1329–1347, 2019.

[5] J. P. Castagna, M. L. Batzle, and R. I. Eastwood, "Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks," *Geophysics*, vol. 50, no. 4, pp. 571–581, 1985.

[6] M. L. A. Han, A. N. De-hua, and D. Morgan, "Effects of porosity and clay content on wave velocities in sandstones," *Geophysics*, vol. 51, no. 11, pp. 2093–2107, 1986.

[7] J. Dvorkin, M. Prasad, A. Sakai, and D. Lavoie, "Elasticity of marine sediments: rock physics modeling," *Geophysical Research Letters*, vol. 26, no. 12, pp. 1781–1784, 1999.

[8] B. H. Russell, K. Hedlin, F. J. Hilterman, and L. R. Lines, "Fluid-property discrimination with AVO: a Biot-Gassmann perspective," *Geophysics*, vol. 68, no. 1, pp. 29–39, 2003.

[9] Z. Hossain, T. Mukerji, and I. L. Fabricius, "Vp-Vs relationship and amplitude variation with offset modelling of glauconitic greensand‡," *Geophysical Prospecting*, vol. 60, pp. 117–137, 2012.

[10] W. M. Mabrouk and W. D. Pennington, "Compressional and shear wave velocity in terms of petrophysical parameters in clean formations," *Journal of Petroleum Science and Engineering*, vol. 65, pp. 62–66, 2009.

[11] M. Ojha and K. Sain, "Velocity-porosity and velocity-density relationship for shallow sediments in the Kerala-Konkan basin of western Indian margin," *Journal of the Geological Society of India*, vol. 84, no. 2, pp. 187–191, 2014.

[12] O. Oloruntobi, D. Onalo, S. Adedigb, L. James, and S. Butt, "Data-driven shear wave velocity prediction model for siliciclastic rocks," *Journal of Petroleum Science and Engineering*, vol. 183, article 106293, 2019.

[13] A. Jørstad, T. Mukerji, and G. Mavko, "Model-based shear-wave velocity estimation versus empirical regressions," *Geophysical Prospecting.*, vol. 47, no. 5, pp. 785–797, 1999.

[14] S. Xu and R. E. White, "A new velocity model for clay-sand mixtures1," *Geophysical Prospecting.*, vol. 43, no. 1, pp. 91–118, 1995.

[15] G. T. Kuster and M. N. Toksöz, "Velocity and attenuation of seismic waves in two-phase media: part I. Theoretical formulations," *Geophysics*, vol. 39, no. 5, pp. 587–606, 1974.

[16] C. H. Chen and M. N. Toksöz, "Inversion of seismic velocities for the pore aspect ratio spectrum of a rock," *Journal of Geophysical Research*, vol. 84, no. B13, pp. 7533–7543, 1979.

[17] F. Gassmann, "Elastic waves through a packing of spheres," *Geophysics*, vol. 16, no. 4, pp. 673–685, 1951.

[18] J. Y. Bai, Z. X. Song, L. Su, W. G. Yang, and S. J. Li, "Error analysis of shear-velocity prediction by the Xu-White model," *Chinese Journal of Geophysics*, vol. 55, no. 2, pp. 589–595, 2012.

[19] J. Y. Bai, C. Q. Yue, Y. Q. Liang et al., "Variable aspect ratio method in the Xu-White model for shear-wave velocity estimation," *Journal of Geophysics and Engineering*, vol. 10, no. 3, p. 035008, 2013.

[20] Z. S. Liu, S. Sun, and D. Zan, "The differential Kuster–Toksoz rock physics model for predicting S-wave velocity," *Journal of Geophysics and Engineering*, vol. 12, pp. 839–848, 2015.

[21] Y. Yang, X. Yin, G. Gao, Z. Gui, and B. Zhao, "Shear-wave velocity estimation for calciferous sandy shale formation," *Journal of Geophysics and Engineering*, vol. 16, no. 1, pp. 105–115, 2019.

[22] S. Xu and M. A. Payne, "Modeling elastic properties in carbonate rocks," *The Leading Edge*, vol. 28, no. 1, pp. 66–74, 2009.

[23] G. Z. Zhang, C. C. Li, X. Y. Yin, and J. Q. Zhang, "Corrected Xu-White model for carbonate S-wave velocity estimation," *Oil Geophysical Prospecting (in Chinese)*, vol. 32, no. 5, pp. 717–722, 2012.

[24] L. Huang, R. R. Stewart, S. Samilk, and N. Dyaur, "Fluid substitution effects on seismic anisotropy," *SEG Technical Program Expanded Abstracts.*, vol. 64, pp. 330–335, 2014.

[25] J. C. Gui, T. S. Ma, and P. Chen, "Rock physics modeling of transversely isotropic shale: an example of the Longmaxi formation in the Sichuan basin," *Chinese Journal of Geophysics*, vol. 63, no. 11, pp. 4188–4204, 2020.

[26] Z. S. Liu, H. Lu, J. Z. Liu, Q. Z. Bao, and Z. Y. Wang, "Estimation of S-wave velocity from wire-line logs for organic-rich rocks," *Journal of Petroleum Science and Engineering*, vol. 205, p. 108928, 2021.

[27] M. R. Rezaee, A. K. Ilkhchi, and A. Barabadi, "Prediction of shear wave velocity from petrophysical data utilizing intelligent systems: an example from a sandstone reservoir of Carnarvon Basin, Australia," *Journal of Petroleum Science & Engineering*, vol. 55, no. 3-4, pp. 201–212, 2007.

[28] M. Asoodeh and P. Bagheripour, "Prediction of compressional, shear and Stoneley wave velocities from conventional well log data using a committee machine with intelligent systems," *Rock Mechanics and Rock Engineering*, vol. 45, no. 1, pp. 45–63, 2012.

[29] H. Akhundi, M. Ghafoori, and G. R. Lashkaripour, "Prediction of shear wave velocity using artificial neural network technique, multiple regression and petrophysical data: a case study in Asmari reservoir (SW Iran)," *Open Journal of Geology*, vol. 4, no. 7, pp. 303–313, 2014.

[30] P. Bagheripour, A. Gholami, M. Asoodeh, and M. Vaezzadeh-Asadi, "Support vector regression based determination of shear wave velocity," *Journal of Petroleum Science and Engineering*, vol. 125, pp. 95–99, 2015.

[31] H. Fattahi, H. Nazari, and A. Molaghab, "Hybrid ANFIS with ant colony optimization algorithm for prediction of shear wave velocity from a carbonate reservoir in Iran,"

International Journal of Mining and Geological Engineering, vol. 50, pp. 231–238, 2016.

[32] N. Weijun, L. Qi, and F. Tao, "Prediction of shear wave velocity in shale reservoir based on support vector machine," *Journal of Xi'an Shiyou University (Natural Science)*, vol. 32, no. 4, pp. 46–49, 2017.

[33] M. Anemangely, A. Ramezanzadeh, H. Amiri, and S. Hoseinpour, "Machine learning technique for the prediction of shear wave velocity using petrophysical logs," *Journal of Petroleum Science and Engineering*, vol. 174, 2019.

[34] P. Wang and S. Peng, "On a new method of estimating shear wave velocity from conventional well logs," *Journal of Petroleum Science and Engineering.*, vol. 180, pp. 105–123, 2019.

[35] Q. Y. Ma, X. Zhang, C. L. Zhang, H. Zhou, and Z. Y. Wu, "Shear wave velocity prediction based on one-dimensional convolutional neural network," *Lithologic Reservoirs*, vol. 33, no. 4, pp. 111–120, 2021.

[36] W. Xiong, X. Ji, Y. Ma et al., "Seismic fault detection with convolutional neural network," *Geophysics*, vol. 83, no. 5, pp. O97–O103, 2018.

[37] Z. Zhong and T. R. Carr, "Geostatistical 3D geological model construction to estimate the capacity of commercial scale injection and storage of CO_2 in Jacksonburg-Stringtown oil field, West Virginia, USA," *International Journal of Greenhouse Gas Control*, vol. 80, pp. 61–75, 2019.

[38] R. H. Feng, T. M. Hansen, D. Grana, and N. Balling, "An unsupervised deep-learning method for porosity estimation based on post-stack seismic data," *Geophysics*, vol. 85, no. 6, pp. M97–M105, 2020.

[39] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Convolutional neural network for seismic impedance inversion," *Geophysics*, vol. 84, no. 6, pp. R869–R880, 2019.

[40] V. Das and T. Mukerji, "Petrophysical properties prediction from prestack seismic data using convolutional neural networks," *Geophysics*, vol. 85, no. 5, pp. N41–N55, 2020.

[41] M. Alfarraj and G. Alregib, "Semi-supervised learning for acoustic impedance inversion," *SEG Technical Program Expanded Abstracts*, 2019.

[42] A. Mustafa, M. Alfarraj, and G. Alregib, "Estimation of acoustic impedance from seismic data using temporal convolutional network," *SEG Technical Program Expanded Abstracts.*, 2019.

[43] K. X. Wang, Q. H. Huang, and S. H. Wu, "Application of long short-term memory neural network in geoelectric field data processing," *Chinese Journal of Geophysics*, vol. 63, pp. 3015–3024, 2020.

[44] N. Pham, X. M. Wu, and E. Z. Naeini, "Missing well log prediction using convolutional long short-term memory network," *Geophysics*, vol. 85, no. 4, pp. WA159–WA171, 2020.

[45] Y. Zhang, H. Zhong, Z. Wu, H. Zhou, and Q. Ma, "Improvement of petrophysical workflow for shear wave velocity prediction based on machine learning methods for complex carbonate reservoirs," *Journal of Petroleum Science and Engineering*, vol. 192, article 107234, 2020.

[46] J. You, J. Cao, X. Wang, and W. Liu, "Shear wave velocity prediction based on LSTM and its application for morphology identification and saturation inversion of gas hydrate," *Journal of Petroleum Science and Engineering*, vol. 205, p. 109027, 2021.

[47] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649, Vancouver, BC, Canada, 2013.

[48] Y. H. Sun and Y. Liu, "Prediction of S-wave velocity based on GRU neural network," *Oil Geophysical Prospecting*, vol. 55, no. 3, pp. 484–492, 2020.

[49] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[50] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[52] K. Cho, B. V. Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: encoder-decoder approaches," *Computer Science*, vol. 1409, p. 1259, 2014.

[53] H. Yan, G. Chen, Y. Tong, B. Ji, and B. C. Hu, "Human rehabilitation motion recognition based on pose estimation and GRU network," *Computer Engineering*, vol. 47, no. 1, pp. 12–20, 2021.

[54] J. Wang, Z. X. Gao, and Y. M. Zhu, "Prediction of Yellow River water quality based on CNN-LSTM model," *People's Yellow River*, vol. 43, no. 5, p. 96-99+109, 2021.

[55] K. B. Zhou, J. Y. Zhang, Y. S. Ren, H. Zhen, and L. X. Zhao, "A gradient boosting decision tree algorithm combining synthetic minority over-sampling technique for lithology identification," *Geophysics*, vol. 85, no. 4, pp. WA147–WA158, 2020.

[56] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, http://arxiv.org/abs/1412.6980.

Hindawi
Geofluids
Volume 2022, Article ID 4100638, 8 pages
https://doi.org/10.1155/2022/4100638

Research Article

Application of Artificial Intelligence Models to Predict the Tensile Strength of Glass Fiber-Modified Cemented Backfill Materials during the Mine Backfill Process

Lei Zhu,[1] **Wenzhe Gu**,[1,2] **Zhicheng Liu**,[1] **and Fengqi Qiu**[1,3]

[1]*China Coal Energy Research Institute Co., Ltd, Xi'an, Shaanxi 710054, China*
[2]*China University of Mining & Technology (Beijing), Beijing 100083, China*
[3]*Xi'an University of Science & Technology, Xi'an, Shaanxi 710054, China*

Correspondence should be addressed to Lei Zhu; zhulei@chinacoal.com and Wenzhe Gu; guwenzhe@chinacoal.com

Received 12 January 2022; Accepted 18 April 2022; Published 1 June 2022

Academic Editor: Bicheng Yan

Cemented backfill coal mining technology is gradually becoming a key technology for green mining of coal resources. And cemented backfill materials generally have congenital defects such as poor crack resistance, poor durability, and high brittleness, which restrict the promotion and application of cemented backfill coal mining technology. Due to the complex stress environment of in situ stress, mining stress, water pressure, and gas pressure, cemented backfill materials need to have good mechanical properties, and glass fiber is usually used to mix into cemented backfill materials to improve its performance, but there are many problems including complex testing process, high cost, and long time-consuming in the study of mechanical properties of glass fiber-modified cemented backfill materials (GFCBM) by laboratory tests. Consequently, this study proposed and compared four artificial intelligence models to forecast the tensile strength of GFCBM. Firstly, the laboratory tests of tensile properties of GFCBM under different influence factors were implemented to supply the prediction model with dataset. The input variables are aeolian sand content, cement content, glass fiber length, and glass fiber content, and the output variable is the tensile strength of GFCBM. The correlation coefficient (R), mean absolute error (MAE), and root mean square error (RMSE) are selected to assess the estimated performance of the hybrid intelligent model. The results indicate that the four hybrid artificial intelligence models show a latent capacity for forecasting the tensile strength of GFCBM, and according to the order from high to low, the prediction ability of the four prediction models is as follows: ABC-SVM, GA-SVM, SSA-SVM, and DE-SVM, and the corresponding R values are 0.9555, 0.9539, 0.9413, and 0.9359, respectively. The research findings are beneficial to promote the application of cemented backfill coal mining technology.

1. Introduction

The surface subsidence, soil erosion, and other ecological environment damage caused by coal mining have been a major problem, which seriously restricts the harmonious development of coal exploitation and ecoenvironment protection [1, 2]. As a key technology for coal resource green mining, backfill mining technology uses the supporting function of backfill body to control the overlying strata movement and reduce surface subsidence, which has become one of the important ways to achieve safe, efficient, economic, and sustainable development of coal resources [3, 4]. Because of its low cost and mature technology, cemented backfill mining technology has been extensively used in mining areas in the Northwest of China [5]. As a typical cement-based material, cemented backfill materials generally have congenital defects such as poor crack resistance, poor durability, and high brittleness. Since the cemented backfill body has the "arching effect" after it is backfilled into the goaf, it requires not only sufficient compressive strength, but also good tensile and shear properties [6]. At present, simply increasing the amount of cement to make up for

the above shortcomings not only brings huge economic and safety costs to the mine but also largely restricts the wider promotion and application of cemented backfill mining technology.

Researchers have proposed that adding glass fiber materials to the cemented backfill materials can increase the strength, toughness, and ductility of the materials and optimize its mechanical properties [7, 8]. The role of glass fiber materials in cemented backfill materials can be summarized as crack resistance, reinforcement, and toughening. The specific reinforcement effect is related to the orientation, the ratio of length to diameter, and the volume content of the glass fiber materials. Qin et al. [9] analyzed the feasibility of polypropylene fiber fabrics to enhance the mechanical properties of concrete and compared the compressive strength and multiscale failure characteristics of ordinary concrete and fiber-modified concrete, and it is found that polypropylene fiber fabric is beneficial to improve the compressive strength of concrete. Elkatatny et al. [10] tested the effect of glass fiber materials on the tensile strength, porosity, and permeability of cement under high temperature and high pressure. The results show that glass fiber materials do not significantly affect the rheology, density, and water content of cement but can significantly improve its tensile strength and compressive strength. Yi et al. [11] studied the internal failure mechanism of cemented backfill materials with glass fiber through uniaxial compression test and X-ray computed tomography (CT), and the results indicate that the mechanical strength of cemented backfill materials with glass fiber is increased by about 70%~90%, and the glass fiber can effectively prevent the propagation of internal cracks. It can be seen that the incorporation of glass fiber materials is able to indeed change the mechanical properties of cemented backfill materials, and the means of its optimization process is mainly laboratory test. However, this method is currently faced with the problems of complex test process, high cost, and long time-consuming, which restrict the development of this research. Therefore, how to find other methods to conveniently obtain the changes in the properties of glass fiber-modified cemented backfill materials (GFCBM) is of great significance.

At present, artificial intelligence technology has been gradually applied in many engineering fields [12, 13]. It can achieve better prediction results on the basis of comprehensive consideration of various influencing factors. Yan et al. [14] proposed an intelligence model named BPNN-GA-AdaBoost to predict the change of coal strength after CO_2 injection into coal seam; Han et al. [15] integrated random forest and particle swarm optimization algorithm to evaluate the fracture performance of concrete; Jalal et al. [16] estimated the swelling strength of expansive soil through ANN, GEP, and ANFIS methods. It can be seen that intelligent prediction has been used in various engineering directions and has achieved excellent results, but currently, it is still facing the following problems: (1) At present, there is almost no intelligence model to forecast the tensile performance of GFCBM, especially the use of support vector machine (support vector machine has obvious advantages in small sample data). (2) There is a lack of comparative studies on the tensile strength of cemented backfill materials using different prediction models. Therefore, it is of urgent significance to implement a comparative study on the intelligent prediction of the tensile properties of GFCBM.

This paper puts forward four hybrid artificial intelligence models, namely, ABC-SVM, DE-SVM, GA-SVM, and SSA-SVM, to predict the tensile strength of GFCBM. Among them, the support vector machine (SVM) is mainly employed in analyzing the function relation between the tensile strength of GFCBM and various influence parameters. The dataset of the model is gained through laboratory tests on tensile properties of GFCBM. The input variables of the model are aeolian sand content, cement content, glass fiber length, and glass fiber content, and the output variable is the tensile strength of GFCBM. The R, MAE, and RMSE were selected to evaluate and compare the prediction performance of these hybrid intelligent model. Finally, the optimal model for predicting the tensile performance of GFCBM was obtained. The research findings are beneficial to promote the application of cemented backfill mining technology.

2. Experiment

2.1. Materials. The aeolian sand and fly ash, together with a small amount of cement and quicklime, are chosen as the cemented backfill materials. The aeolian sand is mainly taken from the mining area in Northern Shaanxi, where the surface is covered with a large amount of aeolian sand, and the fly ash comes from the power plant. Glass fiber is selected as the doped fiber material, mainly considering the low cost. Among them, the length of glass fiber is 3 mm, 6 mm, and 15 mm, and the maximum tensile strength of single glass fiber is 2800 MPa. It is often used as the filler of cement or concrete to improve the strength, impact resistance, tensile strength, bending resistance, and durability of materials, and it is an ideal multifunctional reinforcing material. The photograph of glass fiber is shown in Figure 1.

2.2. Preparation Process. Firstly, glass fiber and raw materials of backfill materials (fly ash, aeolian sand, cement, and quicklime) were weighed, mixed, and stirred evenly, and then, water was added for mixing. Then, the mixed mortar was poured into self-made abrasives and put into the curing box for curing; eventually, a cube-like concrete specimen is formed. The size of the abrasive tool is $7.07 \times 7.07 \times 7.07$ cm, and the WAW-1000D servo press machine was selected to test the tensile strength of the cemented backfill materials at different ages.

2.3. Experimental Scheme. In current study, the influence of aeolian sand content, fly ash content, cement content, quicklime content, glass fiber length, and glass fiber content on the tensile strength of cemented backfill materials is mainly considered. The content of each material refers to the weight proportion, and the specific design scheme is shown in Table 1. Among them, there are 4 kinds of cemented backfill ratios, 3 kinds of glass fiber lengths, 5 kinds of glass fiber content, and a total of 60 schemes. Each scheme carries out 3 experiments for a total of 180 experiments.

Figure 1: The photograph of glass fiber.

Table 1: Experimental scheme design [17].

Group	Aeolian sand : fly ash: cement: quicklime (%)	Glass fiber length (mm)	Glass fiber content (‰)
1	47.5 : 35 : 12.5 : 5	3, 6, 15	1, 3, 5, 10, 15
2	55 : 35 : 5 : 5	3, 6, 15	1, 3, 5, 10, 15
3	21.5 : 35 : 38.5 : 5	3, 6, 15	1, 3, 5, 10, 15
4	30 : 35 : 30 : 5	3, 6, 15	1, 3, 5, 10, 15

The design of the experimental scheme is mainly based on the results of previous studies. Aeolian sand content, cement content, glass fiber length, and glass fiber content are the main variables in this scheme. Therefore, the input variable of the artificial intelligence prediction model selects the above variables. Since glass fiber can significantly modify the tensile strength of cemented backfill materials, this study mainly tests the tensile strength of GFCBM, and it is also the output variable of the artificial intelligence prediction model.

3. Machine Learning Algorithms

In this study, four artificial intelligence models are used to predict the tensile properties of GFCBM, which are ABC-SVM, DE-SVM, GA-SVM, and SSA-SVM. Among them, SVM is employed in analyzing the function relation between the tensile strength of GFCBM and various influence parameters, and ABC, DE, GA, and SSA are employed in optimizing the parameters of the SVM.

3.1. Support Vector Machine. Support vector machine (SVM) [18] is a machine learning means raised by Vapnik. It can be divided into support vector classification and support vector regression to solve classification and regression problems, respectively. As shown in Figure 2, the core idea of SVM is the conversion from low-dimensional spatial data points to high-dimensional spatial data points through nonlinear mapping and adopt the principle of structural risk minimization, and then, classify and predict the data in the high-dimensional space. SVM can effectively avoid the local

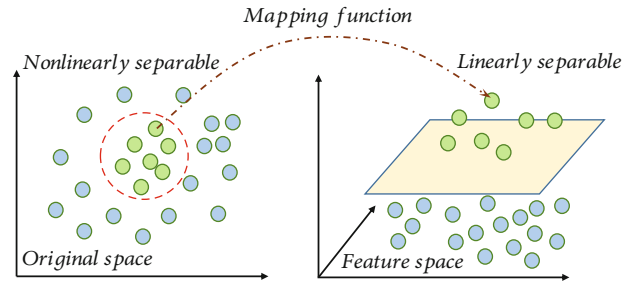


Figure 2: The SVM solution principle.

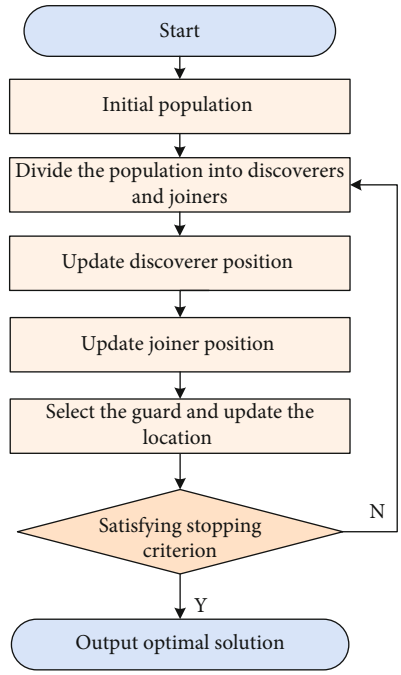


Figure 3: SSA algorithm flow chart.

extremum problem, maximize the prediction accuracy, and prevent the data from overfitting. According to the restricted sample data, it can obtain the optimal value between model complexity and forecast accuracy and improve its generalization ability.

The selection of kernel function and related parameter setting are the key of SVM. In this study, the SVR is selected, and the widely used RBF kernel function is adopted, because it is suitable for different samples and various dimension problems and has strong nonlinear mapping ability. The hyperparameters (C and g) of the SVM model are closely related to its predictive ability, and the optimal solution needs to be gained through optimization algorithms.

3.2. Artificial Bee Colony Algorithm. Artificial bee colony (ABC) is a new optimization algorithm in view of swarm intelligence put forward by Vasquez and Garro [19]. The artificial bee colony algorithm model mainly includes the following elements: one is the nectar source, that is, the group goal. The composition of the group is dedicated to

Table 2: Basic parameter statistics of the dataset.

Parameter	Minimum	Maximum	Unit	Variable
Aeolian sand content	21.50	55	%	Input
Cement content	5	38.50	%	Input
Glass fiber length	3	15	mm	Input
Glass fiber content	1	15	‰	Input
Tensile strength of GFCBM	0.17	1.23	MPa	Output

finding the best nectar source and continuously updated after mining; the second is the composition and division of the bee colony. The hired bee is dedicated to discovering and sharing nectar source information with the follower bees, while the scout bee is dedicated to the update of the nectar source. It is always transformed from the hired bee when the quality of the nectar source drops. The third is colony behavior, that is, the recruitment of new bees and the abandonment of low-value nectar sources.

3.3. Differential Evolution Algorithm. The DE algorithm was put forward by Yuan et al. [20]. This algorithm solves the optimization problem by means of the cooperation and competition of individuals in the whole population and has a strong global convergence potentiality. The process of DE algorithm is similar to other evolutionary algorithms, including mutation, selection, and crossover operations, but compared with other algorithms, DE algorithm runs stably, converges quickly and has low complexity. The DE algorithm begins from the initial population, after mutation, selection, and crossover operations; the best individual is saved in the new population and then iterates until the termination condition is met.

3.4. Genetic Algorithm. Genetic algorithm (GA) is a type of evolutionary calculation, which is a method to imitate Darwin's genetic selection and natural elimination of biotic evolution process [21]. The algorithm is simple, general, and robust and is suitable for parallel processing. The GA mainly transfers the better genes to the next generation by the means of the selection operator and expands the search range by the means of the crossover operator, and the mutation operator accelerates the convergence speed, so as to achieve the goal of global search.

3.5. Sparrow Search Algorithm. Sparrow search algorithm (SSA) is a latest swarm intelligence optimization algorithm, put forward in 2020 [22]. During the process of foraging for sparrows, it is compartmentalized into discoverers and joiners. The discoverers are responsible for finding food in the population and provide search directions for the whole population, while joiners use the discoverers' guidance to obtain food. In order to obtain food, sparrows can usually forage for food using two behavioral strategies: discoverer and joiner. Individuals in the population will be alert to the other individual behaviors, and attackers in the population will compete with high-intake companions for food resources to increase their predation rate. The specific flow chart is shown in Figure 3.

4. Methodology

4.1. Dataset Preparation. In current study, the tensile properties of GFCBM from the experimental test are used as the dataset for the artificial intelligence. As mentioned above, there are 180 series of data for training and testing. According to early research experience [23, 24], compared with other models such as ANN, SVM has obvious advantages in dealing with small samples and nonlinear problems. It is unnecessary to be large for the dataset required for the training and testing of SVM-based model, and the artificial intelligence models can be well trained and tested by the 180 series of data in this paper. According to the experimental scheme, the aeolian sand content and cement content in the GFCBM are variables, and the glass fiber length and content are also variables. Therefore, the input variables are aeolian sand content, cement content, glass fiber length, and glass fiber content, and the output variable is the tensile strength of GFCBM. Table 2 summarizes the data statistics for the whole dataset.

In the process of modeling, the whole dataset will be separated into a training set and a testing set on the basis of a certain ratio [25]. In this paper, the ratio is selected to be 7 : 3, that is, 126 series of data are selected for the training set, and the 54 series of data are selected for the testing set.

4.2. Model Establishment. Figure 4 presents four hybrid artificial intelligence models, that is, ABC-SVM, DE-SVM, GA-SVM, and SSA-SVM. Among them, SVM is employed in analyzing the function relation between the tensile strength of GFCBM and various influence parameters, and ABC, DE, GA, and SSA are employed in optimizing the parameters of the SVM. For comparison and analysis purposes, based on the optimal effect and stable convergence of various machine learning algorithms, it is better to keep the model parameters consistent. The kernel function of SVM is RBF radial basis function, the population size of each algorithm is given with 50, and the maximum number of iterations is given with 100.

4.3. Model Validation and Evaluation. The intelligence model validation and evaluation is an important link for the development of the model. In this study, for the sake of assessing the dependability of the hybrid model effectively, the function relationship between the predicted value and the measured value is described by the correlation coefficient (R), mean absolute error (MAE), and root mean

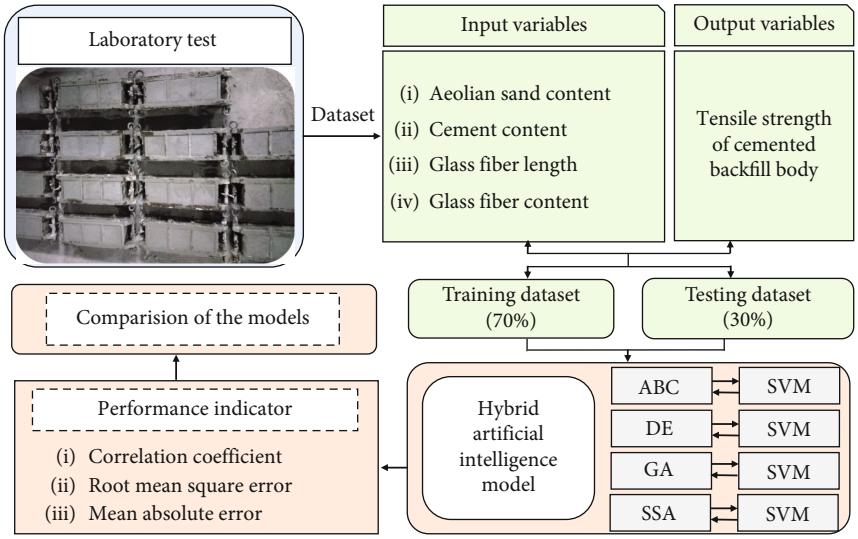


Figure 4: Overall analysis process.

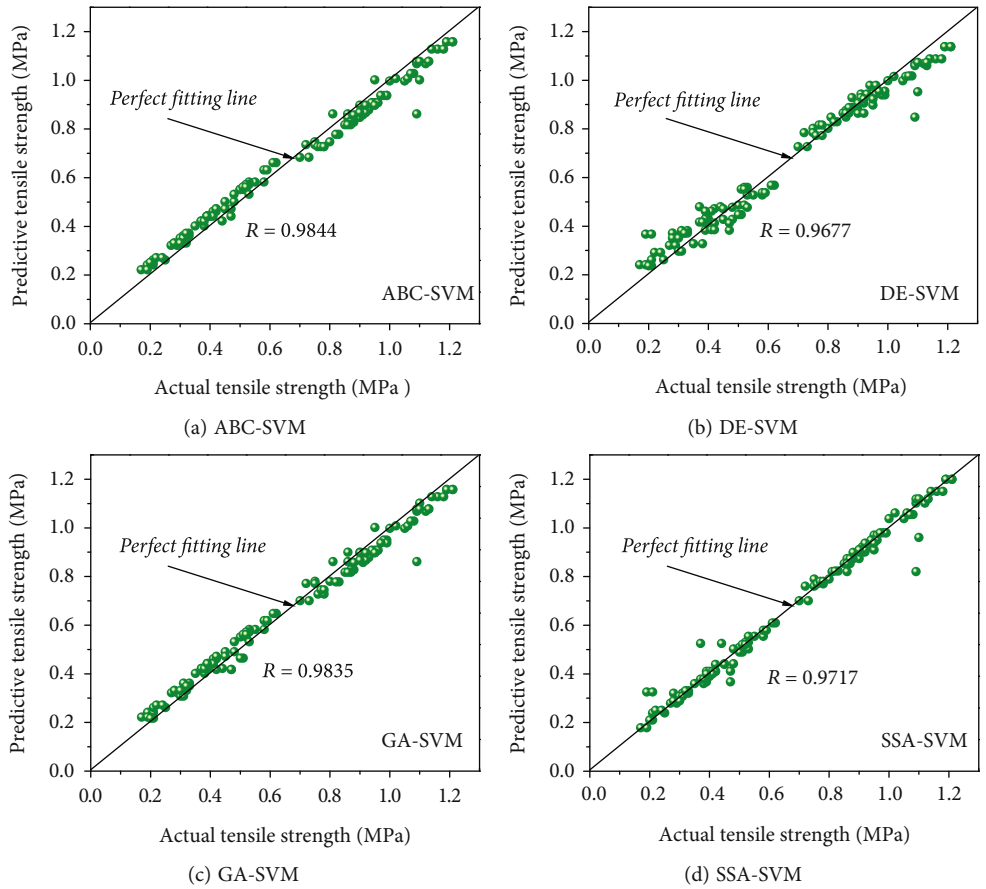


Figure 5: The training effects of different prediction model for training set.

square error (RMSE), respectively, [26]. The closer the value of R is to 1, the better correlation between predicted value and measured value; the smaller the MAE and RMSE, the smaller the error between the predicted value and the measured value. The calculation formula of the three evaluation indexes is as follows:

$$R = \frac{\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}, \quad (1)$$

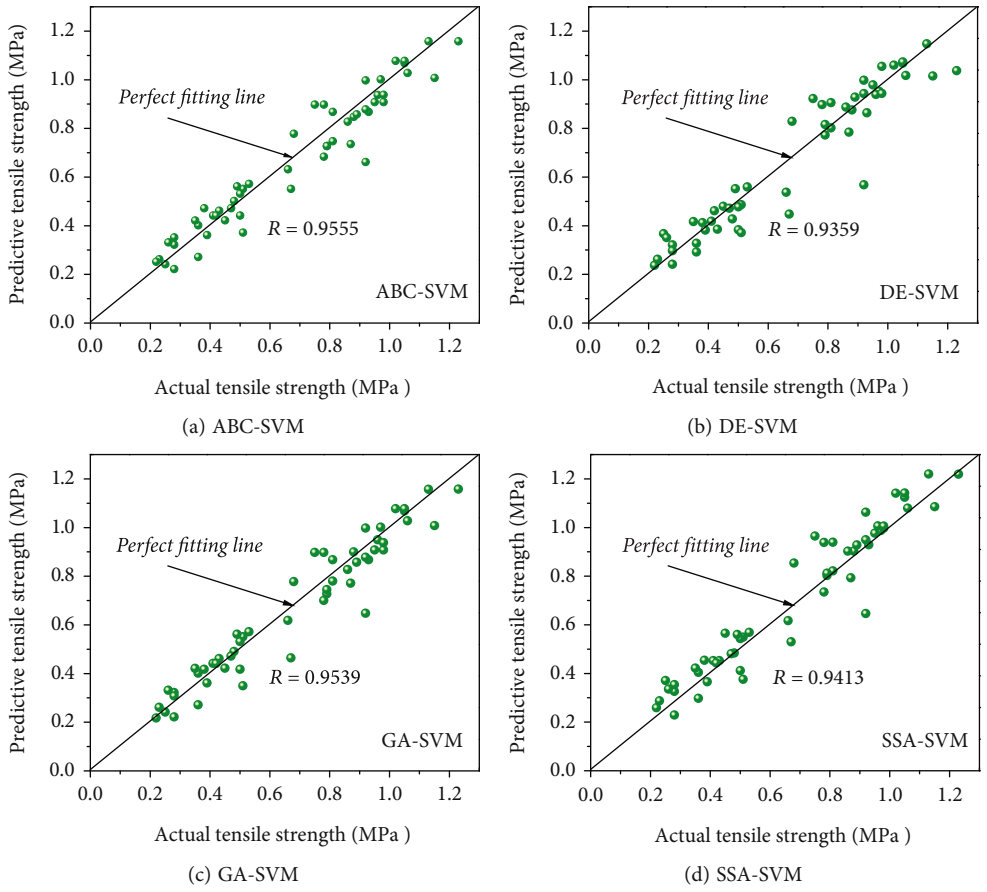


Figure 6: The prediction effects of different prediction model for testing set.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i^* - y_i|, \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2}, \tag{3}$$

where n refers to the number of datasets, y_i^* refers to the predicted value, y_i refers to the measured value, $\overline{y^*}$ refers to the average of the predicted value, and $\bar{y}$ refers to the average of the measured value.

5. Results and Discussion

5.1. Comparative Analysis of Different Prediction Models. This study mainly analyzes and compares the forecast performance of the above four hybrid artificial intelligence models in the tensile strength of GFCBM from two parts of training set and testing set.

Figure 5 shows the training effects of different prediction model for training set. It can be demonstrated that the four hybrid artificial intelligence models have obtained good training effects, and the sample data are essentially near the ideal fitting line (measured value = predicted value), and only a few sample points deviate from the fitting line. From the perspective of the R value, the training effect of ABC-SVM is the best, its R value is 0.9844, followed by GA-SVM, SSA-SVM, and DE-SVM, and its R values are 0.9835, 0.9717, and 0.9677, respectively. In general, the training effects of the four hybrid artificial intelligence prediction models have reached high accuracy.

When completing the model training, the trained model is used for prediction. Figure 6 shows the prediction effects of different prediction model for testing set. By analyzing the data distribution, it is clear that the sample data of testing set is also essentially near the ideal fitting line (measured value = predicted value). According to the order from high to low, the prediction ability of the four prediction models is as follows: ABC-SVM, GA-SVM, SSA-SVM, and DE-SVM, and the corresponding R values are 0.9555, 0.9539, 0.9413, and 0.9359, respectively. Consequently, ABC-SVM has the best predictive ability in terms of the tensile strength of GFCBM.

To better analyze and compare different forecast model performance, the performance indicators of different prediction models are summarized, as shown in Table 3, and it can be drawn that compared with the other three models, the prediction accuracy of ABC-SVM hybrid model is higher in training set and testing set. Its R value is the largest, and its RMSE and MAE values are also very small. This shows that the ABC-SVM intelligent model not only gives full play to the superiority of SVM in handling problems with few samples but also gives full play to the characteristics of ABC in hyperparameter optimization. Considering

Table 3: The performance indicators of different prediction models.

	Training set			Testing set		
	R	RMSE	MAE	R	RMSE	MAE
ABC-SVM	0.9844	0.0474	0.0413	0.9555	0.0749	0.0605
DE-SVM	0.9677	0.0541	0.0515	0.9359	0.0967	0.0707
GA-SVM	0.9835	0.0453	0.0384	0.9539	0.0769	0.0626
SSA-SVM	0.9717	0.0501	0.0507	0.9413	0.0860	0.0661

comprehensively, the ABC-SVM intelligent model has better learning and predictive capabilities. Consequently, this study suggests using the ABC-SVM intelligent model to forecast the tensile strength of GFCBM.

5.2. Contributions and Shortcomings. The innovations and main contributions of this research are as follows: (1) it is proposed to use artificial intelligence technology to predict the tensile strength of GFCBM, which effectively avoids the disadvantages of complex laboratory testing process, long time-consuming, and high cost; (2) considering that SVM has many unique advantages in solving small sample and nonlinear and high-dimensional pattern recognition, it is proposed to use support vector machine to construct the prediction model, which solves the defect of small sample data; (3) systematic comparative research on intelligent optimization algorithms to optimize the performance of SVM has been carried out.

This study is the initial exploration of artificial intelligence technology to forecast the GFCBM mechanical properties. In the future, it is necessary to use artificial intelligence model to forecast the compressive strength and shear strength of GFCBM. Meanwhile, the dataset of artificial intelligence prediction model needs to be continuously enriched, so as to make the better predictive ability.

6. Conclusions

In this research, there are four hybrid artificial intelligence models to be proposed and compared for predicting the tensile strength of GFCBM, that is, ABC-SVM, DE-SVM, GA-SVM, and SSA-SVM. The dataset of the model is built through laboratory tests on tensile properties of GFCBM. The input variables are aeolian sand content, cement content, glass fiber length, and glass fiber content, and the output variable is the tensile strength of GFCBM. The R, RMSE, and MAE are selected to assess the estimated performance of the hybrid intelligent model. The main findings are as follows:

(1) Through laboratory tests of different fly ash content, aeolian sand content, cement content, quicklime content, glass fiber length, and glass fiber content on the tensile strength of GFCBM, it is found that the glass fiber can effectively change the tensile strength of cemented backfill materials

(2) The four hybrid artificial intelligence models proposed in this study show a latent capacity for forecasting the tensile strength of GFCBM, and ABC, DE, GA, and SSA have good effects on SVM hyperparameter optimization

(3) According to the order from high to low, the prediction ability of the four prediction models is as follows: ABC-SVM, GA-SVM, SSA-SVM, and DE-SVM, and the corresponding R values are 0.9555, 0.9539, 0.9413, and 0.9359, respectively. In this study, ABC-SVM intelligent model is suggested to forecast the tensile strength of GFCBM

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by Key Science and Technology Projects of China Coal Group "Research on theory, technology and equipment of four-dimensional collaborative backfill of slurry pipeline in mining overburden fracture area of coal mine" (ZMYXM*CK-023).

References

[1] M. B. Chi, D. S. Zhang, Q. Zhao, W. Yu, and S. S. Liang, "Determining the scale of coal mining in an ecologically fragile mining area under the constraint of water resources carrying capacity," *Journal of Environmental Management*, vol. 279, article 111621, 2021.

[2] Y. G. Yang, T. T. Guo, and W. T. Jiao, "Destruction processes of mining on water environment in the mining area combining isotopic and hydrochemical tracer," *Environmental Pollution*, vol. 237, pp. 356–365, 2018.

[3] Y. Zhang, S. G. Cao, N. Zhang, and C. Z. Zhao, "The application of short-wall block backfill mining to preserve surface water resources in northwest China," *Journal of Cleaner Production*, vol. 261, article 121232, 2020.

[4] Q. Zhang, J. X. Zhang, Y. L. Huang, and F. Ju, "Backfilling technology and strata behaviors in fully mechanized coal mining working face," *International Journal of Mining Science and Technology*, vol. 22, no. 2, pp. 151–157, 2012.

[5] C. C. Qi and A. Fourie, "Cemented paste backfill for mineral tailings management: rReview and future perspectives," *Minerals Engineering*, vol. 144, article 106025, 2019.

[6] B. Y. Li, Y. Hao, J. X. Zhang, and N. Zhou, "Compaction property prediction of mixed gangue backfill materials using hybrid intelligence models: a new approach," *Construction and Building Materials*, vol. 247, article 118633, 2020.

[7] N. Zhou, E. B. Du, J. X. Zhang, C. L. Zhu, and H. Q. Zhou, "Mechanical properties improvement of sand-based cemented backfill body by adding glass fibers of different lengths and ratios," *Construction and Building Materials*, vol. 280, article 122408, 2021.

[8] K. L. Cheng, B. B. Tu, L. Liu, B. Zhang, and H. F. Qiu, "Damage strengthening constitutive model of cemented paste backfill," *Advances in Civil Engineering*, vol. 2021, Article ID 5593983, 2021.

[9] Y. Qin, X. W. Zhang, J. R. Chai, Z. G. Xu, and S. Y. Li, "Experimental study of compressive behavior of polypropylene-fiber-reinforced and polypropylene-fiber-fabric-reinforced concrete," *Construction and Building Materials*, vol. 194, pp. 216–225, 2019.

[10] S. Elkatatny, R. Gajbhiye, A. Ahmed, and A. A. Mahmoud, "Enhancing the cement quality using polypropylene fiber," *Journal of Petroleum Exploration and Production Technology*, vol. 10, no. 3, pp. 1097–1107, 2020.

[11] X. W. Yi, G. W. Ma, and A. Fourie, "Compressive behaviour of fibre-reinforced cemented paste backfill," *Geotextiles and Geomembranes*, vol. 43, no. 3, pp. 207–215, 2015.

[12] J. D. Niu, B. Wang, H. F. Wang et al., "An intelligent prediction method of the karst curtain grouting volume based on support vector machine," *Geofluids*, vol. 2020, Article ID 8892106, 2020.

[13] H. Yan, J. X. Zhang, S. S. Rahman, N. Zhou, and Y. Suo, "Predicting permeability changes with injecting CO_2 in coal seams during CO_2 geological sequestration: a comparative study among six SVM-based hybrid models," *Science of the Total Environment*, vol. 705, article 135941, 2020.

[14] H. Yan, J. X. Zhang, N. Zhou, and M. Li, "Application of hybrid artificial intelligence model to predict coal strength alteration during CO_2 geological sequestration in coal seams," *Science of the Total Environment*, vol. 711, article 135029, 2020.

[15] X. Y. Han, Q. H. Xiao, K. Cui et al., "Predicting the fracture behavior of concrete using artificial intelligence approaches and closed-form solution," *Theoretical and Applied Fracture Mechanics*, vol. 112, article 102892, 2020.

[16] F. E. Jalal, Y. F. Xu, M. Iqbal, M. F. Javed, and B. Jamhiri, "Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP," *Journal of Environmental Management*, vol. 289, article 112420, 2021.

[17] N. Zhou, J. X. Zhang, S. Y. Ouyang, X. J. Deng, C. W. Dong, and E. B. Du, "Feasibility study and performance optimization of sand-based cemented paste backfill materials," *Journal of Cleaner Production*, vol. 259, article 120798, 2020.

[18] D. Y. Dou, W. Z. Wu, J. G. Yang, and Y. Zhang, "Classification of coal and gangue under multiple surface conditions via machine vision and relief-SVM," *Powder Technology*, vol. 356, pp. 1024–1028, 2019.

[19] R. A. Vazquez and B. A. Garro, "Training spiking neural models using artificial bee colony," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 947098, 2015.

[20] Y. Y. Yuan, X. S. Wang, X. Y. Meng, Z. Zhang, and J. M. Cao, "A strategy for helical coils multi-objective optimization using differential evolution algorithm based on entropy generation theory," *International Journal of Thermal Sciences*, vol. 164, article 106867, 2021.

[21] A. A. Motlagh, N. Shabakhty, and A. Kaveh, "Design optimization of jacket offshore platform considering fatigue damage using Genetic Algorithm," *Ocean Engineering*, vol. 227, article 108869, 2021.

[22] J. K. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.

[23] C. C. Qi and X. L. Tang, "Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study," *Computers & Industrial Engineering*, vol. 118, pp. 112–122, 2018.

[24] H. Xiao, Y. Chen, C. Dou et al., "Prediction of ash-deformation temperature based on grey-wolf algorithm and support-vector machine," *Fuel*, vol. 241, pp. 304–310, 2019.

[25] A. R. Khoei, H. Moslemi, and M. R. Seddighian, "An efficient stress recovery technique in adaptive finite element method using artificial neural network," *Engineering Fracture Mechanics*, vol. 237, article 107231, 2020.

[26] Q. W. Li, J. Wu, and H. Q. Wei, "Reduction of elemental mercury in coal-fired boiler flue gas with computational intelligence approach," *Energy*, vol. 160, pp. 753–762, 2018.

Hindawi
Geofluids
Volume 2022, Article ID 3503585, 7 pages
https://doi.org/10.1155/2022/3503585

Research Article

Study on the Imbibition Characteristics of Different Types of Pore-Throat Based on Nuclear Magnetic Resonance Technology

Xiong Liu,[1] **Yang Zhang**,[1] **Ziming Zhang**,[2] **Jinze Xu**,[1,3] **Desheng Zhou**,[1] **Jian Su**,[2] **and Ying Tang**[1]

[1]*School of Petroleum Engineering, Xi'an Shiyou University, Xi'an, China*
[2]*Drilling & Production Technology Research Institute, PetroChina Liaohe Oilfield Company, Panjin, China*
[3]*Department of Chemical and Petroleum Engineering, University of Calgary, Calgary, Canada*

Correspondence should be addressed to Jinze Xu; jinzxu@ucalgary.ca

Received 14 November 2021; Revised 12 February 2022; Accepted 7 April 2022; Published 27 April 2022

Academic Editor: Andri Stefansson

"Fracturing network+imbibition oil production" is a new attempt to effectively develop low-permeability tight reservoirs. Fracturing fluid is not only a carrier for sand carrying but also a tool in the process of imbibition. On the basis of imbibition experiments, combined with nuclear magnetic resonance and pseudo-color processing technology, this paper clarified the dominant forces of different types of pore-throat and quantitatively characterized the contribution of different levels of pore-throat to imbibition oil recovery. The results show that gravity is the main controlling force of imbibition for reservoirs with higher permeability. Fluid replacement mainly occurs in the early period of imbibition. Macropores contribute most of the imbibition recovery, mesopores have a weak contribution, and the contribution of micropores and pinholes can be ignored. For the reservoirs with low permeability, capillary force is the main controlling force of imbibition. Fluid replacement mainly occurs in the later period of imbibition. Macropores contribute most of the imbibition recovery rate, mesopores contribute a small part of the imbibition recovery factor, and the contribution of micropores and pinholes can be ignored. This paper clarified that macropores and mesopores are the main sources of the contribution of imbibition recovery efficiency, and oil content and connectivity are key factors for the imbibition recovery efficiency.

1. Introduction

The development potential of low-permeability tight oil reservoirs is huge [1, 2], and imbibition oil recovery has become the most effective method for enhancing oil recovery for this type of oil reservoirs [3, 4]. Its principle is to use capillary force as the main displacement power to enhance the oil-water replacement capacity between fractures and rock matrix [5, 6]. It can finally achieve the purpose of increasing the degree of crude oil production in the matrix [7, 8]. It is significant to clarify the main controlling factors of different levels of pore-throat and analyze the imbibition characteristics of different types of pore-throat for improving the productivity of a single well [9–11].

At present, there are many scholars who study the law of imbibition oil production based on nuclear magnetic resonance technology: Mason et al. [12–14] compared the imbibition characteristics of rocks under the different boundary conditions. The results show that the imbibition rate of rocks increases with the increase in the total surface area and the recovery increases with the increase in the rock shape factor. Bertoncello et al. took advantage of nuclear magnetic resonance technology to study the influence factors on imbibition [15, 16]. The results show that the greater the permeability and the longer the shut-in time, the higher the imbibition replacement efficiency; the higher the initial water saturation and the longer the imbibition equilibrium time, the lower the imbibition replacement efficiency; and the higher the permeability and injection ratio, the longer the imbibition distance [17–19]. Chen et al. evaluated the influence physical parameters on the dynamic imbibition efficiency [20, 21]. The results show that the average pore radius, permeability, porosity, average pore-throat ratio, medium diameter, average roar radius, and specific surface

Table 1: The physical parameters of cores.

Core number	Porosity (%)	The gas measurement permeability (10^{-3} μm^2)	Length (cm)	Diameter (cm)	Dry weight (g)
1	19.57	14.140	3.95	2.50	39.232
2	2.78	0.059	4.00	2.50	49.745

area are the main factors that affect the dynamic imbibition efficiency, and the degree of influence of those on the imbibition efficiency is gradually poor. The degree of influence of the sorting coefficient on the imbibition efficiency is relatively small [22, 23]. Liu et al. [24] conducted experiments on the imbibition characteristics of tight sandstone reservoirs through nuclear magnetic resonance and mercury intrusion porosity testing, which studied the pore-throat limit of imbibition, examined the relationship between pore throat and reservoir parameters, and proposed the imbibition mechanism of tight sandstone reservoirs. Oort et al. made a profound study on oil-water displacement under the action of osmotic pressure and capillary force in shale reservoirs [25, 26] and obtained the displacement characteristics of pore fluid in the process of imbibition [27, 28]. However, most of the above studies are on the mechanism and influencing factors of imbibition. It is basically a qualitative analysis, without clarifying the main control force of different types of pore-throat and without quantifying the contribution of different types of pore-throat to imbibition oil recovery. The innovations of this paper are as follows: based on the imbibition experiment of cores, combined with the technology of nuclear magnetic resonance and pseudo-color processing, this paper is aimed at achieving the purpose of clarifying the main controlling force of different types of pore-throat and quantitatively characterizing the contribution of different types of pore-throat to imbibition oil recovery.

2. Core Characteristics

In order to study the characteristics of different types of pores and throats, two cores with relatively different physical parameters were selected for experimental research. As shown in Table 1, the gas measurement permeability of core 1 is 14.140×10^{-3} μm^2 and the porosity of core 1 is 19.57%, which belongs to the low-porosity and low-permeability core. And core 2 has a gas measurement permeability of 0.059×10^{-3} μm^2 and a porosity of 2.78%, which belongs to the ultra-low-porosity and tight core.

Figure 1 shows the test results of core 1 and core 2 casting thin slices and scanning electron microscopy. Through the analysis, we can get the conclusion as follows: core 1 is mainly composed of siliceous self-formed filling pores, and the metasomatism of silty crystal dolomite is obvious. It mainly replaces volcanic rock debris. The feldspar and volcanic debris dissolve significantly. Kaolinite fills the pores with a small amount of metasomatic debris. Tuffaceous hydration is positive, and it exhibits wavy extinction characteristics under cross-lighting. The eruptive rock of core 1 accounts for 47%, and the total aperture ratio of core 1 is 7.0%. The pores are mainly intercrystalline pores and intergranular pores. The sorting is good. The pore type is mainly intergranular residual pores. The diagenesis authigenic minerals are composed of clay minerals such as kaolinite and are mostly produced in the form of pore filling; core 2 is mainly quartz schist and metamorphic quartzite cuttings. Its main componet is iron calcite with a small amount of illite filling pores. The features of core 2 are poor pore development and poor sorting. Schist of core 2 accounts for 60%, and total porosity of it is 1.2%. The main types are residual pores and fissures between grains. The diagenetic authigenic minerals are composed of clay minerals such as imon mixed layer and illite and are produced in the form of pore filling and cushioning.

Figure 2 is a bar chart of the pore-throat distribution of the two cores. It can be seen that the peak of the pore-throat size distribution of core 1 is 2.5~6.3 μm. Pore-throat size less than 0.1 μm accounts for 17.96%, the pore size of 0.1~1.0 μm accounts for 25.77%, and the proportion of the pore size more than 1.0 μm is 56.27%. The quality of core 2 is poor, most of the pore-throat size is less than 0.1 μm, and the proportion of pore-throat in this part is as high as 66.37%.

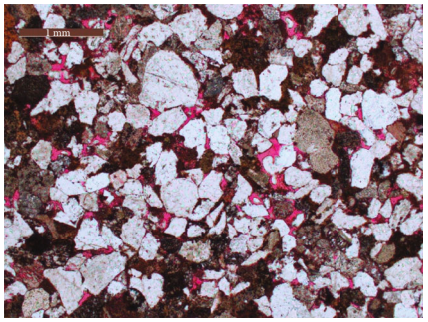
3. Experimental Process

This paper is based on the core imbibition experiments, combined with the technology of the nuclear magnetic resonance (MesoMR23-60H-I medium-size nuclear magnetic resonance spectrometer) and pseudo-color processing to study the imbibition characteristics of different types of pore-throat. The experimental process is as follows:

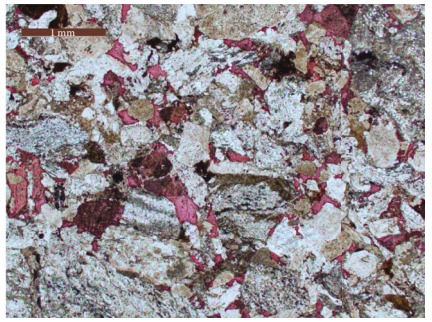
(1) Cut core 1 and core 2 into small cores with a length of about 4.00 cm and a diameter of 2.50 cm; then, clean them simply and maintain the original wettability

(2) Place the core in the displacement device and pressurize it to displace saturated distilled water until it is fully saturated; then, let it stand for a period of time; perform the first NMR scan with a nuclear magnetic resonance instrument on the core, and record the core signal when it is fully saturated with distilled water

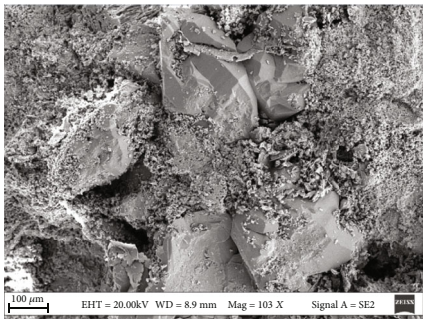
(3) Use fluorine oil to reversely drive the saturated core until no distilled water is produced. The fluorine oil used meets the condition of no nuclear magnetic signal. Let it stand for a period of time to ensure that the core fluid is fully balanced under the action of capillary force. Then, perform the second NMR scan with an MRI machine on the core to record the core signal after reverse saturation of fluorine oil



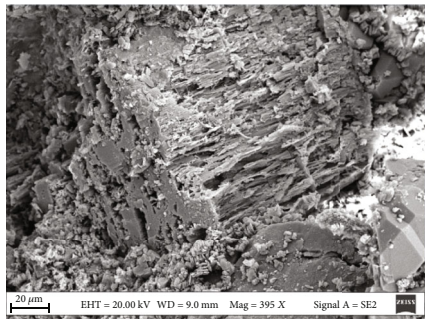
(a) Casting thin section of the core (intergranular pores and dissolved pores)



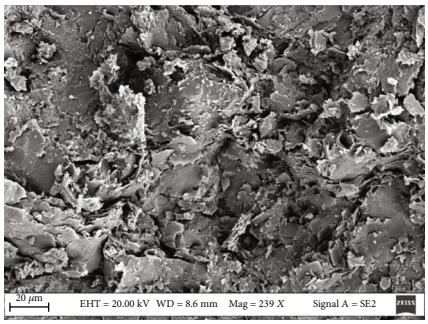
(b) Casting thin section of core 2 (overall dense)



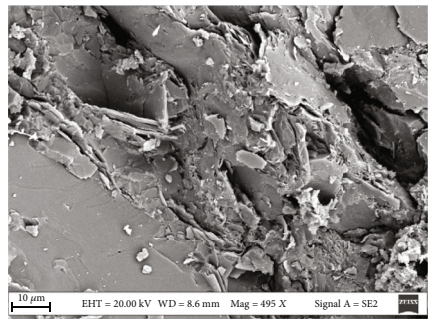
(c) Scanning electron microscope of core 1 (intergranular pores)



(d) Scanning electron microscope of core 1 (detrital particles, dissolved pores)



(e) Scanning electron microscope of core 2 (intergranular residual pores)



(f) Scanning electron microscope of core 2 (intergranular flake illite)

Figure 1: Cast thin section and scanning electron microscope.

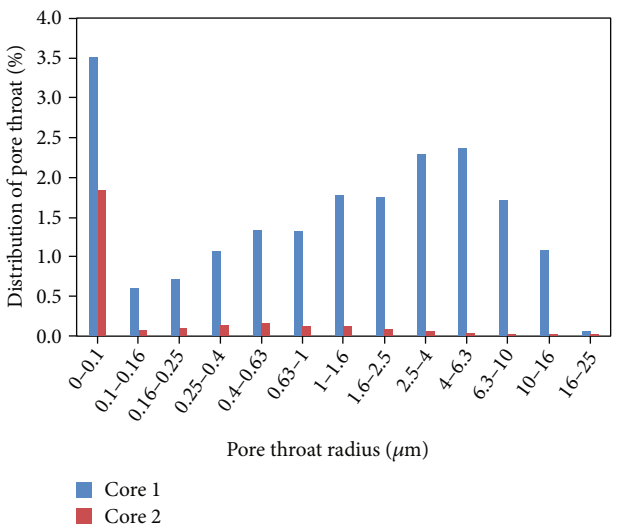


Figure 2: The pore-throat distribution of the core.

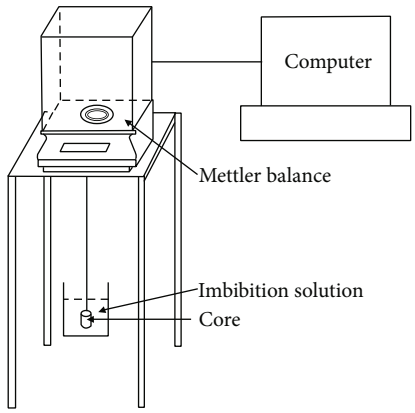
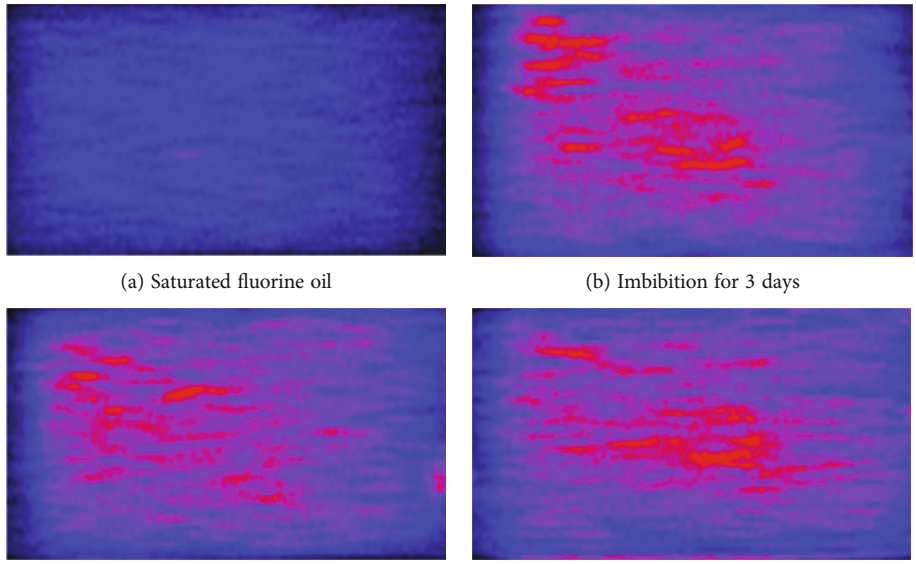


Figure 3: Self-imbibition experimental device.



(a) Saturated fluorine oil (b) Imbibition for 3 days

(c) Imbibition for 5 days (d) Imbibition for 7 days

Figure 4: Pseudo-color image of core 1 (red is the water phase signal).

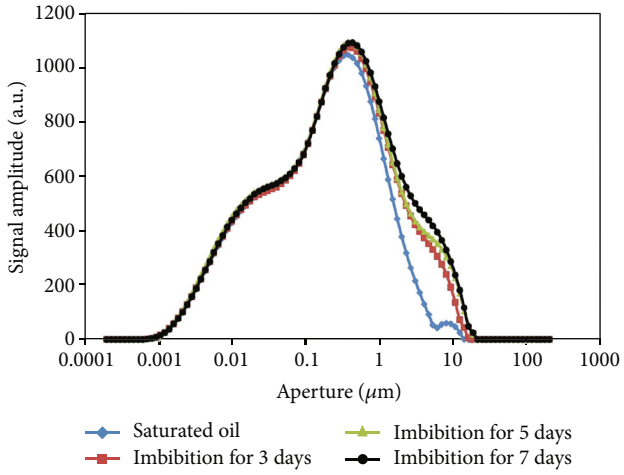


Figure 5: NMR T2 spectrum curve of core 1.

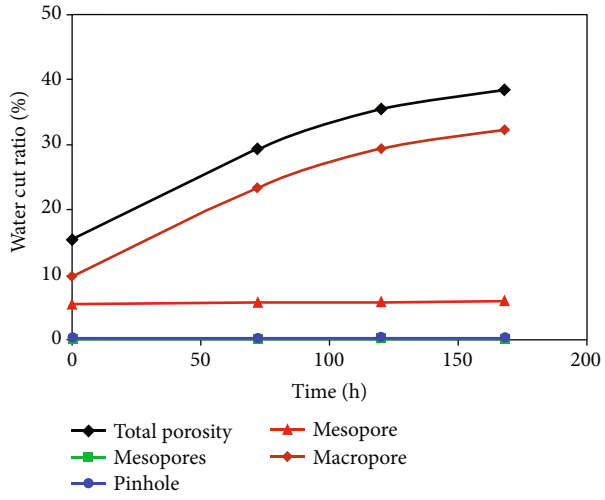


Figure 6: Variation curve of water cut of different types of pore-throat in core 1 with imbibition time.

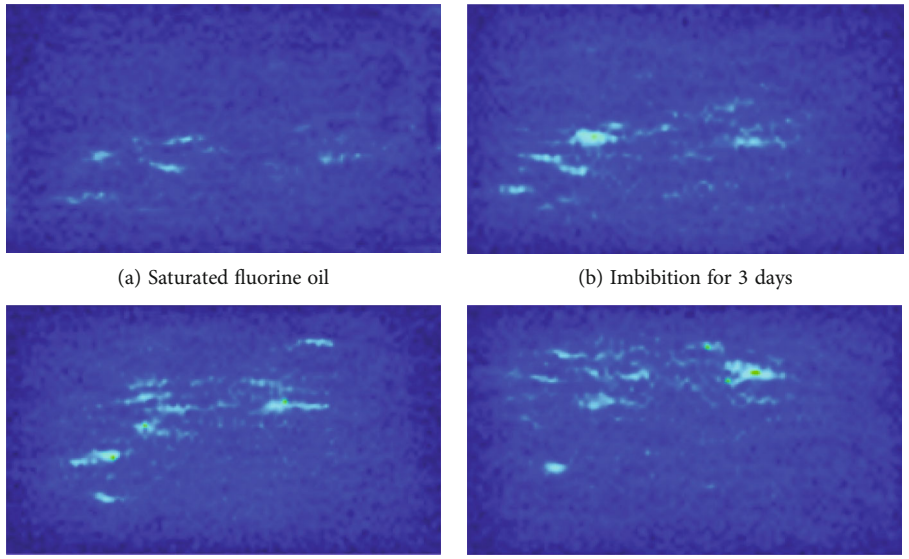
(4) The core self-imbibition experiment system is constructed as shown in Figure 3. The core saturated with fluorine oil was suspended vertically in the distilled water imbibition liquid, and the core is taken out after the imbibition for 3 days, 5 days, and 7 days. Then, the core was subjected to resonance scanning, and core signals are recorded at different time points of imbibition

(5) Process and analyze nuclear magnetic resonance data and pseudo-color visualization images

4. Result Analysis

4.1. Core 1. Figure 4 shows the pseudo-color images generated by scanning of core 1 after being saturated with fluorine oil for 3 days of imbibition, 5 days of imbibition, and 7 days of imbibition. The red area is the area with a strong nuclear magnetic signal. It can be seen from the evolution of the pseudo-color pictures that as the imbibition time increases, the red area gradually spreads, which indicates that during the process of imbibition, the core and the distilled water in the beaker have a significant fluid replacement. Because the core is placed vertically, the left end of the picture is the upper end of the vertical core. It is not difficult to see that the red area spreads from top to bottom along the core axis, indicating that for core 1, the gravity effect is the most obvious during the self-imbibition process.

Figure 5 shows the corresponding NMR T2 spectrum curve of core 1 saturated with fluorine oil for 3 days of imbibition, 5 days of imbibition, and 7 days of imbibition. From the comparison of the curve differences, it is not difficult to see that the fluid imbibition displacement of core 1 mainly occurs in pore-throat of 0.25~20 μm, and fluid replacement



(a) Saturated fluorine oil (b) Imbibition for 3 days

(c) Imbibition for 5 days (d) Imbibition for 7 days

Figure 7: Pseudo-color image of core 2 (bright color is the water phase signal).

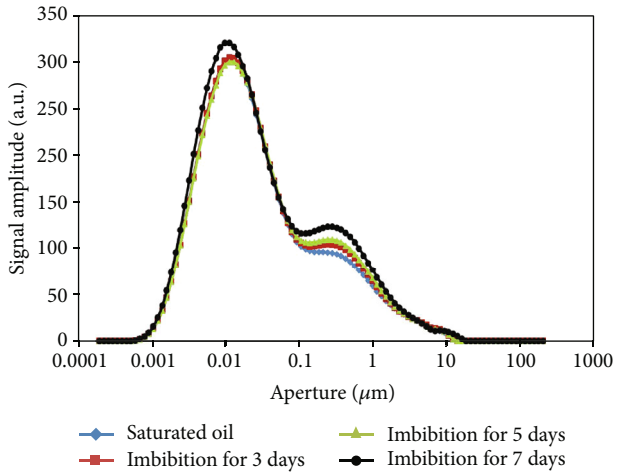


Figure 8: NMR T2 spectrum curve of core 2.

of pore throats of 1.6~10 μm is the most significant. The phenomenon is more obvious in the first 3 days of imbibition replacement, and the efficiency of imbibition replacement is relatively weakened in the next 4 days.

In order to quantitatively characterize the contribution of different types of pore-throat to imbibition oil recovery, the core pore throats are subdivided into four types, including micropores (pore diameter $\leq$ 0.025 μm), pinhole (pore diameter: 0.025~0.1 μm), mesopores (pore diameter: 0.1~1 μm), and macropores (pore diameter > 1 μm). Figure 6 shows the change curve of water cut of different types of pore-throat of core 1 with time. Statistics show that the total water cut of core 1 increased from 15.31% to 38.37%, and the 168-hour imbibition recovery rate of the core was 23.06%. The macroporous water cut increased from the initial 9.63% to 32.29%, which contributed 98.26% of the total oil production by imbibition; the water cut of mesopores increased from the initial 5.35% to 5.75%, which contributed 1.73% of the total oil production by imbibition; the contribution of micropores and pinholes was negligible.

4.2. Core 2. Figure 7 shows the pseudo-color images generated by scanning of core 2 after being saturated with fluorine oil for 3 days of imbibition, 5 days of imbibition, and 7 days of imbibition. It can be seen that with the increase in the imbibition time, the bright-colored area diffuses irregularly. Compared with the axial direction of the core, the lateral movement of the fluid is more obvious, which indicates that the capillary force plays a major role in core 2.

Figure 8 shows the corresponding NMR T2 spectrum curve of core 2 saturated with fluorine oil for 3 days of imbibition, 5 days of imbibition, and 7 days of imbibition. From the comparison of the curves, it is not difficult to see that there are mainly two intervals of pore-throat for the fluid imbibition displacement of core 2, namely, 0.001~0.017 μm interval and 0.08~2.5 μm interval. The fluid replacement in 0.08~2.5 μm pore-throat is relatively obvious. What is more, the imbibition replacement is overall stable; the fluid replacement phenomenon is more obvious in the later 2 days. In the first 5 days, the efficiency of imbibition replacement was relatively weakened, indicating that the time of imbibition replacement in cores with poor pore throats was relatively slower.

Figure 9 shows the change curve of water cut of different types of pore-throat of core 2 over time. Statistics show that the total water cut of core 2 increased from 40.42% to 49.33%, and the 168-hour imbibition recovery rate of core 2 was 8.91%. The macroporous water cut increased from the initial 27.81% to 33.63%, contributing 65.35% of the total imbibition oil production; the mesoporous water cut increased from the initial 10.29% to 13.29%, contributing 33.67% of the total imbibition oil production. The contribution of micropores and pinholes are both around 0.5%, and mesopores have been assumed to have more contribution.

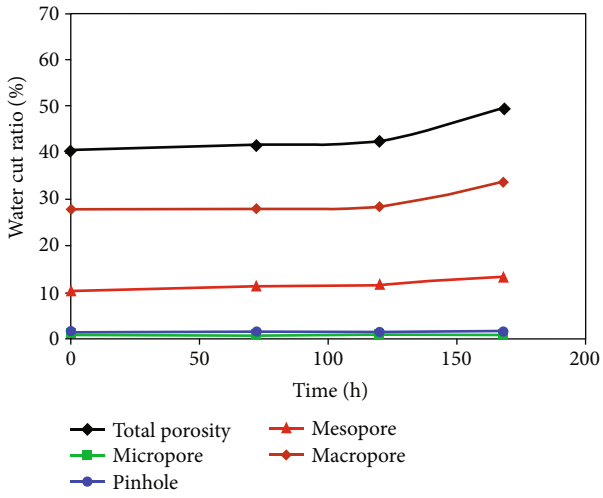


Figure 9: Variation curve of water cut of different types of pore-throat in core 2 with imbibition time.

5. Conclusions

(1) For reservoirs with high permeability, gravity is the main controlling force in the process of imbibition. The fluid replacement mainly occurs in the early period of imbibition, and the effect is relatively weak in the later period. Macropores (pore diameter $> 1\,\mu$m) contribute most of the imbibition, and mesopores (pore diameter: 0.1~1 μm) have a weak contribution. The contribution of micropores and pinholes can be ignored

(2) For reservoirs with low permeability, capillary force is the main controlling force in the process of imbibition. The fluid replacement mainly occurs in the later period of imbibition. Early fluid replacement is relatively weak. Macropores (pore diameter $> 1\,\mu$m) contribute most of the imbibition, and mesopores (pore diameter: 0.1~1 μm) contribute a small part of the imbibition. The contribution of micropores and pinholes can be ignored

(3) Regardless of whether the main controlling force is gravity or capillary force, macropores (pore diameter $> 1\,\mu$m) and mesopores (pore diameter 0.1~1 μm) are the main sources of contribution to imbibition recovery, indicating that oil content and connectivity are two crucial factors that affect imbibition recovery

Data Availability

It will be available based on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the National Natural Science Foundation of China (Nos. 51804257, 52004220, 51934005, and 51874242), the Open Fund of State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (Chengdu University of Technology) (PLC20210316), the PetroChina Innovation Foundation (No. 2020D-5007-0202), and the Innovation and Practice Training Program for Graduate Students (YCS21211006).

References

[1] H. Zheng, R. Q. Liao, N. Cheng, and S. Z. Shi, "Microscopic mechanism of fracturing fluid imbibition in stimulated tight oil reservoir," *Journal of Petroleum Science and Engineering*, vol. 202, article 108533, 2021.

[2] K. Zhang, K. Sebakhy, K. Wu et al., "Future trends for tight oil exploitation," in *SPE 175699, Paper presented at the SPE North Africa Technical Conference and Exhibition*, Cairo, Egypt, September 2015.

[3] H. L. Chen, L. R. Lucas, L. A. D. Nogaret, H. D. Yang, and D. E. Kenyon, "Laboratory monitoring of surfactant imbibition with computerized tomography," *SPE Reservoir Evaluation and Engineering*, vol. 4, no. 1, pp. 16–25, 2001.

[4] X. L. Peng, X. Z. Wang, X. Zhou, Z. Y. Lin, F. H. Zeng, and X. L. Huang, "Lab-on-a-chip systems in imbibition processes: a review and applications/issues for studying tight formations," *Fuel*, vol. 306, article 121603, 2021.

[5] L. Cuice, B. Bourbiaux, and F. Kalaydjlan, "Oil recovery by imbibition in low-permeability chalk," *SPE Formation Evaluation (Society of Petroleum Engineers)*, vol. 9, no. 3, pp. 200–208, 1994.

[6] P. M. Oen, M. E. Jensen, and A. A. Barendregt, "Skjold field, Danish North Sea: early evaluations of oil recovery through water imbibition in a fractured reservoir," *SPE Reservoir Evaluation and Engineering*, vol. 3, no. 1, pp. 17–22, 1988.

[7] J. Wang, H. Q. Liu, J. Xia et al., "Mechanism simulation of oil displacement by imbibition in fractured reservoirs," *Petroleum Exploration and Development*, vol. 44, no. 5, pp. 805–814, 2017.

[8] X. Xu, Y. J. Wan, X. Z. Li et al., "Microscopic imbibition characterization of sandstone reservoirs and theoretical model optimization," *Scientific Reports*, vol. 11, no. 1, pp. 8509–8509, 2021.

[9] S. Liu, J. Ni, X. L. Wen et al., "A dual-porous and dual-permeable media model for imbibition in tight sandstone reservoirs," *Journal of Petroleum Science and Engineering*, vol. 194, article 107477, 2020.

[10] Z. Y. Wang, H. X. Li, X. M. Lan, K. Wang, Y. F. Yang, and V. Lisitsa, "Formation damage mechanism of a sandstone reservoir based on micro-computed tomography," *Advances in Geo-Energy Research*, vol. 5, no. 1, pp. 25–38, 2021.

[11] Q. H. Feng, S. Q. Xu, X. D. Xing, W. Zhang, and S. Wang, "Advances and challenges in shale oil development: a critical review," *Advances in Geo-Energy Research*, vol. 4, no. 4, pp. 406–418, 2020.

[12] H. O. Yildiz, M. Gokmen, and Y. Cesur, "Effect of shape factor, characteristic length, and boundary conditions on spontaneous imbibition," *Journal of Petroleum Science and Engineering*, vol. 53, no. 3-4, pp. 158–170, 2006.

[13] G. Mason, H. Fischer, N. R. Morrow et al., "Oil production by spontaneous imbibition from sandstone and chalk cylindrical cores with two ends open," *Energy and Fuels*, vol. 24, no. 2, pp. 1164–1169, 2010.

[14] C. H. Lyu, Z. F. Ning, M. Q. Chen, and Q. Wang, "Experimental study of boundary condition effects on spontaneous imbibition in tight sandstones," *Fuel*, vol. 235, pp. 374–383, 2019.

[15] Z. M. Yang, X. W. Liu, H. B. Li, Q. H. Lei, Y. T. Luo, and X. Y. Wang, "Analysis on the influencing factors of imbibition and the effect evaluation of imbibition in tight reservoirs," *Petroleum Exploration and Development*, vol. 46, no. 4, pp. 779–785, 2019.

[16] Y. Jiang, Y. Shi, G. Q. Xu et al., "Experimental study on spontaneous imbibition under confining pressure in tight sandstone cores based on low-field nuclear magnetic resonance measurements," *Energy and Fuels*, vol. 32, no. 3, pp. 3152–3162, 2018.

[17] A. Habibi, H. Dehghanpour, M. Binazadeh, D. Bryan, and G. Uswak, "Advances in understanding wettability of tight oil formations: a Montney case study," *SPE Reservoir Evaluation and Engineering*, vol. 19, no. 4, pp. 583–603, 2016.

[18] L. Yang, N. H. Dou, X. B. Lu et al., "Advances in understanding imbibition characteristics of shale using an NMR technique: a comparative study of marine and continental shale," *Journal of Geophysics and Engineering*, vol. 15, no. 4, pp. 1363–1375, 2018.

[19] A. Bertoncello, J. Wallace, C. Blyton, M. Honarpour, and C. S. S. Kabir, "Imbibition and water blockage in unconventional reservoirs: well-management implications during flowback and early production," *SPE Reservoir Evaluation and Engineering*, vol. 17, no. 4, pp. 497–506, 2014.

[20] Z. L. Chen, Z. F. Ning, X. F. Yu, Q. Wang, and W. T. Zhang, "New insights into spontaneous imbibition in tight oil sandstones with NMR," *Journal of Petroleum Science and Engineering*, vol. 179, pp. 455–464, 2019.

[21] L. H. Gao, Z. M. Yang, and Y. Shi, "Experimental study on spontaneous imbibition characteristics of tight rocks," *Advances in Geo-Energy Research*, vol. 2, no. 3, pp. 292–304, 2018.

[22] F. H. Zeng, Q. Zhang, J. C. Guo, Y. Meng, X. Z. Shao, and Y. J. Zheng, "Capillary imbibition of confined water in nanopores," *Capillarity*, vol. 3, no. 1, pp. 8–15, 2020.

[23] F. P. Lai, Z. P. Li, T. T. Zhang, A. Q. Zhou, and B. Gong, "Characteristics of microscopic pore structure and its influence on spontaneous imbibition of tight gas reservoir in the Ordos Basin, China," *Journal of Petroleum Science and Engineering*, vol. 172, pp. 23–31, 2019.

[24] Y. F. Liu, Y. H. Shi, L. Liu, X. C. Yan, D. S. Zhou, and S. Liu, "Determination of the pore-throat limits for water imbibition in tight sandstone reservoirs through NMR analysis," *Journal of Geophysics and Engineering*, vol. 16, no. 1, pp. 253–261, 2019.

[25] X. P. Li, H. Abass, T. W. Teklu, and Q. Cui, "A shale matrix imbibition model-interplay between capillary pressure and osmotic pressure," in *SPE181407, Paper presented at the SPE Annual Technical Conference and Exhibition*, Dubai, UAE, 2016.

[26] E. V. Oort, M. Ahmad, R. Spencer, and N. Legacy, "ROP enhancement in shales through osmotic processes," in *SPE/IADC 173138,Paper presented at the Drilling Conference and Exhibition*, London, England, UK, 2015.

[27] S. Morsy and J. J. Sheng, "Imbibition characteristics of the Barnett shale formation," in *SPE167698, Paper presented at the SPE Unconventional Resources Conference*, Woodlands, Texas, USA, 2014.

[28] X. Liu, D. S. Zhou, L. Yan, S. Liu, and Y. F. Liu, "On the imbibition model for oil-water replacement of tight sandstone oil reservoirs," *Geofluids*, vol. 2021, Article ID 8846132, 7 pages, 2021.

Hindawi
Geofluids
Volume 2022, Article ID 1760065, 9 pages
https://doi.org/10.1155/2022/1760065

Research Article

Probabilistic Evaluation of Hydraulic Fracture Performance Using Ensemble Machine Learning

Xiaoping Xu,[1] Xianlin Ma,[2] and Jie Zhan[2]

[1]*Shengli Oilfield, SINOPEC, Dongying 257001, China*
[2]*School of Petroleum Engineering, Xi'an Shiyou University, Xi'an 710065, China*

Correspondence should be addressed to Xianlin Ma; xianlinm@126.com and Jie Zhan; zhanjie@xsyu.edu.cn

Received 7 January 2022; Accepted 30 March 2022; Published 16 April 2022

Academic Editor: Yating Wang

Evaluation of hydraulic fracture (HF) performances is critical to develop unconventional resources such as tight oil and gas. We present a probabilistic evaluation approach that integrates ensemble machine learning with Monte Carlo simulation. In the method, we employ the ensemble learning to develop a predictive model between well productivity and its influential factors including both geological properties and HF treatment parameters. Next, coupling the built prediction model with Monte Carlo simulation, an empirical cumulative probability distribution of the well productivity is generated. The well HF performance is assessed by estimating its cumulative probability value. The proposed method is applied to evaluate the HF performances in a developed block of the eastern Sulige region. The study shows that 19% of the wells were fractured with good quality and 55% of the wells were fractured with average quality, while the rest were stimulated with poor quality. The evaluations provide a guideline for optimization of HF designs of wells that have not been hydraulically stimulated in the region.

1. Introduction

Tight sand gas has become a new exploration target and important source of natural gas supply in China [1–6]. Sulige gas field, located in the central part of the Ordos Basin, is the largest tight gas field with estimated gas reserves of 18.85 Tcf in China [7–9]. The eastern Sulige is one of the main regions of natural gas exploration and development in the gas field [10, 11]. Due to the low quality of rock properties such as permeability and porosity in the gas-bearing formations, hydraulic fracturing stimulations were carried out so that natural gas can be produced economically from the tight formation.

The hydraulic fracturing is a process which injects large amounts of water, proppants (sand), and additives via a wellbore at high pressure into the low permeable formations to break rock and create cracks (hydraulic fractures). The injected proppants keep fractures open [12–14]. These fractures could extend several hundred feet away the wellbore and form high permeable flow paths, which ease the natural gas flow from the tight formation toward the wellbore and enhance the gas recovery. In the past several years, hundreds of wells were hydraulically stimulated and put on production in the eastern Sulige region. However, the significant differences in the gas production have exhibited in the fractured wells [15]. To further develop the gas field, it is necessary to evaluate the hydraulic fracture performances and to identify the best hydraulic fracturing practices to maximize the well gas production in the field.

Upon completing of a HF job, well production performance such as absolute open flow potential (AOFP) or cumulative gas production is employed as an indicator for the hydraulic fracture performance of the well. The well production performance is controlled by multiple variables. In this work, we consider geological properties and fracturing treatment parameters. The geological properties include gas saturation, pay thickness, porosity, matrix permeability, and presence of nature fractures. The treatment parameters mainly contain fluid injection volumes, injection rates and pressure, and proppant volumes, which influence fracture geometries. To conduct the performance evaluation, it is essential to build a predictive model to quantify the well

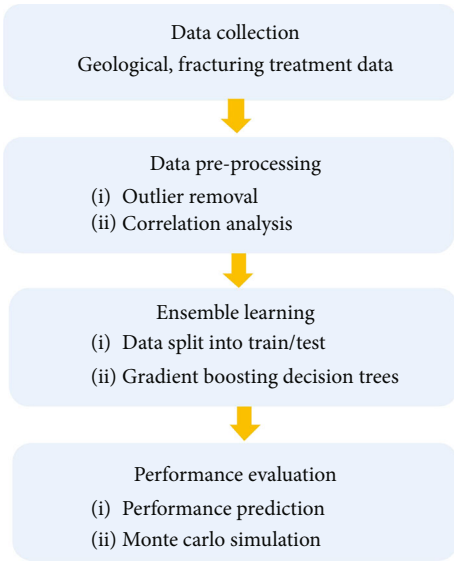


Figure 1: Workflow of HF performance evaluation.

production performance with these influential factors. A numerical reservoir simulator is usually employed to build such model by simulating natural gas flow in both tight matrix and the fractures [16–19]. However, the accuracy of the hydrocarbon production predictions highly relies on the characterization of the tight gas reservoirs and representation of induced fracture geometry. Due to limited formation data and subsurface complexities, it is impossible to quantify the properties precisely. Meanwhile, the underlying flow mechanism is so complex in the tight formations that multiple simplified assumptions have been made in the mechanistic modelling [20]. These challenges have made that the estimated hydrocarbon production from the reservoir simulations often differs from actual production data [21]. In addition, the numerical simulation runs are computationally intensive by solving large partial deferential equations and make them unfeasible for the probabilistic evaluation of the HF performance, where hundreds of simulations are often required. Numerous data have been collected from various sources and available in the eastern Sulige region. A machine learning method could provide potential solutions to overcome the drawbacks.

In recent years, the machine learning method has become a powerful tool for predicting the production performance. For instance, Nejad et al. [22] developed a neural network model to examine the effect of the completion and fracture treatment designs on gas production in the Eagle Ford Shale. Mohaghegh [23] proposed the concept of shale analytics to describe the applications of artificial intelligence to shale gas development. The shale analytics was applied to investigate how different reservoir and completion parameters affect gas production in several shale plays of the United States. Montgomery and O'Sullivan [24] developed a spatial error model and regression-Kriging to forecast tight oil production using a large well dataset from the Williston Basin in North Dakota. Wang and Chen [25] investigated the performances of four machine-learning algorithms in forecasting first-year oil production of Montney tight reservoirs and concluded that random forest outperformed other algorithms. Porras et al. [26] developed a random forest model to predict first-year oil production using geological and completion design parameters. The developed model was used to evaluate hydraulic fracture performance of the horizontal wells in the Viking Formation, Canada. Although the machine learning techniques have shown promising results in the production prediction in the above studies, these studies did not consider the uncertainties related to the complexities of the HF process. The quantification of these uncertainties can provide valuable information to assess the HF performances.

In this study, we develop a probabilistic approach to evaluate the HF performance. Our approach integrates ensemble learning with Monte Carlo simulations to consider the uncertainties caused by the HF treatment complexities. The approach was inspired by Mohaghegh's combination of neural networks and the Monte Carlo simulations [23]. We applied the developed approach to obtain an empirical cumulative probability distribution of the well productivity. The well HF performance is assessed by estimating its probability value.

2. Methodology

The probabilistic evaluation workflow of the HF performance consists of four main steps, as shown in Figure 1.

(Step 1) Data collection: collect raw data from a database. The independent variables include geological properties and hydraulic fracturing treatment parameters. The target variable represents post-treatment well production performance.

(Step 2) Data preprocessing: make raw data ready for building a predictive model. The data preprocessing is composed of two parts: data cleaning and feature selection.

(Step 3) Ensemble learning: build a model to predict the well production performance using ensemble machine learning.

(Step 4) Probabilistic evaluation: combine the predictive model with the Monte Carlo simulation to assess the HF performance.

2.1. Data Collection and Preprocessing. We extracted 743 vertical fractured wells from a private database. These wells were chosen without missing values and from the same developed block in the eastern Sulige. As one of China's natural gas development key areas, the eastern Sulige gas field stretches from the Ordos district in the Inner Mongolia Autonomous Region to the Yulin district of Shaanxi province in China. The reservoir rocks of the gas field mainly consist of the Upper Paleozoic fluvial and deltaic sands [27]. The main gas-producing layers are Permian-age Shihezi formation, Shanxi formation, Benxi formation, and

Table 1: Lists of modelling variables.

Type	Parameters	Unit	Abbreviation
Geological properties	Formation thickness	m	TH
	Well true vertical depth	m	TVD
	Formation porosity	%	PO
	Matrix permeability	mD	PERM
	Gas saturation	%	SG
Fracture treatment	Pad fluid volume	m^3	PFV
	Slurry fluid volume	m^3	SFV
	Proppant fluid ratio	%	PFR
	Total fluid volume	m^3	TFV
	Average injection rate	m^3/min	AIR
	Average injection pressure	MPa	AIP
Production	Absolute open flow potential	10^4m^3/day	AOFP

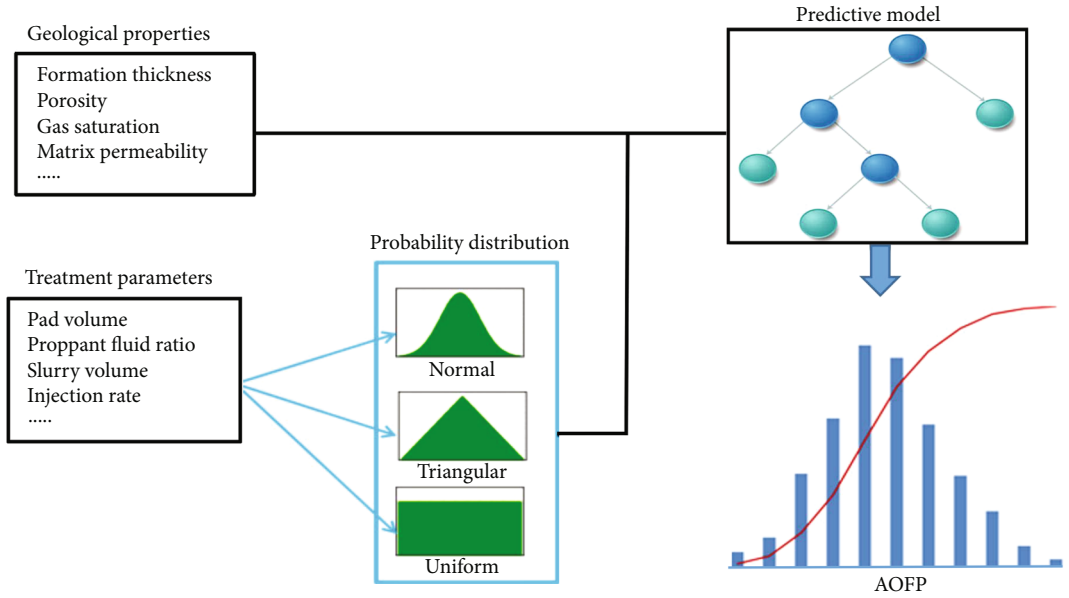


Figure 2: Probabilistic evaluation of a fractured well through obtaining cumulative probability distribution of AOFP (modified from [23])

Majiagou formation. The 8th member of Shihezi formation (He8) is one of the most gas productive zones with formation thickness between 45 m and 60 m. The average depth of the tight formations in the study area ranges from 2300 m to 3800 m. The reservoir porosity ranges from 5% to 15%, and the matrix permeability varies from 0.5 to 20 mD [28]. The multistage hydraulic fractured wells have been applied recently in the field. However, the field is still dominated by vertical well development, accounting for more than 80% of the total number of wells [29]. The vertical wells were drilled into the formations He8 and 1st member of Shanxi formation (Shan1) between 2009 and 2016. Based on geological interpretations, most of the vertical wells have several gas-bearing layers with average thickness less than 5 m. To stimulate the separate layer, multilayer staged fracturing techniques such as mechanical packers, casing sleeves, and coiled tubing were developed to enhance fracturing efficiency and increase single-well production. However, the fractured wells could only keep stable gas production for 1 to 1.5 years; then the production declined rapidly. Therefore, the gas flow rates were allocated properly to maintain a relatively long stable production [30].

For each vertical well, there are twelve variables listed in Table 1.

The target variable is absolute open flow potential (AOFP) to quantify the hydraulic fracture performance. The number of independent variables is eleven. Two kinds of independent variables are identified the most important to the production performance of hydraulically fractured wells: geological properties and HF treatment parameters. The geological properties include formation thickness, true vertical depth (TVD), porosity, matrix permeability, and gas saturation. The HF treatment parameters such as fluid volumes, injection rates, and pressure determine a fracture geometry and conductivity, which play an important role in natural gas flow.

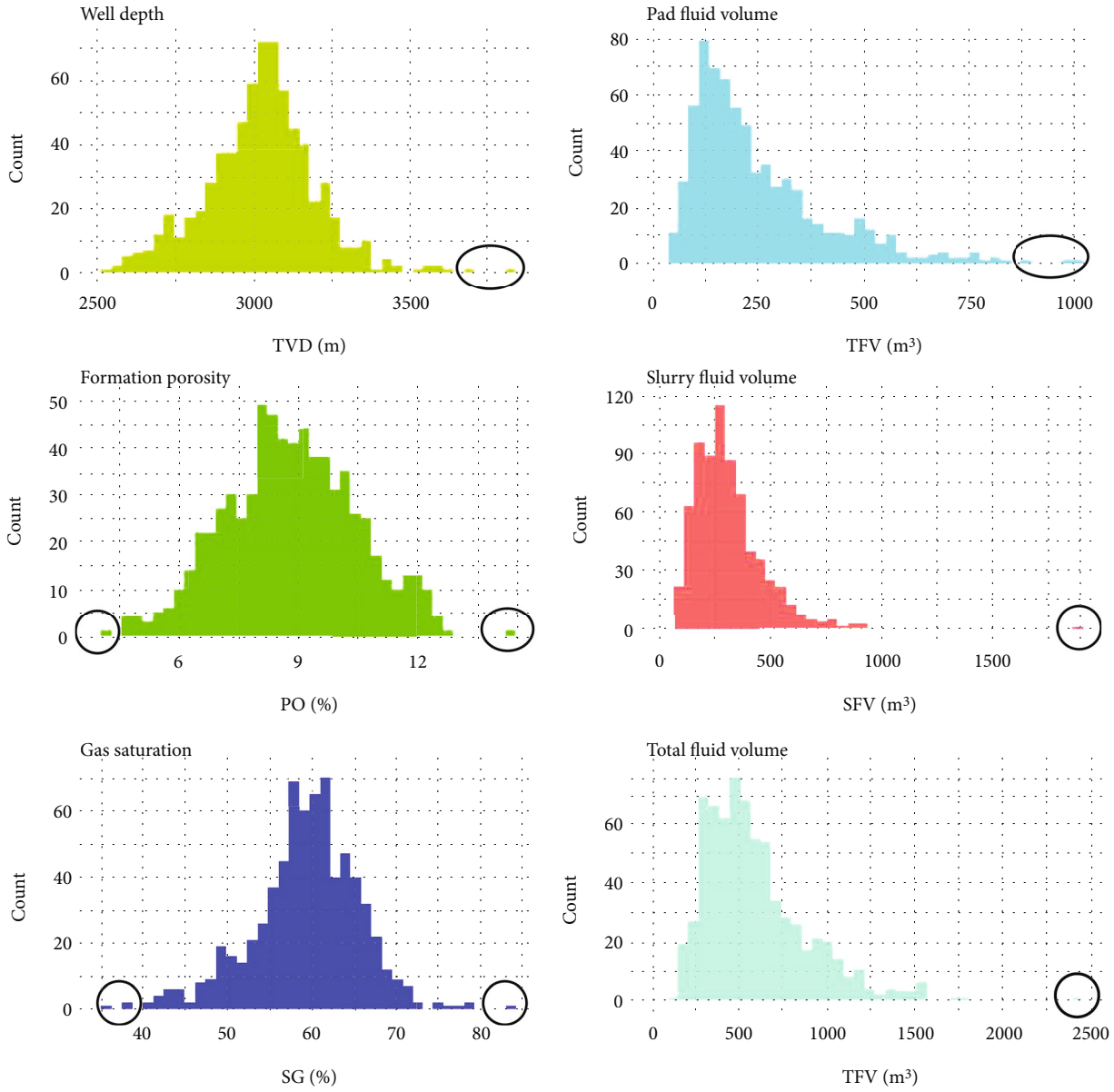


Figure 3: Histograms of the main variables.

The collected raw data usually include missing data, incorrect formats, and abnormal data, and the direct usage of the raw data affects later predictive modelling. Data preprocessing is employed to solve the issues, which mainly includes data cleaning, data transformation, and feature selection, which are described in Section 3.

2.2. Ensemble Learning. The ensemble learning method is an advancement in machine learning technique by building multiple learners and combining the outputs of these learners to obtain robust predictions [31]. It has been confirmed that the ensemble learner has better performances than a single learner and has already been applied in many regression and classification problems [32]. Boosting is one of effective ensemble learning methods. Friedman [33, 34] regarded the boosting as the optimization of a loss function and introduced the concept of gradient boosting. Gradient boosting decision trees (GBDT) is an ensemble algorithm to combine many decision tree models, where each tree is built to minimize the residual error of the previous tree iteratively. The final prediction is integrated from the outputs of all trees. The GBDT algorithm is described briefly as follows.

Assuming that x is a set of features (geological properties and HF treatment parameters) and $F_M(x)$ is a predicted function of the target variable y (absolute open flow potential). Given training data set $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, a GBDT model $F_M(x)$ is built as the summation of M additive functions $f_m(x)$ based on decision trees $t(x; \mu_m)$ [35, 36]:

$$F_M(x) = \sum_{m=1}^{M} f_m(x) = \sum_{m=1}^{M} \theta_m t(x; \mu_m), \quad (1)$$

Table 2: Statistical properties of the variables.

Parameter	Range	Mean	Standard deviation
Absolute open flow (10^4 m^3/day)	0.4~21.5	5.5	3.2
Formation thickness (m)	1.7~38.0	14.6	6.6
True vertical depth (m)	2534~3699	3025	175.7
Formation porosity (%)	4.8~12.7	8.8	1.7
Matrix permeability (mD)	0.1~5.1	0.8	0.7
Gas saturation (%)	30.5~78.5	59.2	6.3
Pad fluid volume (m^3)	39.0~830.0	249.8	158.5
Slurry fluid volume (m^3)	88.2~933.7	310.5	153.2
Proppant fluid ratio (%)	3.1~45.6	27.4	9.3
Total fluid volume (m^3)	144.2~1560.0	589.7	292.6
Injection rate (m^3/min)	1.4~11.0	3.2	1.1

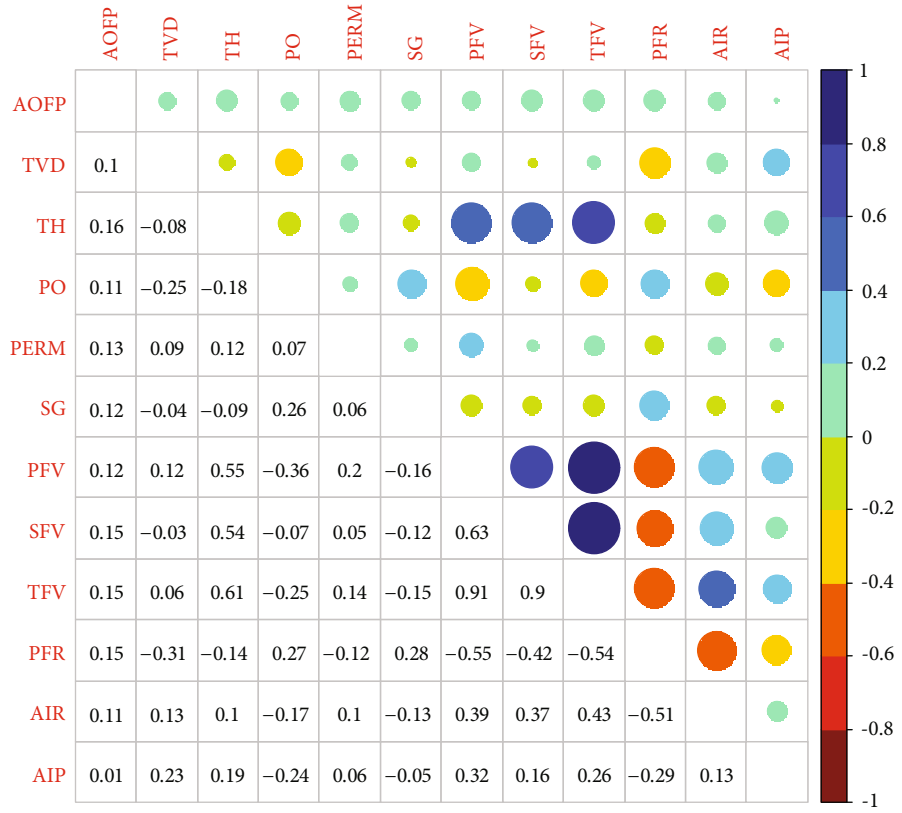


Figure 4: The Pearson covariance matrix between features.

where μ_m is the mean of split locations and the terminal nodes of a tree $t(x;\mu_m)$ and θ_m is estimated by minimizing a specified loss function $L(y, F(x)) = (y - F(x))^2$:

$$f_0(x) = \gamma = \text{argmin}_\gamma \sum_{i=1}^{N} L(y_i, \gamma). \tag{2}$$

The number of iteration m starts from 1 to M, and negative gradient g_{im} in the current prediction model is calculated by

$$g_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}. \tag{3}$$

The tree $t(x;\mu_m)$ is used to approximate the negative gradient, and a descent step size θ_m is computed by

$$\theta_m = \text{argmin}_\theta \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \theta t(x;\mu_m)). \tag{4}$$

Table 3: Optimal hyperparameter values of the GBDT.

Hyperparameter	Name	Ranges	Optimal values
No. of trees	N_estimators	1~200	110
Max. depth	max_depth	2~8	3
No. of features to split	max_features	sqrt; auto	sqrt
Min. number of samples	min_samples_split	2~12	4
Learning rate	learning_rate	0.08~0.13	0.11

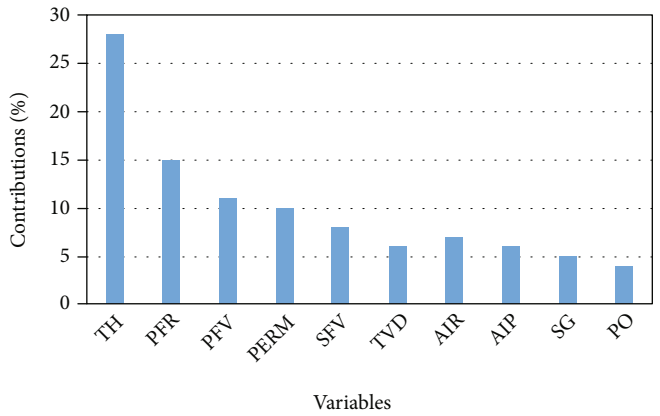


Figure 5: Relative variable importance.

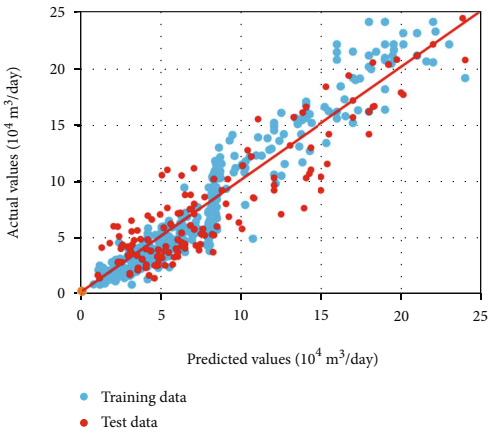


Figure 6: Comparison of actual and predicted AOFP for training and test data.

The model is updated based on Equations (2)–(4):

$$F(x) = F_{m-1}(x) + \theta_m t(x; \mu_m). \tag{5}$$

The GBDT has a few advantages, including the ability to cope with skewed variables, computational robustness, and high scalability [37]. More details of the algorithm can be found from Zhu et al. [27].

2.3. Probabilistic Evaluation. Due to the complexity of hydraulic fracturing process, the production prediction has some uncertainty, which strongly affects the HF performance evaluation.

Table 4: Properties of well Sudong42-48.

Parameter	Value
Formation thickness (m)	17.2
True vertical depth (m)	3138.9
Formation porosity (%)	7.2
Matrix permeability (mD)	0.4
Gas saturation (%)	54.3
Pad fluid volume (m^3)	425
Slurry fluid volume (m^3)	380
Proppant fluid ratio (%)	9.7
Injection rate (m^3/min)	3.5
Injection pressure (MPa)	67.5
Absolute open flow (10^4 m^3/day)	8.5

Monte Carlo simulations can be used to quantify the prediction uncertainty by estimating an empirical cumulative probability distribution of the well productivity instead of a determined value. As shown in Figure 2, we fixed the values of the geological properties of a selected well and assumed the fracture treatment parameters as random variables to follow a certain probability distribution such as triangular or Gaussian or uniform. The ranges, means, and variances of these distributions are estimated from the dataset. We run the predictive model thousands of times by randomly sampling from the given distributions. At the end of the runs, an empirical cumulative probability distribution of the selected well AOFP is generated. Through the probability distribution, the HF performance of the well could be evaluated by estimating its probability value. The details are described in the next section.

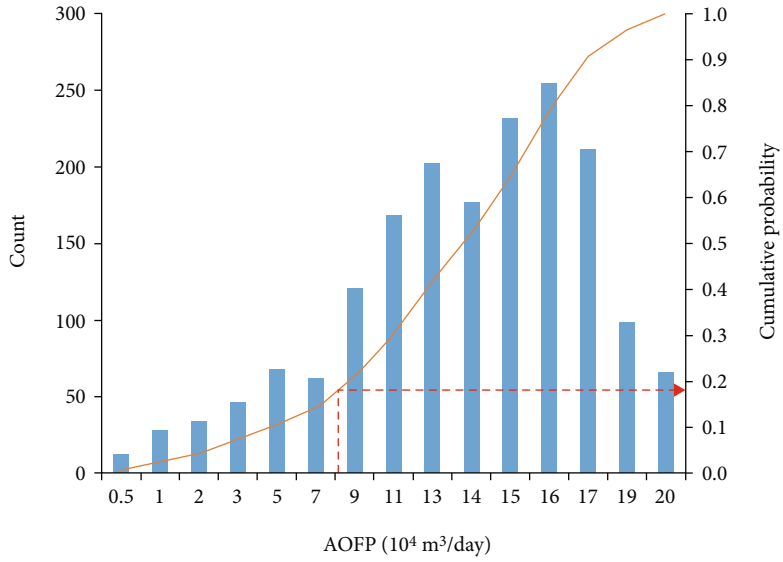


Figure 7: Histogram and probability distribution of a well AOFP.

3. Results

3.1. Outlier Removal. Figure 3 displays the presence of outliers through histograms of the main input variables. The outliers in the data are circled in black for the variables. We also applied Rosner's test to verify the outliers identified in the histograms [38]. A total of 39 wells with the outliers were detected from the dataset and removed. 704 wells were kept to the following predictive modelling.

Table 2 summarizes the range and statistical properties of the independent and dependent variables. The absolute open flow potential per well ranges from 0.4 to 21.5 (10^4 m^3/day), averaging 5.5×10^4 m^3/day. The total injection fluid volume and proppant fluid ratio per well are averaged to be 589.7 m^3 (27.4%), respectively.

3.2. Correlation Analysis. Figure 4 shows the Pearson covariance matrix to quantify the degree of linear correlation among the variables. It is noted that all treatment parameters have a positive correlation with the well productivity (AOFP). The increasing in the fracturing fluid volumes enlarges the reservoir stimulated zones, which creates more contact area between the wellbore and reservoir. As expected, there is the multicollinearity among the pad fluid volume (PFV), slurry fluid volume (SFV), and total fluid volume (TFV) since total fluid volume consists of the pad fluid volume and slurry fluid volume. Therefore, the total fluid volume will be dropped in the following predictive modelling. In addition, the geological properties also have a positive correlation with the AOFP. The formation thickness has the largest value of the Pearson coefficient. More sophisticated feature selection methods including stepwise regression and recursive feature elimination are being investigated.

3.3. Production Forecasting. The 704 wells in the dataset were split into training and test data sets in the ratio of 80 : 20. The main hyperparameters for a GBDT model include learning rate, number of trees, minimum number of samples required at a leaf node, maximum depth, and number of features for the best split. The hyperparameter tuning were performed on the training data using grid searching combined with five-fold cross-validation (CV). The optimal hyperparameter values are listed in Table 3.

Figure 5 shows the results of the relative importance of 10 input variables on the AOFP. The formation thickness and proppant fluid ratio are two most important variables, followed by the pad fluid volume and matrix permeability. The variable importance indicts that the formation thickness has a significant impact on the well productivity. Therefore, when fracturing a gas well, we need to evaluate its formation thickness and matrix permeability. In terms of the fracturing operation, we may add more proppant to enhance the fracture conductivity and increase the pad fluid volume to improve the fracturing effectiveness.

Figure 6 compares the actual and predicted AOFP for the training and test data sets using the built GBDT model.

In the figure, the data points are grouped along the 45-degree straight line, and the values of coefficient of determination R^2 for the training and test data set are calculated to be 0.91 and 0.74, respectively. The results show that the predicted AOFP have a good match with the actual AOFP values for both the training and test data sets, which indicates the developed AOFP forecasting model is robust for the evaluation of the hydraulic fracture performance.

3.4. Well Hydraulic Performance Evaluation. We choose well Sudong42-48 to assess its hydraulic fracturing performance as an example. Table 4 shows the input variable and AOFP values of this well.

We fixed the values of the geological properties of this well and assumed the fracture treatment parameters as random variables to follow a certain probability distribution such as triangular or Gaussian or uniform. We run the predictive model thousands of times by sampling the treatment

Table 5: Evaluation criteria.

Hydraulic fracture performance	Cumulative probability
Good	>0 .6
Fair	[0.3, 0.6]
Poor	<0.3

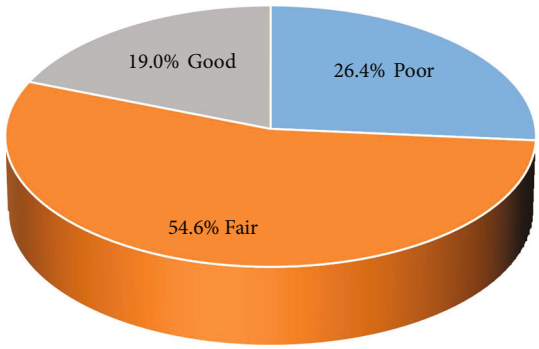


Figure 8: Evaluation results of hydraulic fracture performance in the block of the eastern Sulige region.

parameters from the given distributions. An estimated AOFP was obtained from each run. At the end of the runs, a histogram and empirical cumulative probability distribution of the selected well were generated as shown in Figure 7. The well hydraulic performance is assessed by estimating its cumulative probability based on its AOFP, as shown in Figure 7. The well AOFP is $8.5 \times 10^4\ m^3$/day, which is marked as a filling circle in the x-axis. The corresponding cumulative probability is 0.18 to determine the hydraulic fracturing performance. The hydraulic fracturing quality of this well was classified as "poor" according to the evaluation criteria given in Table 5 because its cumulative probability value is less than 0.3. Taking the same evaluation process, we have assessed the hydraulic fracturing performances of other vertical wells in the developed block of the eastern Sulige region.

From the variable importance analysis, the proppant fluid ratio (FPR) is the most influential factor among the fracture treatment parameters. If we increased FPR value from 9.7% (original) to 23%, the well productivity was increased by 29%, and the hydraulic fracturing quality of this well could be classified as "Fair." Therefore, the evaluation is useful to optimize the HF treatment.

We performed the assessments for all 704 fractured wells. The evaluation results of the block are shown in Figure 8. 54.6% of the wells in the block have been fractured with "Fair" quality, 26.4% wells with "Poor" quality, and only 19% wells with "Good" quality. The results indicate that the better HF designs are required to improve the HF performances in the gas field.

4. Conclusions

In the paper, we have developed a probabilistic workflow to assess the hydraulic fracture performance by integrating the ensemble learning with the Monte Carlo simulation. Using the data from a developed block in the eastern Sulige region, we have applied the workflow to evaluate the hydraulic fracture performance of the wells.

(1) The absolute open flow potential is regarded as the response variable, while ten geological and fracture treatment parameters are chosen as the input variables

(2) An ensemble learning model is built to quantify complex relationship between the geological properties, treatment parameters, and the absolute open flow potential. Results indict that the formation thickness has the most important effect on the well productivity, followed by the proppant fluid ratio

(3) Among 704 fractured vertical wells, 19% of the wells have been stimulated with "Good" quality, 54.6% of the wells with "Fair" quality, and 26.4% wells with "Poor" quality

(4) The proposed workflow could be applied to evaluate other well stimulation performances such as multistage hydraulic fracturing in horizontal wells or acid fracturing

Data Availability

Data is available upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research is funded by the Project of National Natural Science Foundation of China (Grant Nos. 51974253 and 51934005), the Youth Project of National Natural Science Foundation of China (Grant No. 52004219), the Scientific Research Program Funded by Education Department of Shaanxi Province (Grant Nos. 18JS085 and 20JS117), and the Natural Science Basic Research Program of Shaanxi (Grant Nos. 2020JQ-781 and 2017JM5109).

References

[1] H. Li, A. Jia, and D. He, "Technology advancement and prospect of natural gas development in Petrochina," *Natural Gas Industry*, vol. 30, no. 1, pp. 5–7, 2010.

[2] C. Zou, R. Zhu, S. Wu Songtao et al., "Types, characteristics, genesis and prospects of conventional and unconventional hydrocarbon accumulations: taking tight oil and gas in China as an instance," *Acta Petrolei Sinica*, vol. 33, no. 2, pp. 173–187, 2012.

[3] C. Zou, G. Zhai, G. Zhang et al., "Formation, distribution, potential and prediction of global conventional and unconventional hydrocarbon resources," *Petroleum Exploration and Development*, vol. 42, no. 1, pp. 14–28, 2015.

[4] X. Ma, A. Jia, J. Tan, and H. Dongbo, "Tight sand gas development technology and practices in China," *Petroleum Exploration and Development*, vol. 39, no. 5, pp. 611–618, 2012.

[5] S. Longde, Z. Caineng, J. Ailin et al., "Development characteristics and orientation of tight oil and gas in China," *Petroleum Exploration and Development*, vol. 46, no. 6, pp. 1015–1026, 2019.

[6] T. Lu, Y. Liu, L. Wu, and X. Wang, "Challenges to and countermeasures for the production stabilization of tight sandstone gas reservoirs of the Sulige gas field, Ordos Basin," *Natural Gas Industry*, vol. 35, no. 6, pp. 43–52, 2015.

[7] H. Yang, J. Fu, and X. Liu, "Accumulation conditions and exploration and development of tight gas in the Upper Paleozoic of the Ordos Basin," *Petroleum Exploration and Development*, vol. 39, no. 3, pp. 315–324, 2012.

[8] Z. Tan, T. Lu, Y. Liu, L. Wu, and Y. Yang, "Technical ideas of recovery enhancement in the Sulige Gasfield during the 13th Five-Year Plan," *Natural Gas Industry*, vol. 3, no. 3, pp. 234–244, 2016.

[9] G. Ji, A. Jia, D. Meng et al., "Technical strategies for effective development and gas recovery enhancement of a large tight gas field: a case study of Sulige gas field, Ordos Basin, NW China," *Petroleum Exploration and Development*, vol. 46, no. 3, pp. 629–641, 2019.

[10] J. Wang, C. Zhang, and J. Li, "Tight sandstone gas reservoirs in the Sulige gas field: development understandings and stable-production proposals," *Natural Gas Industry*, vol. 41, no. 2, pp. 100–110, 2021.

[11] J. Guo, Z. Guo, and Y. Cui, "Recovery factor calculation method of giant tight sandstone gas field," *Acta Petrolei Sinica*, vol. 39, no. 12, pp. 1389–1396, 2018.

[12] H. Belyadi, E. Fathi, and F. Belyadi, *Hydraulic Fracturing Fluid Systems*, Gulf Professional Publishing, 2017.

[13] T. Jiang, *Fracturing Technology and Case Analysis of Complex and Difficult Oil and Gas Reservoirs*, Sinopec Press, Beijing, 2017.

[14] J. Li, J. Bai, and L. Zhu, "Volume fracturing and its practices in Sulige tight sandstone gas reservoirs, Ordos Basin," *Natural Gas Industry*, vol. 33, no. 9, pp. 65–69, 2013.

[15] Z. Liu, Q. Wang, L. Zhang, J. Chuanjie, and G. Haipeng, "Laws for gas well production decline in east area of Sulige gas field, Ordos Basin," *Xinjiang Petroleum Geology*, vol. 36, no. 1, pp. 82–85, 2015.

[16] C. L. Cipolla, E. P. Lolon, J. C. Erdle, and B. Rubin, "Reservoir modeling in shale-gas reservoirs," *SPE Reservoir Evaluation Engineering*, vol. 13, no. 4, pp. 638–653, 2010.

[17] W. Yu, B. Gao, and K. Sepehrnoori, "Numerical study of the impact of complex fracture patterns on well performance in shale gas reservoirs," *Journal of Petroleum Science Research*, vol. 3, no. 2, pp. 83–89, 2014.

[18] X. Li, D. Zhang, and S. Li, "A multi-continuum multiple flow mechanism simulator for unconventional oil and gas recovery," *Journal of Natural Gas Science and Engineering*, vol. 26, pp. 652–669, 2015.

[19] Y. Xu, J. S. A. Cavalcante Filho, W. Yu, and K. Sepehrnoori, "Discrete-fracture modeling of complex hydraulic-fracture geometries in reservoir simulators," *SPE Reservoir Evaluation Engineering*, vol. 20, no. 2, pp. 403–422, 2017.

[20] C. R. Clarkson, "Production data analysis of unconventional gas wells: review of theory and best practices," *International Journal of Coal Geology*, vol. 109-110, pp. 101–146, 2013.

[21] S. D. Mohaghegh, "Reservoir simulation and modeling based on pattern recognition," in *Paper presented at the SPE Digital Energy Conference and Exhibition*, the Woodlands, Texas, USA, 2011.

[22] A. M. Nejad, S. Sheludko, T. Hodgson, R. McFall, and R. F. Shelley, "A case history: evaluating well completions in the Eagle Ford Shale using a data-driven approach," in *Paper presented at the SPE Hydraulic Fracturing Technology Conference*, the Woodlands, Texas, USA, 2015.

[23] S. D. Mohaghegh, *Shale Analytics: Data-Driven Analytics in Unconventional Resources*, Springer, Cham, Switzerland, 2017.

[24] J. B. Montgomery and F. M. O'Sullivan, "Spatial variability of tight oil well productivity and the impact of technology," *Applied Energy*, vol. 195, pp. 344–355, 2017.

[25] S. Wang and S. Chen, "Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling," *Journal of Petroleum Science and Engineering*, vol. 174, no. 6, pp. 682–695, 2019.

[26] L. Porras, C. Hawkes, and I. Arshad, "Evaluation and optimization of completion design using machine learning in an unconventional light oil play," in *Proceedings of the 8th Unconventional Resources Technology Conference*, 2020.

[27] X. Zhu, C. Liu, D. Zhong, and F. Han, "Diagenesis and their succession of gasbearing and non-gas-bearing reservoirs in the Sulige gas field of Ordos Basin, China," *Acta Geologica Sinica-English Edition*, vol. 83, no. 6, pp. 1202–1213, 2009.

[28] M. Wang, H. Tang, F. Zhao et al., "Controlling factor analysis and prediction of the quality of tight sandstone reservoirs: a case study of the He8 Member in the eastern Sulige Gas Field, Ordos Basin, China," *Journal of Natural Gas Science and Engineering*, vol. 46, pp. 680–698, 2017.

[29] J. Fu, L. Fan, X. Liu, and D. Huang, "Gas accumulation conditions and key exploration & development technologies in Sulige gas field," *Acta Petrolei Sinica*, vol. 40, pp. 240–255, 2019.

[30] A. Jia, D. He, and J. Tang, "The development stategies for Gas fields of low permeability, low abundance and in heterogeneity," in *International Oil & Gas Conference and Exhibition in China*, Beijing, China, 2006.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, second edition, 2017.

[32] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*, CRC Press, 2012.

[33] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–12327, 2001.

[34] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics and Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[35] M. Liu, H. Chen, D. Wei, Y. Wu, and C. Li, "Nonlinear relationship between urban form and street-level PM2.5 and CO based on mobile measurements and gradient boosting decision tree models," *Building and Environment*, vol. 205, article 108265, 2021.

[36] C. Ding, D. Wang, X. Ma, and H. Li, "Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees," *Sustainability*, vol. 8, no. 11, 2016.

[37] J. Tang, B. Fan, L. Xiao et al., "A new ensemble machine-learning framework for searching sweet spots in shale reservoirs," *SPE Journal*, vol. 26, no. 1, pp. 482–497, 2021.

[38] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, Chichester, UK, third edition, 1995.

Hindawi
Geofluids
Volume 2022, Article ID 5916616, 8 pages
https://doi.org/10.1155/2022/5916616

Research Article

Measurement of Total Flow Rates in Horizontal Well Oil-Water Two-Phase Flows by the Application of BP Neural Network Algorithm to Production Array Logs

Xin Zhang,[1] **Hongwei Song**,[1,2] **Haimin Guo**,[1] **and Xinlei Shi**[3]

[1]*School of Geophysics and Oil Resources, Yangtze University, Wuhan, Hubei 430100, China*
[2]*Research Office of Yangtze University, Key Laboratory of Well Logging, China National Petroleum Corporation, Wuhan, Hubei 430010, China*
[3]*CNOOC China Limited, Tianjin Branch, Tianjin 300459, China*

Correspondence should be addressed to Hongwei Song; shw98wj@yangtzeu.edu.cn

Received 1 January 2022; Revised 28 February 2022; Accepted 1 March 2022; Published 11 April 2022

Academic Editor: Bicheng Yan

In the modern petroleum industry system, oil-water two-phase flows exist widely. Among them, the total flow rate of mixture fluid in a horizontal well is difficult to obtain due to the phase segregation caused by gravity. Therefore, it is a difficult and hot issue. To obtain the total flow rate of oil-water two-phase flows in horizontal wells, in this paper, Multiple Array Production Suite (MAPS), which is also called Production Array Logs (PAL), is used to conduct simulation experiments, uses BP neural network (BPNN) algorithm to train the data samples, and establishes the prediction models of the total flow rates of oil-water two-phase flows in horizontal wells. The results showed that the average relative error was less than 10%, which justify that the BPNN has good practicability in using data of MAPS in oil-water two-phase flows horizontal wells to predict the total flow rates, and it provides a new method and theoretical support for obtaining flow rates in horizontal wells.

1. Introduction

With the maturity and wide application of horizontal well oil development technology, it has become an important research direction to study its related production logging technology, including obtaining various flow parameters in wells. Among them, the study of oil-water two-phase flows is a key problem [1, 2].

In the logging industry, some researchers have tried to provide the solution: multiple-sensor array tools (MAPS), which can be used to detect and analyze multiphase flow in horizontal wells. For the multiphase flow of horizontal well, these tools measure the fluid properties at multiple locations around the cross-sectional area of the wellbore, providing a distributed measurement that helps to relate the measurements to obtain flow rates and holdups [3, 4]. According to the measured value, auxiliary computer analysis software can be used to reconstruct it, which can reflect the flow pattern in the horizontal well. At the same time, the relevant special software can reconstruct the results of well logging into simulation images. It should be used in conjunction with neural network technology to obtain better results.

The neural network has shone in the oil-related sector. During the COVID-19 outbreak, the convolutional neural network which extracts online oil news can be used to predict the fluctuation of the oil market [5]. The deep neural network is used to process seismic wave data for reservoir prediction [6].

In addition, some scholars have made researches on the application of processing well logging data with neural networks. Guo et al. [7] studied the use of feedforward artificial neural networks for oil-water two-phase flow production logging interpretation methods and obtained the design of corresponding software. Strict experiments were carried out on the three-phase flow simulation device of Daqing Oilfield. The 105 sets of sample data obtained in the experiment were substituted into the network for training and learning. 93 sets

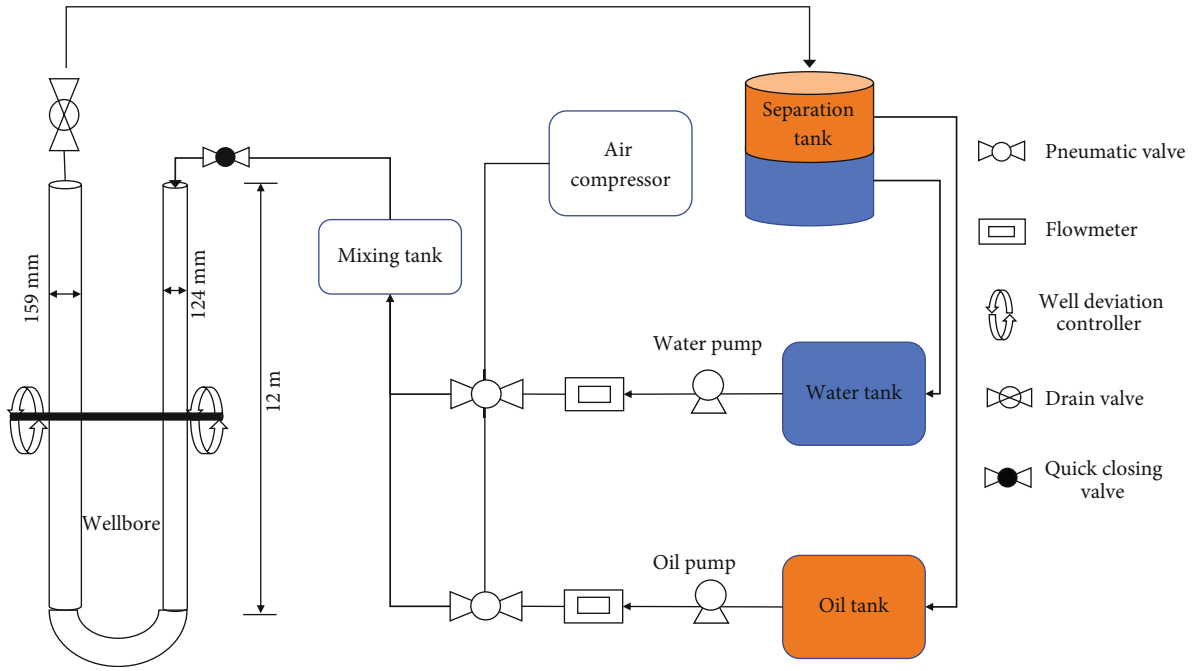


Figure 1: Multiphase flow experimental set-up.

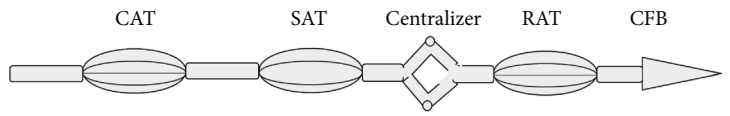


Figure 2: MAPS+CFB.

Figure 3: SAT.

Figure 4: CAT.

of data were selected to train the BP neural network and RBF network. 12 multiple sets of sample data were used for testing and prediction, the value was compared with the actual value, and the error was within the acceptable range [8]; Chen et al. [9] used BP neural network to predict the startup velocity of the continuous flowmeter in the wellbore and obtained excellent prediction results, which showed the potential of this method in conventional vertical well logging interpretation. In general, it is very effective to use great algorithms combined with computer tools to improve productivity [1, 10].

In horizontal wells, the fluid is stratified under the influence of gravity, which makes the instruments used in conventional vertical wells such as inline flowmeters and capacitance holdup meters not applicable anymore. In order to solve these problems, this paper took the MAPS array logging tools as the research object and the experimental data obtained from the oil-water two-phase simulation experiment on the production logging experiment platform of Yangtze University as samples, adopting BPNN that can realize the nonlinear mapping to establish models and predict flow rates. It provides a new method for obtaining flow rates of horizontal wells.

2. Production Logging Simulation Experiment in Horizontal Well

2.1. Experimental Set-Up. The experimental set-up (Figure 1) was a multiphase flow simulation experiment

Figure 5: CFB.

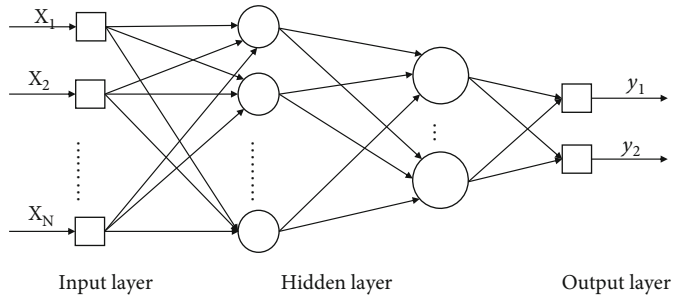


Figure 6: Structure of BP.

Table 1: Cable speed-flow conversion table.

Experimental control flow (m^3/d)	Cable speed (m/min)	Equivalent total flow (m^3/d)
100	0	100.00
	10	385.91
	15	528.87
	20	671.82
300	0	300.00
	10	585.91
	15	728.87
	20	871.82
600	0	600.00
	10	885.91
	15	1028.87
	20	1171.82

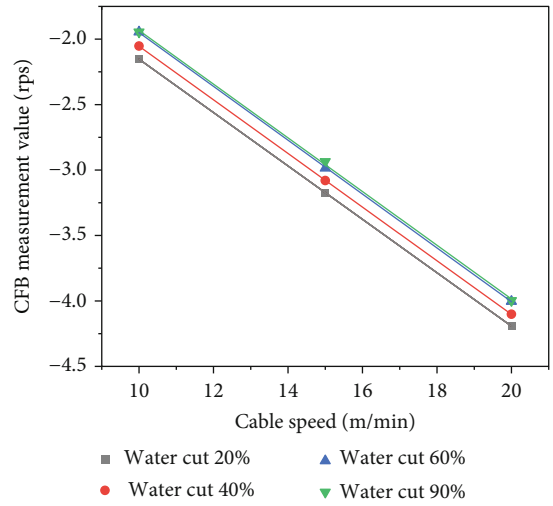


Figure 7: Relational graph of the measured value of 93° inclination of 100 m^3/d CFB and cable speed.

platform of Yangtze University, which included two transparent glass tubes with one inner diameter of $D = 124$ mm and another inner diameter of $D = 159$ mm, which both permitted visual observation of the flow, besides a large liquid storage tank for oil and water storage, an air compressor for pneumatic control valves and gas for experiments, a hydraulic machine for adjusting the angle of the wellbore, and the relevant complete set of pumps, transportation pipelines, and a master console for controlling the input fluid and observing the state of the entire experimental set-up. The simulation experiment used a 12 m long 159 mm inner diameter wellbore. The fluid was an oil-water two-phase flow, the water was tap water, and the oil was No. 10 industrial white oil. The oil-water ratio was set to 20%, 40%, 60%, 80%, and 90% water cut. The flow rates were 100 m^3/d, 300 m^3/d, and 600 m^3/d. The wellbore angles were set to near-horizontal (93°, 90°, and 85°). The measurement methods were divided into spot measurements and continuous measurements. During continuous measurements, set the cable speed to 0, 10 m/min, 15 m/min, and 20 m/min.

2.2. Production Array Logs. The experimental tools MAPS (Figure 2) include spinner array tool for obtaining flow rates, resistance array tool, and capacitance array tool for obtaining holdup, in addition to a caged fullbore flowmeter [11].

The spinner array tool (SAT) (Figure 3) is composed of six microspinners distributed on the same section, which can measure the flow at different positions on the wellbore area; the capacitance array tool (CAT) (Figure 4) and the resistivity array tool (RAT) have similar shapes: twelve microsensors are evenly distributed to measure the fluid nearby [12]. The caged fullbore (CFB) (Figure 5) flowmeter,

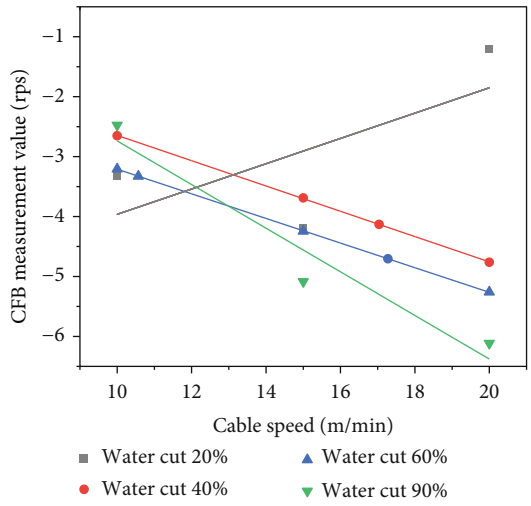


Figure 8: Relational graph of the measured value of 93° inclination of 300 m^3/d CFB and cable speed.

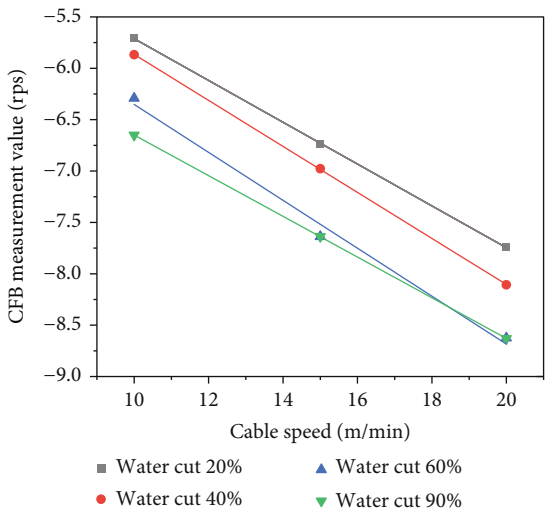


Figure 9: Relational graph of the measured value of 93° inclination of 600 m^3/d CFB and cable speed.

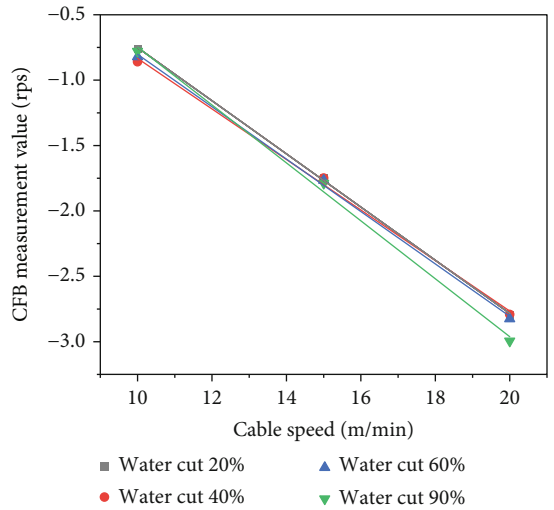


Figure 10: Relational graph of the measured value of 90° inclination of 100 m^3/d CFB and cable speed.

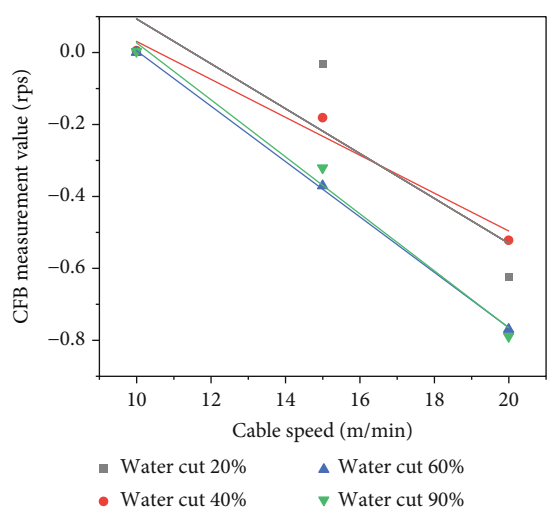


Figure 11: Relational graph of the measured value of 90° inclination of 300 m^3/d CFB and cable speed.

which uses retractable metal blades to measure the flow velocity and resides in the middle of the wellbore, can accurately measure the total flow rate [13].

3. Backpropagation Neural Network

Backpropagation neural network is a multilayer feedforward neural network that uses a backpropagation learning algorithm to adjust the weights, generally uses a sigmoid function to transmit signals between neurons, and can realize any nonlinear mapping from input to output.

BP neural network is a kind of neural network, which is powerful and widely used. It consists of an input layer, a hidden layer, and an output layer. The layers are fully interconnected, and the same layers are not connected. The hidden layer can be one or more layers. Figure 6 is a typical three-layer BP neural network structure picture.

The multilayer network design enables the network to calculate errors more accurately during operation and complete more complex tasks. At the same time, it uses backpropagation algorithm for learning. In the BP neural network, the data signal is propagated back layer by layer through the input layer and hidden layer. When training the network weights, the signal is in the direction of reducing errors in the network structure, from the output layer to the middle layers, and forwards layer by layer. Modify the connection weight of the network [14]. With the continuous progress of learning, the final error becomes smaller and smaller and finally reaches the set ideal error. At this time, a series of neuron weights containing information is obtained, and a model that can solve the problem is completed.

In summary, the BP neural network can realize complex nonlinear mapping, can approximate complex functions

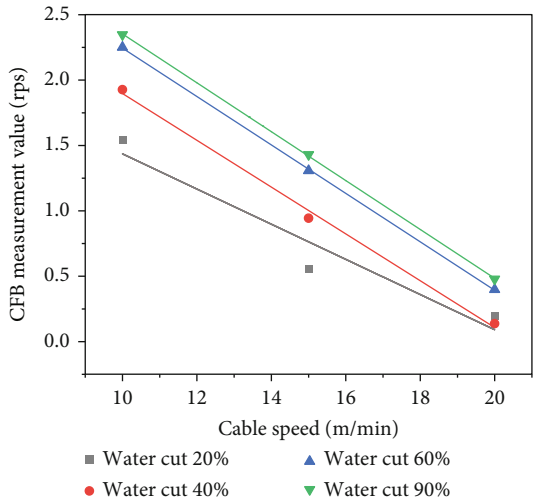


Figure 12: Relational graph of the measured value of 90° inclination of 600 m^3/d CFB and cable speed.

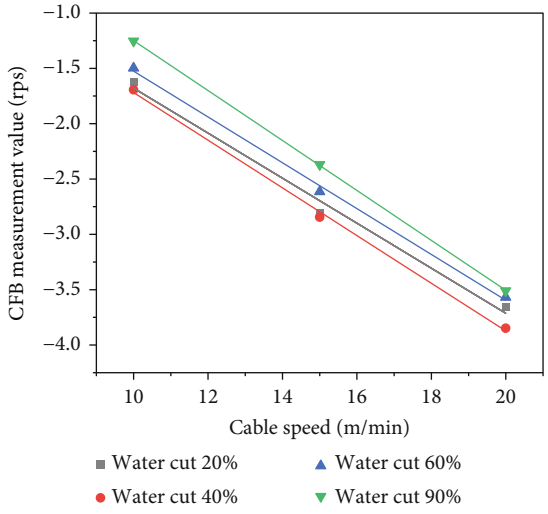


Figure 13: Relational graph of the measured value of 85° inclination of 100 m^3/d CFB and cable speed.

suitable for problem-solving and processing for the fusion of multiple input data, and is suitable for processing MAPS well logging data to predict the flow rates.

4. The Realization of Predicted Flow Rates

4.1. Sample Data Training and Prediction. In the laboratory environment, total input flow rates, water cuts, and wellbore inclinations are controlled; the corresponding MAPS data are used as the original data. Among them, all data samples are divided into two categories according to whether the measurement method is the continuous measurement or spot measurement, and only the CAT corresponding to SAT is included in the spot measurement data.

In the wellbore, pulling the cable to drive the instrument to the direction of fluid entry is equivalent to increasing the flow rate. Therefore, Equation (1) and Equation (2) can be

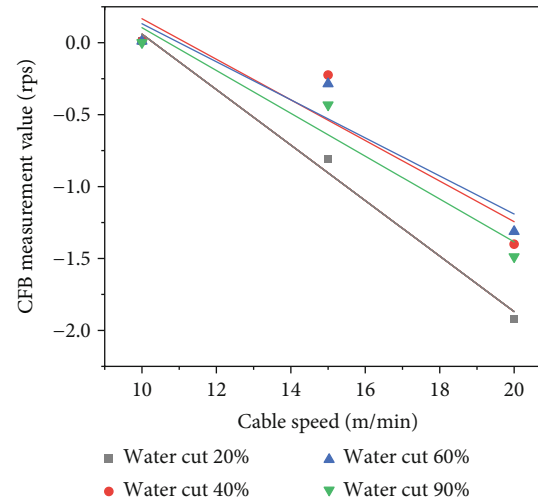


Figure 14: Relational graph of the measured value of 85° inclination of 300 m^3/d CFB and cable speed.

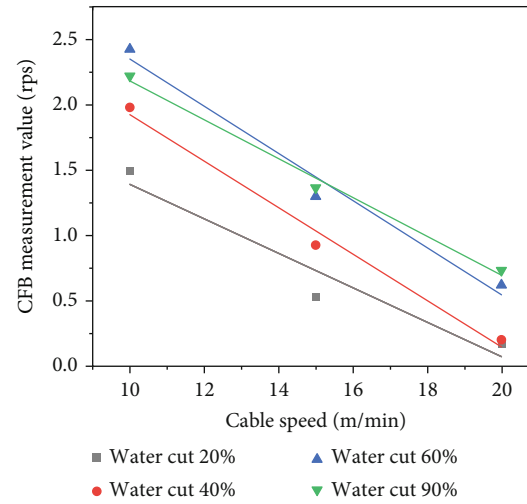


Figure 15: Relational graph of the measured value of 85° inclination of 600 m^3/d CFB and cable speed.

used to convert the cable speed into flow rate:

$$P_c = \frac{1}{4}\pi D^2 \times 3600 \times 24 \times 10^{-6}\, m^3/d, \tag{1}$$

$$Q = v * P_C * \frac{5}{3}, \tag{2}$$

where Q represents the conversion flow (m^3/d), v represents cable speed (m/min), and P_C is pipe constant ((m^3/d)/(cm/s)).

After unit conversion and calculation according to the above formula, the result was obtained and made into Table 1.

The array spinners of SAT are closely distributed in the wellbore wall, and the CFB is located in the center of the wellbore. The combination of the two can measure the flow data in the horizontal well. The conventional processing of these data is to consider the spinner starting velocity fitting

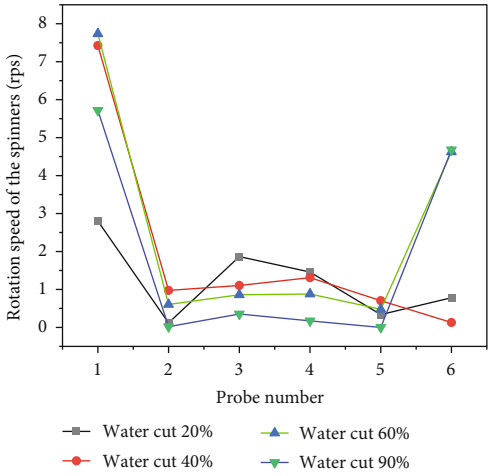


Figure 16: The response diagram of SAT spinners with flow rate = 100 m^3/d in horizontal well in different water cuts.

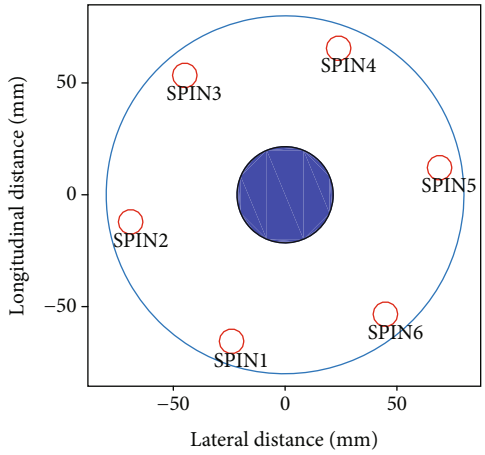


Figure 17: The wellbore distribution map in horizontal well with flow rate = 100 m^3/d of SAT.

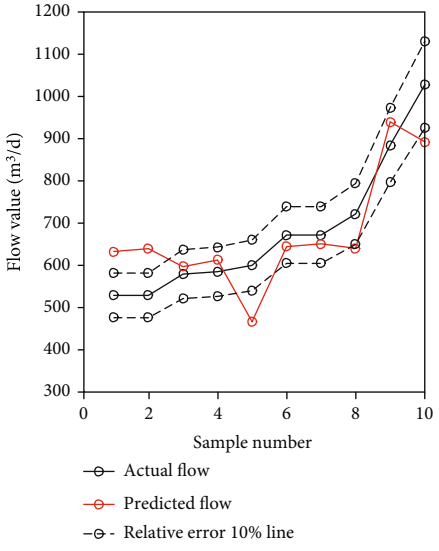


Figure 18: Histogram of the predicted value error of model 1.

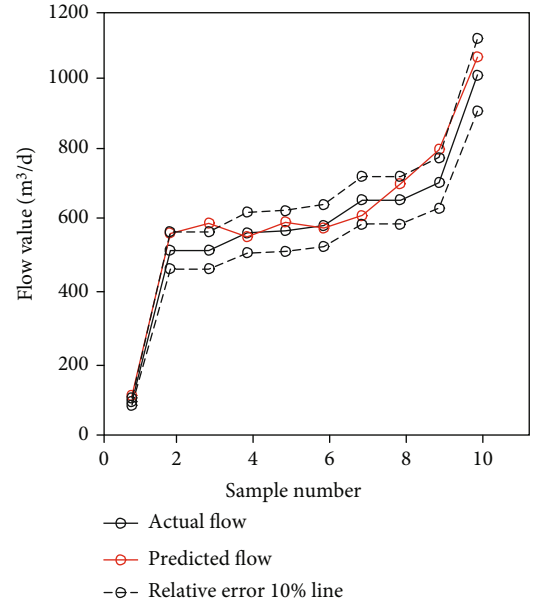


Figure 19: Histogram of the predicted value error of model 2.

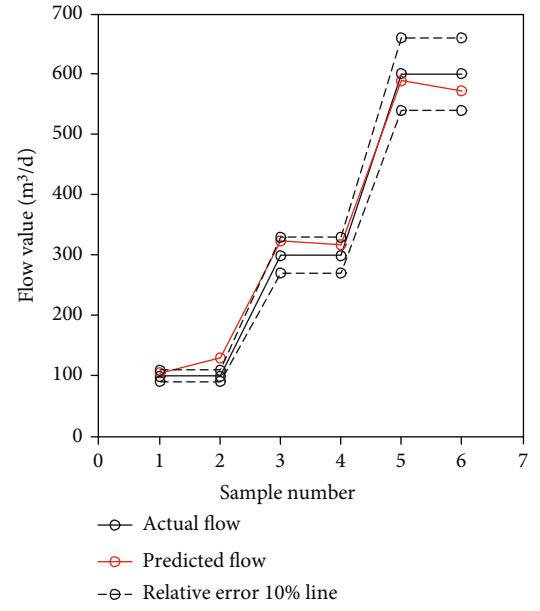


Figure 20: Histogram of the predicted value error of model 3.

to obtain the apparent velocity and obtain the flow data after deviation correction. Using the BP neural network to predict the flow of these data can omit the fitting step and ignore the influence of the correction factor, as shown below in the figures.

Analyzing Figures 7–17, we can get this conclusion: the response values of CFB with different water cuts have a different slope or intercept with the measuring velocity relational graph. There is also a significant difference in SAT response value in oil and water. It can be seen that SAT and CFB have different instrument constants and start-up velocities in different fluids. The data of CAT is used to give the phase states of different positions in the wellbore. In fact,

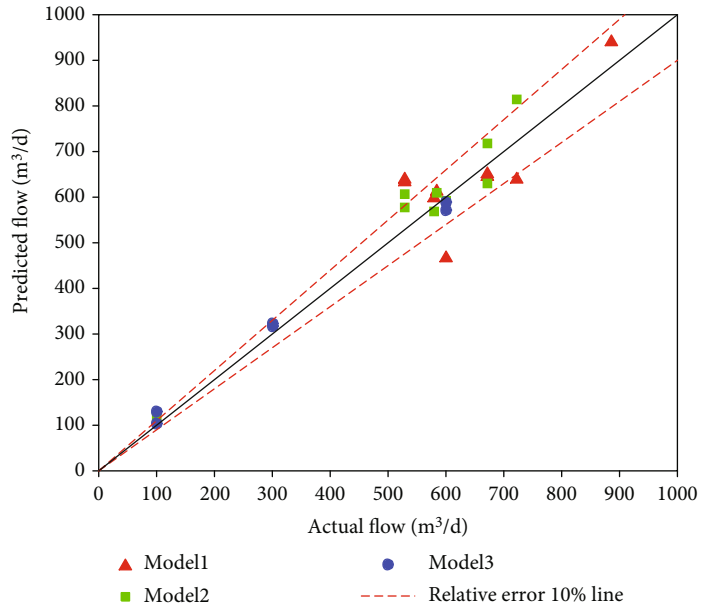


Figure 21: Comparison chart of the prediction effect of the three models.

the BP neural network is trained with different oil and water SAT and CFB data as weights. Therefore, CAT data can be added to the training to improve accuracy. According to the existing data and algorithm requirements, the following three prediction models have been set up:

(1) Data of SAT as inputs: response value of spinner array tool as inputs (129 sets for training, 10 sets for test)

(2) Data of SAT and CFB as inputs: response value of spinner array tool and caged fullbore flowmeter as inputs (129 sets for training, 10 sets for test)

(3) Data of SAT, CAT, and CFB as inputs: response value of spinner array tool and capacitance array tool and caged fullbore flowmeter as inputs (24 sets for training, 6 sets for test)

The above three different input models all use actual flow rates as output for the BP neural network algorithm to adjust connection weights. When importing data into the algorithm, select several groups of data to keep (about 10% of the total amount of data samples) for use in testing algorithms, analyzing errors, and evaluating effects. Benefitting from the advantages of the BP neural network, the error can be continuously corrected to analyze the data by itself, so the data can be used directly without correction processing; on the other hand, due to the advantages of the BP network, when inputting the measured value of capacitance array tool CAT instrument, it is not necessary to normalize to obtain the water holdup, but the network can obtain the relationship through calculation by itself, which makes it quick and convenient to predict the flow rates on the way. The results obtained by running the above three models are as follows:

(1) Data of SAT as inputs

The average relative error is 10.84% and according to Figure 18, there are many jump points. It may be related to insufficient experimental data, and the bigger reason is the limitation of the single SAT value as input that can be analyzed from Figure 18: only 5 of the predicted values can meet the requirement of less than 10% error, accounting for half of the total number of test samples, and the effect is not as expected.

(2) Data of SAT and CFB as inputs

According to Figure 19, there are few jump points and basically, they meet the requirements, and the average relative error is 7.94%; according to the visual error chart, the effect is good, most of the predicted values fluctuate within the ±10% error line, and occasionally, the predicted values exceed these lines. It can be seen that the absolute error is not large at the lower flow rate, the relative error is higher, and the rest of the data is better, which achieves the purpose of predicting the flow rates.

(3) Data of SAT, CAT, and CFB as inputs

This model has less data, which is not conducive to the neural network but still achieves relatively good results according to Figure 20. Overall, there is one and only one jump point, which may be caused by too little data or insufficient experimental conditions. Only one of the six predicted values has a relative error of 29.9%, and the rest are within 10%, the average relative error is 9.03%, which realizes the prediction of flow rate. After increasing the number of data samples, it is bound to further improve the accuracy and reduce the error. It also shows the superiority of model 3.

4.2. Result Analysis. The BP neural network has been used to process the raw data of the array logging tool, and different models have been designed to run. From the results of the three models, according to Figure 21, the accuracy of model 1 is slightly worse, and the average relative error of model 1 is 10.84%, but the consistency is poor, while the average error of model 2 is 7.94% and model 3 is 9.03%. The accuracies are better than model 1. They are limited by the number of output data and a small number of total data sets, so their accuracies may still be improved. In general, this design can be used in actual production with improvement.

5. Conclusions

(1) It is not feasible to predict the flow rates by using only a single spinner array tool SAT instrument response value as the data input. The other two prediction models have higher accuracy. After more data samples and more improvement of multiple output, directions and algorithms can be put into actual production interpretation and complement other methods of calculating flow

(2) Although the BP neural network does not need to calibrate the data due to its unique operating principle, the increased input parameters, for example, in model three, the introduction of capacitance array tool CAT measurement data is equivalent to giving the array spinner tool data with the water holdup parameter corresponding to the flowmeter which adds weight to the network when processing different spinners. As a result, the predicted flow rates are closer to the true value. It can be concluded that when using the BP network, increasing the related different parameters such as the corresponding RAT and CAT data will improve the efficiency and accuracy of the network, which is of great significance to the actual production

(3) BPNN is a traditional and mature neural network algorithm, which is easy to implement and obtain. However, it should be noted that BPNN has disadvantages such as difficult parameters to determine and dependence on samples and it takes many operations to determine the optimal parameter. In future studies, more advanced new algorithms will be used or extended to deep learning

Data Availability

The data are available by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (42174155).

References

[1] H. S. Skjefstad and M. Stanko, "Experimental performance evaluation and design optimization of a horizontal multi-pipe separator for subsea oil-water bulk separation," *Journal of Petroleum Science & Engineering*, vol. 176, pp. 203–219, 2019.

[2] H. W. Song, H. Guo, S. Guo, and H. Shi, "Partial phase flow rate measurements for stratified oil-water flow in horizontal wells," *Petroleum Exploration and Development*, vol. 47, no. 3, pp. 613–622, 2020.

[3] A. Zett, M. Webster, A. Noordermeer et al., "New sensor development helps optimise production logging data acquisition in horizontal wells," *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description*, vol. 53, no. 2, pp. 70–85, 2012.

[4] L. Liao, D. Zhu, and N. Yoshida, "Interpretation of array production logging measurements in horizontal wells for flow profile," in *SPE Annual Technical Conference and Exhibition*, pp. 1–12, Morial Convention Center, New Orleans, Louisiana, USA, 2013.

[5] B. Wu, L. Wang, S. Wang, and Y.-R. Zeng, "Forecasting the U.S. oil markets based on social media information during the COVID-19 pandemic," *Energy*, vol. 226, article 120403, 2021.

[6] Y. J. Tian, J. H. Gao, D. X. Wang, and D. Y. Chen, "Removing strong seismic reflection based on the deep neural network," *Chinese Journal of Geophysics*, vol. 64, no. 8, pp. 2780–2794, 2021.

[7] G. Haimin, W. Xiangyang, Z. Yujie, and L. Juru, "Application of neural network in the analysis of oil-water two-phase flow," *Journal of Jianghan Petroleum Institute*, vol. 3, pp. 48–51, 1994.

[8] D. Yubo, L. Jicheng, and W. Qiong, "Application of RBF neural network in ultrasonic measurement of oil-water two-phase flow holdup," *Information and Control*, vol. 4, pp. 476–480, 2005.

[9] C. Kegui, M. Chenfei, D. Jun, C. Yuanyuan, H. Changbing, and W. Xiaozhun, "Start-up velocity prediction of wellbore continuous flowmeter based on BP neural network," *Logging Technology*, vol. 42, no. 4, pp. 444–448, 2018.

[10] X. Guo, Y. Jin, J. Zi, and B. Lin, "Numerical investigation of the gas production efficiency and induced geomechanical responses in marine methane hydrate-bearing sediments exploited by depressurization through hydraulic fractures," *Energy & Fuels*, vol. 35, no. 22, pp. 18441–18458, 2021.

[11] A. Muratbek, K. Ainagul, H. Peter et al., "The application of multi-sensor production logging and spectral noise logging tools in optimising water shut-off in a carbonate environment," in *SPE Annual Caspian Technical Conference & Exhibition*, Astana, Kazakhstan, November 2016.

[12] G. J. Frisch, D. F. Dorffer, M. Jung, A. Zett, and M. J. Webster, "Improving the process of understanding multiprobe production logging tools from the field to final answer," *Computer-Aided Design*, vol. 43, no. 9, pp. 1099–1109, 2009.

[13] B. Xinbing, W. Ying, and Z. Qiang, "Practical application of full borehole flowmeter in oil field," *Petrochemical Technology*, vol. 23, no. 9, p. 86, 2016.

[14] M. Chen, *MATLAB Neural Network Principles and Examples*, Tsinghua University Press, Beijing, 2013.

Hindawi
Geofluids
Volume 2022, Article ID 8441075, 15 pages
https://doi.org/10.1155/2022/8441075

Research Article

Development of Decline Curve Analysis Parameters for Tight Oil Wells Using a Machine Learning Algorithm

Weirong Li,[1] **Zhenzhen Dong**,[1] **John W. Lee**,[2] **Xianlin Ma**,[1] **and Shihao Qian**[1]

[1]*Xi'an Shiyou University, China*
[2]*Texas A&M University, USA*

Correspondence should be addressed to Zhenzhen Dong; dongzhenzhen1120@hotmail.com

Received 15 January 2022; Accepted 17 March 2022; Published 4 April 2022

Academic Editor: Jose Antonio Torres

To obtain a reliable production forecast, one has to establish a geological model with well logs and seismic data. The geological model usually has to be upscaled using certain upscaling techniques. Then, a dynamic reservoir model is constructed with another dataset, including completion data, production data, fluid properties, and relative permeability curves. At last, the dynamic model needs to be validated by a history matching process. This approach is data-intensive, time-consuming, and often not rigorously accomplished due to the lack of skillset and time. In this study, 10,000 groups of reservoir/completion input data were generated by Latin hypercube sampling method, and then, 10,000 groups of output (oil rate and cumulative production data) were obtained by numerical simulation. Next, a machine learning technique was applied to establish a model between the input data and determining parameters of a decline curve analysis model by fitting the generated cumulative production rate. Overall coefficients of determination (R^2) of the three Arps decline curve factors were 0.966, 0.990, and 0.945. The validation result shows that the production rate and cumulative production predicted by the proposed machine learning–decline curve analysis (ML-DCA) model agreed well with those simulated by reservoir simulation. As a result of the ML-DCA regression model, a complete understanding can be established of the impact of reservoir properties on the DCA model. The proposed ML-DCA model not only provides a quick and robust method for petroleum engineers to estimate production performance for unconventional reservoirs from reservoir and completion properties without full-field geocellular modeling but also can be used to optimize the completion and operation parameters for wells of interest.

1. Introduction

Unconventional oil and gas reservoirs have been able to be developed economically with advancements in horizontal well drilling and multistage hydraulic fracturing technology. Management and operation of unconventional oil and gas reservoirs demand accurate prediction of production rates, which facilitates better development strategy, more economic feasibility, more reliable reserve evaluations, and more informed business decisions.

Many efforts have been made to develop efficient numerical models for simulation of unconventional oil and gas production considering complex hydraulic and natural fracture geometries and multiple gas transport mechanisms in nanopores [1–4]. High-resolution, three-dimensional (3D), geocellular models characterize geological features and fracture network complexities with grid blocks and their related rock and fluid properties, and they yield a comprehensive geographic distribution of pressure and saturation over a period of time. However, such a model is computationally prohibitive as a large number of simulations are required for history matching and production optimization in a close-loop reservoir management and decision-making context. The challenges become even more discouraging with a lack of rock and fluid data and insufficient production history. Researchers have tried to accelerate simulations through techniques such as upscaling [5, 6], multiscaling [7–9], and streamline methods [10, 11]. However, all of these speed-ups require a full physics-based model as a starting point for simplification.

Decline curve analysis (DCA) as an alternative technique to forecast oil and gas production has been applied in the oil

industry for a long time. The most commonly used curve-fit-based decline curve models for unconventional oil and gas reservoirs include the multisegment Arps decline model [12–15], modified Arps DCA model [16], transient hyperbolic model [17], stretched exponential decline method (SEDM) [18], Duong method [19], and power law exponential (PLE) method [20]. It is well known that curve-fit-based DCA models are easy to apply for unconventional reservoirs once the flow regimes related to depth of investigation are appropriately identified, but they fall short of prediction accuracy as they cannot capture the reservoir features or properties that are highly relevant to production performance.

With the introduction of data analytics in the oil and gas industry, such practical challenges have caused emerging data-driven solutions in the area of data mining and machine learning (ML), in which geological data, completion data, and dynamic data are input to the ML algorithm to unravel hidden physical relationships contained in the data but not represented in existing simulation models to output future production performance. Li and Han [21] applied a neural network to study the correlation between selected reservoir/completion properties and the determining parameters of the logistic growth model (LGM). Thus, a trained ML-DCA model can be used to predict the production trends for both an existing well and a new well according to the given reservoir/completion properties. Sun et al. [22] proposed a recurrent neural network- (RNN-) based sequence-to-sequence model to forecast production. Mukherjee et al. [23] tested and performed various ML algorithms, including linear regression (LR), principal component analysis (PCA), neural network regression (NNR), boosted decision tree (BDT), and binned decision tree (BiDT), to forecast gas production. Tamimi et al. [24] presented a comprehensive intelligent decision support system (IDSS) to forecast and determine parameters of Arps decline curve model from 3,400 unconventional wells. Temizel et al. [25] applied the long short-term memory (LSTM) method for predicting long-term production behavior in unconventional shale reservoirs. Xue et al. [26] proposed a multiobjective random forest (MORF) algorithm to forecast the production rate, with reservoir and completion characteristics as input. Gross et al. [27] proposed a physics-informed ML workflow combining fast reduced-order models (ROMs) with reservoir simulation models to predict production as a function of pressure management in a fractured Marcellus shale reservoir. Doan and Vo [28] used ML techniques to enhance the accuracy of production forecasting in the North Malay Basin.

Data-driven models may not capture many geological features of a reservoir, but they run much faster than full-physics models, and the trained model on real data can avoid uncertainty or making hypothesis. They can provide reliable predictions with enough data in the calibration process. However, when the number of parameters becomes large, data-driven models require a certain number of samples for training and testing before they can be effectively used.

The unique nature of this study is using a simulation model as a basis for the production dataset, which provides an alternative for utilizing the machine learning algorithms when actual data are insufficient or unavailable. This research offers less experienced engineers with an effective ML-DCA model that correlates reservoir and fluid properties, as well as completion parameters, with the DCA model to forecast the production rates without full-field geocellular modeling for unconventional reservoirs. Therefore, without even working with complicated geomodels, less experienced engineers can make robust estimations about the production forecast once the ML-DCA model is established and delivered by experienced reservoir engineers. The implementation between DCA and the simplified reservoir model with ML algorithm makes it easy to extend the workflow for any other type of reservoirs, such as thermal or compositional.

Table 1: Cumulative production-time relationship of Arps decline curve.

Exponential $b = 0$	$Q = \frac{1}{D_i}(q_i - q_t)$
Hyperbolic or superhyperbolic	$Q = \left[\frac{q_i}{D_i(1-b)}\right]\left[1 - \left(\frac{q_t}{q_i}\right)^{1-b}\right]$
Harmonic $b = 1$	$Q = \left[\frac{q_i}{D_i}\right] \ln\left(\frac{q_t}{q_i}\right)$

This study is organized as follows. Firstly, the methods in the proposed ML-DCA workflow, including decline curve model and artificial neural network were introduced, followed by synthetic production profiles generated and established from single-well numerical simulation considering 15 geological and completion parameters that included matrix and fracture network properties for data-driven study. Next, the ML-DCA model was trained and tested by the synthetic production profiles; then, a case study was given to evaluate the prediction performance of the ML-DCA model, which shows the power and accuracy the proposed ML-DCA workflow; and a sensitivity analysis was performed to investigate the effect of geological and reservoir parameters on the production with the ML-DCA model; finally, the limitations and future work were discussed and conclusions were drawn.

2. Methodology

2.1. Decline Curve Analysis. The Arps decline curve is the most common DCA. The Arps hyperbolic decline curve model is [29].

$$q = \frac{q_i}{(1 + bD_i t)^{1/b}}, \tag{1}$$

where q is the predicted production, q_i is the initial production, t is time, b is a constant, and D_i is the initial decline rate. When the constant loss ratio b is 0, the decline curve reduces to an exponential decline; if b is 1, the decline curve becomes harmonic; if b is greater than 0 but less than 1, the decline curve is hyperbolic; and if b is greater than 1, often during transient flow, the decline curve is called superhyperbolic. When one deals with low- and ultralow-permeability wells, transient flow lasting months to years is common,

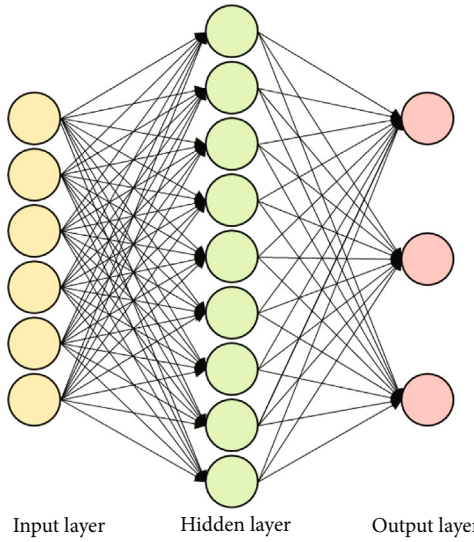


Figure 1: General ANN structure.

$$\begin{pmatrix} \theta_1^{(1)} & \cdots & k_{f1}^{(P)} \\ \vdots & \ddots & \vdots \\ \theta_N^{(1)} & \cdots & k_{fN}^{(P)} \end{pmatrix} \rightarrow \theta(Neural\ Network) \rightarrow \begin{pmatrix} (b, D_i, q_i)_1 \\ \cdots \\ (b, D_i, q_i)_N \end{pmatrix}$$

Figure 2: ML-DCA regression with neural network.

followed by a transition flow regime and later BDF. In practice, b is constant for a long time during transient flow in hydraulically fractured wells and during BDF, so one can apply the Arps model with confidence to those time periods during which b is constant [12].

In this study, cumulative oil production was selected to perform DCA regression as it is much smoother than the production rate. This approach also mitigates the effect of irregular and noisy production data, especially for real field data. Cumulative oil production is the integral of the production rate (Equation (1)), which is defined as

$$Q = \int_{t_0}^{t} q dt = \int_{t_0}^{t} \frac{q_i}{(1 + bD_i t)^{1/b}}, \tag{2}$$

where Q is the cumulative production. Integrating and rearranging Equation (2), the cumulative production-time relationship for different b values can be derived and shown as in Table 1.

2.2. Artificial Neural Network. An artificial neural network (ANN) is a common ML technique inspired by the structure of neurons in the human brain. The ANN was selected as the ML method because it has been successfully applied in many engineering and science problems to extract complex and nonlinear relationships among process variables. In this study, a typical three-layer back-propagation ANN structure was established, consisting of an input layer, a hidden layer, and an output layer, as shown in Figure 1. Several key variables, including hidden layer neurons, training algorithm, and transfer functions, were considered to get the optimal

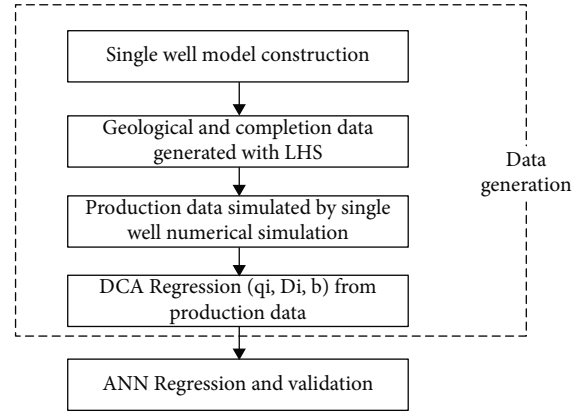


Figure 3: Flowchart of the ML-DCA model for production prediction based on ANN.

ANN structure. The neuron number in the input layer equals the input feature number. The neuron number in the hidden layer was tuned to optimize the objective function. Each neuron received the input information from the output of the neurons in the previous layer and generated and passed output to the next layer. The mathematical model of the i^{th} neuron is expressed as

$$y_i = \phi\left(\sum_{j=1}^{n} w_{ij} x_j + b_i\right), \tag{3}$$

where y_i is the output of the i^{th} neuron on the next layer, x_j is the input from the previous layer, w_{ij} is the weight, n is the input number from the previous layer, and ϕ is the activation function. The weight was tuned for each neuron through optimizing the loss function in the model training process. The R-squared score (R^2) is commonly used as the loss function. It is a statistical measure that represents the proportion of the variance of a dependent variable that is explained by an independent variable or variable in a regression model. R^2 score is given as

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2}, \tag{4}$$

where $\hat{y}_i$ is the predicted value for i, and $\bar{y}_i$ is the mean of y_i.

2.3. Machine Learning-Decline Curve Analysis Algorithm. The ML algorithm provides a statistical technique to analyze a system with an existing dataset without being explicitly programmed. To introduce how the ML algorithm solves a regression problem, one usually defines the input vector, x_i, as

$$x_i = \left(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(P)}\right), i = 1, \cdots, N, \tag{5}$$

where N is the sample size or the number of input vectors and P is the number of input parameters, such as permeability and

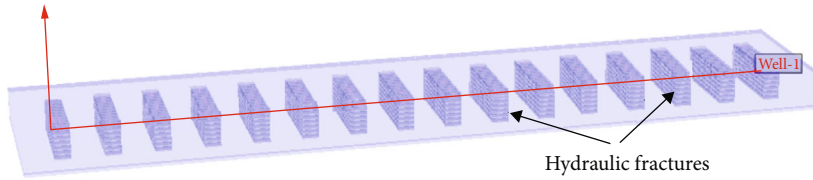


Figure 4: MFHW model to generate the simulated oil and gas production data for DCA.

Table 2: Parameters and associated distribution to generate the input dataset.

Parameters	Minimum value	Maximum value	Distribution function
Grid size, X direction(DI) (ft)	75	125	Uniform[75,125]
Grid size, Y direction(DJ) (ft)	30	80	Uniform[30,80
Grid size, Z direction(DK) (ft)	1	5	Uniform[1,5]
Matrix permeability (mD)	0.0001	1	Power[10,Random[-4,0]]
Porosity	0.05	0.15	Uniform[0.05,0.15]
Horizontal well length (ft)	1800	6000	Triangle[1800,3500,6000]
Fracture half-length (ft)	100	850	Triangle[100,380,850]
Fracture spacing (ft)	75	500	Uniform$[1, 4] \times$ DI
Effective fracture permeability (mD)	1	100	Uniform[1,100]
Layer-up	1	3	Uniform[1,3]
Layer-down	1	3	Uniform[1,3]
Monitored oil rate (bbl/day)	1.5	2.5	Triangle[1.5, 2, 2.5]
Initial reservoir pressure (psi)	2000	6000	Uniform[2000,6000]
Bubble point pressure, psi	400	6000	Initial reservoir pressure/random[1,15]
Operating BHP (psi)	200	3000	Initial reservoir pressure/random[10,30]

Note: the range of each parameter was set based on Chinese oilfield practice, such as Ordos basin [31–33].

porosity. Thus, the input matrix $\mathbf{X}$ can be defined as

$$\mathbf{X} = (x_1, x_2, \cdots x_i, \cdots, x_N)^T = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(P)} \\ \vdots & \ddots & \vdots \\ x_N^{(1)} & \cdots & x_N^{(P)} \end{pmatrix}. \quad (6)$$

In this study, the input was a column matrix composed of reservoir, completion, and operation parameters, such as fracture half-length, fracture width, fracture permeability, and porosity.

The corresponding output parameters can be given as

$$\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots \boldsymbol{y}_i, \cdots, \boldsymbol{y}_N)^T \quad (7)$$

where the output vector y_i is the determining parameters of the decline curve model corresponding to the input vector x_i .

$$y_i = \left(y_i^{(1)}, y_i^{(2)}, \cdots, y_i^{(t)}\right), i = 1, \cdots, N, \quad (8)$$

where t is the number of determining parameters for the decline curve model. Clearly, there are three parameters in the Arps decline curve model; thus, the output y_i is a column matrix of b, D_i, and q_i.

The existing measurement dataset was used to train the ML model. The training dataset is given as

$$(\mathbf{X}, \mathbf{Y}) = \{(x_1, y_1), (x_2, y_2), \cdots, (x_k, y_k), \cdots, (x_1, y_1)\}. \quad (9)$$

The purpose of ML is to establish a mapping function, $\varnothing(\bullet)$, from the training dataset

$$\mathbf{Y} = \varnothing(\mathbf{X}). \quad (10)$$

The proposed ML-DCA model can be expressed and established as Figure 2 with the training dataset.

$$\begin{pmatrix} \varnothing_1^{(1)} & \cdots & k_{\mathrm{f1}}^{(P)} \\ \vdots & \ddots & \vdots \\ \varnothing_N^{(1)} & \cdots & k_{fN}^{(P)} \end{pmatrix} \longrightarrow \varnothing(\text{Neural Network}) \longrightarrow \begin{pmatrix} (b, D_i, q_i)_1 \\ \cdots \\ (b, D_i, q_i)_N \end{pmatrix}. \quad (11)$$

2.4. Pearson Correlation. The most commonly used type of correlation that describes the degree of relationship between two variables is the Pearson correlation. Pearson's r measures the linear relationship between two variables, say X and Y. A correlation of 1 indicates that the data points perfectly lie on a line for which Y increases as X increases. A value of -1 also implies that the data points lie on a line;

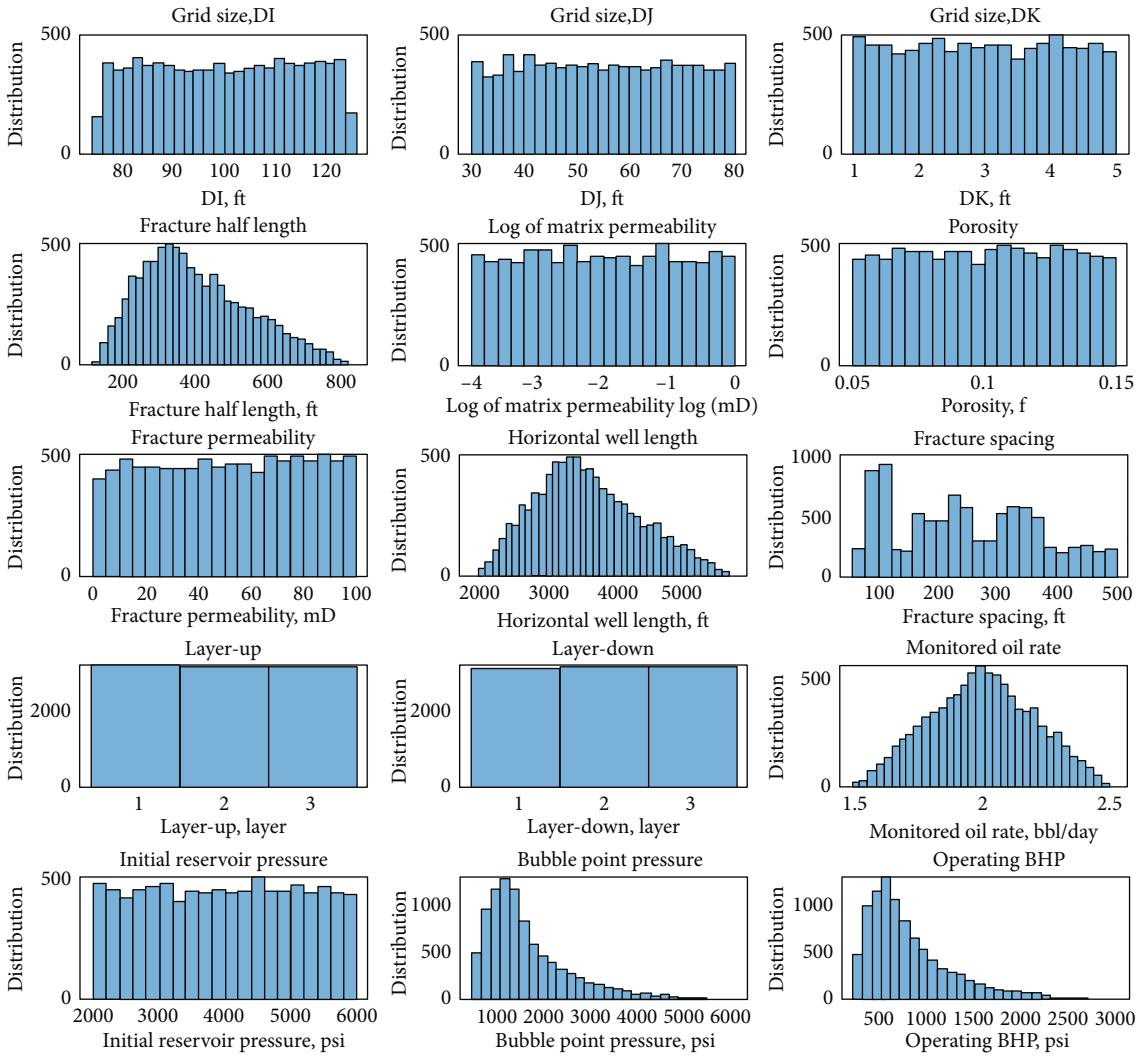


Figure 5: Histograms of geological and completion parameters generated through LHS.

however, Y decreases as X increases. The formular for r is

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}. \quad (12)$$

Figure 3 shows the proposed workflow of ML-DCA. First, a 3D numerical model was established for a typical multistage fractured horizontal well (MFHW). Then, the Latin hypercube sampling (LHS) method was performed to generate enough (e.g., 10,000) experimental designs within the predetermined distribution type and ranges of certain input parameters, which were then used to simulate the cumulative oil production rates by running reservoir simulation. Next, the DCA regression algorithm was then performed to fit the simulated cumulative oil production rates and obtain the determining parameters of the DCA model. Finally, the ML algorithm (e.g., ANN, but it can be other algorithms in future work) was trained and tested to investigate the correlations or mapping function between geological and completion parameters and the determining parameters of the DCA model.

Note that the proposed workflow can be extended to investigate various decline curve models, such as SEDM, PLE, and Duong model, by simply replacing the decline model. In addition, all related algorithms were developed with open-source libraries that can be easily integrated with an in-house or commercial simulator.

3. Data Generation

Building a data-driven model requires a large set of geological features and completion data (production, pressure, well logs, etc.) as inputs and the factors of the DCA model as output. The ML method was then used to establish the correlation between reservoir features and production. Ideally, it is better to obtain the dataset used to build the data-driven model from the actual field. However, synthetic data generated from numerical or analytical models can alternatively be used if good-quality, actual data are insufficient or unavailable [26, 30]. In this study, we used a numerical

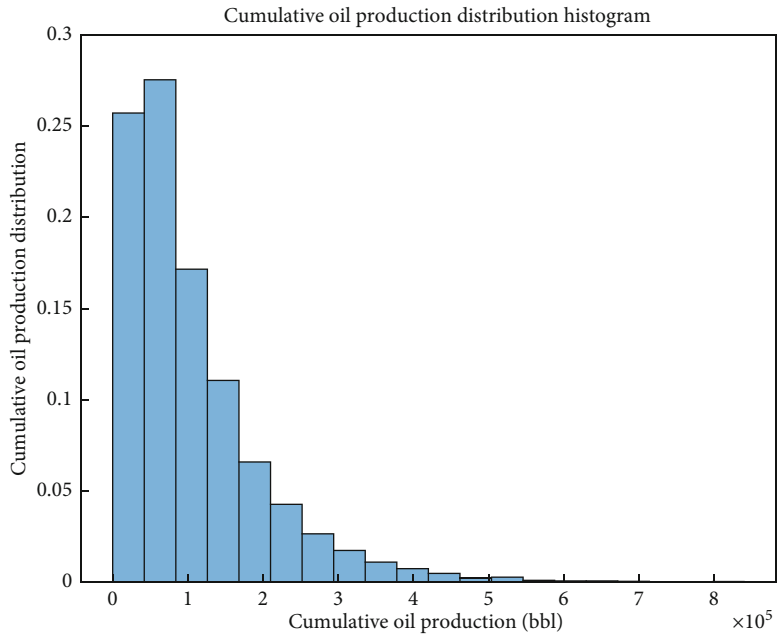


Figure 6: Distribution of 10-year cumulative oil production calculated by numerical simulation.

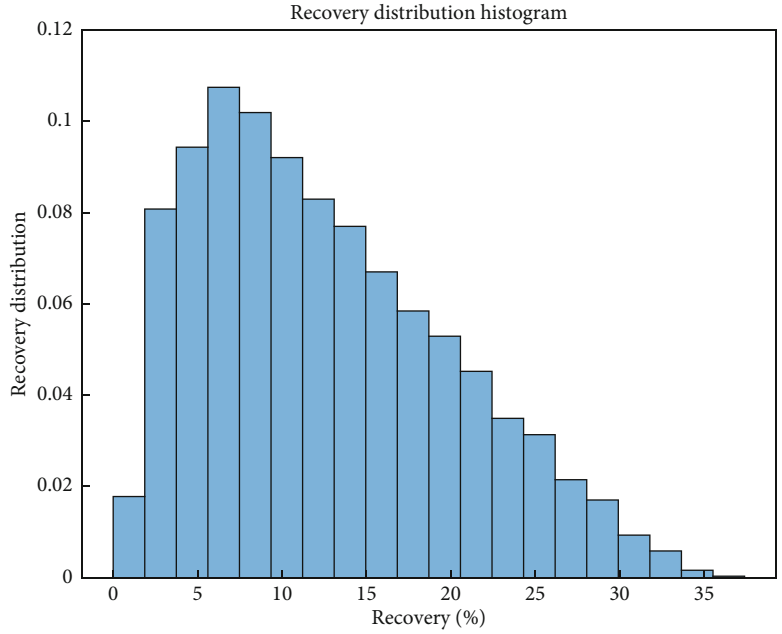


Figure 7: Distribution of 10-year recovery calculated by numerical simulation.

model to generate cumulative production profiles with randomly generated geological and completion parameters. Later, the simulated cumulative production rates were used to establish the relationship between reservoir properties and the DCA model.

3.1. Single-Well Reservoir Model. The MFHW model to simulate production from tight oil reservoirs was a three-phase, 3D rectangular model that was established with a CMG simulator. A tartan grid was used to model the MFHW as it is the best way to catch transient behavior. Otherwise, ones need to use local grid refinement (LGR), which is more time consuming. The grid number ($Nx \times Ny \times Nz$) of the model is set to be $50 \times 21 \times 7$. The grid size for each direction (DI, DJ, and DK) was one of the uncertain parameters that were generated by the sampling method later. Thus, the well control area and drainage area ($Lx = Nx \times DI$, $Ly = Ny \times DJ$, $Lz = Nz \times DK$) was changed with the grid size. The horizontal well was placed in the center of the reservoir model and produced under the constraint

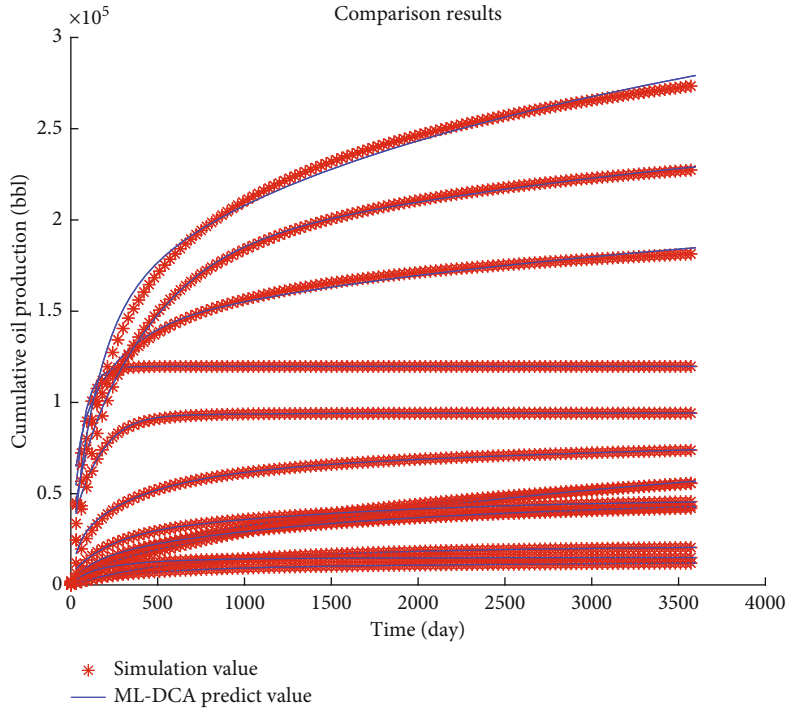


Figure 8: Part of the DCA fitting results.

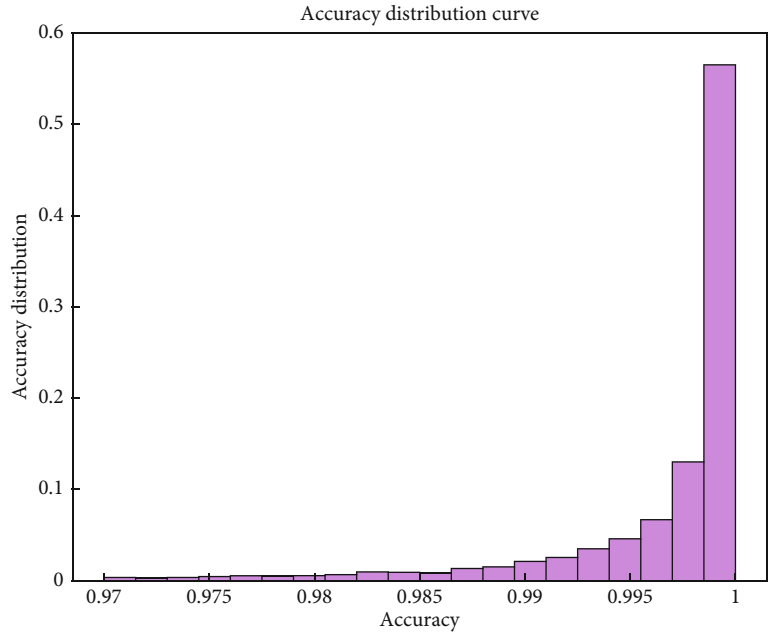
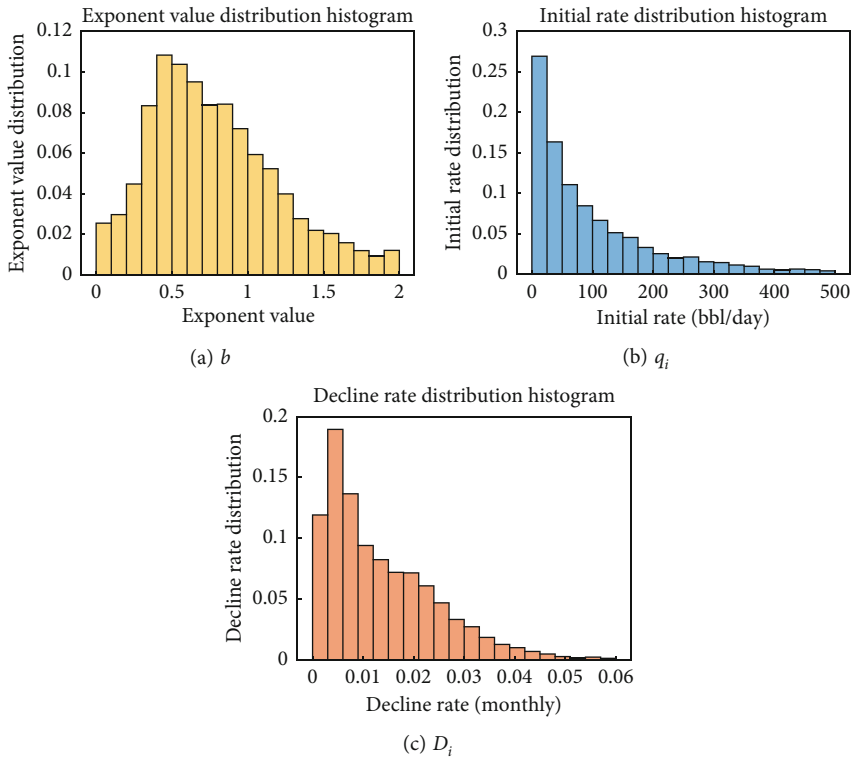


Figure 9: Distribution of R^2 of DCA regression.

of constant flowing BHP. Duration of production was set to be 10 years, which is considered a realistic tight oil production scenario. Figure 4 shows the sideview of the MFHW model, in which the darkened sections of the grid blocks represent the hydraulic fractures.

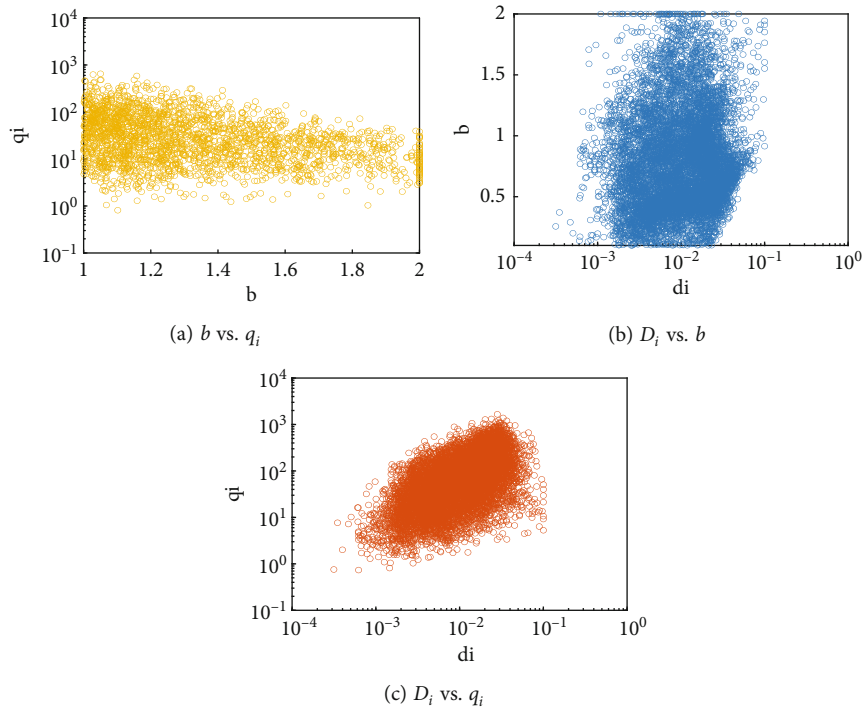
3.2. Latin Hypercube Sampling. Once the base model was established, the experimental design was carried out, varying the design variables to generate experiments that would be used to calibrate the model. In this study, 15 geological and completion parameters were investigated, including grid size in X, Y, and Z directions (DI, DJ, and DK), matrix permeability, porosity, horizontal well length, fracture half-length, fracture spacing, fracture effective permeability, layer-up, layer-down, bubble point pressure, initial pressure, monitored oil rate, and operating bottom-hole-pressure



(a) b (b) q_i

(c) D_i

Figure 10: Probability distribution of the estimated parameters of the DCA model.



(a) b vs. q_i (b) D_i vs. b

(c) D_i vs. q_i

Figure 11: Cross-plot of the fitted DCA parameters for 10,000wells.

(BHP) (Table 2). Among the 15 parameters, the first three features were used to define the drainage area. Matrix permeability and porosity were the tight oil reservoir properties. The following four features, including horizontal well length, fracture half-length, fracture spacing, and fracture effective permeability, define the horizontal well and completion parameters. The layer-up and layer-down specified the number of layers that defines the fracture penetrate above/below the horizontal well. Thus, these two parameters were related to the fracture height. Bubble point pressure was one of the

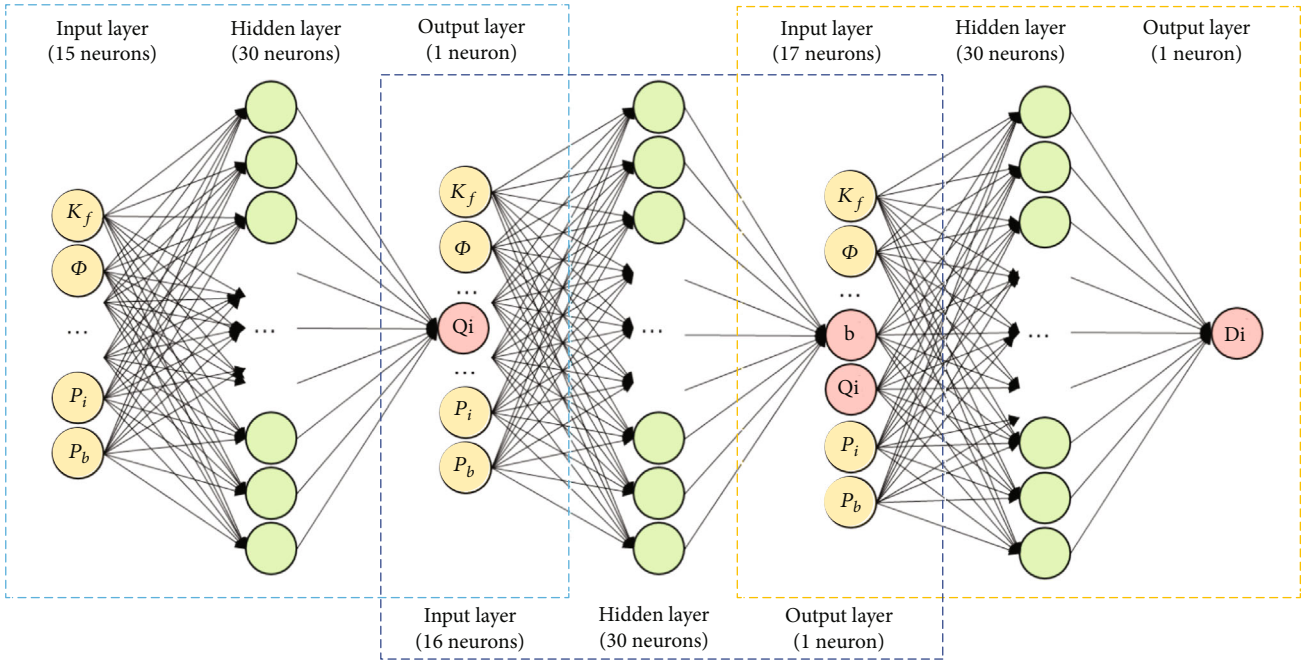


Figure 12: ANN model used in this study to estimate the three Arps parameters from geological and operational parameters.

Table 3: The comparison of overall R^2.

Variables	Sequential network	Simultaneous network
b	0.990	0.923
q_i	0.966	0.904
D_i	0.945	0.912

fluid properties, and initial pressure defined the initial reservoir condition. Bubble point pressure was set to be lower than the initial reservoir pressure for each scenario. The water saturation of the reservoir was constant in this study, and the value was set to 0.4. Relative permeability curve was predefined. The last two features were operation parameters during production. Oil production changed with the operating BHP, which was set to lower than the initial reservoir pressure. The well was shut off when the oil rate was lower than the monitored oil rate defined by well constraints. To obtain effective training and testing of data-driven models, 10,000 samples of each parameter were generated through the LHS method, with the ranges and distribution type listed in Table 2. The probability distributions of each parameter are shown in Figure 5.

3.3. Generation of Cumulative Production. The corresponding 10-year monthly oil production for each combination was then simulated by the numerical model. The distribution of the 10,000 cumulative oil productions and recoveries are displayed in Figures 6 and 7. Basically, the cumulative production of 50% of wells was less than 10×10^4 bbl. The typical oil recovery usually was less than 35%, mainly distributed between 5% and 15%.

3.4. Decline Curve Analysis Regression. To obtain the best predictive model, production data from 10,000 wells were transformed into DCA space. DCA best fit curves were usually computed with the least-squares regression [34]. Thus, least-squares regression was programmed to estimate the three determining parameters of the Arps decline model from the 10,000 synthetic cumulative production profiles generated by the represented reservoir simulation. Figure 8 shows parts of 10,000 fitting results as an example. It shows good agreement between the cumulative oil production calculated by the DCA model and the synthetic cumulative production simulated by the numerical model. Figure 9 is the distribution of coefficient of determination (R^2). 99% of the DCA regression had a R^2 greater than 0.99. In other words, the DCA curve can be a proxy model of the single-well simulation to forecast tight oil production in this study.

Figure 10 gives the histogram of the determined initial rate (q_i), b value, and initial decline rate (D_i) by the ML-DCA model. The distribution can be used for further study if one performs production uncertainty analysis with the DCA model, such as to forecast P10, P50, and P90. b value affects long-term production but does not make much difference in short-time production. Initial production rate and initial decline rate affect the short-term rate. As shown in Figure 11, there was no apparent correlation between b and initial production rate (q_i), meaning that initial production alone cannot represent how well or how poorly a well will produce in the long term. However, b showed a significant correlation with the initial decline rate (D_i) (Figure 11(b)). This observation was similar to Tamimi et al.'s [24] study based on more than 3,400 unconventional wells in DJ basin. Simply put, a higher initial decline rate indicates a higher b value. Thus, it seems a higher decline rate may indicate a bad well, but eventually it produces more in the long term.

Table 4: The optimized network structure for ML-DCA model.

Variables	q_i	B	D_i
Number of neurons in input layer	15 parameters	15 parameters and q_i	15 parameters, q_i and b
Number of neurons in hidden layer	40	50	60
Output in output layer	q_i	B	D_i
Data split		70% for training, 30% for testing	
Function performance		R^2	
Training algorithm		Bayesian regularization	
Iterations to achieve optimal structure	185	250	324
Overall R^2	0.966	0.990	0.945

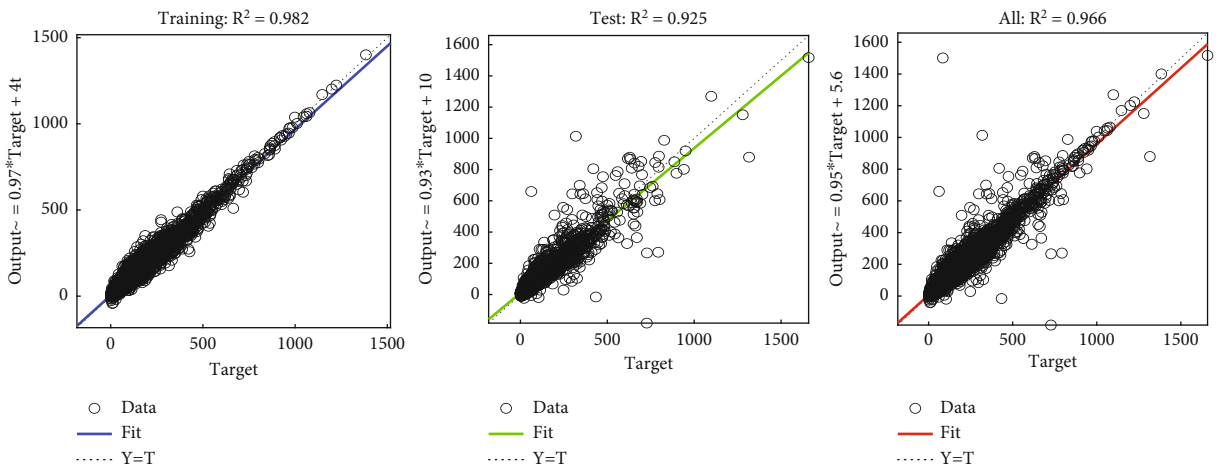


Figure 13: Regression result of parameter q_i with ANN.

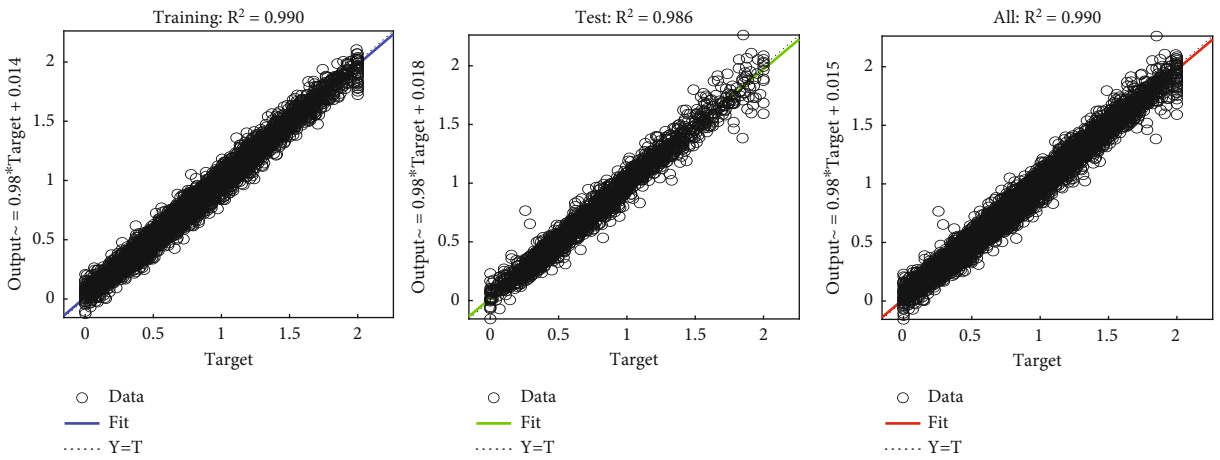


Figure 14: Regression results of parameter b with ANN.

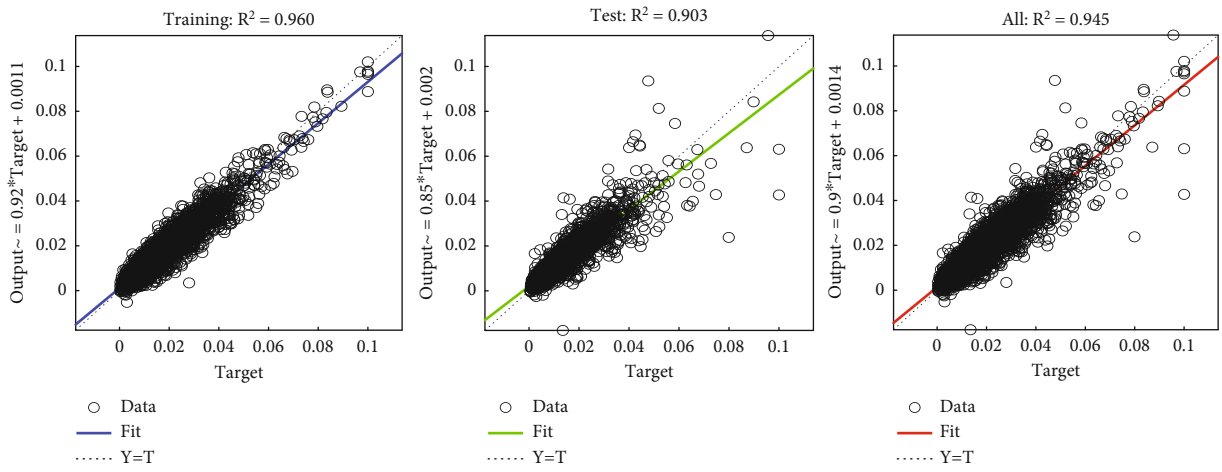


Figure 15: Regression results of parameter D_i with ANN.

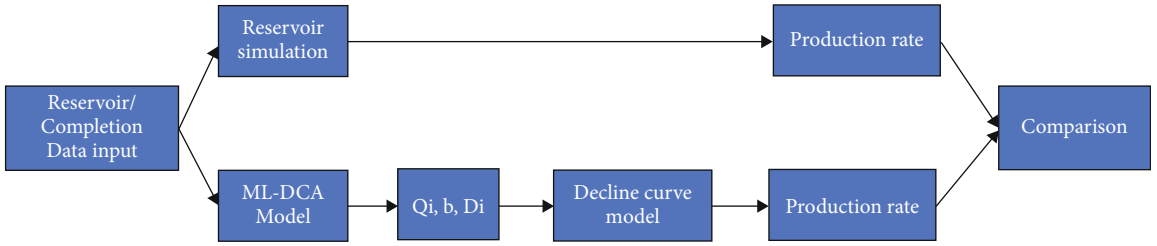


Figure 16: Flowchart to evaluate the production prediction performance.

Table 5: Reservoir features used as inputs for performance validation.

Feature	Value	Units	Feature	Value	Units
Grid size, X direction	84.08	ft	Fracture effective permeability	52	mD
Grid size, Y direction	57.43	ft	Layer-up	2	Layer
Grid size, Z direction	3.21	ft	Layer-down	1	Layer
Matrix permeability	0.0004	mD	Bubble point pressure	845	psi
Porosity	0.08		Initial pressure	3767	psi
Well length	2775	ft	Monitored oil rate	2	bbl/day
Fracture half-length	546	ft	Operating BHP	699	psi
Fracture spacing	252	ft			

4. Results and Analysis

4.1. Neural Network Regression. In this study, three-layer back-propagation neural network was established, as shown in Figure 12. As it can been seen, the three factors of the DCA model were solved sequentially rather than simultaneously, as the sequential network shown as Figure 12 achieved the highest overall R^2 and proven to be the optimal regression workflow to determine the factors of the DCA model (Table 3).

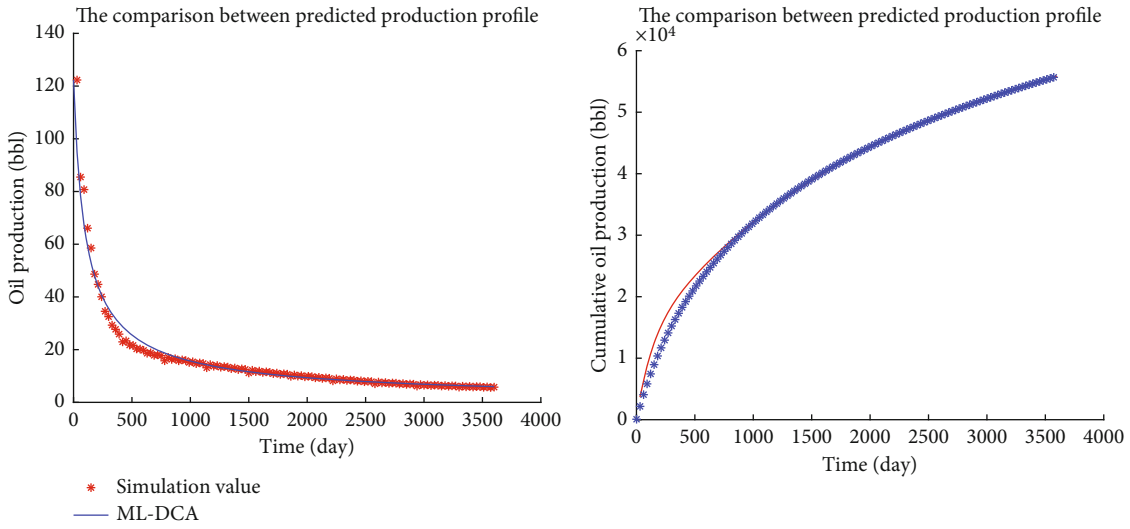
The optimal regression workflow is summarized as follows:

(i) Step 1: initially, there are 15 neurons in the input layer and 1 neuron in the output layer

(ii) Step 2: Bayesian regularization is selected as the training algorithm. Thus, in the 10,000 combinations of input parameters and corresponding cumulative production profiles, 70% of them are used as the training dataset to establish the data-driven model, and 30% of them are used as the testing dataset to determine the DCA factors

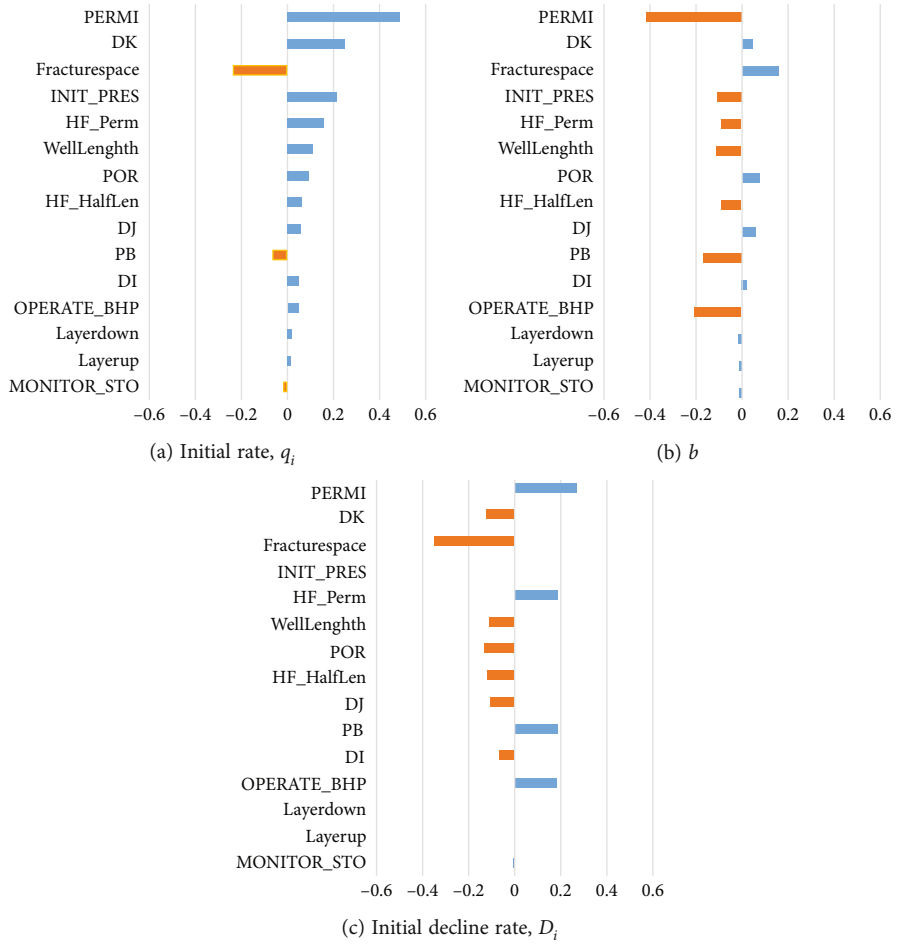
(iii) Step 3: by changing the number of neurons in the hidden layer, the different results for the target DCA factor are obtained

(iv) Step 4: Bayesian regularization, which minimizes a linear combination of squared errors, is performed to modify the connective weights and biases to achieve more accurate results [35]



(a) Oil production (bbl/day) (b) Cumulative oil production (bbl)

Figure 17: Comparison between the production simulated by the numerical model and predicted by ML-DCA algorithm.



(a) Initial rate, q_i (b) b

(c) Initial decline rate, D_i

Figure 18: Sensitivity analysis results of the reservoir and geological parameters on the decline curve model.

(v) Step 5: the ANN's training process continues interactively until the desired level of error or maximum iteration number is reached, and the first final network used to determine the vector of q_i is achieved.

(vi) Step 6: the vector of q_i is added to the input matrix, and Step 2 to Step 5 are repeated to obtain the vector of b

(vii) Step 7: the vector of b is added to the input matrix, and Step 2 to Step 6 are repeated to determine the vector of D_i

The optimum network was achieved and is given in Table 4.

Figures 13–15 show R^2 of the training set, the testing set, and all sets of the three determining parameters predicted by the proposed ML-DCA model, respectively. The axes of the abscissa and ordinate are the actual determining parameters of Arps decline curves and their predicted values, respectively. The smaller the difference between the actual value and the predicted value, the closer the data point is to the 45-degree line. The data in Figures 13–15 (both the training set and testing set) is densely distributed near the 45-degree line indicating the high prediction accuracy of the ML-DCA model. The overall prediction errors of the three factors (q_i, b, and D_i) are 0.966, 0.990, and 0.945, respectively. With the trained and validated ML-DCA model, one can easily and quickly output a decline curve model to forecast the production rate of new wells by only knowing the reservoir and completion parameters.

4.2. Prediction Performance Evaluation. Figure 16 shows the process to evaluate the prediction performance of the ML-DCA model by comparing the production profiles with numerical simulation. Table 5 presents the values of the 15 selected key features used to forecast the production profile of a tight oil well. The comparison result of the well is depicted in Figure 17. Overall, both the production rate and cumulative production between these two methods agreed with each other very well, indicating that the performance of the ML-DCA model is acceptable and reliable for science and engineering applications. It can be concluded that the ML-DCA model has the same accuracy as single-well numerical simulation. Thus, instead of establishing a complex numerical model, engineers who have less experience with numerical simulation can use the ML-DCA model as a proxy model to forecast tight oil production with low cost.

4.3. Sensitivity Analysis. Pearson correlation coefficient was calculated to evaluate the significance of each property in the three determining parameters (q_i, b, and D_i) (Figure 18). Figure 18(a) reflects permeability, grid size in the vertical direction (DK), and fracture spacing to be the top three factors having the most significant effect on the initial rate, followed by initial pressure, fracture permeability, well length, matrix porosity, fracture half-length, etc. Among them, fracture spacing had a significant negative impact on the initial rate. And operating BHP, monitored oil rate, layer-up, and layer-down had a minor effect on the initial rate.

Figure 18(b) indicates that permeability has mainly a negative influence on b. The higher the permeability, the lower the b value. It also can be concluded that operating BHP, bubble point pressure, well length, initial pressure, fracture half-length, and fracture permeability had a negative impact on b value, while fracture spacing, porosity, and grid size had a positive effect on b value.

Figure 18(c) shows the correlation between the initial decline rate (D_i) and reservoir and completion parameters. Fracture spacing, used to define the number of fractures in the reservoir model, had a prominent influence on the initial decline rate D_i. Meanwhile, the grid size (DK, DI, and DJ), well length, porosity, and fracture half-length had a negative influence. Results also indicate matrix permeability, initial pressure, fracture permeability, bubble point pressure, and operating BHP to have a positive impact on the initial decline rate.

5. Discussion and Future Work

It is well known that ML-DCA can be powerful if the data quantity and quality can be improved by including actual field data. However, with most tight oil wells having less than 60 months of production history, these data cannot be directly used to perform history matching. This study proves that this simulation-based proxy tool is reliable with a well-maintained database generated from a single-well reservoir simulation, and its efficiency in computational time allows the practicing engineer to achieve modeling objectives and to reduce uncertainty in a rapid way. As a result, the ML-DCA algorithm not only can be used as a tool to determine DCA factors and predict production, once the initial reservoir conditions, rock properties, and completion and operation parameters are given, but also can be used to optimize the completion and operation parameters for a target reservoir, such as fracture spacing, fracture half-length, and operating BHP. As for future work, further studies can be conducted to investigate the effect of measurement errors and sample size on production prediction performance.

6. Conclusions

In this study, a ML-DCA algorithm, integrating a ML technique with a DCA model, was developed to predict tight oil production performance, serving as a proxy for analytical/numerical reservoir simulation. The following critical conclusions can be summarized:

(1) With the reservoir and completion parameters as the inputs and the DCA factors as outputs, the ML-DCA model can be trained to determine DCA factors accurately, with the overall prediction errors (R^2) of the three Arps decline curve factors being 0.966, 0.990, and 0.945

(2) The production rate and cumulative production predicted by the proposed ML-DCA model agree well with those simulated by reservoir simulation

(3) ML-DCA outperforms traditional DCA methods, especially for a new well or reservoir, as ML-DCA captures the production trend of the training dataset, as well as considering the rock, fluid, completion, and wellbore properties

(4) The proposed ML-DCA outperforms full-field numerical simulation due to its simplicity and low cost

(5) As a result of sensitivity analysis, the reservoir, completion, and operation parameters that affect the three DCA factors can be determined. Among them, fracture spacing and matrix permeability have an essential effect on the DCA model, while the monitored oil rate, layer-up, and layer-down have a minor effect on the DCA model

Nomenclature

b: Arps hyperbolic or superhyperbolic decline exponent, dimensionless
D_i: Initial decline rate in the Arps decline model, D^{-1}
N: Sample size or the number of input vectors in the ML-DCA model
n: Input number from the previous layer in the ANN model
P: Number of input parameters in the ML-DCA model
Q: Cumulative production rate, MSCF or BBL
q: Predicted production rate, MSCF/D or BBL/D
q_i: Initial production in the Arps decline model, MSCF/D or BBL/D
t: Production time, days
x_i: Input from the previous layer of ANN model
y_i: Output of the ith neuron on the next layer in an ANN model
$\widehat{y}_i$: Predicted value
$\bar{y}_i$: Mean value of y_i
w_{ij}: Weight of ANN model.

Greek variables

ϕ: Activation function in the ANN model
μ: Mean value
σ: Standard deviation.

Data Availability

Data is available on request by reaching @dongzhenzhen1120@hotmail.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to thank the Project "Shale Oil Development Study of Chang7 Panke Field" and Project "Fracturing Design Optimization of Multistage Fractured Horizontal Wells in the Lower Temple Bay Field, Yanchang Oilfield" for their support and valuable discussion.

References

[1] W. Li, Z. Dong, and G. Lei, "Integrating embedded discrete fracture and dual-porosity dual-permeability methods to simulate fluid flow in shale oil reservoirs," *Energies*, vol. 10, no. 10, p. 1471, 2017.

[2] B. K. Tiwari, S. Al-Sayegh, H. Al-Muraikhi, P. Kumar, P. V. Cueille, and F. Lislaud, "Modelling hydraulic fractures in a full-field dynamic modeling using DPDP simulation techniques-an unconventional approach applied in a tight carbonate oil reservoir, Kuwait, Middle-East," in *Paper SPE-195156-MS presented at the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2019.

[3] M. Vo, S. Su, J. Lv, and C. Xiao, "Reservoir modeling and production history matching in a Triassic naturally fractured carbonate reservoir in Sichuan, China," *Improved Oil and Gas Recovery*, vol. 4, no. 2020, pp. 1–14, 2020.

[4] W. Yu, K. Wu, M. Liu, K. Sepehrnoori, and J. Miao, "Production forecasting for shale gas reservoirs with nanopores and complex fracture geometries using an innovative non-intrusive EDFM method," in *Paper SPE-191666-MS presented at the SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, September 2018.

[5] E. Alfataierge, E. Chesnokov, and Y. Gorb, "Introduction of upscaling methods derived from the simple averaging method and a comparison with the backus method of upscaling," in *Paper presented at SEG International Exposition and 86th Annual Meeting*, Dallas, Texas, USA, 2016.

[6] M. S. Islam and T. Manzocchi, "The inclusion of two-phase fault rock properties into unscaled models: a novel flow-based geometrical upscaling approach," in *Paper SPE-187228-MS presented at the SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, 2017.

[7] M. Correia and D. Schiozer, "Special connections for representing multiscale heterogeneities in reservoir simulation," in *Paper SPE-200572-MS presented at the SPE Europe featured at 82nd EAGE Conference and Exhibition*, Amsterdam, Netherlands, 2020.

[8] R. Moraes, R. M. Fonseca, M. Helici, A. W. Heemink, and J. D. Jansen, "Improving the computational efficiency of approximate gradients using a multiscale reservoir simulation framework," in *Paper SPE-182620-MS presented at the SPE Reservoir Simulation Conference*, Montgomery, Texas, USA, 2017.

[9] Y. Yang, M. Ghasemi, E. Gildin, Y. Efendiev, and V. Calo, "Fast multiscale reservoir simulations with POD-DEIM model reduction," *SPE Journal*, vol. 21, no. 6, pp. 2141–2154, 2016.

[10] A. Datta-Gupta and M. King, "Streamline stimulation: theory and practice," in *SPE Textbook Series, 11*, Society of Petroleum Engineers, 2007.

[11] S. Tanaka, D. Kam, A. Datta-Gupta, and M. King, "Streamline-based history matching of arrival times and bottomhole pressure data for multicomponent compositional system," in *Paper SPE-174750-MS presented at the SPE Annual Technical Conference and Exhibition*, Houston, Texas, USA, 2015.

[12] J. Lee, "Establishing the basis for multi-segment Arps decline models," in *Paper URTec-5393 Presented at the Unconventional Resources Technology Conference*, Houston, Texas, USA, 2021.

[13] M. F. Tugan and R. Weijermars, "Improved EUR prediction for multi-fractured hydrocarbon wells based on 3-segment DCA: implications for production forecasting of parent and child wells," *Journal of Petroleum Science and Engineering*, vol. 187, no. 2019, article 106692, 2019.

[14] M. F. Tugan and R. Weijermars, "Variation in *b*-sigmoids with flow regime transitions in support of a new 3-segment DCA method: Improved production forecasting for tight oil and gas wells," *Journal of Petroleum Science and Engineering*, vol. 192, no. 2020, article 107243, 2020.

[15] M. F. Tugan and R. Weijermars, "Searching for the root cause of shale well rate variance: highly variable fracture treatment response," *Journal of Petroleum Science and Engineering*, vol. 192, no. 2022, article 109919, 2022.

[16] I. Gupta, C. Rai, C. Sondergeld, and D. Devegowda, "Variable exponential decline: modified Arps to characterize unconventional-shale production performance," *SPE Reservoir Evaluation & Engineering*, vol. 21, no. 4, pp. 1045–1057, 2018.

[17] D. S. Fulford and T. A. Blasingame, "Evaluation of time-rate performance of shale wells using the transient hyperbolic relation," in *SPE-167242-MS presented at the SPE unconventional resources conference Canada*, Alberta, Canada, 2013.

[18] P. P. Valko and W. J. Lee, "A better way to forecast production from unconventional gas wells," in *SPE-134231-MS presented at SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.

[19] A. N. Duong, "An unconventional rate decline approach for tight and fracture-dominated gas wells," in *SPE-137748-MS presented at Canadian Unconventional Resources and International Petroleum Conference*, Calgary, 2010.

[20] D. Ilk, J. A. Rushing, A. D. Perogo, and T. A. Blasingame, "Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using Arps' decline curves," in *Paper SPE-116731-MS presented at the SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA, 2008.

[21] H. Li and Y. Han, "Decline curve analysis for production forecasting based on machine learning," in *Paper SPE-189205-MS presented at the SPE symposium: production enhancement and cost optimization*, Kuala Lumpur, Malaysia, 2017.

[22] J. Sun, X. Ma, and M. Kazi, "Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells," in *Paper SPE-190104-MS Presented at the SPE Western Regional Meeting*, Garden Grove, California, USA, 2018.

[23] T. Mukherjee, T. Burgett, T. Ghanchi, C. Donegan, and T. Ward, "Predicting gas production using machine learning methods: a case study," in *Paper SEG-2019-3215692 presented at SEG International Exposition and Annual Meeting*, San Antonio, Texas, USA, 2019.

[24] N. Tamimi, S. Samani, M. Minaei, and F. Harirchi, "An artificial intelligence decision support system for unconventional field development design," in *Paper URTeC-2019-249 presented at the Unconventional Resources Technology Conference*, Denver, CO, USA, 2019.

[25] C. Temizel, C. H. Canbaz, O. Saracoglu et al., "Production forecasting in shale reservoirs through conventional DCA and machine/deep learning methods," in *Paper URTEC-2020-2878-MS presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference*, Virtual, July 2020.

[26] L. Xue, Y. Liu, Y. Xiong, Y. Liu, X. Cui, and G. Lei, "A data-driven shale gas production forecasting method based on the multi- objective random forest regression," *Journal of Petroleum Science and Engineering*, vol. 196, no. 2021, pp. 107801–107813, 2021.

[27] M. R. Gross, J. D. Hyman, S. Srinivasan et al., "A physics-informed machine learning workflow to forecast production in a fractured Marcellus shale reservoir," in *Paper URTEC-2021-5644-MS presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, Texas, USA, 2021.

[28] T. Doan and M. Vo, "Using machine learning techniques for enhancing production forecast in north Malay Basin," *Improved Oil and Gas Recovery*, vol. 5, no. 2021, pp. 1–7, 2021.

[29] J. J. Arps, "Analysis of decline curves," *Transactions of the AIME*, vol. 160, pp. 228–247, 1945.

[30] B. Kulga, E. Artun, and T. Ertekin, "Development of a data-driven forecasting tool for hydraulically fractured, horizontal wells in tight-gas sands," *Computational Geosciences*, vol. 103, pp. 99–110, 2017.

[31] Z. Lei, S. Wu, T. Yu et al., "Simulation and optimization of CO2 huff-n-puff processes in tight oil reservoir: a case study of Chang-7 tight oil reservoirs in Ordos Basin," in *Paper SPE-191873-MS presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition*, Brisbane, Australia, 2018.

[32] R. Zhang, W. Lv, Y. Zhou, B. Qu, Y. He, and X. Ma, "Research on oil-CO2-water relative permeability of the low permeability reservoir based on history matching," *Improved Oil and Gas Recovery*, vol. 2021, no. 5, pp. 1–12, 2020.

[33] Z. Dong, W. Tian, Y. Yang et al., "Assisted history-matching based on multiobjective surrogate reservoir model for tight oil reservoirs," *IOP Conference Series Earth and Environmental Science*, vol. 781, no. 2, article 022086, 2021.

[34] M. Paryani, O. O. Awoleke, M. Ahmadi, C. Hanks, and R. Barry, "Approximate Bayesian computation for probabilistic decline-curve analysis in unconventional reservoirs," *SPE Reservoir Evaluation & Engineering*, vol. 20, no. 2, pp. 478–485, 2017.

[35] M. Sayyafzadeh, M. Haghighi, and J. N. Carter, "Regularization in history matching using multi-objective genetic algorithm and Bayesian framework (SPE 154544)," in *Paper SPE-154544-MS Presented at the SPE Europec/EAGE Annual Conference*, Copenhagen, Denmark, 2012.

Hindawi
Geofluids
Volume 2022, Article ID 1983303, 12 pages
https://doi.org/10.1155/2022/1983303

Research Article

Data-Driven Method for Predicting Soil Pressure of Foot Blades within a Large Underwater Caisson

Can Huang,[1,2] Hao Zhu,[1] Kunyao Li,[1] Jianxin Zheng,[1] Hao Li,[1] Jiaming Li,[3] and Yao Xiao[1]

[1]*Research and Development Center of Transport Industry of Intelligent Manufacturing Technologies of Transport Infrastructure, CCCC Second Harbor Engineering Company Ltd., Wuhan 430000, China*
[2]*CCCC Highway Bridge National Engineering Research Centre Co. Ltd., Beijing 100120, China*
[3]*State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Jiaming Li; jiaming_li@mail.dlut.edu.cn

Received 24 October 2021; Revised 21 December 2021; Accepted 4 March 2022; Published 19 March 2022

Academic Editor: Keliu Wu

The soil pressure on the bottom surface of the foot blades is an important monitoring point during the sinking process of large underwater caissons. Complex soil-structure interactions occur during the sinking process, making it difficult to accurately predict the soil pressure of foot blades. Accurate construction processes often rely on data from the soil pressure of foot blades in the field. In this study, a data-driven approach is used to establish the relationship between the amount of sinking of the caisson and the soil pressure of foot blades. Furthermore, by improving the splitting method of the original Classification and Regression Tree (CART) algorithm, a single model's numerical prediction of 80-foot blades soil pressures is realized. The improved CART model, multilayer perceptron (MLP), long short-term memory (LSTM), and a linear regression model are compared through a comprehensive multiparameter evaluation method. Finally, this article discusses the deployment scheme of the model by comparing and analyzing the data in the time period of 10 : 00 on July 29, 2020, and 23 : 00 on August 7, 2020. The experimental results can satisfy the engineering demands and provide a basis for further data-driven intelligent control of large caisson sinking.

1. Introduction

As the main bridge engineering deep foundation, a caisson has the advantages of excellent integrity, high load-bearing capacity, superior structural stiffness, small floor area, and good seismic performance [1–3]. By extracting soil from the well, the caisson uses its gravity and sinking aid to overcome buoyancy and soil resistance in order to sink. The sinking process can be analyzed using the soil pressure of foot blades to calculate the end resistance, thus, providing a basis for instructions during construction [4]. However, with the gradual increase in the number of large caissons being built, the theories developed for small foundations may not predict the soil pressure of large caissons accurately [5].

At present, some achievements have been made in studying caisson foot blades. For example, Jiang et al. [6] found that the sand migration during the sinking of the Hutong Yangtze River Bridge has a relatively significant effect on the soil pressure of foot blades through traditional model experiments. Yan et al. [7] studied the caisson of the Oujiang River North Estuary Bridge as an example to demonstrate the effective reduction of the soil pressure of foot blades by layered excavation during the construction of the caisson. Zhang et al. [8] and Baogang et al. [9] found that the sudden sinking of the caisson during construction was related to the short time decrease of the soil pressure of foot blades. Yea and Kim [10] investigated the three-dimensional distribution pattern of the soil pressure of foot blades during sinking for the caisson of Youngjong Grand Bridge by field experiments. The characteristics of soil pressure of caisson in different engineering backgrounds have been well studied. However, due to complex soil-structure interactions, mechanical properties have significant nonlinearity and plasticity during the sinking of the caisson [11]. Existing descriptive design approaches do not readily

capture the accurate prediction of the soil pressure of foot blades. In recent years, during the construction of a caisson, a series of sensors are usually installed to obtain information on the dynamics of the sinking foot blades of the caisson [12]. However, the data analysis capability is not robust; therefore, it cannot predict the soil pressure state of the caisson's foot blades, which increases the uncertainty of caissons construction.

With the development of artificial intelligence and machine learning technology, data-driven applications are being researched and applied in many fields. Most of the current data-driven bridge engineering applications are focused on bridge health monitoring [13]. For example, a data-driven and computer vision-based approach to automatically identify pitting corrosion [14], crack recognition based on a convolutional neural network [15], and a hybrid artificial neural network-based imperial competitive algorithm used to predict damage of slab-on-girder bridge structures [16]. However, there are minimal studies on data-driven methods for bridge construction processes, especially for large caisson construction. In terms of algorithm analysis, most data-driven bridge engineering-based applications focus on the output of a single target. Prediction of multiclass damage [17] and crack width [18] by algorithms such as convolutional neural networks. Output values such as these have only one category or a single regression objective, making it difficult to solve the current problem of the multilabel regression of the caisson soil pressure of foot blades. There are two main solutions to the multilabel problem. One type of processing is through problem transformation, which focuses on manipulating and processing the dataset in order to transform the multilabel learning problem into one or more single-label problems [19–21]. The other is via algorithmic adaptation methods, by improving existing methods that are needed to be directly suitable for learning multilabel datasets [22–24]. In the caisson sinking dataset, the values of the multiple soil pressures of foot blades are predicted simultaneously, and this kind of multilabel processing during the bridge construction process needs to be further studied. Regarding model evaluation indexes, regression problems are generally evaluated by a single index such as mean square error or a fitting coefficient. However, during multimodel evaluation, it is easy to have inconsistent index tendencies of different models, making it difficult for multiple models to conduct comprehensive and accurate quantitative evaluations. Therefore, the comprehensive judgment of multiple models is conducive to comparing and selecting models during the experimental process.

An algorithm based on an improved Classification and Regression Tree (CART) implemented for multilabel prediction will be investigated in detail within this study. Specifically, the primary contributions of this study are as follows: (1) the performance of neural network methods and improved CART models for multilabel foot blades regression prediction of soil pressure is investigated by comparison. Among them, multilayer perceptron (MLP) and long short-term memory (LSTM) are chosen as typical neural network representatives. (2) A multilabel comprehensive evaluation method is improved, and a model comparison and a parameter optimization are performed through comprehensive evaluation indexes. (3) The results of the impact of the model on different learning approaches during field tests are discussed. Based on the natural advantages of the data-driven approach, the present method can be easily extended to other scenarios of caisson construction after sufficient data is collected.

The framework of this study is shown in Figure 1. The second section will explain the engineering background as well as the data acquisition and preprocessing of the GPS data and the soil pressure of foot blades. The third section will investigate the improved CART algorithm. The fourth section focuses on parameter optimization and model comparison based on the improved multiparameter integrated evaluation index. In the fifth part, the deployment scheme of the model and the field prediction results is discussed.

2. Engineering Background and Data Preparation

2.1. Engineering Background. The main channel bridge of the Changtai Yangtze River Bridge is a double-layer cable-stayed bridge. The upper layer of the bridge is a highway, and the lower layer is an intercity railway and ordinary highway. The two pylons of the main channel bridge adopt a large-scale steel caisson foundation, as shown in Figure 2. The foundation plane of the caisson at pier #5 of the main bridge is round-end, the elevation is stepped, and the width of the step is 9.0 m. The bottom surface of the caisson is 95.0 m long, 57.8 m wide, and the radius of the round end is 28.9 m. The top surface of the caisson is 77.0 m in length, 39.8 m in width, and 19.9 m in radius at the round end. The outer wall of the caisson is 1.8 m thick and 43 m high, the inner wall thickness is 2.0 m, and the height is 64 m. The inner and outer ring partition walls are both 1.4 m thick, the outer ring partition wall is 64 m high, and the inner ring partition wall is 39 m high. The standard size of the inner wellbore is 11 m in length and 11 m in width. The partition and inner shaft wall are inverted by 1.5 m in length and 1.5 m in width. The steel shell structure has 28 compartments. The river section where the steel caisson is located is a tidal section of the lower reaches of the Yangtze River. The tidal level is affected by both the Yangtze River runoff and the tide. The 20-year encounter bridge cross-section vertical average maximum velocity ranges from 1.93 m/s to 2.1 m/s. The dry period vertical average maximum velocity is less than 1.05 m/s. The steel caisson is located on the north side of the main channel area. The topography of the pier is relatively stable. The surface layer of the riverbed is loose silt, with an uneven layer thickness that ranges from 11.6 to 4.8 m thick and has poor engineering properties. Most of the sandy soil layers drilled into pier #5 reveals a sandy gravel cement layer, which is a nonlayered structure and is distributed sporadically, revealing that the depth primarily ranges from -35 to -45 m under the river bed.

The location hydrology and geological conditions of Changtai Yangtze River Bridge are complex, the volume of the caisson is large, and the structure form is particularly novel. The specific characteristics and construction

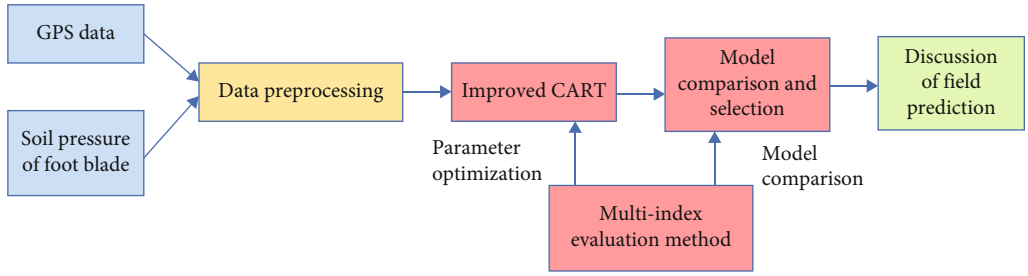


Figure 1: The flow chart of the soil pressure of foot blades prediction.

Figure 2: Caisson under construction.

difficulties are mainly concentrated in that the safety risk and attitude control of initial caisson are difficult. As well as during construction, the caisson has been in a state of dynamic balance, requiring many monitoring items and a large number of measuring points. Traditional monitoring systems have lagging information feedback, which cannot predict the soil pressure of foot blades accurately, thus, increasing the uncertainty of caisson construction.

2.2. Data Preparation. In order to monitor the sinking posture of the caisson and adjust the construction plan in time, Beidou GPS sensors are arranged on the top of the caisson, and the results of the manual high-precision measurement are checked every day to ensure that the monitoring data is accurate and reliable. The installation position of the Beidou control points is shown in Figure 3(a). Figure 3(b) shows the installation positions of the soil pressure of foot blades sensors.

In Figure 3(a), h_1, h_2, h_3, h_4, and h_5 are the measured GPS values, and h_5 is the average value of the four measured points. The sinking amount (SA) is calculated by the following equation:

$$\mathrm{SA}_i = h_{i,t} - h_{i,t-1}, \tag{1}$$

where i is 1-5, representing the sinking amount of 5 locations, $h_{i,t}$ is the vertical value of the i-th monitoring position at time t, and $h_{i,t-1}$ is the monitoring data of the vertical direction at time $t-1$ of the i-th monitoring position.

The sinking data from July 19, 2020, to July 29, 2020, and the soil pressure of foot blades data for the corresponding times formed the data set, as shown in Tables 1 and 2.

The sinking data is recorded every 10 minutes, and the soil pressure of foot blades is recorded every 30 minutes. During the complex construction of the caisson, some of the foot blades sensors are damaged despite the many protections made to protect the sensors; therefore, 80 of them with normal sensors were selected for learning and prediction. Data aggregation of the sinking amount data and the 80 foot blades soil pressure data is based on a 1 h period; thus, the raw dataset was constructed. The raw dataset contains a total of 249 samples and 85 features (5 features for the sinking amount and 80 features for the soil pressure of foot blades). The model was tested on 5 out of the 249 samples from July 29 at 4 : 00 to July 29 at 8 : 00. The remaining 244 were tested for model training and validation. The training and validation sets were split according to 80% (195 samples) and 20% (49 samples), respectively. The model was trained via the training set and the validation set evaluated the model. Mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the fitted coefficient (R^2) were calculated for each model.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)|, \tag{3}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{5}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the true value of the sample, and n is the number of samples.

3. Improved CART Algorithm

A decision tree is an example-based inductive learning approach that constructs a tree-like regression model from the given samples. It is a relatively simple algorithm with excellent robustness compared to other regressions.

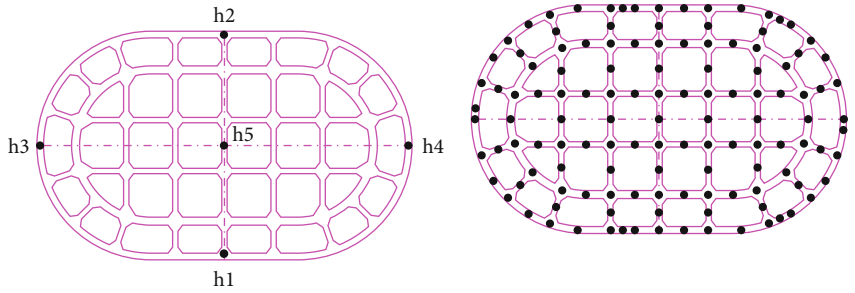


Figure 3: Sensor locations. (a) Beidou GPS sensor locations. (b) Bottom soil pressure of foot blades sensor locations.

Table 1: Sample sinking amount data.

Time	SA-1	SA-2	SA-3	SA-4	SA-5
2020-07-19 00 : 00	5.701	5.413	5.554	5.534	5.568
2020-07-19 00 : 10	5.734	5.429	5.582	5.534	5.632
2020-07-19 00 : 20	5.715	5.386	5.553	5.511	5.601
2020-07-19 00 : 30	5.749	5.442	5.582	5.531	5.605
2020-07-19 00 : 40	5.738	5.435	5.577	5.536	5.599
2020-07-19 00 : 50	5.767	5.456	5.604	5.562	5.633

Table 2: Sample soil pressure of foot blades data.

Time	RF-1	RF-2	RF-16	RF-124	RF-125
2020-07-19 00 : 00	0.100	0.100	0.775	2.968	2.204
2020-07-19 00 : 30	0.100	0.100	0.806	2.964	2.201
2020-07-19 01 : 00	0.100	0.100	0.929	2.959	2.195
2020-07-19 01 : 30	0.100	0.100	0.796	2.953	2.189
2020-07-19 02 : 00	0.100	0.100	0.638	2.941	2.180
2020-07-19 02 : 30	0.100	0.100	1.008	2.936	2.172

Research on the soil pressure of foot blades prediction based on CART [25] is carried out in this study and introduces how to use the decision trees in order to solve the regression problem; lets us define the input matrix x as

$$x = \left(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}\right) = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & x_1^{(4)} & x_1^{(5)} \\ x_2^{(1)} & \ddots & & \vdots & \\ \vdots & & \ddots & & \vdots \\ x_n^{(1)} & & & \ddots & \vdots \\ \vdots & & & & \vdots \\ x_N^{(1)} & x_N^{(2)} & x_N^{(3)} & x_N^{(4)} & x_N^{(5)} \end{pmatrix}, \tag{6}$$

where $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}$ represent the input monitoring SA_1, SA_2, SA_3, SA_4, and SA_5 sinking amount data, also called features, and N represents the number of samples. Each sample strip corresponds to 80 foot blades soil pressure points in the following equation.

$$y_n = \left\{y_{n,1}, \cdots, y_{n,q}, \cdots, y_{n,80}\right\}. \tag{7}$$

Figure 4: Training methods.

It is not possible to directly output 80 foot blades soil pressure values with conventional CART. For the multiobjective regression problem, there are two general solutions: the first one is to convert the multiple regression problem into a single-objective regression, such that if 80 foot blades soil pressure values need to be predicted, then 80 single regression models are constructed, and the soil pressure values at different locations are output. However, this method is too cumbersome, and the training complexity is large and time consuming. The alternative way is to build a multiobjective CART regression model [26]:

$$\min_{p,a}\left[\min_{c_1}\sum_{x_n\in R_1(p,a)}(y_n - c_1)^2 + \min_{c_2}\sum_{x_n\in R_2(p,a)}(y_n - c_2)^2\right], \tag{8}$$

$$\min_{p,a}\left[\min_{c_1}\sum_{x_n\in R_1(p,a)}\left(y_{n,q} - c_{1,q}\right)^2 + \min_{c_2}\sum_{x_n\in R_2(p,a)}\left(y_{n,q} - c_{2,q}\right)^2\right]. \tag{9}$$

Equation (8) is the traditional CART regression node division method, where yn is the target value, $c1$ and $c2$ are the predicted values within the intervals of $R1$ and $R2$, respectively. The minimum point a is chosen as the splitting point by calculating the mean square error of $R1$ and $R2$. The target values in this method are not multidimensional, and the splitting process cannot calculate the loss under the multidimensional data, so the traditional CART cannot be carried out for multiobjective regression. With the improved splitting methods of Eq. (8), Eq. (9) integrates

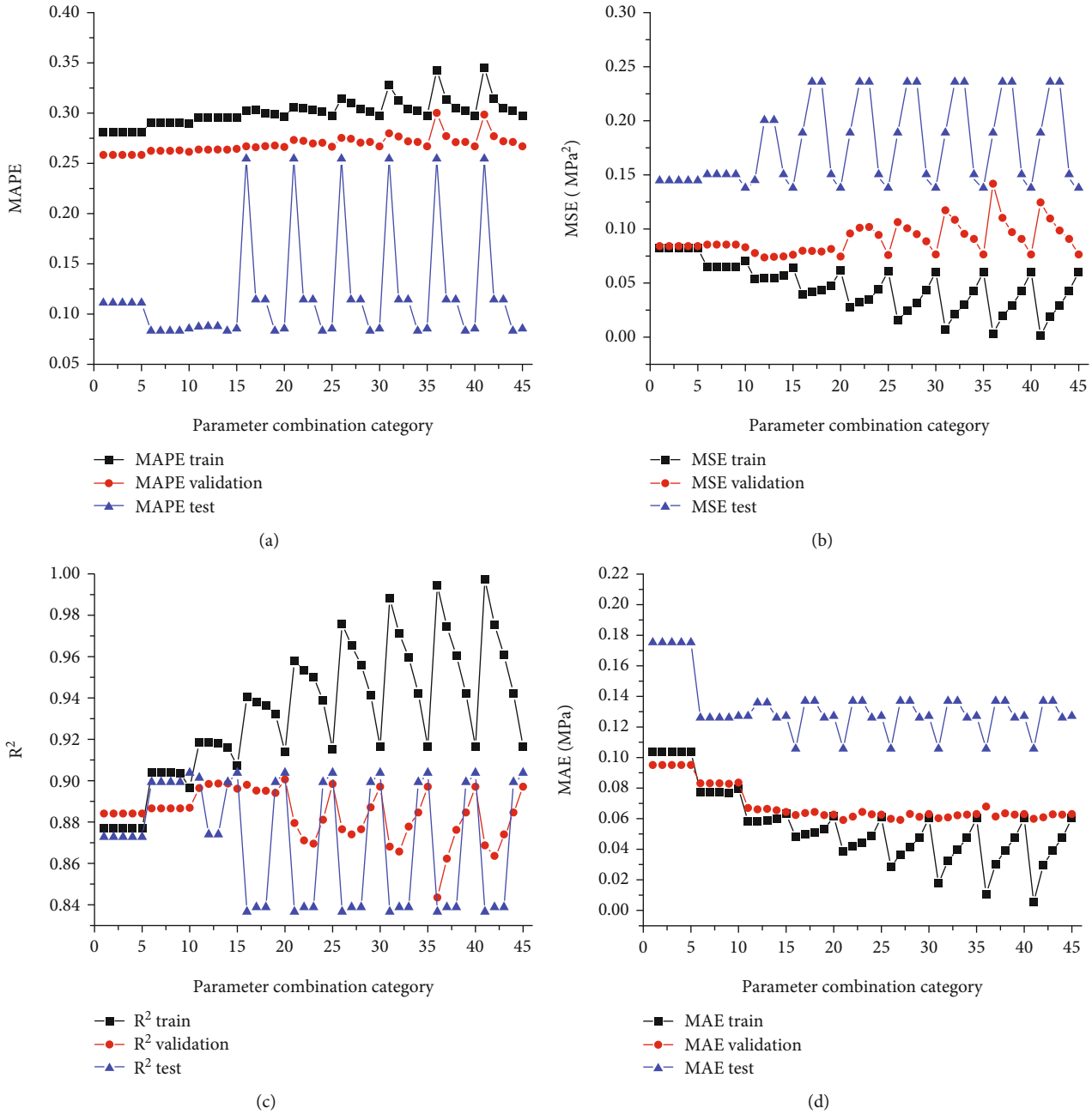


Figure 5: Continued.

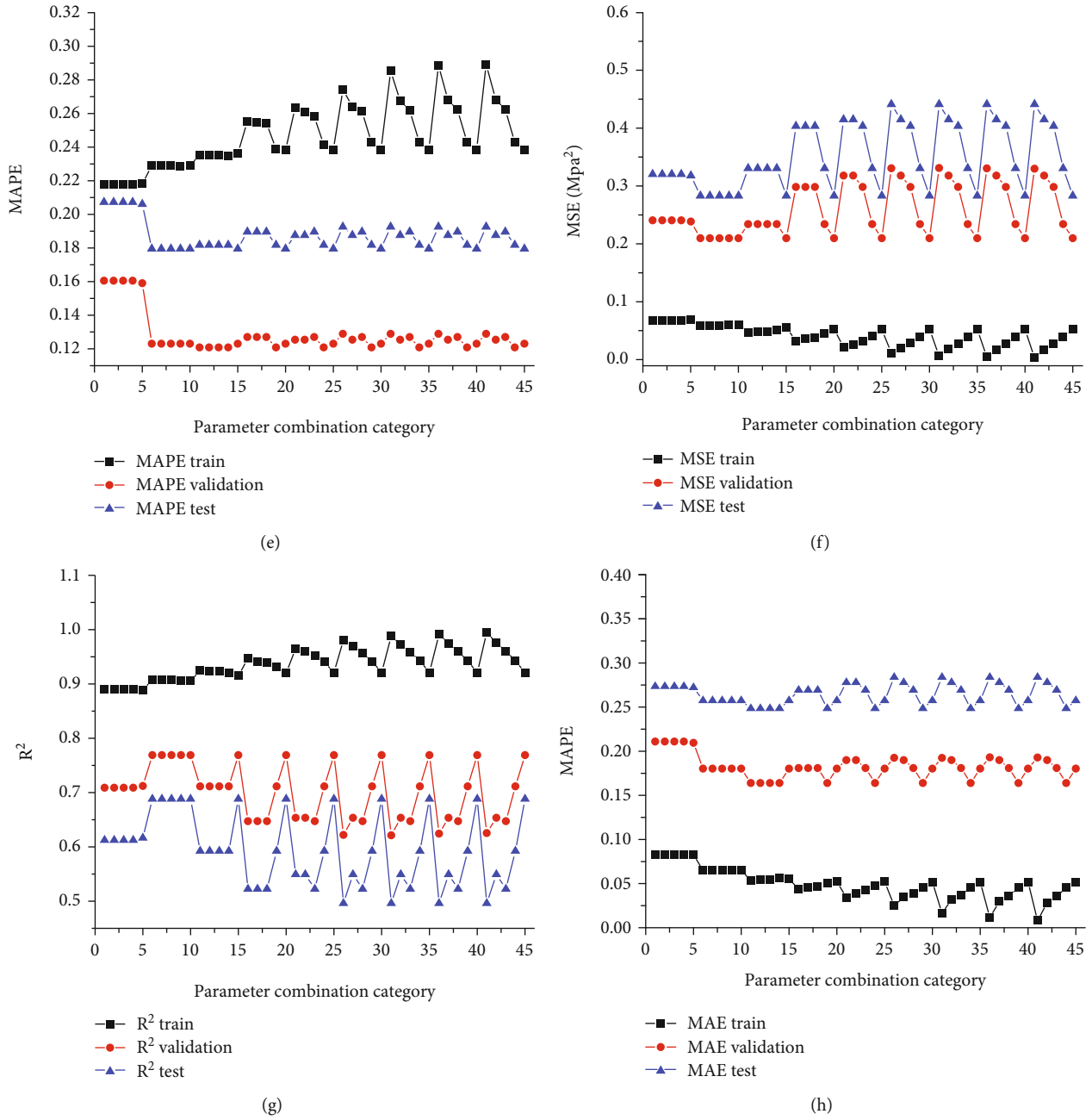


Figure 5: Results of different training methods. (a) Disordered training method MAPE value. (b) Disordered training method MSE value. (c) Disordered training method R^2 value. (d) Disordered training method MAE value. (e) Ordered training method MAPE value. (f) Ordered training method MSE value. (g) Ordered training method R^2 value. (h) Ordered training method MAE value.

the minimum loss of the q-dimensional output value as the splitting point when calculating the loss. The leaf nodes of CART can be written explicitly as

$$\hat{c}_1 = \frac{1}{N_m} \sum_{x_n \in R_1(p,a)} y_n \text{ and } \hat{c}_2 = \frac{1}{N_m} \sum_{x_n \in R_2(p,a)} y_n, \tag{10}$$

$$\hat{c}_{1,q} = \frac{1}{N_m} \sum_{x_n \in R_1(p,a)} y_{n,q} \text{ and } \hat{c}_{2,q} = \frac{1}{N_m} \sum_{x_n \in R_2(p,a)} y_{n,q}. \tag{11}$$

Equation (10) is the output form of the traditional model that calculates the average value of the $R1$ and $R2$ regions as the output of the target values. Improving the above method in order to obtain Eq. (11), the average value of the output target in q dimensions was calculated as the output of multiple targets within the $R1$ and $R2$ regions, respectively.

4. Analysis of the Experimental Results

4.1. Improved CART Training Methods. We split the data according to the temporal order for the prediction of the soil

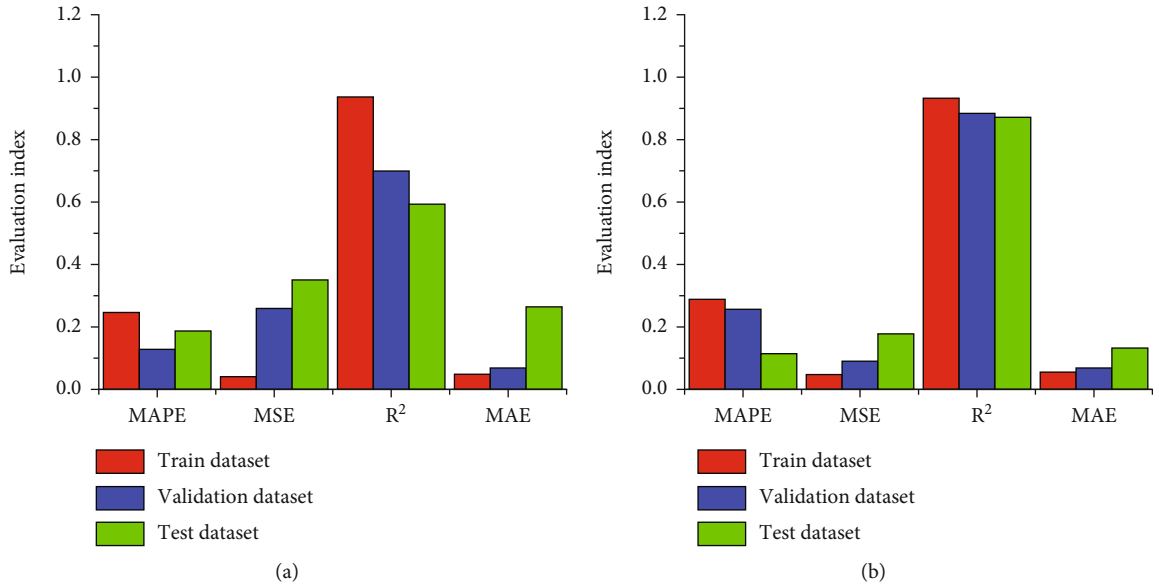


Figure 6: The average value of the index under different training methods. (a) Ordered training of the mean value of each index. (b) Disordered training of the mean value of each index.

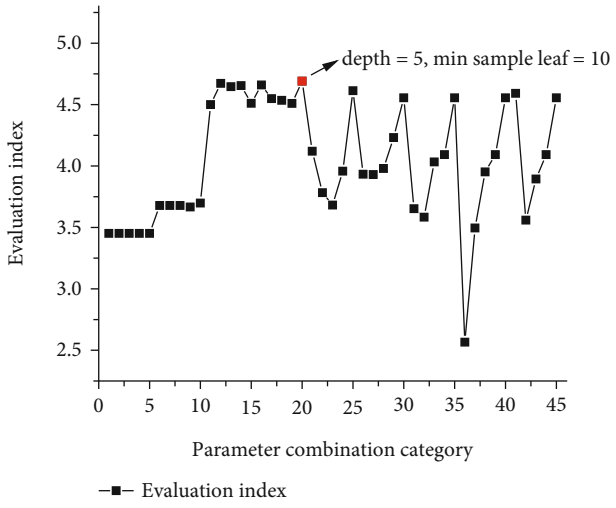


Figure 7: Integrated evaluation index.

pressure of foot blades, where the training was performed separately according to whether the data set was shuffled or not. Examples are shown in Figure 4(a) below, blue 1-10 are the 10 samples on the timeline, and red 1-5 are the test samples. We used two training methods, the first one is shown in Figure 4(b), according to the time order, the first 80% of the data is used as the training set (data samples 1-8 in the yellow area) and 20% as the validation set (samples 9-10 in the green area). Alternatively, as shown in Figure 4(c) ignoring, the temporal order, the overall random selection divides the training and validation sets, i.e., 1-10 in the original data are shuffled, 80% of the data is randomly selected as the training set and 20% as the validation set.

A grid search method is used to select the parameters for the maximum depth of the decision tree and the minimum number of leaf node samples, where the maximum depth is 2, 3, 4, 5, 6, 7, 8, 9, and 10, and the minimum number of leaf node samples is 1, 2, 3, 5, and 10. There are 45 combinations of the maximum depth and the minimum number of leaf node samples, such as <maximum depth is 2 and the minimum number of leaf node samples is 1>, and <maximum depth is 2 and the minimum number of leaf node samples is 2>. In Figure 5, (a), (b), (c), and (d) are the results of training method 2 (disordered method) in Figure 5, and (e), (f), (g), and (h) are the results of training method 1 (ordered method) in Figure 5, respectively.

As the depth increases, the nonlinear representation of the tree model is gradually strengthened. The metrics of both ordered and disordered training on the training set show a step-up trend. The accuracy of the ordered training, however, gradually decreases on the validation and test sets. For the regression coefficient, for example, when the depth is 2, and the minimum number of leaf node samples 1, the fitting coefficient of the training set is 0.8900, and when the depth is 10, and the minimum number of leaf node samples is 1, the fitting coefficient of the training set is 0.9947. The accuracy improved by 11.76%; however, the validation and test sets decreased by 37.16%, and the accuracy of the test set decreased by 50.17%. In the disordered training, the fit coefficient of the training set is 0.8768 when the depth is 2, and the number of minimum leaf samples is 1. It increased to 0.9973 when the depth is increased to 10, while the maximum decrease in the fit coefficient of the validation set and the test set is approximately 16%. As the depth increases, the performance of the training set gradually increases, and the performance of both the validation and test sets decreases, especially within the ordered training.

Increases in the minimum number of leaf node samples at the same depth show an improvement in the robustness of the model, in the model at a depth of 10, the number of minimum leaf samples in the validation set ranges from 1 to 10 in the disordered training, and the fit coefficient increases

Table 3: Cross-validation results.

Validation set / Model	MAPE	MSE (MPa^2)	R^2	MAE (MPa)	Test set MAPE	Test set MSE (MPa^2)	Test set R^2	Test set MAE (MPa)
MLP	0.1127	0.0959	0.8545	0.1273	0.1852	0.2186	0.8634	0.2584
LSTM	0.1988	0.1482	0.7754	0.2109	0.6275	0.9932	-4.3973	0.6421
Linear regression	0.0900	0.0868	0.8764	0.0931	0.9598	0.2793	0.8509	0.3126
Improved CART	0.0487	0.0487	0.8862	0.0671	0.1114	0.1516	0.8632	0.1658

Table 4: Multiparameter model evaluation results.

Model	Validation set $\overline{MAPE}$	Validation set $\overline{MSE}$	Validation set $\overline{R^2}$	Validation set $\overline{MAE}$	Test set $\overline{MAPE}$	Test set $\overline{MSE}$	Test set $\overline{R^2}$	Test set $\overline{MAE}$	Integrated evaluation index value	Ranking
BP	0.2479	0.2669	0.7139	0.3064	0.5492	0.6383	1.0000	0.5169	4.2395	2
LSTM	0.0000	0.0000	0.0000	0.0000	0.0701	0.0000	0.0000	0.0000	0.0701	4
Linear regression	0.3922	0.3462	0.9116	0.5904	0.0000	0.4604	0.9976	0.3669	4.0653	3
Improved CART	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	8.0000	1

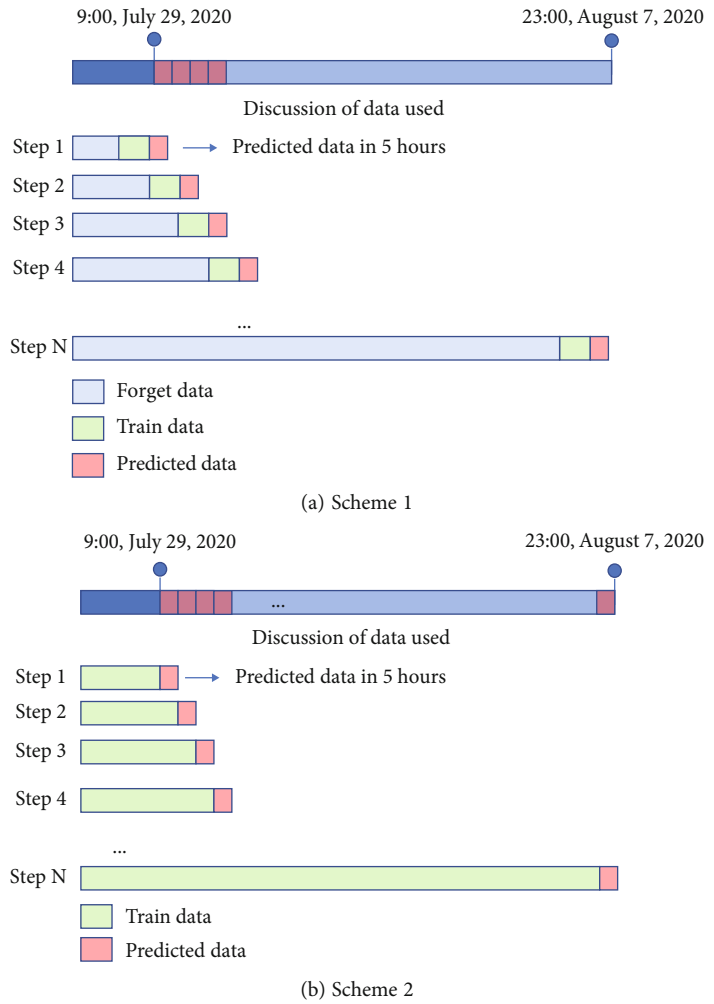


Figure 8: Prediction scheme.

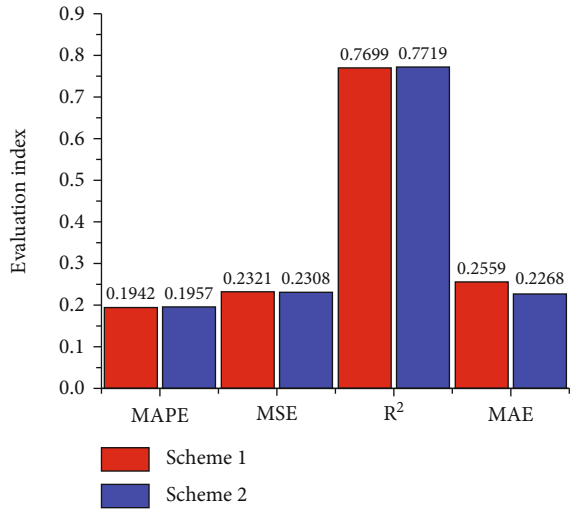


Figure 9: Results of different forecasting schemes.

from 0.8687 to 0.9038. It indicates that in the depth determined case, the accuracy of the validation model and test sets can be improved by increasing the number of minimum leaf samples.

In the ordered training method, since the data set is split according to the temporal order, the construction conditions change more significantly over time, and the data learning in the preceding time period may not be sufficient to support the later predictions. Prediction ability is significantly reduced in the training, validation, and test sets in Figure 6(a), realizing that MSE, R^2, and MAE indicators show the best results in the training set, followed by the validation set, and the worst was observed in the test set. The gap between the training and test set indicators is relatively large; in terms of the fit coefficient, the test set is reduced by 36.65% compared to the training set. In Figure 6(b), the average value of the disorder training is shown. Because the model learns the data characteristics of extended time periods and multiple working conditions during training, the model realizes a better robustness. The average fitting coefficient decreases from 0.9325 in the training set to 0.8716, which is 6.53% lower, and the stability of the model is better compared with the ordered training method.

In the evaluation metrics, the closer the R^2 is to 1, the closer the MSE, MAE, and MAPE are to 0, the better the model performance is. Different indicators do not have the same tendency to evaluate the predictive capability of the model. Furthermore, in this study, the MSE, MAE, and MAPE are transformed as follows:

$$\mathrm{MSE}' = \frac{1}{100 \times \mathrm{MSE}}, \tag{12}$$

$$\mathrm{MAE}' = \frac{1}{100 \times \mathrm{MAE}}, \tag{13}$$

$$\mathrm{MAPE}' = \frac{1}{100 \times \mathrm{MAPE}}. \tag{14}$$

In the regression models, different evaluation indicators are focused on different aspects. The evaluation of different models via a single indicator lacks comprehensiveness. Accordingly, a comprehensive evaluation of different models with multiple indicators is needed. The ranking method proposed by Zorlu et al. [27] in 2008 is a commonly used multi-index comprehensive evaluation method. Zhang et al. [28] carried out optimization in terms of the tendency uniformity and normalization based on Zorlu. However, there is no comprehensive consideration of the different data sets, and the following steps further improve the above study.

Step 1. Calculate the evaluation indexes of MSE, MAE, MAPE, and R^2 for the training set and validation set, respectively.

Step 2. MSE′, MAE′, and MAPE′ by tendency uniform conversion of MSE, MAE, and MAPE.

Step 3. MSE′, MAE′, MAPE′, and R^2 are normalized by the formula:

$\bar{I}_M = I_M - \max(I)/\max(I) - \min(I)$ to get $\overline{\mathrm{MSE}}$, $\overline{\mathrm{MAE}}$, $\overline{\mathrm{MAPE}}$, and $\overline{R^2}$.

Step 4. $\overline{\mathrm{MSE}}$, $\overline{\mathrm{MAE}}$, $\overline{\mathrm{MAPE}}$, and $\overline{R^2}$ of the training and validation set are summed to get the comprehensive evaluation index.

The evaluation metrics for different parameter combinations of the disordered training under the training and validation sets are calculated according to the above comprehensive evaluation algorithm to obtain Figure 7, showing that the model has the highest comprehensive evaluation metrics at a depth of 5 and a minimum number of leaf node samples of 10.

4.2. Comparison of Model Performances. A comparison of the test results of MLP [29], LSTM [30], and the linear regression [31] models are analyzed in this study. The 5-fold cross-validation method is used to divide the sample into 5 equal parts according to 20%, and 4 parts are taken for training and the other 1 part for validation. The average of the five results is used as the evaluation result of the model validation set. The test set evaluation results are obtained by the last five hours of data.

Average values of the 5-fold cross-validation of the MLP, LSTM, linear regression model, and the improved CART model are shown in Table 3. The linear regression model, which requires the modeling of each soil pressure of foot blade feature; thus, constructing 80 linear regression models to meet the engineering requirements, is more complex and computationally intensive compared to the improved CART model. Table 4 is obtained through tendency transformation, and after normalization, the multiparameter model

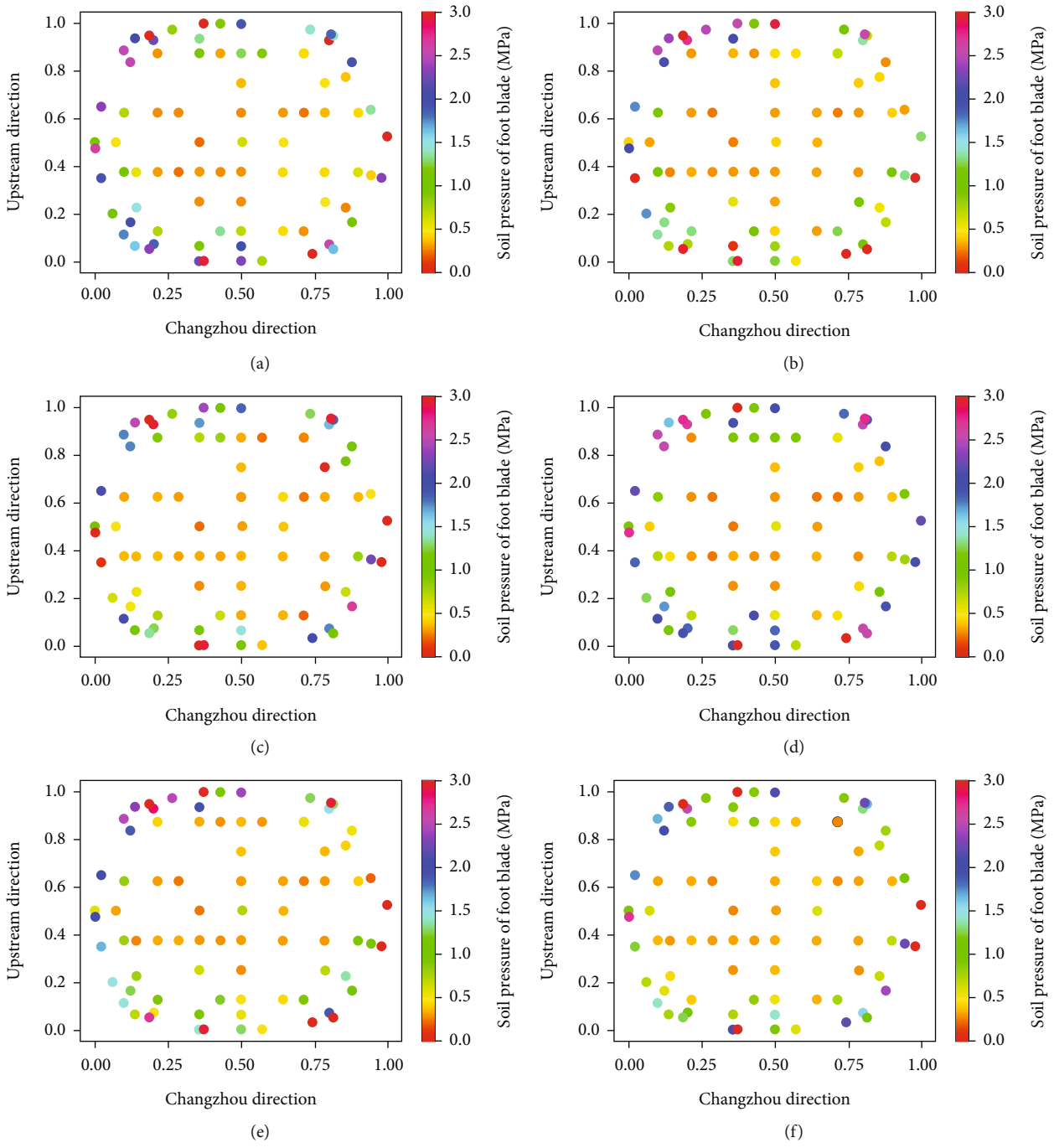


Figure 10: Real and predicted values of the model at different moments in time. (a) July 30, 2020 20 : 00 real value; (b) August 3, 2020 20 : 00 real value; (c) August 7, 2020 18 : 00 real value; (d) July 30, 2020 20 : 00 predicted value; (e) August 3, 2020 20 : 00 predicted value; (f) August 7, 2020 18 : 00 predicted value.

evaluation results are shown in Table 3. Various index results of the validation and the test sets were added to combine and evaluate the robustness of the model simultaneously. Finally, the results were ranked. Based on the evaluation results of the multi-index model in Table 4, it is clear that the improved CART model performs the best in all indexes compared with other models in the validation and test sets in terms of the soil pressure of foot blades prediction it was ranked first overall.

5. Discussion

Previous analytical studies confirm the forecasting schemes and the forecasting algorithms, but whether the model has

sufficient capacity for field forecasting is subject to further discussion. A total of 230 data points from July 29, 2020, 10 : 00 to August 7, 2020, 23 : 00, are used in the discussion as the evaluation data set for the long-term prediction of the model. Due to the differences in the sinking conditions of the caissons, two schemes were selected for the field prediction. In Figure 8(a) scheme one, forgetting is used for the first scheme, where the model learns 43 hours of data, predicts the next five hours of data, and updates the model every five hours. Figure 8(b) scheme two, continuous accumulation of the data set, each time is predicting the next five hours of data, and then updating the model once.

The prediction results of schemes 1 and 2 are shown in Figure 9. It is found that there is no significant difference between the two schemes with a fitting coefficient of 0.7699 for scheme 1 and 0.7719 for scheme 2, under the detection of 230 data points. It shows that in the field forecasting process, it is only necessary to update the model at regular intervals in order to achieve a relatively good forecasting effect. We extracted the true and predicted values of the data at 20 : 00 on July 30, 2020, 20 : 00 on August 3, 2020, and 18 : 00 on August 7, 2020, for visualization. The model's predicted values for the soil pressure of foot blades match the real values, as shown in Figure 10.

6. Conclusion

During the sinking of large caissons, monitoring the soil pressure of foot blades plays a key role during its bridge engineering-based applications safe and smooth sinking. In this study, research based on the data-driven prediction of the soil pressure of foot blades during sinking was developed using the Changtai Yangtze River Bridge caisson as an example, several conclusions were reached as follows:

(1) The multilabel task of a single model for the soil pressure of foot blades of a caisson was achieved by improving the splitting rule of CART. A multiparameter model evaluation algorithm was implemented to select parameters for the maximum depth and a minimum number of samples of the minimum leaf node of the improved CART. The optimal combination of parameters with a maximum depth of 5 and a minimum number of leaf node samples of 10 was selected. In the test set, MAPE is 0.1114, MSE is 0.1516 MPa^2, R^2 is 0.8632, and MAE is 0.1658 MPa

(2) The improved multiparameter model evaluation algorithm compared and analyzed the improved CART, MLP, LSTM, and the linear regression models. It is concluded that the CART model is more suitable for predicting the soil pressure of foot blades during the sinking of the caisson

(3) A total of 230 samples from July 29, 2020, 10 : 00 to August 7, 2020, 23 : 00, were used to continue the evaluation of the improved CART model. Comparison of the two prediction schemes for whether the data were forgotten or not revealed no significant differences, with the average fit coefficient being approximately 0.77. The results of the experiments can be satisfied with the engineering requirements

(4) A shift from empirical decision making to a data-driven based approach needs to be further investigated. Data-driven predictions of the soil pressure of foot blades are part of the overall intelligent construction of the caisson

Data Availability

The data in this study are available from the corresponding author upon reasonable request.

Additional Points

Computer Code Availability. The code in this study is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Kunyao Li and Jiaming Li wrote the manuscript. Can Huang provided suggestions for the research method and revised the article completely. Hao Li and Yao Xiao calculated and analyzed data. Hao Zhu, Jianxin Zheng, Hao Li, and Yao Xiao processed the basic data.

Acknowledgments

This research was funded by the National Development Plan (2021YFF0500900 and 2021YFF0500902).

References

[1] F. E. Lefler and V. D. R. Romero, "Reinforced concrete caissons for port structures in Spain," *Proceedings of the Institution of Civil Engineers: Maritime Engineering*, vol. 162, no. 2, pp. 73–81, 2009.

[2] S. Chakrabarti, P. Chakrabarti, and M. Krishna, "Design, construction, and installation of a floating caisson used as a bridge pier," *Journal of Waterway, Port, Coastal, and Ocean Engineering*, vol. 132, no. 3, pp. 143–156, 2006.

[3] T. Matsuda, K. Maeda, M. Miyake, J. Miyamoto, H. Sumida, and K. Tsurugasaki, "Instability of a caisson-type breakwater induced by an earthquake-tsunami event," *International Journal of Geomechanics*, vol. 16, no. 5, article C4016003, 2016.

[4] M. Guo, X. Dong, S. Wang, and G. Lan, "Study on the change law of resistance at the sinking end of soil extraction of super-large caisson foundation," *Journal of Rock Mechanics and Engineering*, vol. 40, no. S1, pp. 2976–2985, 2021.

[5] M. Guo, X. Dong, and J. Li, "Study on the earth pressure during sinking stage of super large caisson foundation," *Applied Sciences*, vol. 11, no. 21, article 10488, 2021.

[6] B. N. Jiang, M. T. Wang, T. Chen, L. L. Zhang, and J. L. Ma, "Experimental study on the migration regularity of sand outside a large, deep- water, open caisson during sinking," *Ocean Engineering*, vol. 193, p. 106601, 2019.

[7] X. Yan, W. Zhan, Z. Hu, D. Xiao, Y. Yu, and J. Wang, "Field study on deformation and stress characteristics of large open caisson during excavation in deep marine soft clay," *Advances in Civil Engineering*, vol. 2021, Article ID 7656068, 2021.

[8] D. Y. Zhang Zhicheng, F. Zheng, and J. Wang, "Analysis on sudden sinking behaviors of massive open caisson in deep-thick soft clay area," *Chinese Journal of Underground Space and Engineering*, vol. 16, no. 3, pp. 933–943, 2020.

[9] M. U. Baogang, W. Yan, Z. H. U. Jianmin, W. Gong, and O. Zuliang, "Analysis of large caisson sinking measured resistance," *Journal of Civil and Environmental Engineering*, vol. 34, no. S1, pp. 107–115, 2012.

[10] G. G. Yea and T. H. Kim, "Vertical cutting edge forces measured during the sinking of pneumatic caisson," *Marine Georesources & Geotechnology*, vol. 30, no. 2, pp. 103–121, 2012.

[11] K. Sobhan, "Principles of Geotechnical Engineering: Eighth Edition," in *Pws*, pp. 1–13, Cengage Learning, USA, 2012.

[12] R. Royston, B. B. Sheil, and B. W. Byrne, "Monitoring the construction of a large-diameter caisson in sand," *Proceedings of the Institution of Civil Engineers: Geotechnical Engineering*, vol. 12, pp. 1–48, 2020.

[13] X. H. Zhou and X. G. Zhang, "Thoughts on the development of bridge technology in China," *Engineering*, vol. 5, no. 6, pp. 1120–1130, 2019.

[14] N. D. Hoang, "Image processing-based pitting corrosion detection using metaheuristic optimized multilevel image thresholding and machine-learning approaches," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6765274, 2020.

[15] X. Kong and J. Li, "Vision-based fatigue crack detection of steel structures using video feature tracking," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 783–799, 2018.

[16] M. Gordan, H. A. Razak, Z. B. Ismail, K. Ghaedi, Z. X. Tan, and H. H. Ghayeb, "A hybrid ANN-based imperial competitive algorithm methodology for structural damage identification of slab-on-girder bridge using data mining," *Applied Soft Computing*, vol. 88, p. 106013, 2020.

[17] S. Sony, S. Gamage, A. Sadhu, and J. Samarabandu, "Multiclass damage identification in a full-scale bridge using optimally tuned one-dimensional convolutional neural network," *Journal of Computing in Civil Engineering*, vol. 36, no. 2, article 04021035, 2022.

[18] K. Yang, Y. Ding, P. Sun, H. Jiang, and Z. Wang, "Computer vision-based crack width identification using F-CNN model and pixel nonlinear calibration," *Structure and Infrastructure Engineering*, pp. 1–12, 2021.

[19] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.

[20] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker, "Label ranking by learning pairwise preferences," *Artificial Intelligence*, vol. 172, no. 16-17, pp. 1897–1916, 2008.

[21] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Random k-labelsets for multilabel classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1079–1089, 2011.

[22] A. Clare and R. D. King, *Knowledge Discovery in Multi-Label Phenotype Data*, Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, 2001.

[23] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in pp. 681–687, Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, 2001.

[24] M. L. Zhang and Z. H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[25] X. Liang, L. Yuetian, X. Yifei, L. Yanli, C. Xuehui, and L. Gang, "A data-driven shale gas production forecasting method based on the multi-objective random forest regression," *Journal of Petroleum Science and Engineering*, vol. 196, article 107801, 2021.

[26] D. Kocev, C. Vens, J. Struyf, and S. Dzeroski, "Ensembles of multi-objective decision trees," *European Conference on Machine Learning*, pp. 624–631, 2007.

[27] K. Zorlu, C. Gokceoglu, F. Ocakoglu, H. A. Nefeslioglu, and S. Acikalin, "Prediction of uniaxial compressive strength of sandstones using petrography- based models," *Engineering Geology*, vol. 96, no. 3-4, pp. 141–158, 2008.

[28] T. Zhang, Q. Liu, and X. Huang, "Study on TBM penetration rate prediction model and multi-index evaluation method," *Coal engineering*, vol. 53, no. 5, pp. 107–113, 2021.

[29] R. Y. M. Li, B. Tang, and K. W. Chau, "Sustainable construction safety knowledge sharing: a partial least square-structural equation modeling and a feedforward neural network approach," *Sustainability*, vol. 11, no. 20, p. 5831, 2019.

[30] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[31] L. S. Aiken, S. G. West, and S. C. Pitts, "Multiple linear regression," *Handbook of Psychology*, pp. 481–507, 2003.

Hindawi
Geofluids
Volume 2022, Article ID 5637971, 15 pages
https://doi.org/10.1155/2022/5637971

Research Article

Data-Driven Methodology for the Prediction of Fluid Flow in Ultrasonic Production Logging Data Processing

Hongwei Song,[1,2] **Ming Li**,[3] **Chaoquan Wu**,[3] **Qingchuan Wang**,[3] **Shunke Wei**,[3] **Mingxing Wang**,[1] **and Wenhui Ma**[3]

[1]*College of Geophysics and Petroleum Resources, Yangtze University, Wuhan, Hubei 430100, China*
[2]*Research Office of Yangtze University, Key Laboratory of Well Logging, China National Petroleum Corporation, Wuhan, Hubei 430010, China*
[3]*China National Petroleum Corporation, Beijing 100000, China*

Correspondence should be addressed to Hongwei Song; shw98wj@yangtzeu.edu.cn

Received 4 January 2022; Revised 12 February 2022; Accepted 3 March 2022; Published 15 March 2022

Academic Editor: Zhenzhen Wang

A new method for the determination of oil and water flow rates in vertical upward oil-water two-phase pipe flows has been proposed. This method consists of an application of machine learning techniques on the probability density function (PDF) and the power spectral density (PSD) of the power spectrum output of an ultrasonic Doppler sensor in the pipe. The power spectrum characteristic parameters of the two-phase flow are first determined by the probability density function (PDF) method. Then, the transducer signal is preprocessed by distance correlation analysis (DCA), and independent features are extracted by principal component analysis (PCA). The extracted features are used as input to a least-squares fit, which gave the oil flow rates as output. In the same way, the transducer signal is also preprocessed by partial correlation analysis (PCA), and independent features were extracted using independent component analysis (ICA). The extracted features were used as inputs to multilayer back-propagation neural networks, which water cuts as output. The present method was used to calibrate an ultrasonic Doppler sensor to estimate the flow rates of both phases in oil–water flow in a vertical pipe of diameter 159 mm. Predictions of the present method were in good agreement with direct flow rate measurements. Compared to previously used methods of feature extraction from the ultrasonic Doppler power spectrum signals, the present method provides a theoretical basis for the interpretation of ultrasonic multiphase flow logging data. Ultrasonic multiphase flow logging has potential application value in the production profile logging and interpretation evaluation of production wells with low fluid production and high water cut.

1. Introduction

Production logging is a major means of oil well dynamic monitoring. It is an important issue in the logging industry to understand the production status of the oil well production layer, the remaining oil production of the reservoir, the evaluation of the reservoir reconstruction effect, and the adjustment and improvement of the development plan [1]. The production logging technology also plays a very important role in the development of oil fields.

However, oil fields with low porosity and low permeability in the middle-late mining stage are characterized by low flow rate, high water cut, and sand out. The oil-water two-phase flow in the oil well has complex flow regimes with random and variable oil-water interface, as well as a serious slip effect between the oil and water phases. These problems have caused major challenges to traditional production profile logging, such as spinner flowmeter responds poorly under low flow conditions, and the fluid capacitance has poor response under high water cut conditions [2]. Ultrasonic multiphase flow logging tool is a logging method that uses the ultrasonic Doppler effect and the difference in acoustic impedance between oil, gas, and water and uses the spectral characteristics of the discrete phase reflected sound waves to obtain the flow of fluids in each phase [3]. Ultrasonic multiphase flow logging technology solves the

problems of two-phase flow and three-phase flow logging in complex well conditions (low production fluid, sand out, highly deviated well, etc.), and the logging data provides a reliable basis for oilfield development engineers [4]. At this stage, the realization of the instrument method is relatively mature, and the difficulty lies in how to extract information that can accurately reflect the flow of oil, gas, and water from a single ultrasonic frequency spectrum data [5].

Due to the nonlinearity of the data, there are many difficulties in establishing an accurate prediction model. The measurement precision of split-phase flow is lower. Therefore, it is significant to study on the soft measurement method of split-phase flow. In recent years, machine learning models have begun to be applied in various fields. Several investigators suggested the artificial neural network (ANN) methods to solve this problem for multiphase flow [6–8]. ANN techniques have been proposed as a powerful and computational tool to model and solve the complex problems that cannot be described with simple mathematical models [9–11]. Osman presented an ANN model for prediction of pressure drop in horizontal and near-horizontal gas–liquid flow [12]. Zhao et al. established an ANN prediction model based on the conductance signal obtained by measuring the oil-water two-phase flow with electrical methods, where water cut was from 51% to 91%, and good prediction results were obtained [13].

Dimensionality reduction methods are an essential step in any machine learning model pipeline since they will have major importance regarding the accuracy of the classification or regression algorithm applied to the data [14]. Principal component analysis (PCA) is the most commonly used dimensionality reduction method. PCA was originally introduced by Pearson [15] and developed independently by Hotelling [16]. PCA is an unsupervised linear mapping based on an eigenvector search and suitable for Gaussian data. PCA provides different strategies for reducing the dimensionality of feature space and preserves the maximum amount of variance of the original data [17, 18]. PCA can be computed using different algorithms including eigenvalues, latent variable analysis, factor analysis, or linear regression (LR) [19]. Major applications of PCA include image and speech processing, visualization, exploratory data analysis, and robotic sensor data [20].

Aiming at the oil-water two-phase production characteristics of low production, high water cut, and obvious slippage, in order to improve the interpretation accuracy and dig out more ultrasonic multiphase flow logging information, this paper carried out the oil-water two-phase simulation logging experiment of the ultrasonic multiphase flow logging tool to study the ultrasonic frequency spectrum information of the oil-water two-phase flow and the oil and water flow interpretation model. It provides a new interpretation method for the ultrasonic multiphase flow logging of oil-water two-phase production wells with low production liquids.

2. Experimental Facility and Logging Tool

All logging experiments were conducted in the oil-water flow facility at the Yangtze University in China (Figure 1), and the ultrasonic multiphase abortion profile production logging instrument is shown in Figure 2.

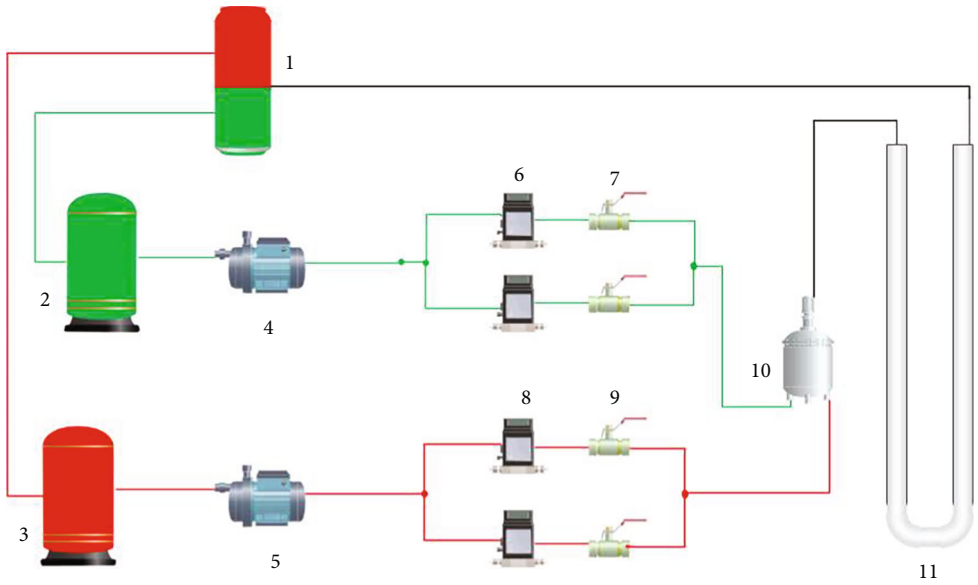
This facility was mainly composed of clear Perspex tubing with an inner diameter of 159 mm, which permitted visual observation of the flow, and the wellbore angle is vertical 90° (relative to the ground). The water of the experimental medium is tap water (density 988 kg/m^3, viscosity 1.16 mPa·s), and the oil is 10# industrial white oil (density 826 kg/m^3, viscosity 8.29 mPa·s). The total flow rate (Q_m) in the experiment ranges from 0.5 m^3/d to 40 m^3/d; the water cut (C_w) varies from 0% to 100%. The oil and water are transported by peristaltic pumps to the pressure-stabilized irrigation and then enter the metering pipeline to ensure that the fluid flow in the oil-water transportation pipeline can ignore the impact of the pump pulsation. The oil and water volumetric flow rates were controlled by butterfly valves, and they were measured by high-precision mass flowmeter (KLB-CMFI-DN6). A total of 42 sets of experimental operating conditions for oil-water two-phase flow with different total flow rates and different water cuts were designed for the experiment. In order to ensure that the fluid flow is sufficiently stable, the measurement of the ultrasonic logging instrument probe was set at 6.0 meters from the inlet of flow fluid; after the oil-water flow rate of each experimental point stabilized for 30 minutes, the ultrasonic instrument started the test and continued at least 3 pulse periods, and the test scenario is shown in Figure 3. All experiments were conducted normal temperature and under the atmospheric pressure.

The structure of the ultrasonic multiphase flowmeter is shown in Figure 2. The ultrasonic probe is located at the lower end of the instrument string during well logging, and the instrument string is connected to the centralizer to make the instrument centered for measurement. The ultrasonic probe adopts spontaneous and self-receiving measurement, and the transmitting and receiving surfaces are conical [4]. Basic principle is the ultrasonic probe that is the transmission of ultrasonic signal by cone, and sound waves meet continuous water phase in the discrete phase (oil bubble, bubble) reflects, when the ultrasonic wave propagation direction and oil bubble movement direction are 90°, ultrasonic reflection in oil bubble surface, and the frequency of the reflected ultrasonic wave relative to change in the frequency of the ultrasonic before reflection, this change is the Doppler frequency shift.

Transducer T emits ultrasonic waves of frequency f_0 to the fluid, and transducer R receives the waves scattered by a particle in the sample volume [21]. Owing to the relative motion between the particle and transducer T, the frequency f_1 of incident waves received by the particle is modulated according to the Doppler effect:

$$f_1 = \frac{c + u\cos\theta}{c} f_0, \tag{1}$$

where c is the speed of sound in fluid, u is the particle velocity in the main flow direction, and θ is the angle (Doppler angle) between the sound beam axis and particle flow direction. For the scattering waves, the moving particle is considered a



1. Oil-water separation tank; 2. Water piggy tank; 3. Oil piggy bank; 4. Water pump; 5. Oil pump; 6 Mass flowmeter; 7. Waterway control valve; 8. Mass flowmeter; 9. Oilway control valve; 10. Oil-water mixer; 11. Simulated wellbore;

Figure 1: Schematic diagram of multiphase flow simulation well experiment device.

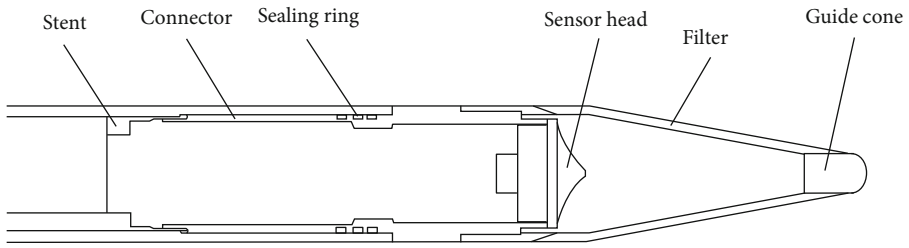


Figure 2: Schematic diagram of ultrasonic multiphase flow logging tool structure.

Figure 3: Simulation logging experiment diagram.

secondary ultrasound source, and the relative motion between the particle and transducer R produces a second Doppler effect. Hence, the frequency f_r of scattering waves received by transducer R is also modulated by the relative motion between it and the particle, which can be expressed as

$$f_r = \frac{c}{c - u\cos\theta} f_1 = \left(1 + \frac{2u\cos\theta}{c - u\cos\theta}\right) f_0. \tag{2}$$

Because the flow velocity u is usually much lower than c, the term $(c - u\cos\theta)$ in Eq. (2) can be approximated as c, which simplifies frequency f_r to

$$f_r \approx \left(1 + \frac{2u\cos\theta}{c}\right) f_0. \tag{3}$$

As a result, the Doppler shift of a single oil bubble can be calculated from [22]

$$f_d = f_r - f_0 = \frac{2u\cos\theta}{c} f_0. \tag{4}$$

The signal reflected by a large number of oil bubbles in the wellbore and the acoustic signal of the wellbore straight lotus are superimposed on the receiving transducer. The signal is transmitted and amplified by telemetry. The final difference frequency signal obtained is the Doppler signal of the ultrasonic multiphase flowmeter [23]. The amount of frequency change is related to the movement speed of the oil bubble, and the amount of reflection of the ultrasonic wave on the surface of the oil bubble corresponds to the number of oil bubbles. Ultrasonic multiphase flow meters use the Doppler frequency domain effect of reflected and scattered ultrasonic signals to obtain fluid flow information [24].

The center frequency of the probe is set to 750 KHz, the measurement method adopts the static point measurement of the instrument, the test time of each experimental measurement point is 3 minutes, and at least 3 cycles of data are collected. The probe is at the foremost end of the instrument string, the fluid is in a noncollecting condition, and the state of the fluid to be measured is basically unchanged. When it flows through the ultrasonic probe, measurement is performed to collect data.

3. Analysis of Experimental Results

3.1. Characteristics and Spectrum Analysis of Instruments Influencing Oil-Water Two-Phase Flow. In order to study the response characteristics of the ultrasonic flowmeter to the oil bubble flow, the logging data collected by the ultrasonic flowmeter was analyzed and processed by the power spectrum, the hydrostatic oil injection (the wellbore is filled with water, increasing the flow of oil), and power spectrum curve (as shown in Figure 4). The power spectrum curve of the same total flow and different water cuts is as the oil-water two-phase flow (as shown in Figure 5). From Figures 4 and 5, the flow characteristics of the fluid at the measuring point can be qualitatively analyzed by the curve change trend.

In the case of hydrostatic oil injection, the amplitude of the power spectrum curve increases with the increase of oil flow. The greater the oil flow, the stronger the reflected signal. When the oil flow rate is very low (less than 1.5 m^3/d), most of the acoustic waves emitted by the ultrasonic sensor are dispersed, the reflected waves are weak, the amplitude is low, and the measurement effect is not obvious. With the increase of oil flow, discrete oil bubbles in the continuous water phase increase, and part of the sound wave is reflected on the surface of the oil bubble. The amplitude of the reflected sound wave is relative to the number of oil bubbles. The more oil bubbles (the larger the oil holdup), the stronger the reflected wave, and the larger the amplitude. Because the water is still, the oil bubbles move upwards at a static drift speed in the water, and the slip phenomenon is obvious [25]; so, the relationship between the change of the center frequency and the change of the oil phase flow rate is not obvious. When the oil flow rate is greater than 20 m^3/d, the oil flow rate increases, but the center frequency decreases. That is, under low flow conditions, the increase in oil flow is mainly due to the increase in the number of oil bubbles (increased oil holdup), and the speed of oil bubbles hardly increases or even decreases. Therefore, in the case of low flow, especially when the water is static or the flow is very low, the center frequency has little correlation with the oil flow.

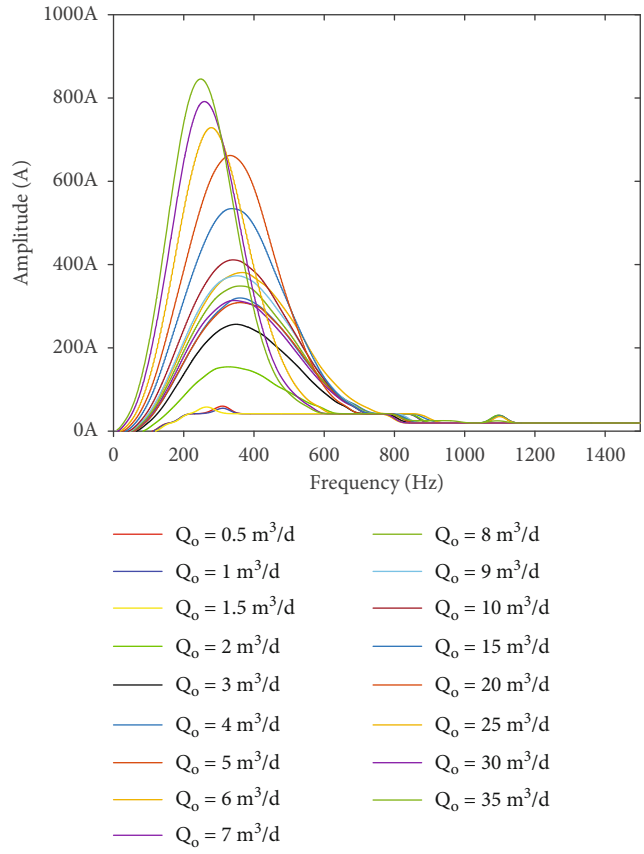


Figure 4: Hydrostatic oil injection power spectrum diagram.

In the case of oil-water two-phase flow, the amplitude of power spectrum curve, left and right attenuation coefficient, and center frequency has obvious changes under the same total flow and different water cuts. The center frequency moved to the right with the increase of water cut, and the peak amplitude decreased with the increase of water cut. This is mainly due to the increase in water cut and the decrease in oil flow. There are fewer discrete oil bubbles in the continuous water phase. Part of the sound waves reflects weakly on the surface of the oil bubbles, and the reflected wave received by the probe decreases, resulting in a drop-in amplitude peaks. The amplitude peak value and center frequency are proportional to the flow rate, but their proportional relationship is uncertain under different total flow well conditions. Therefore, it is necessary to use the statistical results of the power spectrum data measured by the experiment to find the relationship between the amplitude, frequency, and the oil and water flow and obtain the corresponding calculation model or make the relationship chart.

3.2. Measurement of Oil Flow Rates and Description of the Algorithm. For the flow rate prediction, the PDF and PSD are first calculated from the input signal, and appropriate features are extracted by using PCA. The preprocessing steps

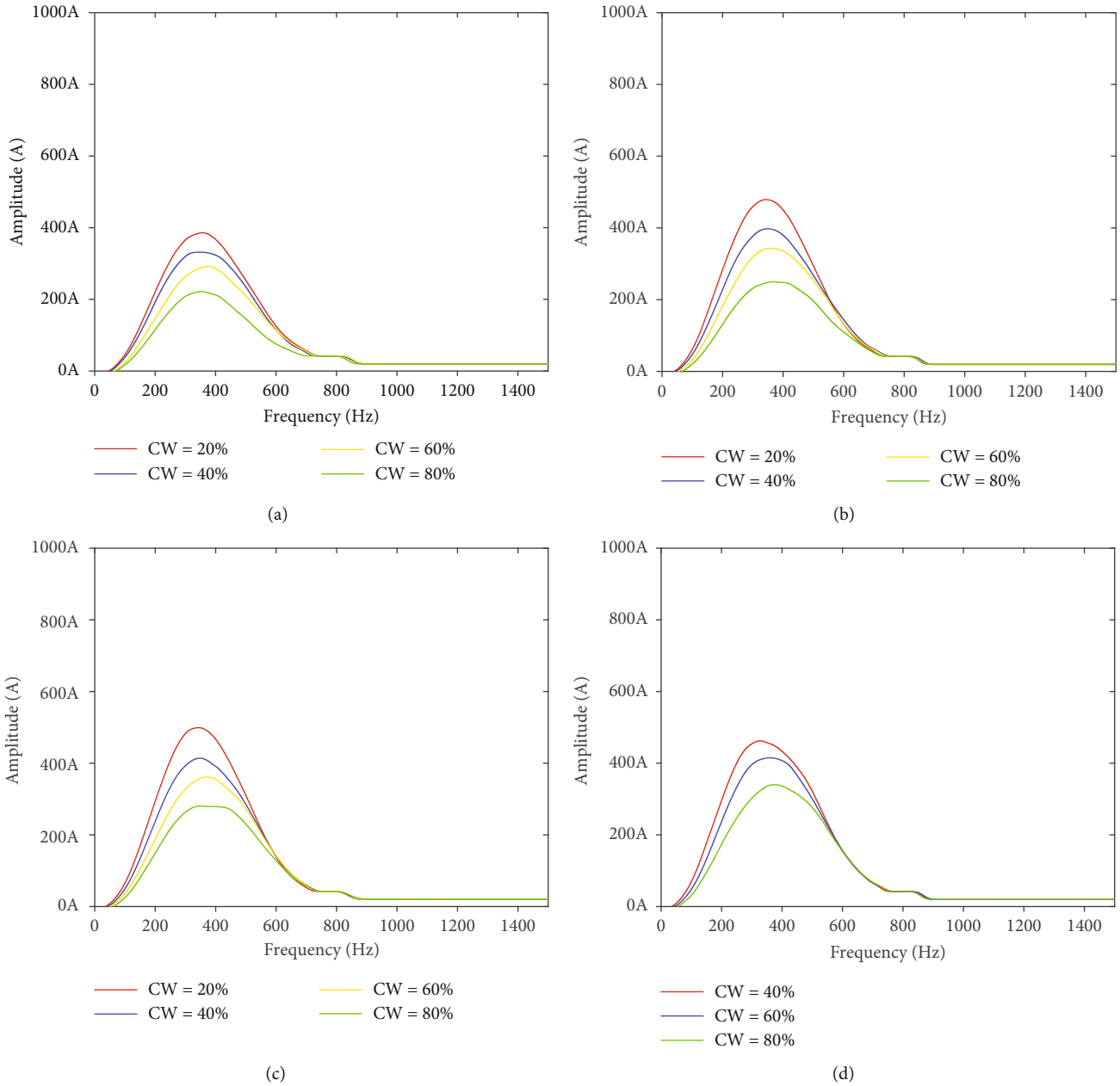


Figure 5: Continued.

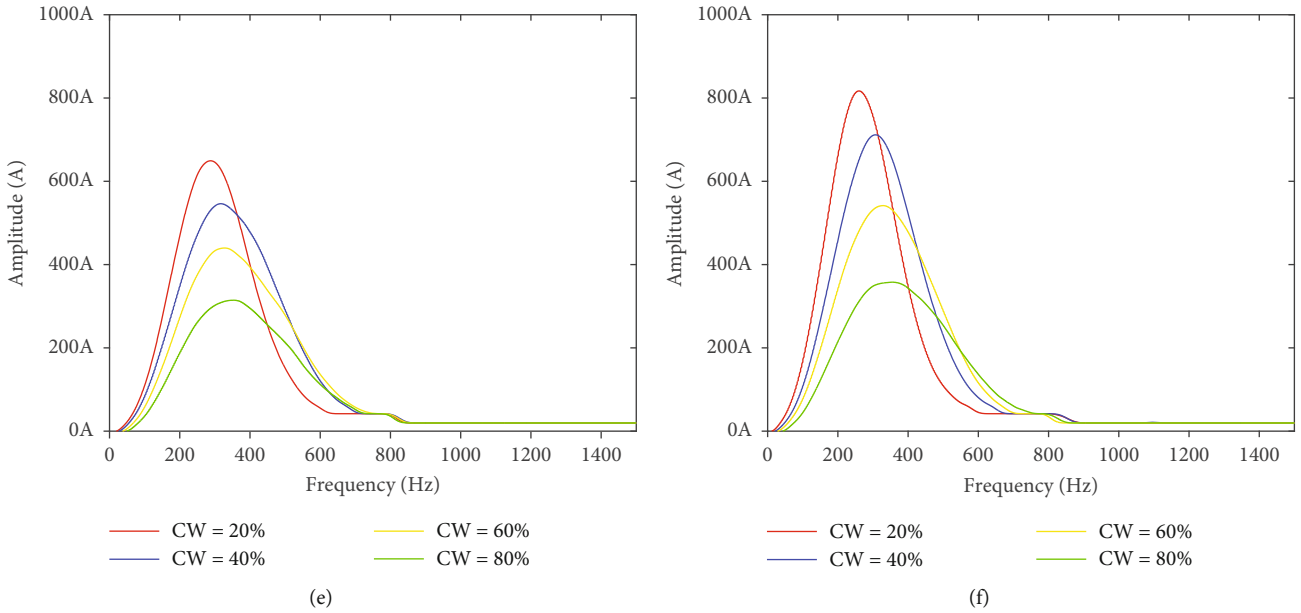


Figure 5: Power spectrum curve of oil-water two-phase flow: (a) Qm = 10 m^3/d, (b) Qm = 15 m^3/d, (c) Qm = 20 m^3/d, (d) Qm = 25 m^3/d, (e) Qm = 30 m^3/d, and (f) Qm = 40 m^3/d.

Table 1: Matrix table of approximate values of variables related to hydrostatic oil injection.

	Correlation between vectors of power spectrum							
	Oil flow	Peak amplitude	Center frequency	Amplitude of oil	Frequency of oil	Variance of oil	Area of oil peak	Logarithm of amplitude ratio
Oil flow	1.000	.960	-.610	.958	-.582	.048	.857	.787
Peak amplitude	.960	1.000	-.385	1.000	-.354	.294	.962	.920
Center frequency	-.610	-.385	1.000	-.377	.984	.693	-.140	-.029
Amplitude of oil	.958	1.000	-.377	1.000	-.348	.299	.966	.922
Frequency of oil	-.582	-.354	.984	-.348	1.000	.763	-.116	.024
Variance of oil	.048	.294	.693	.299	.763	1.000	.497	.641
Area of oil peak	.857	.962	-.140	.966	-.116	.497	1.000	.973
Logarithm of amplitude ratio	.787	.920	-.029	.922	.024	.641	.973	1.000

Table 2: Results of PCA.

Component	Eigenvalue	Variance contribution rate/%	Cumulative variance contribution rate/%
1	6.026	86.087	86.087
2	0.773	11.038	97.125
3	0.168	2.397	99.522
4	0.024	0.340	99.862
5	0.006	0.084	99.946
6	0.004	0.050	99.996
7	0.000	0.004	100.000

employ methods that preserve as much of the information contained in the differential the characteristic parameters as possible, albeit extracting features with a relatively small dimension.

In order to mine more power spectrum signal connotation, the one-dimensional probability density function (PDF) was used to fit the original measurement curve, and the characteristic parameters such as amplitude of oil, frequency of oil, variance of oil, and peak area of reaction oil bubble distribution and flow velocity information were obtained [26]. In order to reduce the complexity of the problem, the correlation analysis method is used to analyze the distribution parameters of the oil-water two-phase flow experiment and the characteristic parameters of the ultrasonic power spectrum to find the relationship between the characteristic parameters of the ultrasonic power spectrum and the oil-water flow parameters.

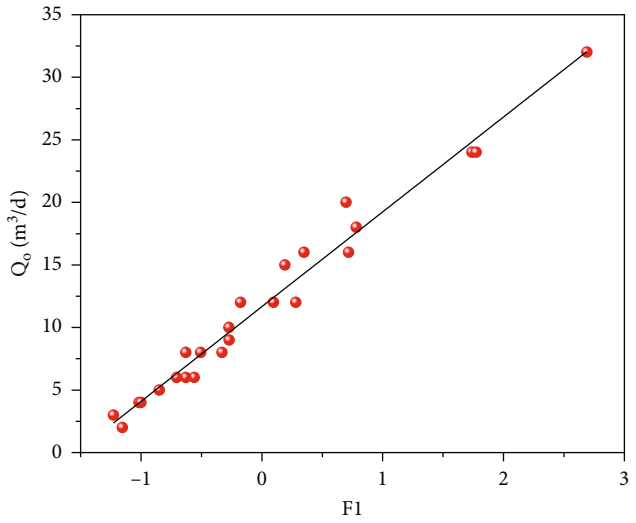


Figure 6: Diagram of relationship between Qo and F1.

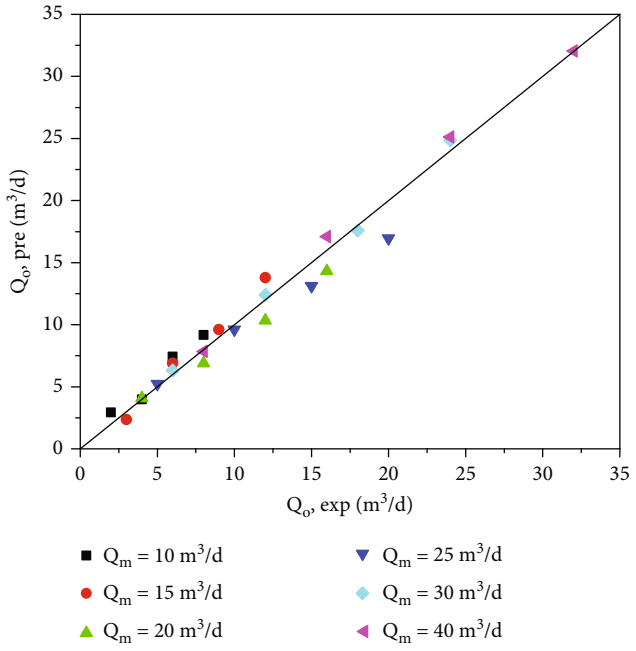


Figure 7: Forecast results of oil flow in oil-water two-phase flow.

First, the correlation analysis of the experimental rationing parameters of hydrostatic oil injection and the characteristic parameters of the ultrasonic power spectrum are carried out. The correlation analysis of the variables related to hydrostatic oil injection is based on the distance process correlation analysis. The approximate value matrix is shown in Table 1. The strength of the distance relationship between variables can be observed through the approximate matrix [27]. The results of the correlation analysis show that the oil flow has a good correlation with many characteristic parameters, if only one or two parameters for the traditional linear or nonlinear fitting to obtain the formula to calculate the oil flow, the results obtained error is larger, but too many variables will inevitably exist data duplication and superposition, resulting in the complexity of the algorithm to enhance. The basic principle of principal component analysis (PCA) is to integrate the original variables into several principal components, replacing a large number of variables with fewer combined variables, to minimize the loss of information carried by the variables, and to make them uncorrelated with each other [28]. The mathematical model of principal component analysis is as follows.

Suppose there are n samples, and each sample has p variables: $x_1, x_2, \ldots, x_p$, the original data observation matrix is as follows:

$$X_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \tag{5}$$

Establish the correlation coefficient matrix R of variables and find the characteristic root of R ($\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p > 0$) and its corresponding unit characteristic vector $(e_1, e_2, \cdots, e_p)$.

Then, determine the number of principal components, define the contribution rate of the principal components which is $\lambda_i / \sum_{k=1}^{p} \lambda_k$, $(i = 1, 2, \cdots, p)$, and the cumulative contribution rate is $\sum_{k=1}^{i} \lambda_k / \sum_{k=1}^{p} \lambda_k$, $(i = 1, 2, \cdots, p)$.

Generally, the eigenvalues $\lambda_1, \lambda_2, \cdots \lambda_m$ corresponding to the first, second, ..., mth $(m \leq p)$ principal components whose cumulative contribution rate reaches 85% or more are taken. The PCA model can be formulated as

$$\begin{cases} f_1 = e_{11}x_1 + e_{12}x_2 + e_{13}x_3 + \cdots + e_{1p}x_p, \\ f_2 = e_{21}x_1 + e_{22}x_2 + e_{23}x_3 + \cdots + e_{2p}x_p, \\ \vdots \\ f_m = e_{m1}x_1 + e_{m2}x_2 + e_{m3}x_3 + \cdots + e_{mp}x_p, \end{cases} \tag{6}$$

where e_{ip} is the p-dimensional eigenvector corresponding to the i-th eigenvalue of the correlation matrix of the original variables; $[x_1\ x_2 \cdots x_p]^T$ is the p-dimensional initial input variable.

Principal component analysis was performed on the characteristic parameters of the oil-water two-phase ultrasonic power spectrum, and the results are shown in Table 2.

It can be seen from Table 2 that the cumulative contribution rate of the first component reaches 86.087%, which exceeds 85%, which can well summarize the original variables. Therefore, the principal component of the characteristic parameters of the oil-water two-phase ultrasonic power spectrum is the first extracted. According to the component score coefficient matrix, the expression can be obtained as

$$F_1 = 0.163X_1 - 0.146X_2 + 0.162X_3 - 0.155X_4 - 0.144X_5 + 0.148X_6 + 0.159X_7. \tag{7}$$

Table 3: Approximate value matrix table of related variables for oil-water two-phase flow.

	Correlation between vectors of power spectrum								
	Total flow	Water cut	Peak amplitude	Center frequency	Amplitude of oil	Frequency of oil	Variance of oil	Area of oil peak	Amplitude ratio
Total flow	1.000	.000	.667	-.555	.668	-.573	-.455	.651	.659
Water cut	.000	1.000	-.651	.601	-.648	.656	.556	-.647	-.690
Peak amplitude	.667	-.651	1.000	-.778	.999	-.852	-.807	.939	.975
Center frequency	-.555	.601	-.778	1.000	-.766	.959	.853	-.630	-.719
Amplitude of oil	.668	-.648	.999	-.766	1.000	-.837	-.790	.947	.979
Frequency of oil	-.573	.656	-.852	.959	-.837	1.000	.933	-.687	-.784
Variance of oil	-.455	.556	-.807	.853	-.790	.933	1.000	-.572	-.684
Area of oil peak	.651	-.647	.939	-.630	.947	-.687	-.572	1.000	.983
Amplitude ratio	.659	-.690	.975	-.719	.979	-.784	-.684	.983	1.000

Table 4: Results of PCA for water cut.

Component	Eigenvalue	Variance contribution rate/%	Cumulative variance contribution rate/%
1	6.09	87.002	87.002
2	0.675	9.639	96.641
3	0.18	2.565	99.207
4	0.035	0.502	99.709
5	0.013	0.19	99.899
6	0.005	0.072	99.971
7	0.002	0.029	100

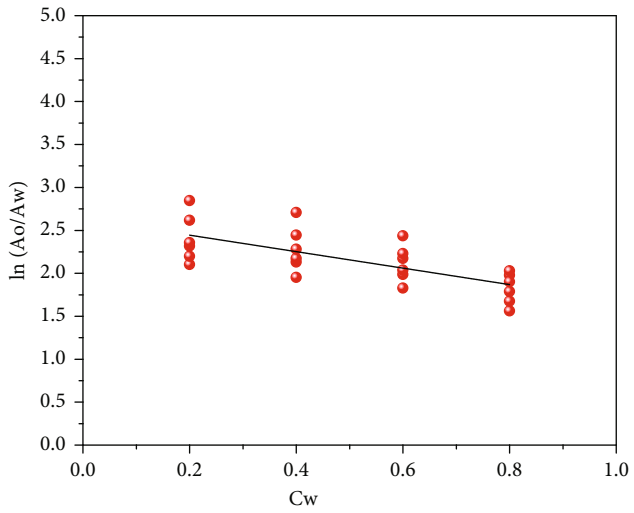


Figure 8: Intersection diagram of water cut and amplitude ratio.

Therefore, the oil flow rate and the principal components extracted from the characteristic parameters of the ultrasonic power spectrum are used for the rendezvous analysis to predict the oil flow rate. The result is shown in Figure 6.

Figure 6 is fitted to get the relationship between the oil flow rate (Q_o) and the principal component value (F_1) measurement model which is

$$Q_o = 7.5778F_1 + 11.667. \tag{8}$$

The oil flow rate increases linearly with the principal component value, and it has very good monotonicity. When the oil flow rate is small, the data correlation is obviously stronger than when the oil flow rate is high, indicating that the ultrasonic flow logging tool is more sensitive to small oil flow changes under the conditions of low oil flow and high water holdup. This is because when the oil flow is low and the wellbore water holdup is high, the size and number of oil bubbles are small, and the oil phase is evenly distributed, which is conducive to ultrasonic Doppler measurement. When the oil flow rate increases and water holdup decreases, the size and number of oil bubbles in the wellbore increase, small oil bubble collision and aggregation become larger, oil phase distribution is uneven, ultrasonic Doppler measurement sensitivity decreases, reflection intensity increases, and amplitude increases. Therefore, ultrasonic Doppler measurement has high sensitivity to high water cut.

Substituting the data obtained from the ultrasonic measurement experiment of the oil-water two-phase flow into Eq. (8), the oil flow rate prediction results as shown in Figure 7 is obtained. In the case of different total flow, the oil flow value in the ultrasonic multiphase flow simulation experiment is in good agreement with the oil flow prediction value.

Table 5: Analysis table of partial correlation of variables related to oil-water two-phase flow.

Correlation Control variable			Water cut	Amplitude of oil	Frequency of oil	Area of oil peak	Amplitude ratio
Total flow	Water cut	Correlation	1.000	-.872	.800	-.852	-.918
		Significance (two-tailed)	.	.000	.000	.000	.000
		Degree of freedom	0	21	21	21	21
	Amplitude of oil	Correlation	-.872	1.000	-.745	.907	.963
		Significance (two-tailed)	.000	.	.000	.000	.000
		Degree of freedom	21	0	21	21	21
	Frequency of oil	Correlation	.800	-.745	1.000	-.505	-.660
		Significance (two-tailed)	.000	.000	.	.014	.001
		Degree of freedom	21	21	0	21	21
	Area of oil peak	Correlation	-.852	.907	-.505	1.000	.971
		Significance (two-tailed)	.000	.000	.014	.	.000
		Degree of freedom	21	21	21	0	21
	Amplitude ratio	Correlation	-.918	.963	-.660	.971	1.000
		Significance (two-tailed)	.000	.000	.001	.000	.
		Degree of freedom	21	21	21	21	0

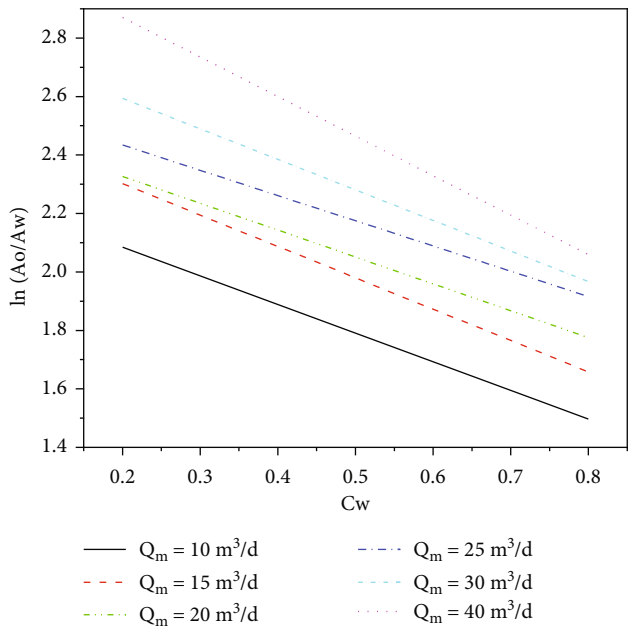


Figure 9: The relationship between water cut and amplitude ratio at the same total flow rate.

On the whole, the oil flow rate prediction model established by the hydrostatic oil injection experiment data is effective in calculating the oil flow rate under the oil-water two-phase flow. The overall average absolute error is 0.92 m^3/d, and the average relative error is 10.36%.

Although the oil flow rate in oil-water two-phase flow can be accurately calculated by extracting principal component value of the characteristic parameters of the ultrasonic power spectrum, the main reason is that the total flow rate is small, the oil phase velocity is close to the static drift velocity of oil in water, and the change of oil flow rate mainly depends on the number of oil bubbles (oil holdup), which is strongly related to the amplitude of oil and area of oil peak. The total flow does not change, the water cut decreases, and as the oil flow increases, the error value increases. That is to say, the prediction model (8) has higher prediction accuracy in the case of high water cut and less oil bubbles. With the increase of oil bubbles, the flow pattern changes from discrete bubble flow to emulsion flow, and the measurement effect becomes worse.

3.3. Water Cut Measurement and Description of the Algorithm. In the previous section, it was shown that the PDF and PSD of the differential ultrasonic power spectrum characteristic parameters followed certain systematic trends as the oil, and water flow rates were changed [29].

The water cut reflects the relationship between the water flow and the total flow in the oil-water two-phase flow and has a complicated relationship with factors such as water holdup and oil-water slip velocity. In the oil-water two-phase flow of low-yield liquid, the water holdup is generally high due to the serious water logging in the wellbore, which brings more difficulties to the calculation of water cut. The conventional production profile interpretation model (drift flux model, slip model) is greatly affected by the fluctuation of water holdup. Ultrasonic multiphase flow logging data processing uses the power spectrum related characteristic parameters to compare and analyze the water cut ratio of the experiment to determine the water cut calculation model. In the same way, for the water prediction, the PDF and PSD are also calculated from the input signal, appropriate features are extracted by using DCA followed by PCA, and regression is performed using artificial neural networks.

It can be seen from the power spectrum analysis of ultrasonic measurement in Figures 4 and 5 that the total flow rate has no obvious dependence on the related characteristic parameters, when the total flow rate remains unchanged, the water cut increases, the amplitude of the corresponding

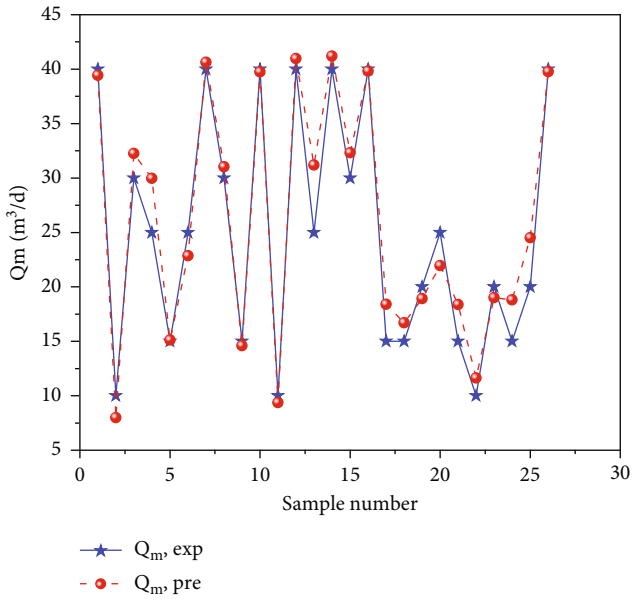


Figure 10: Comparison chart of predicted flow and real flow.

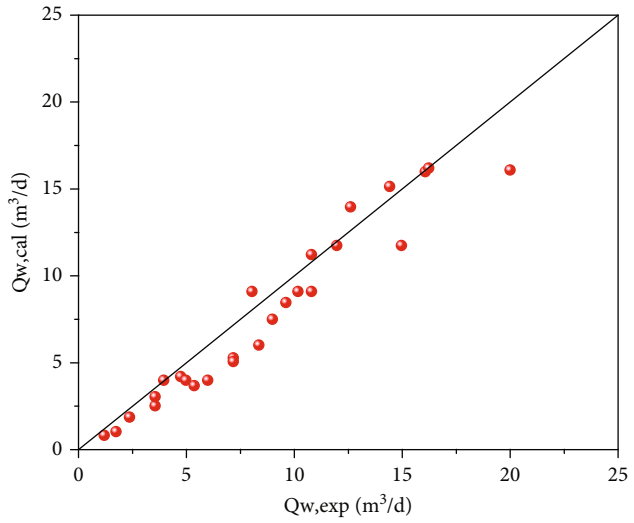


Figure 11: Comparison diagram of experimental and calculated water phase flow.

power spectrum decreases, and the center frequency shifts to the right. Therefore, the distance correlation analysis (DCA) method is also used to perform correlation analysis on the experimental data of the oil-water two-phase flow. The analysis results are shown in Table 3. According to the analysis results, Pearson correlation coefficients between total flow and amplitude peak, center frequency, amplitude of oil, frequency of oil, variance of oil, area of oil peak, and amplitude ratio are not high, all less than 0.7, indicating that the distance correlation intensity between them is very weak, which is consistent with the qualitative analysis results of power spectrum in Figures 4 and 5.

The Pearson correlation coefficient between water cut and ultrasonic characteristic parameters is not high, which

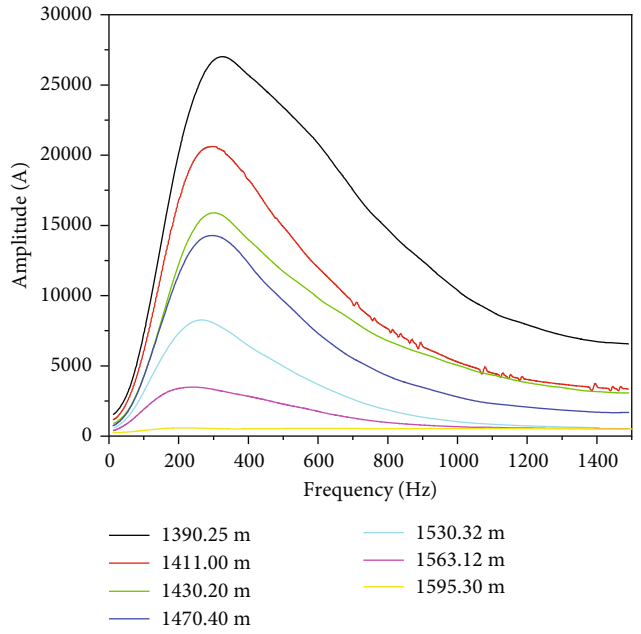


Figure 12: Well 13-1-2 ultrasonic multiphase flow test power spectrum curve.

is positively correlated with the center frequency, frequency of oil and variance of oil, but negatively correlated with the amplitude peak, amplitude of oil, area of oil peak, and amplitude ratio. Pearson correlation coefficient with the total flow is 0, and there is no correlation between water cut and the total flow. In the case of oil-water two-phase flow, the peak amplitude is positively correlated with the oil amplitude, and the center frequency is positively correlated with the frequency of oil. The correlation coefficient is relatively close, and the correlation coefficient with the oil-water amplitude ratio ($\ln\ (A_o/A_w)$) is 0.69.

To investigate the effect of representing the total information in the original dataset with a smaller number of features, principal component analysis was performed on the characteristic parameters of the oil-water two-phase ultrasonic power spectrum, and the results are shown in Table 4.

According to the component score coefficient matrix, the expression can be obtained as

$$F_2 = 0.158X_1 - 0.149X_2 + 0.163X_3 - 0.156X_4 - 0.144X_5 + 0.143X_6 + 0.158X_7. \quad (9)$$

Figure 8 is the intersection of the water cut and the oil-water amplitude ratio, and the fitting formula is Eq. (10):

$$\mathrm{Ln}\left(\frac{A_o}{A_w}\right) = 0.098974C_w + 2.60833. \quad (10)$$

The linear fitting correlation coefficient ($R^2 = 0.45784$) is low, and the water cut calculated by Eq. (10) has a large error, which can be used as the pseudowater cut calculation formula.

Table 6: Well 13-1-2 power spectrum characteristic parameter table.

Layer number	Measuring point depth (m)	Temperature (°C)	Pressure (MPa)	Peak amplitude (μv/Hz^2)	Center frequency (Hz)	Amplitude of oil (μv/Hz^2)	Frequency of oil (Hz)	Fluid phase
1	1390.25	68.00	2.90	1067.00	326.34	812.51	361.70	Oil-water two-phase
2	1411.00	68.80	3.10	812.53	297.02	636.01	322.33	Oil-water two-phase
3	1430.20	69.80	3.30	623.64	301.73	469.24	326.78	Oil-water two-phase
4	1470.40	71.40	3.80	559.80	295.98	453.81	314.15	Oil-water two-phase
5	1530.32	73.50	4.40	314.81	268.70	236.23	261.76	Oil-water two-phase
6	1563.12	74.50	4.80	129.34	241.65	106.64	283.13	Oil-water two-phase
7	1595.30	76.64	5.14	13.46	221.14	0.00	0.00	Static water

Further analysis of moisture content and the relationship between ultrasonic power spectrum parameters, extraction, and analysis of the results of correlation coefficient absolute value are greater than 0.6 variable frequency amplitude (amplitude of oil, frequency of oil, area of oil peak, amplitude ratio), the partial correlation analysis (PCA) method, analysis of traffic is fixed, and the water cut and the correlation between ultrasonic power spectral characteristic parameters process the results as shown in Table 5. It can be seen from the results that when controlling the total flow, the partial correlation coefficient between the water cut and the amplitude ratio is -0.918, and the probability (significance) that they are not correlated is $p = 0$. It can be concluded that under the condition of a certain total flow, there is a significant negative correlation between water cut (C_w) and oil-water amplitude ratio ($\ln(A_o/A_w)$). Figure 9 is a graph showing the relationship between water cut and amplitude ratio when the total flow rate is constant under the oil-water two-phase flow.

It can be seen from Figure 9 that under a certain total flow rate, the water cut and the oil-water amplitude ratio have a good linear correlation. The larger the amplitude ratio, the smaller the water cut. Using this chart, you can calculate the water cut based on the amplitude ratio interpolation when the total flow rate is determined, and you can also calculate the total flow rate based on the amplitude ratio interpolation when the water cut rate is determined. For the flow rate prediction, water cut (C_w) and oil-water amplitude ratio ($\ln(A_o/A_w)$) were used as inputs to multilayer back-propagation neural networks, which gave the total flow rates as output. Figure 10 presents the ratio of predicted and measured flow rates for the ANN; the vast majority of predictions were within ±10% of the measured values.

3.4. Water Phase Flow Calculation. From the above power spectrum Figure 5 and the distances process correlation analysis result Table 3, it can be seen that it is difficult to directly find the relationship between the water phase flow rate and the relevant characteristic parameters of the power spectrum. However, under a certain total flow rate, the relationship between the ratio of the water phase flow rate and the amplitude obtained from the water cut is very obvious, and the linear relationship is better. Therefore, based on the above Eq. (5) of the oil phase flow calculation model and the water cut calculation chart under a certain total flow, the following method is proposed to calculate the water phase flow.

(1) First, calculate the oil phase flow rate Q_o from the principal component value (F_1) using Eq. (8)

(2) From the oil-water amplitude ratio (ln (Ao/Aw)), use Eq. (9) to calculate the pseudo water cut C_w'

(3) Then, use the oil-water amplitude ratio (ln (Ao/Aw)) and the pseudowater cut C_w' obtained in step (2) into the same water cut and amplitude ratio relationship chart (Figure 9) for the total flow rate for ANN to obtain the total flow rate Q_m. Recalculate the water cut according to the definition formula of water cut

$$C_w = \frac{Q_m - Q_o}{Q_m}. \tag{11}$$

(4) Then calculate the error between the water cut calculated by Eq. (7) and the pseudowater cut calculated by Eq. (6). If it is within the error range, the total flow Q_m and water cut C_w can be obtained, and the water phase flow can be calculated by the relationship between them

$$Q_w = Q_m \cdot C_w. \tag{12}$$

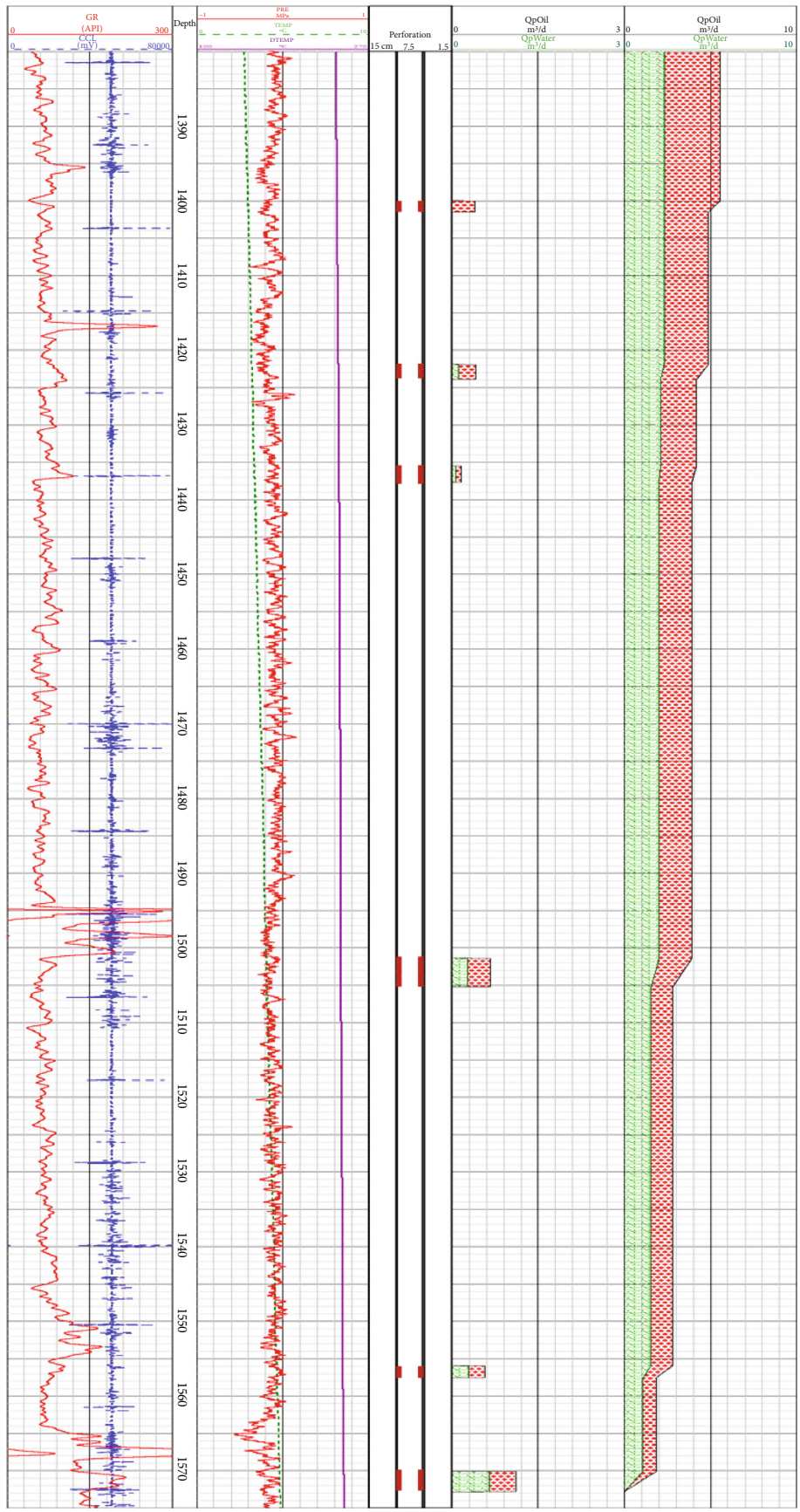


Figure 13: Interpretation results of well 13-1-2.

Table 7: Well 13-1-2 interpretation result table.

Serial number	Interpretation level	Perforated well section (m)	Oil yield (m^3/d)	Water yield (m^3/d)	Total fluid production (m^3/d)	Water cut (%)
1	IV-25	1400.00-1401.50	0.68	0	0.68	0
2	IV-28	1421.90-1423.90	0.5	0.19	0.69	27.09
3	IV-30	1435.47-1437.70	0.17	0.1	0.27	36.98
4	IV-40	1501.38-1505.22	0.65	0.46	1.11	41.64
5	IV-47	1555.95-1557.55	0.49	0.48	0.97	49.42
6	IV-49	1570.10-1572.80	0.77	1.07	1.84	58.36
Total			3.26	2.3	5.56	41.37

(5) If the error is greater than the error limit, adjust the pseudowater cut C_w' according to the relationship between the water cut C_w calculated on the plate and the pseudowater cut C_w' and continue to insert it into the Figure 9 for interpolation calculations until the error is less than the error limit

In order to check the calculation accuracy of the method, the logging data of the ultrasonic Doppler flowmeter in the full flow layer of 26 wells were measured by power spectrum processing to extract the relevant characteristic parameters. Because the calibration conditions of the logging instrument laboratory are different from the working conditions of downhole logging, and the physical properties of oil and water are different, it is necessary to calibrate the amplitude and center frequency of the total flow power spectrum curve of the downhole test to the calibration conditions of the same flow rate for explanation [4]. The calibration coefficient is

$$\begin{cases} K_A = A_{\log}/A_{\exp}, \\ K_F = F_{\log}/F_{\exp}. \end{cases} \tag{13}$$

In the formula, K_A is the amplitude correction coefficient, $A_{\log}$ is the actual logging amplitude value, $A_{\exp}$ is the amplitude value measured in the laboratory with the same flow rate, K_F is the frequency correction coefficient, $K_{\log}$ is the actual logging frequency value, and $K_{\exp}$ is the frequency value measured by the laboratory with the same flow rate.

The calibrated power spectrum characteristic parameter values were substituted into the above formula and chart, and the interpretation results were compared with the water flow calculated by the total production and water cut of logging time at that time. The comparison between the actual water flow in the full flow layer of these 26 wells and the water flow calculated by this method is shown in Figure 11. The average absolute error of the water phase flow calculation is 1.502 m^3/d, and the average relative error is 22.79%. The relative error of the water phase flow rate under high water cut conditions is obviously smaller than the relative error value under low water cut conditions, which also shows that the ultrasonic Doppler oil-water two-phase flow measurement has higher sensitivity under high water cut conditions.

4. Field Application

Western Oilfield is a low-porosity and low-permeability reservoir. It is currently in the middle and late stages of development, and most of the oil wells are in a state of low production fluid and high water cut. Well 13-1-2 is a pumped well. Ultrasonic multiphase flow production profile logging was performed on March 23, 2018. Before logging, the daily production fluid at the wellhead was 5.55 m^3/d, and the water cut was 41.4%.

The logging interpretation of ultrasonic polyphase abortion profile is mainly based on the temperature, pressure, and power spectrum curve of downhole fluid to judge the phase state of downhole fluid, and the oil and water production of each producing zone is calculated according to the production status of wellhead and the amplitude and central frequency of oil signal of power spectrum curve. The power spectrum curve of ultrasonic multiphase flow test in well 13-1-2 is shown in Figure 12.

According to the position of the downhole perforation layer, there are a total of 7 test points, and the test point fluid is a two-phase flow of oil and water. The power spectrum curve of each test point is processed to obtain the power spectrum characteristic parameters, as shown in Table 6.

The comprehensive interpretation results are shown in Figure 13, and the interpretation results are shown in Table 7. After comprehensive analysis, the interpretation results are obtained: the main liquid producing layer is IV-49 sublayer, the liquid production volume is 1.84 m^3/d, and the water cut is 58.36%. The secondary liquid producing layers are IV-25, IV-28, IV-30, IV-40, and IV-47, with liquid production of 0.68 m^3/d, 0.69 m^3/d, 0.27 m^3/d, 1.11 m^3/d, and 0.97 m^3/d, respectively. Water cut were 0.0%, 27.09%, 36.98%, 41.64%, and 49.42%, respectively.

Compared with the actual oil and water flow in the field, the accuracy of data-driven artificial intelligence interpretation is higher than that of traditional ultrasonic power spectrum single-factor calculation of flow, and it also solves the problem of water flow calculation.

5. Conclusion

In this paper, the ultrasonic multiphase flow logging tool is used to test the acoustic frequency characteristics of the oil-water two-phase flow in a vertical simulation experimental wellbore, and the data-driven methodology is used for the

prediction of fluid flow in ultrasonic production logging data processing. The following conclusions are obtained through the analysis of the experimental data.

(1) The ultrasonic multiphase flow logging tool mainly detects the movement parameters of the oil bubbles in the continuous water phase. The power spectrum curve obtained from the experimental data can qualitatively analyze the change characteristics of the total flow rate and water cut of the oil-water two-phase flow

(2) Under low flow conditions, the oil bubbles move upward at a static drift speed; so, the oil phase flow is mainly related to the number of oil bubbles and the ultrasonic reflection intensity that reflects the number of oil bubbles

(3) The characteristic parameters extracted from the ultrasonic power spectrum curve are related to the oil flow and water cut, and there is a functional relationship. This paper uses the distance process correlation analysis method and partial correlation analysis method to analyze the correlation of the power spectrum characteristic parameters and establishes the oil phase flow calculation model, the water cut prediction chart, and the water flow calculation method. The research results show that the method has high precision and can be a very good calculating the output profile parameters of oil-water two-phase flow

(4) The data-driven artificial intelligence interpretation accuracy is higher than the traditional single-factor calculation accuracy of ultrasonic power spectrum

Data Availability

Data is available on request. Please contact the corresponding author for the underlying data supporting the results of the research.

Conflicts of Interest

All authors confirm that there is no financial/personal interest or belief that could affect our objectivity, and no conflicts exist.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (42174155).

References

[1] H. W. Song, H. M. Guo, S. Guo, H. Y. Shi et al., "Measurement method of split-phase flow of oil-water two-phase stratified flow in horizontal wells," *Petroleum Exploration & Development*, vol. 47, no. 3, pp. 573–582, 2020.

[2] A. Zhao, Y. F. Han, Y. Y. Ren, L. S. Zhai, and N. D. Jin, "Ultrasonic method for measuring water holdup of low velocity and high-water-cut oil-water two-phase flow," *Applied Geophysics*, vol. 13, no. 1, pp. 179–193, 2016.

[3] H. Zhuang, *Research on downhole ultrasonic testing method of multiphase flow split flow*, Daqing Petroleum Institute, 2007.

[4] H. Zhang, J. Q. Qiu, Q. C. Wang, C. J. Gan, Y. X. Liu, F. Zhou et al., "Ultrasonic-Doppler three-phase flow logging and its application in Qinghai oilfield," *Logging Technology*, vol. 4, pp. 99–105, 2016.

[5] Q. Wang, X. K. Zheng, F. Y. Meng et al., "Ultrasonic Doppler multiphase flow logging interpretation method based on data mining," *Daqing Petroleum Geology and Development*, vol. 37, no. 6, pp. 149–153, 2018.

[6] E. Abro, V. A. Khoryakov, G. A. Johansen, and L. Kochach, "Improved void fraction determination by means of multibeam gamma-ray attenuation measurements," *Measurement Science and Technology*, vol. 10, no. 2, pp. 99–108, 1999.

[7] M. R. Malayeri, H. Müller-Steinhagen, and J. M. Smith, "Neural network analysis of void fraction in air/water two-phase flows at elevated temperatures," *Chemical Engineering and Processing: Process Intensification*, vol. 42, no. 8-9, pp. 587–597, 2003.

[8] A. Á. D. Castillo, E. Santoyo, and O. Garcia-Valladare, "*A* new void fraction correlation inferred from artificial neural networks for modeling two-phase flows in geothermal wells," *Computational Geosciences*, vol. 41, pp. 25–39, 2012.

[9] S. S. Sablani, W. H. Shayyab, and A. Kacimovc, "Explicit calculation of the friction factor in pipeline flow of Bingham plastic fluids: a neural network approach," *Chemical Engineering Science*, vol. 58, no. 1, pp. 99–106, 2003.

[10] B. Goutorbe, F. Lucazeau, and A. Bonneville, "Using neural networks to predict thermal conductivity from geophysical well logs," *Geophysical Journal International*, vol. 166, no. 1, pp. 115–125, 2006.

[11] A. Bassam, D. Ortega-Toledo, J. A. Hernández, J. G. González-Rodriguez, and J. Uruchurtu, "Artificial neural network for the evaluation of CO 2 corrosion in a pipeline steel," *Journal of Solid State Electrochemistry*, vol. 13, no. 5, pp. 773–780, 2009.

[12] E. A. Osman, "Artificial neural networks models for identifying flow regimes and predicting liquid holdup in horizontal multiphase flow," in *Proceedings of the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2001.

[13] X. Zhao, N. D. Jin, and W. Li, "Soft measurement method of phase volume fraction for oil/water two-phase flow," *Journal of Chemical Industry and Engineering (China)*, vol. 56, pp. 1875–1879, 2005.

[14] R. Silva and P. Melo-Pinto, "A review of different dimensionality reduction methods for the prediction of sugar content from hyperspectral images of wine grape berries," *Applied Soft Computing*, vol. 113, article 107889, 2021.

[15] K. Pearson, "LIII.On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.

[16] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Education & Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[17] S. Deegalla, H. Boström, and K. Walgama, "Choice of dimensionality reduction methods for feature and classifier fusion with nearest neighbor classifiers," in *15th International Conference on Information Fusion (FUSION)*, pp. 875–881, Singapore, 2012.

[18] S. Ahmadkhani and P. Adibi, "Face recognition using supervised probabilistic principal component analysis mixture model in dimensionality reduction without loss framework," *IET Computer Vision*, vol. 10, no. 3, pp. 193–201, 2016.

[19] C. Meng, O. A. Zeleznik, G. G. Thallinger, B. Kuster, A. M. Gholami, and A. C. Culhane, "Dimension reduction techniques for the integrative analysis of multi-omics data," *Briefings in Bioinformatics*, vol. 17, no. 4, pp. 628–641, 2016.

[20] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, article 20150202, 2016.

[21] D. W. Baker and W. G. Yates, "Technique for studying the sample volume of ultrasonic Doppler devices," *Medical and Biological Engineering*, vol. 11, no. 6, pp. 766–770, 1973.

[22] C. Tan, M. Yuichi, W. L. Liu, T. Yuji, F. Dong, and T. Yasushi, "Ultrasonic Doppler technique for application to multiphase flows: a review," *International Journal of Multiphase Flow*, vol. 144, no. 2021, article 103811, 2021.

[23] Z. H. Li, L. B. Liu, T. S. Jiang, H. M. Duan, X. Z. Zhai et al., "The spectral characteristics of the ultrasonic Doppler signal in flow measurement and its relationship with the flow velocity in the pipeline," *Acta Metrology*, vol. 16, no. 1, pp. 68–72, 1995.

[24] W. J. Yang, W. Z. Chen, Q. F. Zhao, Z. G. Jiang et al., "Multiphase flow ultrasonic logging simulation experiment research," *Journal of Jianghan Petroleum Institute*, vol. 4, pp. 76-77, 2003.

[25] X. P. Liu, X. Wang, Y. Liu, J. C. Chen, X. Q. Li et al., "Research on slip ratio correction method of water cut and its application in Zhongyuan oilfield," *Fault Block Oil and Gas Field*, vol. 5, pp. 66-67, 2003.

[26] N. D. Jin, W. K. Ren, X. Chen, L. S. Zhai et al., "Gas holdup measurement of oil-gas-water three-phase flow ultrasonic sensor," *Journal of Applied Acoustics*, vol. 39, no. 1, pp. 36–44, 2020.

[27] X. Longhan and S. Tao, *SPSS statistical analysis and data mining*, Publishing House of Electronics Industry, 2012.

[28] Z. Zhihua, *Machine Learning*, Tsinghua University Press, Beijing, 2016.

[29] H. Shaban and S. Tavoularis, "Measurement of gas and liquid flow rates in two-phase pipe flows by the application of machine learning techniques to differential pressure signals," *International Journal of Multiphase Flow*, vol. 67, pp. 106–117, 2014.

Hindawi
Geofluids
Volume 2022, Article ID 2056323, 12 pages
https://doi.org/10.1155/2022/2056323

Research Article

Parameter Optimization Study of Gas Hydrate Reservoir Development Based on a Surrogate Model Assisted Particle Swarm Algorithm

Le Zhang,[1,2] **Xin Huang**,[1,2] **Jiayuan He**,[1,2] **Xueqi Cen**,[1,2] **and Yongge Liu**[3]

[1]*State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Effective Development, Sinopec Petroleum Exploration & Production Research Institute, Beijing 102206, China*
[2]*Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), 511458, China*
[3]*China University of Petroleum (East China), 266580, China*

Correspondence should be addressed to Le Zhang; zhangle2017.syky@sinopec.com

Received 7 November 2021; Accepted 15 December 2021; Published 18 January 2022

Academic Editor: Zhenzhen Wang

Using surrogate model to assist parameter optimization of oil and gas reservoir development can greatly reduce the call times of numerical simulator and accelerate the optimization process. However, for serial simulators or parallel simulators with low speedup ratio, the conventional method is still time-consuming. Firstly, an improved surrogate model assisted particle swarm optimization (PSO) algorithm was proposed in this paper. Then, the performance of the algorithm was analyzed using the Rastrigin function. Finally, the key operation parameters of a gas hydrate reservoir by depressurization−to−hot−water−flooding method were optimized with the new method. The results show that the new method only affects the update of the global optimal particle without interfering with the calculation process of the local optimal particles at the early stage of optimization. It realizes the rapid addition of the particle samples through the good parallel features of the PSO algorithm, and therefore, improve the precision of surrogate model in a short time. At the late stage of optimization, it is transformed into a local surrogate model to achieve rapid convergence, when the training time of the surrogate model exceeds the calculation time of the simulator. Both the optimization of Rastrigin function and operation parameters of gas hydrate development reveal that the new algorithm greatly reduces the number of iterations under the same accuracy and thus successfully accelerates the optimization process.

1. Introduction

Multiparameter optimization is a common problem in oil and gas industry. At present, the optimization methods that can be combined with simulators mainly include gradient-based algorithms, approximate gradient-based algorithms, and intelligent algorithms [1–3]. Gradient-based algorithms need to accurately obtain the gradient of the objective function, and therefore, the simulator must be open source so that the code can be modified to obtain the gradient [4]. Approximate gradient algorithms mainly include Levenberg-Marquardt (LM) algorithm, simultaneous perturbation stochastic approximation (SPSA) algorithm, etc. These kinds of algorithms usually have fast convergence speed, but it is easy to obtain local optimum for nonconvex problems [5, 6]. Intelligent algorithms include genetic algorithm, simulated annealing algorithm, particle swarm optimization (PSO) algorithm, etc. Their optimization process does not depend on the gradient of the objective function, and the differentiability of the objective function is not necessary [7, 8]. In addition, the global search ability of these algorithms is very strong, so compared with the approximate gradient algorithms, the probability of obtaining the global optimum of nonconvex problems is greatly enhanced. Therefore, intelligent optimization algorithms have been widely used in multiparameter optimization problems in the oil and gas industry in recent years [9, 10].

However, many researchers found that the conventional intelligent optimization algorithms have a slow convergence speed. A large number of iterations are required, and the numerical simulator needs to be called hundreds or even thousands of times. Therefore, the calculation time for complex models can be up to several weeks or even longer [11, 12]. In recent years, with the development of machine learning technology, the surrogate model assisted optimization algorithm points out a new direction for rapid optimization [13]. Based on the known information about an objective function, a surrogate model can be trained to obtain the potential location of optimum, which then are verified by the numerical simulator. Compared with conventional intelligent optimization algorithms, the surrogate model assisted optimization algorithm can find the potential positions and accelerate the convergence quickly [14, 15]. Using radial basis function network to train the surrogate model, Yu et al. proposed a surrogate-assisted PSO algorithm and proved the effectiveness of the new method [16]. Cai et al. also proposed a surrogate-assisted PSO algorithm which focus on the balance between the prediction ability of surrogates and global search ability of PSO, and the results show that the new method can handle high-dimensional expensive problems well [17]. Zhang et al. trained the surrogate model with random forest algorithm and predicted the creep index. The results show that the prediction accuracy of the method is significantly higher than that of empirical model [18]. Chen et al. proposed a surrogate model assisted differential evolution method and optimized the operation parameters in waterflooding production. A higher net present value and better convergence speed are achieved by the new algorithm [19].

From the above analysis, it can be seen that many studies have proposed a variety of computing processes for different surrogate models and intelligent optimization algorithms. Meanwhile, the surrogate model shows a good ability to accelerate the convergence speed. However, most of these models focus on reducing the call times of numerical simulators. For serial simulators or parallel simulators with low speedup ratio, the computing resources are often idle in the calculation process, resulting the low computing efficiency. Making full use of the good parallel characteristics of intelligent optimization algorithms and how to design workflow to make rational use of computing resources have not been fully considered. Therefore, based on the parallel characteristics of PSO algorithm, this paper proposed an improved surrogate-assisted particle swarm optimization algorithm (i-SAPSO) and verified it with Rastrigin function. Then, the operation parameters of gas hydrate reservoir developed by depressurization–to–hot–water–flooding method were optimized by the new method. Finally, the performance of different algorithms and rationality of optimization results were analyzed.

2. Surrogate Model Assisted Particle Swarm Optimization

2.1. PSO Algorithm. PSO algorithm was proposed by James Kennedy and Russell Eberhart in 1995. Inspired from the activity behavior of animal clusters, the algorithm combines the individual information of particles together to make the movement of the whole group and produces an evolution process from disorder to order in the problem-solving space [20–23]. The algorithm randomly selects several particles in the N-dimensional search space, in which the position of particle i in the t-th iteration can be expressed as $\vec{x}_i^t = (x_{i1}^t, x_{i2}^t, \cdots, x_{iN}^t)$. The historical optimal position of the particle i in the t-th iteration can be recorded as $\vec{p}_i^t = (p_{i1}^t, p_{i2}^t, \cdots, p_{iN}^t)$, and the optimal position of all particles, that is, the global optimal position, can be recorded as $\vec{p}_g^t = (p_{g1}^t, p_{g2}^t, \cdots, p_{gN}^t)$. According to the PSO algorithm, the particle has the trend of moving to its historical optimal position and to the global optimal position. Thus, the update formula of the particle position can be expressed as follows:

$$\begin{cases} v_{id}^{t+1} = \omega v_{id}^t + c_1 \,\mathrm{rand}_1()\left(p_{id}^t - x_{id}^t\right) + c_2 \,\mathrm{rand}_2()\left(p_{gd}^t - x_{id}^t\right), \\ x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}, \end{cases} \tag{1}$$

where ω is the inertia weight, c_1 and c_2 are the constants, and $\mathrm{rand}_1()$ and $\mathrm{rand}_2()$ are the random functions. Inertia weight ω in this paper is 1.0, and c_1 and c_2 are both defined as 2.0 [20].

2.2. Surrogate Model. From the iterative process of PSO, it can be seen that the PSO algorithm only updates the particle positions by simply recording the global optimal and historical optimal, but it does not mine the information of all the calculated particles. The introduction of surrogate model is to combine the information of all particles together, so as to obtain the potential position of the optimal value in a short time and improve the optimization speed of PSO. The methods of training surrogate model mainly include Gaussian process regression, support vector machine, radial basis function network, regression tree, artificial neural network, etc. [24] Among these methods, Gaussian process regression is a widely used method, and it has been proved that it can obtain satisfactory training performance. Therefore, this paper mainly uses Gaussian process regression to train the surrogate model [25].

Assume that the training data set is

$$D = \left\{ \left(\vec{x}_i, y_i\right) \middle| i = 1, \cdots, M \right\}, \tag{2}$$

where $\vec{x}$ is the variable vector and y is the vector of fitness.

Gaussian process regression assumes that y follows the multivariate normal distribution, that is:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_M \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1M} \\ k_{21} & k_{22} & \cdots & k_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ k_{M1} & k_{M2} & \cdots & k_{MM} \end{bmatrix} \right), \tag{3}$$

where k is the covariance of the variable vector, and the matrix composed of the covariance of each vector can be represented by K.

When there is a new variable vector denoted by $x*$, the Gaussian process regression assumes that it still satisfies the multivariate normal distribution, the following equation can be obtained based on Equation (3):

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(x,x) & K(x,x^*) \\ K(x^*,x) & K(x^*,x^*) \end{bmatrix}\right). \tag{4}$$

The corresponding predicted value $y*$can be obtained from the properties of Gaussian distribution, which can be expressed as

$$y^* = K(x^*,x)K(x,x)^{-1}y. \tag{5}$$

From the above analysis, it can be seen that the key point affecting the regression performance of Gaussian process is the kernel function generating covariance matrix. In this paper, rational quadratic kernel is used for model training.

2.3. The Improvement of Yu's Method. For optimization problems in oil and gas industry, the fitness calculation of an objective function often takes a long time because of the calls of numerical simulators. In order to reduce the calculation time, many researchers improved the PSO algorithm by combining it with a surrogate model. Among them, the model proposed by Yu et al. is a quite typical model, and therefore, Yu's method was selected as the comparative model in this study. For the problem of finding the minimum value of the objective function, Yu's method mainly includes the following steps:

(1) Latin hypercube sampling is used to obtain M samples in the search space [26]. The objective function is called to calculate the fitness of samples, and the samples and corresponding fitness form the initial fitness sample database

(2) The fitness of the samples is arranged in ascending order, and the first N samples are selected to form the initial particle swarm. Then, the historical optimal and global optimal of the initial N particles are obtained according to the principle of PSO algorithm

(3) The first P samples are selected to train the surrogate model by a machine learning method

(4) The optimal value and optimal location of the surrogate model are obtained by PSO algorithm. After calculating the fitness at the optimal location by using the objective function, the optimal location of the surrogate model and its fitness are added into the sample database. Then, the particles are reordered in the sample database according to the fitness

(5) If the first P samples of the sample database have been changed, the surrogate model is retrained by the machine learning method

(6) Update the global best, and then update the particle swarm according to Equation (1)

(7) The finesses of the updated particles are estimated by the surrogate model. If the estimated fitness of a particle is smaller than the current historical optimal, the particle is selected as the potential particle

(8) Call the objective function to calculate the fitness of the potential particles, and add the potential particles and their fitness into the sample database

(9) Update the global best and the historical best position of each particle

(10) Judge whether the convergence condition is met. If not, return to step 3

It can be seen from the above steps that Yu's method greatly reduces the evaluation times of fitness by using surrogate model. However, the significant reduction of the evaluation times of fitness results in the slow growth of the sample number in the database. Therefore, the difference between the surrogate models trained in step 3 and step 5 may be small, and thus, the optimization convergence speed is slow at the late stage of optimization. In addition, Yu's method works well for parallel simulators with high speedup ratios, but it is prone to idle computing resources for serial or low speedup ratio simulators. For example, the number of potential particles screened in step 7 is far less than the total number of particles. Therefore, the number of cores called in the fitness calculation in step 8 is usually far less than the total number of cores of the computer.

Considering the good parallel characteristics of PSO algorithm, an improved surrogate-assisted particle swarm optimization (i-SAPSO) method which is based on Yu's method was proposed to make full use of computing resources and reduce the total number of iterations. The steps are as follows:

(1) Latin hypercube sampling is used to obtain the initial samples in the search space, and the fitness of each sample is calculated according to the objection function. Due to the independence between samples, parallel computing (MPI, CUDA, etc.) can achieve to make full use of computing resources. Considering that the number of samples to be calculated in the subsequent iteration process of i-SAPSO is much higher than that of Yu's method, the initial number of samples in i-SAPSO algorithm can be much less than that of Yu's method. The samples and their fitness form the initial sample database

(2) The fitness of the samples is arranged in ascending order, and the first N samples are selected to form the initial particle swarm

(3) Update the global best and the historical best of each particle according to the principle of PSO algorithm. Generate the new particle swarm according to Equation (1), and then obtain the fitness of each particle according to the objective function. Similarly, due

to the independence between particles, the fitness of each particle can be calculated in parallel. The new particle swarm and fitness are added to the database

(4) While performing step 3, appropriate computing resources are allocated to train the database to obtain the surrogate model. Then, the optimal value and location of the surrogate model are obtained by PSO algorithm

(5) Monitor the consumed time of step 4. In the early stage of optimization, the number of samples in the database is small, and the step 4 only takes a short time. Therefore, after step 4 is completed, use the computing resources occupied in step 4 to call the objective function to calculate the fitness of the optimal location screened by the surrogate model, and put the optimal location and its fitness into the sample database. However, when the number of samples is large in the late stage of optimization, the time spent in step 4 may be close to that of step 3. At this time, the fitness calculation of the optimal position of the surrogate model can be postponed to the next iteration; that is, the surrogate model is trained and optimized in the current iteration, and the fitness of potential particles is calculated in the next iteration. Moreover, when the number of samples is too large, the time spent in step 4 may exceed that of step 3. Then, the computing resources occupied by step 3 may be idle after the calculation is completed. At this time, the particles in the database shall be sorted, and the τ particles with the highest fitness are selected. At the same time, the optimization range is determined according to the following equation:

$$\begin{cases} lb^i = \min\left(x_1^i, x_2^i \cdots, x_\tau^i\right), \\ ub^i = \max\left(x_1^i, x_2^i \cdots, x_\tau^i\right). \end{cases} \tag{6}$$

(6) Judge whether the convergence condition is met. If not, return to step 3

It can be seen from the above steps that the main difference between the i-SAPSO method and Yu's method is that the surrogate model is no longer used to screen the potential historical optimal position. The good parallel feature of particle fitness calculation is used to quickly supplement the sample database, so as to realize the rapid accuracy improvement of the global surrogate model and acceleration of convergence. In addition, when the number of calculated particles is too large, the particles with high fitness are selected, and the local surrogate model is constructed. Due to the consideration of the invocation of computing resources in each step, the improved algorithm can almost achieve high availability of computing resources in the entire optimization process.

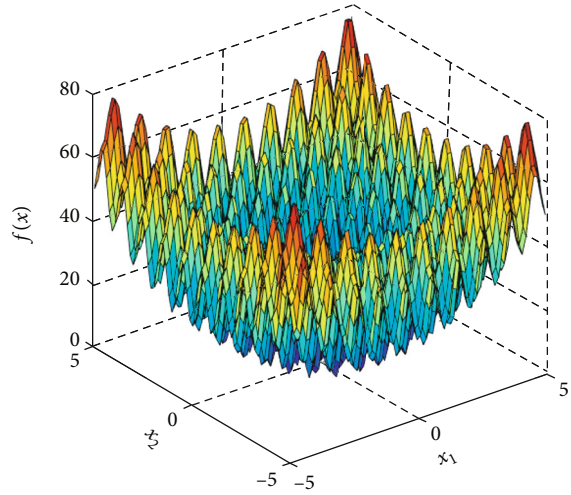


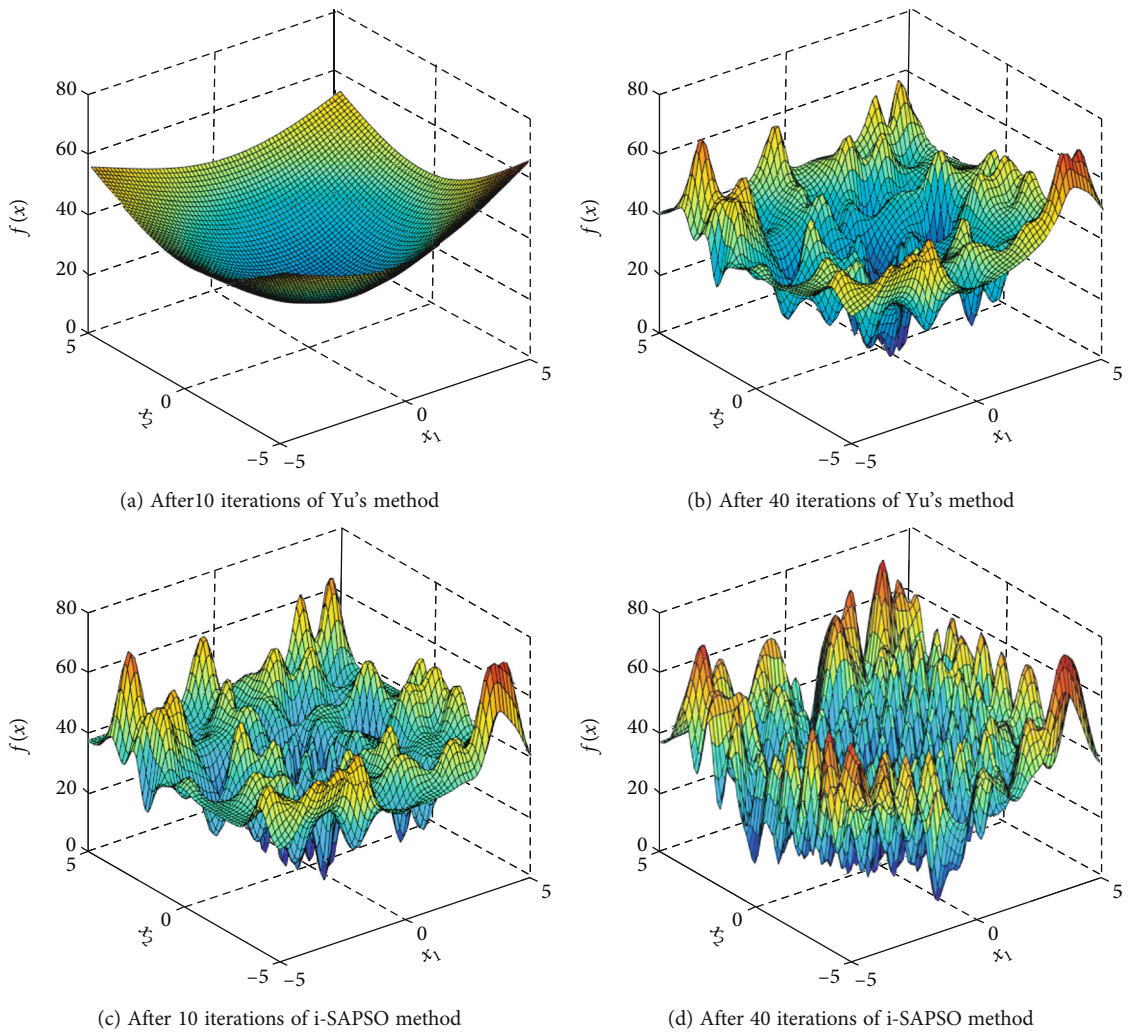
Figure 1: 2D Rastrigin function.

3. Performance Analysis and Comparison

Rastrigin function is a widely used function for testing optimization algorithms, and its expression is

$$f(x) = 10n + \sum_{i=1}^{n}\left[x_i^2 - 10\cos\left(2\pi x_i\right)\right]. \tag{7}$$

The model is a nonconvex function, and the global minimum value 0 is obtained when x_i are all 0. It can be seen from Equation (7) that changing the n value can construct a Rastrigin function of any dimension. In order to more intuitively compare the differences between the Yu's method and i-SAPSO method in terms of surrogate model training, a 2-dimensional Rastrigin function was first used for testing and analysis. Figure 1 shows the values of the 2-dimensional Rastrigin function on the interval [-5 5], from which it can be seen that the function has a multipeaked distribution, and there are many extreme value points, which can effectively test the optimum-seeking ability of global optimization algorithms. The optimization processes of the PSO algorithm, Yu's method, and i-SAPSO method are compared, and the initial Latin hypercube sampling points of the three methods are the same, and the number is 100. The maximum number of iteration is 300, and the iteration is stopped when the objective function value is lower than 1×10^{-10}. Figure 2 shows the comparison of the surrogate model evolution during the iterations of the Yu's method and i-SAPSO method. It can be seen from the figure that the Yu's method greatly reduces the calls of the objective function, and thus, its surrogate model is updated slowly. The surrogate model of Yu's method greatly differs from the Rastrigin function at the 10th iteration, and only after 40 iterations does the surrogate model show more local features of the Rastrigin function. Meanwhile, the surrogate model updates slowly in the subsequent iterations. The surrogate model in the i-SAPSO method can already characterize the local features of the Rastrigin function well after 10



(a) After10 iterations of Yu's method

(b) After 40 iterations of Yu's method

(c) After 10 iterations of i-SAPSO method

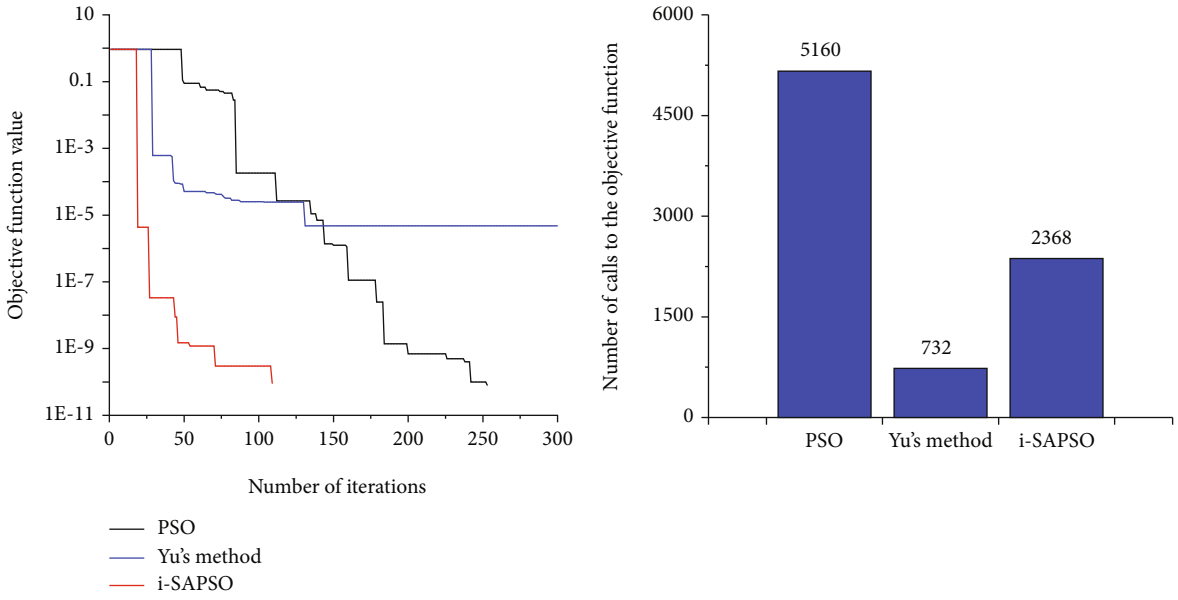
(d) After 40 iterations of i-SAPSO method

Figure 2: Evolution of surrogate model.

iterations, and the local features of the function are more accurately characterized by the surrogate model after 40 iterations.

Figure 3(a) shows the comparison of the change of the objective function value with the number of iterations. From the results, it can be seen that since both the Yu's method and i-SAPSO method use the surrogate model to predict the potential particle positions; the decrease of the objective function is obviously faster than that of PSO at the early stage of optimization. Meanwhile, because the number of particles in the sample database of the i-SAPSO increases rapidly and the surrogate model evolution quickly in the early stage, the probability of obtaining potential particles is much greater than that of Yu's method. Correspondingly, the objective function of i-SAPSO method decreases more rapidly than that of Yu's method. The PSO and i-SAPSO methods reach the preset accuracy after 253 and 109 iterations, respectively, but the Yu's method cannot further reduce the objective function value after 131 iterations. Finally, Yu's method exits the calculation when it reaches the maximum number of iterations of 300. From Figure 3(b), it can be seen that the PSO algorithm has the highest number of objective function calls, while the Yu's model has the lowest number of calls, and the i-SAPSO is in between. From the comparison results, it can be seen that the Yu's method is more suitable for cases where the requirements for optimization results are not critical and the computational process of the objective function is highly parallel. On the contrary, if a higher fitness is expected and the computational process of the objective function is not parallel, i-SAPSO can reasonably organize the computation steps to achieve fast parallel computation.

The number of unknowns can often reach tens or more when optimizing actual oil and gas field development parameters. For this reason, the dimension of the Rastrigin model is changed, and the comparison of the algorithms is carried out. The number of iterations is set to 300, and Table 1 shows the optimized objective function values for the three algorithms. It can be seen from the table that the performance of each algorithm in multidimensional case is basically the same as that in two-dimensional case. That is, under the same number of iterations, the performance of i-



(a) Objective function value

(b) Number of calls to the objective function

Figure 3: Performance comparison of different methods.

Table 1: Optimized objective function values of different methods.

	PSO	Yu's method	i-SAPSO
10-dimension	1.98	2.83	0.21
20-dimension	19.84	12.65	1.78
30-dimension	78.63	110.90	4.73

SAPSO is the best, followed by the PSO algorithm, and the call times of the objective function of Yu's method is the least, so the fitness is the worst under the same number of iterations. In addition, the nonlinearity of the objective function increases with the increase of dimension, and therefore, the probability of falling into the local optimum increases for all three algorithms as the number of dimensions increases.

4. Model Application

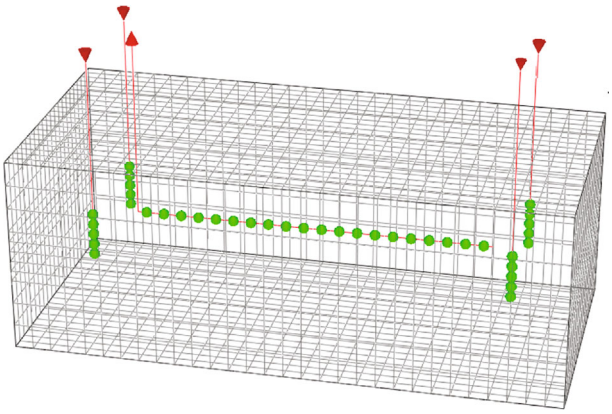
China has implemented two trial production tests of hydrate bearing layers in the Shenhu area, and the average daily gas production rate of the second trial production test is $2.87 \times 104\,m^3$, which is quite lower than the minimum gas production rate required for commercial development [27–29]. Moridis et al. investigated the depressurization performance of the hydrate bearing layers in Mallik area and Alaska North Slope. The results show that the dissociation area is mainly located around the well, and the gas production is low only through depressurization method [29, 30]. Zhang et al. investigated the decomposition conditions of methane hydrate and the effect of hydrate saturation on the methane hydrate dissociation, and the results are similar to those of Moridis [32–34]. In order to enhance gas production, more and more attention has been paid to the combined method of depressurization and thermal stimulation [35, 36]. In this paper, the Tough+Hydrate software was used to establish a numerical simulation model for a Class III hydrate reservoir at station SH7 in the Shenhu area of the South China Sea, and the basic geological parameters of the model are shown in Table 2 [37].

The model is a five-point well pattern composed of four vertical wells and one horizontal well (Figure 4(a)). The grid system is $30 \times 21 \times 11$, and the grid size is $15\,m \times 10\,m \times 2\,m$. The length of the horizontal well in the center is 300 m, and the distance between the vertical well and the nearest perforation of the horizontal well is 75 m. In order to represent the heterogeneity of the hydrate reservoir, a nonuniform permeability distribution is generated by sequential Gaussian simulation method which is a widely used in geostatistics, as shown in Figure 4(b) [38]. The average permeability is 75 md, and the coefficient of variation is 0.4 [39]. Neumann boundary condition is used, and in order to avoid the usage of well fraction, each vertical well has a distance from the boundary. Four vertical wells and one horizontal well first produce at a constant pressure of 4 MPa. The gas production rate continually decreases due to the pressure drawdown and temperature decline. When the gas production rate reaches a critical value (critical gas production rate), the four vertical wells are converted to hot water injection well to enhance gas production. In order to get a better performance, the injection rate of each well is adjusted once during the heat injection process, and thus, the total number of optimized parameters is 11. The range for each parameter is shown in Table 3, where the sum of the injection rates of the four injection wells was always maintained at $300\,m^3/d$.

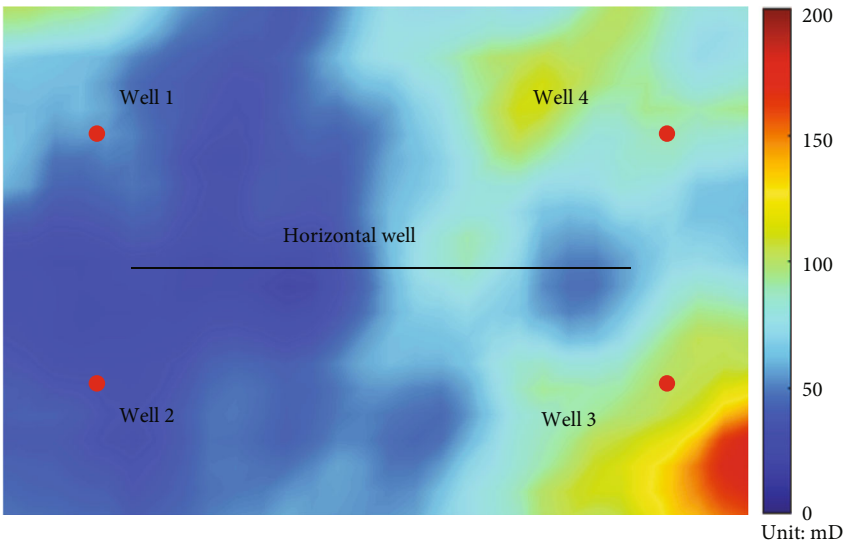
The development of gas hydrate reservoirs needs to consider both productivity and economy. Therefore, the evaluation index is divided into production index and economic index. The production index is mainly the methane recovery,

Table 2: Basic parameters of the hydrate reservoir model.

Parameter	Value	Parameter	Value
Seabed depth	1108 m	Reservoir thickness	22 m
Initial hydrate saturation	0.44	Initial water saturation	0.56
Average permeability	75 mD	Porosity	0.41
Initial pressure	13.83 MPa	Initial temperature	11.7°C



(a) Well layout

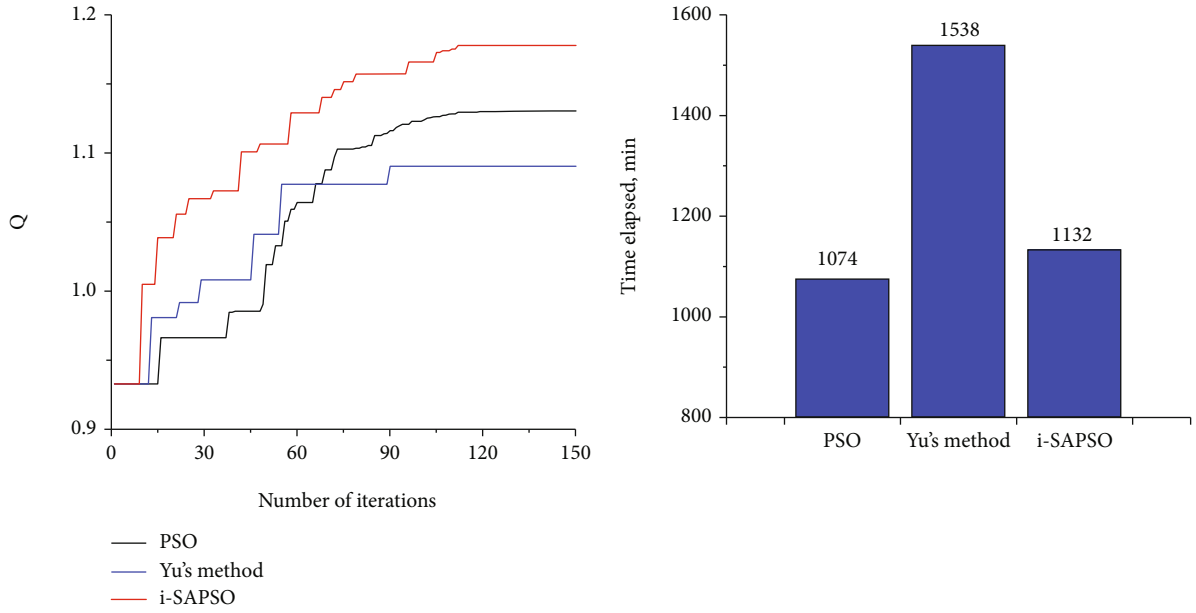


(b) Permeability distribution

Figure 4: Physical model of numerical simulation and permeability distribution.

Table 3: Ranges of optimized parameters.

Parameter	Value	Parameter	Value
Critical gas production rate	4000~20000 m^3/d	Temperature of injected water	20~90°C
Water injection rate of each well	0 ~ 200 m^3/d	Gas production rate when water injection rates are adjusted	4000~20000 m^3/d
Water injection rate of each well after adjustment	0 ~ 200 m^3/d	—	—



(a) Objective function value (b) Time elapsed

Figure 5: Performance comparison of different methods.

Table 4: Optimal parameter values of i-SAPSO method.

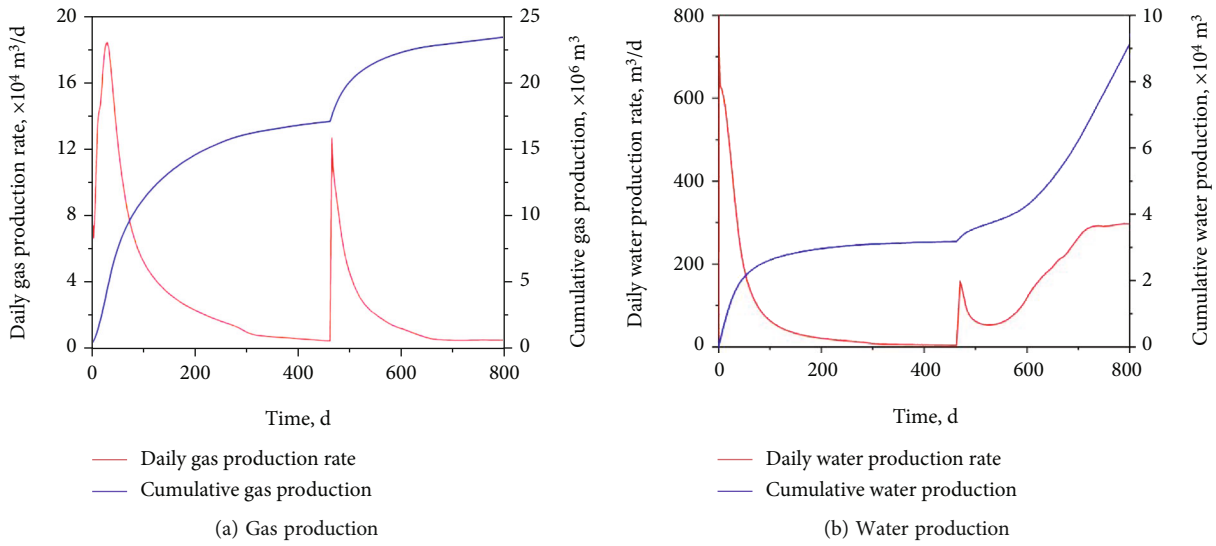
Parameter	Value	Parameter	Value
Critical gas production rate	4254 m^3/d	Temperature of injected water	33°C
Water injection rate of well 1	58 m^3/d	Water injection rate of well 2	51 m^3/d
Water injection rate of well 3	104 m^3/d	Water injection rate of well 4	87 m^3/d
Gas production rate when water injection rates are adjusted	22518 m^3/d	Water injection rate of well 1 after adjustment	73 m^3/d
Water injection rate of well 2 after adjustment	64 m^3/d	Water injection rate of well 3 after adjustment	89 m^3/d
Water injection rate of well 4 after adjustment	74 m^3/d	—	—

and the economic index is mainly the ratio of produced energy to injected energy, that is, energy efficiency ratio. Thus, the objective function of this paper is

$$\begin{cases} Q = E_R + \alpha\eta, \\ E_R = \dfrac{N_D}{N_T}, \\ \eta = \dfrac{E_{\text{out}}}{E_{\text{in}}}, \end{cases} \tag{8}$$

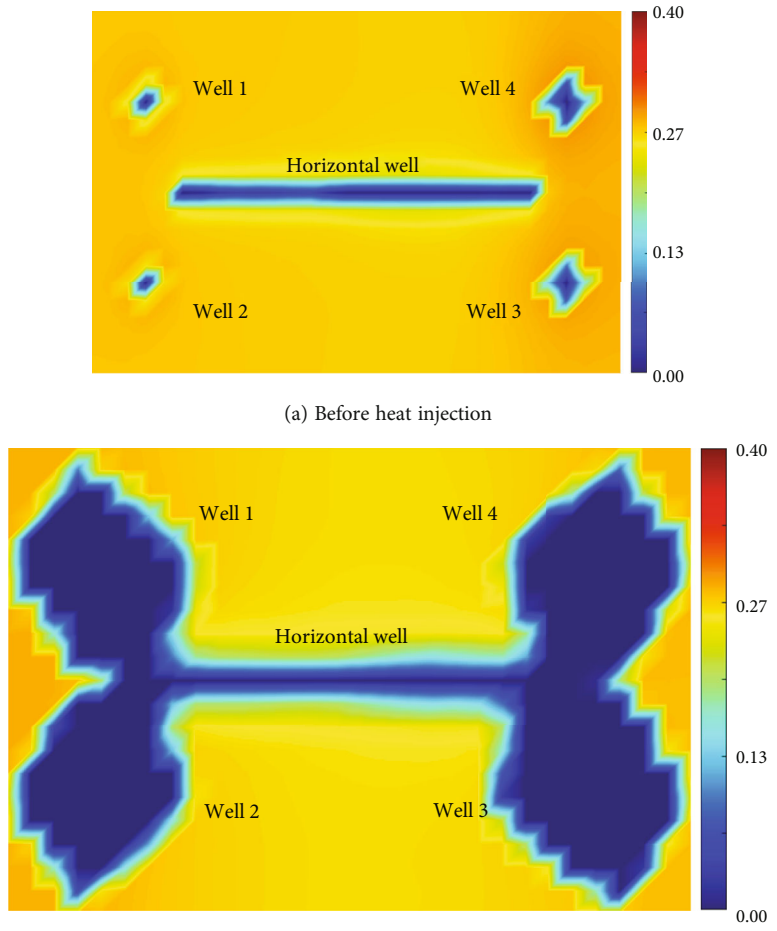
where Q is the objective function; E_R is the methane recovery; η is energy efficiency ratio; N_D is the cumulative volume of methane produced, m^3; N_T is the total volume of methane that can be generated from hydrate dissociation, m^3; E_{out} is the total heat of produced methane, J; E_{in} is the total heat of injected hot water, J; and α is the weight coefficient of E_R and η. According to the research results of Liu et al., the value α of is 0.025 [40].

Three algorithms are compared to optimize the key parameters. The simulated time of depressurization stage is 800 days, and the maximum number of iterations is 150. The test platform is XeonSP, which contains a 40-core processor, and the memory is 64G. From the perspective of making full use of computing resources, the number of particles is set to 35. In parallel computing of i-SAPSO method, each particle occupies one CPU core based on MPI, and thus, 35 cores are occupied by the particles. Three CPU cores are used for surrogate model training, and the remaining two CPU cores are used to process system applications. Figure 5 shows the performance comparison of the three algorithms. It can be seen from the figure that the change law of the objective function is consistent with that when Rastrigin function is used as the objective function. The Yu's method and i-SAPSO method can quickly find the potential position through the surrogate model. Therefore, the fitness quickly increases in the early stage of iteration, but the fitness increase of Yu's method becomes slower in the later stage which means that Yu's method falls into local optimum. Comparing the fitness at the end of the iteration,



(a) Gas production (b) Water production

Figure 6: Gas and water production of the optimal case.



(a) Before heat injection

(b) After heat injection

Figure 7: Hydrate saturation distribution.

it can be concluded that although it is difficult to prove that the optimization result of i-SAPSO method is globally optimal, the performance of i-SAPSO method is better than that of the PSO method and Yu's method under the same number of iterations. Figure 5(b) shows the comparison of simulation time, which indicates that the elapsed time of the PSO method is close to that of i-SAPSO method, while that of Yu's method is the longest. This is mainly due to the good independence between particles of PSO algorithm, which can easily realize parallel computing. Therefore, the elapsed time for 100 iterations is theoretically close to the time required to run the simulator 100 times continuously. Although the surrogate model training is added to the i-SAPSO method, additional computing resources are allocated. At the same time, the parallel computing framework between surrogate model training and particle fitness calculation is considered in the workflow, so the computing time is close to that of PSO method. Although Yu's method greatly reduces the total number of calls to the simulator, the computation process includes one simulator call in both step 4 and step 8, and the two calls cannot be computed parallelly. Therefore, the total calculation time is much higher than the other two algorithms.

Table 4 shows the values of the parameters obtained after the optimization of the i-SAPSO method. Combined with Table 3, it can be seen that the critical gas production rate is 4254 m^3/d, which is only slightly higher than the lower limit of the allowable range (4000~20000 m^3/d) which indicates that the time of depressurization should be long enough. This is because on one hand, the heat contained in rock and fluid can be fully used to promote the dissociation of hydrate if the depressurization period is long, and on the other hand, if hot water is injected too early, the hot water injected is easy to enter the production well directly along the dissociation area, which will lead to a waste of heat energy and reduction of energy efficiency. The optimized injected water temperature is 33°C, which is a relatively low injection temperature. Consistent with the conclusions of other researchers, a relatively low temperature of injected water is conducive to improve energy efficiency. The water injection rates of wells 1 and 2 located in the low permeability area is significantly lower than those of wells 3 and 4 located in the high permeability area, which helps to prevent the excessive pressure rise near the wellbore and the secondary formation of hydrate. Water injection rates of wells 3 and 4 after adjustment are lower than those before the adjustment, which can effectively avoid the rapid flow of injected hot water from the high permeability area to the bottom of the production well. Thus, the energy efficiency ratio can be enhanced.

Figures 6 and 7 show the gas and water production curves and hydrate saturation distributions of the optimal case, respectively. It can be seen that the gas production rate increases significantly after hot water injection. During depressurization development stage, the hydrate dissociation areas around wells 3 and 4 which are located in the high permeability area are significantly larger than those of wells 1 and 2, which indicates that higher permeability can achieve faster reservoir depressurization and hydrate dissociation. There is little difference in the shapes of the dissociation areas formed by the four wells after hot water injection, which means that balanced exploitation of hydrate reservoir can be realized through injection rate adjustment.

In conclusion, i-SAPSO method can effectively accelerate the optimization process, and the optimization results are satisfactory. Therefore, it can be used for parameter optimization of oil and gas reservoir development.

5. Conclusions

Based on the parallel computing framework, this paper proposed an improved surrogate model assisted particle swarm optimization method. Then, the effectiveness of the method was verified by Rastrigin function. Finally, the key parameters of depressurization−to−hot−water−flooding development of natural gas hydrate reservoir are optimized. The main conclusions are as follows:

(1) In a single iteration, the fitness calculation of each particle of PSO algorithm is independent. Therefore, parallel computing can be used to realize the rapid improvement of accuracy of global surrogate model. When the particle and fitness database is large, local surrogate model helps to achieve a quick convergence. The new method fully considers the parallel features of the calculation process and thus can get a better performance

(2) Yu's method greatly reduces the number of calls to the objective function. However, the accuracy of the surrogate model increases slowly with the iteration. Due to the independence of particles in an iteration, the purpose of the new model is to reduce the number of iterations rather than the number of calls to the objective function. The accuracy of the surrogate model increases significantly with the increase of iterations, and therefore, the iteration is expected to quickly converge for the new method

(3) The key parameters of a gas hydrate reservoir by depressurization−to−hot−water−flooding method are optimized. The results show that the fitness of the optimal case of the new method is significantly higher than those of Yu's method and PSO method under the same number of iterations. Meanwhile, the optimization results are satisfactory and consistent with the conclusions of other researchers. Therefore, the effectiveness of the new method is verified

Data Availability

The figure data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Sinopec Excellent Youth Innovation Fund Project (Grant No. P20025), the China Petroleum & Chemical Corporation for financial support (Grant No. P20040-4), and the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0102).

References

[1] V. Machairas, A. Tsangrassoulis, and K. Axarli, "Algorithms for optimization of building design: a review," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 101–112, 2014.

[2] J. Hou, K. Zhou, X. Zhang, X. Kang, and H. Xie, "A review of closed-loop reservoir management," *Petroleum Science*, vol. 84, pp. 363–370, 2016.

[3] B. Pouladi, A. Karkevandi-Talkhooncheh, M. Sharifi, S. Gerami, A. Nourmohammad, and A. Vahidi, "Enhancement of SPSA algorithm performance using reservoir quality maps: application to coupled well placement and control optimization problems," *Journal of Petroleum Science and Engineering*, vol. 189, 2020.

[4] L. Zhang, K. Zhang, Y. Chen et al., "Smart well pattern optimization using gradient algorithm," *Journal of Energy Resources Technology*, vol. 138, no. 1, pp. 1–13, 2016.

[5] T. Foroud, A. Baradaran, and A. Seifi, "A comparative evaluation of global search algorithms in black box optimization of oil production: a case study on Brugge field," *Journal of Petroleum Science and Engineering*, vol. 167, pp. 131–151, 2018.

[6] K. Khan and A. Sahai, "A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context," *International Journal of Intelligent Systems and Applications*, vol. 4, no. 7, pp. 23–29, 2012.

[7] C. Kim, R. Batra, L. Chen, H. Tran, and R. Ramprasad, "Polymer design using genetic algorithm and machine learning," *Computational Materials Science*, vol. 186, 2021.

[8] W. Zhang, A. Maleki, M. Rosen, and J. Liu, "Optimization with a simulated annealing algorithm of a hybrid system for renewable energy including battery and hydrogen storage," *Energy*, vol. 163, pp. 191–207, 2018.

[9] Y. Liu, J. Hou, Z. Chen et al., "Enhancing hot water flooding in hydrate bearing layers through a novel staged production method," *Energy*, vol. 217, 2021.

[10] K. Mohammadi and F. Ameli, "Toward mechanistic understanding of Fast SAGD process in naturally fractured heavy oil reservoirs: application of response surface methodology and genetic algorithm," *Fuel*, vol. 253, no. 1, pp. 840–856, 2019.

[11] Z. Ma and J. Y. Leung, "Design of warm solvent injection processes for heterogeneous heavy oil reservoirs: a hybrid workflow of multi-objective optimization and proxy models," *Journal of Petroleum Science and Engineering*, vol. 191, 2020.

[12] H. Rahmanifard and T. Plaksina, "Application of fast analytical approach and AI optimization techniques to hydraulic fracture stage placement in shale gas reservoirs," *Journal of Natural Gas Science and Engineering*, vol. 52, pp. 367–378, 2018.

[13] A. Bhosekar and M. Lerapetritou, "Advances in surrogate based modeling, feasibility analysis, and optimization: a review," *Computers & Chemical Engineering*, vol. 108, pp. 250–267, 2018.

[14] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, vol. 45, no. 1-3, pp. 50–79, 2009.

[15] A. Golzari, M. H. Sefat, and S. Jamshidi, "Development of an adaptive surrogate model for production optimization," *Journal of Petroleum Science and Engineering*, vol. 133, pp. 677–688, 2015.

[16] H. Yu, Y. Tan, J. Zeng, C. Sun, and Y. Jin, "Surrogate-assisted hierarchical particle swarm optimization," *Information Sciences*, vol. 454-455, pp. 59–72, 2018.

[17] X. Cai, H. Qiu, L. Gao, C. Jiang, and X. Shao, "An efficient surrogate-assisted particle swarm optimization algorithm for high-dimensional expensive problems," *Knowledge-Based Systems*, vol. 184, 2019.

[18] P. Zhang, Z. Yin, Y. Jin, and T. H. T. Chan, "A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest," *Engineering Geology*, vol. 265, 2020.

[19] G. Chen, K. Zhang, L. Zhang et al., "Global and local surrogate-model-assisted differential evolution for waterflooding production optimization," *SPE Journal*, vol. 25, no. 1, pp. 105–118, 2020.

[20] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995.

[21] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, pp. 1945–1950, Washington, DC, USA, 1999.

[22] Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*,, pp. 81–86, Seoul, Korea (South), 2001.

[23] Y. Gong, J. J. Li, Y. Zhou et al., "Genetic learning particle swarm optimization," *IEEE Transactions on Cybernetics*, vol. 46, no. 10, pp. 2277–2290, 2016.

[24] C. Sun, Y. Jin, J. Zeng, and Y. Yu, "A two-layer surrogate-assisted particle swarm optimization algorithm," *Soft Computing*, vol. 19, no. 6, pp. 1461–1475, 2015.

[25] Z. Zhou, Y. S. Ong, M. H. Nguyen, and D. Lim, "A study on polynomial regression and Gaussian process global surrogate model in hierarchical surrogate-assisted evolutionary algorithm," in *2005 IEEE Congress on Evolutionary Computation*, pp. 2832–2839, Edinburgh, UK, 2005.

[26] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.

[27] Y. Liu, J. Hou, H. Zhao, X. Liu, and Z. Xia, "A method to recover natural gas hydrates with geothermal energy conveyed by CO2," *Energy*, vol. 144, no. 1, pp. 265–278, 2018.

[28] Z. Xu, T. Hu, X. Pang et al., "Research progress and challenges of natural gas hydrate resource evaluation in the South China Sea," *Petroleum Science*, 2021.

[29] X. Pang, Z. Chen, C. Jia et al., "Evaluation and re-understanding of the global natural gas hydrate resources," *Petroleum Science*, vol. 18, no. 2, pp. 323–338, 2021.

[30] G. J. Moridis, "Numerical studies of gas production from Class 2 and Class 3 hydrate accumulations at the Mallik site, Mackenzie delta, Canada," *SPE Reservoir Evaluation & Engineering*, vol. 7, no. 3, pp. 175–183, 2004.

[31] G. J. Moridis, S. Silpngarmlert, M. T. Reagan, T. Collett, and K. Zhang, "Gas production from a cold, stratigraphically-bounded gas hydrate deposit at the Mount Elbert Gas Hydrate Stratigraphic Test Well, Alaska North Slope: implications of uncertainties," *Marine and Petroleum Geology*, vol. 28, no. 2, pp. 517–534, 2011.

[32] Y. Zhang, X. Li, Y. Wang, Z. Chen, and G. Li, "Methane hydrate formation in marine sediment from South China Sea with different water saturations," *Energies*, vol. 10, no. 4, pp. 561–561, 2017.

[33] Y. Zhang, X. Li, Y. Wang, Z. Chen, and K. Yan, "Decomposition conditions of methane hydrate in marine sediments from South China Sea," *Fluid Phase Equilibria*, vol. 413, pp. 110–115, 2016.

[34] Y. Zhang, X. Li, Z. Chen, Y. Wang, and X. Ruan, "Effect of hydrate saturation on the methane hydrate dissociation by depressurization in sediments in a cubic hydrate simulator," *Industrial & Engineering Chemistry Research*, vol. 54, no. 10, pp. 2627–2637, 2015.

[35] Q. Wan, H. Si, B. Li et al., "Energy recovery enhancement from gas hydrate based on the optimization of thermal stimulation modes and depressurization," *Applied Energy*, vol. 278, no. 15, pp. 115612–115614, 2020.

[36] Z. Yang, H. Si, and D. Zhong, "AI-based composition model for energy utilization efficiency optimization of gas hydrate recovery by combined method of depressurization and thermal stimulation," *Journal of Natural Gas Science and Engineering*, vol. 92, pp. 104001–104015, 2021.

[37] Y. Song, L. Yang, J. Zhao et al., "The status of natural gas hydrate research in China: a review," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 778–791, 2014.

[38] R. Dimitrakopoulos and X. Luo, "Generalized sequential gaussian simulation on group size v and screen-effect approximations for large field simulations," *Mathematical Geology*, vol. 36, no. 5, pp. 567–591, 2004.

[39] G. F. Reed, F. Lynn, and B. D. Meade, "Use of coefficient of variation in assessing variability of quantitative assays," *Clinical and Vaccine Immunology*, vol. 9, no. 6, pp. 1235–1239, 2002.

[40] Y. Liu, Y. Bai, Z. Xia, and J. Hou, "Parameter optimization of depressurization–to–hot–water–flooding in heterogeneous hydrate bearing layers based on the particle swarm optimization algorithm," *Journal of Natural Gas Science and Engineering*, vol. 53, pp. 403–415, 2018.

Hindawi
Geofluids
Volume 2021, Article ID 2168622, 13 pages
https://doi.org/10.1155/2021/2168622

WILEY | Hindawi

Research Article

A Modeling Study of the Productivity of Horizontal Wells in Hydrocarbon-Bearing Reservoirs: Effects of Fracturing Interference

Lei Huang,[1] Peijia Jiang,[2] Xuyang Zhao,[3] Liang Yang,[4] Jiaying Lin,[1] and Xuyang Guo[1,5]

[1]*State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China*
[2]*Beijing Gas Group Company Limited, Beijing 100035, China*
[3]*Geological Research Institute of CNPC Logging Company Limited, Xi'an, Shaanxi 710000, China*
[4]*Supervision Center of CNPC Tarim Oilfield Company, Korla, Xinjiang 841000, China*
[5]*Department of Petroleum Engineering, China University of Petroleum-Beijing at Karamay, Karamay, Xinjiang 834000, China*

Correspondence should be addressed to Xuyang Guo; xguo@cup.edu.cn

Received 31 October 2021; Accepted 25 November 2021; Published 14 December 2021

Academic Editor: Bicheng Yan

Commercial production from hydrocarbon-bearing reservoirs with low permeability usually requires the use of horizontal well and hydraulic fracturing for the improvement of the fluid diffusivity in the matrix. The hydraulic fracturing process involves the injection of viscous fluid for fracture initiation and propagation, which alters the poroelastic behaviors in the formation and causes fracturing interference. Previous modeling studies usually focused on the effect of fracturing interference on the multicluster fracture geometry, while the related productivity of horizontal wells is not well studied. This study presents a modeling workflow that utilizes abundant field data including petrophysical, geomechanical, and hydraulic fracturing data. It is used for the quantification of fracturing interference and its correlation with horizontal well productivity. It involves finite element and finite difference methods in the numeralization of the fracture propagation mechanism and porous media flow problems. Planar multistage fractures and their resultant horizontal productivity are quantified through the modeling workflow. Results show that the smaller numbers of clusters per stage, closer stage spacings, and lower fracturing fluid injection rates facilitate even growth of fractures in clusters and stages and reduce fracturing interference. Fracturing modeling results are generally correlated with productivity modeling results, while scenarios with stronger fracturing interference and greater stimulation volume/area can still yield better productivity. This study establishes the quantitative correlation between fracturing interference and horizontal well productivity. It provides insights into the prediction of horizontal well productivity based on fracturing design parameters.

1. Introduction

Wells in low permeability reservoirs bearing hydrocarbons typically have low productivity as it is hard for hydrocarbons to efficiently flow. Therefore, horizontal wells and hydraulic fractures are often used to enhance the contact between the wellbore and the matrix [1–6]. Since multistage and multicluster hydraulic fracturing facilitates the establishment of complex fracture networks, this type of fracturing technique is widely used in the development of unconventional reservoirs such as shale oil reservoirs [7–9]. However, due to the stress changes induced by hydraulic fracture initiation and propagation, multistage and multicluster hydraulic fracturing is affected by stress interference, and the geometry of the fracture network can be negatively impacted. It can lead to nonuniform hydraulic fracture growth and unevenly distributed fracture networks [9]. As fracture quality is directly related to horizontal well productivity, it is meaningful to quantitatively understand the relationship between hydraulic fracturing stress interference and horizontal well productivity. In order to quantify the relationship between stress interference in hydraulic fracturing and the production

performance of the fractured horizontal well, it is important to describe the fracture mechanics during hydraulic fracturing and the fluid flow in porous media during horizontal well depletion.

To build an efficient and reliable modeling method for the temporal and spatial evolution of stress, hydraulic fracture width and path, pressure, and hydrocarbon production, adequate modeling techniques should be employed. In geologic media, multiphase fluids flow in the porous media. Hydrocarbon-bearing reservoirs are usually characterized by rock heterogeneity, where fractured and unfractured media play an important role in the evaluation of pressure and stress evolutions [10, 11]. In such problems, the coefficient matrices in multiple time steps in the numerical system are usually huge, and the use of compositional and multiphase flow models and multiporosity assumptions further increases the complexity and leads to higher computational costs [12–14]. Therefore, the accurate modeling of the stress interference and the multistage fractured horizontal well productivity should be specifically investigated.

Hydraulic fracturing involves the use of high-pressure fluid injection into the reservoir to break the rock, forming fractures with high conductivities. This makes it easier for reservoir fluids to flow into the wellbore, thereby increasing oil and gas production. This process can activate natural fractures and increase the complexity of fracture networks. To understand and optimize hydraulic fracturing parameters, researchers have proposed many mathematical models for the simulation of hydraulic fractures. The early researches generally assumed simplified fracture geometries using the 2D plane strain assumption. The widely used Khristinaovic–Geertsma–de Klerk (KGD) model and Perkins–Kern–Nordgren (PKN) model calculate fracture geometries in two-dimensional planes, and fracture width distributions are calculated as a ratio of length to height [15–18]. 2D and 3D models were then derived, and the effect of in-situ stress and fluid flow was incorporated [19–21].

Industrialized exploitation of shale gas and shale oil reservoirs requires more sophisticated models describing the process of multistage and multicluster hydraulic fracturing. In this process, it is necessary to take into account the effect of the evolution of stress fields as there are interstage and intercluster interferences. As a result, hydraulic fractures with nonuniform half-lengths can be generated. Nonplanar and asymmetric fracture geometries can be generated under the influence of stress interference as well [22, 23]. The interference between simultaneously growing fractures can be intensified by decrease fracture spacing, and a minimum spacing should be determined for good fracturing quality [24]. Optimized hydraulic fracture-related parameters including fracturing timing and cluster location were determined in a reservoir-geomechanics-fracturing workflow where the effect of production-induced stress state changes is also considered [25, 26]. In addition to numerical modeling, triaxial tests are usually used in the lab to physically understand the fracturing initiation and propagation process [27, 28]. Monitoring techniques are also developed to better understand the shape and geometry of fractures in shale reservoirs [29, 30].

After the establishment of hydraulic fractures in the horizontal well, numerical simulation techniques are used to compute the fluid flow from the low permeability reservoir to the wellbore [31]. Stress sensitivity was sometimes incorporated in reservoir simulators, and the production prediction was affected by poroelasticity. Typically, the consideration of geomechanical effects tends to decrease the predicted production [32, 33]. In another reservoir simulation model, Moradi et al. [34] pointed out that changes in fracture aperture significantly alter the simulated production rates.

Previous studies usually focused on the stress interference phenomenon during the propagation of multiple hydraulic fractures, and the quantitative impact of this phenomenon on the horizontal well productivity has not been thoroughly investigated. This study employs hydraulic fracturing modeling and reservoir simulation techniques and proposes a modeling workflow quantifying the relationship between stress interference and horizontal well productivity. Parametric studies are also conducted to investigate the stage and cluster parameters on horizontal well productivity. This workflow provides a reference for the optimization of hydraulic fracturing parameters based on fracturing quality and horizontal well productivity modeling.

2. Methodology

In the methodology, a combined modeling workflow consisting of hydraulic fracturing modeling and porous media flow is presented. Finite element methods are used to establish the numerical models. Fracture propagation, geomechanical response, and fluid flows are all considered.

The momentum balance in the stress tensor is used to describe the rock deformation.

$$\nabla \cdot \sigma = 0. \tag{1}$$

The boundary condition for the rock deformation problem has three types. They are the stress boundary, the pressure boundary, and the displacement boundary:

$$\sigma \cdot n = \bar{t} \text{ for } \Gamma_t, \tag{2}$$

$$\sigma \cdot n = -pn \text{ for } \Gamma_p^+ \cap \Gamma_p^-, \tag{3}$$

$$u = \bar{u} \quad \text{for} \quad \Gamma_u, \tag{4}$$

where n represents the unit normal vector, Γ_t represents the traction boundary, $\Gamma_p^+ \cap \Gamma_p^-$ denotes the pressure boundary, and Γ_u represents the displacement boundary.

The Poiseuille's law is employed to compute the incompressible fracturing fluid flow in fractures as follows:

$$q_t = -\frac{w^3}{12\mu}\frac{\partial p}{\partial l}. \tag{5}$$

In Equation (5), w is the width of fracture at location l; p is the pressure; μ is the viscosity; q_t is the flow rate [35].

Then, a lubrication equation can be used to characterize the mass balance in the fracture flow as follows:

$$\frac{\partial w}{\partial t} + \frac{\partial q_{fi}}{\partial s} + q_{ts} = 0, \tag{6}$$

where w is the fracture width, q_{fi} is the flow rate, and q_{ts} is the fluid loss in the fracture flow.

The fluid leak-off into the host rock can be calculated by the following:

$$q_{ts} = c(p_i - p_t), \tag{7}$$

where c is the leak-off coefficient, p_i is the pressure in the fracture, and p_t is the pressure in the formation.

In the fracture mechanism, the normal and shear stresses are depicted by the traction-separation method as follows:

$$\begin{aligned} t_n &= \begin{cases} (1-D)T_n, T_n \geq 0, \\ T_n, T_n < 0, \end{cases} \\ t_s &= (1-D)T_s, \\ t_t &= (1-D)T_t, \end{aligned} \tag{8}$$

where t_n, t_s, and T_t represent the normal and shear components for stresses; D represents the damage. Once the fracture is initiated, the damage factor gradually increases and the fracture propagation can be described [36].

A two-phase black oil model is used to calculate the production in the hydraulically fractured horizontal well. The mass balance equations are as follows:

$$\frac{\partial(\rho_o \phi s_o)}{\partial t} + \nabla \cdot (\rho_o v_o) = \rho_o q_o, \tag{9}$$

$$\frac{\partial(\rho_w \phi s_w)}{\partial t} + \nabla \cdot (\rho_w v_w) = \rho_w q_w, \tag{10}$$

where s_o and s_w are saturation values for oil and water, ρ_o and ρ_w are densities for oil and water, ϕ is porosity, q is the sink/source term, and t is time. The terms of $\partial(\rho_o \phi s_o)/\partial t$ and $\partial(\rho_w \phi s_w)/\partial t$ describe the accumulation of oil and water flows in porous media. $\nabla \cdot (\rho_o v_o)$ and $\nabla \cdot (\rho_w v_w)$ terms represent the fluxes.

Darcy's law is widely used for fluid flows in porous media with low flow rates. It is used in this model as follows:

$$v_o = -\frac{k k_{ro}}{\mu_o}(\nabla p_o - \rho_o g \nabla H), \tag{11}$$

$$v_w = -\frac{k k_{rw}}{\mu_w}(\nabla p_w - \rho_w g \nabla H), \tag{12}$$

where k_r is the relative permeability, μ is the viscosity, v is the velocity, k is the permeability, g is the gravitational acceleration, and H is the depth. In this model, it is assumed that the hydraulic fracture network is fully propped and a fracture permeability is prescribed to denote the fracture conductivity.

Putting Darcy's law and the mass balance together,

$$\frac{\partial(\rho_o \phi s_o)}{\partial t} + \nabla \cdot \left[-\rho_o \frac{k k_{ro}}{\mu_o}(\nabla p_o - \rho_o g \nabla H)\right] = \rho_o q_o, \tag{13}$$

$$\frac{\partial(\rho_w \phi s_w)}{\partial t} + \nabla \cdot \left[-\rho_w \frac{k k_{rw}}{\mu_w}(\nabla p_w - \rho_w g \nabla H)\right] = \rho_w q_w, \tag{14}$$

where more detailed forms of fluid flow diffusivity are obtained.

In the two-phase black oil model, the relationship between water and oil saturations and the initial sink/source rates can be written as follows:

$$\begin{cases} s_o + s_w = 1, \\ q_o(t=0) = q_w(t=0) = 0. \end{cases} \tag{15}$$

Based on the assumption of slightly compressible fluids in the reservoir, the compressibility water and oil can be defined as $C_w = 1/\rho_w \partial \rho_w / \partial p_w$ and $C_o = 1/\rho_o \partial \rho_o / \partial p_o$.

As a result, the flow diffusivities in Equations (12) and (13) can be extended as follows:

$$s_o \phi c_o \frac{\partial p_o}{\partial t} + \rho_o \phi \frac{\partial(1-s_w)}{\partial t} - \rho_o \nabla \cdot \left[\frac{k k_{ro}}{\mu_o}(\nabla p_o - \rho_o g)\right] = 0, \tag{16}$$

$$s_w \phi c_w \frac{\partial p_o}{\partial t} + \rho_w \phi \frac{\partial s_w}{\partial t} - \rho_w \nabla \cdot \left[\frac{k k_{rw}}{\mu_w}(\nabla p_o - \rho_w g)\right] = 0. \tag{17}$$

Neglecting capillary pressure terms, Equations (16) and (17) become the following:

$$\begin{aligned} &[s_w \phi C_w + (1-s_w)\phi C_o]\frac{\partial p_o}{\partial t} + \nabla \\ &\cdot \left[-\left(\frac{k k_{rw}}{\mu_w}\nabla p_o - \frac{k k_{rw}}{\mu_w}\frac{\partial p_c}{\partial s_w} + \frac{k k_{ro}}{\mu_o}\nabla p_o - \frac{k k_{ro}}{\mu_o}\rho_o g - \frac{k k_{rw}}{\mu_w}\rho_w g\right)\right] = 0, \end{aligned} \tag{18}$$

Table 1: Parameters for the base case of hydraulic fracturing modeling.

Parameter	Value
Young's modulus (GPa)	22.41
Poisson's ratio	0.25
Overburden stress (MPa)	64
Initial maximum horizontal stress (MPa)	70
Initial minimum horizontal stress (MPa)	52
Differential stress (MPa)	18
Fracture number in one stage	3
Stage number in a horizontal well	4
Fracture spacing within one stage	15
Spacing between two stages (m)	60
Matrix permeability (m^2)	1×10^{-16}
Matrix porosity	0.1
Total injection rate (m^3/min)	12
Fluid viscosity (mPa·s)	20
Leak-off coefficient (m/Pa·s)	1×10^{-13}

Table 2: Parameters for the base case of reservoir simulation modeling.

Parameter	Value
Reservoir area (m^2)	800000 (1000 × 800)
Pay zone depth (m)	100
Matrix permeability (m^2)	1×10^{-16}
Matrix porosity	0.15
Initial reservoir pressure (MPa)	22
Bottomhole pressure (MPa)	5
Water viscosity (mPa·s)	1
Oil viscosity (mPa·s)	25
Rock density (kg/m^3)	2530
Oil density (kg/m^3)	800
Water density (kg/m^3)	1000
Constant oil compressibility (Pa^{-1})	1×10^{-8}
Constant water compressibility (Pa^{-1})	1×10^{-8}
Fracture permeability (m^2)	1×10^{-14}
Irreducible water saturation	0.1
Residual oil saturation	0.1

For numerical treatment, a matrix form can be obtained as follows:

$$\begin{bmatrix} \phi s_w c_w + \phi(1-s_w)c_w & 0 \\ \phi s_w c_w & \phi \end{bmatrix} \begin{bmatrix} \frac{\partial p_o}{\partial t} \\ \frac{\partial s_w}{\partial t} \end{bmatrix} + \nabla \cdot (-1) \cdot \begin{bmatrix} \frac{kk_{ro}}{\mu_o} + \frac{kk_{rw}}{\mu_w} & 0 \\ \frac{kk_{rw}}{\mu_w} & 0 \end{bmatrix} \begin{bmatrix} \nabla p_o \\ \nabla s_w \end{bmatrix} + \nabla \cdot \begin{bmatrix} -\frac{kk_{ro}}{\mu_o}\rho_o g - \frac{kk_{rw}}{\mu_w}\rho_w g \\ -\frac{kk_{rw}}{\mu_w}\rho_w g \end{bmatrix} = 0. \tag{19}$$

Initial conditions and boundary conditions are needed to solve the porous media flow problem.

An initial condition is as follows:

$$p_o(t=0) = p_{\text{ini}}. \tag{20}$$

The boundary condition is as follows:

$$v \cdot n = 0 \quad \text{on} \quad \partial\Omega, \tag{21}$$

where v is a velocity tensor. Equation (21) represents a no-flow boundary condition for the flow problem.

3. Modeling Study

A synthetic scenario in the development of a shale oil reservoir is established for the numerical study. Modeling parameters in the synthetic case are based on a realistic field in Junggar Basin, northwestern China [37]. Table 1 records the modeling parameters related to hydraulic fracturing while Table 2 shows the parameters for the modeling of fractured horizontal well production in reservoir simulation.

In the modeling workflow, the fracturing and production from a segment in the horizontal wellbore are considered. In the base case, totally four stages are hydraulically fractured in the horizontal wellbore. In each stage, there are three fracture clusters. A fracturing fluid injection rate of 12 m^3/min per stage is used. The base case fracture stage spacing is 60 m. In the parametric study for hydraulic fracturing, the effects of stage number, cluster number, stage spacing, and injection volume are quantified. Then, the horizontal well production and pore pressure depletion in the reservoir corresponding to each hydraulic fracturing scenario are simulated. Thus, the relationship between the stress interference in hydraulic fracturing and the productivity can be quantified using this workflow.

In addition to the hydraulic fracturing modeling and productivity modeling in the aforementioned base case, several parametric studies are also carried out. In the parametric study for the effect of fracture stage and cluster, two scenarios are considered. One scenario has two stages and six clusters in each stage, and the other scenario has three stages with four clusters in each stage. In the parametric study for the effect of fracture stage spacing, another two scenarios of 40 m and 60 m stage spacings are considered. In the parametric study for the effect of fracturing fluid injection rate per stage, two scenarios of 10 m^3/min and 14 m^3/min are modeled.

3.1. Hydraulic Fracturing Modeling. In this section, the effects of fracture stage/cluster number, stage spacing, and fracturing fluid injection rate on the propagation of the hydraulic fracture networks are modeled. The resultant fracture length, fracture width, fracture area, and fracture

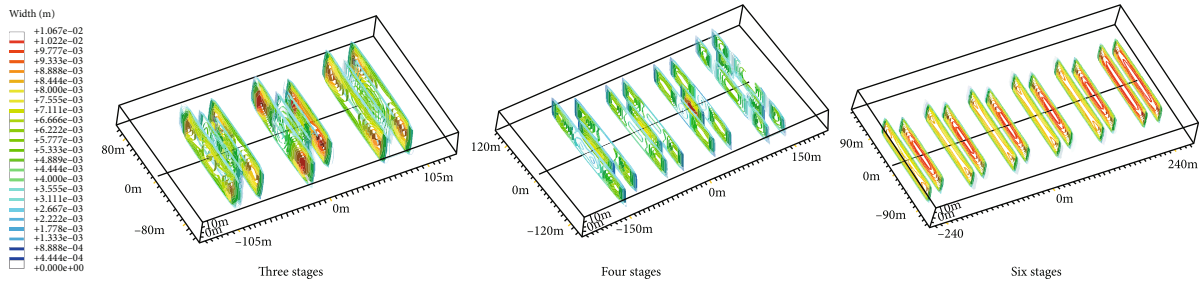


Figure 1: Fracture geometries of different stage and cluster numbers.

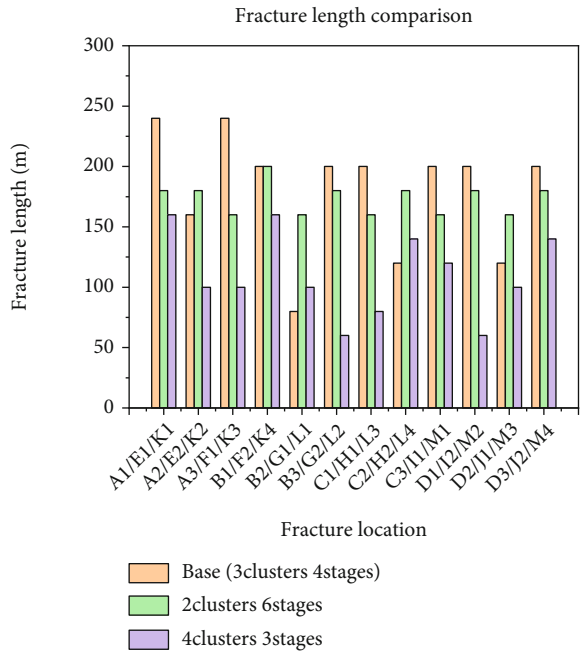


Figure 2: Comparison of fracture length for scenarios with different stage and cluster numbers.

volume are presented. The associated stress evolutions are also discussed.

3.1.1. Fracture Stage and Cluster. Three scenarios of fracture stage and cluster designs are modeled. Totally 12 clusters are kept as a constant, while cluster numbers in each stage are 3, 4, and 6. Other fracturing parameters are kept the same. Figure 1 shows the nonuniform hydraulic fractures in the horizontal well with the three different stage and cluster numbers. In general, the four-cluster design leads to the most nonuniform hydraulic fracture growth in the wellbore, with the two outer fracture clusters longer than the inner fracture clusters. Note that the stage spacing is kept as 60 m in all the scenarios. To better quantify the effect of stage and cluster design on the nonuniform fracture lengths, Figure 2 is plotted where the length of each individual fracture cluster is presented as bars. In Figure 2, the labels A to D represent the four fracture stages, and 1 to 3 represent the clusters in a stage. Note that A represents the first fractured stage while D is the last stage that is fractured in the multistage fracturing job. Similarly, labels E to J represent the six fracture stages in another design, where 1 to 2 are the two clusters in each stage. K to M are the three stages in the 4-cluster scenario. Based on the detailed fracture length results, it is noted that when the cluster number exceeds two in each stage, the inner fractures are always shorter than the outer fractures. This indicates that the stress interference in simultaneously growing clusters inhibits the growth of the inner fractures and makes the outer fracture more competitive in terms of fracture propagation [38]. The 2-cluster scenario results in a more evenly distributed fracture length. This indicates that reducing the clusters in each stage can lower the stress interference effect on the heterogeneous growth of fracture clusters. In this scenario, only two fractures grow simultaneously each time. Therefore, the interference between clusters is reduced, leading to a more uniform growth of fractures. In the 4-cluster scenario, since four fracture clusters are competing in the simultaneous growth, the inhibition on the inner fractures becomes more noticeable. Also, the average fracture lengths are the shortest in the 4-cluster scenario.

Generally, the longest fracture length of an individual fracture is obtained in the first stage in the 3-cluster scenario. This indicates that the first stage is less affected by interstage interference than the stages fractured later on. In addition, in the 2-cluster, 3-cluster, and 4-cluster scenarios, it is noted that the first stage has very symmetric fractures, indicating that the interference does not take effect unless there are sequentially fractured stages. The interstage stress interference is the most significant in the 4-cluster scenario, as the second and third stages have lower fracture lengths. These results show that the interstage interference on hydraulic fracturing interference increases with the cluster number in each stage. However, it is noted that although reducing the number of clusters in each stage helps to form uniform fracture networks, it may increase the cost of the fracturing operations and fracturing time.

Figure 3 presents the temporal evolution of fractured volume and fracture area during the multistage fracturing operation with different fracture stage and cluster designs. This helps to improve the understanding of the correlation between stage and cluster designs and fracture quality. Step-by-step trends are obvious in these results, as stages are sequentially fractured in the operation. The final fractured volumes for the 3-cluster and 2-cluster scenarios are very similar, while the selection of 4-cluster design leads to the lowest fractured volume. This is direct evidence that increased fracturing interference reduces the final fracture

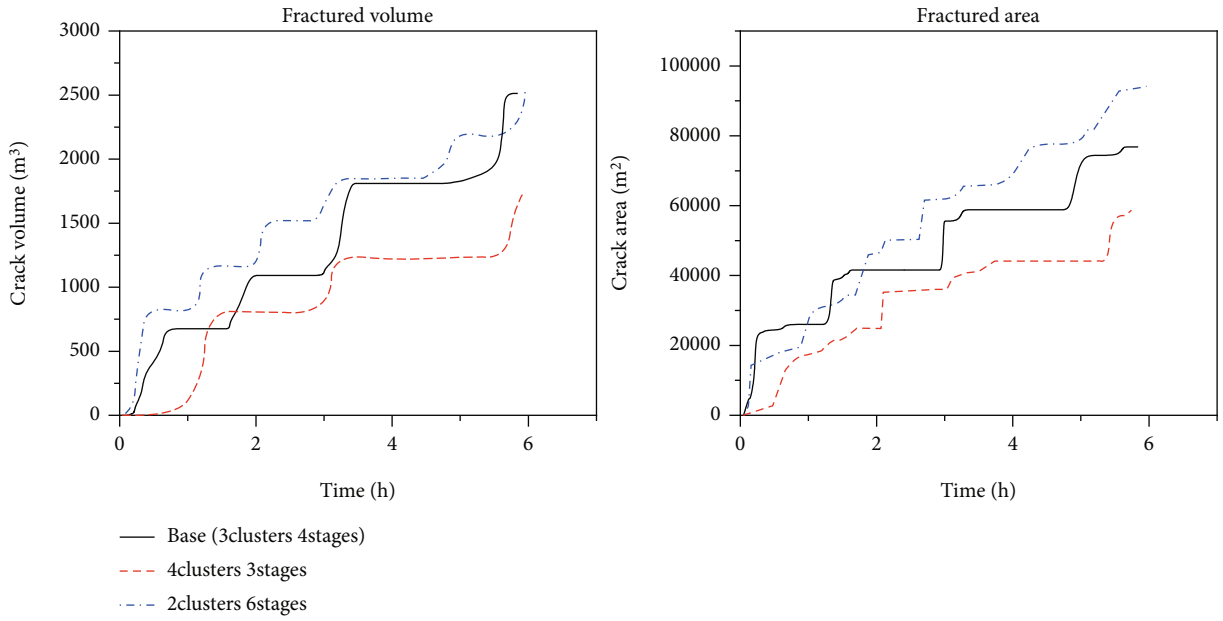


Figure 3: Evolutions of fractured volume and area for scenarios with different stage and cluster numbers.

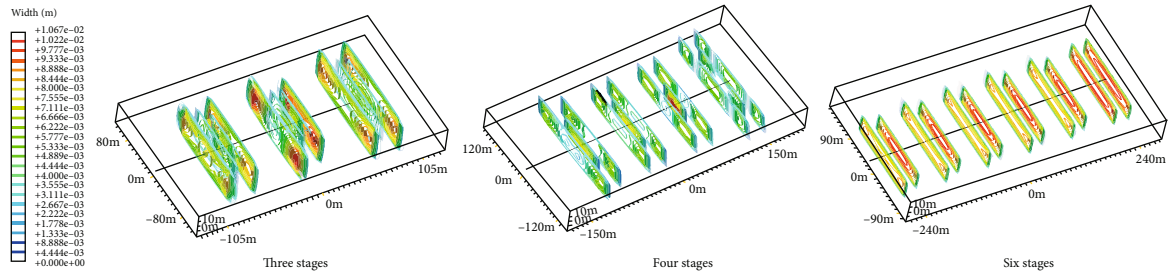


Figure 4: Fracture geometries of different stage spacings.

network quality. Based on the fractured area results, the 2-cluster scenario has the best fracture quality and the 3-cluster scenario has the intermediate quality. Again, the 4-cluster scenario leads to the lowest fracture quality. Note that the change from 3-cluster to 2-cluster design increases the contact area between fractures and the low permeability matrix, implying an improved fracturing performance.

3.1.2. Fracture Stage Spacing. In this section, the effect of the fracture stage on the hydraulic fracture quality is investigated. Three stage spacings of 40 m, 60 m, and 80 m are studied while other fracturing-related parameters are the same as the base case as in Table 1.

Figure 4 describes the fracture geometries from three different stage spacings. Note that the cluster spacing within each stage is kept the same during the parametric study. In general, when the stage spacing is reduced to 40 m, a more nonuniform fracture length pattern is observed. In contrast, when the stage spacing is increased to 80 m, a rather uniform distribution of fracture length is obtained. However, it is noted that the inner fracture growth is always inhibited by fracturing interference even when the spacing is large. It means that the inhibition on the inner fracture is caused by intercluster interference instead of interstage interference.

Figure 5 records the comparison of fracture length of each cluster in scenarios with different stage spacings. Since all three scenarios have the same stage number of four, labels A to D are used to represent the four stages and each stage has three clusters. Based on the fracture length results, the effect of stage spacing is not quite monotonic. The correlation between stage spacing and fracture cluster length is not clear.

Figure 6 shows the temporal changes in the fractured volume and the fractured area during the 4-stage fracturing. The fracturing of each individual stage leads to a sharp increase in the fractured volume and area. The fractured volume results indicate that the 80 m spacing leads to the greatest volume while the 60 m spacing has the lowest fractured volume. The fractured area results show that the 40 m spacing leads to the highest fractured area.

3.1.3. Fracturing Fluid Injection Rate. The fracturing fluid injection rate is a key parameter in designing the hydraulic fracturing operations. It directly governs the amount of fluid injected into the fractures, which are used to establish the net pressure for fracture propagation. In this study, three fluid injection rates of 10 m^3/min, 12 m^3/min, and 14 m^3/min are simulated. Other parameters are the same as the base case.

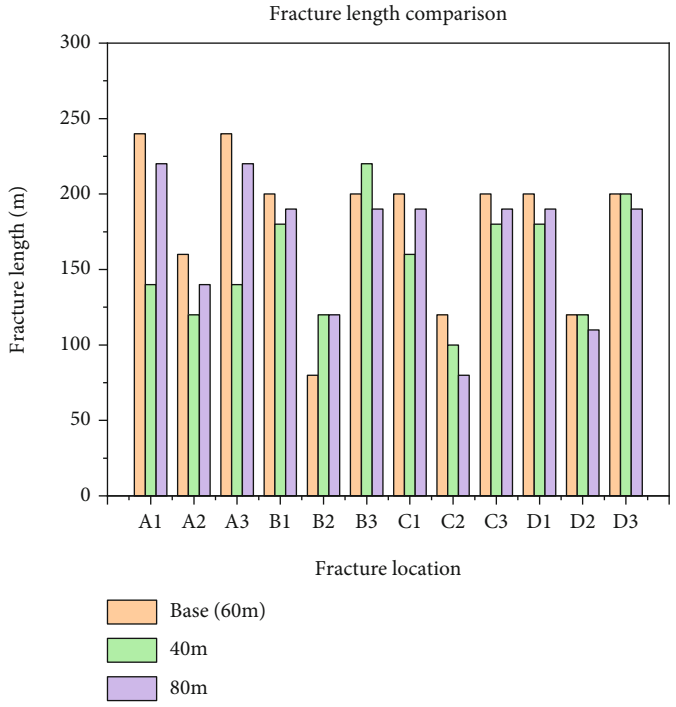


Figure 5: Comparison of fracture length for scenarios with different stage spacings.

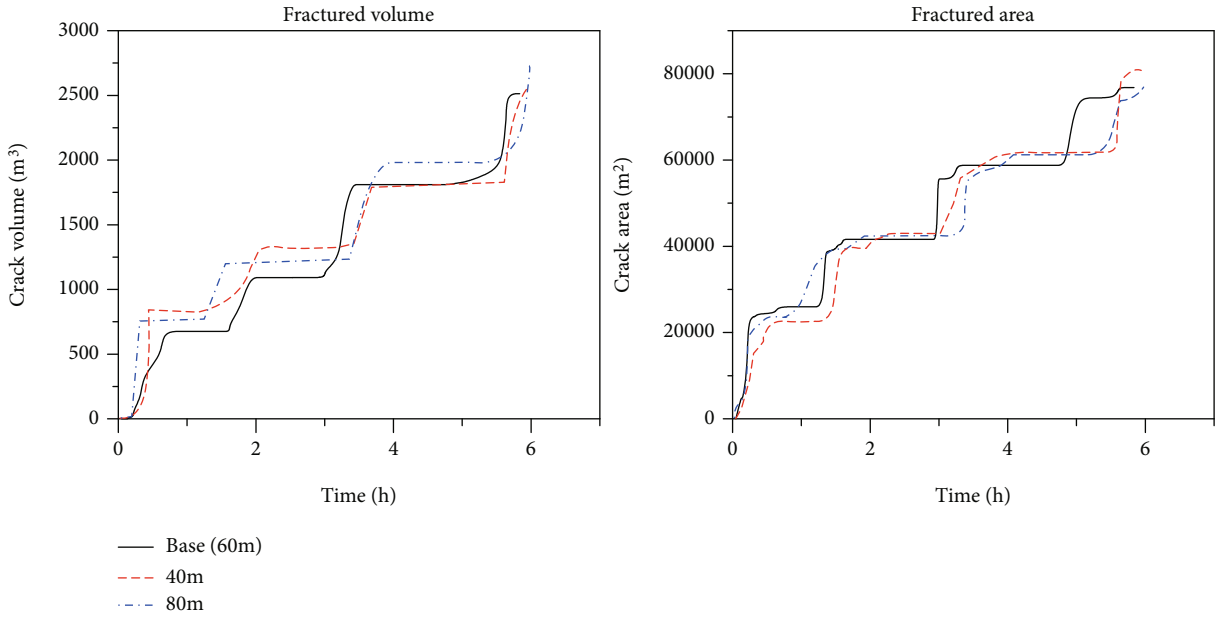


Figure 6: Evolutions of fractured volume and area for scenarios with different stage spacings.

The geometries are shown in Figure 7. Compared to the fracturing fluid injection rates 10 m^3/min and 12 m^3/min, the fracturing fluid injection rate of 14 m^3/min leads to a more nonuniform fracture geometry, and the nonuniform lengths are more significant between the first and the second stages. This shows that an increase in injection rate leads to a more unevenly distributed fracture network, and an increased injection rate corresponds to an elevated fracturing interference between hydraulic fracturing stages.

Figure 8 shows the cluster-by-cluster comparison of fracture length with various fluid injection rates. The correlation between injection rate and fracture length is well in the first stage (stage A). In the first stage, the longest fracture lengths are obtained when the injection rate is 14 m^3/min, and the fracture lengths decrease with the decrease in injection rate.

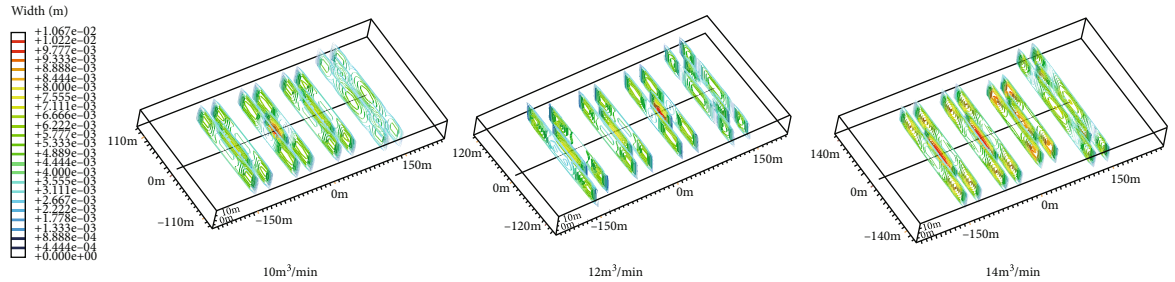


Figure 7: Fracture geometries of different fracturing fluid injection rates per stage.

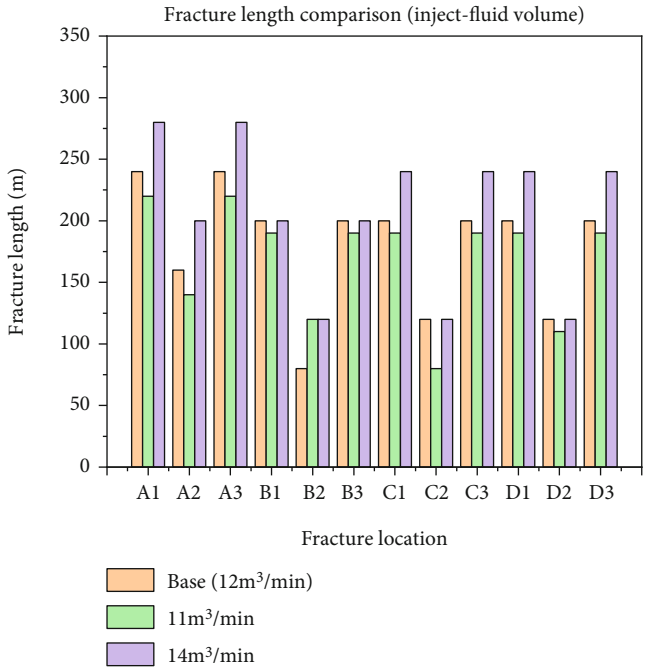


Figure 8: Comparison of fracture length for scenarios with different fracturing fluid injection rates per stage.

The general trend is the same for stages B to D, while oscillations in fracture length are observed. This is caused by the nonuniform results obtained in the simulation for the fracture mechanism.

In Figure 9, the stepwise increases in the fractured volume and area are plotted against fracturing time. In these results, the correlation between the injection rate and the fractured volume/area is clear. With the increase in fracturing fluid injection, the step-by-step fractured volume and area and the final fractured volume and area both increase. Since the total injection time is constant, a higher fracturing fluid injection rate corresponds to a greater volume of fluid injected into the fractures for fracture initiation and propagation.

Based on these parametric studies for the effect of fracture stage/cluster design, stage spacing, and fracturing fluid injection rate, stage spacing is less influential than stage/cluster design and injection rate. The reduction of cluster number per stage and the increase in stage spacing help to establish more evenly propagated fractures. The increase in injection rate can elevate the interstage fracturing interference by making the fracture lengths more uneven, while a greater injection rate improves the overall fractured volume and area.

The study in this section quantifies the hydraulic fracture geometries in the horizontal well, and the effect of fracturing interference is investigated in terms of fracture geometry. However, the investigation of horizontal well productivity requires results more than fracture geometry, and the productivity-related parameters such as pore pressure and hydrocarbon production should be quantified.

3.2. Productivity Modeling. In the previous section, fracture geometry, fractured volume and area, and fracture length are used as the variates to denote the effect of fracturing interference. To further investigate the horizontal well productivity, reservoir simulation techniques are used to calculate the production of the horizontal well with different hydraulic fracture geometries obtained in the previous fracturing modeling. Thus, the relationship between fracturing interference and horizontal well productivity can be established. In the modeling process, the fracture conductivity is assumed to be constant in each fracture and it does not change with time. The fracture conductivity is calculated as the product of the fracture width and a constant permeability value for the fracture. Modeling parameters are shown in Table 2. Productivity over 2 years of production is reported. Figures 10 and 11 compare the cumulative production of oil over 2 years from horizontal wells with three different cluster and stage designs, stage spacings, and fluid injection rates.

In Figure 10, the use of six stages with two clusters in each stage leads to the highest cumulative production. This is because this strategy has the lowest fracturing interference, and it can lead to even depletion within the low permeability reservoir. The use of four stages with three clusters in each stage (the base case) leads to intermediate cumulative production performance. In contrast, the use of four clusters in each stage results in a much lower cumulative oil production curve. Combined with results in Figures 1 and 2, it is noted that the use of four clusters in a stage largely inhibits the growth of fractures, especially for the two inner fractures in each stage. In this scenario, the average fracture length is the lowest, leading to lower production performance.

In Figure 12, how stage spacing affects cumulative production from a horizontal well is presented. In general, the

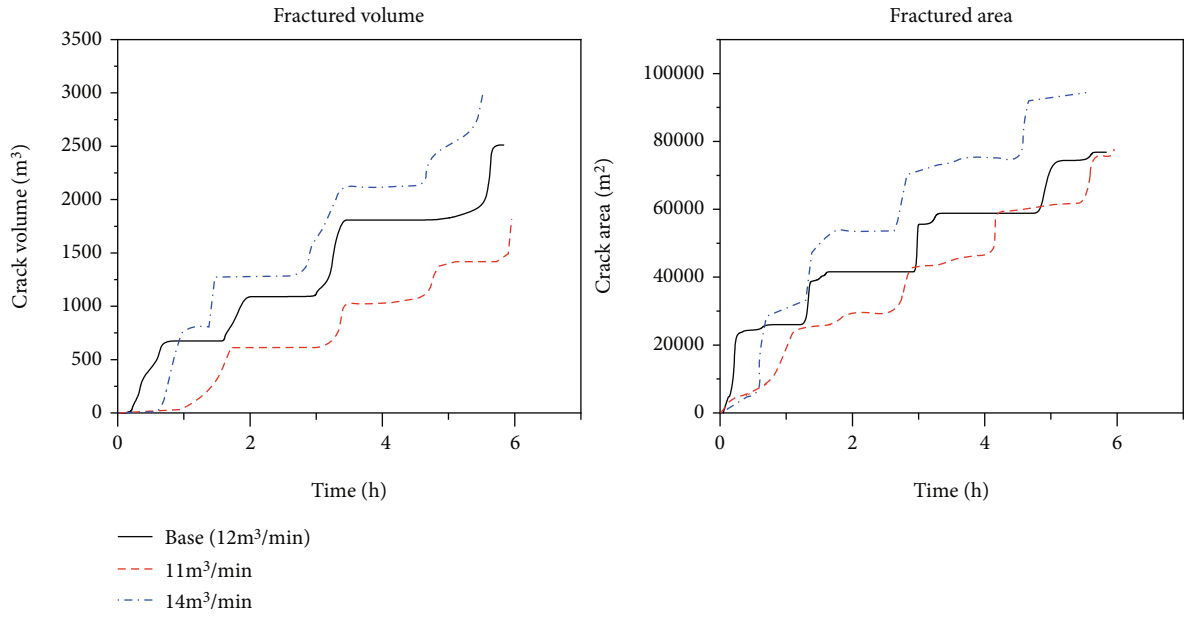


Figure 9: Evolutions of fractured volume and area for scenarios with different fracturing fluid injection rates per stage.

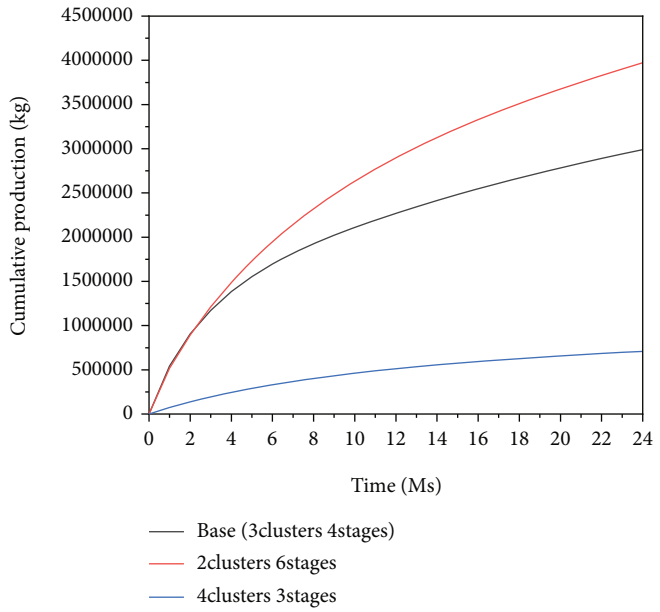


Figure 10: Cumulative oil production from horizontal wells with different stage and cluster designs.

effect of stage spacing is not as great as stage and cluster design, and the differences in cumulative production results are smaller. The base case of a 60 m spacing yields the highest cumulative production, while the stage spacing of 40 m has the lowest cumulative production. Based on the observations in Figures 4 and 5, reducing the stage spacing to 40 m strengthens the fracturing interference and leads to uneven fracture lengths, where the inhibition on the growth of certain fractures is also increased.

Cumulative oil productions from horizontal wells with three fracturing fluid injection rates during the hydraulic fracturing process are compared in Figure 11. Intuitively, the highest injection rate corresponds to the greatest cumulative production, and the lowest injection rate leads to the smallest cumulative production, which is also correlated with the fracture geometries and lengths in Figures 7 and 8.

Based on productivity modeling, the fracture geometry and the horizontal well productivity can be generally correlated: horizontal wells with longer fracture lengths and weaker fracturing interference usually yield higher cumulative production. However, the differences in horizontal well productivity from multiple hydraulic fracturing design

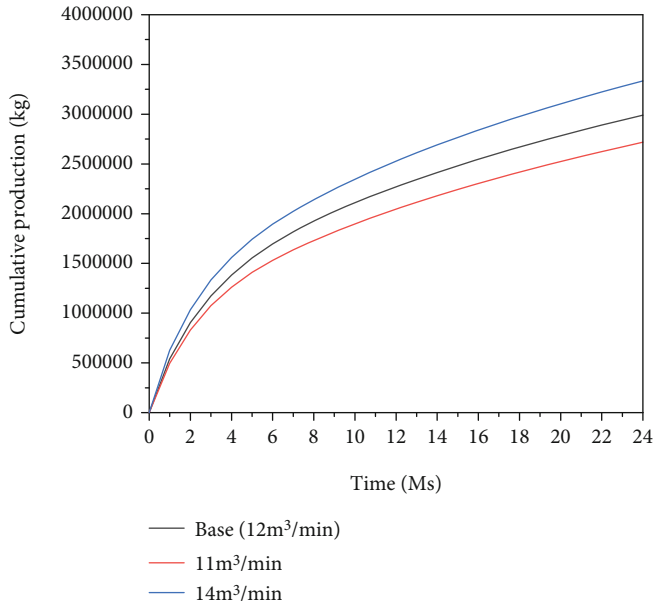


Figure 11: Cumulative oil production from horizontal wells with different fluid injection rates.

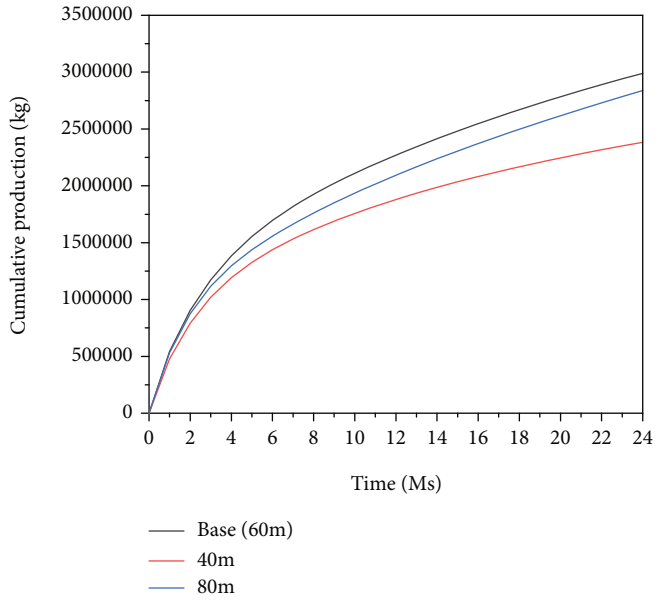


Figure 12: Cumulative oil production from horizontal wells with different stage spacings.

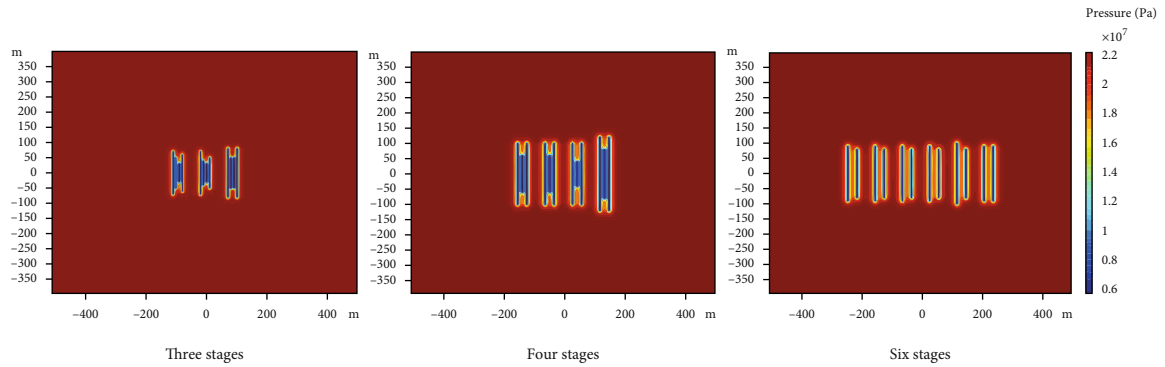


Figure 13: Pressure distribution after three months of production for multistage scenarios of three stages, four stages, and six stages.

scenarios cannot be directly quantified by hydraulic fracturing modeling, and reservoir simulation has to be used to obtain the detailed differences in cumulative oil production.

4. Discussion

Note that in this modeling study, planar hydraulic fractures are considered while it is also possible to have nonplanar hydraulic fractures generated in multistage hydraulically fractured horizontal wells. Nonplanar fracture modeling can better quantify the effect of stress interference on resultant hydraulic fracture geometries. For example, curved fractures with uneven lengths can be obtained using this modeling technique. However, the focus of this study is on the correlation between stress interference and horizontal well productivity. As shown in Figure 13, hydraulic fracture geometries directly govern the drainage area. After three months of depletion, the pressure drop fronts are around 10 m away from the fractures, indicating that the hydrocarbons within this area are being produced. Therefore, the fracture length and the drainage area jointly govern the resultant horizontal well productivity. In this workflow, the effects of curvatures of nonplanar hydraulic fractures on horizontal well productivity are weakened. In consequence, although planar fracture modeling used in this study cannot characterize how fractures are curved by stress interference, the obtained fractures are unevenly distributed and the effects on fracture lengths and the resultant drainage area are still honored.

5. Conclusion

In this modeling study, based on a numerical modeling workflow, the relationship between the fracturing interference during multistage hydraulic fracturing and horizontal well productivity is established, which is the primary contribution in the study. Using hydraulic fracture geometries to quantify the effect of fracturing interference on productivity is not comprehensive. Based on the modeled fracture geometries including fracture length, fractured volume, and fractured area, an estimate can be obtained for productivity. However, quantitative understanding should be obtained using a more comprehensive workflow including fracturing modeling and productivity modeling. In conclusion,

(1) Reducing the cluster number within a stage, increasing the spacing between two stages, and reducing the fracturing fluid injection rate help to decrease the negative impact of fracturing interference on multistage and multicluster fracturing

(2) Fractured volume and fractured area are parameters indicating the magnitude of the stimulated reservoir volume/area. They are generally correlated with fracture lengths. The temporal evolution patterns of fractured volume and area are highly stepwise, which corresponds to the sequentially fractured stages in the field operation

(3) Productivity modeling results are generally correlated with fracturing modeling results. However, productivity modeling is capable of providing the quantitative differences in cumulative production, which directly shows the effect of fracturing interference on horizontal well production performance

Nomenclature

σ: Stress tensor
n: Unit normal vector
$\bar{t}$: Traction boundary
p: Pressure, Pa
u: Displacement vector
w: Fracture width, m
μ: Viscosity, cP
s_o: Oil saturation
s_w: Water saturation
ρ_o: Oil density, kg/m^3
ρ_w: Water density, kg/m^3
ϕ: Porosity
t: Time, s
v: Velocity, m/s
k: Permeability, m^2
g: Gravitational acceleration, m/s^2
C_w: Water compressibility, 1/Pa
C_o: Oil compressibility, 1/Pa.

Data Availability

The data are available by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest

Acknowledgments

The authors acknowledge the financial support from the National Natural Science Foundation of China (No. 51904314, No. 51991362, and No. U19B6003-05), the Fundamental Research Funds for the Central Universities (No. 2462018YJRC031), and the Research Foundation of China University of Petroleum-Beijing at Karamay (No. XQZX20200009).

References

[1] J. Xie, J. Tang, R. Yong et al., "A 3-D hydraulic fracture propagation model applied for shale gas reservoirs with multiple bedding planes," *Engineering Fracture Mechanics*, vol. 2020, no. 228, article 106872, 2020.

[2] X. Zhao, L. Zhou, X. Pu et al., "Formation conditions and enrichment model of retained petroleum in lacustrine shale: a case study of the Paleogene in Huanghua depression, Bohai Bay Basin, China," *Petroleum Exploration and Development*, vol. 47, no. 5, pp. 916–930, 2020.

[3] J. Tang, B. Fan, L. Xiao et al., "A new ensemble machine-learning framework for searching sweet spots in shale reservoirs," *SPE Journal*, vol. 26, no. 1, pp. 482–497, 2021.

[4] S. Mao, Z. Zhang, T. Chun, and K. Wu, "Field-scale numerical investigation of proppant transport among multicluster hydraulic fractures," *SPE Journal*, vol. 26, no. 1, pp. 307–323, 2021.

[5] W. Cheng, G. S. Jiang, J. Y. Xie, Z. J. Wei, Z. D. Zhou, and X. D. Li, "A simulation study comparing the Texas two-step and the multistage consecutive fracturing method," *Petroleum Science*, vol. 16, no. 5, pp. 1121–1133, 2019.

[6] B. Lin, H. Meng, J. Pan, and S. Chen, "Porothermoelastic response of an oil sand formation subjected to injection and micro-fracturing in horizontal wells," *Petroleum Science*, vol. 17, no. 3, pp. 687–700, 2020.

[7] R. Weijermars, N. Sorek, D. Sen, and W. B. Ayers, "Eagle Ford Shale play economics: U.S. versus Mexico," *Journal of Natural Gas Science and Engineering*, vol. 38, pp. 345–372, 2017.

[8] G. Lindsay, G. Miller, T. Xu, D. Shan, and J. Baihly, "Production performance of infill horizontal wells vs. preexisting wells in the major US unconventional basins," in *SPE hydraulic fracturing technology conference and exhibition, the woodlands, Texas, USA, 23–25 January*, USA: Society of Petroleum Engineers, 2018.

[9] S. Mao, P. Siddhamshetty, Z. Zhang et al., "Impact of proppant pumping schedule on well production for slickwater fracturing," *SPE Journal*, vol. 26, no. 1, pp. 342–358, 2021.

[10] C. H. Yew and X. Weng, *Mechanics of hydraulic fracturing, second edition*, Gulf Professional Publishing, 2014.

[11] B. Yan, L. Mi, Z. Chai, Y. Wang, and J. E. Killough, "An enhanced discrete fracture network model for multiphase flow in fractured reservoirs," *Journal of Petroleum Science and Engineering*, vol. 161, pp. 667–682, 2018.

[12] D. R. Harp, D. O'Malley, B. Yan, and R. Pawar, "On the feasibility of using physics-informed machine learning for underground reservoir pressure management," *Expert Systems with Applications*, vol. 178, article 115006, 2021.

[13] X. Guo, Y. Wang, and J. Killough, "The application of static load balancers in parallel compositional reservoir simulation on distributed memory system," *Journal of Natural Gas Science and Engineering*, vol. 28, pp. 447–460, 2016.

[14] B. Yan, L. Mi, Y. Wang, H. Tang, C. An, and J. E. Killough, "Multi-porosity multi-physics compositional simulation for gas storage and transport in highly heterogeneous shales," *Journal of Petroleum Science and Engineering*, vol. 160, pp. 498–509, 2018.

[15] X. Guo, Y. Jin, J. Zi, and B. Lin, "Numerical investigation of the gas production efficiency and induced geomechanical responses in marine methane hydrate-bearing sediments exploited by depressurization through hydraulic fractures," *Energy & Fuels*, vol. 35, no. 22, pp. 18441–18458, 2021.

[16] J. Geertsma and F. De Klerk, "A rapid method of predicting width and extent of hydraulically induced fractures," *Journal of Petroleum Technology*, vol. 21, no. 12, pp. 1571–1581, 1969.

[17] S. A. Khristianovich and Y. P. Zheltov, "Formation of vertical fractures by means of highly viscous liquid," in *4th world petroleum congress*, Rome, Italy, 1955.

[18] R. P. Nordgren, "Propagation of a vertical hydraulic fracture," *Society of Petroleum Engineers Journal*, vol. 12, no. 4, pp. 306–314, 1972.

[19] T. K. Perkins and L. R. Kern, "Widths of hydraulic fractures," *Journal of Petroleum Technology*, vol. 13, no. 9, pp. 937–949, 1961.

[20] R. L. Fung, S. Vilayakumar, and D. E. Cormack, "Calculation of vertical fracture containment in layered formations," *SPE Formation Evaluation*, vol. 2, no. 4, pp. 518–522, 1987.

[21] J. Rungamornrat, M. F. Wheeler, and M. E. Mear, "Coupling of fracture/non-newtonian flow for simulating nonplanar evolution of hydraulic fractures," in *SPE annual technical conference and exhibition*, Society of Petroleum Engineers, Dallas, Texas, USA, 2005.

[22] A. Settari and M. P. Cleary, "Development and testing of a pseudo-three-dimensional model of hydraulic fracture geometry," *SPE Production Engineering*, vol. 1, no. 6, pp. 449–466, 1986.

[23] K. Wu and J. E. Olson, "Simultaneous multifracture treatments: fully coupled fluid flow and fracture mechanics for horizontal wells," *SPE Journal*, vol. 20, no. 2, pp. 337–346, 2015.

[24] X. Guo, J. Ma, S. Wang, T. Zhu, and Y. Jin, "Modeling interwell interference: a study of the effects of parent well depletion on asymmetric fracture propagation in child wells," in *Paper presented at the SPE/IATMI Asia Pacific oil & gas Conference and exhibition*, Bali, Indonesia, 2019.

[25] N. Roussel and S. Mukul, "Strategies to minimize frac spacing and stimulate natural fractures in horizontal completions," in *SPE annual technical conference and exhibition*, Denver, Colorado, USA, 2011.

[26] X. Guo, W. Kan, and J. Killough, "Investigation of production-induced stress changes for infill-well stimulation in Eagle Ford Shale," *SPE Journal*, vol. 23, no. 4, pp. 1372–1388, 2018.

[27] X. Guo, K. Wu, C. An, J. Tang, and J. Killough, "Numerical investigation of effects of subsequent parent-well injection on interwell fracturing interference using reservoir-geomechanics-fracturing modeling," *SPE Journal*, vol. 24, no. 4, pp. 1884–1902, 2019.

[28] B. Hou, Z. Chang, W. Fu, Y. Muhadasi, and M. Chen, "Fracture initiation and propagation in a deep shale gas reservoir subject to an alternating-fluid-injection hydraulic-fracturing treatment," *SPE Journal*, vol. 24, no. 4, pp. 1839–1855, 2019.

[29] G. Xie, L. Luo, X. Liu, L. Liang, W. Jiang, and J. Chang, "Predicting the shape of hydraulic fracture of shale gas horizontal well in Sichuan with log data," *Well Logging Technology*, vol. 41, no. 5, pp. 590–595, 2017.

[30] R. Wen, X. Yang, D. Liu, X. Zheng, and Y. Xu, "Comprehensive evaluation of fracture distribution by near-and far-well logging and monitoring technology," *Well logging Technology*, vol. 43, no. 5, pp. 531–535, 2019.

[31] J. Zhang, H. Song, W. Zhu, and J. Wang, *Liquid transport through nanoscale porous media with strong wettability*, Transport in Porous Media, 2021.

[32] W. Yu and K. Sepehrnoori, "Simulation of gas desorption and geomechanics effects for unconventional gas reservoirs," *Fuel*, vol. 116, pp. 455–464, 2014.

[33] X. Guo, H. Song, K. Wu, and J. Killough, "Pressure characteristics and performance of multi-stage fractured horizontal well in shale gas reservoirs with coupled flow and geomechanics," *Journal of Petroleum Science and Engineering*, vol. 163, pp. 1–15, 2018.

[34] M. Moradi, A. Shamloo, and A. D. Dezfuli, "A sequential implicit discrete fracture model for three-dimensional coupled flow-geomechanics problems in naturally fractured porous media," *Journal of Petroleum Science and Engineering*, vol. 150, pp. 312–322, 2017.

[35] R. Zimmerman and Bodvarsson, "Hydraulic conductivity of rock fractures," *Transport in Porous Media*, vol. 23, no. 1, pp. 1–30, 1996.

[36] X. Shi, Y. Qin, H. Xu et al., "Numerical simulation of hydraulic fracture propagation in conglomerate reservoirs," *Engineering Fracture Mechanics*, vol. 248, article 107738, 2021.

[37] D. Zhi, X. Guo, W. Wang et al., "Fracturing and production analysis of the efficacy of hydraulic fracture stage reduction in the improvement of cost-effectiveness in shale oil development: a case study of Jimsar shale oil, China," *Energy Science & Engineering*, vol. 9, no. 9, pp. 1337–1348, 2021.

[38] A. Chen, X. Guo, H. Yu, L. Huang, S. Shi, and N. Cheng, "A parametric study of hydraulic fracturing interference between fracture clusters and stages based on numerical modeling," *Energy Exploration & Exploitation*, vol. 39, no. 1, pp. 65–85, 2021.

Hindawi
Geofluids
Volume 2021, Article ID 1194186, 12 pages
https://doi.org/10.1155/2021/1194186

Research Article

Intelligent Microfluidics Research on Relative Permeability Measurement and Prediction of Two-Phase Flow in Micropores

Hongqing Song,[1,2] **Changchun Liu**,[1] **Junming Lao**,[1,2] **Jiulong Wang**,[2,3] **Shuyi Du**,[1,2] **and Mingxu Yu**[4]

[1]*School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[2]*National and Local Joint Engineering Laboratory of Big Data Analysis and Computing Technology, Beijing 100190, China*
[3]*Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China*
[4]*Beilong Zeda (Beijing) Data Technology Co., Ltd., Beijing 100190, China*

Correspondence should be addressed to Hongqing Song; songhongqing@ustb.edu.cn

Received 26 October 2021; Revised 18 November 2021; Accepted 22 November 2021; Published 11 December 2021

Academic Editor: Bicheng Yan

Relative permeability is a key index in resource exploitation, energy development, environmental monitoring, and other fields. However, the current determination methods of relative permeability are inefficient and invisible without considering wetting order and pore structure characteristics either. In this study, microfluidic experiments were designed for figuring out key factors impacting on the two-phase relative permeability. The optimized intelligent image recognition was established for saturation extraction. The deep learning was conducted for the prediction of two-phase permeability based on the inputs from microfluidic experiments and image recognition and optimized. Results revealed that phase saturation, wetting order, and pore topology were the key factors influencing the two-phase relative permeability, with the importance of 38.22%, 34.84%, and 26.94%, respectively. The deep learning-based relative permeability model performed well, with MSE < 0.05 and operational efficiency of 3 ms/epoch. Aiming at relative permeability model optimization, on the one hand, the dividing ratio of training set and testing set for flooding phase relative permeability prediction achieved the highest prediction accuracy at 7 : 3, while that for displaced phase was 6 : 4. On the other hand, tanh() activation function performed 40% more accurate than the sigmoid() activation function.

1. Introduction

The relative permeability is a crucial parameter reflecting reservoir rock allocation properties and an indispensable index revealing the characteristics of fluid flow and distribution [1, 2]. Moreover, relative permeability is a key index in resource exploitation, energy development, environmental monitoring, and other fields [3, 4]. The greater the relative permeability of certain fluid in a particular reservoir means that the weaker the reservoir resistance to the fluid, the stronger the mobility of fluid in pores and the more clear the distribution in the reservoir. For instance, in groundwater resource mining and storage, water relative permeability is applied to guide the location of mining sites and to evaluate the risk of groundwater leakage in the reservoir. In oil and gas reservoir development, the relative permeability of crude oil or natural gas is adopted to evaluate the significance and benefits of the reservoir water injection development. In soil environmental monitoring, relative permeability is a key indicator to determine the characteristics of sewage diffusion and transport. The construction of a relative permeability model with high efficiency, high accuracy, high robustness, and extensive applicable scenarios is of great significance to effectively evaluate resource mining efficiency, improve energy recovery, and optimize environmental monitoring and testing [5–8].

The construction of relative permeability models is based on data feedback from adequate relative permeability experiments. However, the current relative permeability determination experiments are still based on the steady-state

multiphase core displacement, which has some shortcomings [9–11]. First, a single core cannot reflect the pore characteristics of the entire reservoir, while equivalent conditional experiments on multiple cores imply significant consumption of time and cost. In addition, the experiments will change the core wetting history or even destroy the internal pore structure of the core, which is not beneficial for improving the experimental accuracy by weakening the stochastic error through multiple experiments. Finally, the experiments require the experimenter to keep recording the flow of each phase fluid and indirectly obtain the saturation at each time, which limits the experimenter resulting in low experimental efficiency [12–15].

Besides, extensive experiments have shown that relative permeability is a function of saturation. Nevertheless, the factors such as wetting order and pore structure characteristics also have negligible effect on relative permeability. Most of the relative permeability models proposed by academia lack the consideration of wetting order and pore structure characteristics. This is largely due to the invisibility of internal flow during conventional steady-state core displacement experiments, leading to the wetting order and the pore structure features' unattainability. In addition, the introduction of wetting order and pore structure characteristics based on the phase saturation will significantly increase the nonlinearity and model complexity of the relative permeability model, which increases the difficulty to build a reliable relative permeability model [16–19].

The nonsteady two-phase microfluidic experiments are characterized by whole-process visualization, automated parameter recording, and the experimental setup closer to the two-phase displacement flow in engineering practice. It is a reliable substitution experiment to obtain the basic data of constructing the relative permeability model. During the microfluidic experiments, the electron ocular records the saturation over time, producing a large number of image data [20–23]. According to the microfluidic experiments, general cognition of impacts from various parameters on relative permeability could be obtained, while the quantitative contribution of each parameter could be clarified basing on the Sobol sensitivity analysis model. Furthermore, by introducing the image recognition algorithm, the intelligent identification, storage and analysis of saturation, wetting order, and pore structure characteristics could be achieved with less time cost comparing to conventional methods [24–27]. Finally, based on the saturation, wetting order, and pore structure data, relative permeability intelligent models are constructed by adopting deep learning algorithm, especially the deep neural network (DNN) [28, 29].

In this study, microfluidic experiments were designed for figuring out key factors impacting on the two-phase relative permeability. The optimized intelligent image recognition was established for saturation extraction. The deep learning was conducted for the prediction of two-phase permeability based on the inputs from microfluidic experiments and image recognition. The microfluidic experiments and deep learning model introduced in this study are of great significance for efficient and reliable research on resource exploitation, energy development, and environmental monitoring.

2. Methodology

The research route for intelligent model-based image recognition and permeability prediction of two-phase flow in micropores is shown in Figure 1. First, experimental images were obtained by the designed microfluidic two-phase flow experiments. Subsequently, the image data was preprocessed via applying the image enhancement algorithm specifically the global equalization. Secondly, the preprocessed image data was input into the convolution neural network (CNN); then, the water and oil saturation in the images was extracted and identified. Output from the CNN, the oil and water saturation together with the wetting order and pore diversity obtained from the microfluidic two-phase flow experiments was input to the deep neural network (DNN) model. Finally, the intelligent prediction model of relative permeability based on the DNN was established and optimized [30–34].

2.1. Physical Equations. In this study, the two-phase flow in micropores specified as the process of oil flooding by water was investigated. The flowing process could be described as the Darcy porous flow [31, 35]. According to the Darcy law, the relationship between the permeability and the pressure and flow, as shown in the following equation [31, 36, 37]:

$$Q = \frac{KA}{\mu}\nabla P, \tag{1}$$

where Q is flow rate, m^3/s; K is the permeability of the porous media, m^2; A is the outlet area, m^2; μ is the viscosity of liquid, Pa·s; and ∇P is the pressure gradient, Pa/m.

Accordingly, the permeability is defined as the format as shown in the following equation:

$$K = \frac{u\mu}{\nabla P}, \tag{2}$$

where u is the liquid velocity, m/s.

Furthermore, the relative permeability of a certain liquid is defined as the format as shown in the following equation:

$$K_{ri} = \frac{u_i\mu_i}{K\nabla P_i}, \tag{3}$$

where K_{ri} is the relative permeability of liquid i, dimensionless; u_i is the velocity of liquid i during multiphase flow, m/s; μ_i is the viscosity of liquid i, Pa·s; and ∇P_i is the pressure gradient on liquid i, Pa/m.

The pressure gradient on liquid i is correlated to the capillary force on which and the absolute driving pressure gradient, as shown in the following equation:

$$\nabla P_i = \frac{P_{c,i} + P_a}{L}, \tag{4}$$

where $P_{c,i}$ is the capillary force on liquid i, Pa; P_a is the absolute driving pressure, Pa; and L is the length of the porous media, m.

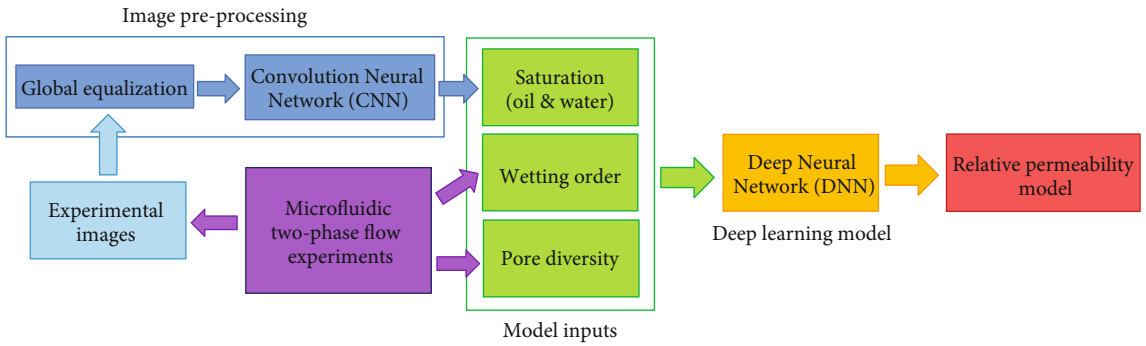


Figure 1: Research route for intelligent model-based image recognition and permeability prediction of two-phase flow in micropores.

2.2. Microfluidic Two-Phase Experiments. The microfluidic two-phase experiments were conducted at a laboratory at room temperature of 24.3°C with 70% relative humidity and reaching stable thermal balance indoors. The atmospheric pressure of the laboratory site was 99.2 kPa. Water used in the experiments was mineral water. Oil used in the experiments was light crude oil. Detailed properties of fluids used in experiments are shown in Table 1. While studying the impact of wetting order on two-phase relative permeability, Light oil $2^{\#}$ was used as flooding phase to flood the mineral water since the microfluidic chips were hydrophobic. Meanwhile, the reason why Light oil $2^{\#}$ as flooding phase was adopted was to maintain the capillary force on either flooding or displaced phase as constant variable. Based on the properties of the experimental fluid, the three-phase porous flow process performed in this research was considered an isothermal incompressible two-dimensional flow [21, 38–40].

The overview of the flooding experiment and data measurement and record is presented in Figure 2. In particular, the driving module, pressure monitoring module, experiment visualization module, and data record module were established. The adopted equipment and materials included two constant-speed microinjection pumps for liquid injection, medical plastic syringes, alloy four-way valve, 16 kPa capsule pressure gauge, PE lines, microfluidic chips, optical microscope with electron lens, waste liquid collection test tube, mobile workstation, and experimental recording camera [41–44]. The left and lower inlets of the four-way valve were connected to the water pump and nitrogen pump, respectively, while the outlet at the top were connected to the pressure gauge and the right outlet connected to the inlet of the microfluidic chip. Before the experiment began, the microfluidic chip was saturated by the displaced phase. The injection rate was also set in the microinjection pump, which corresponded to the injection rate at the inlet of microfluidic chip [36, 45–47].

To quantitatively analyze the contribution of each parameter to relative permeability, the Sobol sensitivity analysis method was adopted. The Sobol method is based on the idea of model decomposition, yielding the sensitivity of parameters 1, 2, and higher, respectively. Usually, 1 sensitivity reflects the main effect of parameters, while 2 and higher sensitivity consider more. Compared with other sensitivity analysis methods, the Sobol method has a relatively stable sampling method, which can grade the sensitivity to the contribution proportion of the output variance through parameters and is a more efficient method to quantitatively identify the sensitivity of different parameters. The specific model of Sobol sensitivity analysis is shown in equations (5)-(9).

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \cdots + \sum f_{1,2,\cdots,n}(x_1, x_2, \cdots x_n), \tag{5}$$

$$D = \sum_i D_i + \sum_{i<j} D_{ij} + \cdots + \sum D_{1,2,\cdots,n}, \tag{6}$$

$$S_{1,2,\cdots,n} = \frac{D_{1,2,\cdots,n}}{D}, \tag{7}$$

$$1 = \sum_i S_i + \sum_{i<j} S_{ij} + \cdots + \sum S_{1,2,\cdots,n}, \tag{8}$$

$$S_{Ti} = 1 - \sum S_{\sim i}. \tag{9}$$

2.3. Image Enhancement and Recognition. Before the saturation recognition and extraction, the images require to be enhanced for highlighting the features while weakening the white noise inside. In particular, smoothing is required to eliminate noise interference in the image or to enhance contrast and sometimes to emphasize the edges and details of the image. For oil-water two-phase microfluidic experimental images, the grayscale distribution of the image is relatively concentrated, resulting in the segmentation difficulty of two-phase saturation. Histogram equalization is achieved by adjusting the gray order distribution of the image so that the distribution on the 0~255 gray order is more balanced, which is an effective method to improve the contrast of the image and simplify the segmentation of oil and water saturation. Shown in Figure 3 is a schematic diagram of the histogram equalization principle. Generally, images obtained from oil-water two-phase microfluidic experiments are suitable for adopting the histogram equalization method to enhance image details.

After the image enhancement is done by histogram equalization, the next step is the saturation recognition and extraction. The convolution neural network (CNN) is adopted in this study for oil and water saturation

Table 1: Phase parameters in the injection experiment (at 24.3°C, 99.2 kPa).

Phase type	Material	Density (kg/m^3)	Viscosity (mPa·s)	Compressibility (1/Pa)
Displaced phase	Light oil 1#	855.297	4.160	5.268×10^{-10}
Flooding/displaced phase	Mineral water	997.323	0.917	4.504×10^{-10}
Flooding phase	Light oil 2#	804.865	1.617	5.602×10^{-10}

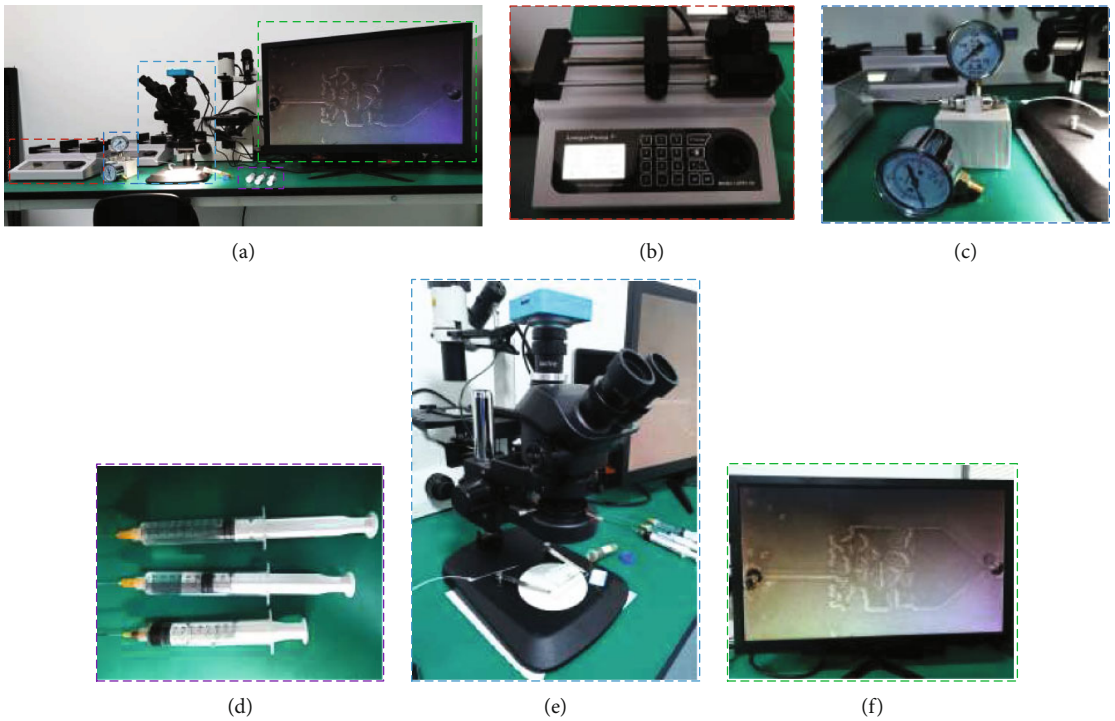


Figure 2: The microfluidic experiment scheme of water-oil flow: (a) overview of experimental equipment; (b) microinjection pump; (c) capsule pressure gauge (16 kPa) and alloy four-way valve; (d) medical plastic syringes; (e) optical microscope with electron lens and microfluidic chip; (f) screen and data record.

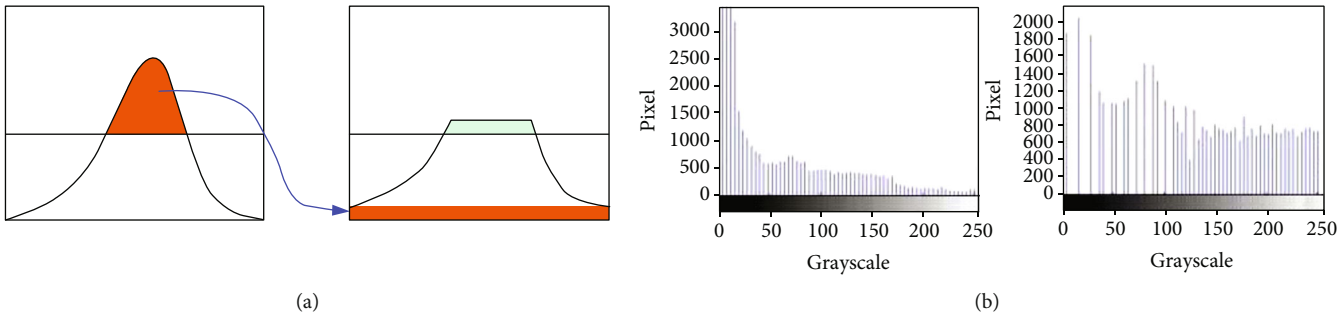


Figure 3: Schematic diagram of the histogram equalization principle. (a) The partial pixels at concentrated grayscale are equally converted to which at diverse grayscale ranging from 0 to 255. (b) The practical processing results based on the principle demonstrated in (a).

recognition and extraction from microfluidic experimental images. The schematic diagram of the CNN is shown in Figure 4. The CNN consists of the following five parts: the input layer, convolution layer, pooling layer, fully connected layer, and the output layer. The key advantage of this method is that it could extract local data features through convolution and pooling operations. Besides, as a supervised intelligent method, the CNN is with high reliability and robustness.

2.4. Deep Learning Model. A narrow definition of deep learning is the neural network with multilayers. Accordingly, the CNN mentioned above and the DNN which is introduced later are both classified as the deep learning model. In this

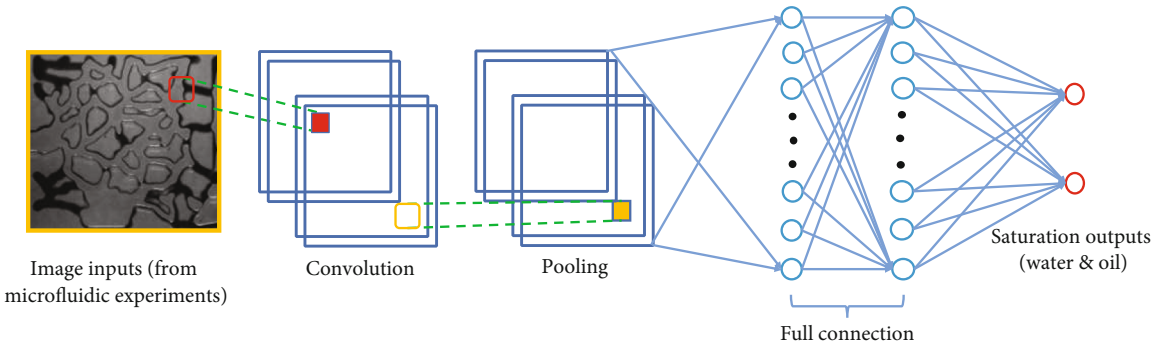


Figure 4: Schematic diagram of the convolution neural network (CNN) for saturation extraction.

study, the deep neural network (DNN) model was proposed for effectively extracting the linear and nonlinear characteristics of the data. The schematic diagram of the DNN for two-phase relative permeability prediction is shown in Figure 5, and the fundamental theories of the DNN are described from equation (10) to equation (14). Specifically, equation (10) to equation (12), respectively, show the operational criterion between the input layer and the first hidden layer, present hidden layer and next hidden layer, and the last hidden layer and output layer, while equation (13) to equation (14) show the activation functions of tanh() and sigmoid(). The parameter matrix of A and b is determined after the DNN is well-trained and validated [48–52].

$$\text{Activation}\left(A^1X + b^1\right) = H^1, \tag{10}$$

$$\text{Activation}\left(A^{n+1}H^n + b^{n+1}\right) = H^{n+1}, \tag{11}$$

$$\text{Activation}\left(A^{N+1}H^N + b^{N+1}\right) = Y, \tag{12}$$

$$\tanh\,(x) = \frac{\exp\,(x) - \exp\,(-x)}{\exp\,(x) + \exp\,(-x)}, \tag{13}$$

$$\text{sigmoid}(x) = \frac{1}{1 + \exp\,(-x)}, \tag{14}$$

where X is the input matrix of this study consisting of saturation, wetting order, and pore diversity; H is the hidden layer matrix; A is the weight matrix while b is the bias matrix, both called the parameter matrix; n is the layer order of the hidden layer; N is the amount of hidden layers; and Y is the output matrix, corresponding to the relative permeability in this study.

In order to control the error of deep learning models during the training learning process, loss function is usually used to reflect the regression training effect of the model. The commonly used loss functions include mean square error (MSE), as shown in equation (15). According to the law of gradient descent and back-propagation of regression error, the parameter matrix A and b are adjusted and a model with high accuracy is finally obtained.

$$\text{MSE} = \frac{1}{M}\sum_{i=1}^{M}\left(y_i^{\text{or}} - y_i^{\text{re}}\right)^2, \tag{15}$$

where M is the total number of the output variable which is the product of variable dimensions and feature numbers; y_i represents the ith output variable, while the superscript or and re represent original value and regressed value, respectively.

For neural networks with multiple input variables, different variables may have different units and values. To eliminate this effect and maintain the relative relationship between the values of the same variables, it is necessary to normalize the input dataset via the following equation:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{16}$$

where X' is the normalized input matrix; X is the original input matrix; and $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the input variables.

3. Results and Discussion

3.1. Characterizations of Two-Phase Flow in Micropores. In this study, 12 sets of microfluidic experiments with different pore structure characteristics, wetting order, and injection rate conditions were designed. The experimental results of microfluidic two-phase flow experiments are shown in Figure 6. Figures 6(a)–6(l), respectively, show experimental images at initial injection, during injection, and injection completion of each group of experiment. Among them, Figures 6(a)–6(c) show experiments with the coefficient of pore diversity of 0.872. The injection rate was 50, 100, and 150 μL/min, respectively. The wetting order was mineral water flooding. It should be noted that when mineral water flooded, the Light crude oil $1^{\#}$ was displaced while when Light crude oil $2^{\#}$ flooded, the mineral water was displaced. Similarly, Figures 4(d)–4(f) show experiments with the coefficient of pore diversity of 3.248. The setting of injection rate and wetting order was consistent with the experiments shown from Figures 6(a)–6(c). Figures 4(g)–4(i) show experiments with the coefficient of pore diversity of 6.965. The setting of injection rate and wetting order was consistent with the experiments shown from Figures 6(a)–6(d).

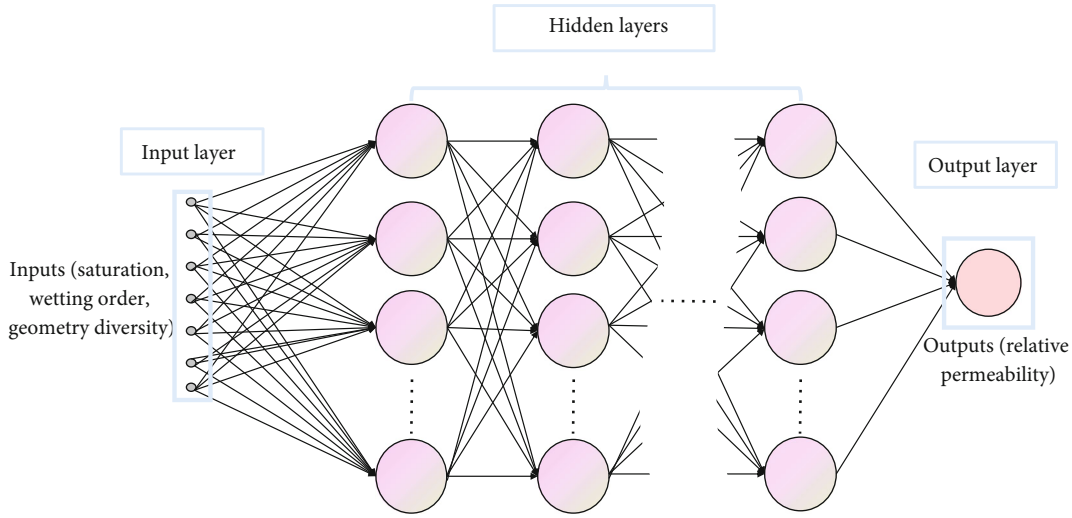


Figure 5: Schematic diagram of the deep neural network (DNN) for two-phase relative permeability prediction.

Figures 4(j)–4(l) show experiments with the coefficient of pore diversity of 0.872, 3.248, and 6.965, respectively. The injection rates were all 100 μL/min. The wetting order was Light crude oil $2^{\#}$ flooding.

According to Figure 6, it could be found that the in addition to the saturation, the wetting order and pore diversity of pores did significantly impact on the two-phase flow in micropores. Hence, the saturation, wetting order, and pore diversity should be contained in the relative permeability model comprehensively.

Though the impact of saturation on relative permeability is hardly possible to visually observe, the effects of wetting order and pore diversity are significantly visible. In particular, at the mineral water flooding condition, it could be seen that the higher pore diversity was conducive to enhance the displacement of displaced phase generally. It is worth noting that the relationship between injection rate and saturation is nonlinear, while at the medium injection rate, 100 μL/min, the most residual displaced phase was left in pores. On the contrary, at the light crude oil flooding condition, the most complete flooding was done at the lowest pore diversity.

The reasons why such phenomena occurred are that on the one hand, the capillary effect is evident during the two-phase flow process in micropores. The pores in microfluidic chips were hydrophobic determined by the materials. When the wetting order was set to be mineral flooding, the mineral water occupied the large pores initially and mainly and seldom permitted into fine pores due to the resistance conducted by capillary force on mineral water. Conversely, when the wetting order was set to be light crude oil flooding, the light crude oil occupied the fine pores initially and immediately and through which the light crude oil escaped instead of occupying the large pores. On the other hand, the saturation of both flooding and displaced phases was governed by Darcy's law of porous flow and capillary effect. Therefore, at a low injection rate particularly 50 μL/min in this study, the capillary effect played the domain role that large pores were more completely occupied. At a high injection rate particularly 150 μL/min in this study, the Darcy law of porous flow performed more significant that flooding phase could permit into partial fine pores with high flooding pressure. Eventually, at the medium injection rate, particularly 100 μL/min in this study, both capillary effect and Darcy's law of porous flow failed to take advantage of themselves resulting in the most suboptimal flooding performance.

According to the microfluidic two-phase flow experiments, the two-phase flow saturation-relative permeability curves were plotted as shown in Figure 7. Specifically, curves from Figure 7(a)–7(l) are corresponding to the experimental results obtained from Figures 6(a)–6(l).

According to Figure 7, it could be found that at the wetting order of mineral water flooding, the equivalent points of relative permeability were all in the interval where the displaced phase saturation was greater than or equal to 0.8. However, at the wetting order of light crude oil flooding, there was no equivalent point of the two phases, while the two-relative permeability interval is shorter than the wetting order of mineral water flooding.

3.2. Sensitivity Analysis of Variable Effects on Relative Permeability. The Sobol full-order weight model was applied to perform the sensitivity analysis of the saturation, wetting order, and pore diversity impact on relative permeability as shown in Figure 8. Notably, due to the saturation of the flooding and displaced phase satisfy the constraints of 1, the present study only considered the displaced phase saturation as the input variable for the Sobol full-order weight model. According to the results of the analysis shown in Figure 8, the saturation has the greatest effect on both the relative permeability of flooding phase and displaced phase. The wetting order was the second key impact factor of the relative permeability followed the saturation. Although the effect of the pore structure characteristics (coefficient of diversity) on the relative permeability was minimal among the three variables, the contribution difference from the saturation was not higher than 50%. Overall, phase

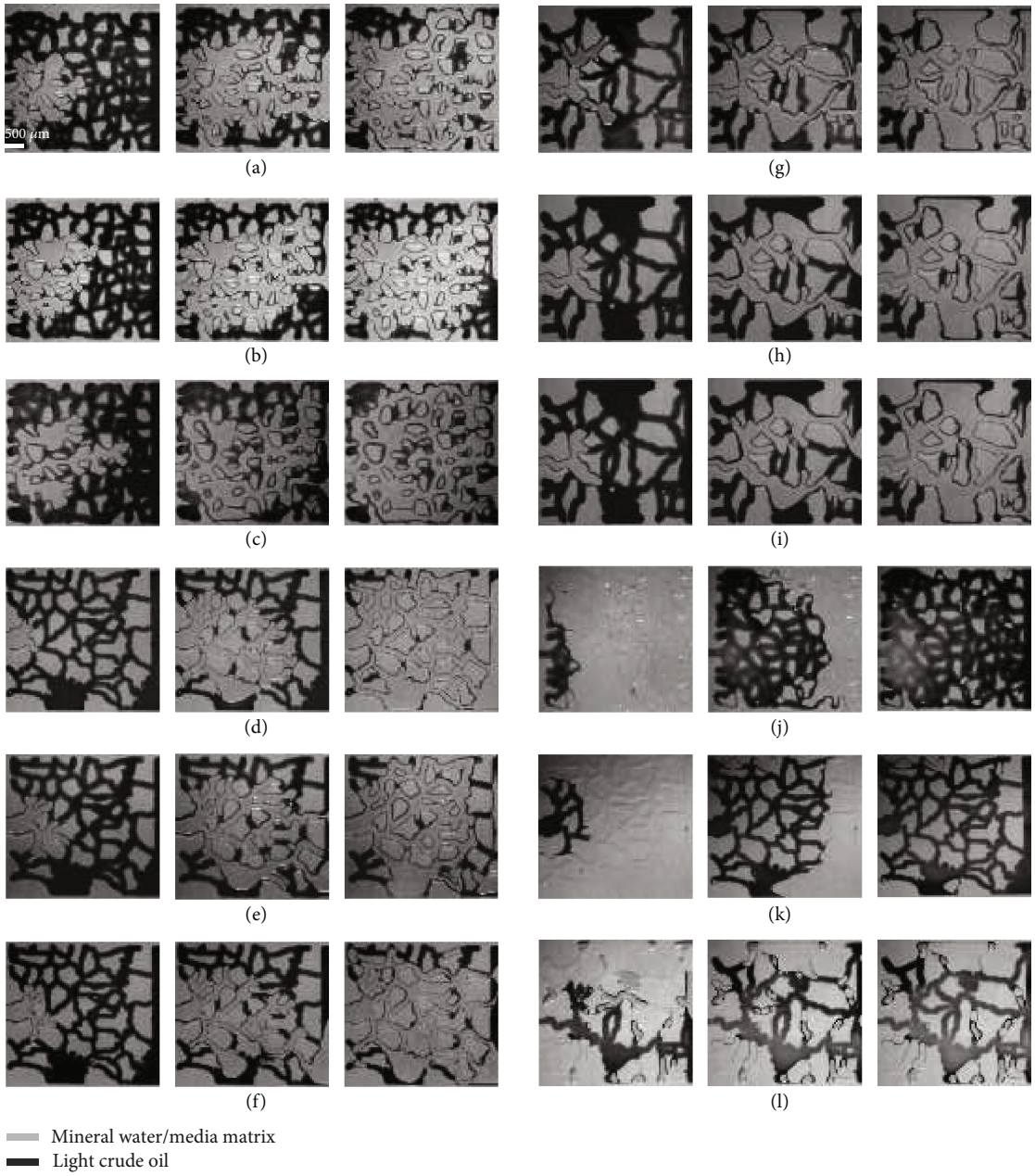


Figure 6: Experimental results of microfluidic two-phase flow experiments.

saturation, wetting order, and pore structure characteristics could be considered key factors in the two-phase flow of oil and water in micropores.

3.3. Performance of Intelligent Relative Permeability Model. The training set of the deep learning-based intelligent relative permeability model was initially divided by the ratio of 8 : 2. The activation function adopted to the model was the tanh() function, and 5 hidden layers were set. The training process and validation results of the intelligent relative permeability model of flooding phase are shown in Figure 9(a), and those of the displaced phase are shown in Figure 9(b).

According to Figures 9(a) and 9(b), it can be found that the training process and validation results of the intelligent relative permeability model of both the flooding phase and displaced phase performed well. Furthermore, Figures 10(a) and 10(b) reveal the specific loss (mean square error, MSE was adopted in this study) at each epoch corresponding to the training process of the intelligent relative permeability model of the flooding phase and displaced phase, respectively. It could be concluded that the intelligent relative permeability models proposed in this study are reliable and accurate.

3.4. Optimization of Intelligent Relative Permeability Model. The intelligent relative permeability model was first optimized by adjusting the dividing ratio of the training set and the test set. As could be seen from the optimization

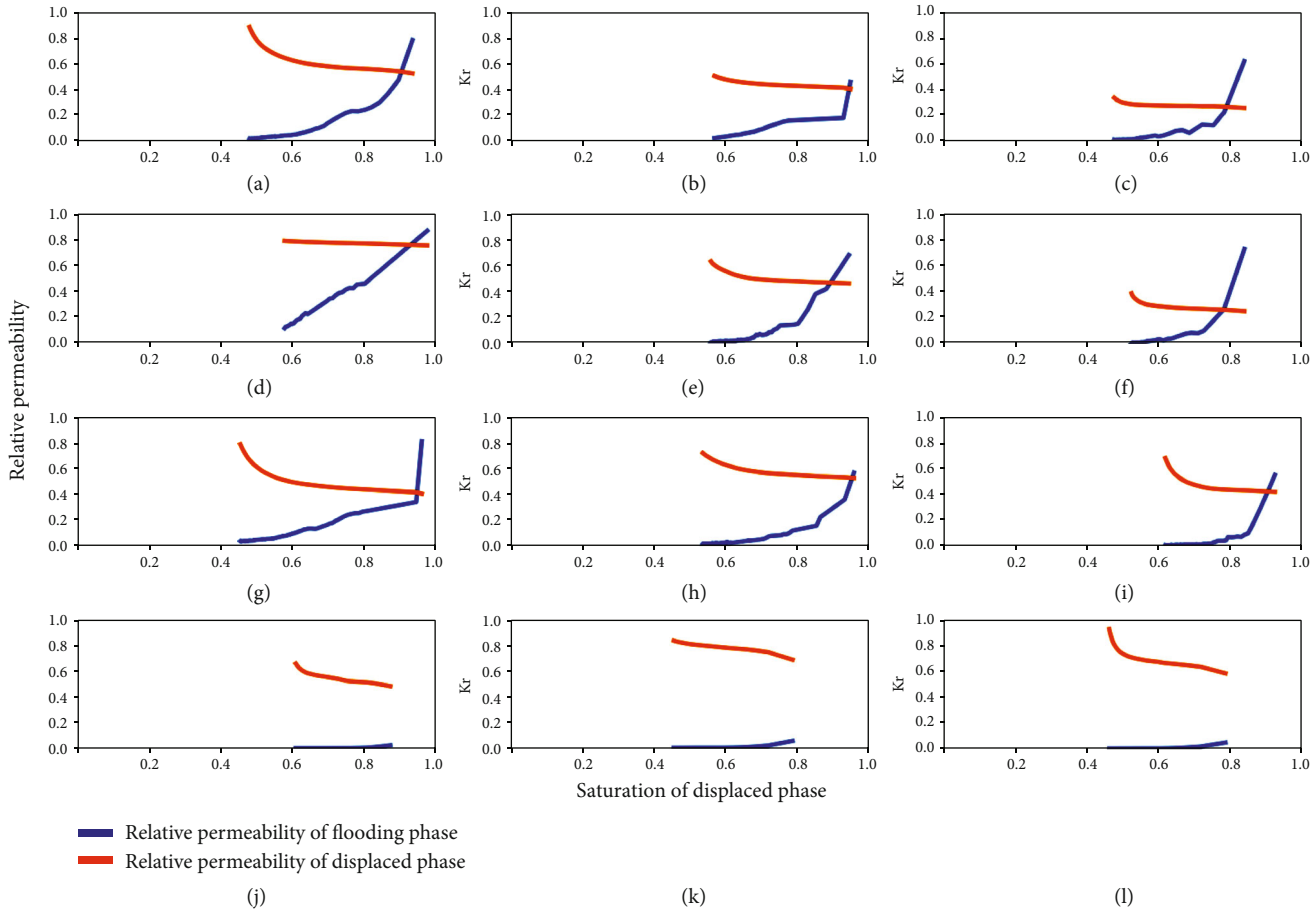


Figure 7: Two-phase relative permeability obtained from microfluidic experiments.

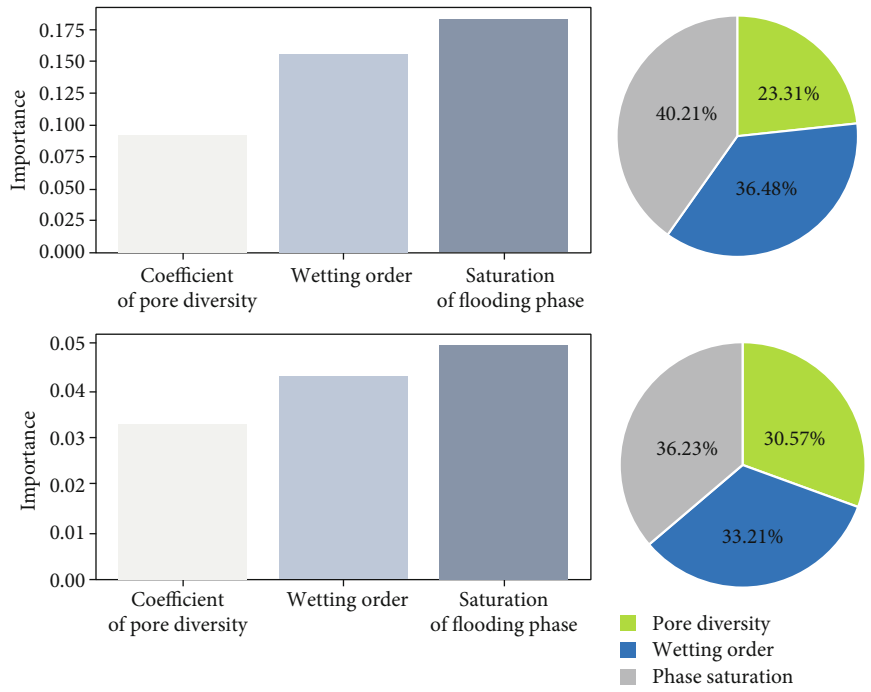


Figure 8: Sobol sensitivity analysis of saturation, wetting order, and pore diversity impact on relative permeability.

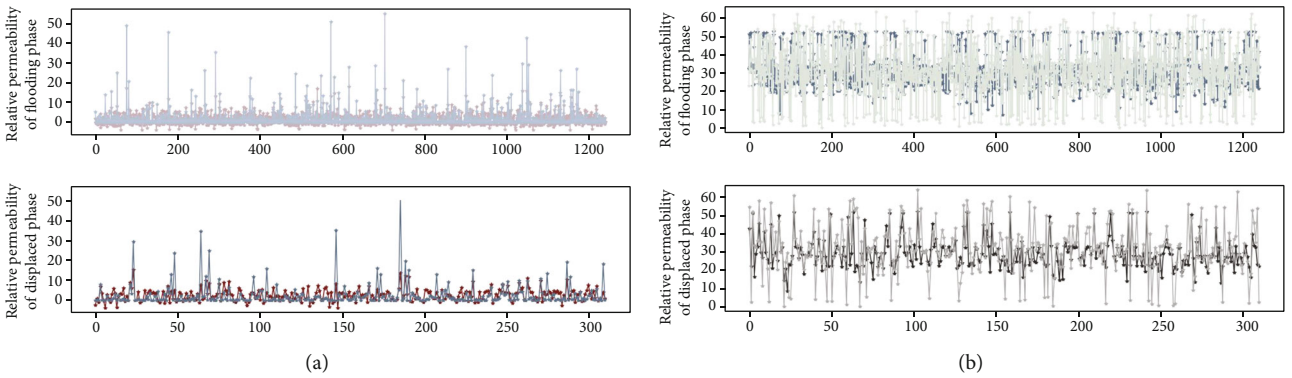


Figure 9: The training and testing results of the DNN: (a) the training process and validation results of the intelligent relative permeability model of the flooding phase; (b) the training process and validation results of the intelligent relative permeability model of the displaced phase.

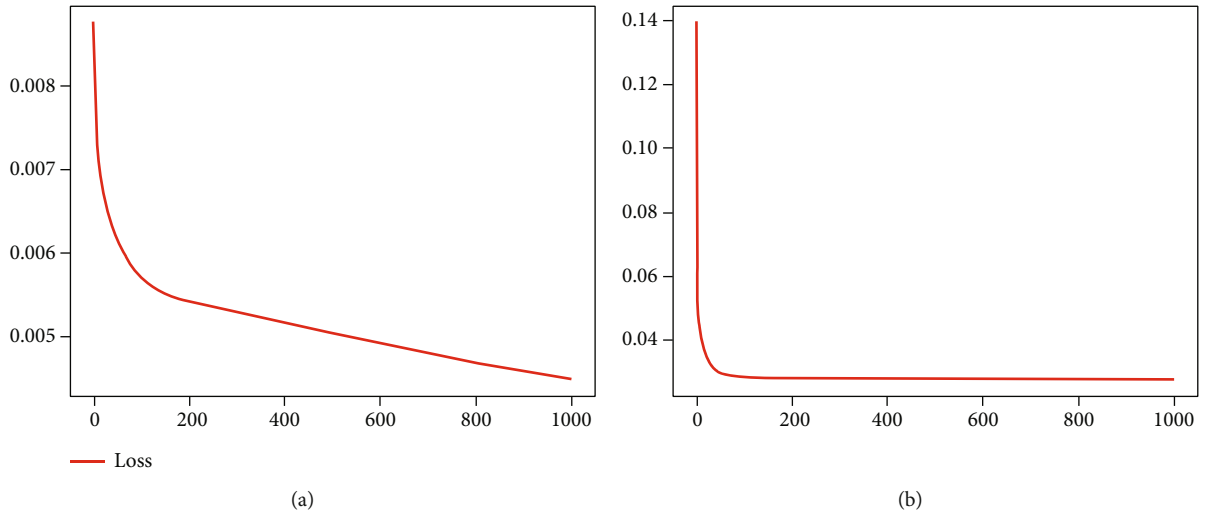


Figure 10: The specific loss (MSE) at each epoch corresponding to the training process of the intelligent relative permeability model of the flooding phase and displaced phase, respectively. (a) The specific loss (MSE) at each epoch corresponding to the training process of the intelligent relative permeability model of the flooding phase. (b) The specific loss (MSE) at each epoch corresponding to the training process of the intelligent relative permeability model of the displaced phase.

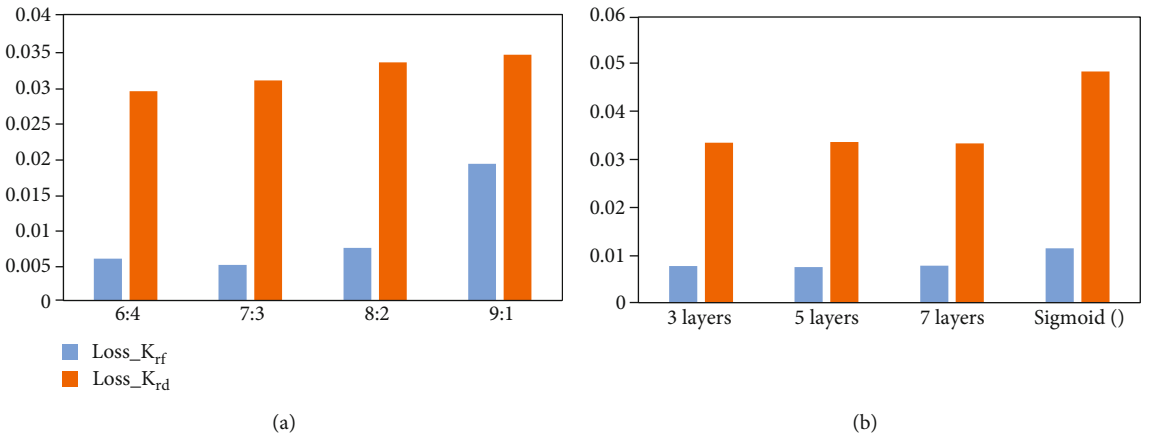


Figure 11: Optimization results of two-phase seepage intelligent model of oil and water. (a) The optimization results of intelligent relative permeability model by adjusting the dividing ratio of the training set and the test set. (b) The optimization results of intelligent relative permeability model by adjusting the number of hidden layers of the deep neural network as well as the activation function.

results in Figure 11(a), the lowest validation loss (MSE) for relative permeability of the flooding phase was at the ratio of 7 : 3 while that for the displaced phase was at the ratio of 6 : 4 compared to the ratio of 8 : 2 for the initial model. Moreover, as the proportion of the training set increased along with the test set scale decreased, the verification loss of the displaced phase relative permeability continuously improved, while that of the flooding phase showed a parabolic trend of falling first and then rising.

Subsequently, the intelligent relative permeability model was optimized by adjusting the number of hidden layers of the deep neural network as well as the activation function. The optimization results in Figure 11(b) show that the adoption of 3 or 7 hidden layers does not bring an obvious optimization effect for the model compared to the initial neural network structure of 5 hidden layers. However, it is worth noting that the tanh() activation function performed 40% more accurate than the sigmoid() activation function while adopting the equivalent 5 hidden layers.

4. Conclusion

Relative permeability is a key index in resource exploitation, energy development, environmental monitoring, and other fields. However, the current determination methods of relative permeability are inefficient and invisible without considering wetting order and pore structure characteristics either. In this study, microfluidic experiments were designed for figuring out key factors impacting on the two-phase relative permeability. The optimized intelligent image recognition was established for saturation extraction. The deep learning was conducted for the prediction of two-phase permeability based on the inputs from microfluidic experiments and image recognition and optimized.

Results revealed that phase saturation, wetting order, and pore topology were the key factors influencing the two-phase relative permeability, with the importance of 38.22%, 34.84%, and 26.94%, respectively. The deep learning model for relative permeability prediction performed well, with MSE < 0.05 and operational efficiency of 3 ms/epoch. Aiming at relative permeability model optimization, on the one hand, the dividing ratio of training set and testing set for flooding phase relative permeability prediction achieved the highest prediction accuracy at 7 : 3, while that for displaced phase was 6 : 4. On the other hand, tanh() activation function performed 40% more accurate than the sigmoid() activation function.

Data Availability

Data is available on request. Please contact the corresponding author for the underlying data supporting the results of the research.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 11972073).

References

[1] Y. Da Wang, R. T. Armstrong, and P. Mostaghimi, "Boosting resolution and recovering texture of 2D and 3D micro-CT images with deep learning," *Water Resources Research*, vol. 56, no. 1, 2020.

[2] J. Zhang, H. Q. Song, W. Y. Zhu, and J. L. Wang, "Liquid transport through nanoscale porous media with strong wettability," *Transport in Porous Media*, vol. 140, no. 3, pp. 697–711, 2021.

[3] F. O. Alpak, M. Araya–Polo, and K. Onyeagoro, "Simplified dynamic modeling of faulted turbidite reservoirs: a deep-learning approach to recovery-factor forecasting for exploration," *SPE Reservoir Evaluation & Engineering*, vol. 22, no. 4, pp. 1240–1255, 2019.

[4] N. J. Alqahtani, T. Chung, Y. D. Wang, R. T. Armstrong, P. Swietojanski, and P. Mostaghimi, "Flow-based characterization of digital rock images using deep learning," *SPE Journal*, vol. 26, no. 4, pp. 1800–1811, 2021.

[5] Z. Zhong, A. Y. Sun, Y. Y. Wang, and B. Ren, "Predicting field production rates for waterflooding using a machine learning-based proxy model," *Journal of Petroleum Science and Engineering*, vol. 194, article 107574, 2020.

[6] B. C. Yan, L. D. Mi, Y. H. Wang, H. W. Tang, C. An, and J. E. Killough, "Multi-porosity multi-physics compositional simulation for gas storage and transport in highly heterogeneous shales," *Journal of Petroleum Science and Engineering*, vol. 160, pp. 498–509, 2018.

[7] B. C. Yan, L. D. Mi, Z. Chai, Y. H. Wang, and J. E. Killough, "An enhanced discrete fracture network model for multiphase flow in fractured reservoirs," *Journal of Petroleum Science and Engineering*, vol. 161, pp. 667–682, 2018.

[8] O. Santamaria, S. H. Lopera, M. Riazi, M. Minale, F. B. Cortes, and C. A. Franco, "Phenomenological study of the micro- and macroscopic mechanisms during polymer flooding with SiO_2 nanoparticles," *Journal of Petroleum Science and Engineering*, vol. 198, p. 108135, 2021.

[9] P. Purswani, Z. T. Karpyn, K. Enab, Y. Xue, and X. L. Huang, "Evaluation of image segmentation techniques for image-based rock property estimation," *Journal of Petroleum Science and Engineering*, vol. 195, article 107890, 2020.

[10] A. Gerami, R. T. Armstrong, Y. Jing, F. A. Wahid, H. Arandiyan, and P. Mostaghimi, "Microscale insights into gas recovery from bright and dull bands in coal," *Journal of Petroleum Science and Engineering*, vol. 172, pp. 373–382, 2019.

[11] O. D. Arigbe, M. B. Oyeneyin, I. Arana, and M. D. Ghazi, "Real-time relative permeability prediction using deep learning," *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 2, pp. 1271–1284, 2019.

[12] J. W. Sanders, H. S. M. Chen, J. M. Johnson et al., "Synthetic generation of DSC-MRI-derived relative CBV maps from DCE MRI of brain tumors," *Magnetic Resonance in Medicine*, vol. 85, no. 1, pp. 469–479, 2021.

[13] W. S. Loos, R. Souza, L. B. Andersen, R. M. Lebel, and R. Frayne, "Extraction of a vascular function for a fully automated dynamic contrast-enhanced magnetic resonance brain

image processing pipeline," *Magnetic Resonance in Medicine*, 2021.

[14] H. Q. Song, H. H. Guo, Y. H. Wang et al., "A novel hybrid energy system for hydrogen production and storage in a depleted oil reservoir," *International Journal of Hydrogen Energy*, vol. 46, no. 34, pp. 18020–18031, 2021.

[15] S. A. Mahoney, T. E. Rufford, V. Rudolph, K. Y. Liu, S. Rodrigues, and K. M. Steel, "Creation of microchannels in Bowen Basin coals using UV laser and reactive ion etching," *International Journal of Coal Geology*, vol. 144, pp. 48–57, 2015.

[16] P. Mostaghimi, R. T. Armstrong, A. Gerami et al., "Cleat-scale characterisation of coal: an overview," *Journal of Natural Gas Science and Engineering*, vol. 39, pp. 143–160, 2017.

[17] J. C. Lv, K. P. Xue, Z. D. Zhang, Z. C. Cheng, Y. Liu, and H. L. Mu, "Pore-scale investigation of hydrate morphology evolution and seepage characteristics in hydrate bearing microfluidic chip," *Journal of Natural Gas Science and Engineering*, vol. 88, p. 103881, 2021.

[18] L. Chen, W. B. Lin, P. Chen, S. Jiang, L. Liu, and H. Y. Hu, "Porosity prediction from well logs using back propagation neural network optimized by genetic algorithm in one heterogeneous oil reservoirs of Ordos Basin, China," *Journal of Earth Science*, vol. 32, no. 4, pp. 828–838, 2021.

[19] H. Q. Song, J. J. Xu, J. Fang, Z. G. Cao, L. Z. Yang, and T. X. Li, "Potential for mine water disposal in coal seam goaf: investigation of storage coefficients in the Shendong mining area," *Journal of Cleaner Production*, vol. 244, p. 118646, 2020.

[20] C. A. Browne, A. Shih, and S. S. Datta, "Pore-scale flow characterization of polymer solutions in microfluidic porous media," *Small*, vol. 16, no. 9, p. 1903944, 2020.

[21] G. Grisanti, D. Caprini, G. Sinibaldi et al., "A microfluidic platform for cavitation-enhanced drug delivery," *Micromachines*, vol. 12, no. 6, p. 658, 2021.

[22] W. J. Yun, C. M. Ross, S. Roman, and A. R. Kovscek, "Creation of a dual-porosity and dual-depth micromodel for the study of multiphase flow in complex porous media," *Lab on a Chip*, vol. 17, no. 8, pp. 1462–1474, 2017.

[23] J. F. Wong and C. A. Simmons, "Microfluidic assay for the on-chip electrochemical measurement of cell monolayer permeability," *Lab on a Chip*, vol. 19, no. 6, pp. 1060–1070, 2019.

[24] H. Y. Wu, W. Z. Fang, Q. J. Kang, W. Q. Tao, and R. Qiao, "Predicting effective diffusivity of porous media from images by deep learning," *Scientific Reports*, vol. 9, no. 1, article 20387, 2019.

[25] Z. N. Wu, Y. H. Zhou, H. L. Wang, and Z. H. Jiang, "Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse," *Science of the Total Environment*, vol. 716, article 137077, 2020.

[26] A. Riaz, R. P. Gandhiraman, I. K. Dimov et al., "Reactive deposition of nano-films in deep polymeric microcavities," *Lab on a Chip*, vol. 12, no. 22, pp. 4877–4883, 2012.

[27] K. Osei-Bonsu, S. Khorsandi, and M. Piri, "Quantitative analysis of phase topology evolution during three-phase displacements in porous media," *Lab on a Chip*, vol. 20, no. 14, pp. 2495–2509, 2020.

[28] P. Kang, J. M. Lao, M. X. Yu, H. Q. Song, and C. Wang, "Reliable prediction on emerging energy supply for National Sustainability and stability: a case study on coal bed gas supply in China based on the dual-LSTM model," *IEEE Access*, vol. 9, pp. 100694–100707, 2021.

[29] S. Du, R. F. Wang, C. J. Wei et al., "The connectivity evaluation among wells in reservoir utilizing machine learning methods," *IEEE Access*, vol. 8, pp. 47209–47219, 2020.

[30] Y. A. Alzahid, P. Mostaghimi, S. D. C. Walsh, and R. T. Armstrong, "Flow regimes during surfactant flooding: the influence of phase behaviour," *Fuel*, vol. 236, pp. 851–860, 2019.

[31] A. H. Alizadeh, M. Khishvand, M. A. Ioannidis, and M. Piri, "Multi-scale experimental study of carbonated water injection: an effective process for mobilization and recovery of trapped oil," *Fuel*, vol. 132, pp. 219–235, 2014.

[32] M. Yue, X. H. Huang, F. M. He, L. Z. Yang, W. Y. Zhu, and Z. X. Chen, "Analysis of the influence of different fracture network structures on the production of shale gas reservoirs," *Geofluids*, vol. 2020, Article ID 8870429, 11 pages, 2020.

[33] M. Yue, W. Y. Zhu, H. Y. Han, H. Q. Song, Y. Q. Long, and Y. Lou, "Experimental research on remaining oil distribution and recovery performances after nano-micron polymer particles injection by direct visualization," *Fuel*, vol. 212, pp. 506–514, 2018.

[34] M. Yue, W. Y. Zhu, Z. Y. Song, Y. Q. Long, and H. Q. Song, "Study on distribution of reservoir endogenous microbe and oil displacement mechanism," *Saudi Journal of Biological Sciences*, vol. 24, no. 2, pp. 263–267, 2017.

[35] H. Q. Song, J. Zhang, D. D. Ni et al., "Investigation on in-situ water ice recovery considering energy efficiency at the lunar south pole," *Applied Energy*, vol. 298, p. 117136, 2021.

[36] J. L. Wang, H. Q. Song, and Y. H. Wang, "Investigation on the micro-flow mechanism of enhanced oil recovery by low- salinity water flooding in carbonate reservoir," *Fuel*, vol. 266, article 117156, 2020.

[37] Q. T. Zhang, C. J. Wei, Y. H. Wang, S. Du, Y. C. Zhou, and H. Q. Song, "Potential for prediction of water saturation distribution in reservoirs utilizing machine learning methods," *Energies*, vol. 12, no. 19, p. 3597, 2019.

[38] N. Wang, Y. Tang, Y. Wu et al., "Dynamic evolution of microstructure morphology in thin-sample solidification: deep learning assisted synchrotron X-ray radiography," *Materials Characterization*, vol. 181, p. 111451, 2021.

[39] S. A. Mahoney, T. E. Rufford, D. Johnson et al., "The effect of rank, lithotype and roughness on contact angle measurements in coal cleats," *International Journal of Coal Geology*, vol. 179, pp. 302–315, 2017.

[40] A. Gerami, P. Mostaghimi, R. T. Armstrong, A. Zamani, and M. E. Warkiani, "A microfluidic framework for studying relative permeability in coal," *International Journal of Coal Geology*, vol. 159, pp. 183–193, 2016.

[41] A. Thomas, S. Q. Wang, S. Sohrabi et al., "Characterization of vascular permeability using a biomimetic microfluidic blood vessel model," *Biomicrofluidics*, vol. 11, no. 2, article 024102, 2017.

[42] S. Perez-Rodriguez, S. A. Huang, C. Borau, J. M. Garcia-Aznar, and W. J. Polacheck, "Microfluidic model of monocyte extravasation reveals the role of hemodynamics and subendothelial matrix mechanics in regulating endothelial integrity," *Biomicrofluidics*, vol. 15, no. 5, article 054102, 2021.

[43] T. Scherr, G. L. Knapp, A. Guitreau et al., "Microfluidics and numerical simulation as methods for standardization of zebrafish sperm cell activation," *Biomedical Microdevices*, vol. 17, no. 3, p. 65, 2015.

[44] Q. Ramadan and L. Jing, "Characterization of tight junction disruption and immune response modulation in a miniaturized Caco-2/U937 coculture-based in vitro model of the human intestinal barrier," *Biomedical Microdevices*, vol. 18, no. 1, 2016.

[45] T. Zhang and S. Y. Sun, "A coupled lattice Boltzmann approach to simulate gas flow and transport in shale reservoirs with dynamic sorption," *Fuel*, vol. 246, pp. 196–203, 2019.

[46] L. Z. Yang, J. J. Xu, J. Fang, Z. G. Cao, T. X. Li, and H. Q. Song, "Risk evaluation of groundwater leakage in coal seam goaf: a case study in the Lingxin Mining Area," *Environmental Science and Pollution Research*, vol. 27, no. 21, pp. 26066–26078, 2020.

[47] K. Xu, D. Agrawal, and Q. Darugar, "Hydrophilic nanoparticle-based enhanced oil recovery: microfluidic investigations on mechanisms," *Energy & Fuels*, vol. 32, no. 11, pp. 11243–11252, 2018.

[48] H. Q. Song, S. du, R. F. Wang et al., "Potential for vertical heterogeneity prediction in reservoir basing on machine learning methods," *Geofluids*, vol. 2020, Article ID 3713525, 12 pages, 2020.

[49] J. Z. Liu, "Potential for evaluation of interwell connectivity under the effect of intraformational bed in reservoirs utilizing machine learning methods," *Geofluids*, vol. 2020, Article ID 1651549, 10 pages, 2020.

[50] Y. D. Wang, M. J. Blunt, R. T. Armstrong, and P. Mostaghimi, "Deep learning in pore scale imaging and modeling," *Earth-Science Reviews*, vol. 215, article 103555, 2021.

[51] Y. M. Liu, S. J. Chen, K. Sagoe-Crentsil, and W. H. Duan, "Predicting the permeability of consolidated silty clay via digital soil reconstruction," *Computers and Geotechnics*, vol. 140, article 104468, 2021.

[52] T. Zhang, Y. Li, Y. T. Li, S. Y. Sun, and X. Gao, "A self-adaptive deep learning algorithm for accelerating multi-component flash calculation," *Computer Methods in Applied Mechanics and Engineering*, vol. 369, article 113207, 2020.

Hindawi
Geofluids
Volume 2021, Article ID 3223530, 9 pages
https://doi.org/10.1155/2021/3223530

Research Article

Prediction of Fluid Viscosity in Multiphase Reservoir Oil System by Machine Learning

Lihua Shao,[1] **Ru Ji**,[2] **Shuyi Du**,[2] **and Hongqing Song**[2]

[1]*School of Mathematics and Physics, University of Science and Technology Beijing, 100083 Beijing, China*
[2]*School of Civil and Resource Engineering, University of Science and Technology Beijing, 100083 Beijing, China*

Correspondence should be addressed to Hongqing Song; songhongqing@ustb.edu.cn

Received 19 October 2021; Accepted 22 November 2021; Published 9 December 2021

Academic Editor: Bicheng Yan

It is important to realize rapid and accurate prediction of fluid viscosity in a multiphase reservoir oil system for improving oil production in petroleum engineering. This study proposed three viscosity prediction models based on machine learning approaches. The prediction accuracy comparison results show that the random forest (RF) model performs accurately in predicting the viscosity of each phase of the reservoir, with the lowest error percentage and highest R^2 values. And the RF model is tremendously fast in a computing time of 0.53 s. In addition, sensitivity analysis indicates that for a multiphase reservoir system, the viscosity of each phase of the reservoir is determined by different factors. Among them, the viscosity of oil is vital for oil production, which is mainly affected by the molar ratio of gas to oil (MR-GO).

1. Introduction

The fluid viscosity of the oil-gas reservoir [1, 2] is the key factor to determine the final development effect and economic benefit of the oil-gas reservoir. Therefore, it has become an important basis performance for formulating oil-gas field development plans [3], studying oil and gas reservoir performance, implementing plan adjustment, and evaluating stimulation [4].

The combination of PVT device and high-pressure falling ball viscometer [5, 6] can realize the laboratory analysis of reservoir samples, to obtain the viscosity value in reservoir environment (high pressure and temperature). PVT device [7, 8] can create specifical temperature and pressure to simulate reservoir environmental conditions; therefore, it has been widely used in the oil industry in recent years [9]. However, the acquisition of such data, including sampling and subsequent analysis, will cost considerable cost and time, which is not desirable [10].

In addition, to simplify and quickly obtain the reservoir fluid viscosity and to analyze the influencing factors of viscosity, a large number of simulation studies on viscosity have appeared in the petroleum industry, and many commonly used viscosity models have been proposed, including LBS viscosity model [11], CS viscosity model, LLS viscosity model, Pt viscosity model, and PR viscosity model [12, 13]. The application of these models can realize the viscosity acquisition of reservoir fluid with specific composition and realize viscosity prediction. But the viscosity of oil-gas reservoir fluid, especially in oil-gas-water multiphase reservoir system, is affected by many factors, including reservoir environmental conditions, oil and gas composition, and water and gas injection [14]. At present, it is impossible to find a general viscosity model to describe the viscosity characteristics of fluids in multiphase reservoir systems.

Therefore, the objective of this study is to establish a reliable and accurate machine learning model for predicting the viscosity of each phase in a multiphase mixed oil-gas-water system. Research shows that deep neural networks (DNN), random forests (RF), and support vector regression (SVR) are very good at capturing and learning the nonlinear feature relationships between data, and they can accurately predict parameters in a data-driven manner without physical models. Compared with some classic machine learning algorithms, these machine learning algorithms can often

maintain high prediction accuracy even under small sample conditions, which map more feature spaces. And trained models have higher portability and can quickly adapt to different application scenarios. This paper, therefore, would choose these three machine learning methods to predict and analyze the viscosity of each fluid in a multiphase mixed oil-gas-water system.

To achieve this research purpose, Tarim reservoir oils were taken as an example. Tarim Oilfield is located in Kuqa County, Xinjiang, China. It is urgent to tackle the key problems of enhanced oil recovery. The early gas injection research recognized that gas injection is the practical technical direction of enhanced oil recovery of reservoir in Tarim. This study collects a large number of experimental and simulation data to provide a large amount of data for machine learning. In addition, the sensitivity analysis of influencing factors plays a vital role in the development guidance of the reservoir oils, so the collected data was sorted and analyzed, according to the reservoir environmental factors (temperature T and pressure P) and the reservoir composition (the molar ratio of gas to oil MR-GO and molar ratio of water to oil MR-WO [15, 16]). Therefore, the developed viscosity prediction model covers wider ranges of input data, which is important to the production of oil reservoirs.

The structure of this paper is as follows. In the following section, the background, governing equations, and development methodology of the three presented models, including RF, DNN, and SVR, are introduced and described in detail. In addition, this section will also give the calculation method of the statistical indicators for evaluating the three models. Next, in Section 3, the accuracy and calculation time comparison of these three developed models will be evaluated by the statistical indicators, and the reliability analysis of the calculation process of the RF model will be given. Moreover, the sensitivity analysis of the influencing factors will be carried out, and the influence weight of each influencing factor on the output viscosity in the multiphase system will also be given. Finally, Section 4 will present the key findings of this paper.

2. Methodology

2.1. Prediction Models. Three prediction models were used to predict the viscosity of a multiphase reservoir oil system from input data of crude oil systems, such as MR-GO, MR-WO, reservoir environment pressure (P, MPa), and temperature (T, °C). Table 1 indicates the statistical characteristics of the input data.

2.1.1. Random Forest (RF). Random forest [17], as an ensemble learning algorithm based on classification and regression trees, has been widely used in many fields. Ho [18] first proposed the random forest algorithm in 1995 and improved the algorithm by Breiman [19] in 2001. RF is a machine learning method based on statistical learning theory. First, multiple samples are extracted from the original data set through the bootstrap resampling method. Then, a decision tree model is established for

Table 1: Statistical characteristic of the input data.

Parameter	Minimum	Maximum	Mean
P	5	95	46.25
T	30	210	120
MR-GO	0.5	6	1.5
MR-WO	0	2.3	1.292

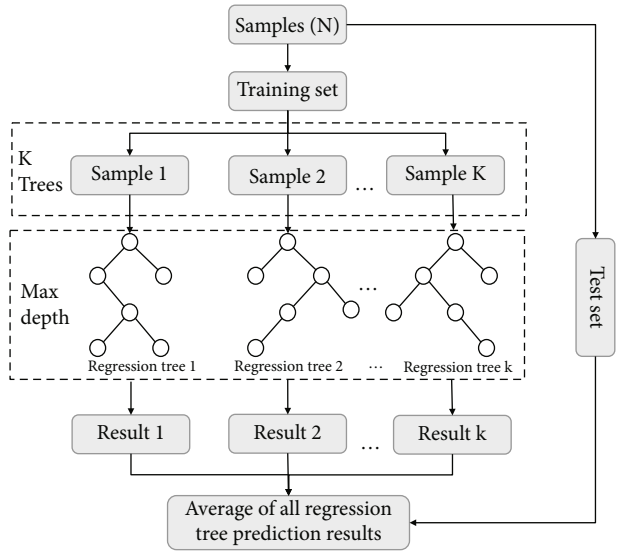


Figure 1: The training process of random forest.

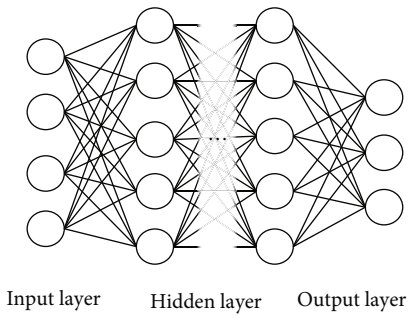
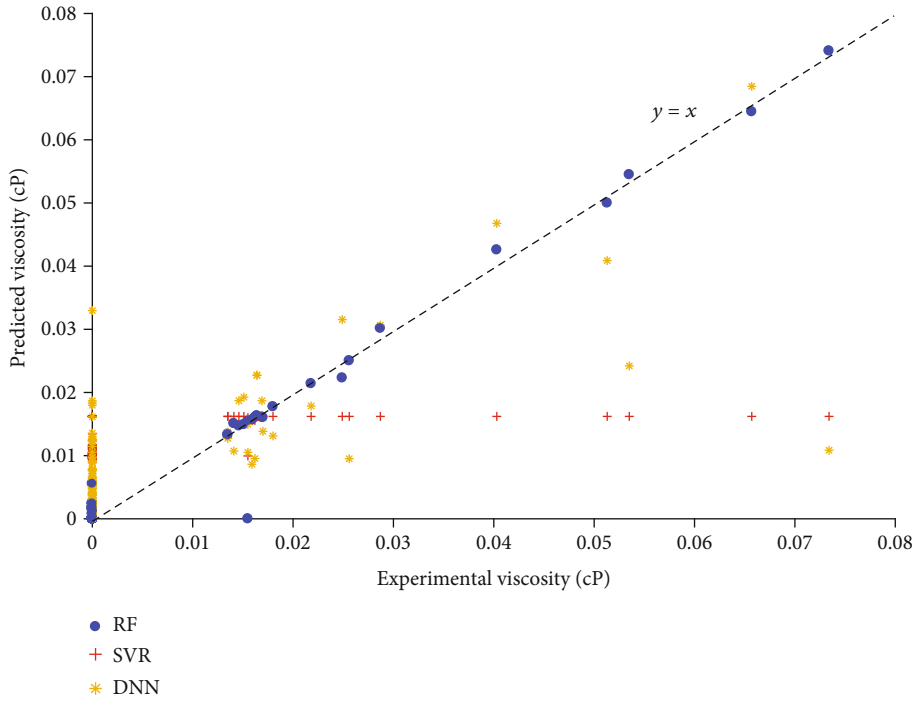


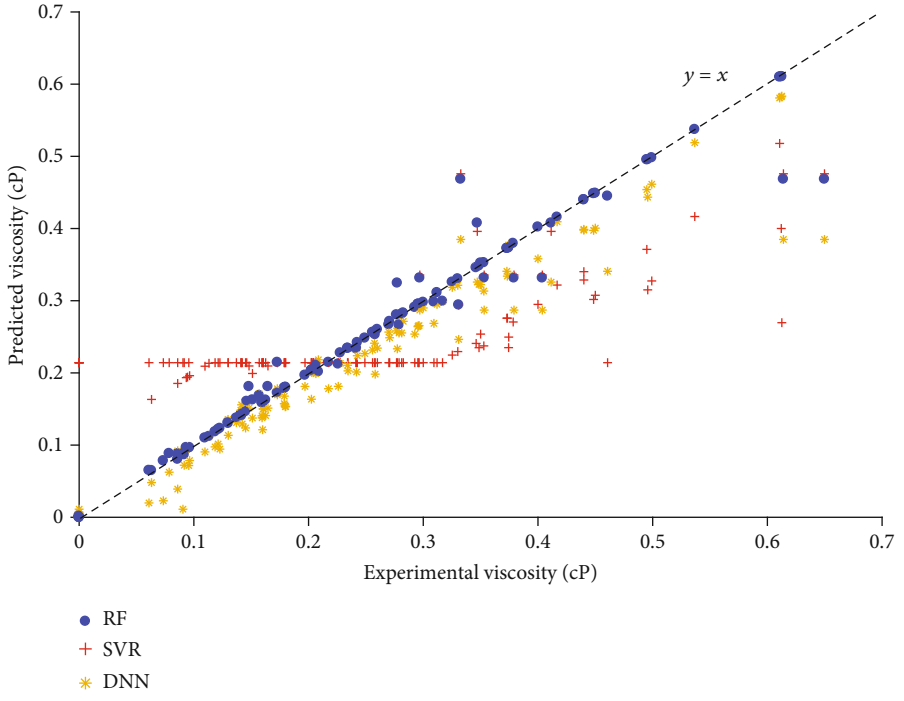
Figure 2: The training process of deep neural network.

Table 2: Statistical evaluation results calculated by three models.

Models		MSE	MAE	R^2
RF	v_g	$2.8979e-06$	0.0005	0.9824
	v_o	0.0008	0.0093	0.9623
	v_w	$4.3221e-06$	0.0008	0.9999
DNN	v_g	$9.2978e-05$	0.0063	0.4364
	v_o	0.0022	0.0301	0.8942
	v_w	0.0099	0.0463	0.7759
SVR	v_g	0.0002	0.0112	-0.1083
	v_o	0.0108	0.0855	0.4798
	v_w	0.0118	0.0767	0.7340

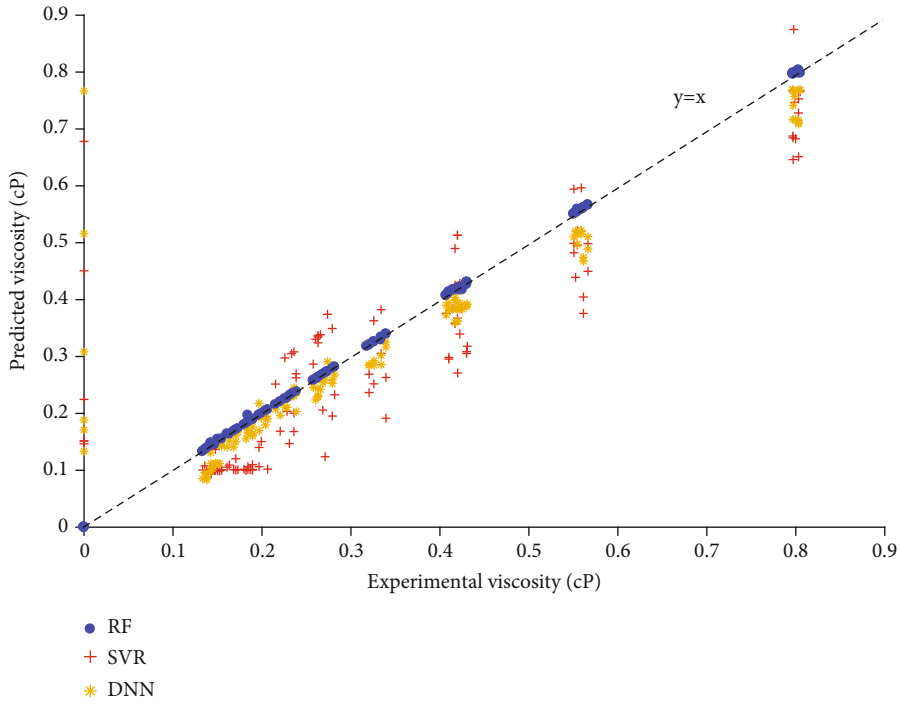


(a) Gas viscosity (υ_g)



(b) Oil viscosity (υ_o)

Figure 3: Continued.



(c) Water viscosity (υ_w)

Figure 3: The cross-plot pictures between model predictions and the corresponding experimental values.

each bootstrap sample. Finally, the predictions of multiple decision trees are combined and averaged to obtain the final prediction result. The ultimate regression decision formula is as follows:

$$M_{\mathrm{rf}}(x) = \frac{1}{K}\sum_{k=1}^{K} t_i(x), \tag{1}$$

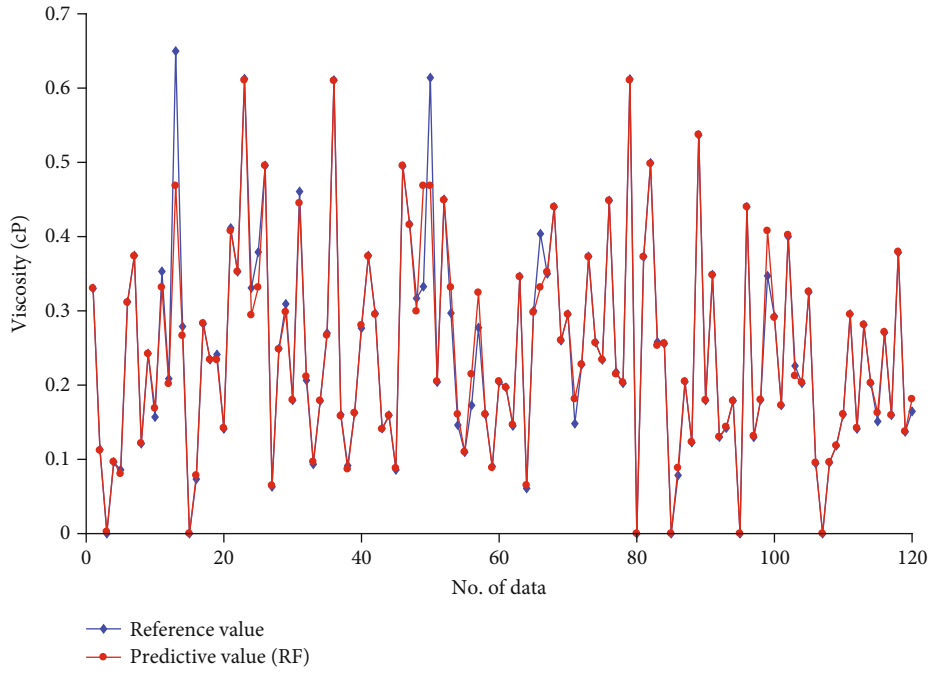
where M_{rf} represents the calculation result of the random forest model, K is the number of regression trees required, and t_i represents a single regression tree model. In the calculation process of the random forest model, there are two extremely critical hyperparameters, which are the number of regression trees (n-estimators) and the number of random variables at the nodes (max depth). Too little number of regression trees will affect the accuracy of the calculation. Similarly, too many numbers will increase the complexity of the calculation. The training process of random forest is shown in Figure 1.

2.1.2. Deep Neural Network (DNN). A deep neural network [20, 21] is a machine learning method that combines a multilayer perceptron structure and a backpropagation algorithm. It is mainly composed of three parts, including the input layer, hidden layer, and output layer (Figure 2). The key structure of a deep neural network is called a neuron, which can characterize the nonlinear mapping relationship between input and output. The output equation of each neuron is as follows:

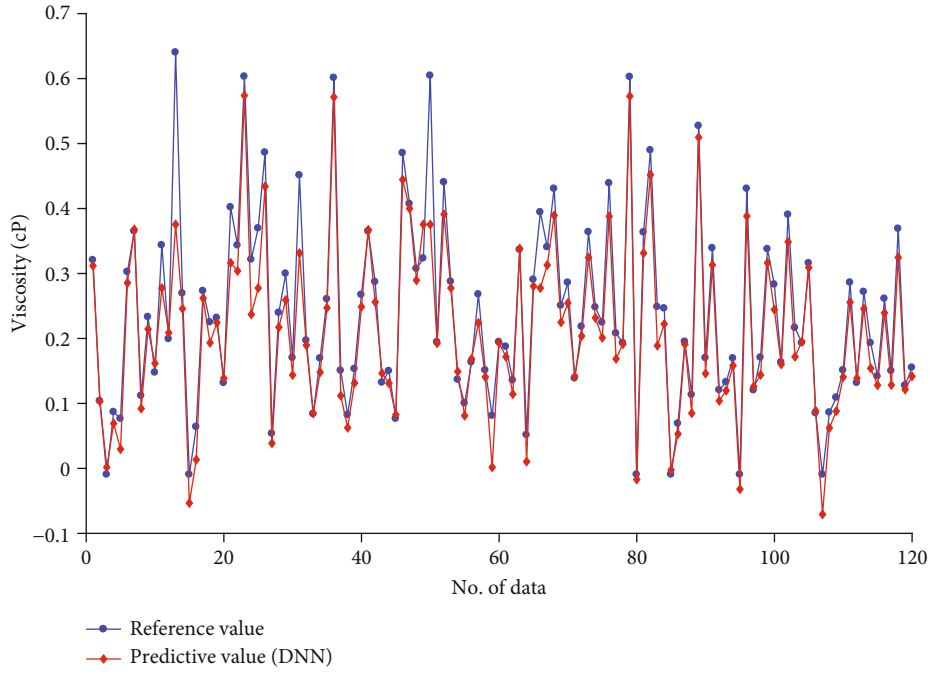
$$O(x) = g\left(\sum_i w_i x_i + b\right), \tag{2}$$

where O is the output value, g is the activation function, w is the weight of the input parameter, and b is the threshold. The training process of the neural network model is divided into two steps: forward propagation and backward propagation. First, forward propagation is used to calculate the predicted value of the model. The error gradient between the predicted value and the true value is obtained by the loss function. According to the error gradient, the weights and thresholds in the neural network are adjusted through the backpropagation algorithm. Repeating this process can make the network continuously learn the hidden features between the data.

2.1.3. Support Vector Regression (SVR). Support vector machine [22–24] is a machine learning method proposed by Cortes and Vapnik [25] in 1995 to learn the mapping relationship between parameters. The core idea of support vector regression is to find the nonlinear mapping relationship between input space and output space. Relying on the nonlinear mapping, data is mapped to a high-dimensional characteristic space. The estimating

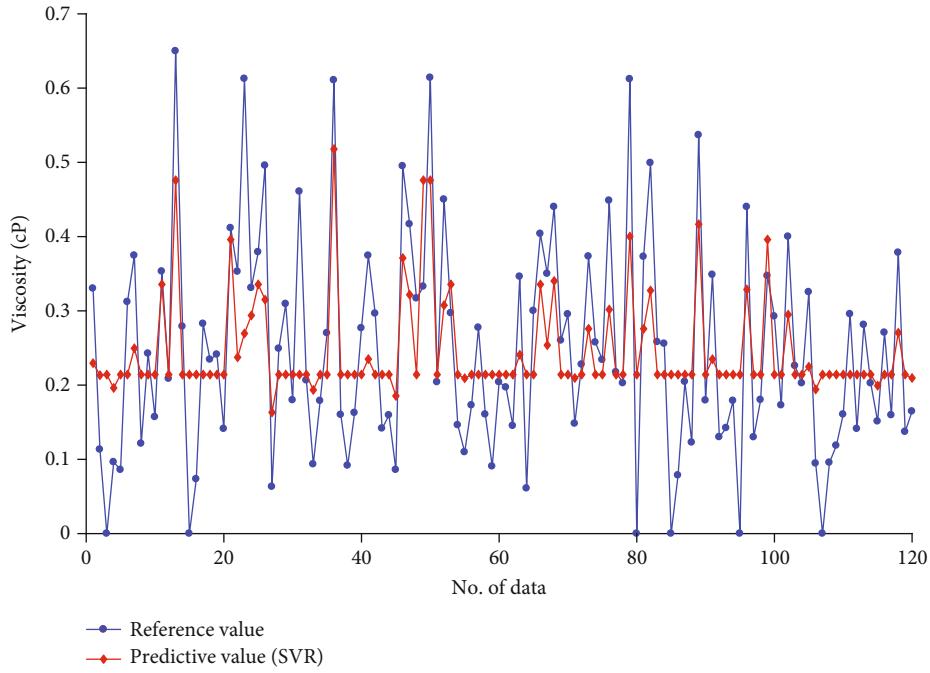


(a) RF model



(b) DNN model

Figure 4: Continued.



(c) SVR model

Figure 4: The comparison diagram between model predictions and the corresponding experimental values of oil viscosity (v_o).

function of linear regression in characteristic space is as follows:

$$f(x) = \langle w, x \rangle + b, \tag{3}$$

where f is the linear function and w and b are the identified weight vector and the bias term, respectively.

In the high-dimensional feature space, the optimization problem for SVR with ε-insensitive loss function is as follows:

$$\begin{aligned} \min\ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \zeta_i) \\ \text{s.t.}\ & \left(w^T x_i + b\right) - y_i \le \varepsilon + \xi_i \\ & y_i - \left(w^T x_i + b\right) \le \varepsilon + \zeta_i \\ & \xi_i, \zeta_i \ge 0, \quad i = 1, 2, \cdots, N, \end{aligned} \tag{4}$$

where $\|w\|^2$ in the objective function is the confidence range reflecting the generalization ability, ξ_i and ζ_i are the slack variables that represent the upper and lower limits of allowable error, $\sum_{i=1}^{N}(\xi_i + \zeta_i)$ denotes the experimental risk reflecting the learning capacity of function, $\varepsilon > 0$ is an insensitive loss coefficient, and parameter $C(C \ge 0)$ is a penalty factor. In SVR, the dual problem of Equation (4) is often derived by using the Lagrange multiplier method, based on which a linear regression function can finally be constructed.

2.2. Statistical Evaluation of Three Models. To evaluate the accuracy of machine learning models, three statistical indicators were used, including mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) [26–28]. This experiment also utilizes these indicators to

Table 3: The prediction calculation time of different models.

Methods	Prediction time
RF	0.53 s
DNN	0.82 s
SVR	0.76 s
TDM	52.8 s

Table 4: Statistical error analysis results calculated by RF model.

Parameter	Value
Training set	
R^2	0.9921
MAE	0.0021
MSE	$9.942E-5$
Testing set	
R^2	0.9816
MAE	0.0035
MSE	0.0003
Total	
R^2	0.9811
MAE	0.0023
MSE	0.0001

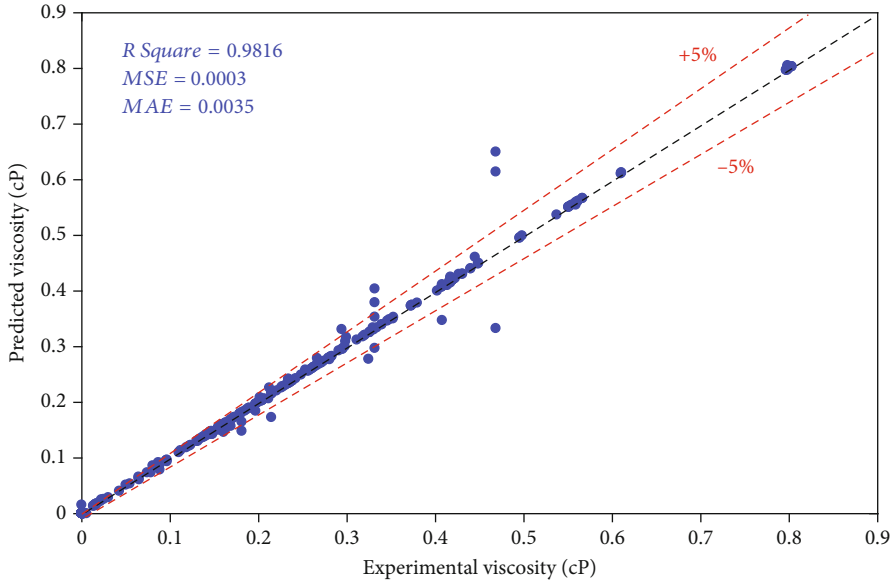


Figure 5: The cross-plot figure of the proposed RF model.

calculate the error and test the robustness of the machine learning model.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i^{\mathrm{data}} - X_i^{\mathrm{model}}\right)^2, \tag{5}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|X_i^{\mathrm{real}} - X_i^{\mathrm{mod\ el}}\right|, \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(X_i^{\mathrm{data}} - X_i^{\mathrm{model}}\right)^2}{\sum_{i=1}^{N}\left(X_i^{\mathrm{data}} - \mathrm{average}\left(X_i^{\mathrm{model}}\right)\right)^2}, \tag{7}$$

where N represents the total number of samples, X_i^{data} is the referenced parameter that is the actual expected value, and X_i^{model} represents the predicted value of the machine learning methods.

3. Results and Discussion

3.1. Comparison of Three Proposed Models. To evaluate the precision of each constructed network, MSE, MAE, and R^2 are calculated, based on different output data of gas viscosity (υ_g), oil viscosity (υ_o), and water viscosity (υ_w), respectively. The calculated evaluation results are presented in Table 2.

According to the results of Table 2, the proper prediction model for viscosity modeling is RF. Take the oil viscosity modeling as an example; among these three prediction models, the presented RF model has the lowest MSE of 0.008, lowest MAE of 0.0093, and highest R^2 value of 0.9623. Compared to the other two developed models, the SVR model has the worst prediction results.

To reveal and visualize the performance of each prediction model, a cross-plot picture between model predictions and the corresponding experimental values is drawn in Figure 3, for gas viscosity (υ_g), oil viscosity (υ_o), and water viscosity (υ_w), respectively. In this curve, there are three color points: blue points for estimated data of RF, red points for estimated data of SVR, and yellow points for estimated data of DNN, and $Y = X$ line for experimental values. In a cross-plot picture, a higher precision is attained when the data is closer to the $Y = X$ line. From Figures 3(a)–3(c), in all prediction results for gas viscosity (υ_g), oil viscosity (υ_o), and water viscosity (υ_w), only in the RF model, there is an adequate closeness of the majority of the data points to the line $Y = X$, showing a very good agreement between model predictions and the corresponding experimental values. For the RF model, the coefficient of determination (R^2) is high to 0.9824, 0.9623, and 0.9999 for gas viscosity (υ_g), oil viscosity (υ_o), and water viscosity (υ_w), respectively, which is much higher than the other models.

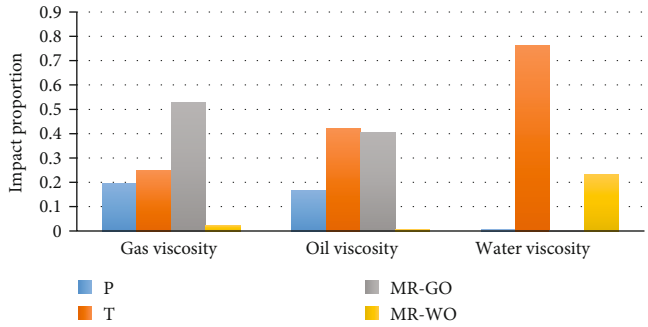


Figure 6: Importance analysis of influencing factors on the viscosity of multiphase gas-oil-water systems.

Moreover, in Figure 4, model predictions and the corresponding experimental values of oil viscosity (υ_o) by three proposed models are also compared and depicted. From these results, the predicted results using the RF model are very close to the corresponding experimental results. It can be seen that the RF model achieves a good accuracy of

predicting the oil viscosity (υ_o) of the multiphase oil-gas-water systems.

Finally, taking 1200 sets of data as an example, the prediction calculation time of machine learning models was compared with the tradition numerical method (TDM), as shown in Table 3. In this paper, the experiment is based on TensorFlow and Sk-learn learning library with Python language. The hardware resources include Intel i7-7700hq@2.8G processor, 16 G memory, and Nvidia GTX 1060 (6 G) graphics card. Table 3 indicates that machine learning models only need 0.53 s, 0.82 s, and 0.76 s for the prediction calculation of RF, DNN, and SVR, respectively. Prediction time by machine learning models is tremendously fast, while for the tradition numerical method, the prediction time is high up to 52.8 s. Machine learning models, therefore, have advantages in terms of computing time.

3.2. Results Analysis by RF Model. Statistical evaluation parameters of the best network RF model for training, testing, and total data sets are presented in Table 4. Table 4 reports that for training set, R^2 of the RF model is high to 0.9921, while MAE and MSE are very low with the values of 0.0021 and $9.942E-5$. For testing set, the R^2, MAE, and MSE values are 0.9816, 0.0035, and 0.0003. While for a total set, the corresponding statistical evaluation results are 0.9811, 0.0023, and 0.0001, respectively, indicating good evaluation parameters for the RF model.

Cross-plot figure of the proposed RF model is also shown in Figure 5. It could be seen that there is a dense point distribution around the $Y = X$ line for all of the data, and the error is basically maintained within 5%, indicating the sufficient accuracy and reliability of the developed RF model.

3.3. Importance Analysis. Importance analysis of influencing factors was carried out to figure out a sensitivity analysis on viscosity in a multiphase gas-oil-water system. For each viscosity, such as gas, oil, and water, each of the independent influencing factors, such as P, T, MR-GO, and MR-WO, was evaluated in this part. The results of the importance analysis are indicated in Figure 6. For each output result, the sum of the influence proportion of the influencing factors (P, T, MR-GO, and MR-WO) is 1. The higher the impact proportion value, the stronger the relationship between the input parameter and the output function.

As it is expected, for gas viscosity, the MR-GO is the most significant factor influencing the output result with an impact proportion of 0.53. The second is environmental factors, including P and T, and the proportions are 0.20 and 0.25, respectively. Here, the water content has a litter effect. For oil viscosity, the output value is mainly affected by ambient temperature T and gas content MR-GO, and the specific impact proportion is 0.42 and 0.40, respectively. Secondly, pressure P also plays an important role with the affecting proportion of 0.17. Finally, the viscosity of water in the multiphase system is mainly determined by the ambient temperature T, and the influence proportion is 0.76, followed by its content MR-WO in the system, and the ratio is 0.23.

4. Conclusions

In this study, three machine learning models, namely, random forest (RF), deep neural network (DNN), and support vector regression (SVR), were proposed for calculating and predicting the phase viscosity of multiphase reservoir oils systems.

To make a judgment of the accuracy of each developed model, various statistical evaluation indicators, including mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2), were applied. The results show that the RF model has higher accuracy compared with the other two models. Specifically, for the RF model, the R^2, MAE, and MSE values are 0.9811, 0.0023, and 0.0001, respectively, indicating good evaluation performance. Moreover, machine learning models have advantages in computing time for the RF model, which only needs 0.53 s for 1200 sets of data prediction.

Moreover, an importance analysis of influencing factors was carried out on viscosity in this multiphase gas-oil-water system. For gas viscosity, the MR-GO is the most significant influencing factor with an impact proportion of 0.53, followed by P and T. Next, oil viscosity is mainly affected by T and MR-WO, and P also has an effect. Finally, for water, the viscosity is mainly determined by T, followed by MR-WO.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

All authors confirm that there is no financial/personal interest or belief that could affect our objectivity, and no conflicts exist.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 11972073).

References

[1] A. Hemmati-Sarapardeh, A. Shokrollahi, A. Tatar, F. Gharagheizi, A. H. Mohammadi, and A. Naseri, "Reservoir oil viscosity determination using a rigorous approach," *Fuel*, vol. 116, no. 1, pp. 39–48, 2014.

[2] H. Chen, "Oil-gas-water system study in fine reservoir description researches—taking Yulou oil-bearing sets in the West Depression in Liaohe Basin in China as an example," *Arabian Journal of Geosciences*, vol. 14, no. 4, pp. 1–16, 2021.

[3] L. Chen, Z. Liu, and N. Ma, "Optimize production allocation for the oil-gas field basing on a novel grey model," *Journal of Natural Gas Geoscience*, vol. 4, no. 2, pp. 121–128, 2019.

[4] S. K. Matthai, "Reservoir simulation: mathematical techniques in oil recovery," *Geofluids*, vol. 8, no. 4, 345 pages, 2008.

[5] B. Calvignac, E. Rodier, J. J. Letourneau, P. Vitoux, C. Aymonier, and J. Fages, "Development of an improved

falling ball viscometer for high-pressure measurements with supercritical CO_2," *The Journal of Supercritical Fluids*, vol. 55, no. 1, pp. 96–106, 2010.

[6] S. Swamura, "An improvement of high-pressure rolling-ball viscometer and evaluation of the equipment constant," *Journal of the Society of Materials Science Japan*, vol. 63, no. 10, pp. 729–734, 2014.

[7] I. A. Munz, B. W. D. Yardley, and S. A. Gleeson, "Petroleum infiltration of high-grade basement, South Norway: pressure-temperature-time-composition (P–T–t–X) constraints," *Geofluids*, vol. 2, no. 1, 53 pages, 2002.

[8] C.-Y. Sun G.-J. Chen et al., "Measurement of interfacial tension for the CO_2 injected crude oil + reservoir water system," *Journal of Chemical & Engineering Data*, vol. 50, no. 3, pp. 936–938, 2005.

[9] R. A. Almehaideb, I. Ashour, and K. A. El-Fattah, "Improved K-value correlation for UAE crude oil components at high pressures using PVT laboratory data," *Fuel*, vol. 82, no. 9, pp. 1057–1065, 2003.

[10] M. G. Rezk and J. Foroozesh, "Phase behavior and fluid interactions of a CO_2-light oil system at high pressures and temperatures," *Heliyon*, vol. 5, no. 7, p. e02057, 2019.

[11] V. A. Tarasov, M. A. Komkov, V. A. Romanenkov, R. V. Boyarskaya, and D. S. Boyarskii, "A rheological model of the influence of thermal exposure on the viscosity of a an LBS-4 varnish during impregnation of a multilayer woven blank of a heat-protective coating," *Polymer Science Series D*, vol. 11, no. 2, pp. 179–182, 2018.

[12] S. Khosharay, M. Pierantozzi, and G. Di Nicola, "Modeling investigation on the viscosity of pure refrigerants and their liquid mixtures by using the Patel–Teja viscosity equation of state," *International Journal of Refrigeration*, vol. 85, pp. 255–267, 2018.

[13] G. Feng, C. Shuang, and X. Guo, "Comparison and analysis of the viscosity forecasting models of reservoir fluids," *Petroleum Geology and Recovery Efficiency*, vol. 5, 2007.

[14] H. Wang, J. G. Wang, F. Gao, and X. Wang, "A two-phase flowback model for multiscale diffusion and flow in fractured shale gas reservoirs," *Geofluids*, vol. 2018, Article ID 5910437, 15 pages, 2018.

[15] J. G. Heuer, C. L. Jacocks, and Inventors; Conoco Inc., "Control of gas-oil ratio in producing wells," *U.S. Patent 3,368,624*, vol. 13, 1968.

[16] K. S. Bajaalah, "New correlations for gas oil ratio and oil formation volume factor for gas condensate reservoirs," *International Journal of Innovation and Applied Studies*, vol. 13, no. 4, 2015.

[17] M. Daviran, A. Maghsoudi, R. Ghezelbash, and B. Pradhan, "A new strategy for spatial predictive mapping of mineral prospectivity: automated hyperparameter tuning of random forest approach," *Computers & Geosciences*, vol. 148, 2021.

[18] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada, 1995.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] M. Hauser, S. Gunn, S. Saab, and A. Ray, "State-space representations of deep neural networks," *Neural Computation*, vol. 31, no. 3, pp. 538–554, 2019.

[21] "Deep neural networks-based classification optimization by reducing the feature dimensionality with the variants of gravitational search algorithm," *International Journal of Modern Physics C*, vol. 32, no. 10, p. 2150137, 2021.

[22] A. C. Serfidan, F. Uzman, and M. Tuerkay, "Optimal estimation of physical properties of the products of an atmospheric distillation column using support vector regression," *Computers & Chemical Engineering*, vol. 134, article 106711, 2020.

[23] L. Wang, Y. Ma, X. Chang, C. Gao, Q. Qu, and X. Chen, "Projection wavelet weighted twin support vector regression for OFDM system channel estimation," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 469–489, 2021.

[24] M. Asadolahi, M. G. Akbari, G. Hesamian, and M. Arefi, "A robust support vector regression with exact predictors and fuzzy responses," *International Journal of Approximate Reasoning*, vol. 132, pp. 206–225, 2021.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[26] R. M. Khalifa, S. Yacout, and S. Bassetto, "Developing machine-learning regression model with logical analysis of data (LAD)," *Computers & Industrial Engineering*, vol. 151, article 106947, 2021.

[27] B. Kumi-Boateng and M. S. Peprah, "Modelling local geometric geoid using soft computing and classical techniques: a case study of the University of Mines and Technology (UMaT) local geodetic reference network," *International Journal of Earth Sciences*, vol. 2, no. 3, pp. 166–177, 2020.

[28] H. Gong, R. Pishgar, and J. H. Tay, "Artificial neural network modelling for organic and total nitrogen removal of aerobic granulation under steady-state condition," *Environmental Technology*, vol. 40, no. 24, pp. 3124–3139, 2019.