Artificial Intelligence in Geosciences 4 (2023) 95–110

Contents lists available at ScienceDirect

Artificial Intelligence in Geosciences

journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences

Optimization of shale gas fracturing parameters based on artificial intelligence algorithm

Shihao Qian [a], Zhenzhen Dong [a], Qianqian Shi [a], Wei Guo [b], Xiaowei Zhang [b], Zhaoxia Liu [b], Lingjun Wang [a], Lei Wu [a], Tianyang Zhang [a], Weirong Li [a,*]

[a] Xi'an Shiyou University, Xi'an, 710065, China

[b] Research Institute of Petroleum Exploration & Development, PetroChina, Beijing, 100083, China

ARTICLE INFO

Keywords:
Shale gas
Parameter optimization
Prediction
GBDT
PSO

ABSTRACT

Resource-rich shale gas plays a pivotal role in new energy types. The key to scientifically and efficiently developing shale gas fields is to clarify the main factors that affect the production of shale gas wells. In this paper, according to the shale gas reservoir characteristic of the Fuling marine Longmaxi Formation, a single-well geological model was established using the reservoir numerical simulation software CMG. Then, 10,000 different reservoir models were randomly generated for different formation physical parameters, completion parameters, and fracturing parameters using the Monte Carlo method, and these 10,000 models were simulated numerically. The machine learning model uses a dataset of 10,000 different geological, completion, and fracturing parameters as input and 10,000 production curves as output. Multiple machine learning regression methods were used to train and test the dataset, and the optimal method (GBDT algorithm) was selected, and the accuracy R^2 of the test set of the GBDT prediction model is 0.96. A fracturing parameter optimization workflow was constructed by combining a production prediction model with a particle swarm optimizer (PSO). The process can quickly optimize the fracturing parameters and predict the production for each time by targeting the cumulative gas production under different geological conditions. The optimized parameters are Fracture Spacing, Fracture Width, Intrinsic Permeability, Fracture Half-length, Langmuir Pressure, and Langmuir Volume. The initial predicted cumulative gas production was 4.59×10^8 m^3, which was optimized to 4.90×10^8 m^3. The proposed PSO-GBDT proxy model can instantly predict the production of shale gas wells with considerable accuracy, reliability, and efficiency, which is a vital tool for optimizing fracture design. This investigation provides a solid foundation for predicting the production of unconventional gas reservoirs and for parameter optimization.

1. Introduction

Shale gas is an unconventional natural gas stored in reservoir rock systems, mainly organic-rich shale. It can be present in a free state in natural fractures and pores but also in an adsorbed condition on the surface of the cheesecake and clay particles, as well as a minor amount stored in a dissolved state in the cheesecake and asphaltene, with the percentage of adsorbed gas generally ranging from 20% to 85% (King, 2010). Shale gas resources are abundant worldwide and have great potential for development. Shale gas resources worldwide are 4.57×10^{14} m^3, of which 1.87×10^{14} m^3 is technically recoverable (Ambrose et al., 2010). In the world, the five nations with the largest technically recoverable resources of shale gas are China (3.6×10^{13} m^3, accounting for about 20%), the United States (2.4×10^{13} m^3, about 13%), Argentina, Mexico, and South Africa. Shale gas is abundant in China, with technically recoverable resources of 3.6×10^{13} m^3, 1.6 times more than conventional gas (Ma et al., 2017; Johnson and Boersma, 2013). As a result, the plentiful shale gas resources in China lay the foundation for the further development of the energy industry and the importance of shale gas as a resource with huge reserves.

Many scholars have conducted sensitivity analysis as well as optimization of physical parameters of reservoirs, well completion parameters, and fracturing parameters in shale gas reservoir development attributed to the different petrophysical properties and production

* Corresponding author.
E-mail addresses: LTXH990111@163.com (S. Qian), dongzz@xsyu.edu.cn (Z. Dong), sqq17765855520@163.com (Q. Shi), weiguo12022@163.com (W. Guo), zhangxw12022@163.com (X. Zhang), liuzhaoxia@petrochina.com.cn (Z. Liu), w894459085@163.com (L. Wang), wuleiyx@aliyun.com (L. Wu), zty16223334@gmail.com (T. Zhang), weirong.li@xsyu.edu.cn (W. Li).

https://doi.org/10.1016/j.aiig.2023.08.001
Received 11 April 2023; Received in revised form 4 July 2023; Accepted 4 August 2023
Available online 5 August 2023
2666-5441/

characteristics of shale gas reservoirs (Onwunalu and Durlofsky, 2009; Williams-Stroud, 2008; Xu et al., 2015). Curtis et al. (Curtis, 2002) introduced the concept of shale and shale gas. They analyzed and compared the maturity, gas adsorption coefficient, reservoir thickness, organic content, and total gas volume of five shale reservoirs in the United States. The study's results indicate that the degree of natural fracture development was an essential controlling factor in shale gas development and that successful extraction of shale gas reservoirs requires hydraulic fracturing. Zhang (Zhang et al., 2009) investigated the effect of reservoir parameters and hydraulic fracturing parameters on the productivity of shale gas reservoirs using Eclipse software. Unsteady gas flow from matrix to fracture and multi-component gas desorption were considered. The matrix-fracture coupling factor (sigma) and fracture permeability were used to characterize the fracture network caused by fracturing. The effects of matrix and fracture permeability, non-Darcy flow coefficient, porosity, matrix sub-grid, channeling coefficient, rock compressibility, half-length, and spacing of prominent fractures on production were analyzed. WeiYu et al. (Yu and Sepehrnoori, 2013) demonstrated the accuracy of multi-stage hydraulic fracture numerical simulations for valuable Barnett Shale production data by factoring in gas desorption effects. Based on the Barnett Shale data, six uncertain parameters within a reasonable range are determined to finalize the optimal design based on NPV maximization for different gas price conditions. This integrated technique optimizes well location and hydraulic fracture treatment design to produce the best drainage area around the well. It gives insight into hydraulic fracture interference between single wells and nearby wells. Yan Xuemei (Yan et al., 2015) developed a production prediction model based on "main fracture + network fracture permeability" using Eclipse software and applied a Plackett-Burman-type linear experimental design method. Parameters such as the practical reconstruction volume, the number of primary fractures, and the length of prior fractures were optimized through sensitivity analysis. Zhu Dawei et al. (Zhu et al., 2020) developed a coupled local grid encryption-embedded discrete fracture model for fractured well production prediction. They adopted an orthogonal design to optimize parameters, including fracture length, fracture conductivity, and the number of fracture sections.

With the fast growth of data science in recent years, big data analysis methodologies have been increasingly employed in oil and gas exploration and production (Ben et al., 2020; Dong et al., 2022; Wu et al., 2021; Zhan et al., 2019). Petroleum engineers have started to use machine learning methods for production prediction and fracturing parameter optimization for unconventional resources, which can not only accurately evaluate the fracturing effect of each well but also optimize the fracturing parameters. Gorucu et al. (Gorucu and Ertekin, 2011) investigated optimizing the design of hydraulically fractured horizontal wells in naturally fractured tight gas-sand reservoir systems. An expert system for planning efficient production improvement methods was created by combining a commercial reservoir simulator with an artificial neural network (ANN). The developed ANN-based production prediction model was utilized by Nejad et al. (2015) to simulate and optimize the fracturing parameters of a fractured well in the Eagle Ford formation. The production of a single well using the optimized combination of fracturing parameters was measured to be 43% higher than the actual one. In 2018, Luo et al. (2018) selected 13 model input parameters from the Bakken shale based on previous experience and careful consideration. The parameter selection was the so-called feature extraction using Pearson correlation coefficient, random forest, RFE method, and Ll parametric method. A neural network prediction model was developed with sound production in the first year as the target variable. Where the hidden layer is four layers with 100 neurons per layer, and the accuracy of the test set is $R^2 = 0.614$, and finally a sensitivity analysis was performed to prove the applicability of the machine learning method. Wang (Wang and Chen, 2019) developed yield prediction models based on RF, AdaBoost, SVM, and ANN algorithms. The yield prediction models were preferentially developed by comparing the prediction accuracy and prediction mean square deviation of multiple models. The coefficient of determination (R^2) of both the training and prediction sets of AdaBoost and RF algorithms were higher and the prediction results were better. However, the mean square error difference between the AdaBoost training and validation sets was large, indicating that the AdaBoost method had the problem of overfitting. As a result, Wang proposed developing the production prediction model utilizing RF rather than the AdaBoost method. Tan Chaodong et al. (Tan et al., 2020) used the fracturing construction history data of 200 existing wells and reservoir physical characteristics to develop a Bayesian neural network model to optimize the fracturing parameters. Principal component analysis (PCA) was employed to decrease the dimensionality even further. The Bayesian neural network model's input parameters are the reduced-dimensional principal components, and the output parameter is the fracturing impact evaluation index. To avoid overfitting the neural network, the Bayesian approach was used to adaptively update the regularization coefficients, and a three-layer Bayesian neural network prediction model was created. The model was trained using 90% of the 200 wells as training data and 10% as test data. The testing findings revealed that the model's relative error in predicting the test set after training was less than 5% and that it could be utilized to improve the fracturing parameters.

In recent years, there has been an increasing trend among scholars to employ machine learning in shale gas production. Specifically, various machine learning methods are utilized to predict key performance indicators, such as gas production and net present value, and optimize both reservoir and fracturing parameters in shale gas reservoir fracturing operations. In their study, Zhao et al. (2022) proposed an innovative approach called VFRBF-FSO, which utilizes an intelligent variable fidelity radial basis function (VFRBF) surrogate model for optimizing fracture stages in a multi-objective framework. The objective functions considered were the net present value (NPV) and cumulative gas production (CGP). The optimization variables selected included fracture half-length, fracture spacing, well spacing, number of fracture strips, and well length. The results of the optimization process revealed that the VFRBF-FSO method demonstrated excellent convergence and versatility. Moreover, it significantly reduced the simulation run time compared to the HF model by approximately five times for two different well and fracture configurations. Wang et al. (2022a) proposed a multi-objective optimization prediction model (MOO-PM) that integrates the least squares support vector regression (LSSVR) prediction model and the non-dominated ranking genetic algorithm II (NSGA-II). The objective functions considered were the fracture fluid return rate (FBR) and the first-month gas production (PROD). Optimization variables included horizontal length, number of fracture sections, fracture length, fracture fluid injection rate, fracture fluid viscosity, fracture fluid volume, and proppant dosage. Furthermore, in the same year, Wang et al. (2022b) introduced a novel integrated optimization method called WSF-MFSVR, which utilized a multi-fidelity support vector regression (MFSVR) surrogate model. The optimization focused on horizontal well spacing and fracture stage placement. Optimization variables included well spacing, fracture half-length, number of fracture strips, fracture spacing, and well length. The optimization objectives were defined as net present value (NPV) and cumulative gas production (CGP). To enhance evaluation accuracy, the researchers employed a particle swarm optimization (PSO) algorithm to determine the optimal hyperparameters of the MFSVR model. In 2023, Zhou and Ran (2023) introduced a modified genetic algorithm approach called Spearman Genetic Algorithm (SGA) for efficient optimization of fracturing parameters in the context of reservoir engineering. The optimization process focused on key parameters including the number of fracture sections, horizontal length, fracture width, and fracture half-length. Additionally, leveraging a dataset obtained from simulations of multi-stage fractured horizontal wells in shale gas reservoirs, the researchers developed a production prediction model using the XGBoost algorithm. This model allowed for accurate forecasting of production outcomes based on the analyzed data.

Although various techniques have been utilized to predict shale reservoir production performance and optimize each parameter, after reviewing the literature and related works, we observed that some problems remain.

(1) The application of machine learning techniques in optimizing shale gas fracturing parameters and predicting shale gas production is currently a trending area of research that demands further exploration to unlock its full potential. (2) Most studies get one data point, e.g., final recovery, daily oil (gas) production, etc. They are not a complete production curve, so it is impossible to use machine learning for parameter optimization and production prediction. (3) Because of the complex physical properties of shale gas reservoirs, most research models tend to be idealized, which significantly weakens prediction accuracy. Geological conditions of shale gas reservoirs and historical production data should be needed to ensure the accuracy of prediction models. (4) In the above research investigation, the small data set resulted in an insufficient amount of data. When machine learning is then performed on the data, overfitting occurs, thus making the established prediction models perform poorly.

This study proposes a complete workflow for optimizing fracturing parameters in horizontal shale gas wells, combining reservoir numerical simulation with machine learning to generate a machine learning model and using particle swarm algorithm (PSO) to optimize fracture parameters. Section 2 describes the machine learning methods used in this study, and the workflow is illustrated. Section 3 develops the reservoir geological and numerical models for the target block. Section 4, a shale gas horizontal well-fracturing dataset is obtained by numerical simulation. A multi-factor sensitivity analysis is performed on the physical, completion, and fracturing parameters. Section 5 the performance of different ML-based yield prediction models was evaluated to compare the results of different models trained on the dataset to optimize the best ML-based capacity prediction model. In section 6 the prediction and optimization of capacity and parameters, respectively, through the PSO-GBDT parameter optimization process.

2. Methods

This section describes the methodological principles and the workflow of the main algorithms used in the study. The methods used include machine learning methods (Wang et al., 2023) (Linear Regression (Kavitha S et al., 2016; Lim, 2019), Support Vector Machines (SVM) (Cios et al., 2007; Vapnik, 1999), Decision Tree (DT) Regression (Wang and Xia, 2017), Gradient Boosting Decision Tree (GBDT) Regression, Random Forest (Brieman, 2001; Gamal et al., 2021)), and Particle Swarm Optimization (PSO).

To predict cumulative gas production, we initially employ machine learning techniques, specifically the gradient-boosted decision tree (GBDT) algorithm, which yields optimal results. Subsequently, we utilize the particle swarm optimization (PSO) algorithm to optimize both the fracturing parameters and the predicted cumulative production. Detailed explanations of the GBDT algorithm and the PSO algorithm are provided below.

2.1. GBDT regression

Machine learning is one of the most sophisticated and cutting-edge data processing research topics. Machine learning, in its broadest sense, is a process that allows a machine to learn and execute activities that cannot be accomplished by direct programming. Machine learning, in practice, is a process that takes a vast quantity of data as input, trains a model, and then utilizes the model to generate predictions. As a simulation of the human brain, the process of "training" and "prediction" is equivalent to the human brain's "induction" and "speculation" (Jordan and Mitchell, 2015). In this paper, we mainly use machine learning is GBDT regression.

MART (Multiple Additive Regression Tree), also known as GBDT, is an iterative decision tree technique that comprises multiple decision trees (He et al., 2014; Friedman, 2001). The learning process of the decision tree is to use the sample features of the training set to divide the data and obtain the predictions of the leaf nodes of each node (as in Fig. 1).

The model is also based on the Boosting algorithm. Each iteration produces a new decision tree in the direction of minimizing residuals and iterates constantly to enhance prediction accuracy.

GBDT is a member of the integrated learning boosting family; however, it is not the same as classic Adaboost (Tang et al., 2020). It is an optimization technique that employs an additive model and a forward distribution algorithm to achieve learning. The algorithm's primary phase is as follows: First, the base learner, a tree with only the root node, is initialized. Then M base learners are generated, and the current model's negative gradient value of the loss function is computed and utilized as an estimate of the residuals. The residual is then fitted with a regression tree CART. The fitted tree's leaf nodes are then searched for a value that minimizes the loss. Finally, the learner is brought up to date.

The input training sample set $T = \{(x_1,y_1),(x_2,y_2),(x_i,y_i),\cdots(x_n,y_n)\}$, $x_i \in X \subseteq R^n$, X is the input sample space, x_i is the evaluation metric, $y_i \in Y \subseteq R$, Y is the compliance case, the loss function is $L(y,f(x))$, and the output is the regression tree $\widehat{f}(x)$. The specific procedure of the GBDT algorithm is as follows.

(1) Initialize the estimation function so that the loss function is minimized.

$$f_0(x) = \arg\ min \sum_{i=1}^{N} L(y_i, c) \tag{1}$$

$f_0(x)$ is the tree with only one root node and $L(y_i, c)$ is the loss function, where c is the constant that minimizes the loss function.

(2) Let the number of iterations be m, and perform (A)-(D) when m ≤M, where (m = 1,2, …, M).
   (A) For sample $i = 1, 2, \ldots$, N, calculate the negative gradient of the loss function and use it as the residual estimate. Calculate the residual r_{mi}.

$$r_{mi} = -[\partial L(y_i, f(x_i))/\partial f(x_i)]_{f(x)=f_{m-1}(x)} \tag{2}$$

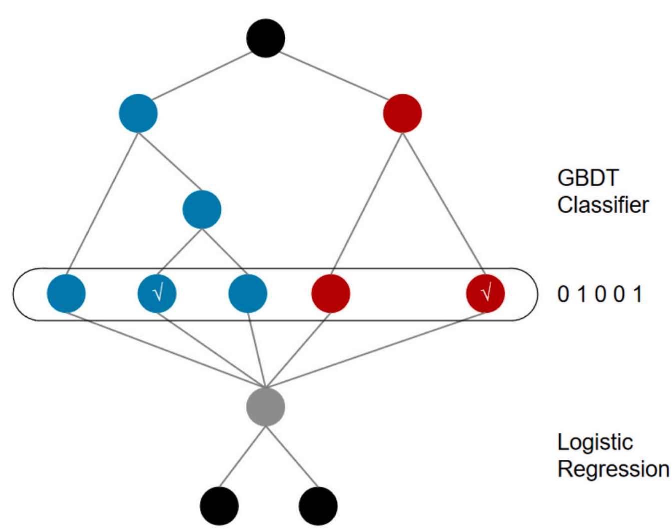


Fig. 1. Hybrid model structure. Each tree's output is handled as a categorical input feature to a sparse linear classifier. Boosted decision trees tend to be quite effective feature transformations.

(B) Fitting the residuals to r_{mj} generates a regression tree to estimate the regression tree leaf node region. The m-th tree node region $R_{mj}, j = 1, 2, \cdots, J$ is obtained.

(C) For $j = 1, 2, \cdots, J$, the value of the leaf node region is estimated using linear search to minimize the loss function.

$$C_{mj} = \arg\ min \sum_{x_i \in R_{mj}} (y_i - (f_{m-1}(x_i) + c))^2 \tag{3}$$

(D) Update Learner $f_m(x)$.

$$f_m(x) = f_{m-1}(x) + \sum_{m=1}^{M} c_{mj} I \quad (x \in R_{mj}) \tag{4}$$

(3) The final regression tree is obtained by accumulating all C_{mj} values in the same leaf node region.

$$\widehat{f}(x) = f_M(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj}\ I \quad (x \in R_{mj}) \tag{5}$$

2.2. Particle Swarm Optimization (PSO)

The particle swarm optimization was first proposed by Kennedy (Kennedy and Eberhart, 1995), an American psychologist, and Ebert Art, an electrical engineer, in 1995 as a new parallel metaheuristic algorithm. The algorithm simulates the mechanism of cooperation in the flock foraging behavior of organisms such as flocks of birds and fish in nature to find the optimal solution to the problem (Fernandez-Martinez et al., 2008).

The heuristic Algorithm is a problem-solving strategy that employs inductive reasoning and experimental investigation. The primary performance criteria for heuristic algorithms are generality, stability, and quick convergence.

The Meta-heuristic Algorithm is a heuristic algorithm modification created by merging a stochastic algorithm with a local search algorithm. Meta-heuristics is an iterative generating process that allows for studying and exploiting the search space with heuristic algorithms via the clever mixing of many notions. Learning tactics are utilized in this process to collect and master knowledge to locate near-optimal solutions effectively.

PSO is a widely used algorithm known for its simplicity and ease of implementation. It does not require gradient information, making it applicable to a range of optimization problems. Its key strength lies in its ability to explore the global search space effectively, finding optimal solutions in complex domains.

In practical applications, PSO is used in function optimization, neural network training, image processing, robot control, and aerospace. In reservoir engineering, PSO has shown its usefulness in parameter optimization, such as determining optimal values for reservoir properties and well operating parameters. It can also be applied to reservoir prediction tasks, helping to estimate future reservoir behavior based on historical data. Additionally, PSO has been employed in well network optimization, optimizing the placement and configuration of wells to maximize hydrocarbon recovery from the reservoir.

Overall, PSO is a versatile algorithm with broad applications in various domains, including reservoir engineering. It offers effective solutions to complex optimization problems when used appropriately and with proper parameter tuning.

In Fig. 2, we present the flow chart of PSO.

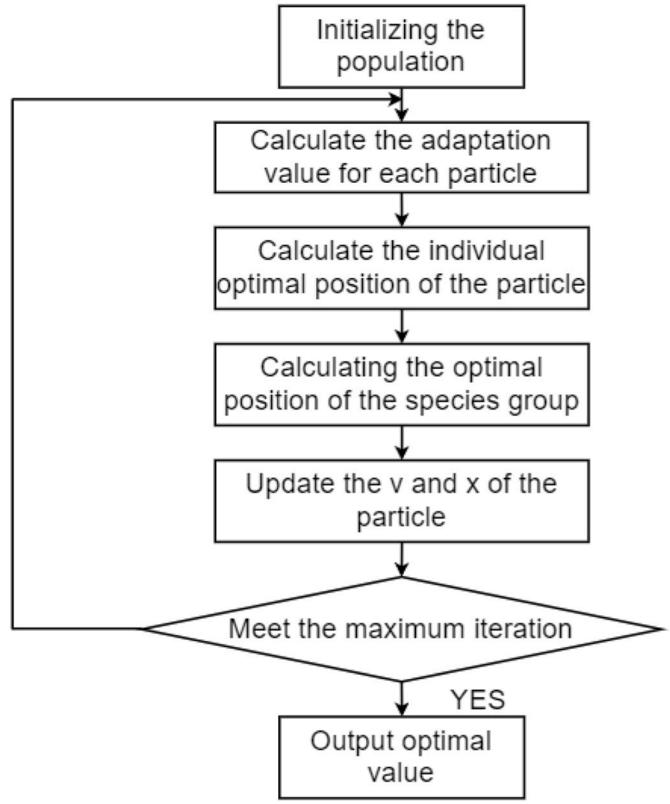


Fig. 2. Flow chart of PSO.

2.3. Data dimensionality reduction

In shale gas reservoirs, the production of fractured horizontal wells is influenced by a multitude of parameters, each with varying magnitudes. These parameters play a crucial role in constructing accurate production prediction models and analyzing data effectively. To ensure the comparability of these parameters, it is necessary to standardize the original data. By applying data standardization techniques, we can eliminate the impact of varying magnitudes among the parameters, facilitating more reliable and meaningful analysis.

In the field of data standardization, two commonly employed methods are Z-score normalization (Jain et al., 2005; Zou et al., 2020) and Min-max normalization (Kim et al., 2021). Z-score normalization, also known as standardization, is a widely recognized data standardization technique. It involves transforming raw data into a standard normal distribution with a mean of 0 and a standard deviation of 1. This method is useful as it eliminates the scale differences among different variables in the dataset, making the data comparable and facilitating meaningful analysis.

By converting the data into a standard normal distribution, the Z-score normalization technique provides several advantages. Firstly, it enables the identification of outliers, which are data points that deviate significantly from the average. These outliers can be easily detected as they fall outside the typical range of values within the standardized distribution. Secondly, standardization allows for more effective data analysis, modeling, and comparison. Since the data now follow a standard distribution, statistical measures such as mean, standard deviation, and correlation coefficients can be readily calculated and interpreted. This, in turn, enhances the reliability and accuracy of any subsequent analyses or modeling performed on the standardized data.

The Z-score normalization is calculated as follows:

1. Calculate the mean and standard deviation of the original data,
2. For each data point, the Z-score is calculated using equation (6), where X is the original data point:

$$Z = \frac{(X - mean)}{\text{standard deviation}} \tag{6a}$$

3. The obtained Z-score value indicates the degree of deviation from the mean for each data point, in units of standard deviation. A positive value indicates that the data point is above the mean and a negative value indicates that the data point is below the mean.

However, when processing data using z-score normalization, it is required that the data needs to satisfy the condition of normal distribution. The parameter data set of this subject does not satisfy the normal distribution, so this subject uses the Min-max normalization method to process the data, and the values of all parameters are mapped uniformly to the interval [0, 1]. The Min-max normalization formula is shown in Equation (7).

$$X_{new} = \frac{X - min(X)}{max(X) - min(X)} \tag{7}$$

2.4. Workflow

As illustrated in Fig. 3, there are four processes in our forecast of the shale gas resources' production performance and parameter optimization.

Step 1: Numerical model building. We obtained an overview of the current status of the application of artificial intelligence in fracturing horizontal wells for shale gas extraction through extensive literature research. Using the reservoir numerical simulation program CMG software, single-well geology and numerical models were created for the Fuling maritime Longmaxi Formation's shale gas reservoir characteristics.

Step 2: Database generation. The 10,000 sets of varying reservoir models are randomly generated and simulated using the Monte Carlo method to obtain the corresponding 10,000 sets of production curves for the orientation of different reservoir physical parameters, completion parameters, and fracturing parameters. Multi-factor sensitivity analysis is performed on these parameters.

Step 3: Machine learning model building and application. The data set from Step 2 is used as the input for the machine learning model with 10,000 different sets of geological, completion, and fracturing parameters and 10,000 different sets of production curves as the output. The best method (GBDT algorithm) is then chosen by training the data set with various machine learning regression techniques.

Step 4: Parameter optimization and prediction. It is possible to

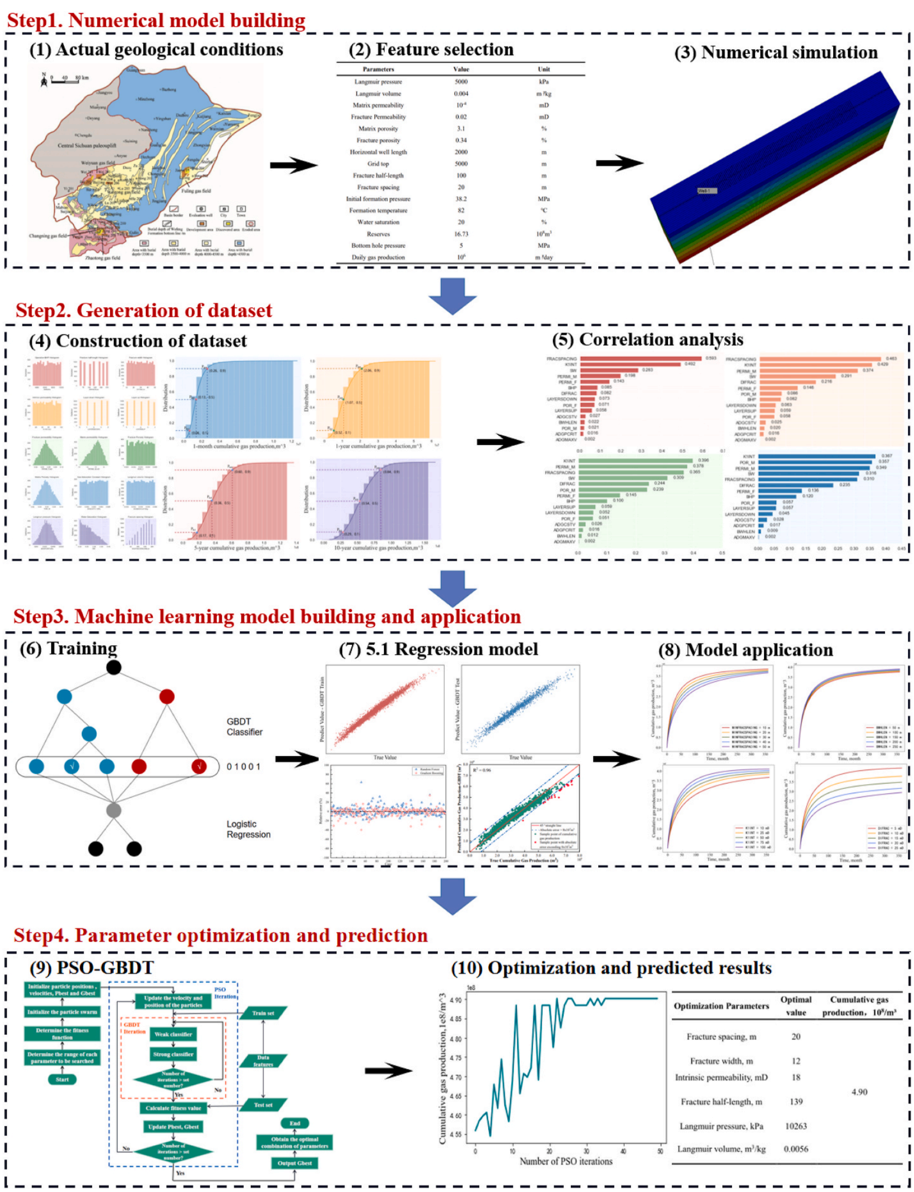


Fig. 3. Workflow of shale gas parameter optimization based on artificial intelligence algorithm.

quickly optimize the fracturing parameters and predict the production for each period based on different geological conditions by combining a production prediction model with a particle swarm optimizer (PSO).

3. Numerical model building

3.1. Regional geological overview

The Sichuan Basin is a vast hydrocarbon-bearing superimposed basin formed from the Upper Yangzi Craton. It is currently encircled by tectonic mountain belts like the Longmen Mountains, the Mixing Mountains, and the Daba Mountains, covering an area of approximately 18×10^4 km^2. The sedimentary rocks of the Sichuan Basin have grown to a thickness of 12,000 m. It is the basin in China with the greatest concentration of minerals, thanks to the formation of nine groups of meridional rocks (Dai et al., 2021). The Late Ordovician–Early Silurian Upper Ordovician Wufeng Formation–Lower Silurian Longmaxi Formation, which was heavily deposited in the Sichuan Basin and its surrounding regions under the influence of tectonic and marine erosion, is rich in penstock fossils. The Wulong Formation-Longmaxi Formation is dominated by siliceous, clayey, calcareous, and silt shales. The high-quality shales at the bottom of the Wubong Formation and a sub-section of the Longmaxi Formation are characterized by thin layers, carbon-rich, silica-rich, deep water, and low deposition rate, which are the core formations for shale gas development and exploration (Fig. 4) (He et al., 2017; Wang et al., 2020).

3.2. Numerical model building for shale gas reservoirs

Shale reservoirs have extremely low permeability and microfractures with different degrees of development. The overall performance is characterized by dual pore, so a dual pore dual permeability model is usually established to simulate the shale numerically. A single-well numerical model of a shale multi-stage fractured horizontal well is built using the geological characteristics of the Upper Ordovician Wufeng Formation reservoir and the engineering parameters of horizontal wells in the Sichuan Basin and its outlying areas.

The grid number is $300 \times 29 \times 9$, as shown in Fig. 5. The numerical model is 3000m (300×10) long, 440m ($\sum(10 \times 20, 10, 8.5, 1, 0.4, 0.2, 0.4, 1, 8.5, 10, 20 \times 10)$) wide and 45m ($9 \times 5$) deep. The reservoir type is dual pore dual permeability, and the total geological reserves of shale gas are 16.73×10^8 m^3, including 5.04×10^8 m^3 of free gas reserves, and the geological, construction, and fracture parameters are shown in Table 1.

4. Generation of dataset

4.1. Construction of dataset

After building the geological model, many datasets are required to be generated to train the predictive model built using machine learning. This study uses a numerical model to randomly generate cumulative production data for a set of 10,000 geological and completion parameters using a Monte Carlo approach. In this study, 15 geological and completion parameters were studied, including matrix permeability,

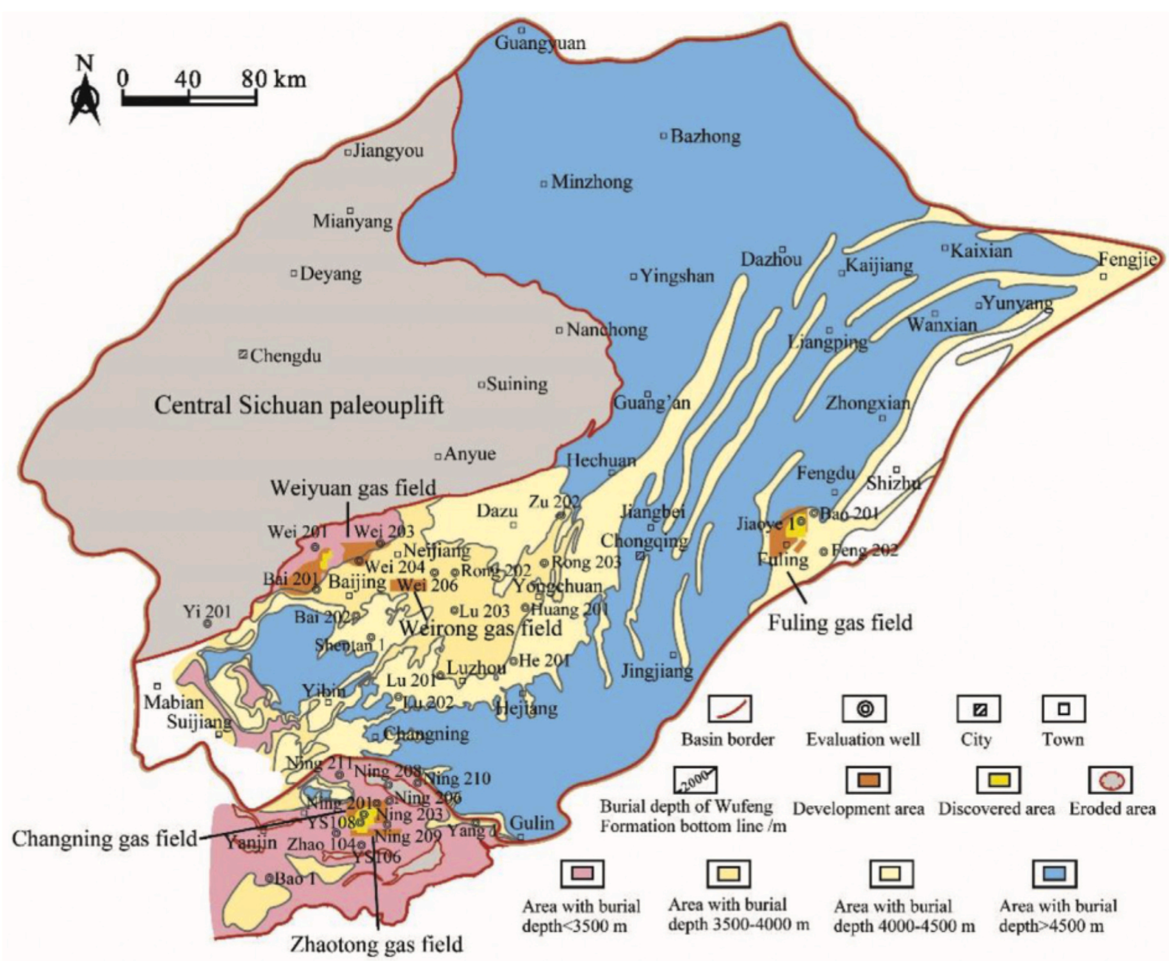


Fig. 4. Distribution map of shale gas fields in the Sichuan Basin and its southern edge, Wufeng Formation-Longmaxi Formation (Dai et al., 2020).

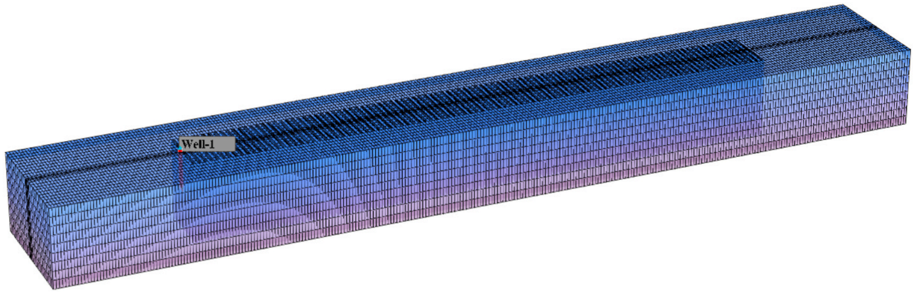


Fig. 5. Numerical modeling of shale gas multi-stage fractured horizontal wells.

Table 1
Basic parameters of the shale gas single well numerical model.

Parameters	Value	Unit
Langmuir pressure	5000	kPa
Langmuir volume	0.004	m^3/kg
Matrix permeability	10^{-4}	mD
Fracture Permeability	0.02	mD
Matrix porosity	3.1	%
Fracture porosity	0.34	%
Horizontal well length	2000	m
Grid top	5000	m
Fracture half-length	100	m
Fracture spacing	20	m
Initial formation pressure	38.2	MPa
Formation temperature	82	°C
Water saturation	20	%
Reserves	16.73	$10^8 m^3$
Bottom hole pressure	5	MPa
Daily gas production	10^6	m^3/day

fracture permeability, matrix porosity, fracture porosity, water saturation, upper, lower, bottom hole pressure, native permeability, fracture half-length, fracture spacing, fracture width, Lang pressure, Lang

Table 2
Parameters and associated distribution to study.

Parameters	Minimum value	Maximum value	Distribution type	Symbol in Fig. 6
Matrix permeability, mD	0.000001	0.001	lognormal	PERMI_M
Fracture permeability, mD	0.001	0.1	Triangle	PERMI_F
Matrix Porosity	0.02	0.1	Triangle	POR_M
Fracture Porosity	0.005	0.001	Uniform	POR_F
Water Saturation	0.2	0.6	Uniform	S_W
Layer-up	0	4	Uniform	LAYERSUP
Layer-down	0	4	Uniform	LAYERSDOWN
Operation BHP, kPa	3000	10000	Uniform	BHP
Intrinsic permeability, mD	10	100	Uniform	K1INT
Fracture half-length, m	60	180	Uniform	BWHLEN
Fracture spacing, m	20	220	Uniform	FRACSPACING
Fracture width, m	10	50	Uniform	DIFRAC
Langmuir pressure, kPa	3000	20000	Triangle	ADGPCRIT
Langmuir volume, m^3/kg	5×10^{-7}	0.008	Uniform	ADGMAXV
Gas Adsorption Constant, 1/kPa	0.001	0.005	Uniform	ADGCSTV

volume, gas adsorption constants (Table 2). And the probability distribution of each parameter is shown in Fig. 6.

The applicable range of each parameter in the table is mainly set according to the practical situation of Chinese shale gas reservoirs. Among these 15 parameters, gas reservoir parameters: matrix permeability and porosity, fracture permeability and porosity, Langmuir pressure, Langmuir volume and gas adsorption constant, water saturation; fracturing parameters: intrinsic permeability, fracture spacing, fracture half-length, fracture width, upper and lower layers; and working regime: BHP.

The cumulative shale gas production is the output parameter employed in this study, which is simulated for each combination corresponding to the monthly shale gas production estimated using numerical simulation model simulations. Fig. 7 depicts the cumulative gas production distribution over one month, one year, five years, and ten years. And is explained in Table 3.

4.2. Correlation analysis

The correlation analysis of various parameters on cumulative gas production for a one-month production period is presented in Fig. 8(a). It can be seen that among these parameters, Matrix Permeability, Water Saturation, Intrinsic Permeability, and Fracture Spacing correlation coefficients of 0.198, 0.283, 0.492, and 0.593, respectively, have a strong correlation on the cumulative gas production compared to the other parameters. While Matrix Porosity, Fracture Half-length, Langmuir Pressure, Langmuir Volume, and Gas Adsorption Constant have weak correlations. Fig. 8(b) demonstrates that Matrix Permeability, Water Saturation, Intrinsic Permeability, and Fracture Spacing still maintain strong correlations with cumulative gas production, while the correlation value for DIFRAC changes from 0.082 to 0.216, indicating a transition from weak to strong correlation as production time increases. In Fig. 8(c), Matrix Permeability, Fracture Permeability, Matrix Porosity, Water Saturation, Intrinsic Permeability, Fracture Spacing, and Fracture Width display strong correlations. The correlation value for Matrix Permeability increases, while Fracture Spacing exhibits a slight decrease compared to the previous plot. Fig. 8(d) shows that the correlation value for Matrix Permeability changes from 0.021 to 0.357, while Fracture Spacing decreases from 0.593 to 0.310.

Overall, among the parameters to be optimized, Intrinsic Permeability, Fracture Half-length, Fracture Spacing, Fracture Width, Langmuir Pressure, and Langmuir Volume, the strong correlations are Intrinsic Permeability, Fracture Spacing, and Fracture Width. At the beginning of production, without considering geological conditions, the main influencing factors for cumulative gas production are Intrinsic Permeability and Fracture Spacing; at the middle and late production, the main influencing factors of cumulative gas production are Intrinsic Permeability, Fracture Spacing, and Fracture Width.

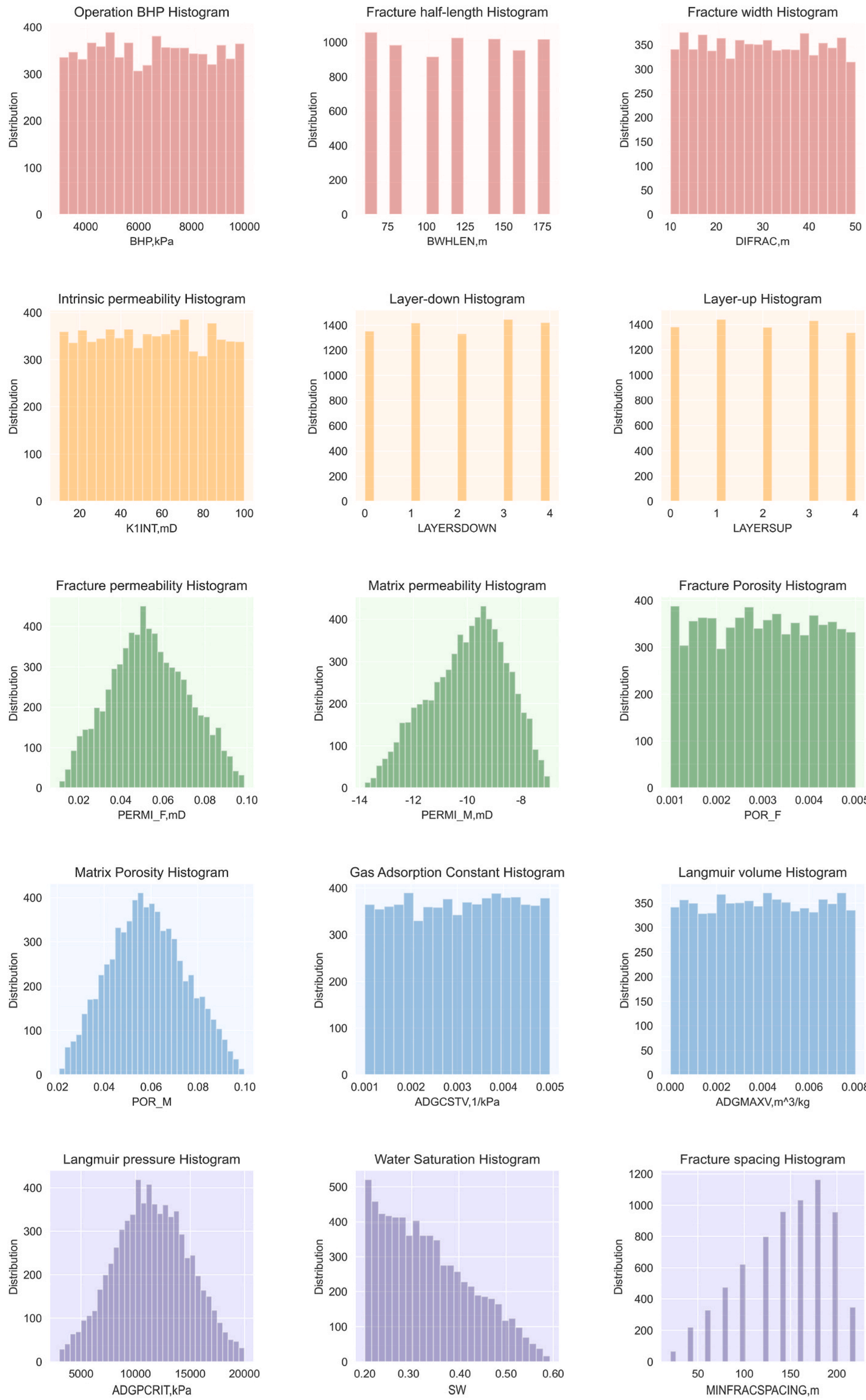


Fig. 6. Histograms of geological and completion parameters generated.

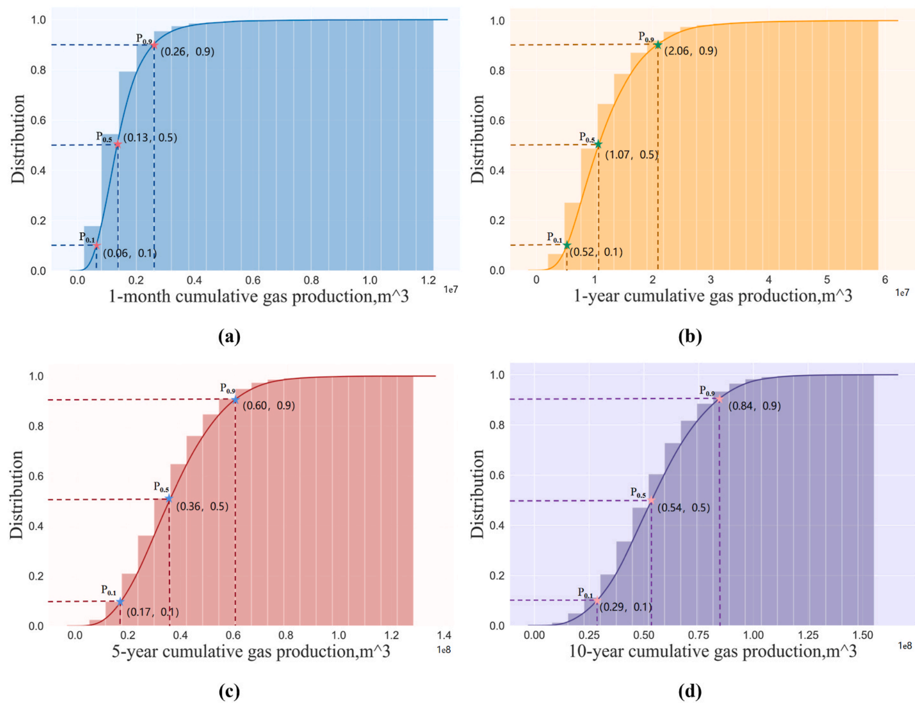
5. Machine learning model and application

5.1. Regression model

In this study, 75% of the dataset obtained in Section 4 was utilized as the training set, while 25% of the dataset was used as the test set. Five machine learning prediction models were developed: Linear Regression, Decision Tree (DT) Regression, Gradient Boosting (GBDT) Regression, Support Vector Machine (SVM), and Random Forest. The datasets were also processed, and the effects of the different models were evaluated by R^2.

The correlation index (R^2) was chosen as a criterion to assess the prediction accuracy of the neural network model developed in this study (Ottah et al., 2015). The value of R^2 runs from 0 to 1, and the higher the number, the better the model fit. The formula for calculating R^2 is as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{K} (\widehat{y_i} - y_i)^2}{\sum_{i=1}^{K} (\overline{y_i} - y_i)^2} \tag{6b}$$



(a) (b) (c) (d)

Fig. 7. Distribution of different times of cumulative gas production calculated by numerical simulation.

Table 3
Cumulative gas production distribution table at different times.

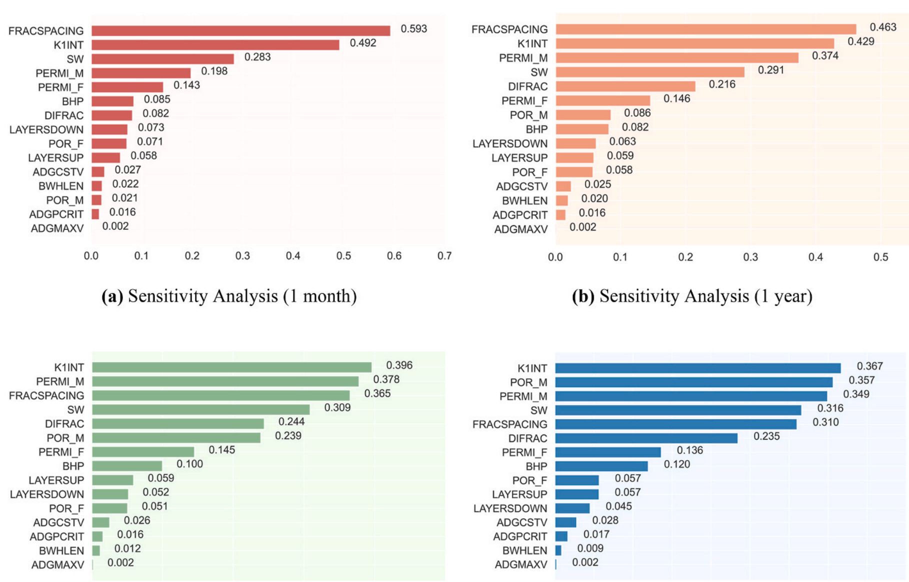
Cumulative gas production ($\times$ 10^7m^3)	P = 0.1	P = 0.5	P = 0.9
1-month	0.06	0.13	0.26
1-year	0.52	1.07	2.06
5-year	1.70	3.60	6.00
10-year	2.90	5.40	8.40

Note: When P = 0.1, P = 0.5 and P = 0.9, it indicates that with a 10%, 50% and 90% probability.

Where, $\overline{y_i}$ is the average value of y_i.

Fig. 9 shows the scatter plots of the prediction results and the actual results of the five prediction models studied in this study. The correlation index (R^2) results of the selected prediction models are shown in Table 4. The linear regression model, the support vector machine model, and the decision tree moderator model have scattered predicted and actual value points, low R^2 values, large errors, and poor prediction results.

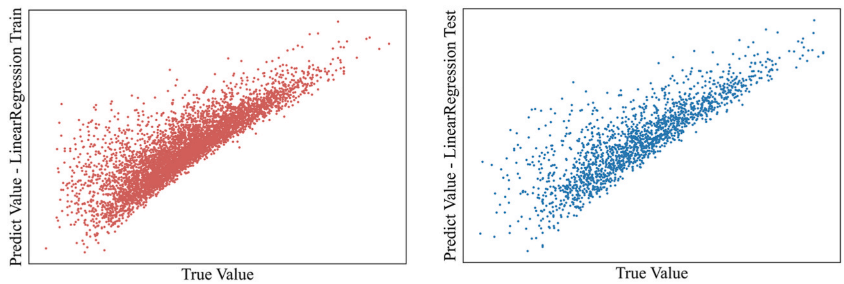
Fig. 10 shows the relative errors of the prediction results of 200 randomly selected validation cases combined with Table 4. It can be seen that Gradient Boosting and Random Forest both have R^2 over 0.9 with more minor errors compared to the other three machine learning



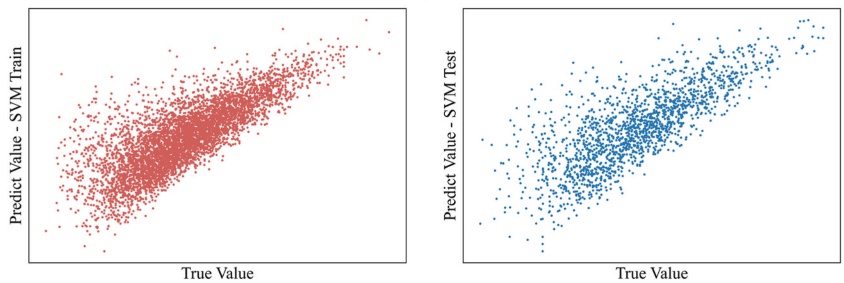
(a) Sensitivity Analysis (1 month) **(b)** Sensitivity Analysis (1 year)

(c) Sensitivity Analysis (5 years) **(d)** Sensitivity Analysis (10 years)

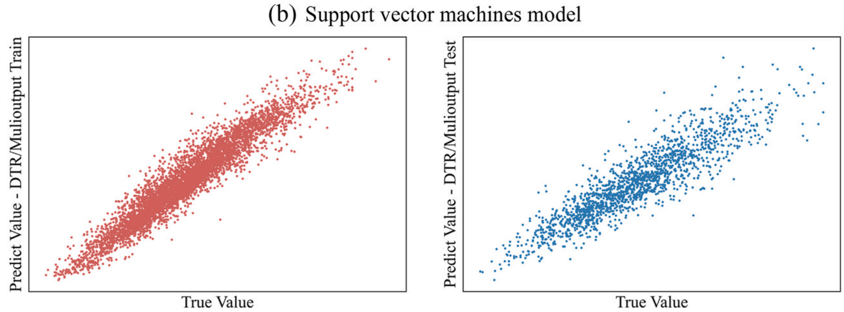
Fig. 8. Histogram of sensitivity of each parameter to capacity for 1 month, 1 year, 5 years, and 10 years.



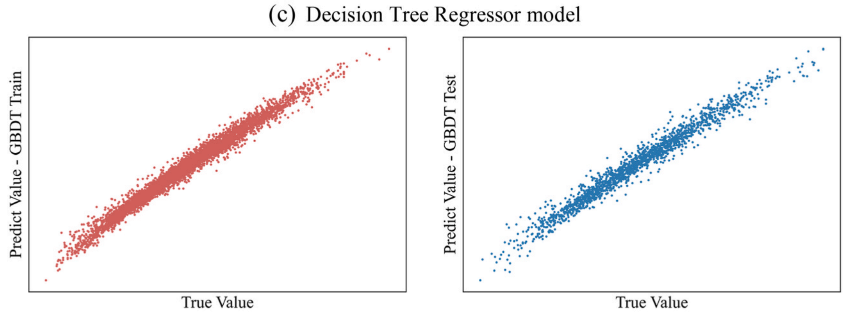
(a) Linear Regression model



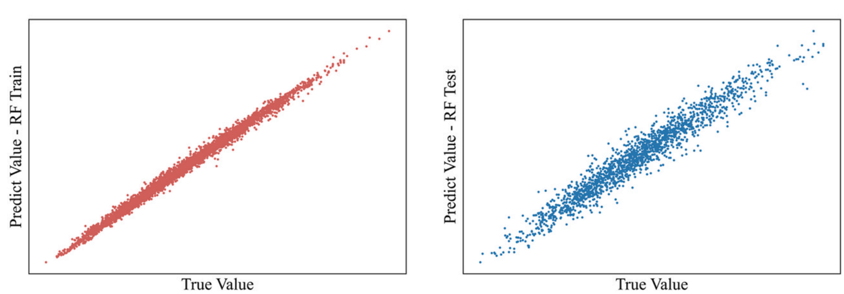
(b) Support vector machines model



(c) Decision Tree Regressor model



(d) Gradient Boosting Regression model



(e) Random Forest Regressor model

Fig. 9. Model performance of each model on the training and test set.

models. The results obtained are more concentrated around the x-axis, and the error distribution range is much smaller than the other methods, which is far superior to the other models.

Therefore, to filter out more suitable ML prediction models from gradient boosting and random forest, we compared their predicted values from machine learning with the actual values from numerical simulations for one month, one year, five years, and ten years, respectively (as shown in Fig. 11). The horizontal coordinates in Fig. 11

Table 4
Comparison of the prediction performance of various models.

Machine Learning Algorithms	Train R^2	Test R^2
Linear	0.69	0.69
SVM	0.55	0.54
Decision Tree Regressor	0.90	0.74
Gradient Boosting (GBDT)	0.97	0.96
Random Forest	0.98	0.93

represent the actual cumulative gas production of distinct samples, while the vertical coordinates represent the anticipated values of the Gradient Boosting (GBDT) and Random Forest algorithms, respectively. The lesser the discrepancy between the model prediction and the actual sample, the closer the sample points are to the red 45-degree line. Most of the points plotted in Fig. 11 are distributed near the 45-degree line. However, whether the prediction is for one month, one year, five years, or ten years, the prediction of Gradient Boosting has a higher aggregation than that of Random Forest. The calculated Gradient Boosting's R^2 is larger than those of Random Forest, indicating that the prediction performance of Gradient Boosting is better than that of Random Forest.

5.2. Model application

This work analyzes the sensitivity analysis of different fracturing parameters on cumulative shale gas production to evaluate the influence of different fracturing parameters on cumulative gas production, as illustrated in Fig. 12.

Before performing the sensitivity analysis on the cumulative gas production, a set of basic groups with parameters must be set, as shown in Table 5. Each parameter is set with 5 different sets of values from small to large in the distribution range of Table 2, respectively.

Fig.(a) demonstrates that cumulative gas production generally increases with decreasing fracture spacing, with a notable increase occurring early in the mining process. It indicates that fracture spacing is strongly sensitive to cumulative gas production and is negatively correlated.

According to Fig.(b), the cumulative gas production increases as the fracture half-length increases, but this rise is less pronounced in the parameter distribution range. It suggests a positive correlation and that the fracture half-length is less susceptible to total gas output.

According to Fig.(c), the cumulative gas production grows dramatically with increasing intrinsic permeability, peaks at the beginning of extraction, and is proportional to intrinsic permeability. It suggests that

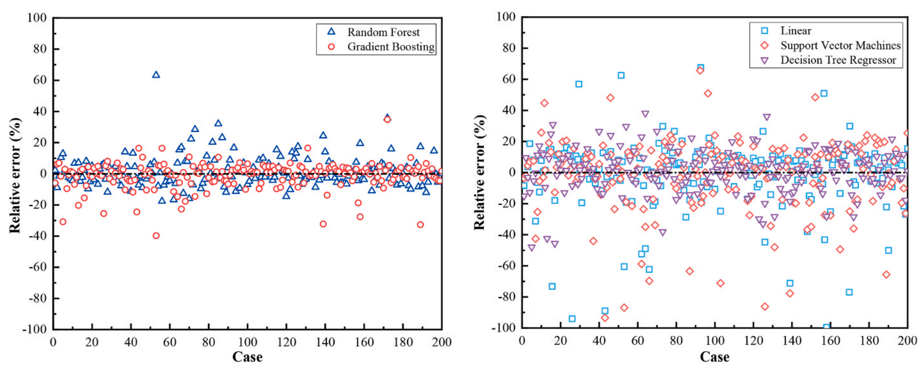


Fig. 10. Relative error of prediction results for selected 200 verification cases.

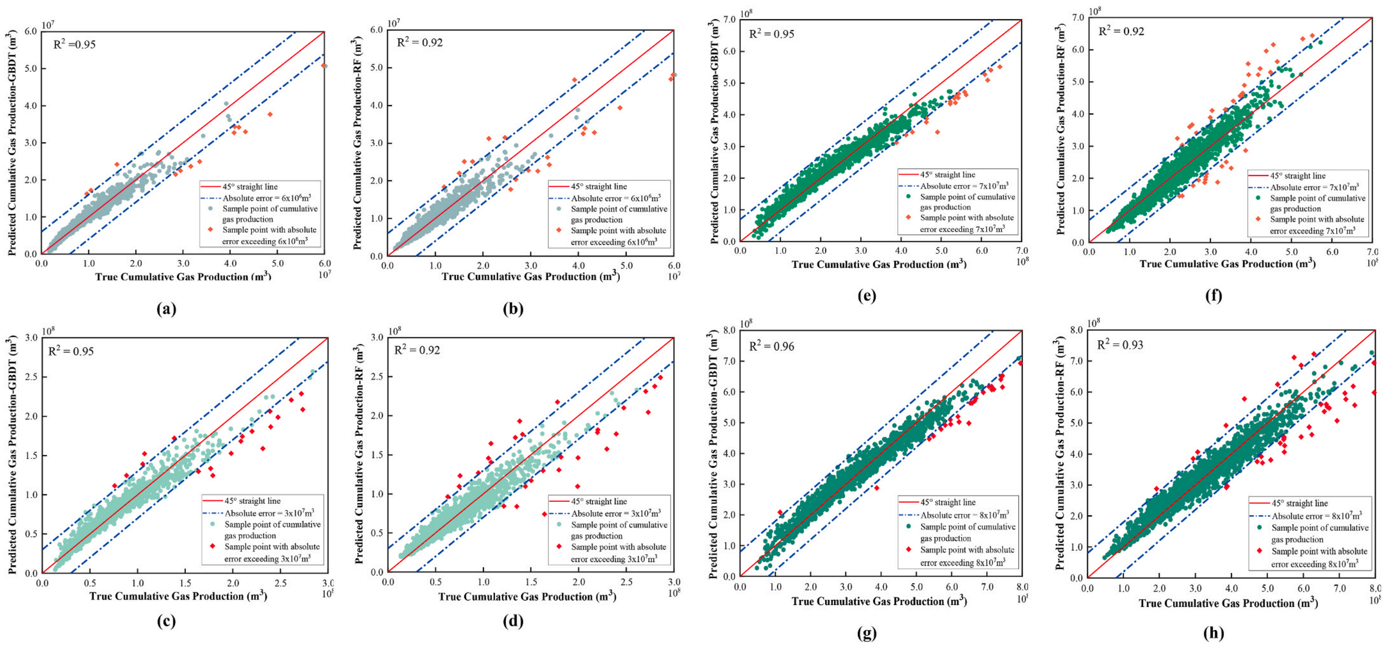


Fig. 11. Comparisons of the true values and the predicted cumulative gas productions, Gradient Boosting (GBDT) is shown on the left of the figure, and Random Forest is shown on the right. **(a, b)** for 1 month of production, **(c, d)** for 1 year of production, **(e, f)** for 5 years of production, and **(g, h)** for 10 years of production.

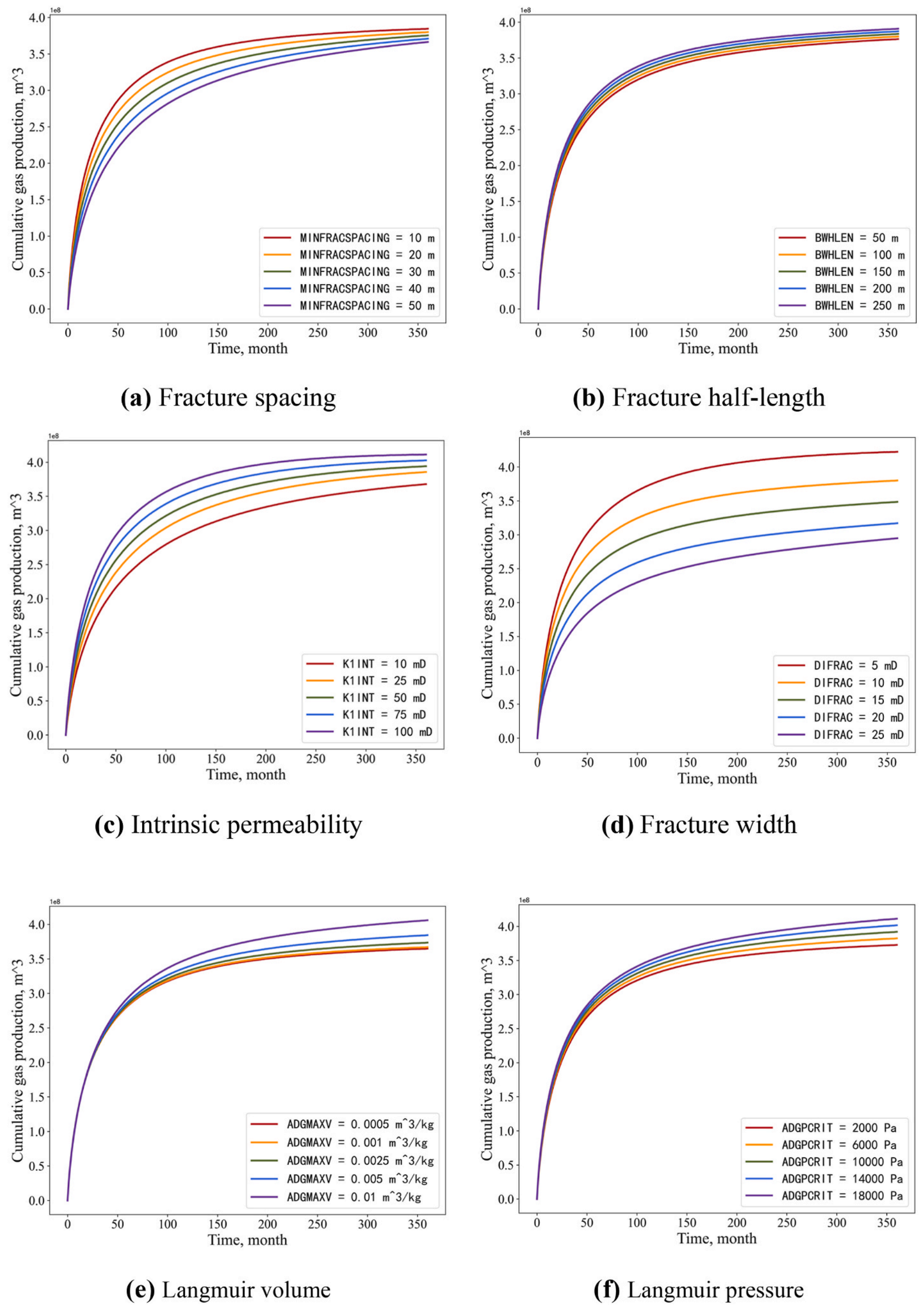


Fig. 12. Sensitivity analysis of cumulative gas production to fracturing parameters.

Table 5
Basic case parameters and values.

Parameters	value	Parameters	value
Matrix permeability, mD	10^{-4}	Fracture permeability, mD	0.02
Matrix Porosity,%	3.1	Fracture Porosity,%	0.34
Water Saturation,%	20	Operation BHP, kPa	5000
Layer-up	4	Intrinsic permeability, mD	30
Layer-down	4	Fracture half-length, m	100
Fracture spacing, m	20	Langmuir pressure, kPa	5000
Fracture width, m	10	Langmuir volume, m^3/kg	0.004
Gas Adsorption Constant, 1/kPa	0.0035		

intrinsic permeability and cumulative gas production have a positive correlation and are extremely sensitive to each other.

The cumulative gas production grows as the fracture width increases in Fig. (d), and the fracture width significantly increases the ultimate cumulative gas production of gas wells. It shows a positive correlation between fracture width and cumulative gas output.

Fig.(e) shows that Langmuir volume positively correlates with cumulative gas production. But, when Langmuir volume is generally at a low level, the contribution of increasing Langmuir volume to gas well production is minimal—indicating that the sensitivity of Langmuir volume to gas well production is weak. However, when Langmuir volume increases to a certain level, a further increase in Langmuir volume significantly improves gas well production. As the Langmuir volume increases, the proportion of adsorbed gas in the matrix increases, although the free gas content does not change. With 30 years of cumulative production, the adsorbed gas recovery is increasing as the free gas is continuously recovered.

According to Fig.(f), the cumulative gas production of gas wells increases comparatively significantly as Langmuir pressure rises. It suggests that the Langmuir pressure is somewhat sensitive to gas well production. This is mainly because improved gas well production results from quicker desorption of adsorbed gas from shale reservoirs when the Langmuir pressure constant increases. The figure also shows that at the beginning of gas well extraction, the contribution of the Langmuir pressure constant to gas production is not immediately apparent. This is primarily because during the initial stages of gas well extraction, and free gas is extracted first, then adsorbed gas is desorption. As a result, the Langmuir pressure constant primarily affects the intermediate and final stages of gas reservoir extraction.

6. Parameter optimization and prediction

6.1. Coupling of PSO optimization and GBDT regression model

The Particle Swarm Optimization protocol (PSO), a well-known meta-heuristic global optimizer, to determine the optimal design of fracturing parameters. The PSO algorithm is combined with a trained Gradient Boosting Decision Tree (GBDT) model, which serves as a fitness evaluator for a large set of project design parameters. By employing the GBDT model, the computational burden of the optimization process is significantly reduced, allowing for a larger number of PSO iterations.

To provide a comprehensive understanding, Table 6 presents the final parameters of the PSO model, including the GBDT parameters. Additionally, Fig. 13 illustrates the flow of the optimization process.

By integrating the PSO algorithm with the GBDT model, we aim to achieve an efficient and effective optimization of the fracturing parameter design. This approach enables us to explore a wide range of design possibilities and identify the most favorable parameter configurations.

6.2. Optimization and predicted results

Through the sensitivity analysis of cumulative gas production, various fracturing parameters and gas reservoir parameters affect the cumulative gas production when fracturing horizontal wells for gas recovery from shale. Therefore, in this study, the PSO algorithm is used to optimize the fracturing parameters among the input parameters with the objective of optimal cumulative gas production. Fig. 14 shows the trend of cumulative gas production during the PSO optimization. The cumulative gas production progressively rises and eventually reaches a plateau as the number of iterations rises. This indicates that the cumulative gas production is closer to the optimal result through PSO iterations.

Based on the specified basic parameters and using cumulative gas production as the objective function, the optimization results for fracturing parameters and gas reservoir parameters are presented in Table 7. The optimization process led to several changes. The fracture spacing remained unchanged, but the fracture width was optimized from 10m to 12m. The intrinsic permeability was reduced from 30mD to 18mD, while the base value of fracture half-length increased from 100m to 139m after optimization. Regarding the gas reservoir parameters, the initial values for Langmuir pressure and Langmuir volume were 5000 kPa and 0.004 m^3/kg, respectively. Through optimization, these values were adjusted to 10263 kPa and 0.0056 m^3/kg, respectively. The machine learning algorithm predicted a cumulative gas production volume of 4.59×10^8 m^3. However, by employing the PSO algorithm to optimize the parameters, the optimal cumulative gas production volume was achieved at 4.90×10^8 m^3. Thus, through parameter optimization using the PSO algorithm, the optimal cumulative gas production was significantly improved.

Table 6
Hyperparameters of PSO algorithm and GBDT model.

Parameters in PSO algorithm	Value	Parameters of GBDT model	Value
Population number group size	15	Loss	deviance
Maximum number of iterations maximum	50	Learning_rate	0.1
Inertia weight(ω)	0.8	max_depth	3
Learning factor (c_1)	2	Criterion	gini
Learning factor (c_2)	2	n_estimators	100

In summary, by applying the PSO algorithm for parameter optimization, the fracturing and reservoir parameters were fine-tuned, resulting in improved cumulative gas production. The optimized values demonstrate the effectiveness of the PSO algorithm in achieving better production outcomes in gas reservoirs.

7. Discussion

Machine learning has become a widely adopted research methodology for data processing, and the predictive model employed in this study holds immense potential for further refinement. In this investigation, we have leveraged the Gradient Boosting Decision Tree (GBDT) model, continuously fine-tuning crucial aspects such as the number of hidden layers, neuron count, learning rate, and other parameters to achieve optimal performance. By iteratively adjusting these hyperparameters, we aim to enhance the model's prediction accuracy and achieve more reliable results.

To bolster the predictive capabilities of the GBDT model, we can incorporate advanced optimization techniques like Particle Swarm Optimization (PSO) or Genetic Algorithms (GA) to optimize the hyperparameters. This integration of state-of-the-art optimization methods aids in identifying the most favorable parameter settings, thereby elevating the model's overall performance and predictive power.

Furthermore, it is important to acknowledge that the evaluation and optimization undertaken in this study have primarily focused on a single objective function, specifically the cumulative gas production. However, future research endeavors should consider employing a multi-objective optimization approach, such as the Pareto method. By incorporating additional functions, such as cumulative gas production and net present value (NPV), a more comprehensive and robust assessment of the reservoir's performance and economic viability can be attained.

Through a continuous cycle of refinement, exploration of cutting-edge optimization techniques, and integration of multi-objective optimization strategies, our aim as reservoir engineers is to push the boundaries of machine learning applications in reservoir engineering. By doing so, we aspire to elevate the precision and effectiveness of decision-making processes, ultimately leading to the maximization of reservoir performance and economic success.

8. Conclusion

This study proposes a cumulative gas production evaluation workflow based on coupled particle swarm optimization and the GBDT model for fracturing parameter optimization for capacity prediction and fracturing parameter optimization, leading to the following conclusions.

(1) Machine learning methods can efficiently process field data and solve non-linear problems compared to traditional simulation and prediction methods. And synthesize a variety of factors, such as geology, fracturing construction, and production, which significantly improves the model's efficiency and prediction accuracy.
(2) Correlation and sensitivity analyses reveal that intrinsic permeability and fracture width are the dominant factors influencing cumulative gas production in shale gas reservoirs. These parameters exhibit strong correlations with cumulative gas production, with correlation coefficients of 0.367 and 0.316, respectively, over a 10-year production period. On the other hand, the impact of fracture half-length on cumulative gas production is found to be minimal, as indicated by a low correlation coefficient of only 0.009 after 10 years of production.

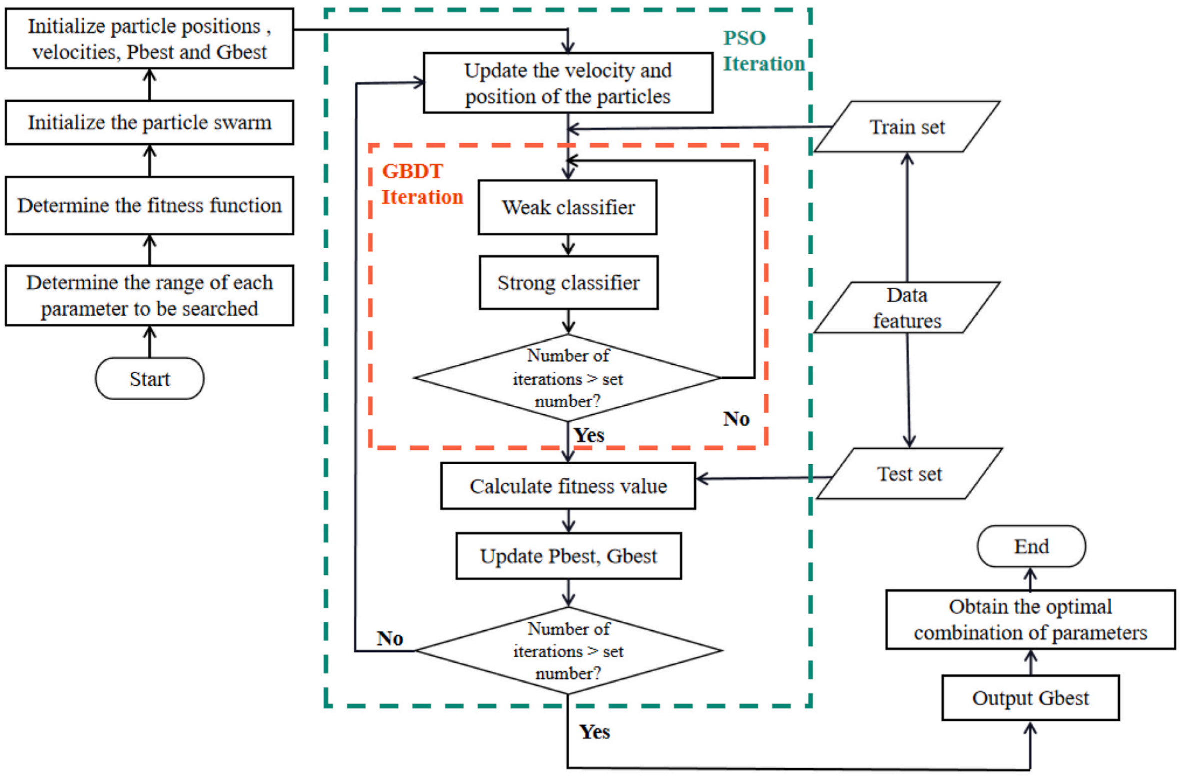


Fig. 13. Optimization workflow.

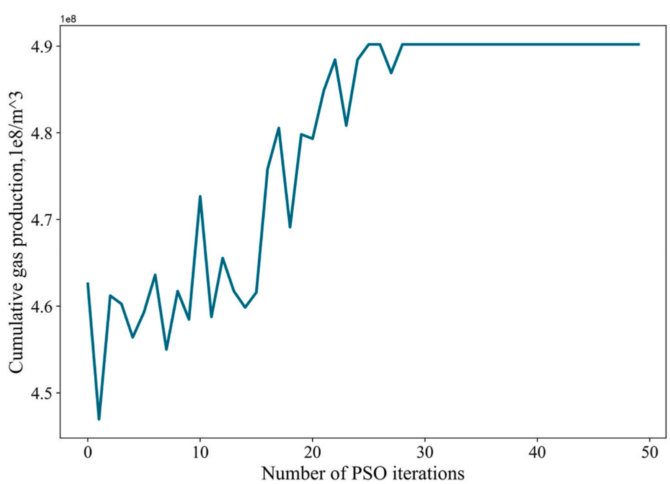


Fig. 14. The trend of Cumulative gas production during PSO iteration.

Table 7
Optimal parameters.

Basic Parameters	value	Optimization Parameters	value	Optimal value
Matrix permeability, mD	10^{-4}	Fracture spacing, m	20	20
Matrix Porosity,%	3.1	Fracture width, m	10	12
Water Saturation,%	20	Intrinsic permeability, mD	30	18
Layer-up	4	Fracture half-length, m	100	139
Layer-down	4	Langmuir pressure, kPa	5000	10263
Fracture permeability, mD	0.02	Langmuir volume, m^3/kg	0.004	0.0056
Fracture Porosity,%	0.34	Cumulative gas production, $10^8/m^3$	4.59	4.90
Operation BHP, kPa	5000			
Gas Adsorption Constant, 1/kPa	0.0035			

(3) Different machine learning models have other production prediction effects. By comparing five production prediction models, we concluded that GBDT and Random Forest yielded R^2 of 0.96 and 0.93, respectively, and they predicted models with better prediction performance than Linear, SVM, and Decision Tree Regressor. Through more in-depth testing and training, the final study concluded that the GBDT model was used to predict single well capacity and evaluate fracturing effectiveness.

(4) The combination of PSO and the trained GBDT model enables efficient optimization of fracturing parameters. PSO effectively identifies the optimal fracturing parameters that maximize cumulative gas production within the search domain, resulting in significant computational savings. PSO successfully optimized various parameters, including fracture spacing, fracture width, intrinsic permeability, fracture half-length, Langmuir pressure, and Langmuir volume. As a result of the optimization, the cumulative gas production, which was initially predicted to be 4.59 $\times 10^8$ m^3, was improved to 4.90 $\times 10^8$ m^3.

Declaration of interests

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.

CRediT authorship contribution statement

Shihao Qian: Data curation, Visualization, Writing – original draft, Methodology, Software. **Zhenzhen Dong:** Conceptualization, Writing – review & editing. **Wei Guo:** Project administration. **Xiaowei Zhang:** Project administration. **Zhaoxia Liu:** Visualization. **Lingjun Wang:** Software. **Lei Wu:** Software. **Tianyang Zhang:** Resources. **Weirong Li:** Supervision.

Acknowledgment

Special thanks to Professors for their careful guidance on the selection, collection, and writing of this thesis to its final draft; to fellow lab members for their essential technical help; and to Xi'an Shiyou University for funding the Graduate Student Innovation and Practical Skills Training Program (No. YCS21213174).

References

Ambrose, R.J., Hartman, R.C., Diaz-Campos, M., Akkutlu, I.Y., Sondergeld, C.H., 2010. New pore-scale considerations for shale gas in place calculations. In: All Days. Presented at the SPE Unconventional Gas Conference. SPE, Pittsburgh, Pennsylvania, USA. https://doi.org/10.2118/131772-MS. SPE-131772-MS.

Ben, Y., Perrotte, M., Ezzatabadipour, M., Ali, I., Sankaran, S., Harlin, C., Cao, D., 2020. Real-time hydraulic fracturing pressure prediction with machine learning. In: Day 3 Thu, February 06, 2020. Presented at the SPE Hydraulic Fracturing Technology Conference and Exhibition. SPE, The Woodlands, Texas, USA. https://doi.org/10.2118/199699-MS. D031S008R003.

Brieman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://link.springer.com/article/10.1023/A:1010933404324.

Cios, K.J., Kurgan, L.A., Witold, P., Swiniarski, R.W., SpringerVerlag, 2007. Data mining: a knowledge discovery approach. Data Mining: Knowl. Discov. Approach. https://doi.org/10.1007/978-0-387-36795-8.

Curtis, J.B., 2002. Fractured shale-gas systems. AAPG (Am. Assoc. Pet. Geol.) Bull. 86, 1921–1938. https://doi.org/10.1306/61EEDDBE-173E-11D7-8645000102C1865D.

Dai, J., Dong, D., Ni, Y., Hong, F., Zhang, S., Zhang, Y., Ding, L., 2020. Several essential geological and geochemical issues regarding shale gas research in China. J. Nat. Gas Geosci. 5, 169–184. https://doi.org/10.1016/j.jnggs.2020.07.004.

Dai, J., Ni, Y., Liu, Q., Wu, X., Gong, D., Hong, F., Zhang, Y., Liao, F., Yan, Z., Li, H., 2021. Sichuan super gas basin in southwest China. Petrol. Explor. Dev. 48, 1081–1088. https://doi.org/10.11698/PED.2021.06.01.

Dong, Z., Wu, L., Wang, L., Li, W., Wang, Z., Liu, Z., 2022. Optimization of fracturing parameters with machine-learning and evolutionary algorithm methods. Energies 15, 6063. https://doi.org/10.3390/en15166063.

Fernandez-Martinez, J.L., Álvarez, J.P.F., García-Gonzalo, M.E., Pérez, C.O.M., Kuzma, H.A., 2008. Particle Swarm Optimization (PSO): a simple and powerful algorithm family for geophysical inversion. In: SEG Technical Program Expanded Abstracts 2008, SEG Technical Program Expanded Abstracts. Society of Exploration Geophysicists, pp. 3568–3571. https://doi.org/10.1190/1.3064068.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gamal, H., Alsaihati, A., Elkatatny, S., Haidary, S., Abdulraheem, A., 2021. Rock strength prediction in real-time while drilling employing random forest and functional network techniques. J. Energy Resour. Technol. 143 https://doi.org/10.1115/1.4050843.

Gorucu, S.E., Ertekin, T., 2011. Optimization of the design of transverse hydraulic fractures in horizontal wells placed in dual porosity tight gas reservoirs. In: All Days. Presented at the SPE Middle East Unconventional Gas Conference and Exhibition. SPE, Muscat, Oman. https://doi.org/10.2118/142040-MS. SPE-142040-MS.

He, X., Bowers, S., Candela, J.Q., Pan, J., Jin, O., Xu, Tianbing, Liu, B., Xu, Tao, Shi, Y., Atallah, A., Herbrich, R., 2014. Practical Lessons from Predicting Clicks on Ads at Facebook. Presented at the 20th ACM SIGKDD Conference. ACM Press, New York, NY, USA, pp. 1–9. https://doi.org/10.1145/2648584.2648589.

He, Z., Hu, Z., Nie, H., Li, S., Xu, J., 2017. Characterization of shale gas enrichment in the Wufeng Formation–Longmaxi Formation in the Sichuan Basin of China and evaluation of its geological construction–transformation evolution sequence. J. Nat. Gas Geosci. 2, 1–10. https://doi.org/10.1016/j.jnggs.2017.03.002.

Jain, A., Nandakumar, K., Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern Recogn. 38, 2270–2285. https://doi.org/10.1016/j.patcog.2005.01.012.

Johnson, C., Boersma, T., 2013. Energy (in)security in Poland the case of shale gas. Energy Pol. 53, 389–399. https://doi.org/10.1016/j.enpol.2012.10.068.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349, 255–260. https://doi.org/10.1126/science.aaa8415.

Kavitha, S., Varuna, S., Ramya, R., 2016. A comparative analysis on linear regression and support vector regression. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). Presented at the 2016 Online International Conference on Green Engineering and Technologies (IC-GET). IEEE, Coimbatore, India, pp. 1–5. https://doi.org/10.1109/GET.2016.7916627.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Presented at the ICNN'95 - International Conference on Neural Networks. IEEE, Perth, WA, Australia, pp. 1942–1948. https://doi.org/10.1109/ICNN.1995.488968.

Kim, H.-J., Baek, J.-W., Chung, K., 2021. Associative knowledge graph using fuzzy clustering and min-max normalization in video contents. IEEE Access 9, 74802–74816. https://doi.org/10.1109/ACCESS.2021.3080180.

King, G.E., 2010. Thirty years of gas shale fracturing: what have we learned?. In: All Days. Presented at the SPE Annual Technical Conference and Exhibition. SPE, Florence, Italy. https://doi.org/10.2118/133456-MS. SPE-133456-MS.

Lim, H.-I., 2019. A linear regression approach to modeling software characteristics for classifying similar software. In: Presented at the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). IEEE, Milwaukee, WI, USA, pp. 942–943. https://doi.org/10.1109/COMPSAC.2019.00152.

Luo, G., Tian, Y., Bychina, M., Ehlig-Economides, C., 2018. Production optimization using machine learning in Bakken shale. In: Proceedings of the 6th Unconventional Resources Technology Conference. Presented at the Unconventional Resources Technology Conference. American Association of Petroleum Geologists, Houston, Texas, USA. https://doi.org/10.15530/urtec-2018-2902505.

Ma, Z., Pi, G., Dong, X., Chen, C., 2017. The situation analysis of shale gas development in China-based on Structural Equation Modeling. Renew. Sustain. Energy Rev. 67, 1300–1307. https://doi.org/10.1016/j.rser.2016.06.085.

Nejad, A.M., Sheludko, S., Shelley, R.F., Hodgson, T., McFall, R., 2015. A case history: evaluating well completions in the Eagle Ford shale using a data-driven approach. In: Day 2 Wed, February 04, 2015. Presented at the SPE Hydraulic Fracturing Technology Conference. SPE, The Woodlands, Texas, USA. https://doi.org/10.2118/SPE-173336-MS. D021S004R004.

Onwunalu, J.E., Durlofsky, L.J., 2009. Development and application of a new well pattern optimization algorithm for optimizing large-scale field development. New Orleans, Louisiana. In: All Days. Presented at the SPE Annual Technical Conference and Exhibition, SPE. https://doi.org/10.2118/124364-MS. SPE-124364-MS.

Ottah, D.G., Ikiensikimama, S.S., Matemilola, S.A., 2015. Aquifer matching with material balance using particle swarm optimization algorithm – PSO. SPE, Lagos, Nigeria. In: All Days. Presented at the SPE Nigeria Annual International Conference and Exhibition. https://doi.org/10.2118/178319-MS. SPE-178319-MS.

Tan, C., He, J., Zhou, T., Liu, J., Song, W., 2020. A Study on the Optimization of Fracturing Operation Parameters Based on PCA-BN, vol. 42. Joumal of South wast Petroleum University (Science & Technology Edition, pp. 56–62. https://doi.org/10.11885/j.issn.1674-5086.2020.05.12.05.

Tang, J., Fan, B., Xu, G., Xiao, L., Tian, S., Luo, S., Weitz, D., 2020. A new tool for searching sweet spots by using gradient boosting decision trees and generative adversarial networks. In: Presented at the International Petroleum Technology Conference. OnePetro. https://doi.org/10.2523/IPTC-19941-Abstract.

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Trans. Neural Network. 10, 988–999. https://doi.org/10.1109/72.788640.

Wang, H., Shi, Z., Zhao, Q., Liu, D., Sun, S., Guo, W., Liang, F., Lin, C., Wang, X., 2020. Stratigraphic framework of the Wufeng - Longmaxi shale in and around the Sichuan Basin, China: implications for targeting shale gas. Energy Geosci. 1, 124–133. https://doi.org/10.1016/j.engeos.2020.05.006.

Wang, L., Yao, Y., Luo, X., Daniel Adenutsi, C., Zhao, G., Lai, F., 2023. A critical review on intelligent optimization algorithms and surrogate models for conventional and unconventional reservoir production optimization. Fuel 350, 128826. https://doi.org/10.1016/j.fuel.2023.128826.

Wang, L., Yao, Y., Wang, K., Adenutsi, C.D., Zhao, G., Lai, F., 2022a. Data-driven multi-objective optimization design method for shale gas fracturing parameters. J. Nat. Gas Sci. Eng. 99, 104420 https://doi.org/10.1016/j.jngse.2022.104420.

Wang, L., Yao, Y., Wang, W., Adenutsi, C.D., Zhao, G., Lai, F., 2022b. Integrated optimization design for horizontal well spacing and fracture stage placement in shale gas reservoir. J. Nat. Gas Sci. Eng. 105, 104706 https://doi.org/10.1016/j.jngse.2022.104706.

Wang, S., Chen, S., 2019. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. J. Petrol. Sci. Eng. 174, 682–695. https://doi.org/10.1016/j.petrol.2018.11.076.

Wang, Y., Xia, S.-T., 2017. Unifying attribute splitting criteria of decision trees by Tsallis entropy. In: Presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, New Orleans, LA, pp. 2507–2511. https://doi.org/10.1109/ICASSP.2017.7952608.

Williams-Stroud, S., 2008. Using microseismic events to constrain fracture network models and implications for generating fracture flow properties for reservoir simulation. In: All Days. Presented at the SPE Shale Gas Production Conference. SPE, Fort Worth, Texas, USA. https://doi.org/10.2118/119895-MS. SPE-119895-MS.

Wu, L., Dong, Z., Li, W., Jing, C., Qu, B., 2021. Well-logging prediction based on hybrid neural network model. Energies 14, 8583. https://doi.org/10.3390/en14248583.

Xu, J., Guo, C., Wei, M., Jiang, R., 2015. Production performance analysis for composite shale gas reservoir considering multiple transport mechanisms. J. Nat. Gas Sci. Eng. 26, 382–395. https://doi.org/10.1016/j.jngse.2015.05.033.

Yan, X., Wang, X., Zhang, H., Wang, Y., Duan, Y., 2015. Analysis of sensitive parameter in numerical simulation of shale gas reservoir with hydraulic fractures, 37, pp. 127–132. https://doi.org/10.11885/j.issn.1674-5086.2013.06.14.04.

Yu, W., Sepehrnoori, K., 2013. Optimization of multiple hydraulically fractured horizontal wells in unconventional gas reservoirs. J. Petrol. Eng. 1–16. https://doi.org/10.1155/2013/151898, 2013.

Zhan, C., Sankaran, S., LeMoine, V., Graybill, J., Sher Mey, D.-O., 2019. Application of machine learning for production forecasting for unconventional resources. In: Proceedings of the 7th Unconventional Resources Technology Conference. Presented at the Unconventional Resources Technology Conference. American Association of Petroleum Geologists, Denver, Colorado, USA. https://doi.org/10.15530/urtec-2019-47.

Zhang, X., Du, C., Deimbacher, F., Crick, M., Harikesavanallur, A., 2009. Sensitivity studies of horizontal wells with hydraulic fractures in shale gas reservoirs. In: Presented at the IPTC 2009: International Petroleum Technology Conference. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609-pdb.151.iptc13338 cp.

Zhao, G., Yao, Y., Wang, L., Adenutsi, C.D., Feng, D., Wu, W., 2022. Optimization design of horizontal well fracture stage placement in shale gas reservoirs based on an efficient variable-fidelity surrogate model and intelligent algorithm. Energy Rep. 8, 3589–3599. https://doi.org/10.1016/j.egyr.2022.02.228.

Zhou, X., Ran, Q., 2023. Optimization of fracturing parameters by modified genetic algorithm in shale gas reservoir. Energies 16, 2868. https://doi.org/10.3390/en16062868.

Zhu, D., Hu, Y., Cui, M., Chen, Y., Liang, C., Cai, W., He, Y., Wang, X., Chen, H., Li, X., 2020. Productivity simulation of hydraulically fractured wells based on hybrid local grid refinement and embedded discrete fracture model. Petrol. Explor. Dev. 47, 365–373. https://doi.org/10.1016/S1876-3804(20)60053-2.

Zou, Z., Yang, Y., Fan, Z., Tang, H., Zou, M., Hu, X., Xiong, C., Ma, J., 2020. Suitability of data preprocessing methods for landslide displacement forecasting. Stoch. Environ. Res. Risk Assess. 34, 1105–1119. https://doi.org/10.1007/s00477-020-01824-x.