Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 3, 2024

Shale Gas Fracturing Construction Parameter Optimization Based on SSA-XGBoost

Jixin Dai, He Zhang

School of Mechatronic and Electrical Engineering, Southwest Petroleum University, Chengdu 610500, China

Abstract: To address the challenges of poor model performance, low prediction accuracy, and suboptimal application of intelligent optimization algorithms for hydraulic fracturing parameter optimization in shale gas production, this study proposes a novel optimization method based on the SSA-XGBoost algorithm. First, using production data from 120 wells with complete records in the Y block, grey relational analysis (GRA) was employed to reduce dimensionality among 14 factors affecting fracturing performance, resulting in the selection of nine primary controlling factors. Based on the mapping relationship between these controlling factors and the fracturing performance evaluation index (daily gas production), the Sparrow Search Algorithm (SSA) was applied to optimize the hyperparameters of the Extreme Gradient Boosting (XGBoost) model. The resulting SSA-XGBoost-based shale gas production prediction model improved the coefficient of determination (R2) by 5.6% compared to the original model, achieving significantly enhanced prediction accuracy and outperforming existing models in the field. Finally, based on the production predictions of the optimized model, the optimal parameter ranges for hydraulic fracturing were determined, yielding normalized daily gas production 27.50% higher than that of benchmark wells. These findings demonstrate the rationality and accuracy of the proposed optimization approach, providing valuable reference for on-site fracturing parameter design.

Keywords: Hydraulic Fracturing; Grey Relational Analysis; Sparrow Search Algorithm; Extreme Gradient Boosting; Production Prediction; Parameter Optimization.

1. Introduction

Shale gas is a critically important oil and gas resource in China, with immense development potential[1]. In shale gas reservoir development, horizontal drilling and multi-stage hydraulic fracturing play pivotal roles. Optimizing horizontal fracturing parameters prior to production is one of the most crucial steps in enhancing shale gas production. The effectiveness of production enhancement through fracturing largely depends on the optimization of fracturing strategies[2]. By effectively optimizing shale gas fracturing parameters, not only can the production capacity of shale gas wells be significantly improved, but the damage caused by fracturing fluids to the reservoir can also be minimized. Therefore, research on shale gas hydraulic fracturing parameter optimization methods is of great significance.

Extensive research has been conducted by scholars worldwide on optimizing hydraulic fracturing parameters. Traditional methods for fracturing parameter optimization often involve constructing fracture propagation models to simulate the fracture extension process during hydraulic fracturing, thereby guiding parameter design[3,4]. However, such simulation models simplify the fracture propagation process and fail to accurately represent the complex fracture networks in reservoirs. Another approach utilizes numerical simulation and multivariate analysis methods for optimizing fracturing parameters[5,6]. While these methods provide a more comprehensive perspective, the formation of diverse and complex fracture networks during large-scale hydraulic fracturing significantly increases the computational time required for shale reservoir simulations, thereby reducing the efficiency of parameter optimization.

With the widespread application of artificial intelligence technology in the oil and gas industry, a large amount of fracturing operation and production performance data is now available for reference[7,8]. Significant progress has been made in the research on establishing shale gas production prediction models through data mining and machine learning, as well as optimizing fracturing construction parameters[9]. Lee et al.[10]used the LSTM algorithm to predict the future production rate of shale gas wells based on two features of production data and cycles. Tan et al[11] established shale gas production prediction models using six machine learning algorithms, including BP neural network, random forest, support vector regression, LightGBM, XGBoost, and multiple linear regression. They selected the best-performing model and optimized fracturing construction parameters considering both production and cost-profit ratio. Wang et al[12]built a prediction model based on the LSSVR algorithm and applied a multi-objective method to optimize construction parameters. Liu et al [13] developed a comprehensive production evaluation model using random forest algorithm and coordinated principal component analysis to evaluate fracturing effects in horizontal wells, combining orthogonal experimental design methods to optimize fracturing parameter design. Qian et al[14]integrated reservoir numerical simulation with the GBDT algorithm to create a machine learning model, utilizing particle swarm optimization (PSO) to optimize fracture parameters. This method aimed to optimize fracturing parameters and predict production based on cumulative gas production under different geological conditions. Dong et al[15] compared the optimization effects of four evolutionary algorithms—genetic algorithm, differential evolution algorithm, simulated annealing algorithm, and particle swarm optimization—on tight oil horizontal well fracturing parameters. Zhou et al[16]established a regression prediction model to determine the relationship between parameters and production capacity, optimizing fracturing parameters based on an improved

genetic algorithm with production as the target. Yang et al[17] studied a vertical fracturing well in a block of the Sulige gas field, using k-means clustering, classification enhancement, extreme gradient boosting, and LightGBM algorithms to build regression models. They then constructed an optimization model using a genetic algorithm, focusing on four factors: fracturing fluid, proppant, liquid nitrogen, and construction displacement quantity.

However, several substantive issues remain in the current research. First, the relationship between reservoir parameters, operational parameters, and production performance is highly nonlinear, with strong interdependencies among influencing factors. The use of multiple factors as input parameters complicates model structures. Second, existing studies often overlook parameter optimization during the development of prediction models, and single-model approaches fail to achieve optimal prediction accuracy. Additionally, many studies adopt data-driven models and directly use intelligent optimization algorithms to optimize fracturing parameters without addressing challenges such as low optimization accuracy, poor search capabilities, and limitations imposed by the dimensionality of field data. These issues make it difficult to ensure precise optimization of operational parameters.

To address these challenges and achieve the goal of enhancing shale gas well production, this study proposes a hydraulic fracturing parameter optimization method based on an SSA-XGBoost prediction model. The method employs Grey Relational Analysis (GRA) to reduce the dimensionality of fracturing parameters affecting gas production, selecting key influencing factors as inputs. It then constructs an Extreme Gradient Boosting (XGBoost) prediction model with a key evaluation metric, the normalized daily gas production, as the output. The Sparrow Search Algorithm (SSA), known for its strong optimization capabilities, is used to fine-tune the model, achieving optimal prediction accuracy. Finally, by predicting gas production under various fracturing scenarios, the method identifies the optimal operational parameters, providing actionable optimization recommendations for field fracturing operations.

2. Research Methodology Theory

Guided by data mining and machine learning algorithms, the optimization of fracturing operational parameters for shale gas is achieved by building a production prediction model. The primary theoretical methods relied upon in the model construction process include Grey Relational Analysis , Extreme Gradient Boosting, and the Sparrow Search Algorithm . These methods work together to improve the accuracy and efficiency of the optimization process, ensuring precise prediction of shale gas production and optimal fracturing parameters.

2.1. Grey Relational Analysis

Currently, the main methods used for dimensionality reduction are Principal Component Analysis (PCA), Factor Analysis, and Grey Relational Analysis (GRA). Grey Relational Analysis (GRA) calculates the grey relational degree to describe the strength of correlation between sequences. By ranking these relational degrees, GRA reduces the dimensionality of the data, effectively selecting the key influencing factors from the feature data sequences, eliminating redundant information, and comprehensively extracting valuable information from the dataset[18].

When processing field datasets, due to the complex nonlinear relationships between influencing factors and the fracturing effect evaluation index (daily gas production), directly using all the influencing factor data as input parameters to construct the model can make the model overly complex, leading to overfitting or underfitting, and increasing the difficulty of model training. By employing Grey Relational Analysis to select the primary control factors, this method not only retains the original information to the greatest extent but also simplifies the problem.In this study, the dataset consists of 120 samples, each containing 14 variables (fracturing effect influencing factors) and 1 dependent variable (daily gas production).

The prediction model can be influenced by the dimensionality of the data, which may result in inaccurate prediction outcomes. Therefore, before conducting Grey Relational Analysis, the data must be preprocessed. The first step is to perform dimensionless processing (range normalization) on all data, ensuring that each variable is treated consistently and in comparable units.

$$x^* = \frac{x_i - \min\left(x_i\right)}{\max\left(x_i\right) - \min\left(x_i\right)} \tag{1}$$

Where: x_i is the value of the sample feature, x^* is the normalized value, $\max\left(x_i\right)$ are the maximum value of the sample features, $\min\left(x_i\right)$ is the minimum value of the sample feature.

The specific steps of Grey Relational Analysis (GRA) are as follows [19]:

1) Determine the reference sequence and comparison sequences;

2) Analyze the sequences, where the sequence that reflects the system's behavioral characteristics is called the reference sequence, and the sequence composed of factors influencing the system's behavior is called the comparison sequence.

$$Y = \left\{Y\left(k\right) \middle| k = 1, 2, \cdots, n\right\} \tag{2}$$

$$X_i = \left\{X_i\left(k\right) \middle| k = 1, 2, \cdots, n\right\}, i = 1, 2, \cdots, m \tag{3}$$

Where: Y is the reference sequence (daily gas production); X_i is the comparison sequence (influencing factors); k is the number of data points in the sequence, k =120; i refers to the number of comparison sequences, i=14

3) Dimensionality Reduction

Prior to analyzing the sequences, normalization has been performed on the data.

4) Calculating the Correlation Coefficient

The correlation coefficient reflects the degree of association between a specific data point in the comparison sequence and its corresponding data point in the reference sequence. The expression is as follows:

$$\xi_i = \frac{\min\limits_i \min\limits_k \left|y\left(k\right) - x_i\left(k\right)\right| + \rho \max\limits_i \max\limits_k \left|y\left(k\right) - x_i\left(k\right)\right|}{\left|y\left(k\right) - x_i\left(k\right)\right| + \rho \max\limits_i \max\limits_k \left|y\left(k\right) - x_i\left(k\right)\right|} \tag{4}$$

Where: ξ_i is the correlation coefficient; $y(k)$ is the data of the reference sequence; $x_i(k)$ is the data of the comparison sequence; $\rho \in [0,1]$ is the distinguishing coefficient, typically set to 0.5.

5) Calculating the Relational Degree

The relational degree reflects the degree of association between the comparison sequence and the reference sequence, denoted as r_i. The formula is as follows:

$$r_i = \frac{1}{n}\sum_{k=1}^{n}\xi_i(k) \tag{5}$$

Where: $r_i \in (0,1)$.

6) Sorting the Relational Degree

The relational degrees can be sorted in descending order to assess the influence of each factor, allowing for the selection of the dominant factors based on their impact.

2.2. Extreme Gradient Boosting Algorithm

The Extreme Gradient Boosting algorithm(XGBoost) is a parallel ensemble learning method that uses the idea of boosting, where subtrees are generated in parallel and the final results are obtained through a weighted aggregation of these trees. It is highly effective in handling complex datasets[20].

Due to the complexity of shale gas production datasets, traditional machine learning techniques often fail to meet the practical requirements of engineering applications. Therefore, by examining the relationship between various fracturing effect influencing factors and normalized daily gas production, XGBoost can be used to construct a shale gas production prediction model, achieving superior prediction results.

In predicting normalized daily gas production, the results from K decision trees are aggregated to obtain the final prediction, as shown in the following formula[21]:

$$\hat{y}_i = \sum_{k=1}^{k} f_k(x_i), f_k \in F \tag{6}$$

Where: $\hat{y}_j$ is the final predicted value of daily gas production from the model; f_k is the prediction from each decision tree; F is the regression tree space; $f_k(x_i)$ is the predicted value of f_k given the input factors x_i (fracturing effect influencing factors). Therefore, the loss function L can be expressed as the difference between the actual value y_i and the predicted value $\hat{y}_i$:

$$L = \sum_{i=1}^{n} loss(y_i, \hat{y}_i) \tag{7}$$

To prevent overfitting and improve the model's generalization ability, a regularization term Ω is usually added to the equation. The loss function can be expressed as follows, where the regularization term $\Omega(f_k)$ determines the depth of the tree to be adjusted (max_depth), as shown in the following expression:

$$Obj = \sum_{i=1}^{n} loss(y_i, \hat{y}_i) + \sum_{k=1}^{k}\Omega(f_k) \tag{8}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \tag{9}$$

Where: T is the number of leaf nodes, w_j is the weight of the j-th leaf node.

XGBoost uses a forward boosting strategy, where each time a new decision tree is added, it learns a new function and its coefficients to fit the residuals of the previous step. The calculation expression is as follows:

$$\hat{y}_i^k = y_i^{k-1} + f_k(x_i) \tag{10}$$

$$Obj^{(k)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}^{k-1} + f_k(x_i)\right) + \Omega(f_k) \tag{11}$$

The above equation is expanded using the Taylor series as follows:

$$Obj^{(k)} \approx \sum_{i=1}^{n}\left[L(y_i, \hat{y}^{k-1}) + g_i f_k(x_i) + \frac{1}{2}h_i f_k^2(x_i)\right] + \Omega(f_k) \tag{12}$$

Where: g_i is the first-order derivative of the loss function $L(y_i, \hat{y}^{k-1})$ with respect to $\hat{y}^{k-1}$, and h_i is the second-order derivative of $(y_i, \hat{y}^{k-1})$ with respect to $\hat{y}^{k-1}$, The total of all h_i is sum of the minimum sample weights for the leaf nodes that need to be adjusted(min_child_weight). Additionally, it is possible to merge the same function values at the same leaf node. Therefore, the final derived result is as follows:

$$w_j = -\frac{G_j}{H_j + \lambda} \tag{13}$$

Substituting w_j into the objective function simplifies to:

$$Obj^{(k)} = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \lambda T + C \tag{14}$$

The mapping relationship between daily gas production and various fracturing effect influencing factors serves as the objective function to obtain a simplified equation, ultimately yielding the predicted value of normalized daily gas production.

In the ensemble learning framework of XGBoost, several hyperparameters influence model performance, such as the number of weak classifiers (n_estimators), the minimum loss reduction required for further splits (gamma), with higher values leading to more conservative algorithms, the maximum depth of the trees (max_depth), the minimum sum of instance weights (min_child_weight), subsampling rate (subsample), regularization parameter (lambda), and the maximum number of leaf nodes (max_leaves). These hyperparameters can adjust the model's fitting ability.

Therefore, to achieve accurate prediction of normalized daily gas production, it is crucial to quickly, simply, and accurately select the optimal hyperparameters for the model.

2.3. Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is a new population-based intelligent optimization algorithm proposed by Xue et al. in 2020. Its core idea is to simulate the foraging and predator-avoidance behaviors of sparrow populations[22].

Compared to widely used optimization algorithms, such as Genetic Algorithm, Particle Swarm Optimization, and Ant Colony Optimization, SSA offers several advantages: efficient, fast, parallel, adaptive, robust, simple in computation, and globally converging to the optimal solution[23]. The application of SSA can further improve the prediction accuracy of models and reduce the processing time for complex field data. It is less likely to fall into local optima, greatly enhancing the model's ability to be applied in real-world scenarios, making it better suited to the needs of wellsite operations and providing reliable support for the optimization of fracturing construction plans.

The SSA algorithm is employed to automatically optimize the hyperparameters of the XGBoost algorithm, with a focus on adjusting four key hyperparameters: the number of trees (n_estimators), maximum tree depth (max_depth), learning rate (learning_rate), and subsample rate (subsample).

The specific steps of the SSA-optimized XGBoost algorithm are as follows:[24]:

1) Input the dataset and build the shale gas production prediction model based on XGBoost.

2) Initialize the sparrow population and set the maximum number of iterations.

3) Train the model, calculate the fitness value of each sparrow, and rank them.

4) Select the next generation until the entire sparrow population converges. At this point, the optimal values of the four hyperparameters—gamma, subsample, colsample_ bytree, alpha, and lambda—are obtained. If the population does not converge and the maximum number of iterations has not been reached, return to step 3 and continue iterating.

5) Determine the optimal values of the four hyperparameters in the XGBoost algorithm and obtain the shale gas production prediction model based on SSA-XGBoost, see Figure 1.

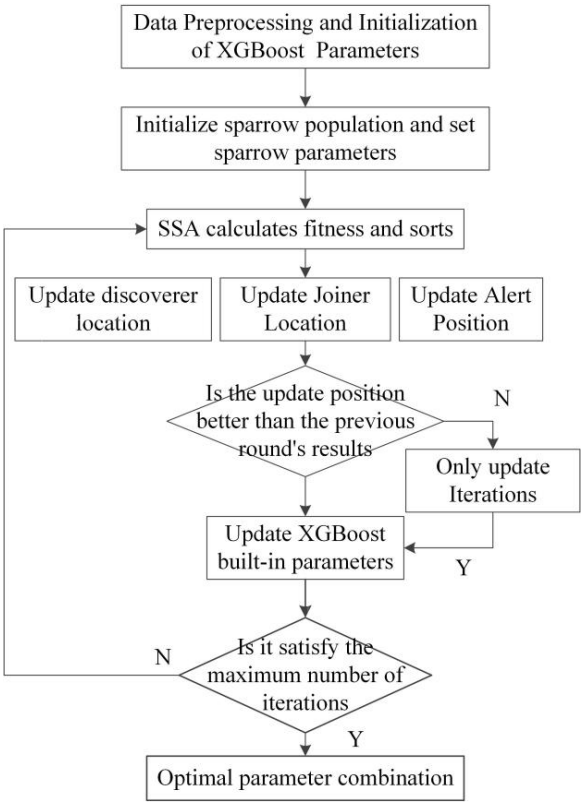

Figure 1. SSA optimizes the XGBoost flowchart

3. Modeling Experiment Analysis

Grey Relational Analysis (GRA) is first applied to optimize and extract the main influencing factors from the geological and operational factors affecting fracturing performance. The selected key factors are used as input variables, with the fracturing performance evaluation index (normalized daily gas production) serving as the output variable to construct the XGBoost model. The Sparrow Search Algorithm (SSA) is then utilized to optimize the hyperparameters of the model, resulting in the development of an SSA-XGBoost-based shale gas production prediction model. Finally, the trained prediction model is employed to optimize and analyze the fracturing operation parameters.

3.1. Selection of Key Variables

Geological, operational, and flowback data from 120 shale gas wells in the Y block, covering the period from 2018 to 2022, were collected, organized, and processed. The dataset is complete, and 120 samples were selected for model construction and analysis. daily gas production is chosen as the evaluation index for fracturing performance, and the influencing factors are analyzed using GRA. The dataset contains 15 variables (14 independent variables and 1 dependent variable), including Total Organic Carbon (TOC), porosity, gas content, average single-stage length, horizontal section length, number of stages, total cluster number, average number of clusters per stage, operation flow rate, total fracturing fluid volume, proppant intensity, fluid intensity, total proppant volume, and normalized daily gas production.The grey relational degree between each factor and normalized daily gas production is calculated, and the factors are ranked based on their relational degrees, see Figure 2.

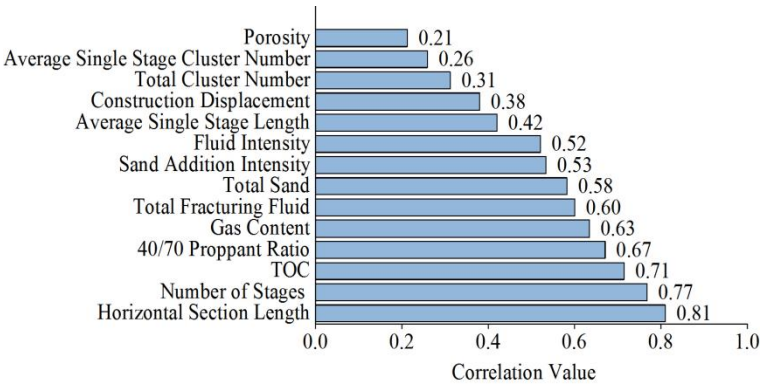

Figure 2. Grey Relational Degree Analysis Results

From Figure 2, it can be seen that nine factors have a relational degree value above 0.5. Among these, the effective horizontal section length and the number of stages have the highest relational degree values of 0.81 and 0.77, respectively, indicating that these two factors have a significant positive impact on normalized daily gas production. Additionally, physical parameters such as TOC and gas content show strong correlations, with relational degree values of 0.71 and 0.63, respectively, suggesting that reservoirs with better physical properties tend to have higher production rates. The factors with relational degree values above 0.5 contain most of the original information. Therefore, these nine factors are selected as the main input parameters, with normalized daily gas production as the output variable, for the construction of the shale gas production prediction model.

3.2. Model Evaluation Metrics

To evaluate the performance of the models, commonly used regression evaluation metrics, including Mean Squared Error (MSE), Mean Relative Error (MRE), and Coefficient of Determination (R^2), are applied. The formulas for these metrics are as follows[25]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{15}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i - y_i\right|}{y_i} \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{17}$$

Where: $\hat{y}_i$ is the predicted value, y_i is the actual value, and $\bar{y}$ represents the mean value. Among them, MSE reflects the degree of variation in the data; the closer it is to 0, the higher the model's accuracy. MRE represents the relative error between the predicted and actual values; the closer it is to 0, the greater the model's accuracy and stability. R^2 indicates the goodness of fit between the predicted and actual values; the closer the value is to 1, the better the model's performance.

represents the predicted value, represents the actual value, and represents the mean value. Among them, intuitively reflects the degree of variation in the data; the closer it is to 0, the higher the model's accuracy. reflects the relative error between the predicted and actual values; the closer it is to 0, the higher the model's accuracy and stability. characterizes the model's goodness of fit between the predicted and actual values; the closer the value is to 1, the better the model performance.

3.3. Establishment of the SSA-XGBoost Shale Gas Production Prediction Model

Based on the selected 120 samples, 70% of the samples are used as the training set and 30% as the test set to ensure that the model has sufficient learning capacity for generalization. After feature selection through grey relational analysis, there are 9 input vectors and 1 output vector. The XGBoost model for shale gas production prediction is constructed using the training set, and the SSA algorithm is applied to optimize the values of 4 hyperparameters in the XGBoost model: n_estimators, max_depth, learning_rate, and subsample.

The sparrow population size is initially set to 30, with 200 iterations. The mean squared error (MSE) is used as the fitness evaluation metric for the model during iterative training. The curve of MSE values with respect to the number of iterations for the XGBoost model, see Figure 3. The horizontal axis represents the number of iterations, and the vertical axis represents the MSE value. As the SSA algorithm continues to optimize through iterations, the MSE value gradually decreases. After approximately 40 iterations, the MSE value stabilizes around 0.05. When the MSE value reaches its minimum, the optimal values for n_estimators, max_depth, learning_rate, and subsample are obtained, with the optimal results being 230, 13, 0.215, and 0.8, respectively, see Figure 4.

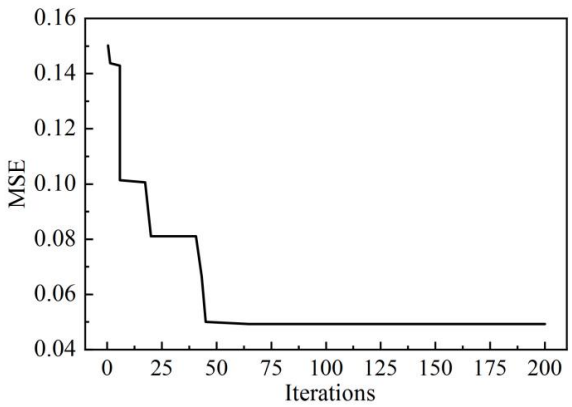

Figure 3. XGBoost model optimization iteration process graph

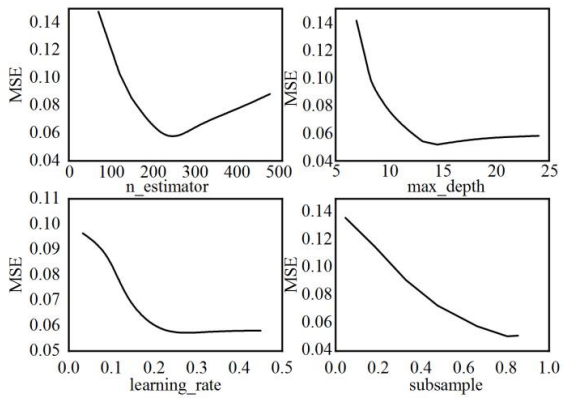

Figure 4. Model hyperparameter optimization process graph

After optimizing the hyperparameters of the XGBoost model using the SSA algorithm, the SSA-XGBoost shale gas production prediction model is successfully obtained.

3.4. Prediction Results Analysis

After the model was established, to validate the performance of the SSA-XGBoost shale gas production prediction model, five prediction models, including SSA-XGBoost, XGBoost (default parameters), Gradient Boosting Decision Tree (GBDT), Support Vector Machine Regression (SVR), and Random Forest (RF), were used to predict the same daily gas production data and analyze their predictive capabilities. The XGBoost model, even without optimization, already demonstrated strong predictive ability, outperforming the other three models. After optimization using SSA, the SSA-XGBoost model exhibited a significant improvement in prediction accuracy, achieving the best performance. Specifically, the Mean Squared Error (MSE) and Mean Relative Error (MRE) decreased by 66% and 63%, respectively, while the Coefficient of Determination (R^2) increased by 5.6%. This enhancement provides more reliable predictive support for the optimization of fracturing operation schemes, see Figure 5/Table 1.

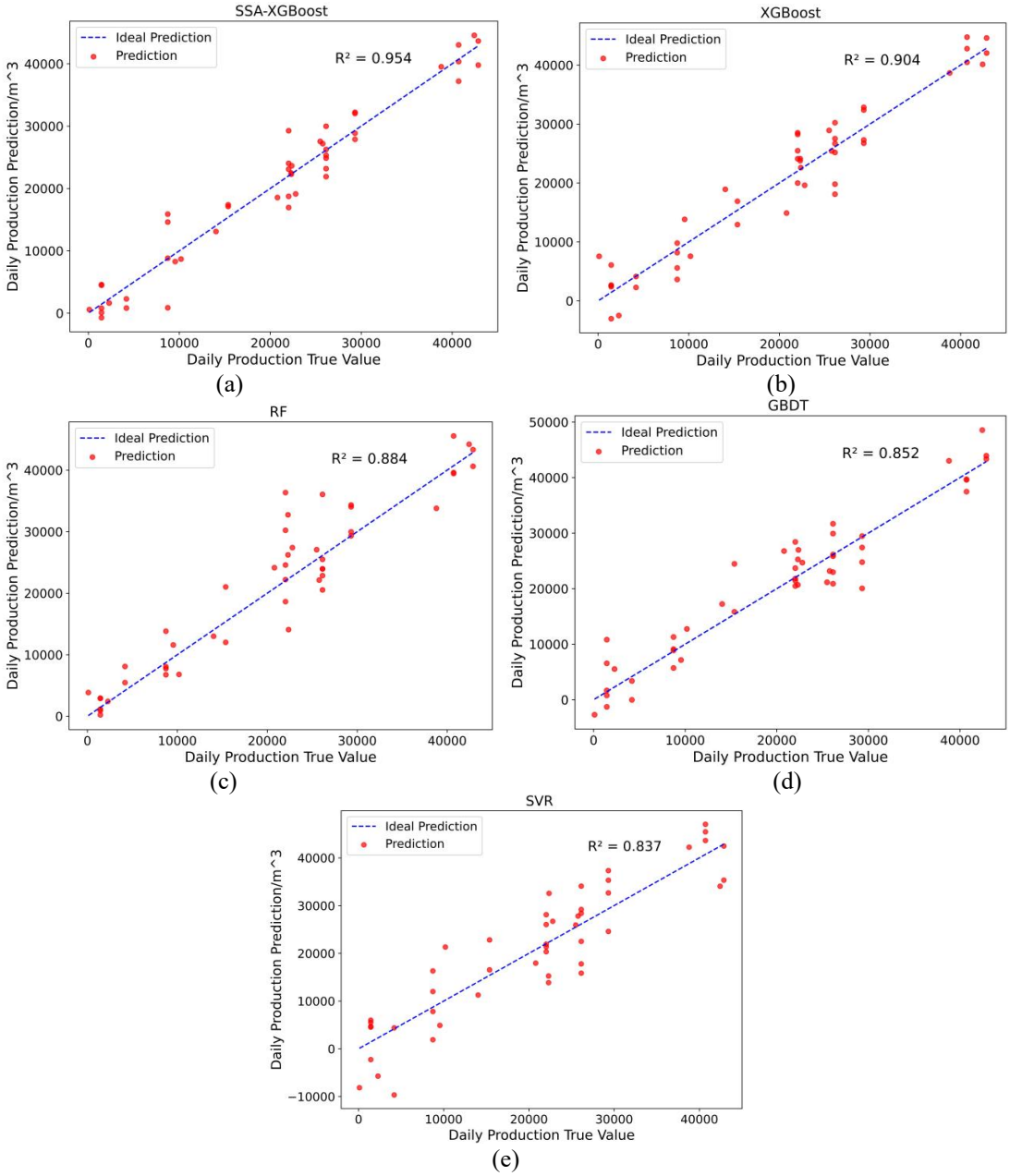

Figure 5. Model Test Set Prediction Results
(a)SSA-XGBoost(b)XGBoost(c)RF(d)GBDT(e)SVR

Table 1. Model Prediction Performance

Numble	SSA-XGBoost	XGBoost	RF	GBDT	SVR
MSE	0.050	0.149	0.174	0.231	0.277
MRE	0.095	0.260	2.237	1.133	0.921
R2	0.954	0.904	0.884	0.852	0.837

To further validate the optimization capability of the Sparrow Search Algorithm (SSA) in improving the prediction accuracy of the model, three widely used optimization algorithms—Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO)—were selected as benchmarks. These algorithms were applied for hyperparameter optimization of the XGBoost model to perform performance optimization tests. The experimental results show that SSA-XGBoost outperforms the other three optimization models in all three evaluation metrics, demonstrating the strong optimization capability of SSA. It also exhibits superior performance in improving prediction accuracy, significantly enhancing the model's predictive precision and proving its strong practical value, see Figure 6.

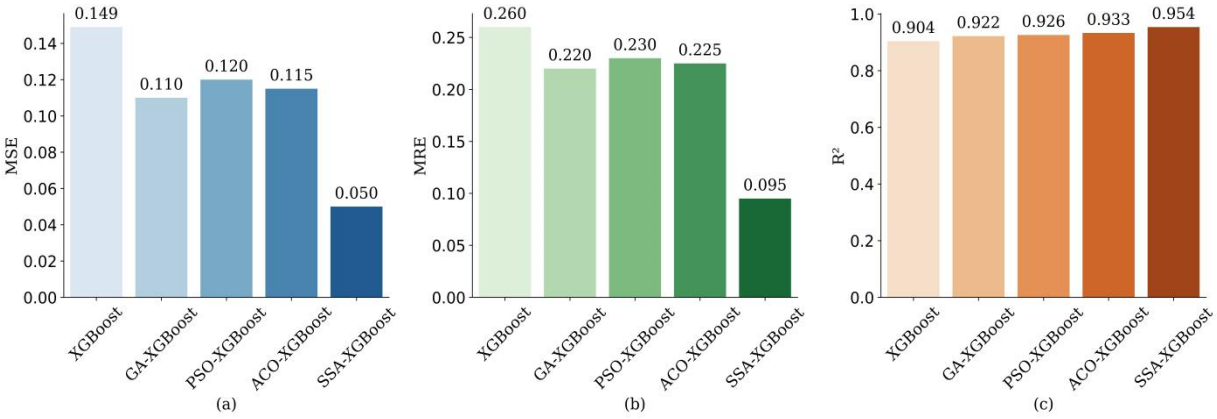

Figure 6. Comparison of optimization algorithm performance

3.5. Optimization of Fracturing Construction Parameters Analysis

Using the trained SSA-XGBoost shale gas production prediction model, the daily gas production for different fracturing construction plans was predicted. These predictions were compared to the predicted production from other plans, with the top-performing plans identified as the optimal ones. With the recent advancements in fracturing technology, current construction plans have undergone significant changes compared to previous ones. Therefore, based on the fracturing operations in the shallow shale gas wells of the Y block from 2018 to 2022, and considering the results across different plans, 200 construction plans were designed. Geological parameters (TOC, gas content) and the average horizontal well length and number of stages for the wells in the block were used as input variables. The optimal fracturing construction plans were identified by selecting the top ten plans with the highest predicted gas production, see Table 2.

The following abbreviations are used to represent key parameters in the shale gas fracturing operation:

1)TOC: Total Organic Carbon
2)GC: Gas Content
3)HSL: Horizontal Section Length
4)NS: Number of Stages
5)TFF: Total Fracturing Fluid
6)TS: Total Sand
7)40/70 PR: 40/70 Proppant Ratio
8)FI: Fluid Intensity
9)SAI: Sand Addition Intensity
10)PDGP: Predicted Daily Gas Production

Table 2. Optimal scheme for fracturing construction

Scheme	TOC (%)	GC (m³/t)	HSL (m)	NS	TFF (m³)	TS (t)	40/70PR (%)	FI (m³/m)	SAI (t/m)	PDGP (m³)
1	4.133	3.675	926	13.2	21741.13	2752.29	63.99	31.20	3.912	29318.05
2					21740.96	2752.23	70.19	31.41	3.90	29317.87
3					25814.50	3307.60	49.26	33.99	4.12	29196.83
4					34079.02	4158.16	48.07	30.51	3.64	26140.22
5					34078.72	4158.08	63.17	30.49	3.62	26139.94
6					35429.89	4121.10	62.40	30.17	4.10	23952.61
7					24250.25	3940.20	46.06	24.84	4.03	23872.76
8					24577.51	4004.60	52.69	24.79	4.03	23156.99
9					34397.55	3957.20	61.42	32.91	3.79	22747.04
10					23872.53	3318.72	60.42	28.89	4.11	22303.95

Table 3. Construction of the standard well

Scheme	TOC (%)	GC (m^3/t)	HSL (m)	NS	TFF (m^3)	TS (t)	40/70PR (%)	FI (m^3/m)	SAI (t/m)	PDGP (m^3)
1	4.41	4.01	1 045	14	26827.77	2322.44	57.04	27.06	2.34	22747.04
2	4.76	3.77	1003.9	14	22608.51	3140.94	56.92	22.84	3.17	27973.14
3	3.89	3.77	991.24	11	34210.40	3740.94	56.92	35.05	3.83	29156.99
4	4.76	4.05	989.58	12	21741.03	4133.94	64.71	22.29	4.23	26432.82
5	4.04	3.71	976	10	24107.20	3768.57	57.67	24.78	3.86	28872.76
6	3.81	3.73	975	12	28801.50	3753.38	58.81	29.54	3.73	32029.82
7	4.73	3.98	963	14	22121.89	4803.14	69.91	23.75	5.15	29658.65
8	4.35	4.69	954.9	13	21564.22	3301.51	49.53	23.72	3.63	22502.07
9	4.29	4.57	931.4	12	33205.06	3523.3	67.40	38.36	5.22	25318.03
10	4.41	4.70	909.1	11	22615.75	2683.29	60.26	24.87	2.95	24748.03
11	4.32	4.65	865.49	12	25082.35	3291.7	64.43	29.82	3.91	26528.62
12	4.08	3.32	862	13	24946.31	3455.64	68.87	23.87	4.45	23901.02
13	4.10	4.16	840.9	10	25814.22	2486.03	47.29	25.71	2.47	21700.16

According to the analysis of the results , the optimal range for fracturing fluid volume is 21,741.13–35,429.89 m^3; the optimal range for total sand volume is 2,752.29–4,158.16 t; the optimal range for 40/70 proppant ratio is 46.06–70.19%; the optimal range for fluid intensity is 24.79–33.99 m^3; and the optimal sand addition intensity is 3.62–4.12 t/m, see Table 3.

To verify the effectiveness of the optimized plans, 13 neighboring wells with similar geological parameters and horizontal well lengths to the wells in the Y block were selected for analysis and comparison. Table 3 shows the construction conditions for the neighboring wells. The results show that Well 6 achieved the highest daily gas production among the 13 neighboring wells, exceeding the average daily production of the 13 wells by 27.50%. Therefore, the construction parameters of Well 6 were identified as the optimal ones among the 13 neighboring wells. Additionally, the construction parameters of Well 6 were found to be very close to the average values of the optimal plan parameters, with relative errors within 5%, see Table 4.

Table 4. Optimization Scheme Benchmarking Analysis

Fracturing Parameters	Average Value	Benchmark Well 6 Fracturing Parameters	Relative Error (%)
TFF (m^3)	27998.21	28801.50	2.3
TS (t)	3653.38	3642.52	2.7
40/70PR (%)	60.17	58.81	2.3
FI (m^3/m)	30.52	29.54	3.2
SAI(t/m)	4.03	3.85	4.4

Since the selected neighboring wells have similar geological parameters and horizontal well lengths to the well in question, the daily gas production of Well 6 closely reflects the rationality of the fracturing construction parameters. The above results indicate that the optimized construction plan derived from the SSA-XGBoost shale gas production prediction model is highly accurate.

4. Summary

Through gray relational analysis of the influencing factors, the dominant factors were selected as input parameters, simplifying the model structure. The SSA algorithm was applied to optimize the XGBoost model, determining the optimal parameter combination. As a result, the SSA-XGBoost model significantly improved the accuracy of shale gas production prediction, with MSE and MRE decreasing by 66% and 63%, respectively, and R^2increasing by 5.6%. Compared to four traditional optimization algorithms, the SSA-optimized model demonstrated superior performance, confirming the significant effectiveness of SSA in optimizing XGBoost parameters and validating its strong practical applicability. When integrated with 200 field-designed fracturing schemes, well benchmarking experiments indicated that optimizing fracturing parameters using the SSA-XGBoost shale gas production prediction model is a reliable and efficient method, with the potential to achieve significant production increases in the Y block.

Despite the limited data available from the Y block, future improvements can incorporate data from additional wells, further refining the model to accommodate fracturing wells in different blocks and enhancing the model's generalization ability and on-site applicability.

References

[1] X.S.Guo, B.J Tenger,X.F.Wei,et al. Occurrence mechanism and exploration potential of deep marine shale gas in Sichuan Basin.Acta Petrolei Sinica.2022,Vol.43(No.4),p.453-468.(In Chinese)

[2] H. Zhang, J. Sheng. Optimization of horizontal well fracturing in shale gas reservoir based on stimulated reservoir volume.Journal of Petroleum Science and Engineering. 2020, Vol.190,p. 107059

[3] T.Y. Luo, J.Z. Zhao, J.H. Wang, et al. Multiple-Transverse-Fracture extension model for Hydraulic fracturing.Natural Gas Industry.2007,Vol. 27 (No. 10) No. 10, p. 75-78 + 139. (In Chinese)

[4] X. Shi, Y.F. Cheng, X. Chang, et al. Establishment and application of the model for the synchronous propagation of multi-cluster fractures in the horizontal section of shale-gas horizontal well.Oil Drilling & Production Technology. 2018, Vol.40 (No.2),p. 247-252. (In Chinese)

[5] C. Liu. Numerical investigating the hydraulic fracturing of horizontal well and the optimization of stimulation parameters,

University of Science and Technology of China, 2017. (In Chinese)

[6] R.H. Zhang, L.H. Zhang, H.Y. Tang, et al. A Simulator for Production Prediction of Multistage Fractured Horizontal Well in Shale Gas Reservoir Considering Complex Fracture Geometry.Journal of Natural Gas Science and Engineering. 2019,Vol. 66 , p. 106537.

[7] Y.W. He, Z.Y. He, Y. Tang, et al. Shale gas well production evaluation and prediction based on machine learning.Oil Drilling & Production Technology.2021, Vol. 43 (No. 4), p. 518-524. (In Chinese)

[8] X.T. Peng, Y. Wang, C. Jia, et al. Fracturing effect prediction based on sparrow search algorithm and BP neural network, Oil Drilling & Production Technology, 2022,Vol. 44 (No.42) , p. 522-528. (In Chinese)

[9] M. Sheng, G.S. Li, S.C. Tian, et al. Research Status and Prospect of Artificial Intelligence in Reservoir Fracturing Stimulation, Drilling & Production Technology, Vol. 45 (2022) No. 4, p. 1-8. (In Chinese)

[10] K. Lee, J. Lim, D. Yoon, et al. Prediction of Shale-Gas Production at Duvernay Formation Using Deep-Learning Algorithm.SPE Journal.2019,Vol. 24 (No. 6) , p. 1317-1325.

[11] C.D. Tan, J.Z. Yang, M.Y. Cui, et al. Fracturing Productivity Prediction Model and Optimization of the Operation Parameters of Shale Gas Well Based on Machine Learning.Lithosphere.2024,Special Issue 4 , p. 2884679.

[12] L. Wang, Y. Yao, K. Wang, et al. Data-driven multi-objective optimization design method for shale gas fracturing parameters. Journal of Natural Gas Science and Engineering. 2022,Vol. 99 ,p. 104420.

[13] X.W. Liu, D.P. Li, Y.P. Jia, et al. Optimizing construction parameters for fractured horizontal wells in shale oil. Frontiers in Earth Science.2023,Vol.10 , p. 1015107.

[14] S. Qian, Z. Dong, Q. Shi, et al. Optimization of shale gas fracturing parameters based on artificial intelligence algorithm.Artificial Intelligence in Geosciences. 2023,Vol. 4, p. 95-110.

[15] Z. Dong, L. Wu, L. Wang, et al. Optimization of Fracturing Parameters with Machine-Learning and Evolutionary Algorithm Methods. Energies.2022,Vol. 15 (No.16), p. 6063.

[16] X. Zhou, Q. Ran. Optimization of fracturing parameters by modified genetic algorithm in shale gas reservoir. Energies. 2023, Vol. 16 (No. 6) , p. 2868.

[17] H. Yang, X. Liu, X. Chu, et al. Optimization of tight gas reservoir fracturing parameters via gradient boosting regression modeling. Heliyon.2024,Vol. 10 , p. e14092.

[18] D.L. Guo, Y.F. Tang, S.G. Li, et al. Optimization of Tight Gas Fracturing Operation Parameters Based on BP-PSO.Science Technology and Engineering.2022,Vol. 22 (No. 19) , p. 8304-8312. (In Chinese)

[19] Z.Z. Zhou, J.Q. Tang, Y.M. Huang, et al. Optimization of Fracturing Parameters for Horizontal Wells in Block A of Aonan Oilfield. Mathematics in Practice and Theory. 2021,Vol. 51 (No. 20) , p. 65-72. (In Chinese)

[20] M. Wang M, G. Hui G,Y. Pang Y, et al. Optimization of machine learning approaches for shale gas production forecast. Geoenergy Science and Engineering.2023, 226: p.211719.

[21] Q. Wang, H.W. Qin, C.S. Qi, et al. Improved XGBoost temperature prediction method based on SA-PSO. Electronic Measurement Technology.2023,Vol. 46 (No.7) , p. 67-72.

[22] J. Xue, B. Shen. A novel swarm intelligence optimization approach: sparrow search algorithm.Systems Science & Control Engineering: An Open Access Journal. 2020,Vol. 8 (No.1) , p. 22-34.

[23] R.C. Liu, J.X. Li, J. Liu, et al. A Survey on Dynamic Multi Objective Optimization. Chinese Journal of Computers. 2020, Vol. 43 (No. 7) , p. 1246-1278.

[24] X.Y. Hou, H.Y. Lu, M.D. Lu, et al. Bidirectional Learning Equilibrium Optimizer Combining Sparrow Search and Random Difference. Computer Science.2023, Vol. 50 (No.11), p. 248-258.

[25] C. Zhang, Y. He, L. Yuan, et al. Capacity Prognostics of Lithium-ion Batteries using EMD Denoising and Multiple Kernel RVM.IEEE Access.2017, p.1-1