

Analysis of Influencing Factors on Shale Gas Well Productivity based on Random Forest

Yu Zhou ^{1, a}, Zaixun Gu ^{2, b}, Changyu He ^{1, c}, Bing Yang ^{2, *} and Jian Xiong ^{2, d}

¹ CNPC Chuanqing Drilling Engineering Company Limited, Chengdu, Sichuan 610051, China

² Research Institute of Big Data, School of Business Administration, Southwestern University of Finance and Economics, Chengdu Sichuan 611130, China

* Corresponding author: Bing Yang (Email: yb1995@foxmail.com), ^a zhouy153@cnpc.com.cn,

^b guzaixun_sc@cnpc.com.cn, ^c hecy_dyy@cnpc.com.cn, ^d xiongjian2017@swufe.edu.cn

Abstract. The objective of this study is to examine the primary controlling factors influencing shale gas well production. To this end, the research employs a random forest model to analyze the production capacity influencing factors based on the production data of a shale gas field in the Sichuan Basin. First, the process of analyzing the production capacity influencing factors based on the random forest model is designed for the high-dimensional, large samples, and multiple covariance characteristics of shale gas production data. The significance of each feature was evaluated through the impurity method and the out-of-bag data method, which identified the pivotal factors influencing the production capacity of gas wells across different time periods. To eliminate the potential influence of production enhancement measures on data analysis, this study further screened the core geological and engineering parameters and evaluated their significance. This research not only offers theoretical insights into shale gas exploration and development but also provides a reference point for predicting the production capacity of shale gas fields with comparable geological characteristics.

Keywords: Random Forest; Shale Gas; Factor Analysis; Capacity Forecasting.

1. Introduction

Shale gas as an unconventional energy source has become a hot spot for oil and gas exploration and development [1, 2]. Shu Honglin et al. [3] established evaluation criteria for deep shale gas favorable areas and preferred favorable areas for deep shale gas exploration. Huang Minghua et al. [4] summarized and briefly discussed the key exploration, drilling, testing and production technologies involved in shale gas production. Zhang Haijie et al [5] established a shale gas reservoir characterization method and analyzed the longitudinal and transverse spreading law of shale gas reservoirs. To further analyses the production capacity of gas wells, Hu Xi et al. [6] studied the geological characteristics of shale gas and the main controlling factors of enrichment and high production. Qing Peng [7] analyzed the decreasing characteristics of gas wells over the whole life cycle, and clarified the production and pressure decreasing law of deep shale gas wells.

Scholars have also conducted a number of studies on the importance analysis of indicators using random forests. Wang Wenjie et al. [8] studied the genetic variable screening method based on random forest variable importance score. Cao Taoyun [9], on the other hand, through the random forest algorithm, elaborated on variable importance measures such as minimum depth, variable split frequency of root nodes, replacement importance and Gini node purity in the regression problem. Based on the predictive performance of random forests, such as substitution importance and node purity, the variable importance measures of random forests can be calculated. In addition, the study by Janitza et al. [10] proposed an evaluation method based on the AUC statistic to overcome the effect of unbalanced category proportions in the dataset. Meanwhile, feature importance can also be calculated based on the construction process of tree models in random forests, such as the minimum depth and split frequency of root nodes.

In order to further explore the relationship between shale gas well production and related well characteristics, this paper takes the production data of a shale gas field in Sichuan as an object of

study, and points out the main well characteristics affecting the cumulative production under different cycles through the shale gas production capacity influencing factor analysis model. The degree of influence of geological and engineering parameters is further analyzed to provide theoretical reference for shale gas exploration and development with similar geological background.

2. Method of Analyzing Factors Influencing Production Capacity

2.1 Capacity Impact Factor Analysis Process

Conventional correlation analysis methods are usually based on econometric models, but the production data in shale gas wells are high-dimensional and large samples, and the variables have multicollinearity, making it difficult to describe the complex mechanism of interaction between variables using conventional linear and polynomial models. Tree models are usually used to study the numerical distribution of each characteristic, and can assess the importance of the characteristic on the response variable based on its information entropy index, which is suitable for research scenarios with a large number of characteristics and large data samples. The technical route of the analysis of production capacity influencing factors is shown in Figure 1.

According to the technical route of capacity influencing factor analysis shown in Figure 1, the problem is first set up as a regression problem, and the key features are taken as independent variables, and the cumulative production of different cycles is taken as the dependent variable to establish a regression tree model (this paper uses random forest). The tree model is then adopted as the basic tool for feature importance analysis, and the importance of each feature in the dataset is analyzed using two evaluation tools: feature impurity method and out-of-bag data method.

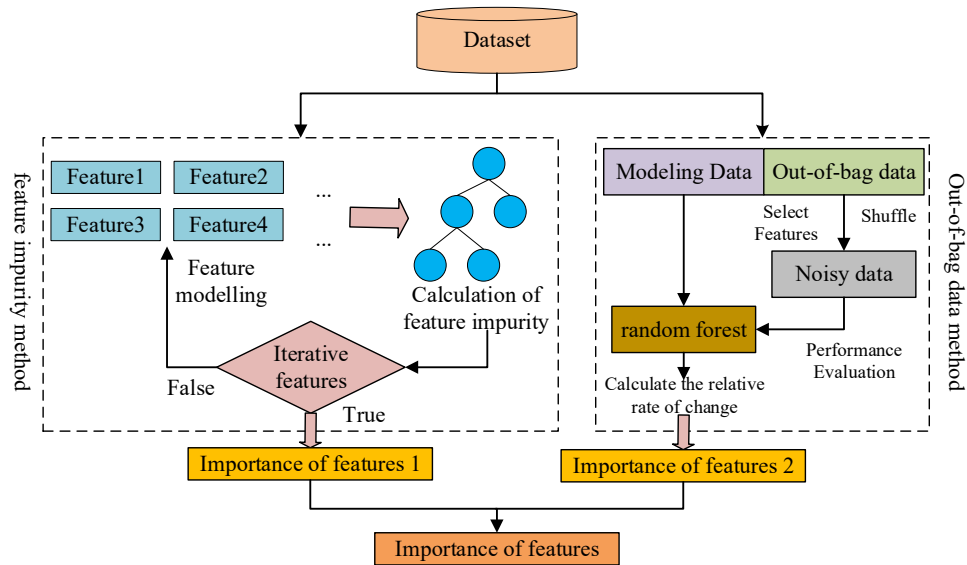


Fig 1. Technical route for analyzing the factors influencing production capacity

2.2 Principles of Random Forest Algorithm

The Random Forest (RF) algorithm is a parallel integrated learning method of many decision trees [11] that can perform variable significance analysis, singular value detection, missing data interpolation and survival analysis, in addition to classification and regression of datasets [12].

2.2.1 Decision Tree

Decision tree is a classic tree-based machine learning model with an intuitive and explanatory structure, so it is widely used in the field of data analysis. According to the difference of the splitting criterion of the nodes, the common decision tree algorithms include ID3 (Iterative Dichotomiser3) algorithm based on information gain, C4.5 algorithm based on information gain ratio, and Cart

(Classification and Regression Tree) algorithm based on Gini coefficient. In the Random Forest algorithm, the binary decision tree is the basic learner and is suitable for classification prediction problems for different types of data.

However, the ID3 algorithm chooses information gain as the loss function, has a selection bias for multi-class features, and is unable to handle continuous features. Although the C4.5 algorithm improves on ID3 by using the information gain rate (Gini index) [13] as the loss function, it uses a single-point discretization for continuous features, resulting in a large time overhead, and is unable to handle regression problems. In contrast, the Cart decision tree uses the Gini coefficient as a splitting criterion to select features by minimizing the Gini coefficient, which avoids the problem of the information gain criterion favoring attributes that take on more values and is computationally efficient.

The Gini coefficient is a measure of the imprecision of the model; the smaller its value, the lower the imprecision, the better the classification ability of the features, i.e. the lower the uncertainty of the features. Assume that the classification of the sample takes the value M , 0 or 1, and P_m indicates the probability that the sample x_i takes the value 1. The Gini coefficient, $Gini(D)$, reflects the probability that two samples are randomly drawn from the data set D and that their category labels do not match. Therefore, the smaller the value of $Gini(D)$, the purer the data set D . For the dichotomous data set D , the Gini coefficient is calculated as follows:

$$Gini(D) = \sum_{m=1}^M P_m(1 - P_m). \quad (1)$$

For the data set D , where the number of data is $|D|$, a binary tree is constructed by dividing the data set D into two parts, the data sets D_1 and D_2 , according to whether the attribute A takes a certain possible value a . Under the condition that the attribute A takes the value a , the Gini coefficient of the sample D is defined as:

$$GiniIndex(D|A=a) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2). \quad (2)$$

According to the formula of Gini coefficient in the above equation, calculate the Gini coefficient $GiniIndex(D|A=a)$ of the two data subsets, if the Gini coefficient is less than the set threshold, stop the recursion and return to the cart regression subtree. At the same time, calculate the Gini coefficient under different values of each feature in the current node, and select the feature with the lowest uncertainty in the Gini coefficient and the value of the feature as the partition point of the optimal splitting feature and the optimal partition feature. Using the optimal split feature and the value of the feature division point, perform the data set division to obtain two data subsets as the corresponding sub nodes. Repeat the above steps for the generated child nodes to obtain the final cart classification tree.

2.2.2 Random Forest

Random Forest is a supervised learning algorithm based on the integration strategy of Cart Tree [14] and Bagging [15]. The Random Forest algorithm introduces random attributes into the model training process. The training data of each cart tree is obtained by random sampling, while each decision tree randomly selects some of the features for training during the training process.

Specifically, Random Forest randomly selects M samples from a training dataset of size N for K selections each time there is a return. Based on these K sub-training data sets, K different decision trees are constructed. When constructing each cart tree, the averaging method is used to decide the final result for regression problems, while the voting method is used to decide the final result for classification problems. The flow of the Random Forest algorithm is shown in Figure 2.

Figure 2 shows the specific flow of the Random Forest algorithm. In the classification problem, the combiner uses the majority classification result of the voting method [16] as the final prediction result. And in the regression problem, the combiner takes the average of multiple regression results as the final prediction result.

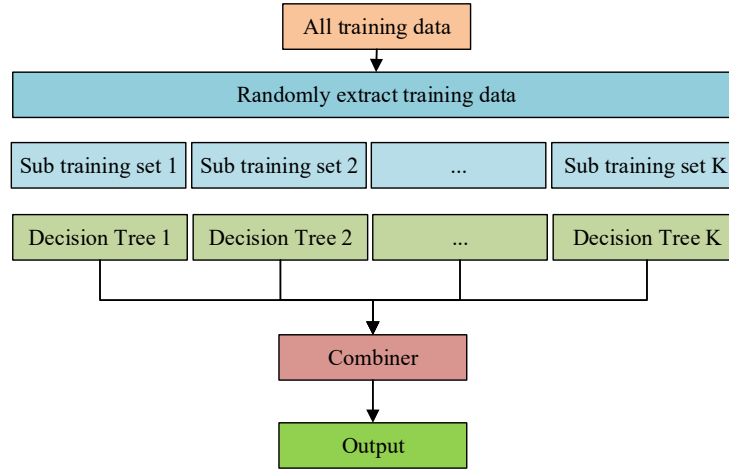


Fig 2. Principle of Random Forest Algorithm

2.2.3 Prediction Model based on Random Forest

The main steps to implement the shale gas production capacity prediction model based on random forest are as follows:

1. Data preparation. Collect shale gas production capacity related training datasets, including the feature data of shale gas wells and the corresponding production capacity data, and preprocess the feature data, including operations such as data cleaning, feature selection and feature conversion. A ratio relationship of 0.8:0.2 is adopted as the training data and prediction data of the regression model;
2. Random forest establishment. Random forest builds multiple decision trees and uses the idea of bagging to fuse them to obtain a more accurate and stable model. When a prediction is needed for a certain sample, the prediction results of each tree in the forest for that sample are counted and the final prediction result is selected from these prediction results by voting method. By integrating the prediction results of multiple decision trees, the random forest can reduce the prediction bias of a single decision tree and improve the stability and generalization ability of the model;
3. Model parameters are judged to be optimal or not. For the Random Forest algorithm, the hyper-parameters, such as the number of decision trees in the Random Forest and the maximum depth of decision tree learning, are cyclically traversed through the grid search method, and the optimal features are selected for node splitting according to the Gini coefficient or other indicators;
4. Model prediction and evaluation. The samples to be predicted are predicted using the constructed random forest model. The metrics function in Sklearn, a model computation library in Python, is used to help evaluate the performance of the model on the capacity prediction task. In addition, by accessing the `feature_importances_` attribute of the model, information on the importance of each feature in the model prediction can be obtained to perform the analysis of capacity influencing factors;
5. Feature Importance Analysis. Using the random forest model, the importance of each feature in the prediction can be calculated to help analyses the influencing factors of shale gas production capacity. Based on the results of feature importance, feature selection, model optimization and further research are carried out.

2.3 Feature Importance Analysis

There are two methods for assessing the importance of features in random forests, the feature impurity method and the out-of-bag data method, and in this paper we average the importance derived from both methods.

2.3.1 Feature Impurity Method

The impurity index can be used to evaluate the importance of features. The principle is to measure the importance of the corresponding features by the difference in impurity before and after the branching of the tree model node.

Calculate the impurity of feature j at node k . The calculation formula is as follows:

$$G(k, j) = \frac{1}{N_{left}} \sum_{y_i \in X_{left}} (y_i - \bar{y}_{left})^2 + \frac{1}{N_{right}} \sum_{y_i \in X_{right}} (y_i - \bar{y}_{right})^2. \quad (3)$$

Calculate the difference in impurity before and after the k -node split to measure feature importance.

$$VIM(k, j) = w_k * G_k - w_{left} * G_{left} - w_{right} * G_{right}. \quad (4)$$

where w_k , w_{left} and w_{right} are the ratios of the number of samples in node k and its left and right child nodes to the total number of samples, and G_k , G_{left} and G_{right} are the impurities of the node and its left and right child nodes, respectively.

If feature j appears in the node set M , then the importance of feature j in the i^{th} tree model is:

$$VIM_i(j) = \sum_{k \in M} VIM_i(j). \quad (5)$$

There are N decision trees in the random forest, and the total importance of feature j is:

$$VIM(j) = \sum_{i=1}^N VIM_i(j). \quad (6)$$

2.3.2 Out-of-bag Data Method

Out-of-bag data refers to the part of the data that is not used for modelling when constructing the sub-dataset through repeated sampling, and thus can be used to evaluate the performance of the tree model, the evaluation metric is called out-of-bag data error.

1. For the trained as-you-go forest model, select the corresponding out-of-bag data to calculate the error, denoted as E_0 ;
2. Add noise perturbation to the j -th feature of the out-of-bag data (re-perturbing the corresponding feature values for each sample) and calculate the out-of-bag data error again, denoted as E_j ;
3. Calculate the relative rate of change of the error, denoted as $(E_0 - E_j)/E_0$, to obtain the importance index of the j th feature;
4. Perform 1. to 3. for all features to obtain the importance index of all features.

If the accuracy of the out-of-bag data is significantly reduced after adding random noise ($E_0 < E_j$), it means that the feature has a great influence on the prediction result of the data and has a high importance.

3. Analysis of Factors Influencing the Productivity of Gas Wells

3.1 Data Description

The data for this study are derived from the production data of a shale gas field in Sichuan, which contains daily production data of 311 wells, with the maximum number of production days being 3138 and the minimum being 48. The data includes three areas: "W2", "W4" and "W5". "W2" includes 102 gas wells, "W4" includes 206 gas wells, and "W5" includes 3 gas wells.

The main features of each gas well are: Class I reservoir thickness, Class I reservoir continuous thickness, Long-11 thickness, Long-11 Class 1 thickness, horizontal length of Long-11, pressure coefficient of Long-11, drilling rate of Long-11, drilling rate of Long-11, horizontal section length, number of clusters in a single section of the main body, number of main holes in a single section, cluster spacing, main perforation density, sand addition intensity, liquid intensity, segment length, casing change, combined pressure length, lost section length, high-quality shale drilling rate, actual number of perforation clusters, converted effective number of sections, 100 mesh quartz sand, 40/70 mesh quartz sand, 40/70 mesh ceramist, 30/50 mesh ceramist, total sand volume, acid, slippery water, linear gel, weak gel, total liquid volume, average single-section sand volume, average single-section liquid volume, ceramist ratio, average pump-off pressure, displacement, pump-off pressure at the end

of fracturing, well soaking time, wellhead casing pressure when starting to discharge, cumulative discharge volume before production, flowback rate before production, well spacing, and reserve abundance. The model is established based on the 90-day cumulative production of gas wells, the first year's gas production, the second year's cumulative gas production, the third year's cumulative gas production, the fourth year's cumulative gas production, the fifth year's cumulative gas production, and the sixth year's cumulative gas production, respectively, to analyze the main factors affecting the gas well productivity at different time lengths. The number of days counted in the unit "year" is 330 days. Since the number of production days for each well is inconsistent, gas wells with a production time of less than 90 days are excluded. Table 1 shows the number of gas wells in different production cycles, and Figure 3 shows the average annual gas production distribution of 242 gas wells in 6 years.

Table 1. Number of gas wells in different production cycles

90-day	First year	Second year	Third year	Fourth year	Fifth year	Sixth year
242	215	173	102	53	36	22

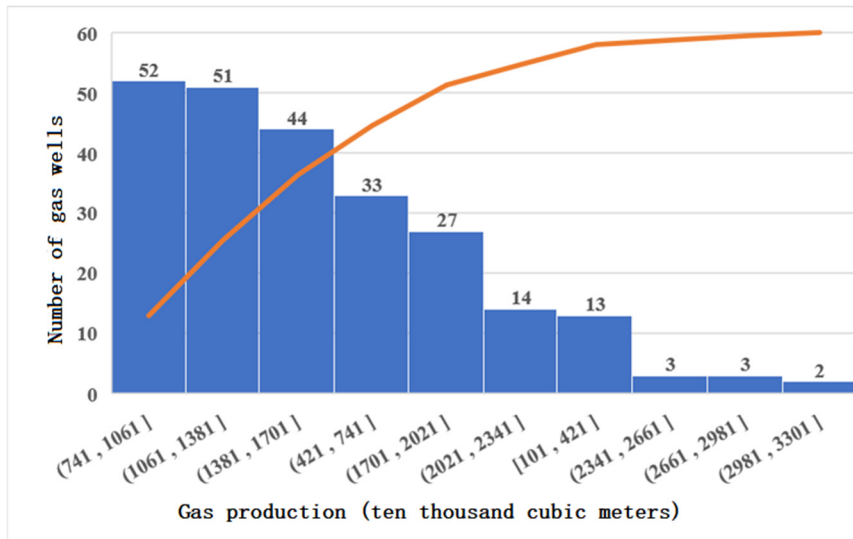


Fig 3. Average annual gas production distribution over 6 years

3.2 Model Training

The Python language has great advantages in data analysis and interaction, exploratory computing and data visualization. It also integrates a rich set of third-party extension libraries. In this paper, the Python library is used to build models, perform prediction and analysis, and compute various evaluation metrics to evaluate the model.

In machine learning models, there are usually some hyperparameters that need to be specified before model training, such as the depth of the decision tree, the number of cart subtrees in the random forest, the maximum number of layers, and the maximum number of nodes in the cart subtree. The choice of these hyperparameters will affect the performance of the model, so it is necessary to adjust the hyperparameters to find the optimal model. The random forest model has many hyperparameters and it is difficult to manually set a good set of model parameters. Therefore, grid search and cross-validation are used to determine the hyperparameters of the model. Grid search [17] is a commonly used hyperparameter fitting method. Its principle is to find the optimal model by traversing all possible hyperparameter combinations. Cross-validation can help evaluate the performance of different hyperparameter combinations and prevent the model from overfitting and underfitting. Typically, K-fold cross-validation [18] is used for evaluation, which means that the data set is divided into K-folds, K-1 folds are used for training each time, and then the remaining fold is used for testing. This is repeated K times, with a different test fold selected each time. Finally, the results of the K

training times are averaged as an evaluation indicator of the model performance. The specific steps are as follows:

1. Specify the value range of the hyperparameters, set the number of decision trees $n_estimators$ to 25-75, set max_depth to the range 5-20, and generate a hyperparameter grid.
2. Use K-fold cross validation to train the model and evaluate the performance of each set of hyperparameter combinations.
3. Select the optimal hyperparameter combination according to the performance evaluation results.
4. Use the optimal hyperparameter combination to train the model, and use the test data on the test set as training samples and the remaining data as prediction samples to evaluate the effect of the model.

Predicting shale gas production capacity can be considered as a regression task. We use the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R-squared), which are commonly used in regression tasks, as indicators to evaluate model performance. The calculation formula is as follows:

$$MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (8)$$

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

During the evaluation process, the cumulative production over different production cycles was evaluated. From the 90-day cumulative production to the sixth-year cumulative production, the evaluation results on the test set were an average RMSE of 42.78, an average MAE of 42.78 and an average R2 of 42.78, which can meet the needs of practical applications.

3.3 Importance Analysis

Figure 4 shows the importance ratio of all gas well parameters, and the overall average importance ranking (Top 5) is: Long-11 Class 1 thickness (27.05%), Class I reservoir thickness (13.67%), 40/70 mesh ceramist (7.03%), reserve abundance (6.99%), horizontal length of Long-11 (3.55%). The importance ratio is prominent in that Long-11 Class 1 thickness is reflected in the cumulative production in the first five years, Class I reservoir thickness is reflected in the cumulative production in the third and fourth years, reserve abundance is reflected in the cumulative production in the sixth year, 40/70 mesh ceramist is reflected in the cumulative production in the first three years, Class I storage layer continuous thickness is reflected in the cumulative production in the fifth and sixth years, and horizontal length of Long-11 is reflected in the cumulative production in 90 days, the first year, and the sixth year.

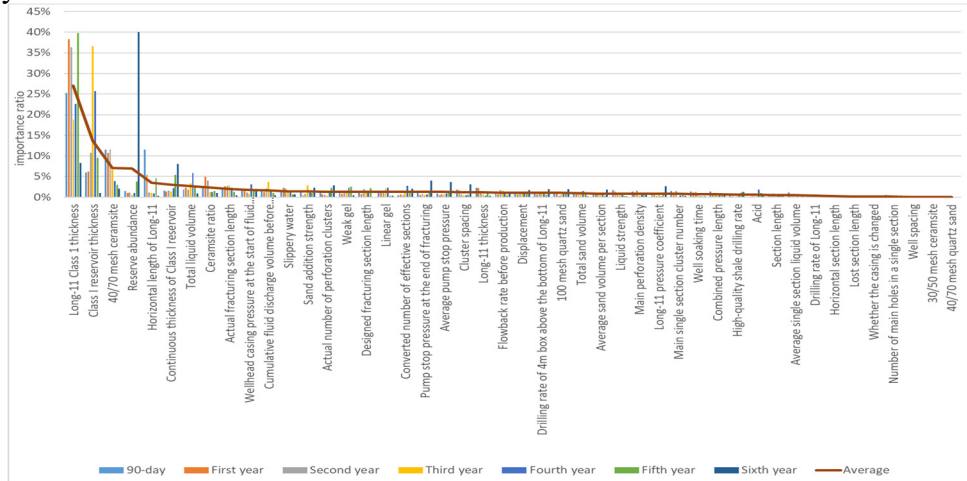


Fig 4. Percentage of importance of each parameter

Furthermore, the importance of all geological and engineering parameters was analyzed and the importance ratios of geological and engineering parameters for different time periods are shown in Table 2. In general, the importance of geological parameters is close to 50% of that of engineering parameters in the short term. As the cumulative production period increases, the importance of geological parameters increases. In other words, geological parameters are the main reason influencing long-term production.

Table 2. Importance ratio of geological parameters and engineering parameters.

Parameters	90-day	First year	Second year	Third year	Fourth year	Fifth year	Sixth year
geological	48.98%	55.15%	52.49%	59.54%	52.81%	64.98%	60.44%
engineering	51.02%	44.85%	47.51%	40.46%	47.19%	35.02%	39.56%

3.4 Analysis of Core Parameters Excluding the Impact of Production Increase Measures

The overlapping importance of some gas well parameters will lead to multicollinearity of the data. Therefore, some core variables with no mutual influence and independent physical meaning are selected to construct a new index system for analysis. In addition, due to the short production time of some gas wells and the addition of stimulation measures after production for a period of time, in order to expand the sample size and exclude the influence of measures, only 90-day and first-year cumulative production are selected as variables to evaluate the importance of some core parameters. The relevant gas well characteristics and the importance of the assessment are shown in Table 3.

Table 3. The importance ratio of core geological parameters and engineering parameters.

Parameters	90-day	First year	Second year
Core geological parameter	48.98%	55.15%	52.49%
Core engineering parameter	51.02%	44.85%	47.51%

According to the analysis of the importance ratio of core geological parameters and engineering parameters after excluding the production-increasing measures, after excluding some exogenous indicators, the importance ratio of geological parameters has improved. In terms of increase, the increase in first-year cumulative production is greater than the increase in 90-day cumulative production.

Table 4. The importance ratio of core engineering parameters.

Parameters	90-day	First year	Average
Thickness-1 sublayer	39.01%	28.75%	33.88%
Actual fracturing section length	15.40%	19.78%	17.59%
Cluster spacing	6.68%	7.47%	7.07%
BRMC4-1 sublayer	6.19%	5.21%	5.70%
Sand addition strength	4.15%	5.51%	4.83%
Fluid strength	4.39%	5.02%	4.71%
TOC-1 sublayer	4.56%	4.35%	4.45%
Drilling rate of Long-11	2.76%	5.77%	4.26%
POR-1 sublayer	4.18%	3.49%	3.84%
Number of main holes in a single section	2.79%	3.74%	3.26%
SG-1 sublayer	3.30%	3.16%	3.23%
Pressure coefficient prediction	2.69%	3.11%	2.90%
QALL-1 sublayer	2.56%	3.22%	2.89%
Number of clusters in a single section of the main body	1.34%	1.43%	1.38%

Table 4 gives the specific importance ratios of the core engineering parameters. The thickness-1 sublayer (geological parameter) is the most important indicator, with an average importance of 33.88%. The second most important indicator is the actual fracturing section length, which is about 17.59%.

4. Summary

This paper reveals the main control factors of shale gas well production capacity through in-depth analysis of production data from a shale gas field in Sichuan. Using the random forest model, combined with the impurity method and the out-of-bag data method, the importance of each characteristic in different production cycles was successfully evaluated, providing a scientific basis for predicting shale gas production capacity. The results show that the importance of geological parameters in long-term production capacity has gradually become prominent, and core parameters such as Long-11 Class 1 thickness and Class I reservoir thickness play a key role in different production stages.

In addition, by excluding the influence of stimulation measures, the paper further screens the core geological parameters and engineering parameters, and analyzes their importance. This work not only helps to improve the accuracy of productivity prediction, but also provides an important reference for the follow-up shale gas exploration and development.

In the future, with the continuous advancement of shale gas exploration and development technology and the continuous accumulation of data, we are expected to have a deeper understanding of the control mechanism of shale gas production capacity and contribute more wisdom and strength to the efficient development of shale gas.

References

- [1] Chen Gengsheng, Shi Xuewen, Liu Yong, et al. New understanding of the controlling factors of deep shale gas enrichment in the Wufeng Formation-Longmaxi Formation in the southern Sichuan Basin [J]. *Natural Gas Industry*, 2024, 44(01): 58-71.
- [2] Li Yuegang, Zhou Anfu, Xie Wei, et al. Lithofacies division and main controlling factors of reservoir development of shale in the Wufeng Formation-Longyi_1 submember in Luzhou area, southern Sichuan Basin [J]. *Natural Gas Industry*, 2022, 42(08): 112-123.
- [3] Shu Honglin, He Fangyu, Li Jilin, et al. Geological characteristics and favorable exploration areas of deep shale in the Wufeng Formation-Longmaxi Formation in Da'an Block, Sichuan Basin [J]. *Natural Gas Industry*, 2023, 43(06): 30-43.
- [4] Huang Minghua, Zheng Wen, Zheng Hui. Key technologies and applications in shale gas extraction process[J]. *Contemporary Chemical Research*, 2024, (08): 83-85. DOI: 10.20087/j.cnki.1672-8114. 2024. 08. 026.
- [5] Zhang Haijie, Mou Yongcheng, Han Guisheng, et al. Characteristics and distribution law of shale gas reservoirs in Upper Ordovician Wufeng Formation-Lower Silurian Longmaxi Formation in Dazu area, Sichuan Basin[J]. *Natural Gas Exploration and Development*, 2024, 47(04): 28-37.
- [6] Hu Xi, Li Yading, Dan Xia, et al. Geological characteristics and main controlling factors of shale gas enrichment and high production in the Wufeng Formation-Longyi_1 sub-member in Luzhou area, Sichuan Basin [J]. *Natural Gas Technology and Economy*, 2024, 18(04): 15-23.
- [7] Qing Peng. Analysis of dynamic characteristics and main controlling factors of deep shale gas [J]. *Petrochemical Technology*, 2024, 31(07): 238-240.
- [8] Wang Wenjie, Ma Jinsha, Gao Qian, et al. Variable screening method based on random forest variable importance score and its application in tumor typing diagnosis [J]. *Chinese Journal of Disease Control and Prevention*, 2023, 27(3): 274-280.
- [9] Cao Taoyun. Study on variable importance based on random forest [J]. *Statistics and Decision*, 2022(4): 4.

- [10] Janitza S, Strobl C, Boulesteix A L. An AUC-based permutation variable importance measure for random forests[J]. BMC bioinformatics, 2013, 14: 1-11.
- [11] Rigatti S J. Random forest[J]. Journal of Insurance Medicine, 2017, 47(1): 31-39.
- [12] Nicodemus K K. On the stability and ranking of predictors from random forest variable importance measures[J]. Briefings in bioinformatics, 2011, 12(4): 369-373.
- [13] Jiao Yanan, Ma Jie. An improved MEP decision tree pruning algorithm[J]. Journal of Hebei University of Technology, 2019, 48(6): 24-29.
- [14] Ma Yankun, Guo Chaoyou, Xu Zhe. Research on fault diagnosis method of marine diesel engine based on random forest[J]. Ship Electronic Engineering, 2020, 40(11): 128-131.
- [15] Xu Xiaoyang, Ji Zhicheng. Research on modeling of selective ensemble extreme learning machine classifier[J]. Computer Applications and Software, 2016, 33(9): 279-283.
- [16] Wang Lijia, Zhu Zhengwei, Zhu Yanping, et al. A multi-strategy multi-objective reinforcement learning algorithm based on voting method[J]. Computer Simulation, 2023, 40(04): 341-345+406.
- [17] Zhang Beibei, Hu Min. Customer satisfaction prediction based on grid search improved random forest[J]. Journal of Beijing Information Science and Technology University (Natural Science Edition), 2021, 36(4): 50-53.
- [18] Li Zhengliang, Peng Sisi, Wang Tao. Surrogate model sequence sampling method based on k-fold cross validation[J]. Chinese Journal of Computational Mechanics, 2022, 39(2): 244-9.