Open Access Library Journal
2023, Volume 10, e9963
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Research on Influencing Factors and Prediction Methods of Shale Gas Content Based on Machine Learning Algorithm

Fanjing Mao[1,2]

[1]School of Earth Sciences and Engineering, Xi'an Shiyou University, Xi'an, China
[2]Shaanxi Key Laboratory of Petroleum Accumulation Geology, Xi'an Shiyou University, Xi'an, China
Email: maofanjing424@163.com

How to cite this paper: Mao, F.J. (2023) Research on Influencing Factors and Prediction Methods of Shale Gas Content Based on Machine Learning Algorithm. *Open Access Library Journal*, **10**: e9963.
https://doi.org/10.4236/oalib.1109963

Received: March 3, 2023
Accepted: April 17, 2023
Published: April 20, 2023

Abstract

In order to serve the research of shale gas reservoir gas content and improve the prediction effect of shale gas reservoir gas content, this paper compares the factors influencing the gas content of shale gas reservoir and general prediction methods, and carries out the analysis of the factors influencing the gas content of shale gas and the response of sensitive logs in M1 well area, determines the main controlling factors of shale gas content, and establishes the prediction model of shale gas content in the well area on this basis. In order to improve the prediction accuracy, the CatBoost algorithm was introduced to build a shale gas content prediction model in the study area and compared with the measured gas content for verification; meanwhile, to verify the applicability of the model, the log data of the neighboring well F1 in the M1 well area were imported into the model to calculate its gas content and compared with its measured gas content for verification. The results show that the main influencing factors of shale gas content in M1 well block are total organic carbon content and pore specific surface area, etc. In the conventional shale gas content prediction model, adsorbed gas and free gas are calculated separately, and the summed gas content is larger than the measured gas content; the highest accuracy of multiple regression analysis is 0.702. The accuracy of the shale gas content prediction model established by applying CatBoost algorithm with well logging and testing data as input features and corresponding measured gas content as output labels is 0.986, which is better than the conventional algorithm; the model also has a higher accuracy in predicting the shale gas content of the neighboring well F1 in M1 well area.

Subject Areas

Geophysics, Petroleum Geology

Keywords

Shale Gas Content, Influencing Factors, Prediction Method, CatBoost Algorithm, M1 Well

1. Introduction

Shale gas refers to a kind of unconventional natural gas that is mainly located in organic shale and its thin interbed (sandstone, siltstone, high-carbon shale, etc.), and occurs in adsorption, free and dissolved state [1] [2]. Shale gas content is a key parameter for shale gas resource evaluation and target optimization, and an important standard for evaluating whether shale has exploitation value. Accurate prediction of shale gas content is particularly important for shale gas exploration and development. The influencing factors of shale gas content are complex [3] [4] [5]. The total organic carbon content, mineral content, porosity and water saturation all affect shale gas content to varying degrees. The size of gas content is the result of the comprehensive action of various factors.

At present, the methods used to evaluate shale gas content mainly include: field analysis, logging calculation, seismic inversion, experimental simulation, etc. [6]. The on-site analytical method can characterize the actual gas content of shale, but it is difficult to reasonably obtain the lost gas content. The experimental simulation isothermal adsorption method can obtain the theoretical adsorption capacity of shale by simulating the natural gas adsorption process, but it cannot reflect the actual gas content. Rick (2004) put forward a calculation model suitable for shale isotherm adsorption gas in the study area based on Lance isotherm adsorption equation; Chen Kang, Zhang Jinchuan (2016) took the shale of Longmaxi Formation in western Hunan and Hubei as the target layer, established a linear relationship between the data obtained from isothermal adsorption experiment and the TOC content, and obtained the adsorption gas content fitting equation. The common method to calculate shale gas content based on logging data is to calculate the content of adsorbed gas and free gas respectively, and then add the two to get shale gas content [6]. Zhang Zuoqing *et al.* [7], Zhao Jinzhou *et al.* [8] and Wen Kang [9] have successively improved the calculation method of adsorbed gas content, obtained the free gas content based on physical properties parameters, and added the two to get the total gas content, thus improving the prediction accuracy of shale gas content. In addition, some scholars directly establish a statistical relationship between logging, seismic and other parameters and gas content to predict shale gas content. Nie Haikuan [10] established a multiple regression equation based on shale reservoirs in Sichuan Basin to calculate shale gas content in the area; Zhang Yong [11] predicted shale gas content in the study area based on a deep neural network, which improved the prediction accuracy of shale gas content; Liu Jun [12] introduced the grey correlation analysis method to evaluate the gas-bearing property of shale reservoir; Sun Jianmeng (2014) proposed the calculation model method of shale adsorbed gas volume based on the research results of coalbed methane. Domestic and foreign scholars have made a lot of beneficial explorations and achieved many important results in the calculation of gas content in shale reservoirs, but the previous calculation methods and prediction models are restricted by certain use conditions and requirements.

In recent years, machine learning methods have been widely used in various geological parameter modeling. Chen Yuanyuan *et al.* predicted the total organic carbon content in western Chongqing based on particle swarm optimization support vector machine algorithm, and achieved good results [13]; Wang Jintao *et al.* predicted logging curves based on CNN-GRU neural network algorithm [14]; Based on two machine learning methods, Yang Zhanwei and others predicted the TOC content of shale from Wufeng Formation to Longmaxi Formation in South Sichuan and achieved good results [15]. Machine learning plays an important role in effectively improving the accuracy of data analysis [16]. As a kind of machine learning, CatBoost algorithm belongs to the integrated learning model based on the tree structure, which can effectively avoid over-fitting. The greedy combination and sorting promotion method used by CatBoost algorithm can enhance the data generalization ability, maximize the effective utilization of data, reduce the weight of abnormal data, and greatly improve the reliability and consideration of the model. Logging information has the advantages of good continuity and high vertical resolution. Combining CatBoost algorithm with logging information can achieve accurate data training and optimize geological parameter modeling.

Based on the analysis of the factors affecting the shale gas content in the study area, this paper applies the conventional prediction method and CatBoost algorithm to predict the shale gas content respectively, so as to optimize the prediction method, with a view to finding a fast and accurate method to predict the shale gas content in the study area using logging data, thus providing a new idea for the calculation of the shale gas content in the study area.

2. Overview

Well block M1 is located in the southeast of North China Block, and is bounded by Taihang Mountain, Zhongtiao Mountain, Huoshan Mountain and Mount Wutai respectively. It is a composite syncline formed in Yanshanian period under the Mesozoic compression background of the North China Plate, which is generally distributed in the direction of near NNE. See **Figure 1** for the geographical location of Well M1. In the Triassic system, the study area is in the stage of stable structural development and evolution, and a set of sandstone, siltstone, shale and other sand-shale interbedding strata have been deposited, with a large thickness of 800 - 2000 m. The shale interval developed in the Triassic system is the potential shale gas interval of the Carboniferous-Permian coal-bearing series, forming a good cap rock. The tectonic deformation conditions in the study area are weak, which is relatively favorable for the preservation of shale gas.

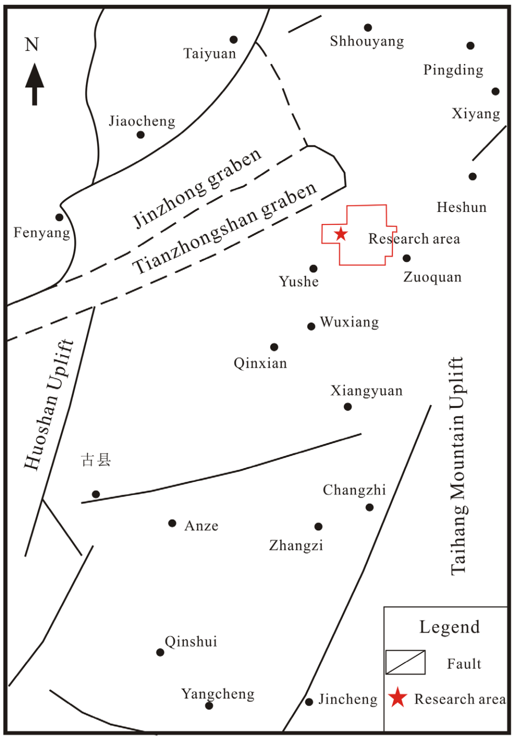

Figure 1. Geographical location of the study area [18].

The Upper Paleozoic Carboniferous-Permian shale is widely developed in the study area, and shale gas has great exploration potential [17]. The coal measures of the Shanxi Formation in the area develop typical marine and continental transitional facies shale, which is developed in the delta plain peat swamp deposits under the epicontinental marine sedimentary background, and transits from the delta estuary bar to the delta plain facies [18]. The instability of the sedimentary environment makes the lithological strata in the study area complex, mainly composed of mudstone, siltstone and sandstone, consisting of several cycles of gray shale, silty shale with white quartz sandstone and coal seams, alternating frequently, and often with coal lines. The vertical lithology changes greatly, the thickness is unstable, and the continuity is poor, which is "sandwich" superposition. The study area has a high abundance of organic matter, *with an* average *TOC* value of 2.98%, which has a strong hydrocarbon generation potential. The type of organic matter in the study area is mainly type III. The main body of shale reservoir is in the mature-high mature stage. The maturity of organic matter *Ro is* 2.56% on average. The organic matter has entered the dry gas window, generating a large amount of methane, which is conducive to the formation of shale gas reservoirs. The coal-measure shale reservoir of marine and continental transitional facies has the characteristics of low porosity and low permeability, and the pore type and pore structure are relatively complex, which is quite different from marine shale. The shale gas content in the area is 0.12 - 1.21 cm^3/g, with an average of 0.629 cm^3/g, reaching the lower limit of shale gas industrial development.

3. Factors Affecting Gas Content of Shale Reservoir

3.1. Influencing Factors and Mechanism of Shale Gas Content

The influencing factors of shale gas content are very complex. From the macro and micro perspectives, the influencing factors of shale gas content can be sorted into the following aspects (**Table 1**).

The content of total organic carbon is one of the most fundamental factors affecting the content of shale gas. It is not only the source base of shale gas generation, but also the core carrier of shale gas occurrence, which determines the gas content and hydrocarbon generation potential of shale. Therefore, the higher the content of total organic carbon, the greater the hydrocarbon generation potential and the higher the gas content. The organic matter maturity of shale reaches a certain degree and enters the gas window, which is the premise for shale formation to become a potential exploration target. With the high maturity of organic matter, the gas yield increases with the increase of gas yield; Clay minerals are another important carrier of shale gas, and have strong adsorption on gas. However, the stratum with high clay mineral content is relatively soft, which is not conducive to fracture development and shale gas exploitation; Shale gas is also mainly adsorbed on the pore surface of organic matter, and large specific

Table 1. Influencing factors and mechanism of shale gas content (according to literature [3] [4] [5] [6] [19]-[34]).

Influence factors		Impact mechanism
Macro factors	Rock type	The adsorption capacity of shale and silty mudstone is relatively high, while that of fine siltstone and limestone is relatively low; Siliceous shale and carbonaceous shale are conducive to the occurrence of shale gas.
	Temperature and pressure	With the increase of formation temperature, the gas content decreases; As the formation pressure increases, the gas content increases. When the pressure increases to a certain extent, the gas content increases slowly.
	Construction conditions	Positive structure and overpressure are conducive to shale gas enrichment.
Micro factors	Total organic carbon content	The higher the organic carbon content, the greater the hydrocarbon generation potential and the higher the shale gas content.
	Kerogen type	Type I kerogen contributes less to shale gas content, while type II and type III kerogen are favorable gas sources and contribute more to gas content.
	Maturity of organic matter	At the initial stage of gas generation, the higher the maturity of organic matter is, the more conducive to the increase of gas content; In the later stage, the gas content has a downward trend with the increase of maturity.
	Clay mineral	The high content of clay minerals is conducive to shale gas adsorption, but not conducive to fracture development.
	Porosity	The larger the pore specific surface area and total pore volume, the greater the content of adsorbed gas; The size of effective pore determines the content of free gas; The influence of fractures on gas content has two sides.
	Water content	The higher the water content, the lower the gas storage capacity and storage space, and the smaller the gas content.

surface area can provide favorable conditions for gas storage and help to improve shale gas content; Temperature, pressure and water content are important factors affecting shale gas content. The temperature of shale reservoir mainly affects the adsorption capacity of shale. Due to the exothermic adsorption process, as the temperature of shale reservoir increases, the gas gradually desorbs, the adsorption capacity decreases, and the gas content decreases. The presence of water in shale will not only reduce the gas adsorption capacity, but also reduce the gas storage space, thus reducing the shale gas content.

3.2. Analysis of Factors Affecting Shale Gas Content in M1 Well Block

The shale in M1 well area is a marine and continental alternating facies shale system. The instability of its sedimentary environment makes the lithology and physical properties of the target layer vary greatly in the vertical direction. Therefore, based on the core test data of different depths collected and collated in the M1 well block, this paper makes a statistical intersection between the shale gas content and the formation temperature (refer to the logging well temperature data), clay mineral content, total organic carbon content, specific surface area, porosity and vitrinite reflectance (Figure 2) to analyze their correlation.

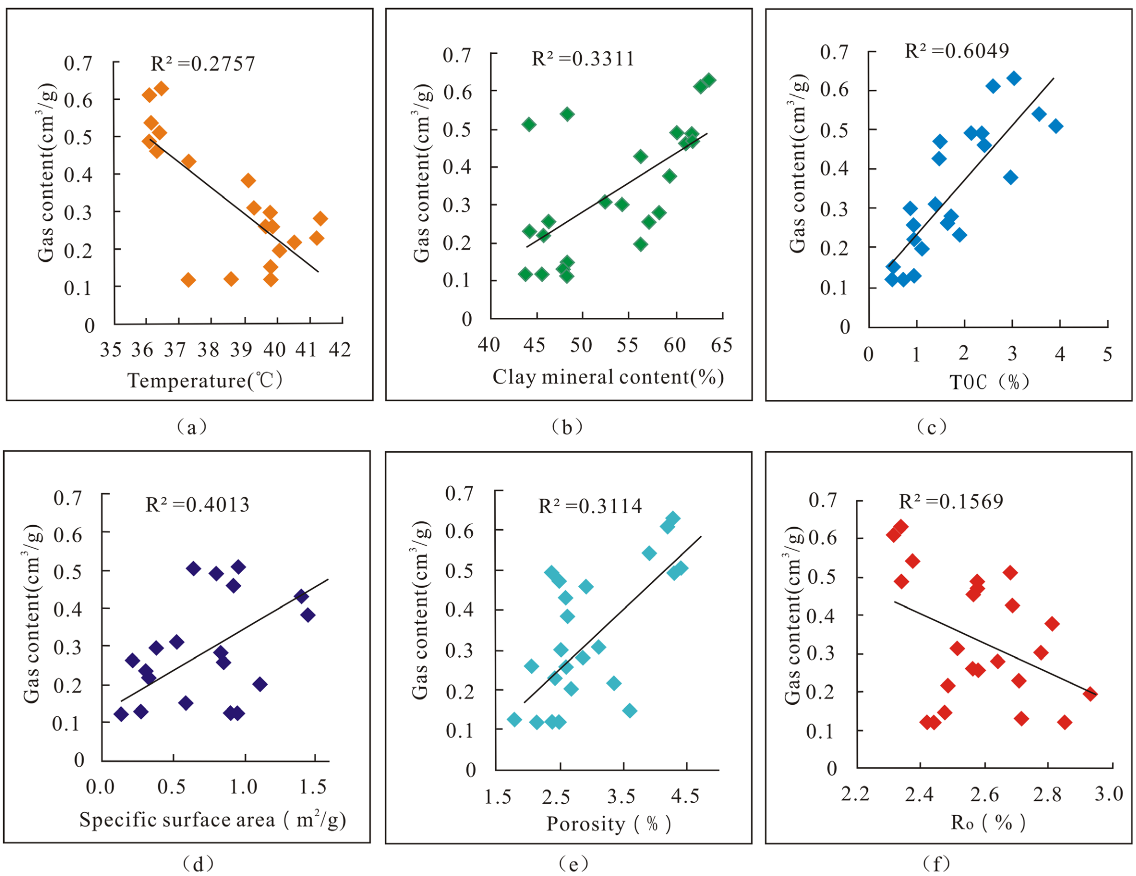

Figure 2. Crossplot of shale gas content and temperature (a), clay mineral content (b), total organic carbon content (c), specific surface area (d), porosity (e), vitrinite reflectance (f).

It can be seen from Figure 2 that the gas content in the study area has a positive correlation with the total organic carbon content, clay mineral content, porosity and specific surface area, and a weak negative correlation with the formation temperature and vitrinite reflectance. Among them, the correlation between gas content and total organic carbon content is the strongest, followed by specific surface area, and the correlation between other parameters is relatively poor. Therefore, the main control factors affecting the shale gas content in the study area are the total organic carbon content and pore specific surface area, and the total organic carbon content has a relatively strong impact.

4. Calculation Method and Principle of Shale Gas Content

At present, shale gas content calculation methods are mainly divided into two types. One is direct method, including analytical method based on laboratory core testing and analysis. Its calculation accuracy is relatively high, but it is subject to the influence of core quantity and analysis and test cost; The second is the indirect method, including the addition method of calculating the sum of adsorbed gas and free gas respectively, and the mathematical statistical model method based on the gas content of core test and logging data. The regression analysis method and nonlinear mathematical method belong to the mathematical statistical model method, and the model accuracy depends on the number of statistical samples and their representativeness.

4.1. Traditional Method for Shale Gas Content Calculation

1) Additive method

The calculation model of adsorbed gas content is as follows [8]:

$$G_s = \left(kP^n - bT\right) \times TOC \times \left(\frac{1}{1+\alpha m}\right) \tag{1}$$

where: G_s is the adsorbed gas content, unit gas content, cm^3/g; P is the reservoir pressure, MPa; T is the reservoir temperature (take the well temperature data obtained from logging), ˚C; TOC is the total organic carbon mass fraction (measured value of laboratory core), %; K, n and b are model coefficients; M is the mass fraction of water, %; α is a coefficient, indicating the influence of water on shale adsorption performance, and the value is 0.3.

Including:

$$m = \frac{S_W \times \varnothing}{\rho_R} \tag{2}$$

$$P = H \times 9.8 \times \left(\rho_R - \rho_w\right) \times 10^{-3} \tag{3}$$

where: ρ_R is shale density, g/cm^3; S_W is water saturation, %; Φ Is formation porosity, %; H is the burial depth of the sample point, m; ρ_W is the density of formation water, g/cm^3.

The calculation model of free gas content is as follows:

$$G_{f'} = \frac{1}{B_g} \times \varnothing S_g \times \frac{1}{\rho_R} \times 10^4 \quad (4)$$

where: $G_{f'}$ is the free gas content before calibration, cm^3/g; B_g is the gas compression coefficient, m^3/m^3; S_g is gas saturation, %.

Including:

$$S_g = 1 - S_w \quad (5)$$

$$B_g = \frac{ZTP'}{293P} \quad (6)$$

where: Z is the gas compression factor; P' is the pressure under standard condition (273.15 K), 0.1 MPa.

The free gas content shall be corrected for the volume of adsorbed gas content [9], and the correction formula is as follows:

$$G_f = \left(1 - \frac{\rho_t}{\rho_R}\right) G_f \quad (7)$$

where: G_f is free gas content, unit: cm^3/g; ρ_t is the density of adsorbed state of methane, taking 0.37 g/cm^3.

The calculation formula of total gas content is:

$$V = G_s + G_f \quad (8)$$

where: V is the total gas content, in cm^3/g.

2) Regression analysis method

The regression analysis method mainly combines the actual situation of the study area, selects the key influencing factors of the gas content in the study area, and uses the unit or multiple linear regression fitting method to fit the gas content according to the laboratory test, so as to establish the gas content calculation model.

$$Y = aX_1 + bX_2 + \cdots + k$$

where: Y is the total gas content, in cm^3/g; X_1, X_2 is the influence factor; a, b and k are coefficients.

4.2. CatBoost Algorithm to Calculate Shale Gas Content

1) Principle of CatBoost algorithm

CatBoost (Category Boosting) algorithm is an integrated learning model based on tree structure [35], which is an implementation of Boosting strategy. Since in the decision tree, the label average value is used as the criterion for node splitting, the expression is [36]:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} y_j}{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}}} \quad (8)$$

where: x_k^i is the i-th category feature of the k-th training sample; y_j for No j forecast tags; I is the indicator function.

The algorithm is based on a certain eigenvalue x_k^i when there is only one record, it is easy to cause over-fitting. The CatBoost model uses symmetric trees to establish mirror nodes. When calculating node gain, a priori value Q is used to reduce low-frequency category noise. The priori value Q *is* set as the average value of the prediction label in the data set, and it is given weight μ. It is helpful to reduce the impact of abnormal data on the overall data and improve the reliability and consideration of the model. The expression formula is:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} y_j + \mu Q}{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} + \mu} \tag{9}$$

where: Q is the added priori, μ is a priori weight, and the value is greater than 0.

This greedy method will cause prediction bias (bias) problem in each step of gradient upgrading [36]. CatBoost algorithm is improved by using the ordered boosting method, that is, by randomly generating a training sample with the order of [1, n], and the number of random sorting is σ. To train n different models $M_1, \cdots, M_n$, M_i only the first i samples in the array are used to learn the model [37]. This process can enhance the robustness of the algorithm and reduce noise interference.

In addition, CatBoost algorithm can continuously adjust the weight of each input feature during training, and the measurement formula is as follows:

$$J_j^2 = \frac{1}{M \sum_{m=1}^{M} J_j^2 \left(T_m\right)} \tag{10}$$

where: M is the number of iterations; J_j^2 represents the global importance of feature j; T_m is the decision tree of m nodes.

Therefore, CatBoost can ensure that all data sets can be used for training and learning by using sorting and upgrading, and can enhance the generalization ability, which can maximize the effective utilization of data compared with the strict splitting of data sets; CatBoost algorithm uses greedy combination, which can effectively improve the prediction accuracy. However, as a kind of machine learning, CatBoost algorithm model builds learning training based on a large number of data, and the accuracy of the model increases with the amount of training data. Therefore, when the training data is too small, the CatBoost algorithm may not be applicable.

2) Implementation steps of CatBoost algorithm

Calculate shale gas content based on CatBoost library in Python language. First, the data set is established according to the existing measured gas content and corresponding logging data for sample input, and it is divided into training set and test set by random sampling. The model is obtained by model training of the training set data with Catboat algorithm, and the test set is calculated with the trained model, and the model prediction result is obtained. The specific implementation steps are shown in **Figure 3**.

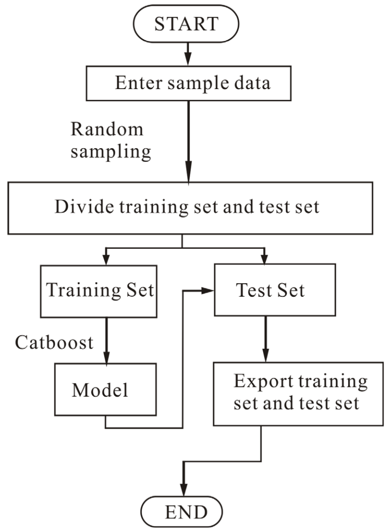

Figure 3. Prediction process based on Catboat algorithm model.

5. Experimental Verification

5.1. Calculation of Gas Content Based on Traditional Methods

Based on the analysis of influencing factors of shale gas content in the above-mentioned area, multiple linear regression analysis is carried out between gas content and various influencing factors (Table 2). It can be seen from Table 2 that the ternary linear regression relationship between shale gas content and total organic carbon content, specific surface area and clay mineral content is relatively good, and the prediction accuracy is 0.702.

Comparing the gas content calculated by the addition method and linear regression method with the measured gas content of the sample (Figure 4), it can be seen that the gas content calculated by the addition method is larger than the actual value, and the linear regression method has relatively good consistency, but the estimation accuracy still needs to be improved.

5.2. Gas Content Calculation Based on CatBoost Algorithm

First, analyze and standardize the logging data (including natural gamma (GR), lateral resistivity (RLLD), compensated neutron (CNL), compensated density (DEN), natural potential (SP), acoustic transit time (AC), etc.) of the target interval in the area, and analyze the correlation between the measured gas content and the corresponding logging response. See Table 3 for the results. It can be seen from Table 3 that the shale gas content has a good correlation with the normalized values of natural gamma, deep lateral resistivity, compensated neutron and compensated density.

In addition, according to the analysis in Figure 2, there is a strong correlation between the shale gas content and the total organic carbon content. In view of the limited sample test data, the total organic carbon content value of the target interval can be obtained indirectly from the logging data [19] [21] [22], which

Table 2. Multivariate regression calculation model analysis of gas content.

Regression project	Mathematical model of correlation analysis	Regression coefficient
Binary linear regression	$V = 0.0977*TOC + 0.1153*S + 0.0564$	$R^2 = 0.5969$
	$V = 0.1170*TOC + 0.0398*\Phi + 0.0147$	$R^2 = 0.5158$
	$V = 0.1157*TOC + 0.0101*C - 0.4028$	$R^2 = 0.6093$
	$V = 0.1712*S + 0.0878*\Phi - 0.0873$	$R^2 = 0.5201$
	$V = 0.0070*C + 0.1673*S - 0.1968$	$R^2 = 0.4277$
	$V = 0.0130*C + 0.1173*\Phi - 0.7052$	$R^2 = 0.5151$
Ternary linear regression	$V = 0.0776*TOC + 0.0397*\Phi + 0.1152*S - 0.0262$	$R^2 = 0.6178$
	$V = 0.0924*TOC + 0.0450*\Phi + 0.0103*C - 0.5059$	$R^2 = 0.6434$
	$V = 0.1023*TOC + 0.0510*S + 0.0082*C - 0.3266$	$R^2 = 0.7021$
	$V = 0.0816*TOC + 0.0741*S - 0.0285*T + 1.2171$	$R^2 = 0.6023$
	$V = 0.0974*TOC + 0.1000*S - 0.1689*Ro + 0.5079$	$R^2 = 0.5944$

Note: V in the table is the calculated gas content; S is the specific surface area; Φ Is porosity; C is clay mineral content; T is temperature; Ro is vitrinite reflectance.

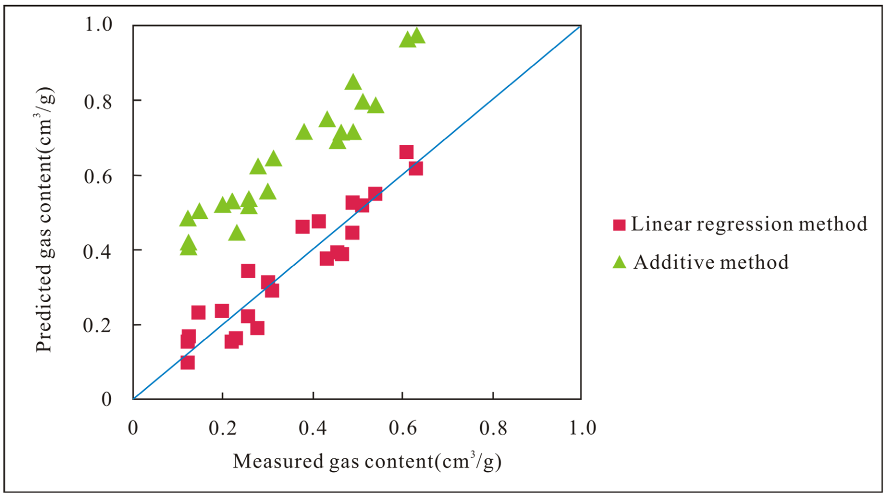

Figure 4. Comparison chart of total gas content prediction effect.

Table 3. Correlation analysis of measured gas content and logging response.

Logging response	GRRV	SPRV	ACRV	DENRV	CNLRV	RLLDRV
Correlation coefficient R^2	0.550	0.104	0.205	0.387	0.356	0.376

can be used as one of the input characteristics of the model to improve the prediction accuracy of the model. Therefore, this paper selects six parameters, namely natural gamma, deep lateral resistivity, compensated neutron, compensated density standardization value and total organic carbon content, as the input characteristics of the model, and the gas content test value is the comparison label Y. The input sample data set includes 375 groups of samples and 2250 data. The above data set is randomly divided into training set and test set, with the division ratio of 4:1, that is, 300 sets of samples for training set and 75 sets of samples for test set.

The setting of tree depth and learning rate will affect the model training results. See Table 4 for the comparison relationship. After many times of debugging, the tree depth is set to 5, the learning rate is 0.05, and the other values are the default values for model training.

The trained model predicts the gas content of the test set. In order to ensure the stability of the model, run the model several times, and end the operation when the prediction accuracy does not change, and get the shale gas content based on the algorithm model (Figure 5). It can be seen from Figure 5 that the prediction accuracy of Catboat algorithm can reach 0.986.

In order to verify the applicability of CatBoost algorithm to the prediction of shale gas content in the area, the logging data and core test data of Well F1 in the vicinity of Well M1 in the study area are selected to carry out the prediction and verification of shale gas content.

The total organic carbon content, natural gamma, deep lateral resistivity, compensated neutron, compensated density, acoustic transit time and other standardized values of the shale reservoir section in Shanxi Formation of Well F1 are imported into the above CatBoost algorithm model, and the original set values are used for data training. The comparison between the gas content

Table 4. Comparison table of different parameter commissioning (part).

Tree depth	4	5	5	5	6	6	6
Learning rate	0.03	0.04	0.05	0.07	0.03	0.04	0.05
Correlation coefficient R^2	0.75	0.903	0.986	0.981	0.950	0.925	0.978

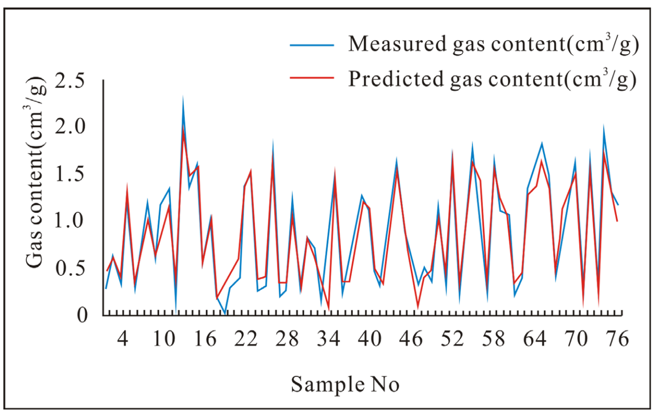

Figure 5. Comparison between predicted gas content and measured gas content.

calculated by the model and the actual gas content of the shale in the well is verified (Figure 6). The two sets of data R^2 is 0.760, indicating that the CatBoost algorithm is ideal for predicting the gas content of the shale in Well F1 in the study area, and further verifies the applicability of the CatBoost algorithm for predicting the gas content of the shale in Well M1 and its adjacent areas in the study area.

5.3. Comparative Analysis

Compare and analyze the calculation results of shale gas content in the study area with the traditional calculation method of shale gas content and CatBoost algorithm, see Table 5. The correlation coefficient R^2 indicates the correlation between the calculated result and the measured value, and the mean square error MSE indicates the error of the calculated result. It can be seen from Table 5 that the prediction accuracy of CatBoost algorithm is significantly higher than that of traditional calculation methods, and the calculation error is small.

To sum up, compared with the conventional calculation method of shale gas content, the CatBoost algorithm applied to the prediction of shale gas content in the study area can effectively improve the prediction accuracy, avoid prediction deviation, and have better applicability, which can provide important technical support for the exploration and development of shale gas in M1 well block and its adjacent areas. Therefore, CatBoost algorithm has certain feasibility and promotion value for predicting shale gas content.

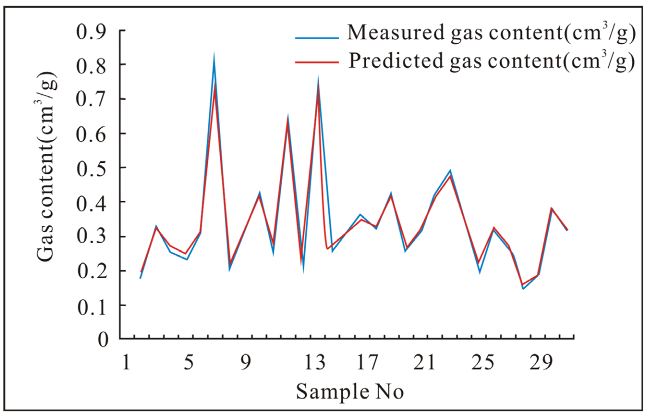

Figure 6. Comparison between predicted gas content and measured gas content of Well F1.

Table 5. Comparison of different calculation methods for shale gas content.

Method	Additive method	Regression analysis method	CatBoost algorithm
Correlation coefficient R^2	0.632	0.702	0.986
Relative error	0.153	0.015	0.002

6. Conclusions and Prospect

1) The influencing factors of shale gas content are complex. The correlation between shale gas content and total organic carbon content in well block M1 is the strongest, followed by the correlation with specific surface area. The main control factors of gas content are total organic carbon content and pore specific surface area. The size of gas content is the result of the comprehensive effect of various factors.

2) Shale gas content calculation methods are mainly divided into direct method and indirect method. In the conventional shale gas content prediction model of M1 well block, the calculation result of additive method is too large and the accuracy is low. The calculation accuracy of regression analysis method is 0.702, which needs to be improved. Using logging data and test data, the prediction accuracy of the shale gas content prediction model established by CatBoost algorithm for M1 well block shale gas content is 0.986, the prediction effect for its adjacent wells is ideal, and the application effect of the model is good.

3) Intelligent analysis has become a powerful tool for modern mathematical analysis and has unique advantages in solving complex nonlinear problems. It is an inevitable trend for future geological research to apply it to a wide range of geological parameter modeling and logging data processing, and also an important direction for the future development of shale gas-bearing evaluation research. At present, CatBoost algorithm is relatively less applied to the modeling of geological parameters of shale gas, so this model has broad application prospects.

Fund Project

Shaanxi Natural Science Basic Research Program (2019JM-359).

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Zhang, J.C., Jin, Z.J. and Yuan, M.S. (2004) Reservoiring Mechanism of Shale Gas and Its Distribution. *Natural Gas Industry*, **24**, 15-18+131-132. (In Chinese)

[2] Dong, D.Z., Qiu, Z., Zhang, L.F., *et al.* (2021) Progress on Sedimentology of Transitional Facies Shales and New Discoveries of Shale Gas. *Acta Sedimentologica Sinica*, **39**, 29-45. (In Chinese)

[3] Wu, J.S., Yu, B.S. and Li, Y.X. (2012) Adsorption Capacity of Shale Gas and Controlling Factors from the Well Yuye 1 at the Southeast of Chongqing. *Journal of Southwest Petroleum University* (*Science & Technology Edition*), **34**, 40-48. (In Chinese)

[4] Jin, P.P., Ou, C.H., Ma, Z.G., *et al.* (2018) Evolution of Montmorillonite and Its Related Clay Minerals and Their Effects on Shale Gas Development. *Geophysical Prospecting for Petroleum*, **57**, 344-355. (In Chinese)

[5] Zhang, J.Z., Li, X.Q., Zou, X.Y., *et al.* (2021) Pore Structure Characteristics of a Marine-Continental Coal-Bearing Shale Reservoir and Its Effect on the Shale Gas-

Containing Property. *Geochimica*, **50**, 478-491. (In Chinese)

[6] Zhang, J.C., Liu, S.G., Wei, X.L., *et al.* (2021) Evaluation of Gas Content in Shale. *Oil & Gas Geology*, **42**, 28-40. (In Chinese)

[7] Zhang, Z.Q., Sun, J.M., Gong, J.S. and Xia, Z.L. (2015) Gas Content Calculation Model of Shale Gas Reservoir. *Lithologic Reservoirs*, **27**, 5-14. (In Chinese)

[8] Zhao, J.Z., Shen, C., Ren, L. and Tan, X.C. (2017) Quantitative Prediction of Gas Contents in Different Occurrence States of Shale Reservoirs: A Case Study of the Jiaoshiba Shale Gasfield in the Sichuan Basin. *Natural Gas Industry*, **37**, 27-33. (In Chinese)

[9] Wen, K., Yan, J.P., Zhong, G.H., *et al.* (2022) New Method of Shale Gas Evaluation of Wufeng-Longmaxi Formation in Changning Area, Southern Sichuan Basin. *Lithologic Reservoirs*, **34**, 95-105. (In Chinese)

[10] Nie, H.K. and Zhang, J.C. (2012) Shale Gas Accumulation Conditions and Gas Content Calculation: A Case Study of Sichuan Basin and Its Periphery in the Lower Paleozoic. *Acta Geologica Sinica*, **86**, 349-361. (In Chinese)

[11] Zhang, Y., Ma, X.D., Li, Y.J. and Cai, J.S. (2021) The Application of Deep Learning to the Shale Gas Content Prediction in Nanchuan (South Sichuan). *Geophysical and Geochemical Exploration*, **45**, 569-575. (In Chinese)

[12] Liu, J. and Zhang, M.Y. (2021) Gas-Bearing Property Evaluation by Petrophysical Logging in Shale Gas Reservoirs: A Case Study in J Area of Fuling Shale Gas Field,sichuan Basin. *Petroleum Geology and Experiment*, **43**, 128-135. (In Chinese)

[13] Chen, Y.Y., Deng, X.J., Wang, X.L., *et al.* (2021) Application of a PSO-SVM Algorithm for Predicting the TOC Content of a Shale Gas Reservoir: A Case Study in Well Z in the Yuxi Area. *Geophysical Prospecting for Petroleum*, **60**, 652-663. (In Chinese)

[14] Wang, J.T., Wen, X.T., He, Y.L., Lan, J.L. and Zhang, C.M. (2022) Logging Curve Prediction Based on a CNN-GRU Neural Network. *Geophysical Prospecting for Petroleum*, **61**, 276-285. (In Chinese)

[15] Yang, Z.W., Jiang, Z.X., Liang, Z.K., *et al.* (2022) Evaluation of Shale TOC Content Based on Two Machine Learning Methods: A Case Study of Wufeng-Longmaxi Formation in Southern Sichuan Basin. *Lithologic Reservoirs*, **34**, 130-138. (In Chinese)

[16] Xiao, L.Z. (2022) The Fusion of Data-Driven Machine Learning with Mechanism Models and Interpretability Issues. *Geophysical Prospecting for Petroleum*, **61**, 205-212. (In Chinese)

[17] Zhu, Y.M., Zhang, Q.H., Qu, X.R., *et al.* (2015) Investigation and Development Potential Evaluation of Deep Shale Gas Resources in Qinshui Basin. Science Press, Beijing, 40-41. (In Chinese)

[18] Fu, J.J., Guo, S.B., Gao, Q.F. and Yang, J. (2016) Reservoir Characteristics and Enrichment Conditions of Shale Gas in the Carboniferous-Permian Coal-Bearing Formations of Qinshui Basin. *Earth Science Frontiers*, **23**, 167-175. (In Chinese)

[19] Zhao, J.L., Wang, Y.S., Yan, H.P., *et al.* (2021) Research on Prediction Method of Shale Reservoir Geochemical Parameters Based on Coal Field Logging Data. *Progress in Geophysics*, **36**, 2143-2150. (In Chinese)

[20] Zhang, Z. (2018) Study on Gas Content Characteristics of Shale Gas Reservoir in the Yushu Community Block of Qinshui Basin. Xi'an Petroleum University, Xi'an. (In Chinese)

[21] Zhang, B.Y. (2017) Research on Logging Characterization Method of Shale Reser-

voir Parameters. China University of Geosciences, Beijing. (In Chinese)

[22] Yu, G. (2016) Comprehensive Interpretation of Shale Gas Geophysics. East China University of Technology Press, Shanghai. (In Chinese)

[23] Zhang, B.M., Cai, Q.S., Chen, X.H., *et al.* (2021) Reservoir Characteristics and Gas-Bearing Capacity of the Wufeng-Longmaxi Formation in the Well Eyy2, East Huangling Uplift, Western Hubei Province. *Geology in China*, **48**, 1485-1498. (In Chinese)

[24] Guo, X.S. (2014) Rules of Two-Factor Enrichment for Marine Shale Gas in Southern China—Understanding from the Longmaxi Formation Shale Gas in Sichuan Basin and Its Surrounding Area. *Acta Geologica Sinica*, **88**, 1209-1218. (In Chinese)

[25] Li, J.Q., Lu, S.F., Li, W.B. and Cai, J.C. (2021) Evaluation on IN-Situ Gas-Bearing Characteristic of Shale Based on Process Analysis Method. *Journal of Southwest Petroleum University* (*Science & Technology Edition*), **43**, 42-55. (In Chinese)

[26] Li, J.Q., Lu, S.F., Zhang, P.F., *et al.* (2020) Quantitative Characterization and Microscopic Occurrence Mechanism of Pore Water in Shale Matrix. *Acta Petrolei Sinica*, **41**, 979-990. (In Chinese)

[27] Hu, Z.M., Duan, X.G., He, Y.B., *et al.* (2018) Influence of Reservoir Primary Water on Shale Gas Occurrence and Flow Capacity. *Natural Gas Industry*, **38**, 44-51. (In Chinese)

[28] Ji, L.M., Qiu, J.L., Zhang, T.W. and Xia, Y.Q. (2012) Experiments on Methane Adsorption of Common Clay Minerals in Shale. *Earth Science* (*Journal of China University of Geosciences*), **37**, 1043-1050. (In Chinese)

[29] Li, W.G., Yang, S.L., Xu, J. and Dong, Q. (2012) A New Model for Shale Adsorptive Gas Amount under a Certain Geological Conditions of Temperature and Pressure. *Natural Gas Geoscience*, **23**, 791-796. (In Chinese)

[30] Zhong, G.H., Xie, B., Zhou, X., Peng, X. and Tian, C. (2016) A Logging Evaluation Method for Gas Content of Shale Gas Reservoirs in the Sichuan Basin. *Natural Gas Industry*, **36**, 43-51. (In Chinese)

[31] Ambrose, R.J., Hartman, R.C., Diaz-Campos, M., Akkutlu, I.Y. and Sondergeld, C.H. (2010) New Pore-Scale Considerations for Shale Gas in Place Calculations. *The SPE Un-Conventional Gas Conference*, Pittsburgh, 23-25 February 2010, 219-229. https://doi.org/10.2118/131772-MS

[32] Li, Z. (2020) Sensitivity and Gas Content Change Effect of Niutitang Formation Shale Reservoir in Cengong Block, Guizhou. China University of Geosciences, Beijing. (In Chinese)

[33] Li, W.B., Lu, S.F., Li, J.Q., *et al.* (2019) The Coupling Relationship between Material Composition and Pore Microstructure of Southern China Marine Shale. *Natural Gas Geoscience*, **30**, 27-38. (In Chinese)

[34] Tang, Y., Li, L.Z. and Jiang, S.X. (2014) A Logging Interpretation Methodology of Gas Content in Shale Reservoirs and Its Application. *Natural Gas Industry*, **34**, 46-54. (In Chinese)

[35] Boschetti, A. and Massaron, L. (2020). Python Data Science Essentials. Trans. By J. W. Yu, Machinery Industry Press, Beijing, 15-158. (In Chinese)

[36] Scientific Platform Serving for Statistics Professional (SPSSPRO) (2021) Version 1.0.11. Online Application Software. https://www.spsspro.com

[37] Yuan, J.R., Zeng, X.J., Wu, H.Y., *et al.* (2021) Analytical Determination of Interwell Connectivity Based on Interwell Influence. *Tsinghua Science and Technology*, **26**, 813-820. https://doi.org/10.26599/TST.2020.9010039