ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Optimized Random Forest Method for 3D Evaluation of Coalbed Methane Content Using Geophysical Logging Data

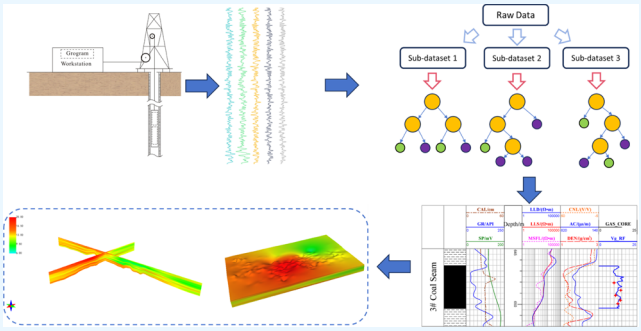
Jianhong Guo, Zhansong Zhang,* Guangshan Guo, Hang Xiao, Qing Zhao, Chaomo Zhang, Hengyang Lv, Zuomin Zhu, and Can Wang

Cite This: *ACS Omega* 2024, 9, 35769−35788

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Accurate evaluation of coalbed methane (CBM) content is crucial for effective exploration and development. Traditional gas content measurement methods based on laboratory analysis of drill core samples are costly, whereas geophysical logging methods offer a cost-effective alternative by providing continuous high-resolution profiles of rock layer physical properties. However, the relationship between CBM content and geophysical logging data is complex and nonlinear, necessitating an advanced prediction method. This study focuses on the No. 3 coal seam in the Shizhuang South Block of the Qinshui Basin, utilizing geophysical logging data and 148 sets of laboratory core samples. We employed the Random Forest (RF) method optimized with a simulated annealing-genetic algorithm (SA-GA) to develop the SA-GA-RF model for evaluating CBM content. The model's performance was validated using test data and new CBM well data, and it was applied to calculate the vertical gas content profiles of No. 3 coal seam across 128 wells. The SA-GA-RF model demonstrated an average relative error of 13.13% in the test data set, outperforming Backpropagation Neural Network (BPNN), Least Squares Support Vector Machine (LSSVM), Extreme Learning Machine (ELM), and multivariate regression (MR) methods. The model also exhibited strong generalizability in new wells and improved model-building efficiency compared to traditional cross-validation grid search methods. The construction of a three-dimensional CBM content model, incorporating well coordinates and elevation data, allowed for detailed identification of high gas content areas and layers. This three-dimensional model offers a more precise characterization than traditional two-dimensional isopleth maps, providing valuable insights for CBM exploration, reserve evaluation, and production optimization.

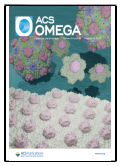


1. INTRODUCTION

Coalbed methane (CBM) resources have garnered global attention due to their potential to mitigate the depletion of traditional oil resources and environmental impacts.[1,2] Particularly amid concerns over coal mining safety and greenhouse gas emissions, the significance of CBM exploration and development has been underscored.[3,4] Substantial research efforts have been devoted to CBM development globally.[5−8] In China, abundant coal resources present extensive development prospects, making CBM exploration a primary focus in unconventional oil and gas resource research in recent years.[9,10] The evaluation of CBM resource productivity is crucial for its exploration and development,[11,12] yet estimating CBM resources remains challenging due to inherent uncertainties. CBM, classified as an unconventional gas resource,[13] exhibits reservoir and seepage mechanisms distinct from conventional natural gas, with resource estimates affected by multifaceted interactions.[14] Moreover, the complex interplay among various CBM reservoir parameters complicates their distribution and quantitative assessment.[15−18] Of these parameters, precise evaluation of gas content is particularly critical.[19,20] Assessing the gas-bearing properties of coal reservoirs spans the entire CBM development lifecycle,[21] from initial CBM selection to adaptive engineering design, production capacity construction, and treatment of low-production, low-efficiency wells, necessitating a thorough understanding of coal seam gas properties.[22,23] Given CBM's status as a clean energy source, accurate evaluation of CBM content represents a critical research frontier with significant implications for coalbed exploration and development.[24,25]

Currently, methods for evaluating CBM content can be classified into direct and indirect approaches. The direct method involves determining the gas content of core samples

Received: May 6, 2024
Revised: July 28, 2024
Accepted: July 31, 2024
Published: August 8, 2024



https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

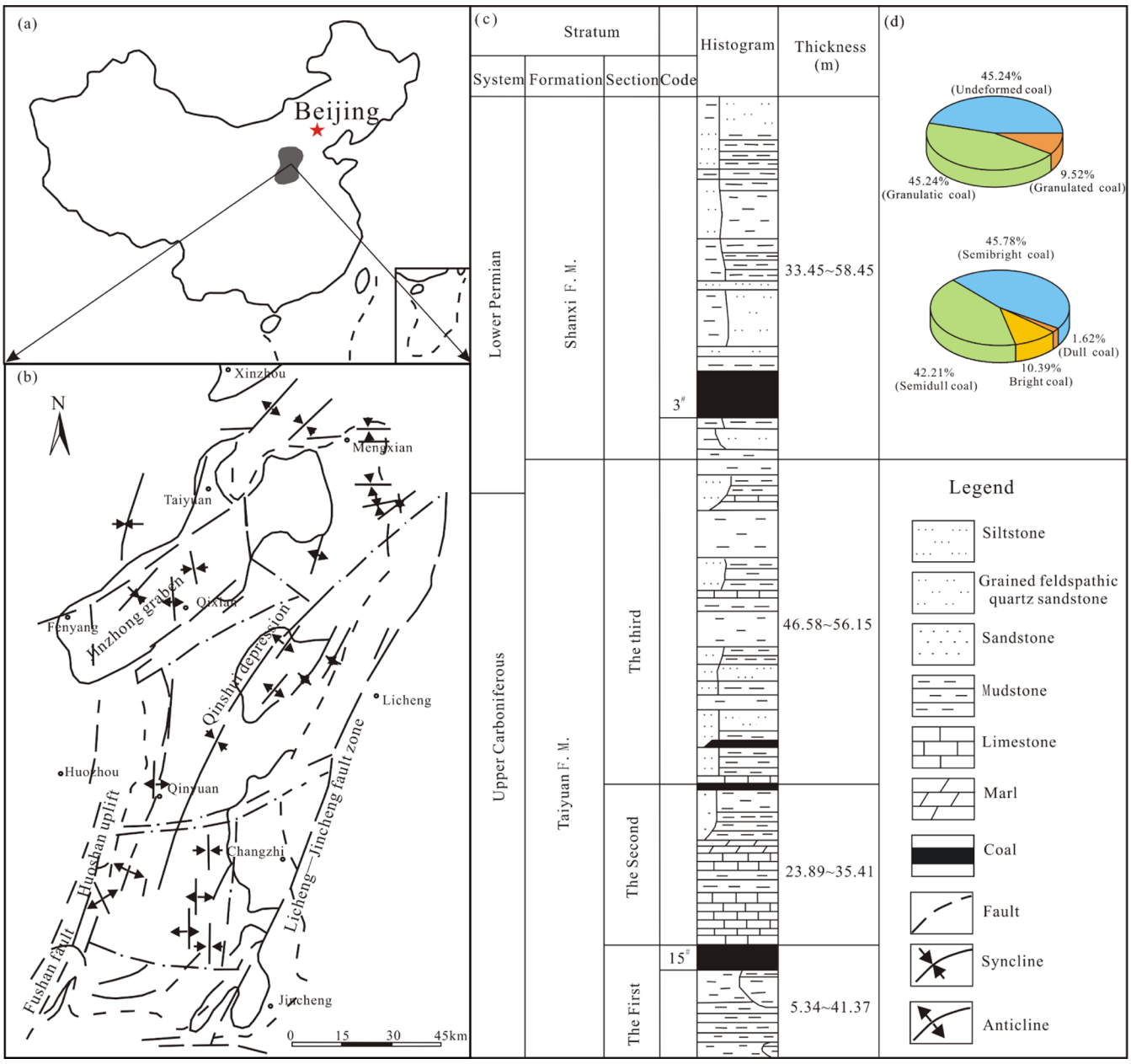


Figure 1. Geological map of the study area. (a) The study area is in Shanxi Province, China; (b) The geological survey map of the study area; (c) The geological histogram of the study area; (d) The pie chart of coal structure and macroscopic coal rock description type in the study area.

through desorption experiments conducted on cores obtained from closed boreholes.[26−28] This method is considered the most direct and accurate.[29] However, due to the poor mechanical strength of coal seams, cores often suffer from low integrity rates and potential damage to the borehole walls, resulting in insufficient experimental data on coal sample desorption.[30] Moreover, this method is costly, time-consuming, and impractical for widespread application since production wells typically do not undergo coring measurements across entire operational areas. In contrast, indirect methods address the limitations of direct methods. These approaches typically involve constructing models to evaluate CBM content using data from isothermal adsorption experiments and geophysical logging. Scholars internationally have developed various methods based on CBM reservoir mechanisms and desorption experiments. Notably, the KIM method and its refined equations calculate coalbed gas content using industrial components and have been proposed as effective indirect evaluation tools.[31] The isothermal adsorption model is established by experiment, and the content of CBM is predicted by the isothermal adsorption line. Subsequently, geophysical logging technology has been gradually applied to the evaluation of CBM content because of its high-cost performance, strong reliability, and high resolution to characterize the changes of vertical physical quantities of boreholes.[32−35] There are two main ways to evaluate the gas content by using geophysical logging data: the first is to combine the isothermal adsorption experiment method, use the geophysical logging data to analyze the coal industrial components, calculate the coefficient in the Langmuir equation based on the analysis results or correct this equation,[36] Kim[37] and Hawkins[38] used geophysical logging data to calculate

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

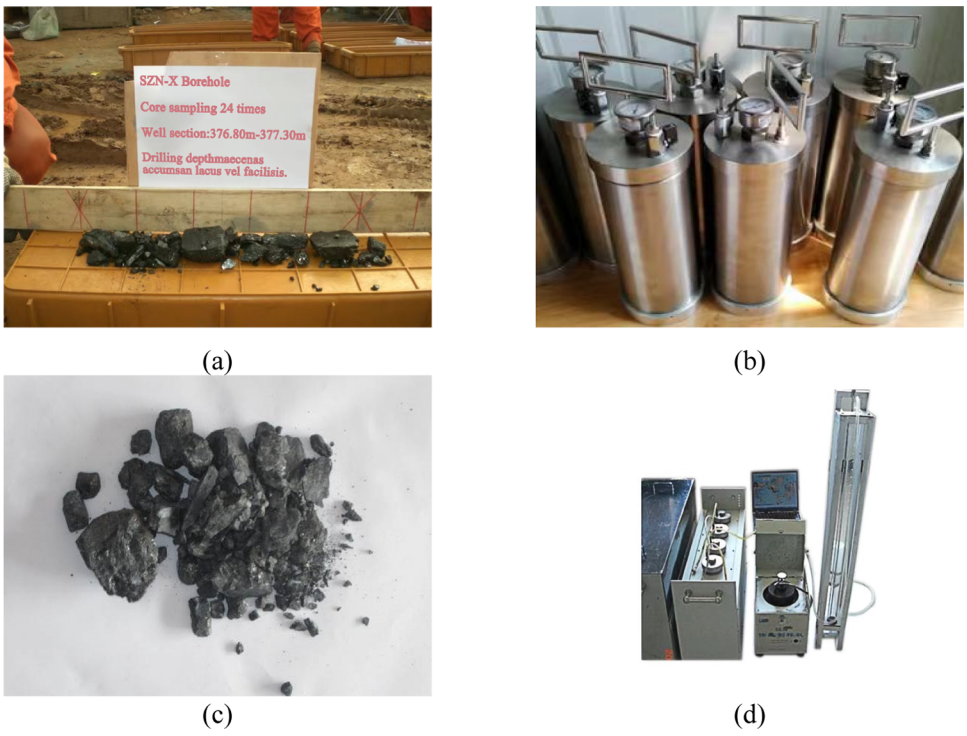


Figure 2. Coal seam coring, sample feeding, desorption and reading calibration diagram. (a) Coal seam core sampling; (b) Laboratory instrument for loading coal seam core; (c) Core desorption samples; (d) Reading of instrument recording parameters. Photograph courtesy of Guangshan Guo. Copyright 2024.

industrial components, and used the ratio of fixed carbon to volatile matter to calculate the coefficient calculation in the isothermal adsorption equation and dry ash-free basis correction, and then predicted the experimental gas content of coal samples and improved it. The precision of the isothermal adsorption method for evaluation is contingent upon field conditions, such as the number of parameter wells and the accuracy of bottom hole flow pressure testing in the field.[39] Another approach involves directly utilizing geophysical logging data to evaluate CBM content, including methods like the volume-model[40] and background-value[41] methods, which were initially prevalent but suffer from less-than-optimal accuracy. The selection of parameters significantly influences outcomes, and as industrial CBM content evaluation accuracy improves over time, the applicability of these methods has gradually diminished. Consequently, various mathematical techniques, encompassing multiple linear regression, nonlinear regression, and machine learning methods, have been employed. These methods combine geophysical logging data with core gas content derived from laboratory desorption to construct CBM content prediction models. Applied across diverse geographical contexts internationally, these approaches consistently demonstrate feasibility.[42−44]

CBM reservoirs are characterized by greater complexity and heterogeneity compared to conventional oil and gas reservoirs.[45] The logging response of coal seams is influenced by various factors, which complicates the relationship between coalbed parameters and geophysical logging data, making it nonlinear and intricate.[46,47] Directly identifying the principal controlling factors of CBM content proves challenging. Machine learning methods show promise in elucidating these nonlinear relationships within data, especially when dealing with the complex interplay between geophysical logging data and CBM content.[48,49] Consequently, these methods often achieve superior predictive accuracy compared to conventional multiple regression techniques. Currently, various machine learning approaches such as Back-propagation Neural Networks (BPNN),[50] Extreme Learning Machines (ELM),[51] and Support Vector Machines (SVM)[52,53] are employed to forecast CBM content. Empirical evidence suggests that these methods typically outperform traditional regression techniques.[54] However, machine learning methodologies have limitations; they require a substantial volume of core experimental data, ideally with a uniform distribution. Obtaining coal core samples that meet these stringent data quality requirements poses a significant challenge, complicating the development of predictive models. Accurately predicting CBM content in individual wells is essential for constructing a three-dimensional model of CBM content. While some scholars have proposed methods for constructing three-dimensional parameters of coal seams in the past,[12,55,56] there is limited research on three-dimensional modeling of CBM content. This study combines machine learning methods with geophysical logging data to construct a CBM content prediction model. To ensure the model's generalizability, a three-dimensional CBM content model was developed by evaluating CBM content from numerous wells across the study area and integrating borehole coordinates and elevation data obtained from remote sensing.

This paper introduces the Random Forest (RF) method and employs the simulated annealing (SA) -genetic algorithm (GA) to optimize its hyperparameters. The RF is known for its robustness with small sample sizes and unbalanced data distributions.[57] The SA-GA ensures efficient tuning of RF hyperparameters. The paper discusses the feasibility and applicability of this hybrid approach in predicting CBM content. It is applied to the Shizhuang South Block of Qinshui Basin and validated against actual production data. Results demonstrate that the enhanced random forest model, utilizing

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

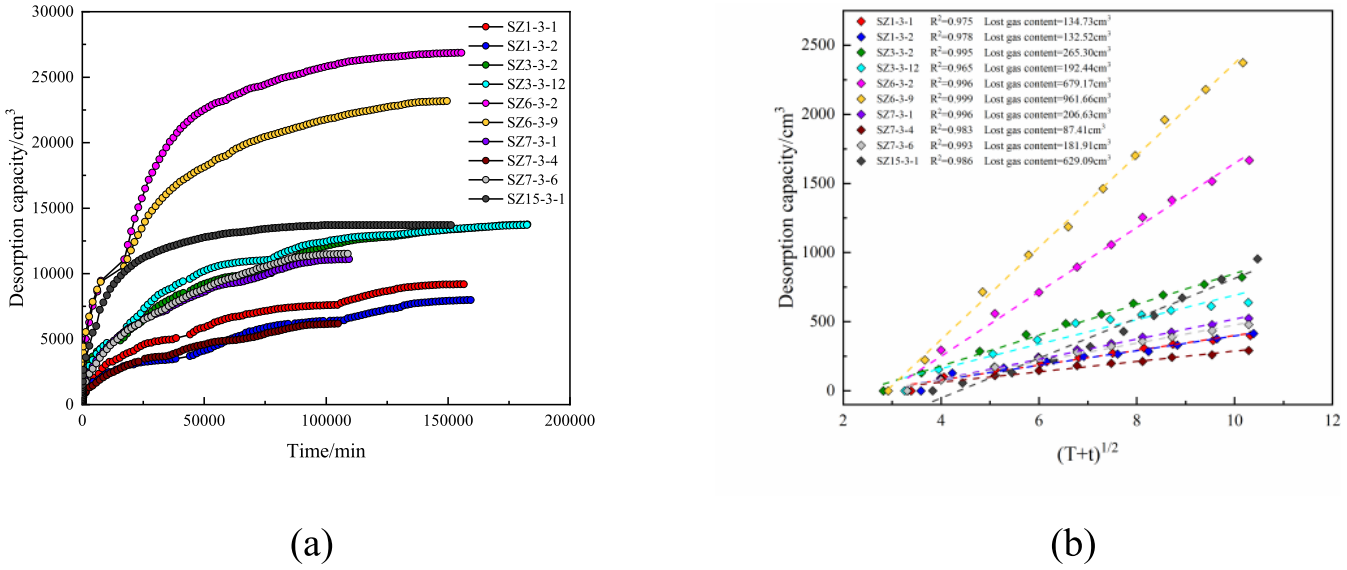


Figure 3. Coal seam core sample desorption data display. (a) The cumulative desorption content curve of coal seam samples; (b) Calculation of lost gas content.

geophysical logging data, effectively assesses CBM content, providing a reliable basis for constructing a three-dimensional model. This model offers valuable guidance for subsequent exploration and development, with practical engineering significance and application value.

2. GEOLOGICAL OVERVIEW

Located in the southeast of the Qinshui Basin (Figure 1a), Shanxi Province, China, the Shizhuang South Block stands out as a prominent area for CBM exploration.[58] Situated at the southern terminus of the Qinshui syncline, the block exhibits a predominantly monoclinic structural configuration (Figure 1b). Secondary folds, trending NNE in the west and NNS in the east, characterize the basin's western and eastern regions,[59] respectively. Fault activity is minimal, with a gentle average strata dip of approximately 5°. The geological history of the region has been marked by various tectonic events, including the Indosinian, Yanshanian, and Himalayan periods. The structural layout of the Shizhuang South Block follows an east−west zoning pattern, with elevations generally decreasing from southeast to northwest.[60] The Shanxi Formation, characterized by a deltaic sedimentary environment,[61,62] hosts the No. 3 Coal Seam, the focus of development efforts and the subject of this study (Figure 1c). The No. 3 Coal Seam typically exhibits a stable thick-ness distribution ranging from 4.0 to 8.0 m, with an average of 6.0 m, and is buried at depths ranging from 400 to 1020 m, averaging 750 m. In the No. 3 coal seam of the study area, the coal structure displays variations in composition. It can be segmented into undeformed coal, cataclastic coal, and granulated coal. Predominantly, undeformed coal and cataclastic coal constitute the major components. Regarding microcoal constituents, semibright coal and semidull coal prevail, with semibright coal being more abundant. Dull coal registers the lowest occurrence among these coal types, as illustrated in Figure 1d. The coal and associated rocks exhibit high maturity, with vitrinite reflectivity ranging from 2.5 to 3.0%, indicative of anthracite coal.[63,64] Permeability tests conducted on four parameter wells reveal generally low permeability within the No. 3 Coal Seam, typically ranging from 0.01 to 0.04 mD,[65] characterizing it as an ultralow permeability reservoir.

3. METHODS AND PRINCIPLES

3.1. Source of CBM Content Data and Geophysical Logging Data. Focusing on the Shizhuang South Block as the study area, this study aims to investigate CBM content under the air-dry base state. Coal rock samples were obtained during drilling operations (Figure 2a) and subsequently transferred to the laboratory (Figure 2b). Desorption processes were conducted on these samples (Figure 2c), with associated data such as determination time, interval time, standard quantitative tube readings, and gas volumes recorded (Figure 2d). Corrections were applied based on laboratory temperature and pressure. Figure 3a illustrates the cumulative desorption curve of the No. 3 coal seam sample in selected parameter wells after correction. Residual gas was quantified, enabling the calculation of lost gas volume for each sample. This involved extracting cumulative desorption time (T) and interval time (t) to derive $\sqrt{(T+t)}$, which was then linearly correlated with the cumulative desorption amount of the initial 2 h desorption phase. The intercept of the resulting linear equation provided the amount of lost gas (Figure 3b). In conjunction with Figure 3, it is evident that the gas content of the No. 3 coal seam in the Shizhuang South Block exhibits high complexity, with significant variation among coal samples of equal weight, underscoring the pronounced heterogeneity of coalbed parameters. Based on experimental results, the average proportions of desorption gas, loss gas, and residual gas in coal samples are 94.2, 3.7, and 2.1%, respectively. The gas content of cores from the No. 3 coal seam in the study area ranges from 4.55 to 26.13 m^3/t, with the majority of samples falling between 5 and 15 m^3/t.

Geophysical well logging involves deploying instruments into the wellbore to conduct real-time measurements of geological parameters using probes. These probes are connected via cables to a surface data acquisition system, transmitting data in real-time to provide geological insights at various depths within the vertical wellbore (see Figure 4a).

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

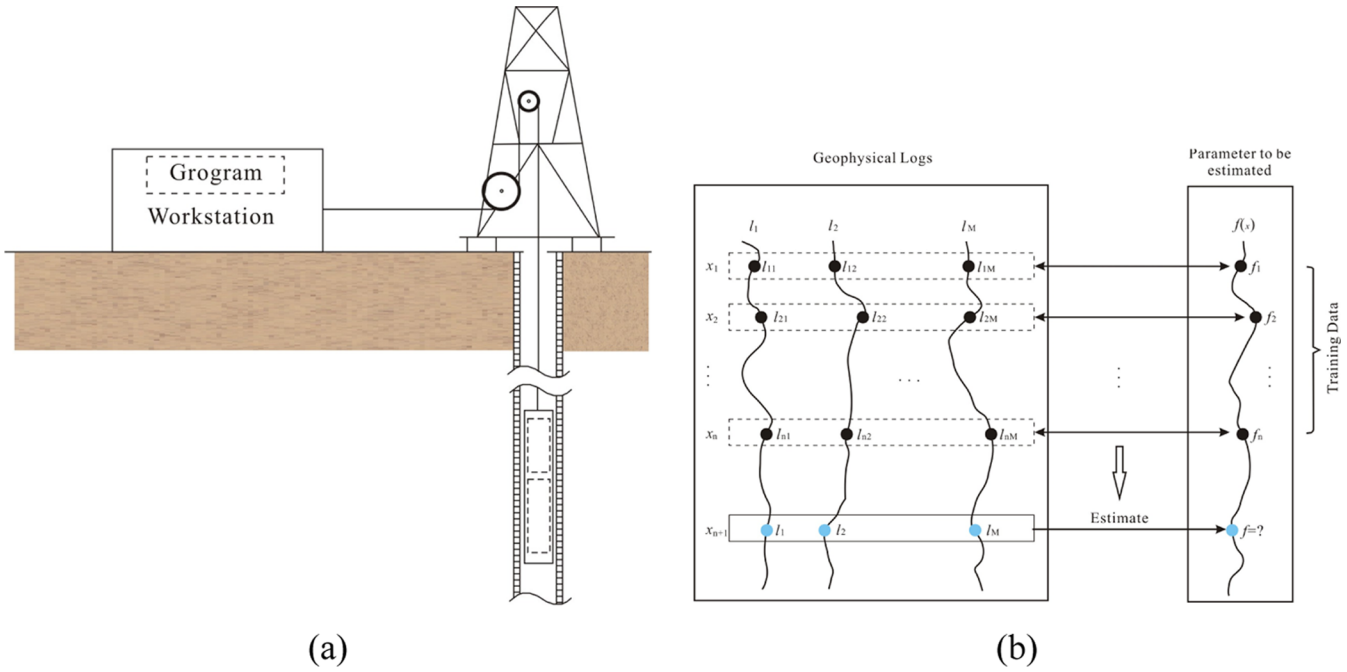


Figure 4. Collection of geophysical logging data and the schematic diagram of the evaluation parameter model constructed by logging curves. (a) Simplified acquisition diagram of geophysical logging curve; (b) The geophysical logging curve is combined with the core data to construct the model, which can be used for the whole depth section with logging data.

This data effectively characterizes physical properties of subsurface rocks such as density, resistivity, and acoustic velocity. Coalbed parameters are integrated with geophysical well logging responses to develop parameter evaluation models. Application of these models across the entire wellbore facilitates the generation of continuous parameter curves (see Figure 4b).

3.2. Principles of Random Forest Based on Simulated Annealing-Genetic Algorithm Improvement. *3.2.1. Principles of the Random Forest Algorithm.* The Random Forest (RF) algorithm, proposed by Breiman in 2001,[66] is a learning method that combines multiple individual learners to form an integrated model. The RF method is composed of decision tree models of the same type, which belongs to homogeneous integration. This method regards a single decision tree as the result of its model for the target and synthesizes it to get a new model. One set of decision trees can be written as $\{h(X, \theta_k), k = 1, 2, \cdots K\}$. Wherein, θ_k is a random variable, obeying independent and identical distribution. X and K respectively represent the number of independent variables and the number of decision trees. The results of RF prediction are obtained by averaging the results of each decision tree:

$$\overline{h}(X) = \frac{1}{K}\sum_{k=1}^{K}\{h(X, \theta_k)\} \tag{1}$$

To improve the predictive accuracy and stability of the model, the Bagging approach is introduced into the RF method. Bagging is a parallel ensemble learning method based on the Bootstrap technique, which generates multiple individual learners by sampling with replacement from the entire data set, thereby enhancing the model's generalization performance.[67,68]

The RF regression model used in this paper is CART tree. The flow of RF method is as follows: given that the original training sample size is N, and the number of features (log curves) participating in modeling is M;

(1) Bootstrap is used to draw samples from the original training samples with replacement to form subtraining set. For each sample in the original training set, the probability of each sample not being drawn is $(1-1/N)^N$. When N approaches infinity, there is

$$\lim_{n\to\infty}(1 - (1/N))^N \approx 1/e \approx 0.368 \tag{2}$$

In eq 2, e is a natural constant. And it can be obtained that when the samples have an enough size, the number of samples that did not participate in the modeling of the decision tree approaches to 36.8% of the original training samples. This part of data is called OOB (out of bag), which can generally be used to test the effect of the decision tree model.

(2) Using the subtraining set, a decision tree model is built. First, $m(m \le M)$ features are randomly chosen from the complete set of features (geophysical logging curves), and the ones resulting in the highest purity when splitting nodes are selected. Mean square error (MSE) is employed in regression to measure purity, where smaller MSE indicates higher purity. Thus, the feature with the lowest MSE is considered optimal for node splitting. The specific calculation formula of MSE is

$$M(A) = \frac{\sum_{i=1}^{N}(y_i - \overline{y})}{N} \tag{3}$$

Equation 3 calculates the feature with the smallest MSE among m features, where N is the number of samples in the target split node, $\overline{y}$ is the average value of sample features, and A is one of the m features. This process avoids feature vector magnitude differences, eliminating the need for data normalization and preserving data

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

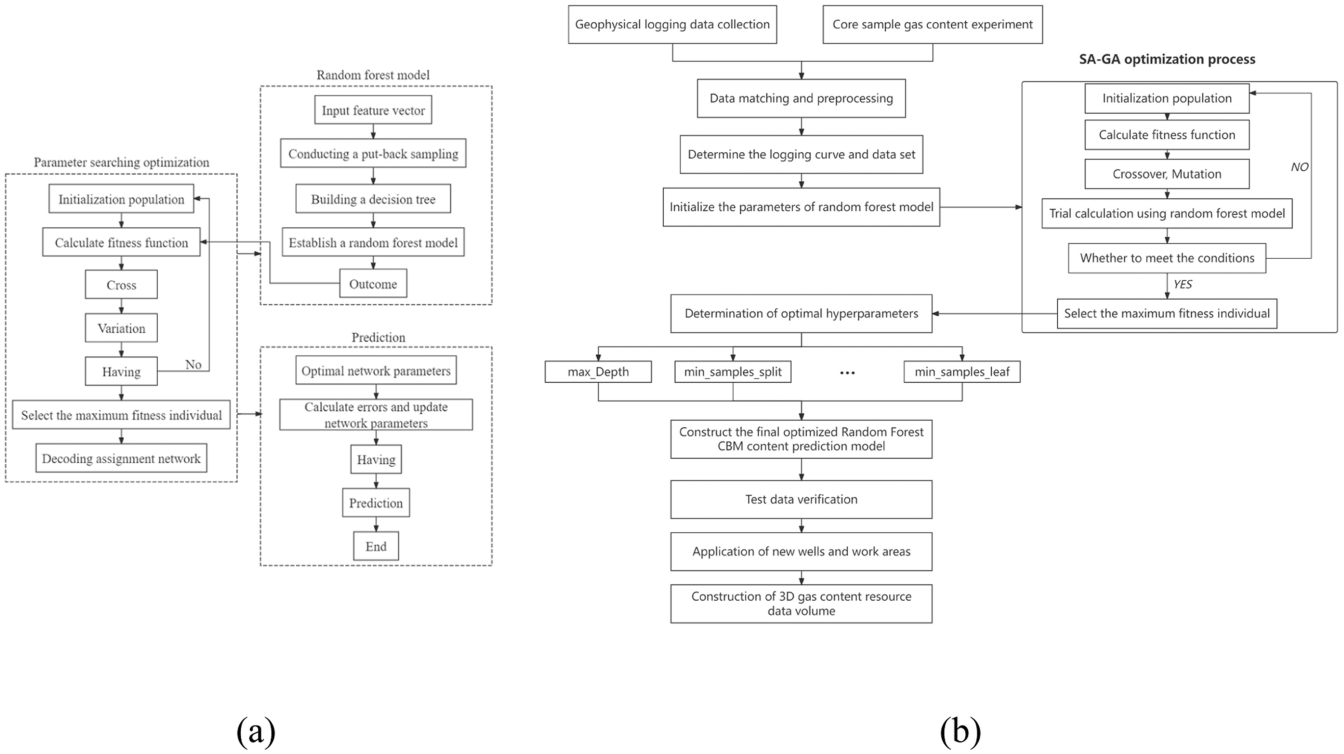


Figure 5. Process diagram. (a) The schematic diagram of SA-GA optimized RF method; (b) Process diagram of coalbed methane content evaluation model construction.

features. Splitting in a binary tree form is based on this principle, with the splitting ending determined by the tree's depth and the minimum number of samples in leaf nodes.

(3) Through repeating steps (1) and (2) for K times, K subtraining sets and their corresponding models can be obtained. These independent models can be integrated to form the Random Forest model.

(4) When the RF classification model is used to predict the test set, each decision tree model will give a prediction result. For the regression type problems, the RF prediction results adopt the average method. That is, the average of the prediction results of K base classifiers is the RF prediction result, as shown in eq 1.

(5) For the generalization error of RF, the random vector (x, y) which follows the independent and identical distribution is taken as an example. Combined with eq 2, the corresponding mean square generalization error of $h(X)$ is

$$E_{X,Y}(Y - h(X))^2 \tag{4}$$

In RF regression, if the number of decision trees tends to be infinite, there is

$$E_{X,Y}(Y - \overline{h}(X, \theta_k))^2 \rightarrow E_{X,Y}(Y - E_\theta h(X, \theta))^2 = \mathrm{PE}^*_{\mathrm{tree}} \tag{5}$$

In eq 5, θ_k is the random variable of the kth decision tree. E_θ corresponds to the mathematical expectation, and $\mathrm{PE}^*_{\mathrm{tree}}$ is the generalization error of RF regression. If the regression decision tree is unbiased for the random variable θ and there is $EY = E_X h(X, \theta)$, then:

$$\mathrm{PE}^*_{\mathrm{forest}} = E_\theta E_{X,Y}(Y - \overline{h}(X, \theta_k))^2 \le \bar{\rho}\,\mathrm{PE}^*_{\mathrm{tree}} \tag{6}$$

In eq 6, $\bar{\rho}$ is the correlation coefficient of the remaining $Y - h(X, \theta)$ and $Y - h(X, \theta')$, θ and θ' are independent of each other. To sum up, the RF method will eventually converge with the increasing number of decision trees, and the generalization error will tend to be a certain value.

RF exhibits randomness in two key aspects: the randomness in training data selection for base classifiers and the randomness in selecting features for node splitting. Utilizing more base classifiers effectively leverages the original training data, while the Bootstrap concept mitigates uneven sample data distribution, rendering RF an efficient and practical nonlinear algorithm.[67]

3.2.2. Principles of the Simulated Annealing-Genetic Algorithm. To address the challenge of hyperparameter selection in the RF algorithm, the simulated annealing (SA) -genetic algorithm (GA) is employed for parameter optimization. SA-GA combines the advantages of SA and GA, allowing effective escape from local optima and rapid convergence,[69,70] while avoiding overfitting and oscillation phenomena, thus achieving global optimization of parameters.

(1) Parameter Initialization: Select input and output parameters, and randomly generate n individuals as the initial population, representing the number of decision trees and feature partitions. Set the maximum number of iterations to $M_{\max}$, and determine the initial and final temperatures to initiate the annealing process.

(2) Fitness Evaluation: Utilize the RF error evaluation function (eq 7) as the fitness function to assess the quality of individuals in the population.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

$$f = \frac{1}{n}\sum_{i=1}^{n}(y_k - y)^2 \tag{7}$$

In eq 7, f is the fitness function, y_k and y correspond to the predicted value and the true value in the kth population, respectively, and n is the number of randomly generated individuals.

(3) Selection, Crossover, Mutation: Retain superior individuals as parents (eq 8) through selection operations like the roulette wheel method, then perform crossover and mutation operations to generate new offspring.

$$p_i = \frac{f_i}{\sum_i^n f_i} \tag{8}$$

In eq 8, p_i is the probability that an individual becomes a parent, and f_i is the fitness value of the ith fitness function, dimensionless.

(4) Local Update: Apply simulated annealing to the new individuals obtained from crossover and mutation along with the parents, using the Metropolis criterion for updating (eq 9). If the probability of the new individual is less than the randomly generated value, keep the offspring; otherwise, retain the parent for the next optimization round.

$$P(\Delta E, T) = e^{\Delta E/K_B T} \tag{9}$$

In eq 9, P is the probability of transition to offspring; ΔE is the difference of fitness between parent and offspring, dimensionless; K_B is Boltzmann constant, dimensionless; T is simulated annealing temperature, °C.

(5) Iteration Termination: Repeat steps 1 to 4 until reaching the maximum iteration count, completing the parameter optimization process.

The above algorithm flow is shown in Figure 5a. Figure 5b is the process diagram of this study.

3.3. CBM Content Evaluation Model Construction Method Process. The process of constructing a CBM content evaluation model based on geophysical logging data involves three key steps, as depicted in Figure 5b.

(1) Data Collection and Preprocessing: Initially, data are collected and preprocessed. This includes aligning geophysical logging data with gas content data, performing depth correction, and cleansing the data. Factors induced by drilling, such as drill pipe deformation and stretching, often cause discrepancies between cumulative drill pipe length and actual drilling depth, necessitating depth matching.[71] Additionally, the fragile mechanical strength of coal seams can lead to borehole wall collapse during drilling, resulting in anomalous responses in geophysical logging data. Therefore, samples showing significant expansion are excluded, along with segments displaying distorted logging curve responses in depth sections, particularly near the initiation and termination of the coal seam. Sections of noncoal rock typically exhibit abnormal geophysical logging responses and are thus removed.[64] These preprocessing steps are crucial to ensure the scientific rigor and accuracy of subsequent modeling.

(2) Hyperparameter Optimization: Utilizing the method outlined in Section 3.2.2, hyperparameter optimization for the RF approach is conducted. The final optimized hyperparameters are then employed to construct the CBM content prediction model.

(3) Model Testing and Application: Following model construction, it undergoes testing and application. This involves its application to both test data and new wells not utilized in the model's initial construction phase. The aim is to develop a comprehensive three-dimensional gas content resource model.

4. THE ACTUAL MODELING PROCESS AND RESULTS

4.1. Modeling Process. *4.1.1. Geophysical Logging Curve Selection.* The coal seam reservoir itself has the characteristics of low natural gamma response and easy expansion. The porosity series shows the characteristics of low density and high acoustic time difference, in which the compensated neutron logging response is high. Compared to the mudstone section of the coal seam roof, the coal seam shows high resistivity, with some sections reaching up to 2000 Ω·m. The geophysical logging data in the study area consists mainly of eight logging curves from seven logging series: Caliper (CAL), Spontaneous potential (SP), Nature gamma (GR), Compensation density (DEN), Compensated neutron (CNL), Acoustic time difference (AC), Deep lateral resistivity (LLD), and Shallow lateral resistivity (LLS). Based on the physical properties reflected by the actual data, the porosity logs, resistivity logs, and natural gamma logs are commonly used for predicting the gas content in coal seams. Considering that the target of the study is the No. 3 coal seam, which is relatively shallow, depth is also incorporated into the model construction. After preprocessing the geophysical logging curves, a correlation analysis was conducted with the gas content in core samples, employing formulas 10−12 to calculate the Pearson correlation index (Pearson), Kendall rank correlation coefficient (Kendall), and Spearman rank correlation coefficient (Spearman). In total, more than 160 sets of core sample gas content data were collected. After data verification and cleaning, 148 sets were retained. The results of the correlation analysis are presented in Table 1.

$$\text{Pearson} = \frac{M\sum_{i=1}^{M} x_i y_i - \sum_{i=1}^{M} x_i \sum_{i=1}^{M} y_i}{\sqrt{M\sum_{i=1}^{M} x_i^2 - \left(\sum_{i=1}^{M} x_i^2\right)}\sqrt{M\sum_{i=1}^{M} y_i^2 - \left(\sum_{i=1}^{M} y_i^2\right)}} \tag{10}$$

$$\text{Kendall} = \frac{4A}{M(M-1)} - 1 \tag{11}$$

$$\text{Spearman} = \frac{M\sum_{i=1}^{M}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{M}(x_i - \bar{x})^2 \sum_{i=1}^{M}(y_i - \bar{y})^2}} \tag{12}$$

In eqs 10−12, x_i and y_i represent the logging response values and core gas content values, respectively, with subscript i

Table 1. Correlation Analysis between Logging Responses and Core Gas Content

CBM content (g/cm³)	AC (μs/m)	CNL Pu	DEN (g/cm³)	GR API	RD (Ω·m)	DEPTH (m)
Pearson	0.55	−0.43	−0.54	−0.59	0.28	−0.63
Kendall	0.35	−0.31	−0.38	−0.44	0.22	−0.39
Spearman	0.49	−0.48	−0.54	−0.64	0.36	−0.58

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

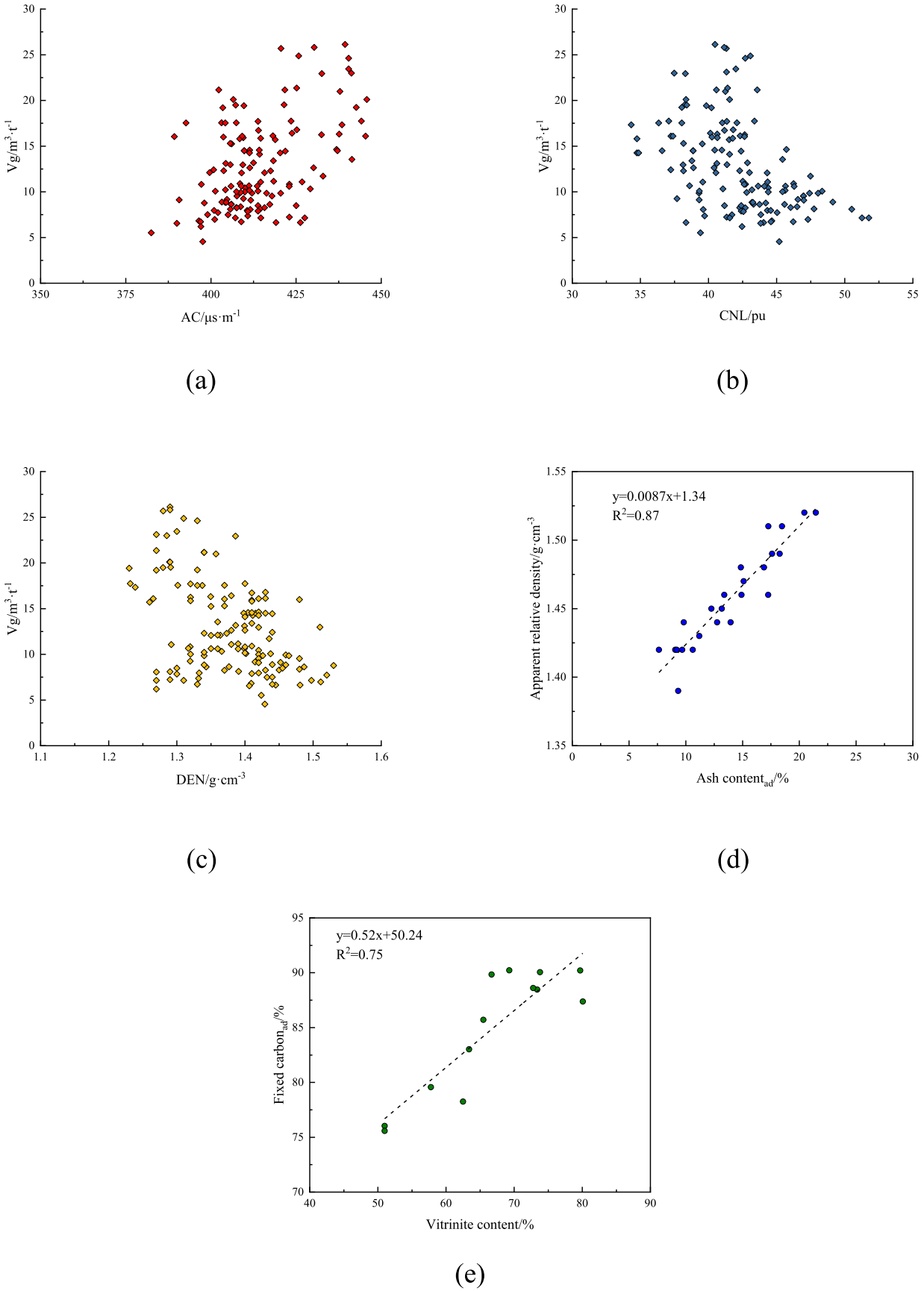


Figure 6. Cross-plots of geophysical logging response versus experimental parameters. (a) Cross-plot of acoustic time difference logging response and core gas content; (b) Cross-plot of compensated neutron logging response and core gas content; (c) Cross-plot of compensated density logging response and core gas content; (d) Cross-plot of core ash content and apparent density; (e) Cross-plot of vitrinite content and fixed carbon content of core.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

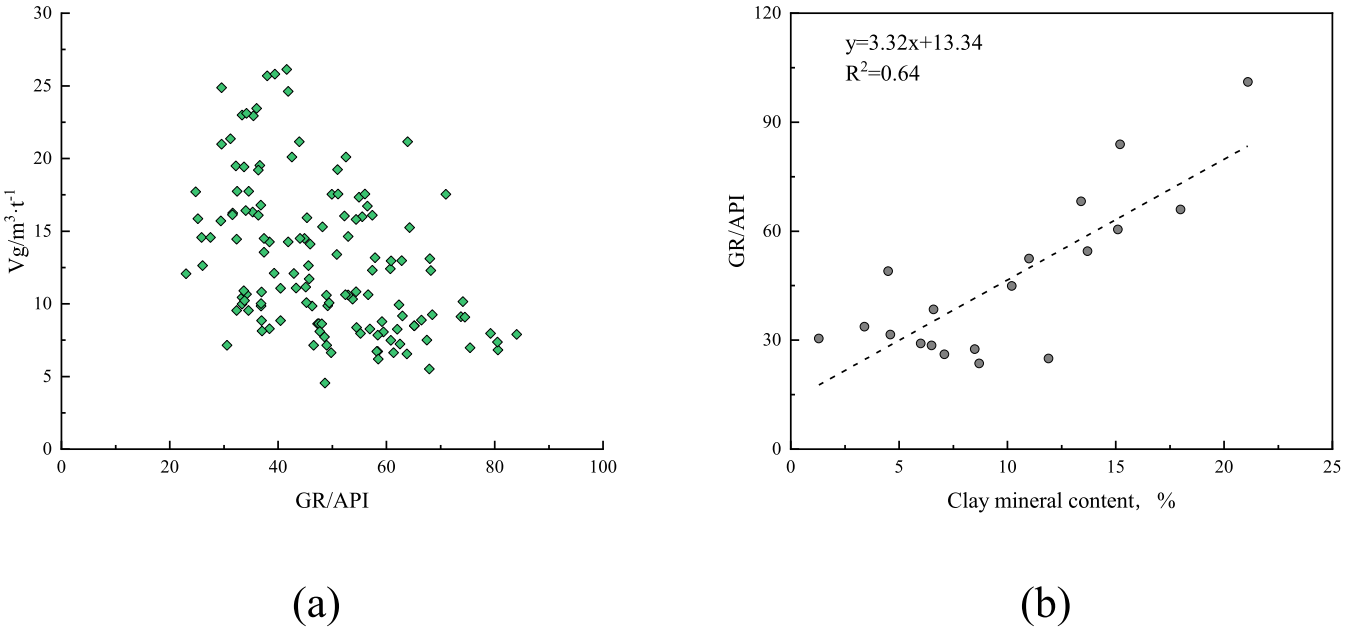


Figure 7. Cross-plot of natural gamma curve and core experimental parameters. (a) Cross-plot of natural gamma curve and core gas content; (b) Cross-plot of core clay mineral content and natural gamma curve.

denoting the sample index, and M representing the number of samples. A is the count of pairs where the ranks of the two attributes agree in order. $\bar{x}$ is the average value of the logging response, and $\bar{y}$ is the average value of the coalbed methane content.

Based on the data presented in Table 1, corresponding cross-plots were generated (Figure 6). Among the porosity logging metrics, the AC curve shows a positive correlation with gas content (Figure 6a), while the CNL and DEN curves demonstrate negative correlations (Figure 6b,c). A significant correlation was also observed between laboratory apparent relative density and air-dried basis ash content (Figure 6d). Increasing coal rock densification correlates with reduced reservoir porosity and coalbed methane content; conversely, less dense coal tends to be softer and typically has higher methane content.[72] Moreover, higher levels of inorganic minerals such as ash can fill pores and fractures, thereby hindering the storage and adsorption of coalbed methane. This phenomenon explains the observed relationship between methane content and compensated density logging responses.[73,74] Additionally, regression analysis was performed between air-dried basis fixed carbon content and vitrinite content in microscopic components (Figure 6e). Both factors facilitate the adsorption and retention of coalbed methane. According to the literature, higher vitrinite content in higher-rank coals enhances fragmentation, thereby increasing the adsorptive surface area for methane and consequently boosting gas content.[75] Furthermore, significant fractures within the coal rock can influence acoustic travel-time logging responses.[76,77] Compensated neutron logging, affected by factors including coal rock structure and gas content, shows a negative correlation with methane content after compensation.[78] It is important to note that coalbed methane reservoirs differ from conventional gas reservoirs in that methane primarily exists in adsorbed form, which complicates the direct correlation between gas content and logging responses, thereby challenging the assessment of gas content.

The GR curve shows a negative correlation with core gas content (Figure 7a). Analysis of the intersection between clay mineral content and the GR curve reveals a positive correlation (Figure 7b). Clay minerals influence the gas content in coal seams by affecting the adsorptive properties of the coal; an increase in clay minerals in coal reduces the methane content, while a higher presence of these minerals enhances the natural radioactivity, which explains the trend observed between coalbed methane content and natural gamma log responses.[79] Resistivity logging is relatively complex; in comparison to other log responses, the resistivity log response is influenced not only by the gas content but also by factors such as the degree of coal metamorphism, coal rock structure, mineral content, and distribution, surrounding rock pressure, temperature, and more.[46] The deep radial exploration of the LLD curve shows a low correlation with gas content (Figure 8), while the LLS curve, affected by invasion, no longer represents the original

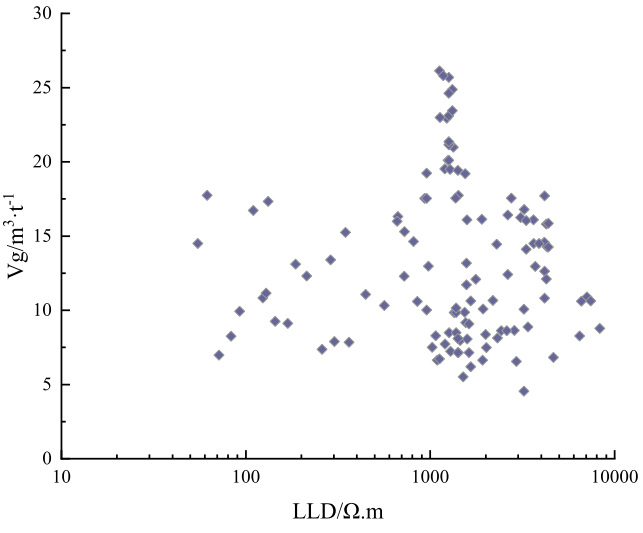


Figure 8. Cross-plot of LLD curve and core gas content.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

formation's resistivity, and is not used for predicting gas content. Effective depth shows the highest correlation coefficient with gas content, indicating that depth is one of the key factors affecting gas content.[29] The depth of the coal seam determines whether the gas produced by coalification can be preserved. Theoretically, at shallower depths, the temperature effect is relatively insignificant. As coal seam depth increases, both the degree of coalification and the quantity of generated hydrocarbons should increase, following Langmuir's adsorption isotherm; however, as depth further increases, the temperature effect becomes more pronounced, and beyond a certain critical depth, this relationship no longer holds. The drilling sites in the study area are near an elevation of 1000 m, with the No. 3 coal seam being relatively shallow and above sea level. Therefore, this paper incorporates conversions of the on-site logging instrument core height, elevation, and logging depth, with converted depths approaching sea level (0) indicating deeper burial. The intersection graph of burial depth and gas content trend, as shown in Figure 9, suggests that the closer to sea level, the deeper the burial depth, the relatively higher the gas content.

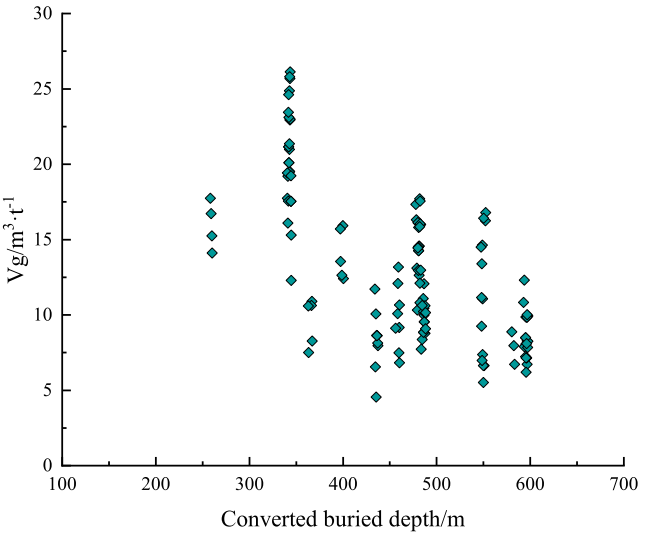


Figure 9. Cross-plot of buried depth and core gas content after conversion.

Based on the correlation analysis tables and cross-plots, combined with experimental data, it is evident that the relationship between geophysical logging data responses and variations in coalbed methane content is nonlinear. When constructing models for coalbed methane content, six logging curves are selected based on the physical properties they represent: AC, DEN, CNL, GR, DEP, and LLD.

4.1.2. Types of Hyperparameters to be Optimized. Before constructing a CBM content model using the RF method, it is crucial to optimize the hyperparameters of the RF approach, i.e., to explore the impact of these hyperparameters on model construction. The hyperparameters studied in this paper are the number of decision trees (n_estimators), the number of features considered for splitting at each leaf node (max_features), the maximum depth of the trees (max_depth), the minimum number of samples required to split an internal node (min_samples_split), and the minimum number of samples required at a leaf node (min_samples_leaf). Based on previous studies' data ratios, the collected gas content data are first divided into training and test data sets in a 7:3 ratio. A controlled variable approach is employed for exploration, systematically varying each parameter while keeping others at their default values. To mitigate result variability, model construction incorporates cross-validation. Cross-validation is a widely accepted and effective technique for evaluating models, commonly employing k-fold cross-validation.[80] This method entails partitioning the training data set into k subsets, with each subset used once as a test set while the remaining $k-1$ subsets serve as training data. This process generates k models, each assessed against its respective test data subset. Evaluation employs an accuracy metric, and the final k-fold cross-validation result is the mean accuracy across these models. This paper employs 3-fold cross-validation, and to mitigate the uncertainty from the model's bagging approach during exploration, a fixed random seed is used for each modeling instance, ensuring consistency in the subsets used. The exploration results are shown in Figure 10. From Figure 10a, it is evident that if the number of decision trees is too low, both cross-validation results and errors in the training and test data sets are high, indicating an underfitted model. When the number of trees exceeds 40, errors are effectively reduced with minor differences between them. This indicates the significant impact of the number of trees, which requires detailed exploration. Figure 10b–d show the maximum tree depth, the minimum number of samples required to split an internal node, and the minimum number of samples required at a leaf node, respectively. These three parameters are interdependent and determine the model's complexity. The default values for CART trees are 2 for min_samples_split and 1 for min_samples_leaf, which are the limits for binary splitting. Thus, after a certain increase in maximum depth, the model tends to stabilize, and further increases have no impact on the model. For the small sample size of gas content data studied in this paper, the model stabilizes at a depth of 13, as shown in Figure 10b, where no fluctuations in cross-validation, training, or test set errors occur, indicating no overfitting. When exploring the minimum number of samples required at internal nodes and leaf nodes, with tree depth set to None, model depth (complexity) is determined solely by one parameter. Increasing the minimum number of samples, which effectively reduces model depth, tends to lead the model toward an underfitted state, worsening performance across cross-validation, training, and test data sets. As seen in Figure 10e, when the number of features to consider per split is set to 1, both cross-validation and test set errors are poor, indicating suboptimal model performance. Furthermore, the strength of the random forest lies in minimizing the correlation between decision trees. When the number of features per split is large, the correlation between trees increases, which can impact model accuracy. The exploration of individual features with other parameters at default values shows that when the number of trees is set to 100, as shown in Figure 10a, the model performs well. This demonstrates that the impact of the number of features per split can be masked when analyzed in isolation. Therefore, setting the number of trees to a lower value, as shown in Figure 10f, where the number of trees is 4, reveals significant differences in the impact of the number of features per split on model accuracy. To establish a reliable random forest classification model, the SA-GA algorithm is utilized for hyperparameter optimization.

4.2. Results. After identifying the logging curves involved in model construction and the hyperparameters requiring

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

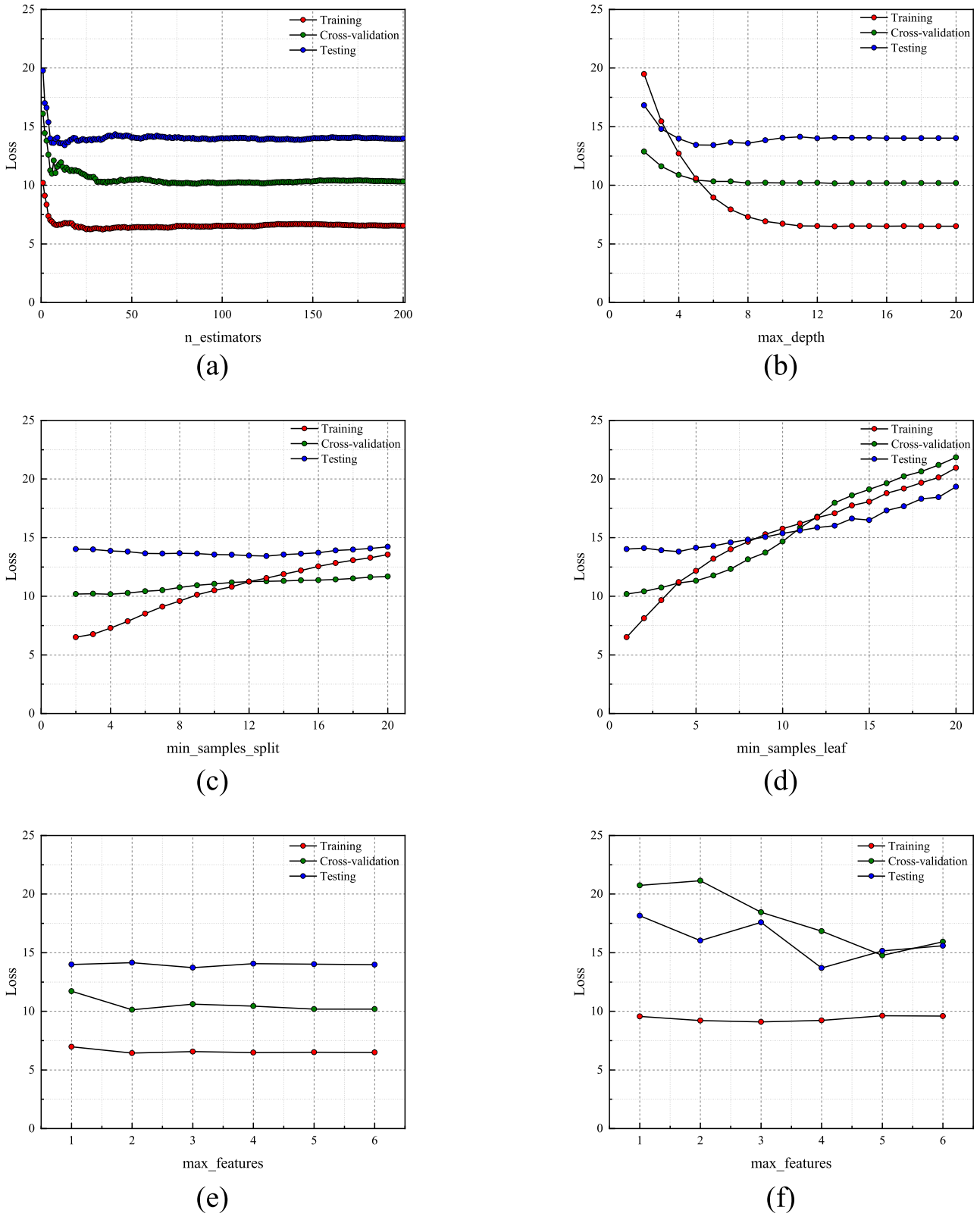


Figure 10. Study on hyperparameters of RF method. (a) The influence of the number of decision trees on model performance; (b) The influence of the maximum depth of the decision tree on model performance; (c) The influence of the minimum number of samples required to divide internal nodes on model performance; (d) The influence of the number of leaf nodes on the model performance; (e) When the number of decision trees is 100, the influence of the number of split features on the performance of the model; (f) When the number of decision trees is 4, the influence of the number of split features on the performance of the model.

optimization, the SA-GA method was employed for hyperparameter tuning. The iteration was set to 100 times, with a population size of 10 and a temperature reduction parameter of 0.98. The error function variation with iteration number is

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

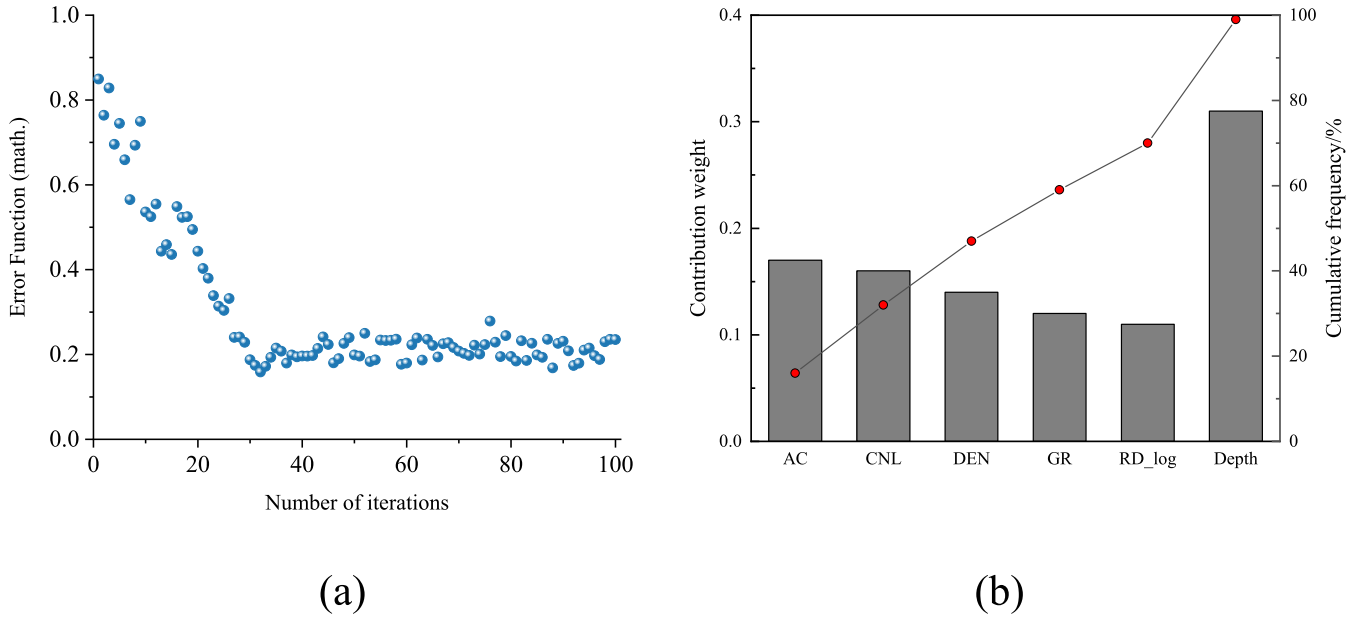


Figure 11. Hyperparameter optimization process and final model contribution weights. (a) The optimization process of SA-GA algorithm for hyperparameters in RF algorithm; (b) Contribution weights for each logging curve in the model.

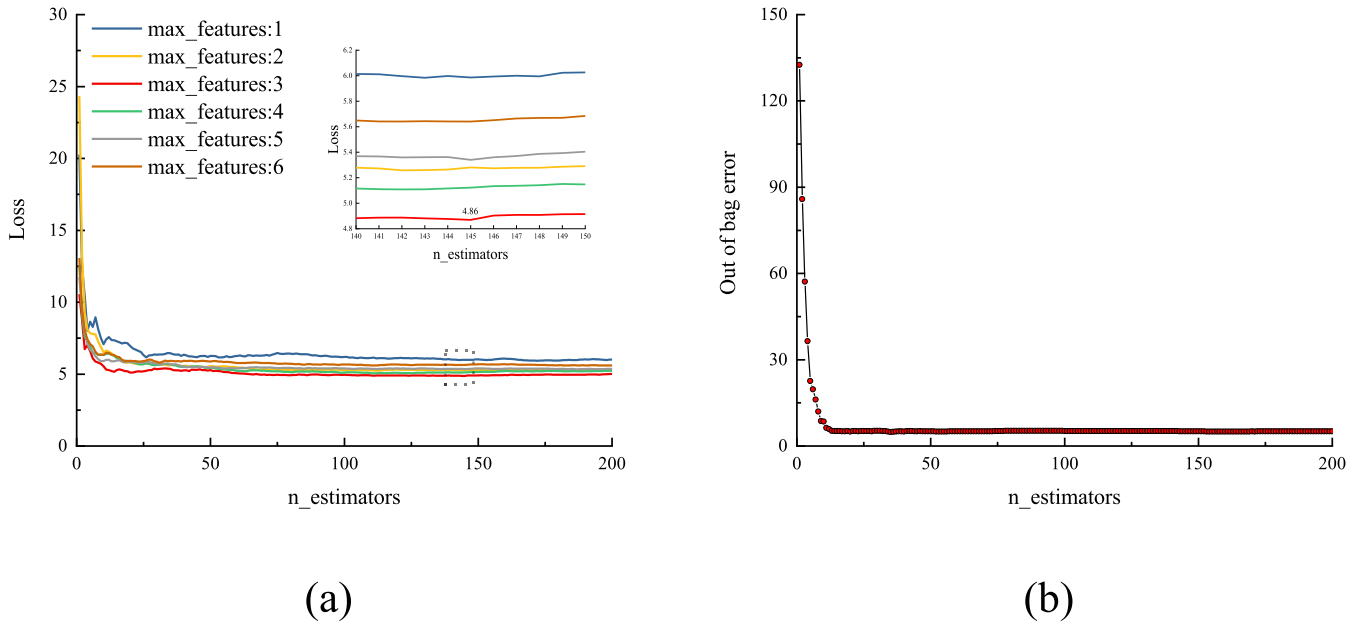


Figure 12. Effectiveness of the model. (a) The relationship between the number of different nodes splitting features and the number of decision trees.; (b) Out-of-bag error curve.

depicted in Figure 11a: after exceeding 30 iterations, the error function stabilized. At this point, the optimized hyperparameters for the random forest included 145 trees, 3 features per split, a tree depth of 14, a minimum of 2 samples required to split an internal node, and a minimum of 1 sample required at a leaf node. Additionally, eq 3 was used to calculate the contribution of each logging sequence to the model, as shown in Figure 11b. Depth made the largest contribution to the construction of the gas content model, followed by AC, CNL, DEN, and GR, with the contributions decreasing sequentially, and RD contributing the least, less than 0.1, consistent with theoretical analyses.

Figure 12a illustrates the relationship between the number of features per split and the number of trees under optimized hyperparameters, achieving a root-mean-square error (RMSE) of 4.86. Figure 12b displays the out-of-bag error curve for the finalized parameters, demonstrating stabilization as the number of decision trees reaches 145, affirming the model's efficacy under these conditions. Figure 13a presents the results of back-analysis on the training data, while Figure 13b illustrates the model's application to the independent test data set, which was not used in model construction. The relative error for the training set back-analysis is 5.99%, with a goodness of fit of 0.98 compared to laboratory results. For the test set, the relative prediction error is 13.13%, with a goodness of fit of 0.84 compared to laboratory results, indicating robust model training without overfitting.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

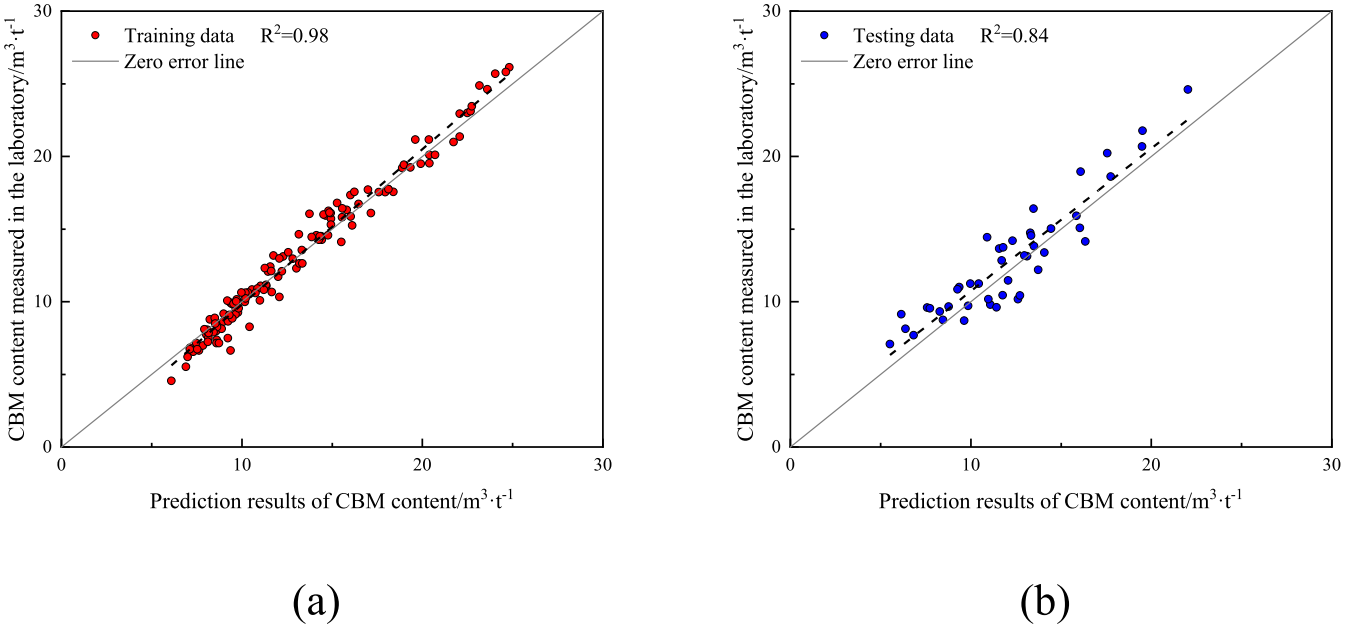


Figure 13. Effect of model application. (a) The effect of training data regression; (b) The application effect of the test data set that is not involved in the model construction.

After constructing the gas content model, two new wells within the same study area were evaluated to assess the model's generalizability; their data were not included in the model's training or testing phases. The CBM content curves predicted using the SA-GA-RF method and laboratory gas content results are displayed in Figure 14. Figure 14a,b present the logging curves and calculated gas content for these two wells, respectively. The first track in the figures provides stratigraphic information, specifically detailing the No. 3 coal seam. The second track displays the lithology profile, indicating that both wells exhibit mudstone at the top and bottom of the No. 3 coal seam. The third track shows the lithology logging data, including CAL, GR, and SP logging curves. The fourth track displays the resistivity logging data with three resistivity curves, while the fifth track shows the porosity logging data with three porosity curves. Finally, the sixth track presents the results, including the gas content curves calculated using the method described in this paper, alongside the core laboratory gas content. Both wells with the No. 3 coal seam had 13 valid experimental samples each, excluding samples from areas of severe diameter expansion. The average relative errors for the eight samples in Well A1 and five samples in Well A2 were 11.09 and 13.61%, respectively. These errors are consistent with the test set error level, demonstrating the generalizability of the established model.

Significant errors (with relative errors exceeding 35%) were observed in the new well, particularly notable with the sample at 796.33 m in Well A1, located where the borehole experienced diameter expansion. This section's logging curve responses can be significantly disturbed, as logging instruments measure from the bottom upward, potentially magnifying judgment errors. As discussed in Section 3.3, despite excluding anomalous data, inherent noise in the data itself can lead to minor discrepancies between core laboratory measurements and log interpretations. Therefore, it is important to note that errors stemming from these factors should not solely be attributed to the predictive capabilities of the SA-GA-RF method.

The completed CBM content model was applied to wells within the Shizhuang South Block, and a three-dimensional methane content model was constructed based on the evaluation results from multiple wells. Considering the area of the study zone and the spatial distribution of coalbed methane wells involved in this modeling, the No. 3 coal seam underwent grid-based three-dimensional geological modeling, with the horizontal grid set at 100 m × 100 m; the vertical grid was controlled at 0.5 m based on the thickness and stability of the No. 3 coal seam in the study area. Using well position coordinates and elevations provided by satellite remote sensing data, combined with actual logging data to determine the position of the No. 3 coal seam, Employing the sequential Gaussian algorithm, a three-dimensional methane content attribute model (Figure 15a) was built using the methane content curves from 128 wells and the Gaussian model, with variogram analysis used to assess the continuity of methane content in space and its anisotropy in various directions. Figure 15b is a grid diagram with the thickness of the three-dimensional model, which can characterize the difference in gas content in different sections of the three-dimensional model. The three-dimensional coalbed methane content model reveals that the methane content distribution within the No. 3 coal seam in the study area ranges from 6.4 to 25.4 m^3/t, consistent with core results. High methane content areas are located in the western and northern parts of the block, with two high methane content layers developing longitudinally within 1.0 m from the top and 2.0 m from the bottom of the seam.

After constructing the three-dimensional CBM content model, the two-dimensional isopleth map of CBM content represents the average values (Figure 16a). From Figure 16a, the averaging of methane content values within individual wells results in minimal overall variability, making it difficult to accurately identify areas of high methane content. In contrast, the three-dimensional methane content model allows for the depiction of variations in methane content between different wells within the No. 3 coal seam according to specific needs

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

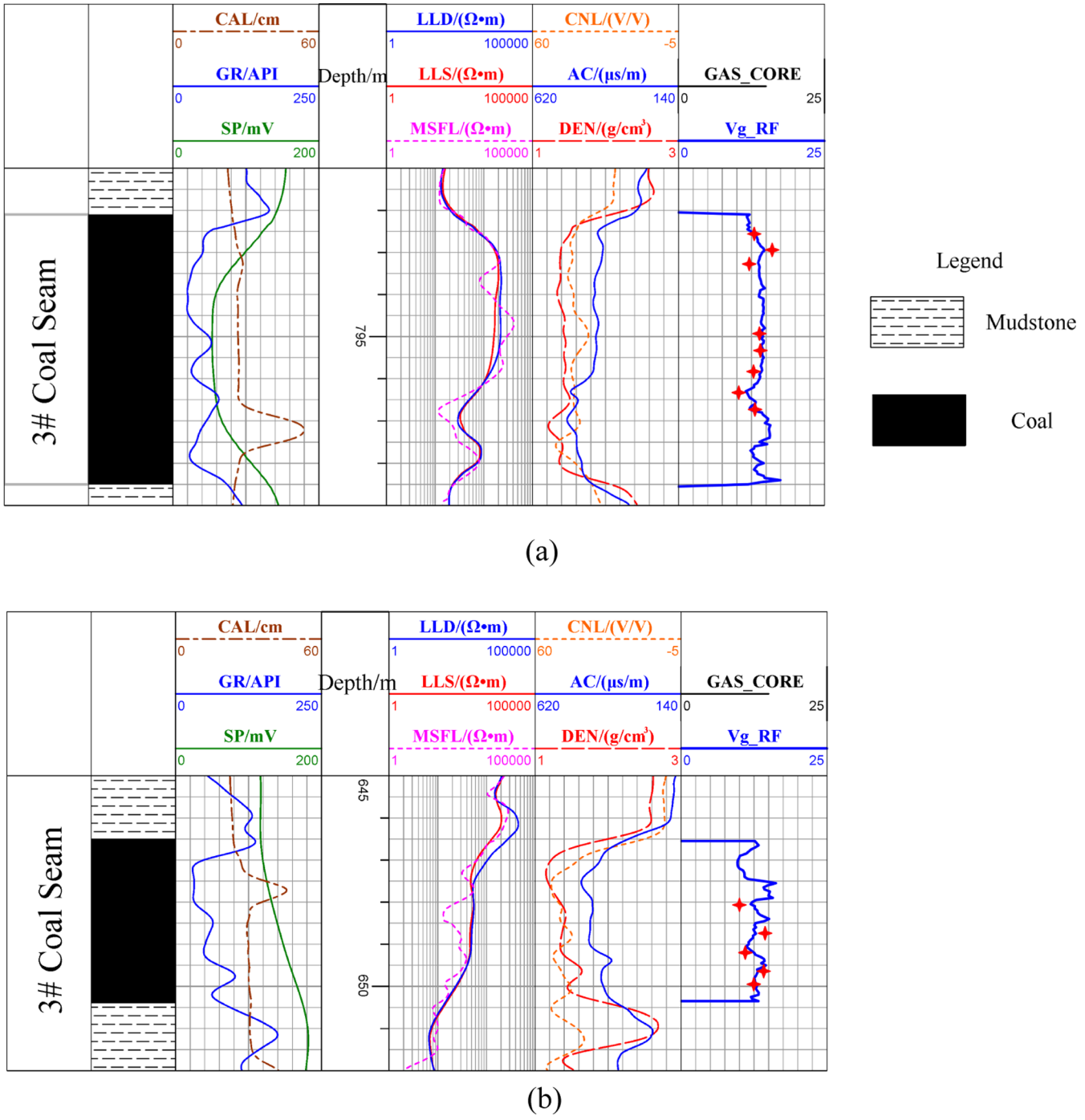


Figure 14. Application effect of the model constructed by this method in the actual new wells. (a) The display of A1 well in the study area; (b) The display of A2 well in the study area.

(Figure 16b) and enables the construction of cross-sectional views of the CBM content across the entire study area (Figure 17).

To further characterize the advantages of the three-dimensional model, 68 production wells in the study area with a stable output over 6 years were selected. The average effective daily gas production of each well was calculated, and contour maps were plotted (Figure 18a). Based on the calculated CBM content curves, the average gas content values of the perforated sections in No. 3 coal seam were extracted and mapped (Figure 18b). During the mapping of averages, the selection range for color mapping was notably restricted. Inappropriate choices for the lower and upper limits could result in a uniform color fill of the isopach, due to the significant loss of information in the average values of CBM content. Utilizing the three-dimensional model, the peak gas content values within the perforated sections and the values within 0.5 m above and below the peak were averaged. This method ensures that the true levels of coal seam gas content are extracted, while additionally including an extra 1 m to guard against outliers, with the results plotted in a contour map (Figure 18c). A comparison of panels a, b, and c in Figure 18 shows a better correlation between high gas production wells and the areas of high gas content depicted in Figure 18c.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

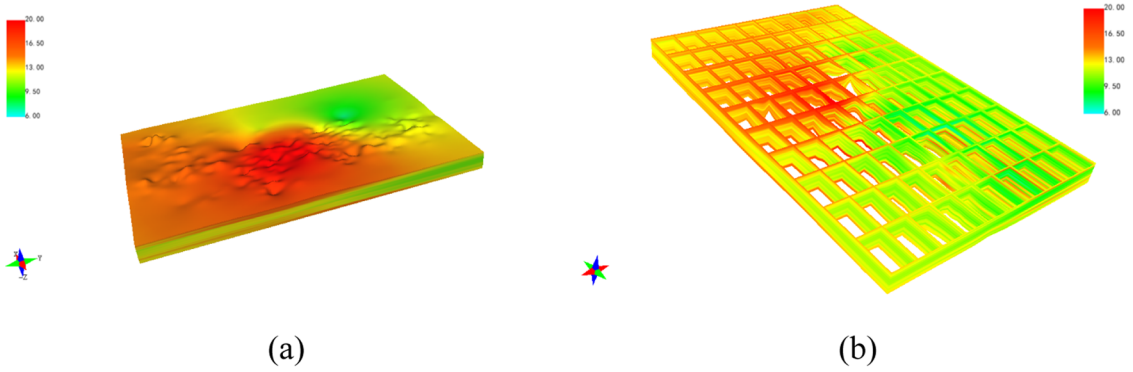


Figure 15. Three-dimensional CBM content model. (a) Three-dimensional model of gas content attribute; (b) Three-dimensional model with thickness grid diagram.

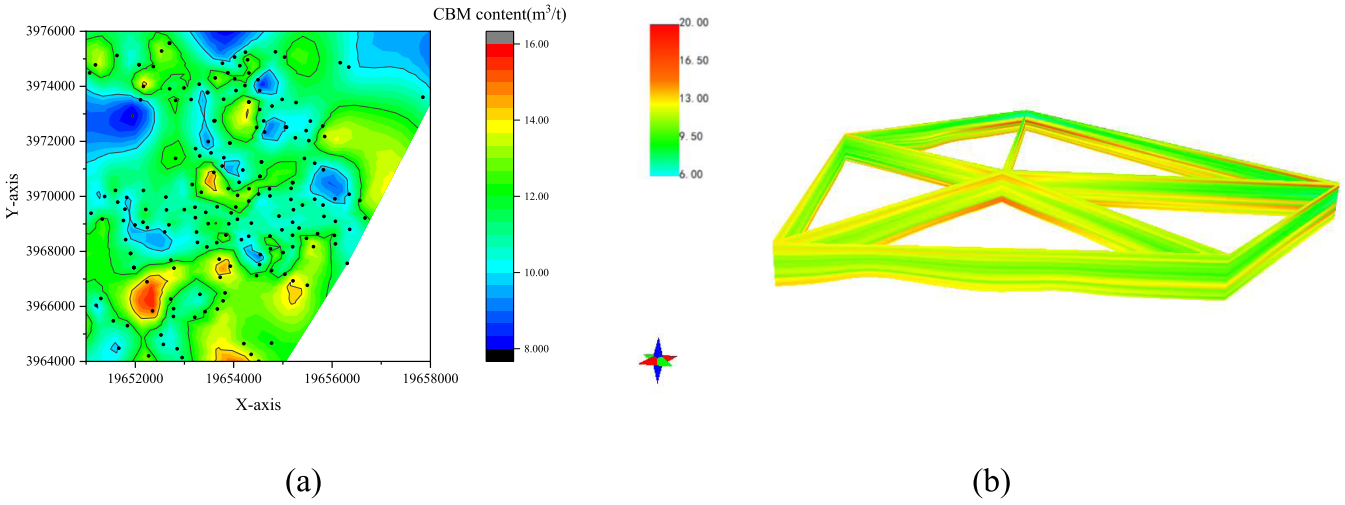


Figure 16. Application of CBM content model. (a) Two-dimensional gas content contour map; (b) Using three-dimensional model to characterize the difference of gas content in No. 3 coal seam between different wells.

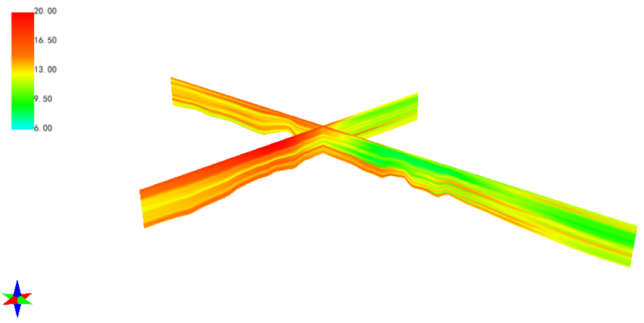


Figure 17. Gas content distribution section of No. 3 coal seam in the study area (Cross Section).

Fitting analysis of the data extracted from the three-dimensional model with gas production (Figure 18d) indicated a positive correlation; the greater the peak segment of gas content, the higher the coalbed methane well output. It should be noted that the goodness of fit was 0.67, as the output of coalbed methane wells is not only related to the level of gas content but also influenced by a multitude of factors including coal matrix structure, reservoir pressure, temperature, and human interventions.[81] In Figure 18d, wells with daily production not reaching 500 m^3 were defined as low-production and inefficient. While some of these wells had low gas content levels, others with peak values exceeding 12 m^3/t could be re-evaluated for potential production enhancement measures, showcasing the informational advantages brought by the three-dimensional model. For low-production and inefficient wells, quality control can be conducted by combining records of fracturing conditions and on-site measures to minimize unfavorable production events such as well shutdowns.

The construction of the three-dimensional CBM content model leverages the accuracy of the methane content evaluation model and the richness of logging data, thereby enhancing the practicality of CBM content evaluation.

5. DISCUSSION

5.1. Advantages of the Methodology. *5.1.1. Advantage of Accuracy.* To verify the accuracy of the method described in this paper compared to other types of machine learning approaches, previous researchers have used BPNN, Least Squares Support Vector Machines (LSSVM), and ELM methods to construct CBM content models.[53,64] Models were built using the same training sets, and predictions were made on validation sets to compare different outcomes. Figure 19 illustrates the cross-plot of the gas content results calculated by multiple methods with the gas content of the cores obtained from laboratory measurements. Table 2 presents the accuracy of the method used in this paper against three other machine learning algorithms and multivariate regression. Analysis of

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

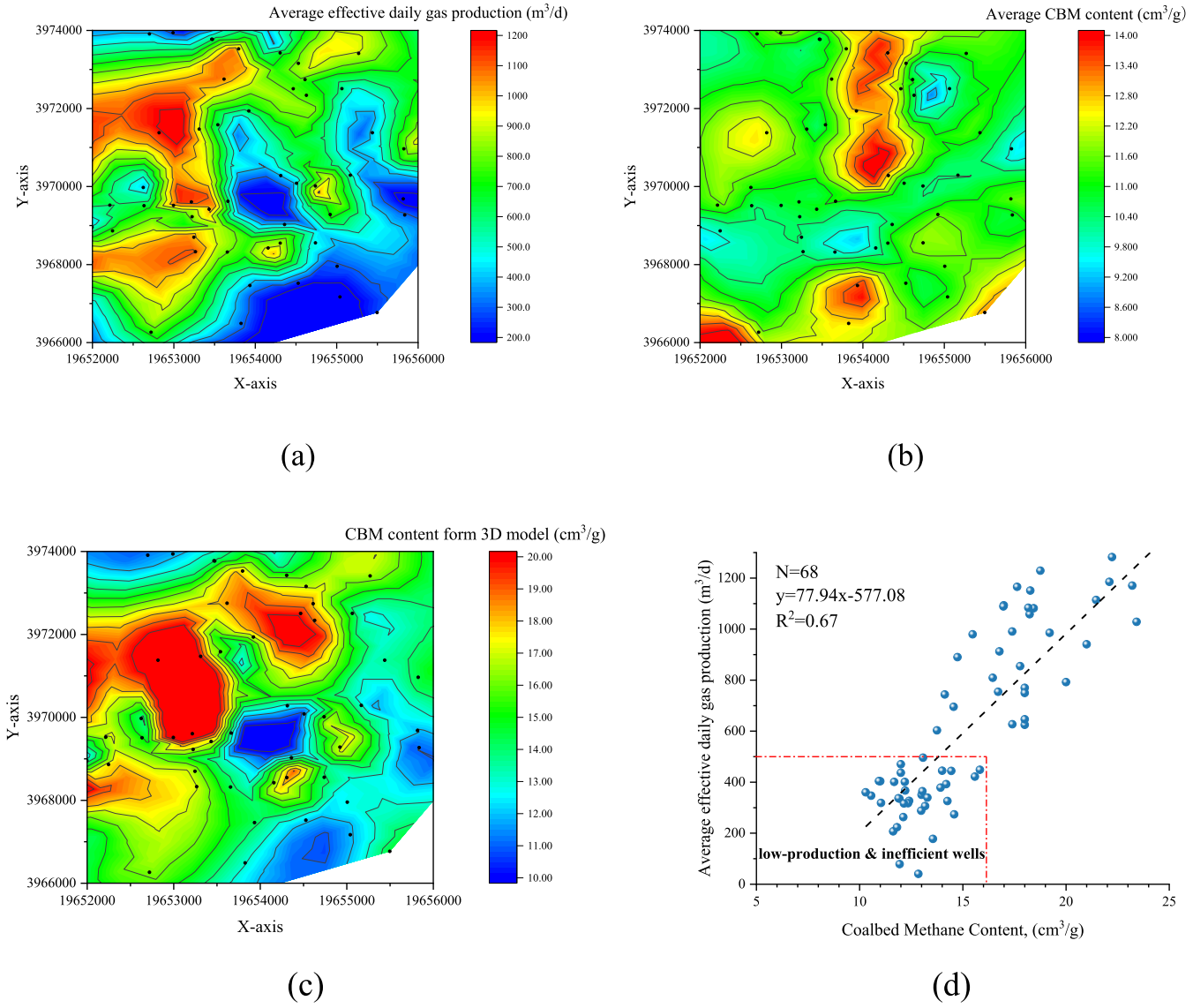


Figure 18. Application comparison effect of two-dimensional model and three-dimensional model. (a) The average effective daily gas production contour map of 68 wells; (b) Contour map of average CBM content of 68 wells; (c) Contour maps of CBM content peaks extracted from 68 wells based on three-dimensional models; (d) The fitting relationship between the peak CBM content and the average effective daily gas production.

Table 2 indicates that the SA-GA-RF method has the lowest average relative error and root-mean-square error, with the highest fit goodness of prediction results to core experimental outcomes, achieving 0.84. Among the other four methods, the least-squares support vector machine method ranked second in effectiveness, while multivariate regression had the highest error. Considering core methane content data and the challenges associated with core sampling, as well as the experimental cycle and economic costs of methane content determination, core data availability is limited, presenting a typical small-sample data problem. Additionally, the distribution of methane content data is uneven. Consequently, for the BPNN method, the small sample size increases the difficulty of model construction. Although the LSSVM method can address small sample issues, it does not solve the problem of imbalanced samples. The ELM has error propagation in the opposite direction to BPNN and, while easier to construct than BPNN, does not yield ideal practical results. As for the multivariate regression method, despite its widespread application, results from the study area indicate that the correlation between logging responses and methane content is not significant, not simply a linear relationship, and the regression method's requirement for a large sample size is also unmet, resulting in suboptimal application effectiveness. The RF method, due to its resampling characteristic, is suitable for small and imbalanced data sets. Additionally, this method does not require normalization of feature vectors, thus preserving the integrity of features. The optimized SA-GA-RF can efficiently and accurately determine hyperparameters, enhancing the applicability of the improved method.

5.1.2. Advantage of Model Building Speed. The SA-GA algorithm can efficiently determine the hyperparameters of the random forest method. For example, comparing it to traditional grid search combined with cross-validation, where the number of decision trees and the number of splitting features are exhaustively searched, the search range for decision trees is set between 1 and 200, with a step size of 1; the search range for the number of splitting features is set between 1 and 6, with a step size of 1. To reduce the randomness of the exploration results, leave-one-out cross-validation is used, with

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

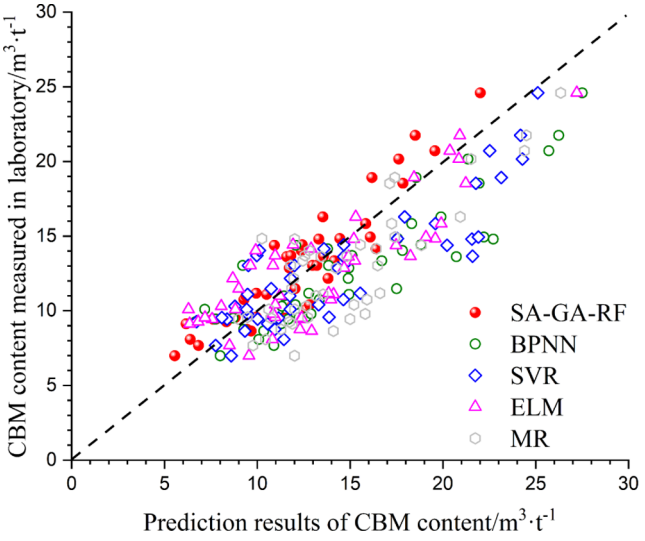


Figure 19. Multimethod calculation of CBM content and core experimental results of the intersection diagram.

Table 2. Application Error Comparison of CBM Content Method Constructed by Different Methods

method	SA-GA-RF	BPNN	SVR	ELM	MR
average relative error (%)	13.13	21.52	19.12	20.84	24.32
RMSE (m^3/t)	1.99	2.91	2.42	2.63	3.02
R^2	0.84	0.75	0.80	0.78	0.69

each search taking 12 h. If further searching of other hyperparameters is required, each additional set of hyperparameters exponentially increases the number of calculations and the computational time. Using the SA-GA-RF method to search for five groups of hyperparameters, the entire search process takes only 0.48 h, significantly enhancing the construction efficiency of the CBM content evaluation model. The construction time of the model is also related to the sample size. Table 3 shows the accuracy under different

Table 3. Accuracy of the Model in the Test Data under Different Training and Verification Data Ratios

	average relative error (%)	
ration (training/testing)	training data	testing data
5:5	7.29	22.42
6:4	6.96	17.20
7:3	5.99	13.13
8:2	5.96	12.98

training and validation data ratios, with the data ratios for the training and test sets set at 5:5, 6:4, 7:3, and 8:2, respectively, using the SA-GA-RF method to build the coalbed methane content prediction model. As the proportion of the training data set increases, the average relative error of the validation data decreases, thus enhancing the model's accuracy. Although all data will be used when constructing the final model, the results in the table indicate that increasing the training samples positively impacts the precision of model evaluation. As exploration and development in the study area deepen, the number of core samples will increase. The continual enrichment of core samples can further upgrade the existing prediction model. The method presented in this paper can accelerate the upgrade of the model, enhancing the efficiency of the approach.

5.2. Limitations of the Methodology. While the gas content prediction model developed using the SA-GA-RF method demonstrates generalizability within the same coal seam of a specific area, it remains susceptible to noise interference. Such noise can be introduced during data collection and preprocessing of logging data, potentially reducing the accuracy of the final model. While these errors are inevitable, they are deemed acceptable within the scope of the model's application. However, the model's applicability may be limited when applied to areas with different geological backgrounds or coal seams at varying depths. Particularly at depths below a critical threshold, the relationship between coal seam depth and gas content changes, which could lead to inaccuracies in methane content predictions. In such scenarios, the model may lose its effectiveness. Nevertheless, the model construction process outlined in this paper can be adapted to develop a tailored CBM content evaluation model suitable for these specific conditions.

In practical applications, numerous technical challenges are encountered. One major issue is the incompleteness of logging curve series in some wells due to previous construction techniques and a lack of clarity in understanding. This deficiency, whether in constructing curves or in remodeling, inevitably results in a loss of accuracy. Specifically, the absence of AC and CNL logging curves, when only the DEN porosity curve is available, significantly impacts model precision. This paper considers the error introduced by substituting other parameters during modeling and therefore does not use coal quality analysis data for modeling. Although scholars have demonstrated that geophysical logging data can effectively construct coal seam industrial components, the approach is difficult to generalize when the logging series in production wells are incomplete. The accuracy of the three-dimensional model depends on the density of CBM wells. The dense CBM wells and the high accuracy of the model can improve the accuracy of the model. In the current production, the distribution of CBM wells is not uniform, and there are no wells to control in some areas. The current three-dimensional model faces challenges in accurately incorporating structural belts, relying primarily on gas content curves generated by the logging evaluation model. This approach still lacks a comprehensive geological perspective and necessitates improvements across multiple scales to enhance model fidelity.[82]

With the advancement of geophysical logging instruments and continuous improvement in experiments, issues related to the instruments' resistance to interference and the measurement of CBM content are gradually being resolved. Using the determination of gas content in coalbed as an exemplar, numerous scholars have made significant advancements. For instance, some have utilized adsorption theory and production dynamics to assess gas content levels,[83] while others have employed desorption mechanisms to identify parameters indicative of CBM content.[84,85] Accurate determination of gas content plays a crucial role in enhancing model precision and provides a robust foundation of data for logging-based CBM content evaluations. Furthermore, with ongoing advancements in evaluation techniques, both coal quality parameters and CBM content assessments are becoming increasingly precise.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

6. CONCLUSIONS

(1) The relationship between CBM content and geophysical logging responses is complex, and the limited number of core sample data makes it challenging to accurately evaluate gas content in coal seams.

(2) Based on the RF method, suitable for addressing issues with uneven sample distribution and small data volumes, the SA-GA method was used to optimize the hyperparameters within the RF method, resulting in the SA-GA-RF method. The SA-GA-RF method offers advantages such as rapid model construction, high precision, and strong generalizability. It performs well with validation data and new wells, outperforming BPNN, LSSVM, ELM algorithms, and multivariate regression method in terms of accuracy and is more efficient than the traditional grid search-cross-validation pattern.

(3) Using the CBM content results calculated by the SA-GA-RF method, a three-dimensional coalbed methane content model for the No. 3 coal seam applicable to the area was constructed. Compared to two-dimensional isopleth maps, this three-dimensional model provides a more detailed description of CBM content, effectively identifying areas with high gas content and the distribution of high-gas-content layers vertically.

Accurate evaluation of CBM content provides a data foundation for exploration, reserve estimation, and production capacity building in coalbed methane fields. It has significant implications for precision selection of CBM areas, optimal choice of perforation layers, and comprehensive management of low-production and inefficient wells. The research methods discussed in this paper hold practical engineering significance.

AUTHOR INFORMATION

Corresponding Author

Zhansong Zhang − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*; Email: Zhangzhs@yangtzeu.edu.cn

Authors

Jianhong Guo − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*; orcid.org/0000-0002-1145-170X

Guangshan Guo − *CNOOC Research Institute, Beijing 100028, China*

Hang Xiao − *Research Institute of Exploration & Development, Sinopec Jianghan Oilfield Company, Wuhan 430223, China*

Qing Zhao − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*

Chaomo Zhang − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*

Hengyang Lv − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*

Zuomin Zhu − *Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education, Yangtze University, Wuhan 430100, China; College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430100, China*

Can Wang − *Hubei Geol Bur, Hydrogeol & Engn Geol Inst, Jingzhou 434007, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c04305

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was financially sponsored by Open Fund of Key Laboratory of Exploration Technologies for Oil and Gas Resources, Ministry of Education (No. K2023-02), and the Hainan Provincial Natural Science Foundation of China (No. 422QN355).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to express their most sincere gratitude to the field workers.

REFERENCES

(1) Hamawand, I.; Yusaf, T.; Hamawand, S. G. Coal seam gas and associated water: A review paper. *Renewable Sustainable Energy Rev.* **2013**, *22*, 550−560.

(2) Palmer, I. Coalbed methane completions: a world view. *Int. J. Coal Geol.* **2010**, *82* (3−4), 184−195.

(3) Karacan, C. Ö.; Ruiz, F. A.; Cotè, M.; Phipps, S. Coal mine methane: a review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. *Int. J. Coal Geol.* **2011**, *86* (2−3), 121−156.

(4) Moore, T. A. Coalbed methane: A review. *Int. J. Coal Geol.* **2012**, *101*, 36−81.

(5) Cai, Y.; Liu, D.; Yao, Y.; Li, J.; Qiu, Y. Geological controls on prediction of coalbed methane of No. 3 coal seam in Southern Qinshui Basin, North China. *Int. J. Coal Geol.* **2011**, *88* (2−3), 101−112.

(6) Fu, H.; Tang, D.; Xu, T.; Xu, H.; Tao, S.; Li, S.; Yin, Z.; Chen, B.; Zhang, C.; Wang, L. Characteristics of pore structure and fractal dimension of low-rank coal: A case study of Lower Jurassic Xishanyao coal in the southern Junggar Basin, NW China. *Fuel* **2017**, *193*, 254−264.

(7) Karacan, C. Ö.; Ulery, J. P.; Goodman, G. V. R. A numerical evaluation on the effects of impermeable faults on degasification efficiency and methane emissions during underground coal mining. *Int. J. Coal Geol.* **2008**, *75* (4), 195−203.

(8) Maricic, N.; Mohaghegh, S. D.; Artun, E. A parametric study on the benefits of drilling horizontal and multilateral wells in coalbed methane reservoirs. *SPE Reservoir Eval. Eng.* **2008**, *11* (06), 976−983.

(9) Jiang, B.; Qu, Z.; Wang, G. G. X.; Li, M. Effects of structural deformation on formation of coalbed methane reservoirs in Huaibei coalfield, China. *Int. J. Coal Geol.* **2010**, *82* (3−4), 175−183.

(10) Zhang, J.; Liu, D.; Cai, Y.; Pan, Z.; Yao, Y.; Wang, Y. Geological and hydrological controls on the accumulation of coalbed methane

within the No. 3 coal seam of the southern Qinshui Basin. *Int. J. Coal Geol.* **2017**, *182*, 94−111.
(11) Zhou, F.; Allinson, G.; Wang, J.; Sun, Q.; Xiong, D.; Cinar, Y. Stochastic modelling of coalbed methane resources: A case study in Southeast Qinshui Basin, China. *Int. J. Coal Geol.* **2012**, *99*, 16−26.
(12) Zhou, F.; Guan, Z. Uncertainty in estimation of coalbed methane resources by geological modelling. *J. Nat. Gas. Sci. Eng.* **2016**, *33*, 988−1001.
(13) Liu, D.; Yao, Y.; Tang, D.; Tang, S.; Che, Y.; Huang, W. Coal reservoir characteristics and coalbed methane resource assessment in Huainan and Huaibei coalfields, Southern North China. *Int. J. Coal Geol.* **2009**, *79* (3), 97−112.
(14) Pashin, J. C.; McIntyre-Redden, M. R.; Mann, S. D.; Kopaska-Merkel, D. C.; Varonka, M.; Orem, W. Relationships between water and gas chemistry in mature coalbed methane reservoirs of the Black Warrior Basin. *Int. J. Coal Geol.* **2014**, *126*, 92−105.
(15) Hildenbrand, A.; Krooss, B. M.; Busch, A.; Gaschnitz, R. Evolution of methane sorption capacity of coal seams as a function of burial history—a case study from the Campine Basin, NE Belgium. *Int. J. Coal Geol.* **2006**, *66* (3), 179−203.
(16) Hou, S.; Wang, X.; Wang, X.; Yuan, Y.; Zhuang, X.; Wang, X. Geological controls on gas saturation in the Yanchuannan coalbed methane field, southeastern Ordos Basin, China. *Mar. Pet. Geol.* **2016**, *78*, 254−270.
(17) Hower, J. C.; Gayer, R. A. Mechanisms of coal metamorphism: case studies from Paleozoic coalfields. *Int. J. Coal Geol.* **2002**, *50* (1−4), 215−245.
(18) Scott, S.; Anderson, B.; Crosdale, P.; Dingwall, J.; Leblang, G. Coal petrology and coal seam gas contents of the Walloon Subgroup—Surat Basin, Queensland, Australia. *Int. J. Coal Geol.* **2007**, *70* (1−3), 209−222.
(19) Yan, T.; Yao, Y.; Liu, D. Critical tectonic events and their geological controls on gas generation, migration, and accumulation in the Weibei coalbed methane field, southeast Ordos basin. *J. Nat. Gas. Sci. Eng.* **2015**, *27*, 1367−1380.
(20) Zhou, F.; Yao, G.; Tyson, S. Impact of geological modeling processes on spatial coalbed methane resource estimation. *Int. J. Coal Geol.* **2015**, *146*, 14−27.
(21) Nolde, J. E.; Spears, D. A preliminary assessment of in place coalbed methane resources in the Virginia portion of the central Appalachian Basin. *Int. J. Coal Geol.* **1998**, *38* (1−2), 115−136.
(22) Ni, X.; Tan, X.; Wang, B.; Fu, X. An evaluation method for types of low-production coalbed methane reservoirs and its application. *Energy Rep.* **2021**, *7*, 5305−5315.
(23) Zuo, S.; Zhang, L.; Deng, K. Experimental study on gas adsorption and drainage of gas-bearing coal subjected to tree-type hydraulic fracturing. *Energy Rep.* **2022**, *8*, 649−660.
(24) Zou, Z.; Liu, D.; Cai, Y.; Wang, Y.; Li, J. Geological factors and reservoir properties affecting the gas content of coal seams in the Gujiao Area, Northwest Qinshui Basin, China. *Energies* **2018**, *11* (5), 1044.
(25) Meng, Y.; Tang, D.; Xu, H.; Li, Y.; Gao, L. Coalbed methane produced water in China: status and environmental issues. *Environ. Sci. Pollut. Res.* **2014**, *21*, 6964−6974.
(26) Alexeev, A. D.; Feldman, E. P.; Vasilenko, T. A. Methane desorption from a coal-bed. *Fuel* **2007**, *86* (16), 2574−2580.
(27) Flores, R. M. Coalbed methane: from hazard to resource. *Int. J. Coal Geol.* **1998**, *35* (1−4), 3−26.
(28) Mardon, S. M.; Eble, C. F.; Hower, J. C.; Takacs, K.; Mastalerz, M.; Bustin, R. M. Organic petrology, geochemistry, gas content and gas composition of Middle Pennsylvanian age coal beds in the Eastern Interior (Illinois) Basin: Implications for CBM development and carbon sequestration. *Int. J. Coal Geol.* **2014**, *127*, 56−74.
(29) Fu, X.; Qin, Y.; Wang, G. G. X.; Rudolph, V. Evaluation of gas content of coalbed methane reservoirs with the aid of geophysical logging technology. *Fuel.* **2009**, *88* (11), 2269−2277.
(30) Wang, Y.; Liu, D.; Cai, Y.; Yao, Y.; Zhou, Y. Evaluation of structured coal evolution and distribution by geophysical logging methods in the Gujiao Block, northwest Qinshui basin, China. *J. Nat. Gas. Sci. Eng.* **2018**, *51*, 210−222.
(31) Kim, A. G. *Estimating methane content of bituminous coalbeds from adsorption data*; Department of the Interior, Bureau of Mines: Washington, United States, 1977.
(32) Ghosh, S.; Chatterjee, R.; Paul, S.; Shanker, P. Designing of plug-in for estimation of coal proximate parameters using statistical analysis and coal seam correlation. *Fuel* **2014**, *134*, 63−73.
(33) Morin, R. H. Hydrologic properties of coal beds in the Powder River Basin, Montana I. Geophysical log analysis. *J. Hydrol.* **2005**, *308* (1−4), 227−241.
(34) Roslin, A.; Esterle, J. S. Electrofacies analysis using high-resolution wireline geophysical data as a proxy for inertinite-rich coal distribution in Late Permian Coal Seams, Bowen Basin. *Int. J. Coal Geol.* **2015**, *152*, 10−18.
(35) Zhou, F.; Yao, G. Sensitivity analysis in permeability estimation using logging and injection-falloff test data for an anthracite coalbed methane reservoir in Southeast Qinshui Basin, China. *Int. J. Coal Geol.* **2014**, *131*, 41−51.
(36) Deng, S.; Hu, Y.; Chen, D.; Ma, Z.; Li, H. Integrated petrophysical log evaluation for coalbed methane in the Hancheng area, China. *J. Geophys. Eng.* **2013**, *10* (3), No. 035009.
(37) Kim, A. G. *Estimating methane content of bituminous coalbeds from adsorption data*, Department of the Interior, Bureau of Mines, 1977, 8245.
(38) Hawkins, J. M.; Schraufnagel, R. A.; Olszewski, A. J. Estimating coalbed gas content and sorption isotherm using well log data. In *SPE Annual Technical Conference and Exhibition*, Washington, D.C., United States, October 04, 1992, Paper ID: SPE-24905-MS. DOI: 10.2118/24905-MS.
(39) Dong, C.; Hegeman, P. S.; Carnegie, A.; Elshahawi, H. Downhole measurement of methane content and GOR in formation fluid samples. *SPE Reservoir Eval. Eng.* **2006**, *9* (01), 7−14.
(40) Shao, X.; Sun, Y.; Sun, J.; Tang, D.; Xu, H.; Dong, X.; Lv, Y. Log interpretation for coal petrologic parameters: A case study of Hancheng mining area, Central China. *Pet. Explor. Dev.* **2013**, *40* (5), 599−605.
(41) Jin, Z.; Xue, H.; Gao, H.; Zhao, X.; Bai, G. Technology for evaluation of CBM reservoir logging and its application. *Coal Geol. Explor.* **2013**, *41* (2), 42−45.
(42) Hu, X.; Yang, S.; Zhou, X.; Zhang, G.; Xie, B. A quantification prediction model of coalbed methane content and its application in Pannan coalfield, Southwest China. *J. Nat. Gas. Sci. Eng.* **2014**, *21*, 900−906.
(43) Yu, J.; Zhu, L.; Qin, R.; Zhang, Z.; Li, L.; Huang, T. Combining k-means clustering and random forest to evaluate the gas content of coalbed bed methane reservoirs. *Geofluids* **2021**, *2021*, 1−8.
(44) Chunrong, W.; Minqiang, X.; Jianhua, S.; Xiang, L.; Chenrun, J. Coal mine gas emission gray dynamic prediction. *Procedia Eng.* **2011**, *26*, 1157−1167.
(45) Huang, Z.; Zou, C.; Yang, Y.; Zhang, G.; Wang, W. Coalbed methane reservoir evaluation from wireline logs in TS district, southern Qinshui Basin. *Geoscience* **2012**, *26* (6), 1275.
(46) Zhu, L.; Ma, Y.; Cai, J.; Zhang, C.; Wu, S.; Zhou, X. Key factors of marine shale conductivity in southern China—Part II: The influence of pore system and the development direction of shale gas saturation models. *J. Pet. Sci. Eng.* **2022**, *209*, No. 109516.
(47) Guo, J.; Zhang, Z.; Xiao, H.; Zhang, C.; Zhu, L.; Wang, C. Quantitative interpretation of coal industrial components using a gray system and geophysical logging data: A case study from the Qinshui Basin, China. *Front. Earth Sci.* **2023**, *10*, No. 1031218.
(48) Zhou, B.; O'Brien, G. Improving coal quality estimation through multiple geophysical log analysis. *Int. J. Coal Geol.* **2016**, *167*, 75−92.
(49) Liu, H.; Sang, S.; Wang, G. G. X.; Li, Y.; Li, M.; Liu, S. Evaluation of the synergetic gas-enrichment and higher-permeability regions for coalbed methane recovery with a fuzzy model. *Energy* **2012**, *39* (1), 426−439.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788

(50) Jiang, S.; Sun, P.; Lyu, F.; Zhu, S.; Zhou, R.; Li, B.; He, T.; Lin, Y.; Gao, Y.; Song, W.; Xu, H. Machine learning (ML) for fluvial lithofacies identification from well logs: A hybrid classification model integrating lithofacies characteristics, logging data distributions, and ML models applicability. *Geoenergy Sci. Eng.* **2024**, *233*, No. 212587.
(51) Liu, H.; Wu, Y.; Cao, Y.; Lv, W.; Han, H.; Li, Z.; Chang, J. Well logging based lithology identification model establishment under data drift: A transfer learning method. *Sensors* **2020**, *20* (13), 3643.
(52) Xia, H.; Qin, Y.; Zhang, L.; Cao, Y.; Xu, J. Forecasting of coalbed methane (CBM) productivity based on rough set and least squares support vector machine. In *2017 25th International Conference on Geoinformatics*; IEEE: Buffalo, NY, USA, 02−04 August 2017; pp 1−6 DOI: 10.1109/GEOINFORMATICS.2017.8090914.
(53) Chen, T.; Zhang, Z.; Zhou, X.; Guo, J.; Xiao, H.; Tan, C.; Qin, R.; Yu, J. Prediction model of coalbed methane content based on well logging parameter optimization. *Coal. Geol. Explor.* **2021**, *49* (3), 30.
(54) Zhu, L.; Zhang, C.; Zhang, C.; Zhang, Z.; Nie, X.; Zhou, X.; Liu, W.; Wang, X. Forming a new small sample deep learning model to predict total organic carbon content by combining unsupervised learning with semisupervised learning. *Appl. Soft Comput.* **2019**, *83*, No. 105596.
(55) Omotilewa, O. J.; Panja, P.; Vega-Ortiz, C.; McLennan, J. Evaluation of enhanced coalbed methane recovery and carbon dioxide sequestration potential in high volatile bituminous coal. *J. Nat. Gas. Sci. Eng.* **2021**, *91*, No. 103979.
(56) Zhou, F.; Oraby, M.; Luft, J.; Guevara, M. O.; Keogh, S.; Lai, W. Coal seam gas reservoir characterisation based on high-resolution image logs from vertical and horizontal wells: A case study. *Int. J. Coal Geol.* **2022**, *262*, No. 104110.
(57) Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Sicre, C. M.; Dedieu, G. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sens.* **2017**, *9* (2), 173.
(58) Xing, L. R.; Yao, Y.; Liu, D.; Liu, J.; Zhou, L.; Li, H. Geological Characteristics of Coalbed Methane Reservoir in Southern Shizhuang Block, Southeastern Qinshui Basin. *Appl. Mech. Mater.* **2013**, *295-298*, 3209−3212.
(59) Cao, Y.; Davis, A.; Liu, R.; Liu, X.; Zhang, Y. The influence of tectonic deformation on some geochemical properties of coals—a possible indicator of outburst potential. *Int. J. Coal Geol.* **2003**, *53* (2), 69−79.
(60) Xu, H.; Tang, D.; Tang, S.; Zhao, J.; Meng, Y.; Tao, S. A dynamic prediction model for gas−water effective permeability based on coalbed methane production data. *Int. J. Coal Geol.* **2014**, *121*, 44−52.
(61) Hou, H.; Shao, L.; Wang, S.; Xiao, Z.; Wang, X.; Li, Z.; Mu, G. Influence of depositional environment on coalbed methane accumulation in the Carboniferous-Permian coal of the Qinshui Basin, northern China. *Front. Earth Sci.* **2019**, *13*, 535−550.
(62) Hou, H.; Shao, L.; Tang, Y.; Li, Y.; Liang, G.; Xin, Y.; Zhang, J. Coal seam correlation in terrestrial basins by sequence stratigraphy and its implications for paleoclimate and paleoenvironment evolution. *J. Earth Sci.* **2023**, *34* (2), 556−570.
(63) Meng, Z.; Yan, J.; Li, G. Controls on gas content and carbon isotopic abundance of methane in Qinnan-East coal bed methane block, Qinshui Basin, China. *Energy Fuels* **2017**, *31* (2), 1502−1511.
(64) Guo, J.; Zhang, Z.; Guo, G.; Xiao, H.; Zhu, L.; Zhang, C.; Tang, X.; Zhou, X.; Zhang, Y.; Wang, C. Evaluation of Coalbed Methane Content by Using Kernel Extreme Learning Machine and Geophysical Logging Data. *Geofluids* **2022**, *2022*, No. 3424367.
(65) Li, J.; Liu, D.; Yao, Y.; Cai, Y.; Qiu, Y. Evaluation of the reservoir permeability of anthracite coals by geophysical logging data. *Int. J. Coal Geol.* **2011**, *87* (2), 121−127.
(66) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(67) Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123−140.
(68) Breiman, L. Using iterated bagging to debias regressions. *Mach. Learn.* **2001**, *45*, 261−277.
(69) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220* (4598), 671−680.
(70) Djurišić, A. B.; Elazar, J. M.; Rakić, A. D. Simulated-annealing-based genetic algorithm for modeling the optical constants of solids. *Appl. Opt.* **1997**, *36* (28), 7097−7103.
(71) Fu, X.; Qin, Y.; Wang, G. G. X.; Rudolph, V. Evaluation of coal structure and permeability with the aid of geophysical logging technology. *Fuel* **2009**, *88* (11), 2278−2285.
(72) Li, S.; Zhang, C.; Hu, A.; Chen, D.; Yan, S.Building porosity model of coalbed using collocated cokriging. In *EAGE Conference on Petroleum Geostatistics*; European Association of Geoscientists & Engineers, 2007 .
(73) Hou, H.; Zhang, H.; Shao, L.; Guo, S.; Zhao, M.; Wang, S. Coal macrolithotype distribution and its genetic analyses in the deep Jiaozuo coalfield using geophysical logging data. *ACS Omega* **2021**, *6* (51), 35523−35537.
(74) Hou, H.; Liang, G.; Shao, L.; Tang, Y.; Mu, G. Coalbed methane enrichment model of low-rank coals in multi-coals superimposed regions: a case study in the middle section of southern Junggar Basin. *Front. Earth Sci.* **2021**, *15* (2), 256−271.
(75) Hou, H.; Shao, L.; Tang, Y.; Li, Z.; Zhao, S.; Yao, M.; Wang, X.; Zhang, J. Pore structure characterization of middle-and high-ranked coal reservoirs in northern China. *AAPG Bull.* **2023**, *107* (2), 213−241.
(76) Zhu, S.; Peng, X.; You, Z.; Li, C.; Deng, P. The effects of cross-formational water flow on production in coal seam gas reservoir: A case study of Qinshui Basin in China. *J. Pet. Sci. Eng.* **2020**, *194*, No. 107516.
(77) Zhang, Z.; Qin, Y.; Wang, G.; Sun, H.; You, Z.; Jin, J.; Yang, Z. Evaluation of coal body structures and their distributions by geophysical logging methods: case study in the Laochang block, eastern Yunnan, China. *Nat. Resour. Res.* **2021**, *30*, 2225−2239.
(78) Feng, G.; Zhao, X.; Wang, M.; Song, Y.; Zheng, S.; He, Y.; You, Z. Fractal pore and its impact on gas adsorption capacity of outburst coal: Geological significance to coalbed gas occurrence and outburst. *Adsorpt. Sci. Technol.* **2022**, *2022*, No. 4273900.
(79) Shi, J.; Zeng, L.; Dong, S.; Wang, J.; Zhang, Y. Identification of coal structures using geophysical logging data in Qinshui Basin, China: Investigation by kernel Fisher discriminant analysis. *Int. J. Coal Geol.* **2020**, *217*, No. 103314.
(80) Mahmood, Z.; Khan, S. On the use of k-fold cross-validation to choose cutoff values and assess the performance of predictive models in stepwise regression. *Int. J. Biostat.* **2009**, *5* (1), 1−19.
(81) Guo, Z.; Zhao, J.; You, Z.; Li, Y.; Zhang, S.; Chen, Y. Prediction of coalbed methane production based on deep learning. *Energy* **2021**, *230*, No. 120847.
(82) Tao, S.; Pan, Z.; Tang, S.; Chen, S. Current status and geological conditions for the applicability of CBM drilling technologies in China: A review. *Int. J. Coal Geol.* **2019**, *202*, 95−108.
(83) Feng, R. A method to evaluated gas content with coalbed methane reservoir based on adsorption theory and production analysis. *Geofluids* **2022**, *2022* (1), No. 7341886.
(84) Zhang, Z.; Wang, B.; Cui, Y. CBM desorption model and stages based on a natural desorption experiment. *ACS Omega* **2022**, *7* (42), 37555−37563.
(85) Tao, S.; Chen, S.; Pan, Z. Current status, challenges, and policy suggestions for coalbed methane industry development in China: A review. *Energy Sci. Eng.* **2019**, *7* (4), 1059−1074.

https://doi.org/10.1021/acsomega.4c04305
ACS Omega 2024, 9, 35769−35788