

Research Article

Combining K-Means Clustering and Random Forest to Evaluate the Gas Content of Coalbed Bed Methane Reservoirs

Jie Yu,^{1,2} Linqi Zhu³,,³ Ruibao Qin,² Zhansong Zhang,¹ Li Li,² and Tao Huang²

¹Yangtze University, Wuhan 430100, China

²CNOOC Research Institute, Beijing 100027, China

³Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China

Correspondence should be addressed to Linqi Zhu; zhulq@idsse.ac.cn

Received 27 June 2021; Revised 19 July 2021; Accepted 30 July 2021; Published 21 August 2021

Academic Editor: Wei Zhao

Copyright © 2021 Jie Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The accurate calculation of the gas content of coalbed bed methane (CBM) reservoirs is of great significance. However, due to the weak correlation between the logging response of coalbed methane reservoirs and the gas content parameters and strong nonlinear characteristics, it is difficult for conventional gas content calculation algorithms to obtain more reliable results. This paper proposes a CBM reservoir gas content assessment method combining K-means clustering and random forest. The K-means clustering is used to divide the reservoirs and distinguish the types to establish a random forest model. Judging from the evaluation effect of the research block, the prediction accuracy of the new method is significantly higher than that of the original method, and more accurate gas content prediction values can be obtained for different types of reservoirs. Studies have shown that this method can help the gas content evaluation of CBM reservoirs, improve the accuracy of gas content evaluation, and better support the exploration and development of CBM reservoirs. The results of this study show that the random forest method based on clustering can effectively distinguish the relationship between different logging responses and gas content. On this basis, the random forest algorithm modeling can effectively characterize the complex relationship between gas content and logging curve response. In the case of poor correlation between gas content and logging curve, the gas content of the reservoir can also be accurately calculated.

1. Introduction

With the continuous progress of exploration and development, research on various unconventional reservoirs such as coalbed bed methane (CBM) and shale gas is in full swing, and they are the key growth points of reserves [1–3]. CBM is natural gas that is generated by biochemistry and pyrolysis during the formation and evolution of coal seams and is stored in coal seams. At present, the United States, Canada, Australia, Russia, India, China, and other countries have all started exploration and development of CBM [4–6].

The gas content of a CBM reservoir is a very important parameter, which determines the reserves and final production of the reservoir [7, 8]. However, compared with other reservoirs, it is more difficult to calculate the gas content of CBM reservoirs, which extremely restricts the determination of high-quality reservoirs and the formulation of develop-

ment plans, resulting in unclear understanding of CBM reservoirs. For the core gas content calculation method, Kim proposes to combine the moisture and ash content with the coal bed temperature and pressure and the equilibrium water state correction amount to calculate the coalbed gas content [9]. Ahmed et al. provided the establishment of an isotherm adsorption model to describe the gas content using an isotherm adsorption experiment [10]. Hawkins et al. proposed to use the Langmuir coal rank equation to predict gas content [11]. However, none of the above methods can predict the vertical continuity of the gas content in the formation. Logging is currently the only method that can accurately predict the vertical gas content change of a single well. It is of great significance to establish an accurate gas content logging evaluation model.

Some scholars have studied the logging calculation method of the gas content of CBM reservoirs. Liu et al. [12]

, Meng et al. [13] and Shao et al. [14] both proposed a statistical method for evaluating the gas content of CBM reservoirs. Jin et al. [15] and Fu et al. [16] also used this method. In addition to the use of statistical models or volume models for gas content evaluation, the relationship between gas content and logging response is too complicated. At present, methods for evaluating coal reservoir gas content using machine learning algorithms have gradually emerged. Hou and Wang [17] used the error back propagation neural network to predict the air content and achieved certain results. Pan and Huang [18] and Wu also used BPNN to predict the air content. Lian et al. [19] introduced support vector machines to the evaluation of air content. Guo et al. [20, 21] used the grey system and random forest to predict the gas content. Xiang et al. [22] proposed the application method of deep learning in CBM logging interpretation and believed that the effect of deep belief network in CBM gas content prediction is better than BPNN, multiple regression, and Langmuir equation method.

Although many scholars have proposed a variety of methods for evaluating the gas content of CBM, it cannot be ignored that the complexity of CBM reservoirs is greater, even higher than that of shale gas reservoirs. The logging response of such a complex reservoir is affected very diversely, and the main controlling factors of logging response of different reservoirs are different. No matter how strong the approximation ability of the model is, it is difficult to accurately evaluate the gas content parameters by establishing a single evaluation model. The innovative part of this article is to use clustering methods to classify data with different feature relationships, so that machine learning algorithms can predict different data more targeted, and use more efficient machine learning algorithms to improve the prediction effect.

Based on this idea, this paper proposes a K-means clustering+random forest air content evaluation method, that is, first collect data and use the clustering method to classify the data. After that, the classified data is used to establish a model separately and evaluate the gas content. Finally, a series of established models are used to apply the logging curve to obtain the final gas content prediction curve for the entire well section. In this way, the influence of different main control factors on the prediction can be eliminated as much as possible, so that the model is more targeted and the prediction effect of the model is improved. Although this approach seems to be a more complicated modeling method, the prediction effect of the actual reservoir gas content has been greatly improved. From the perspective of the prediction effect of the research block, the method proposed in this paper is effective and can help the exploration and development of CBM.

2. Data

The study block is located in the southeast of Qinshui Basin in Central China. Drilling revealed that there are 16 coal seams in Taiyuan formation and Shanxi formation, with the maximum total thickness of 23.6 m. Among them, No. 15 coal seam in Taiyuan formation and No. 3 coal seam in

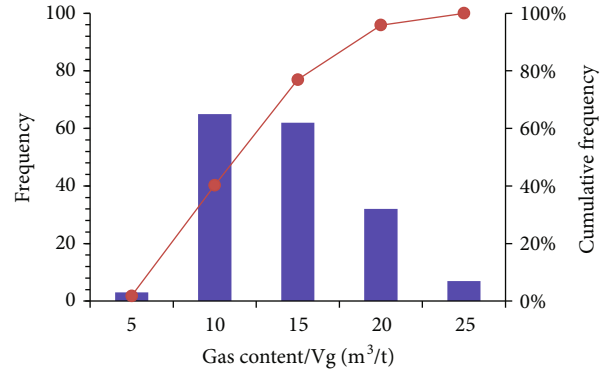


FIGURE 1: Distribution histogram of measured gas content in No. 3 coal.

Shanxi formation are stably distributed in the whole basin. The main mining coal seam is No. 3 coal seam, and No. 15 coal seam is only involved in individual wells. At present, there are more than 1000 coalbed methane wells in the block. The production of wells in different positions varies greatly, and the gas content distribution is unclear, which restricts the exploration and development of coalbed methane. We collected 169 coring gas content measured data from 22 CBM parameter wells in this block, and 6 logging curves including natural gamma ray, spontaneous potential, borehole diameter, deep shallow lateral resistivity, and bulk density. The measured gas content data shows that the gas content of No. 3 coal seam is mainly distributed between 5 and 20 m³/t (Figure 1).

3. Method

3.1. K-Means Clustering. Although we often use classification or regression algorithms in supervised learning methods to predict categories or values, we still often encounter situations where we need to use unsupervised learning methods to obtain a set of data categories. When the amount of data is large, you can consider using clustering algorithms to get different data categories. Clustering is subordinate to unsupervised learning, which does not rely on the defined classes and training examples of class labels. Among them, K-means clustering is a very classic clustering method [23].

At the beginning of K-means clustering, we first need to figure out how to quantitatively calculate the difference between two comparable elements. The smaller the degree of difference, the greater the direct correlation between the two samples, and the more likely it is a rock sample of one type of rock. We define the degree of dissimilarity mathematically here.

Suppose $X = \{x_1, x_2, x_3, \dots, x_n\}$, $Y = \{y_1, y_2, y_3, \dots, y_n\}$, where X and Y are two-element items, each with n measurable characteristic attributes; then, the degree of dissimilarity between X and Y is defined as

$$d(X, Y) = f(X, Y) \longrightarrow R. \quad (1)$$

Among them, R is the real number field. That is to say, the degree of dissimilarity is a mapping of two elements to

the real number field, and the real number quantitatively represents the degree of dissimilarity of the two elements. The calculation of dissimilarity can use Euclidean distance, Manhattan distance, Minkowski distance, and so on. Usually, we use Euclidean distance:

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}. \quad (2)$$

The above method of calculating dissimilarity has a problem, that is, attributes with a large value range have a higher impact on distance than attributes with a small value range. In order to solve this problem, it is generally necessary to normalize the attribute value. The so-called normalization is to map each attribute value proportionally to the same value interval, so as to balance the influence of each attribute on the distance. Usually, each attribute is mapped to the interval [0,1], and the mapping formula is

$$a'_i = \frac{a_i - \min(a_i)}{\max(a_i) - \min(a_i)}. \quad (3)$$

Among them, $\max(a_i)$ and $\min(a_i)$ represent the maximum and minimum values of the i th attribute in all element items. The so-called clustering problem is to give a set of elements D , where each element has n observable attributes, use a certain algorithm to divide D into k subsets, and require the degree of difference between the elements within each subset as much as possible low, and the element dissimilarity of different subsets is as high as possible. Concentration, each subset is called a cluster. Different from classification, classification is exemplary learning, which requires that each category be clarified before classification and that each element is mapped to a category, while clustering is observational learning, and the category may not be known or even the number of categories may not be known before clustering.

K-means tries to find the natural category of the data. The user sets the number of categories to find a good category center. The algorithm flow is as follows:

- (1) Enter the number of data sets and categories K
- (2) Randomly assign the center point of the category
- (3) Put each point into the set of the category center point closest to it
- (4) Move the category center point to the set where it is
- (5) Go to step 3 until convergence

After a number of cycles, the best classification effect can be obtained. Different from marine shale reservoirs, the relationship between gas content of coal reservoirs and logging response of the coal reservoirs is relatively poor, and the laws are inconsistent, which also leads to the unreliability of the final prediction model. This is because coal reservoirs are more complex than shale reservoirs and have worse continuity, which causes the logging of coal seams to be affected by multiple factors. Using the clustering method to obtain multiple categories and establishing corresponding prediction

models based on different categories can greatly improve the prediction results.

3.2. Random Forest. Random forest is a highly flexible machine learning algorithm that has just emerged in the 21st century. It refers to a classifier that contains multiple decision trees. The thinking behind it is similar to group wisdom. In the 1980s, Breiman et al. invented an algorithm for classification trees, which performed classification or regression through repeated dichotomy of data, which greatly reduced the amount of calculation. In 2001, Breiman combined the classification trees into a random forest, that is, randomized the use of variables and the use of data, generated many classification trees, and then summarized the results of the classification trees [24]. Random forest improves the prediction accuracy without a significant increase in the amount of calculation. Random forest is not sensitive to multivariate collinearity, and the results are relatively robust to missing data and unbalanced data and can well predict the effect of thousands of explanatory variables.

Random forest uses a random method to build a forest. There are many decision trees in the forest, and there is no correlation between each decision tree in the random forest. After obtaining the forest, when a new input sample enters, let each decision tree in the forest make a judgment separately to see which category the sample belongs to. The class with the most classification times is the predicted class. Random forest can handle quantities whose attributes are discrete values. The construction process of random forest is as follows:

- (1) If there are N samples, N samples are randomly selected for replacement (one sample is randomly selected each time and then returned to continue selection). Use the selected N samples to train a decision tree as the sample at the root node of the decision tree
- (2) When each sample has M attributes, when each node of the decision tree needs to be split, then m attributes are selected from these M attributes, and the condition $m < M$ is satisfied. Then, from these m attributes, strategies such as information gain are used to select one attribute as the split attribute of the node
- (3) In the process of decision tree formation, each node must be split according to step 2 until it can no longer be split. Note that there is no pruning during the entire decision tree formation process
- (4) Follow steps 1-3 to build a large number of decision trees to form a random forest

In the process of building each decision tree, attention should be paid to the impact of sampling and complete splitting. The first is two random sampling processes. Random forest samples the input data in rows and columns. For line sampling, a replacement method is used, that is, in the sample set obtained by sampling, there may be duplicate samples. Assuming that there are N input samples, there are also N

samples sampled. In this way, when training, the input samples of each tree are not all samples, making it relatively difficult to overfitting. Then, perform column sampling, from M features, select m ($m < M$).

After that, a decision tree is built using a completely split method for the sampled data, so that a certain leaf node of the decision tree cannot continue to split, or all the samples in it point to the same category. Generally, many decision tree algorithms have an important step-pruning, but this is not done here. Since the previous two random sampling processes ensure randomness, even if pruning is not performed, overfitting will not occur. Using a random forest method to predict gas content should be able to achieve better results.

3.3. Combination Method of K-Means Clustering and Random Forest. It is difficult to evaluate the gas content of coal reservoirs, because the logging response has been affected by various factors, resulting in a poor relationship between the logging response and the core. Only by using clustering and other methods to truly combine logging responses for classification, different types of data are affected differently, and the relationship between logging responses and gas content in different categories is closer. Therefore, K-means clustering is performed first, and then based on the results of the clustering, a random forest model of different types is established for final application. In fact, the inherent meaning of this model is similar to that of random forests. It uses K-means clustering combined with random forests to form a “forest group” to predict gas content more accurately. The modeling and forecasting process is as follows:

- (1) Use K-means clustering to divide the data into several categories. The measurement method usually used to compare the results of different K values is the average distance between a data point and its cluster centroid. Since increasing the number of clusters will always reduce the distance to the data point, when K is the same as the number of data points, increasing K will always reduce the metric to zero. Therefore, this indicator cannot be used as the sole target. Conversely, the average distance to the center of mass is plotted as a function of K , and the “elbow point” at which the reduction rate changes sharply can be used to roughly determine the K value
- (2) Use K sets of data and random forest algorithm to train K models. After determining the category of the new data, the corresponding model can be used to calculate the gas content
- (3) When predicting new data, first determine the category of the new data by calculating the Euclidean distance between the sample data and the centroids of multiple classes of data. The new data belongs to the category corresponding to the centroid with the smallest Euclidean distance. After the category is determined, the corresponding model is used for prediction, and the predicted value of the gas content of the sample point is obtained, and the reliability of the

algorithm is determined by comparing with the real value

4. Result

First, the data needs to be further analyzed to clarify the relationship between the gas content of the CBM reservoir and the logging response. The corresponding results are shown in Figure 2.

In Figure 2, V_g refers to the total gas content results obtained through experiments. AC refers to the acoustic time difference curve response, CAL refers to the well diameter curve response, and CNL refers to the neutron porosity curve response. DEN refers to the density curve response, GR refers to the natural gamma curve response, and RD refers to the deep resistivity curve response. It can be clearly seen that the correlation between each curve and gas content is poor, which is obviously different from marine shale reservoirs. From the correlation of each curve, deep resistivity logging, neutron logging, and sonic logging have a relatively good relationship with the gas content parameters of the reservoir. It is recommended that the above curves can be used as the input curve of the model. When the gas content increases, the sonic time difference of the coal seam increases significantly. As the gas content of the reservoir increases the hydrogen index of the coal seam, the neutron porosity also increases. In addition, with the increase of gas content, the response value of deep resistivity logging increased significantly, indicating that the adsorbed gas in coal reservoirs can significantly increase the resistivity of the reservoir and reduce the conductivity of the CBM reservoir. We also recommend that the resistivity curve be logarithmically converted during input. Based on the above data, the K-means clustering research is carried out.

Figure 3 reflects the relationship between the clustering results and the sample-particle distance. It can be clearly seen that when there are more than 3 cluster types, the average distance reduction speed slows down significantly, indicating that it is not necessary to select more than 3 clusters. So here, we choose the number of clusters to be 3. After clustering, the gas content prediction model is established in different categories. Use the established model to predict the modeling samples, and the results are shown in Figure 4.

It can be seen from Figure 4 that, first of all, the correlation between the response of the logging curve and the poor total gas content does have a great impact on the prediction of gas content. Even if the random forest algorithm has strong approximation and generalization capabilities, the prediction effect obtained is very poor. It can be clearly seen from the results on the right that the prediction effect of the classification modeling after clustering is obviously better than that of Figure 4(a). The prediction effect is poor only when the total gas content is less than $5 \text{ cm}^3/\text{g}$, and the reservoirs with total gas content less than $5 \text{ cm}^3/\text{g}$ are not the reservoirs of our concern. Through the clustering method, data with relatively consistent main control factors are unified and classified, and the model established on this basis is more targeted. Judging from the core prediction results, the ideas

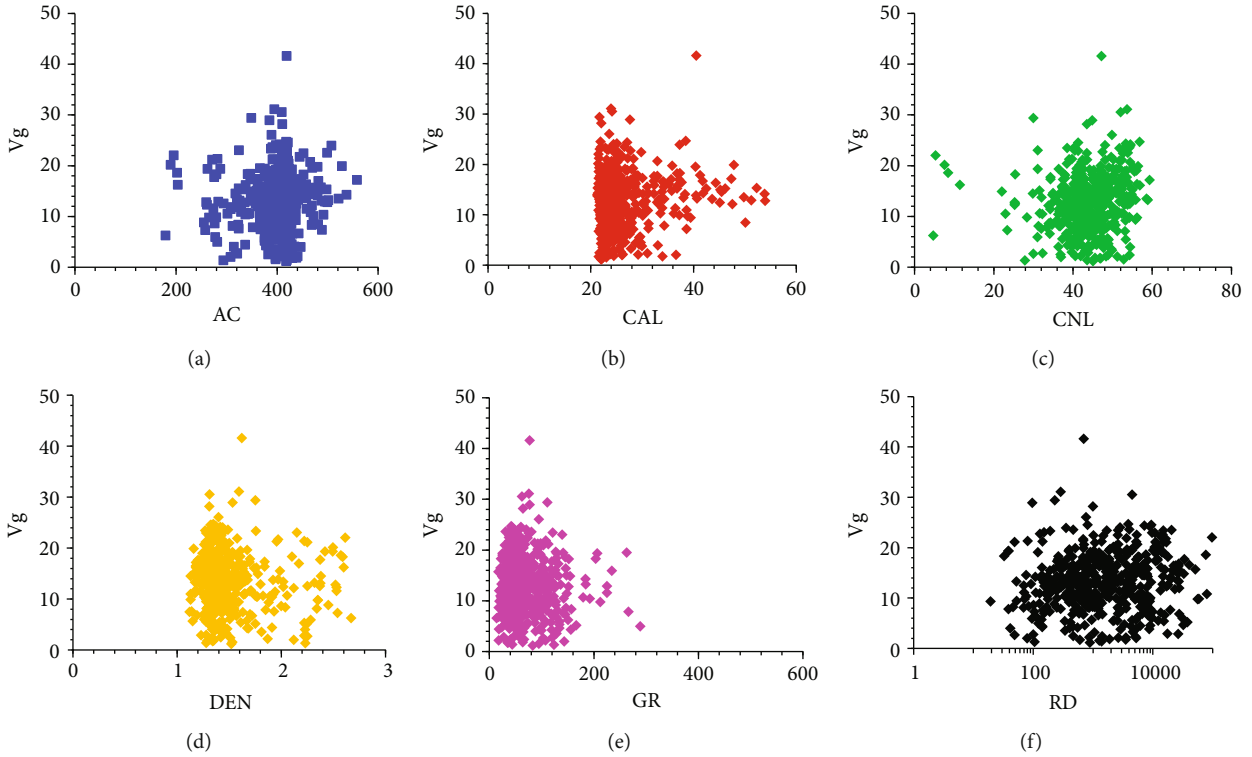


FIGURE 2: Display of the relationship between logging curve response and gas content.

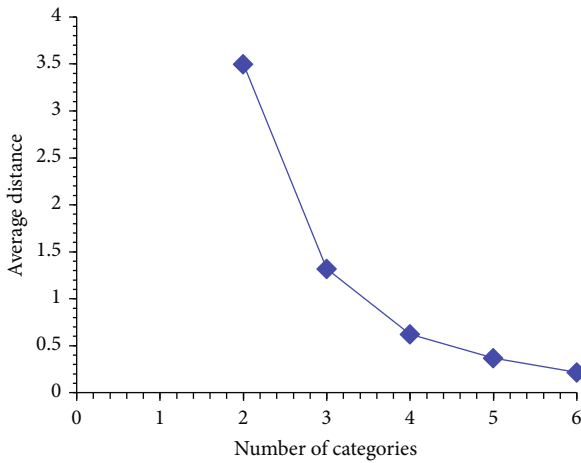


FIGURE 3: Display the relationship between clustering results and the distance between sample and mass point.

proposed in this article are very helpful for the gas content prediction of coal-measure reservoirs.

This method is used to predict the gas content of test wells A and B in the study area. The results are shown in Figures 5 and 6, respectively.

In Figures 5 and 6, the first track is the depth track, and the second track is the caliper curve measured by the four-arm caliper tool. In the third track, the SP curve is the spontaneous potential logging curve, the GR curve is the natural gamma logging curve, Rxo is the microsphere focused resistivity logging curve, RS is the shallow lateral resistivity log-

ging curve, and RD is the deep lateral resistance. Rate logging curve. In the fifth track, DEN is the density logging curve, AC is the sonic logging curve, and CNL is the neutron porosity logging curve. In the sixth channel, Vg_{RF} is the gas content curve directly predicted by random forest, and Vg_{core} is the gas content value of the core. In the seventh track, Vg_{KRF} is the gas content curve obtained by random forest prediction after clustering, Category is the clustering result of the curve, and Conclusion is the interpretation conclusion of the CBM reservoir.

It can be seen from Figure 5 that the correspondence between the Vg_{KRF} curve and the core is much higher than the Vg_{RF} curve, indicating that the random forest modeling effect based on the clustering method is better. Through the analysis of the curve, it can be seen that the gas content of type I reservoirs is relatively low, and the corresponding natural gamma response value of coal reservoirs corresponding to category III is low, the density response value is low, the acoustic wave response value is relatively high, and the resistivity response value is relatively high. This shows that type III coal reservoirs are high-quality coal reservoirs with higher coal content, and their gas content should also be higher than other reservoirs. From the perspective of the prediction effect, it is obvious that the gas content prediction results directly based on the random forest algorithm predict low gas content in type III reservoirs, especially in the interval

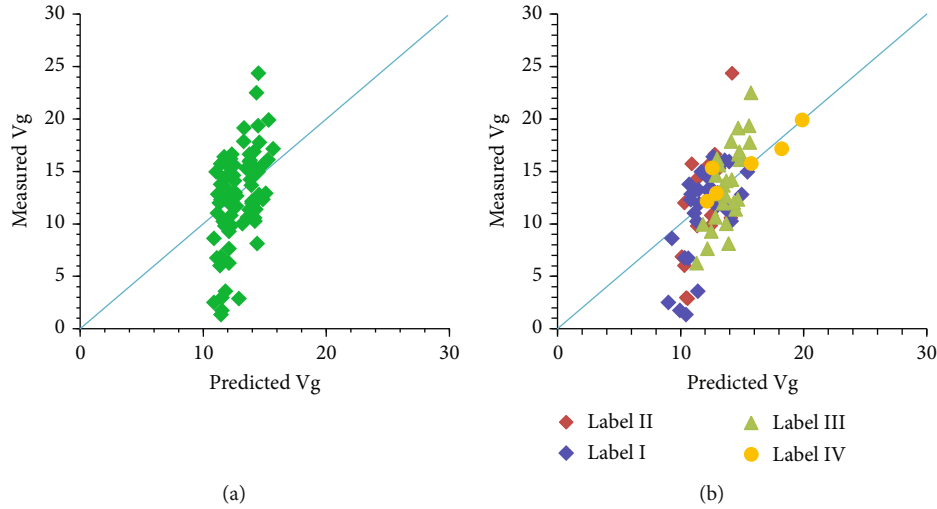


FIGURE 4: The predicted total gas content is compared with the experimental results. (a) The result of direct use of random forest algorithm prediction, and (b) the prediction result obtained by combining K-means clustering and random forest algorithm.

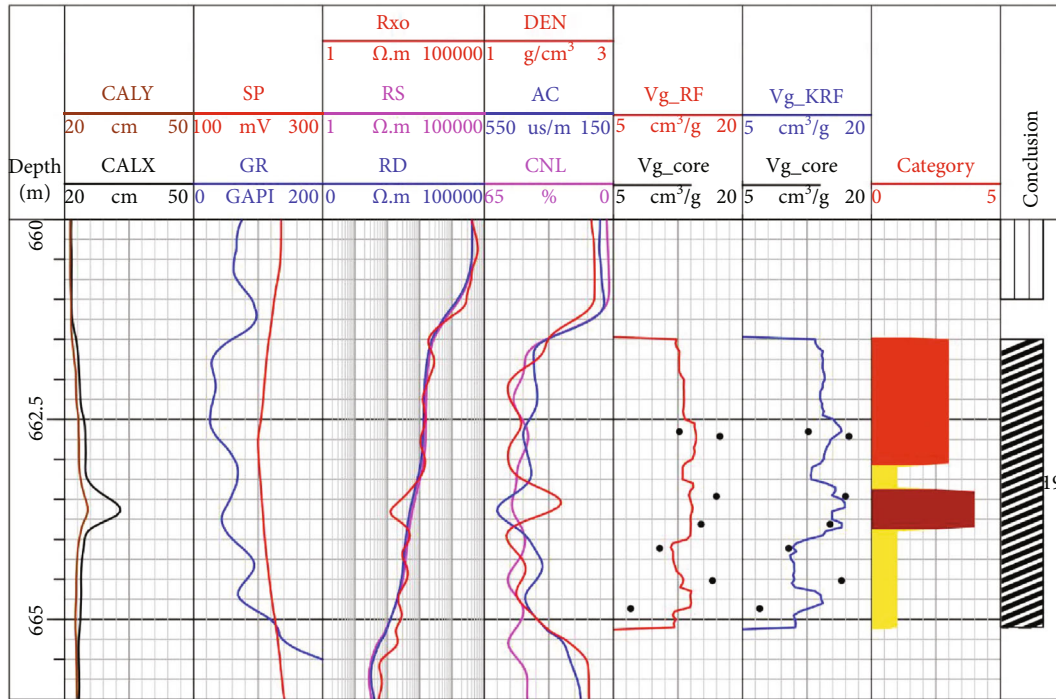


FIGURE 5: Gas content prediction results of well A in the study area.

with very high gas content. In reality, this will make it difficult for us to find the best quality reservoirs. In addition, it has to be mentioned that although we only use the acoustic log response value, resistivity log response value, and neutron log response value, other curves also have a good corresponding relationship with the category, which proves that the accuracy of the class. Type IV reservoirs obviously correspond to the expanded diameter interval, and targeted modeling for this interval can enhance the reliability of the model as much as possible. Therefore, from the application effect of well A, the gas content evaluation method proposed in this paper is more reliable than previous methods.

Figure 6 shows the importance of targeted models. The coal reservoir in Figure 6 is basically a type I reservoir. The gas content curve obtained by directly using random forest for modeling and prediction has very small fluctuations and is not very specific, which makes it difficult for us to directly use the results for high-quality reservoir recognition. The prediction effect of Vg_KRF is relatively more accurate and can be used for high-precision characterization of gas content. However, it can be seen that at 1243.1 m-1245.5 m, the predicted result of Vg_KRF is too small, but the predicted trend is consistent with the actual core trend, indicating that the clustering results need to be adjusted. Or, due to the

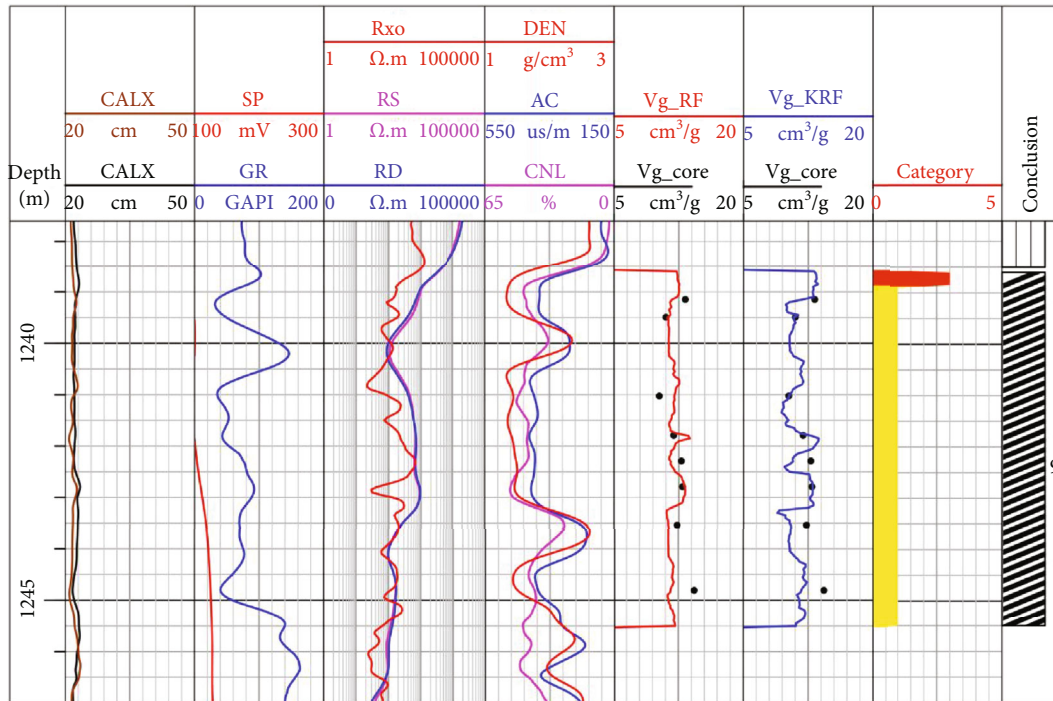


FIGURE 6: Gas content prediction results of well B in the study area.

saving of CBM reservoir development costs, the resolution of the logging tools used for measurement is not enough, and the logging response is disturbed when the vertical change of the reservoir is severe, which ultimately leads to inaccurate classification. Therefore, in the next step of the study, we can focus on the study of log curve superresolution based on wavelet transform and other methods to further improve the prediction effect. In general, the method proposed in this paper is of great help to the gas content evaluation of coal reservoirs with poor correlation between logging response and reservoir parameters.

5. Summary and Conclusions

The calculation of the gas content of CBM reservoirs is more complicated than other reservoirs. Various characteristics of coal reservoirs will have a series of effects on the logging response. This paper proposes a combination of K-means clustering and random forest algorithm to solve the difficult problem of calculating the gas content of CBM reservoirs. The research conclusions are as follows:

- (1) The fact that the calculation of the gas content of CBM reservoirs is more complicated is that the CBM reservoir itself is relatively complex, which causes the logging response to be affected by multiple factors. It further affects the relationship between CBM logging response and gas content, making the correlation poor
- (2) First, the samples are clustered, and the gas content prediction model of random forest is established for each type of sample. Through clustering results com-

bined with actual logging curve analysis, it can be clearly seen that through clustering algorithms, different types of CBM reservoirs can be effectively divided. Furthermore, by comparing the prediction results of CBM gas content, it can be seen that the gas content model established after clustering is more targeted and can evaluate the gas content more accurately. The method proposed in this paper can improve the calculation accuracy of pure CBM gas content and provide a way of thinking for parameter evaluation when the relationship between logging response and reservoir parameters is poor

Nomenclature

- CBM: Coalbed bed methane
- RF: Random forest
- Vg: Gas content
- AC: Sonic log
- CAL: Caliper log
- CNL: Neutron log
- DEN: Density log
- GR: Gamma log
- RD: Resistivity log.

Data Availability

All our data has been fully displayed in the pictures in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the CNOOC information construction major project (No. 2019-KJZC-010); the Laboratory for Marine Geology, Qingdao National Laboratory for Marine Science and Technology (No. 2MGQNLN-KF202004); China Postdoctoral Science Foundation (Nos. 2021M690161 and 2021T140691); Post-doctorate Funded Project in Hainan Province; the Chinese Academy of Sciences-Special Research Assistant Project; China Engineering Technology Development Strategy Hainan Research Institute Consulting Research Project (No. 20-HN-ZT-01); and Open Fund of Key Laboratory of Exploration Technologies for Oil and Gas Resources (Yangtze University), Ministry of Education (Nos. K2021-03 and K2021-08).

References

- [1] L. Tian, Y. Cao, S. Liu, B. Shi, J. Liu, and D. Elsworth, "Coalbed methane reservoir fracture evaluation through the novel passive microseismic survey and its implications on permeable and gas production," *Journal of Natural Gas Science and Engineering*, vol. 76, p. 103181, 2020.
- [2] L. He, H. Mei, X. Hu, M. Dejam, Z. Kou, and M. Zhang, "Advanced flowing material balance to determine original gas in place of shale gas considering adsorption hysteresis," *SPE Reservoir Evaluation and Engineering*, vol. 22, no. 4, pp. 1282–1292, 2019.
- [3] L. Zhang, Z. Kou, H. Wang et al., "Performance analysis for a model of a multi-wing hydraulically fractured vertical well in a coalbed methane gas reservoir," *Journal of Petroleum Science and Engineering*, vol. 166, pp. 104–120, 2018.
- [4] J. Kumar, V. A. Mendhe, A. D. Kamble et al., "Coalbed methane reservoir characteristics of coal seams of south Karanpura coalfield, Jharkhand, India," *International Journal of Coal Geology*, vol. 196, pp. 185–200, 2018.
- [5] R. Baouche and D. A. Wood, "Characterization and estimation of gas-bearing properties of Devonian coals using well log data from five Illizi Basin wells (Algeria)," *Advances in Geo-Energy Research*, vol. 4, pp. 356–371, 2020.
- [6] Z. Yang, Y. Li, Y. Qin et al., "Development unit division and favorable area evaluation for joint mining CBM," *Petroleum Exploration and Development*, vol. 46, no. 3, pp. 559–568, 2019.
- [7] O. Esen, S. C. Özer, A. Soylu, A. Ramazani Rend, and A. Fişne, "Geological controls on gas content distribution of coal seams in the Kimik coalfield, Soma Basin, Turkey," *International Journal of Coal Geology*, vol. 231, p. 103602, 2020.
- [8] Q. Chen, F. Zhang, F. Qiu et al., "Quantitatively determining gas content using pulsed neutron logging technique in closed gas reservoir," *Journal of Petroleum Science and Engineering*, vol. 198, p. 108149, 2021.
- [9] A. G. Kim, "Estimating methane content of bituminous coal beds from adsorption data," *United States Department of the Interior, Report of Investigations-Bureau of Mines*, vol. 8245, pp. 1–11, 1977.
- [10] U. Ahmed, D. Johnston, and L. Colson, "An advanced and integrated approach to coal formation evaluation," in *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, October 1991.
- [11] J. M. Hawkins, R. A. Schraufnagel, and A. J. Olszewsk, "Estimating coal bed gas content and sorption isotherm using well log data," in *SPE Annual Technical Conference and Exhibition*, Washington, D.C., October 1992.
- [12] R. Liu, J. Wang, W. Liu, S. Guo, and Z. Li, "CBM logging evaluation method based on coal seam structure," *China Coalbed Methane*, vol. 11, no. 4, pp. 22–25, 2014.
- [13] Z. Meng, Y. Guo, and J. Zhang, "Application and prediction model of CBM content based on logging parameters," *Coal Science and Technology*, vol. 42, no. 6, pp. 25–30, 2014.
- [14] X. Shao, Y. Sun, J. Sun et al., "Logging interpretation of coal petrologic parameters: a case study of Hancheng mining area," *Petroleum Exploration and Development*, vol. 40, no. 5, pp. 559–565, 2013.
- [15] Z. Jin, H. Xue, H. Gao, X. Zhao, and G. Bai, "Technology for evaluation of CBM reservoir logging and its application," *Coal Geology & Exploration*, vol. 41, no. 2, pp. 42–45, 2013.
- [16] X. Fu, Y. Qin, G. G. X. Wang, and V. Rudolph, "Evaluation of gas content of coalbed methane reservoirs with the aid of geophysical logging technology," *Fuel*, vol. 88, no. 11, pp. 2269–2277, 2009.
- [17] J. Hou and Y. Wang, "Interpretation of well logging data for CBM using BP neural network," *Fuel*, vol. 35, no. 3, pp. 41–45, 1999.
- [18] H. Pan and Z. Huang, "Discussion on the interpretation method of CBM content," *Coal Geology and Exploration*, vol. 26, no. 2, pp. 58–60, 1998.
- [19] C. Lian, Y. Zhao, H. Li, F. Qu, F. Cai, and J. Zhang, "Prediction of coalbed gas content based on support vector machine regression," *Journal of Xi'an University of Science and Technology*, vol. 28, no. 4, pp. 707–710, 2008.
- [20] J. Guo, Z. Zhang, and C. Zhang, "Prediction and application of CBM content based on gray system and logging method," *Geophysical and Geochemical Exploration*, vol. 44, no. 5, pp. 1190–1200, 2020.
- [21] J. Guo, Z. Zhang, and C. Zhang, "The exploration of predicting CBM content by geophysical logging data: a case study based on slope correlation random forest method," *Geophysical and Geochemical Exploration*, vol. 45, no. 1, pp. 18–28, 2021.
- [22] M. Xiang, X. Qi, F. Zhang, and R. An, "Application of deep learning to CBM logging interpretation," *Geology and Exploration*, vol. 56, no. 6, pp. 1305–1312, 2020.
- [23] H. Yang, H. Pan, H. Ma, A. A. Konaté, J. Yao, and B. Guo, "Performance of the synergetic wavelet transform and modified K-means clustering in lithology classification using nuclear log," *Journal of Petroleum Science and Engineering*, vol. 144, pp. 1–9, 2016.
- [24] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," *Journal of Petroleum Science and Engineering*, vol. 174, pp. 776–789, 2019.