

# Research on prediction of coalbed methane production based on Radial Basis Function Network

Hongxing Huang<sup>1,2</sup>, Ming Huang<sup>3</sup>, Dong Cheng<sup>1,2</sup>, Bing Sun<sup>2</sup>, Yi Zhang<sup>1,2</sup>

<sup>1</sup>China United Coalbed Methane National Engineering Research Center Co., Ltd.beijing,100000, China

<sup>2</sup>PetroChina Coalbed Methane Co., Ltd.beijing,100000, China

<sup>3</sup>School of Science, Southwest Petroleum University, Chengdu, Sichuan, 610500, China

**Abstract.** My country's coal seam permeability is generally low, and it is difficult to carry out large-scale development and utilization. For different coal seam blocks, the use of abundant field data to predict gas production can not only provide effective guidance for on-site construction, but also significantly save development costs. This paper presents a prediction model of coalbed methane production based on radial basis function network. According to the field data of X area, the correlation analysis of the factors affecting the gas production is carried out, and the main control factors with significant correlation are selected. Then, taking these main control factors as input and gas production as output, a series of radial basis function networks with different precisions were constructed through trial calculation of different expansion coefficients. Finally, with the goal of maximizing precision, a highly accurate fitting was obtained. The optimal network with good prediction effect.

## 1 Introduction

According to the internationally accepted development criteria, the most suitable development permeability of CBM is  $(3\sim 4)\times 10^{-3}\text{um}^2$  and cannot be lower than  $1\times 10^{-3}\text{um}^2$  [1], and the coal field with the largest permeability in my country is only  $(0.54\sim 3.8)\times 10^{-3}\text{um}^2$ , which brings severe challenges to the mining of CBM in my country. Therefore, saving development costs from different directions and expanding production capacity have become key issues in CBM exploitation.

For different coalfield blocks, a large amount of on-site data is used to achieve accurate predictions of coal seam gas production and coalbed gas production. The targeted and vigorous exploitation of high-yield wells will reduce the development investment of low-yield wells, which will effectively save development costs while bringing more benefits. More economic benefits. Radial basis function is a method of interpolation in high-dimensional space. After the development of many scholars, a new method of neural network learning is derived, namely, radial basis function network (RBF). RBF network has the characteristics of small amount of calculation, high precision, flexible node, simple format, etc., and has been widely used in many fields. At present, the more popular neural network algorithm is the BP network. Chen Lianjun compared the training and learning of the two networks in the literature [2], and the result is that the RBF network prediction model has fewer learning times and faster calculations. It effectively avoids the tedious calculation of BP network.

This paper proposes a method for forecasting coalbed methane production based on RBF network. Correlation analysis of 18 factors influencing coalbed methane production was carried out, and 4 main control factors with significant correlations were screened out. Then, with the average daily gas production of a single well as the goal, the RBF network was built and predicted using the selected main control factors, which provided guidance for the targeted development of coalbed methane on-site.

## 2 Correlation analysis

Before establishing and predicting the gas production model, it is necessary to select the appropriate main control factors. The selection of more variables will lead to a decrease in the prediction accuracy, and when the significant variables are not included in the model, the prediction accuracy will also be affected.

Correlation analysis is the process of describing the closeness of the relationship between objective things and expressing it with appropriate statistical indicators [3]. By calculating the correlation coefficients between different variables and gas production and sorting the results, we can understand which variables have the greatest impact on gas production and which variables have nothing to do with gas production.

Single correlation analysis is to analyze the linear correlation degree of two variables. The statistical measure it uses is single correlation coefficient, or correlation coefficient for short. The overall correlation coefficient is a statistic that expresses the degree of linear influence between two variables. The expression is a

\*Corresponding author's e-mail: [huangming962477@163.com](mailto:huangming962477@163.com)

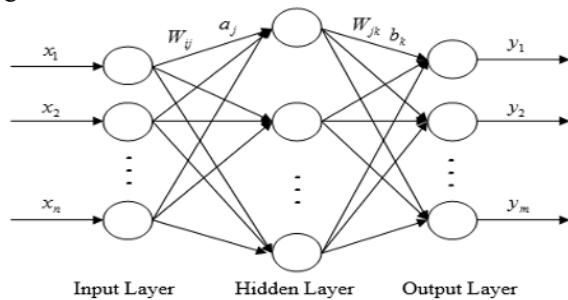
constant, and its definition is as follows:

$$\gamma = \frac{Cov(x, y)}{\sqrt{Var(X)Var(Y)}} \quad (1)$$

Where  $Var(x)$  and  $Var(y)$  are the variances of the parameters  $X$  and  $Y$ , and  $Cov(x, y)$  is the covariance of the parameters  $X$  and  $Y$ .

### 3 Radial Basis Function Network

The radial basis function network is composed of three layers of results. The first layer is the input layer and is composed of signal source nodes. The number of neurons depends on the number of main control factors. The second layer is the hidden layer, and its neurons can be adjusted adaptively during the RBF network learning process until the target error requirement is reached. The third time is the output layer, which responds to the effect of the input mode, and its transfer function is a linear function. The prediction network structure is shown in Fig.1 below:



**Fig1.** RBF network prediction model of coalbed methane production

The network mainly learns through two stages. The first stage is the training and learning from the input layer to the hidden layer. In order to determine the network center vector and radius, a clustering learning method without teacher guidance is used; the second stage is the weight adjustment between the hidden layer and the output layer. A teacher-guided recursive least square method is used to determine the weight vector [4].

For the network input  $X=(x_1, x_2, x_3, x_4)$  the basis function is selected as the Gaussian function. The radial basis function indicates the nonlinear mapping relationship from the input layer to the hidden layer. The formula is:

$$\varphi_j(x) = \exp\left(-\frac{\|X - X_j^p\|^2}{2\delta_j^2}\right) \quad (2)$$

Where:  $X_j^p$  is the center point of the  $j$ -th basis function, which can be determined by clustering analysis based on

the input sample, and  $\delta_j$  is the width of the basis function.

The neurons in the output layer of the network are the weighted summation of the outputs of the hidden layer neurons. Since the activation function is a pure linear function, the output is:

$$F^*(x) = \sum_{j=1}^N \varphi_j(x) \times w_j \quad (3)$$

The weight adjustment is:

$$w_j(L+1) = w_j(L) + \beta(F^{*d} - F^*(L)) * \alpha_j(x) / [\alpha^T(x)\alpha(x)] \quad (4)$$

Where:  $F^{*d}$  is the expected value of the dependent variable,  $F^*(L)$  is the output value of the  $L$ -th calculation of the dependent variable,  $\alpha$  is the learning rate, and  $\alpha(x)$  is the hidden layer basis function mapping vector.

In order to improve the promotion ability of the network and prevent the phenomenon of over-fitting, according to the research results of literature [5] and literature [6], the criterion that characterizes the promotion ability is added to the prediction model:

$$\Delta F \geq \frac{n\lambda}{\Delta w * 2 \log\left(1 + \frac{F}{E}\right)} \quad (5)$$

Where:  $\Delta w$  is the difference between the arithmetic mean of the link weights of the input nodes when the sample set is trained twice in the training process, and  $\Delta F$  is the average of the new sample identification output results in response to the weights at the completion of a certain training Root error,  $n$  is the number of input nodes,  $\lambda$  is taken as  $1 \times 10^{-4}$ ,  $F$  is the mean value of the sample output after normalization, and is  $E$  the variance of the sample output after normalization [7].

### 4 RBF network gas production and forecast

18 parameters summarized in the field data of coalbed methane in the X area: Total Liquid Volume, Production & Drainage Interval, Dynamic Fluid Level, Displacement, Casing Pressure, Amount of Sand Added, CAL1, DEN, DT, CNL, GR, SP, Preflush Rate, Bottomhole Pressure, RD, Discharge Strength, RS, MSFL, first carry out the standardization of the data, and standardize all the data to the 0-1 interval through dispersion standardization. Then, according to the 18 factors that affect the production of coalbed methane, the four main control factors with the largest correlation coefficients are selected through the method of correlation analysis. The correlation analysis results of all factors are as follows:

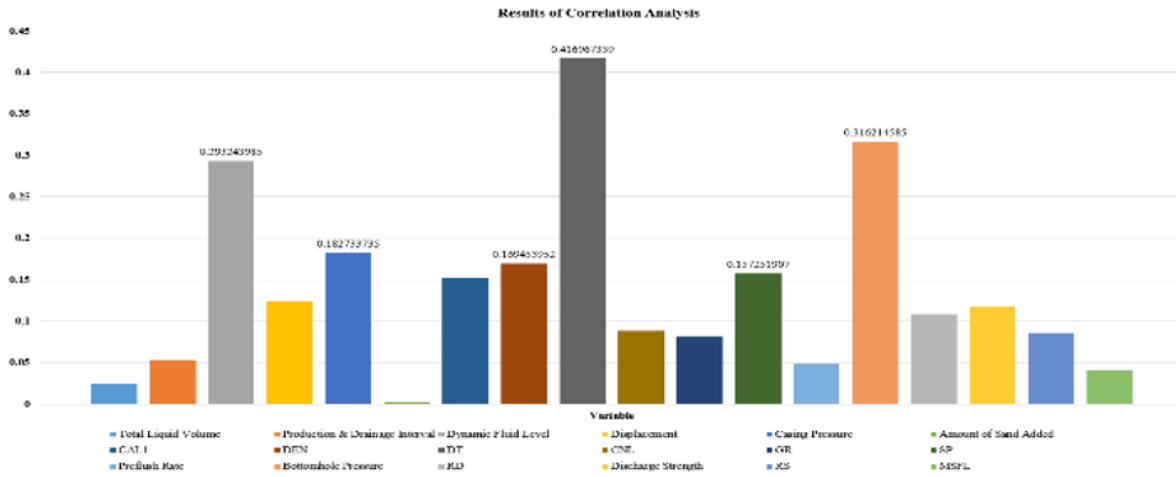


Fig2. Results of correlation analysis of all variables

It can be seen from Fig.2 that the most significant variables related to CBM production obtained through correlation analysis are DT, Bottomhole Pressure and Casing Pressure. The more relevant ones are Casing Pressure, DEN, SP, CAL1, Displacement, Discharge Strength. The correlation between the remaining 8 variables and gas production is at a low level. Since too many variables are selected, the prediction error of the model will increase, and too few selections will also affect the prediction accuracy of the model. Therefore, the top 4 displacement, CAL1, DT and bottom hole pressure are the four parameters as the main control factors. The following uses these 4 main control factors as input and gas production as output to build a radial basis function network model<sup>[8]</sup>.

First, 73 oil and gas well data are divided into training data set and test data set. In this paper, 10 data are randomly selected as the test set, and the other 63 data samples are used as the training set. Use the data samples of the training set to build a radial basis function network, adjust the expansion speed factor of the network, start from 0.1 with a step size of 0.1 to end at 50, and get 500 different radial basis function networks and their corresponding fitting accuracy. It is generally believed that the prediction can be made only when the fitting accuracy is higher than 90%, and unless the value of the expansion speed factor is extremely abnormal (too large), the fitting accuracy of the radial basis function network is generally 100% or infinitely close 100% is obviously better than 90%, so all 500 networks are suitable for direct prediction.

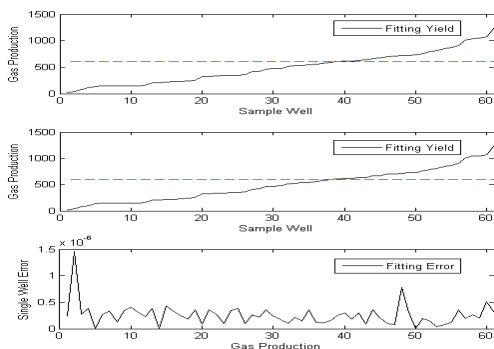


Fig3. Optimal Radial Basis Function network fitting result graph

Before forecasting, we first classify CBM wells into two types: high-yield wells and non-high-yield wells. If the production is in the [0,600) interval, it is regarded as a non-high-yield well; if the output is above 600, it is regarded as a high-yield well. Then the 12 wells under the prediction data set are substituted into the above radial basis function network model. When the predicted production is in the same interval as the real production of oil and gas wells, that is to say, the classification result is the same as the real category, we think the prediction is successful. Finally, with the goal of maximizing classification prediction accuracy, an optimal radial basis function network is obtained. The fitting results and classification prediction results of the network are as follows:

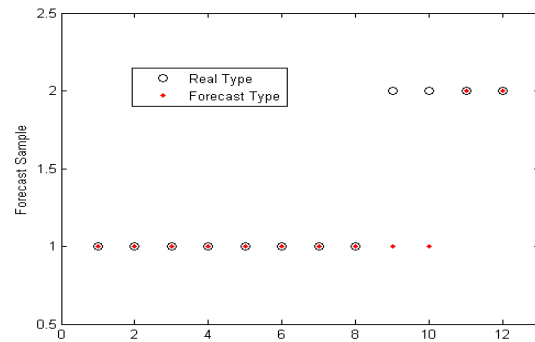


Fig4. Optimal radial basis function network prediction result graph

It can be seen from Fig.3 that the final selected radial basis function network has extremely high fitting accuracy, and the single well fitting error is below  $1.5 \times 10^{-6}$ . It can be seen from Fig.4 that for the 8 low-non-high-yield wells and 4 high-yield wells in the test set, only two wells failed in type prediction, with an accuracy rate of 83.33% and good prediction results.

## 5 Conclusion

Based on the results and discussions presented above, the conclusions are obtained as below:

- (1) The field data of coalbed methane has a large

dimension and contains many factors with little influence. In this paper, the real data is reduced in dimensionality, and the correlation coefficient method is used to retain the four most significant factors related to the production of coalbed methane, which ensures the accuracy of prediction while reducing the dimensionality of the factors.

(2) The radial basis function network has an excellent fitting effect. The method in this paper is used to classify and predict the production of 12 CBM wells, and the prediction accuracy rate reaches 83.33%.

## Acknowledgement

This paper was supported financially by the National Science and Technology Major Project of China (No. 2016ZX05043002).

## References

1. Guoxin Chen.(1995) Coal mine gas drainage technology and analysis[J]. Safety in Coal Mines (05):2-6(in Chinese).
2. Lianjun Chen, Caiquan Su. (2009) Radial basis function network prediction model of mine gas emission[J].Journal of Shandong University of Science and Technology(Natural Science), 28(04):93-96+101(in Chinese).
3. Xiaoou Liu, Shaoyun Ge. (2020) The definition of energy efficiency of regional integrated energy system and its correlation analysis[J]. Automation of Electric power Systems, 44(8):8(in Chinese).
4. Fangfang Wang. (2014)Using Radial Basis Function Neural Network to Predict Fluid Accumulation in Gas Well[D]. Southwest Petroleum University (in Chinese).
5. Guiren Liu, Baoming Ren.(2001) Rotating mining method for fully mechanized mining face in gently inclined coal seam[J]. Journal of Xi'an University of Science and Technology, 021(004):319-321(in Chinese).
6. Huaili Zhang,Pingchuan Wang.(2003)Rotary mining technology of 7715 fully mechanized caving face in Xingtai Mine[J].Hebei Coal:20-21(in Chinese).
7. Yongming Han, Chenyu Fan, Zhiqiang Geng, et al. (2020) Energy efficient building envelope using novel RBF neural network integrated affinity propagation. 209
8. Surabhi Chandra, Prerna Gaur, Diwaker Pathak. (2020)Radial basis function neural network Based Maximum power point tracking for photovoltaic brushless DC motor connected water pumping System.86