

# Research on Oil Well Production Prediction Based on Radial Basis Function Network

Chuanxin Zhang<sup>1</sup>, Yunwei Kang<sup>2\*</sup>, Jin Chen<sup>1</sup>, Yunxiang Zhao<sup>2</sup>, Junxiu Ma<sup>1</sup>

<sup>1</sup> Engineering Technology Research Institute, Xinjiang Oilfield Company, 834000, China

<sup>2</sup> School of Sciences, Southwest Petroleum University, Chengdu, Sichuan, 610500, China

**Abstract:** Selection of well and reservoir is an important step in the process of stimulation and transformation of oil fields. Good measures can effectively save the cost in the process of oil field development and greatly increase the production of oil fields. Aiming at the problem of well and reservoir selection in petroleum engineering, a method of oil well production prediction based on radial basis function network is proposed in this paper. According to the field data of Xinjiang oilfield, the main controlling factors with greater influence are selected by correlation analysis after data pretreatment. Then we randomly divide the data into training data set and prediction data set, and use the training data set to create a radial basis function network. Finally, we use the radial basis function network to predict the prediction data set, and the final prediction accuracy reaches 80%.

## 1 Introduction

In the process of exploitation and development of oil fields, due to the immaturity of previous science and technology [1], the exploitation of subsidiary resources such as oil often consumes a lot of human and financial resources. Sometimes even with the risk of unsatisfactory production of oil wells. And most oil fields in China [2] have the characteristics of high adsorption, low permeability and low pressure, which makes oil wells have a low production. Therefore, in order to improve the energy production of oil wells and reduce the waste of energy and human and financial resources, it is urgent to carry out stimulation and transformation measures for oil wells. Selection of well and reservoir centered on production prediction will play an important role in it.

At present, some methods, such as hydraulic fracturing and gas injection displacement, are the main stimulation measures at home and abroad [3]. Hydraulic fracturing technology is one of the most important and widely used means in this process. Therefore, how to select wells and layers will greatly affect the production of oil wells. Radial Basis Function (RBF), as a method of interpolation in high dimensional space, was proposed by Lowe and Broomhead and developed continuously. Later, some scholars put forward a new learning method of neural network, namely radial basis function network. Although the development history of radial basis function network is very short, it has the characteristics of small computation, high accuracy, flexible nodes, simple format and so on [4]. It plays an important role in many aspects and has been successfully applied in many fields, such as function approximation, signal processing,

pattern recognition and so on. Many formerly seemingly intractable and complex problems may be solved easily by using radial basis function networks.

In this paper, the radial basis function network is applied to oil well production prediction, which significantly improves the fitting effect of field data and provides a basis for well and reservoir selection. Firstly, we optimize the field data of oil wells, and then establish a radial basis function network model, which is applied to predict oil well production, providing some guidance for fracturing and stimulation in engineering.

## 2 Material and Methods

### 2.1 Gray Correlation Degree Analysis

Grey relational degree analysis is a method of grey system analysis and processing of random variables [5], as well as a mapping of data to data. Firstly, we quantify the qualitative data, supplement the incomplete data and normalize the range. Let the data after pretreatment of the selected parameters be  $x_i$ , and the correlation between data  $x_j$  and target data  $x_i$  (stable daily gas production after fracturing) is  $\zeta_{ij}(k)$ . Among them,  $k$  represent the sampling points of the correlation between data  $x_j$  and data  $x_i$ . The total number of sampling points is  $n$ , then the correlation coefficient can be expressed as follows:

\*Corresponding Author: Yunwei Kang; email: kangyunweiSWPU@163.com; phone:15281856277

$$\zeta_{ij}(k) = \frac{\alpha_{\min} + \alpha_{\max}\rho}{\alpha_{ij}(k) + \alpha_{\max}\rho} \quad (k = 1, 2, \dots, n) \quad (1)$$

In the formula

$$\alpha_{ij}(k) = |x_i(k) - x_j(k)|, \alpha_{\min} = \min_j \min_k \alpha_{ij}(k), \alpha_{\max} = \max_j \max_k \alpha_{ij}(k)$$

$\rho$  is a given constant between 0 and 1, which is usually taken as 0.5 in application.

Thus, the relevance degree  $\gamma_{ij}$  is:

$$\gamma_{ij} = \frac{1}{n} \sum_{k=1}^n \zeta_{ij}(k) \quad (2)$$

Relevance degree  $\gamma_{ij}$  reflects the degree of correlation between data  $x_j$  and data  $x_i$ , and describes the relative changes between data  $x_j$  and data  $x_i$  during the development of the system. By using the ranking order of correlation degree, the influence degree of each parameter on CBM production can be grasped, which lays a foundation for production prediction.

## 2.2 Radial Basis Function Network

### 2.2.1 Radial Basis Function Network Model

The center of RBF neurons [6] is randomly selected in the input sample. Once the center is selected, it is fixed and the output of RBF neurons is determined. The farther the input of the RBF neuron is from the center of the RBF neuron, the more the output of the neuron tends to be zero, that is to say, the lower the activation of the neuron.

If a small number of input vectors in some local areas of the input space will affect the output of the whole neural network, then the network is called a local approximation network. RBF is a kind of local approximation network.

The output expression of this neuron is:

$$a = f(\|W - p\| \cdot b) = \text{radbas}(\|W - p\| \cdot b) \quad (3)$$

Among them,  $\text{radbas}()$  represents the radial basis function. In general, we use the Gaussian function as the radial basis function. The expression of the Gaussian function is as follows:

$$\text{radbas}(n) = e^{-n^2} \quad (4)$$

Because the Gaussian function is very smooth, the form is very simple, and more importantly, it is about radial symmetry, so we often use it.  $\|W - p\|$  is the Euclidean distance, and its expression is as follows:

$$\|W - p\| = \sqrt{\sum_{i=1}^R (w_{1,i} - p_i)^2} = [(W - p^T)(W - p^T)^T]^{-\frac{1}{2}} \quad (5)$$

### 2.2.2 Learning algorithm of radial basis function network

There are three main parameters in the training and learning algorithm of radial basis function network [7], which are the weight between the radial base and the output layer, the center of the basis function and the variance. Self-organizing center selection method is an important learning algorithm for radial basis function networks. Self-organizing center selection method can train the center of radial basis function network model and the weight of radial base to output layer. It can be divided into two independent steps, namely organizational learning stage and supervisory learning stage. The main content of organizational learning stage is training and learning the variance and center of RBF. The main contents of the supervision stage are training and learning the weights of the output layer. The main process of self-organizing center selection method is as follows:

#### (1) Learning Center $t$

K-means clustering is needed in the process of self-organizing center selection. The main steps of using K-means clustering algorithm are as follows:

Step 1: Initialize the cluster center. Firstly,  $I$  different input vectors are randomly selected from the input training data set and used  $t_i(0) (i = 1, 2, \dots, I)$  as the center of the radial basis function neural network at the beginning. At this point, the iteration step is set to  $n = 0$ .

Step 2: Input the sample  $X_k$  in the training data set randomly.

Step 3: Find the nearest center from all the centers to training sample  $X_k$ . That is to say, to find out  $i(X_k)$  which satisfy the following conditions:

$$i(X_k) = \arg \min_i \|X_k - t_i(n)\| \quad (i = 1, 2, \dots, I) \quad (6)$$

In the upper form,  $t_i(n)$  represents the  $k$ -th center of the radial basis function network in the  $n$ -th iteration process.

Step 4: Use the following formula to adjust the center of the radial basis function network:

$$t_i(n+1) = \begin{cases} t_i(n) + \eta[X_k(n) - t_i(n)], & i = i(X_k) \\ t_i(n), & \text{other circumstances} \end{cases} \quad (7)$$

Among them,  $\eta$  represents the learning step, and  $0 < \eta < 1$ .

Step 5: Determine whether all the training sets have been trained and the distribution of the center of the radial basis function network will not change. If not, let  $n = n + 1$ , and then jump to step 2. Otherwise, the final  $t_i(n)$  is the radial basis function network's center  $t_i (i = 1, 2, \dots, I)$ .

#### (2) Calculating variance $\sigma$

After the center of the radial basis function network is determined, we begin to calculate the variance of the radial basis function. Usually we choose the Gauss function:

$$G(\|X_k - t_i\|) = \exp\left(-\frac{1}{2\sigma^2}\|X_k - t_i\|^2\right) (i = 1, 2, \dots, I) \quad (8)$$

In this case, the variance of the radial basis function can be calculated by using the following formula:

$$\sigma_i = \frac{d_{\max}}{\sqrt{2I}}, i = 1, 2, \dots, I \quad (9)$$

Among them,  $d_{\max}$  represents the maximum distance between the centers selected in the previous step;  $I$  represents the number of units in the radial base.

(3) Training and learning weights  $w_{ij}$

The training and learning of weights from hidden layer to output layer of RBF network can be realized by pseudo-inverse method:

$$W = G^+ D \quad (10)$$

Among them,  $D = [d_1, \dots, d_k, \dots, d_N]^T$  responds to expectations;  $G^+$  represents the pseudo-inverse matrix of matrix  $G$ , which is calculated in the following way:

$$G^+ = (G^T G)^{-1} G^T \quad (11)$$

The calculation of matrix  $G$  is as follows:

$$g_{ki} = \exp\left(-\frac{1}{d_{\max}^2}\|X_k - t_i\|^2\right) (k = 1, 2, \dots, N; i = 1, 2, \dots, I) \quad (12)$$

Therefore, the calculated weight matrix  $W$  is as follows:

$$W = \{w_{ij}\} (i = 1, 2, \dots, I; j = 1, 2, \dots, J) \quad (13)$$

It is noteworthy that the final output of the output layer of the RBF network should be the weighted sum of the output data of the RBF network, so the real output  $Y(n)$  of the RBF network should be:

$$\{y_{kj}(n)\} = G(n)W(n) (k = 1, 2, \dots, N; j = 1, 2, \dots, J) \quad (14)$$

### 3 Results

According to the field data of Xinjiang oilfield, we first pre-process the data, then analyze the grey correlation degree of 24 factors affecting the production of oil wells, and calculate the correlation degree between the parameters and the average daily oil production. The concrete results are shown in the following table:

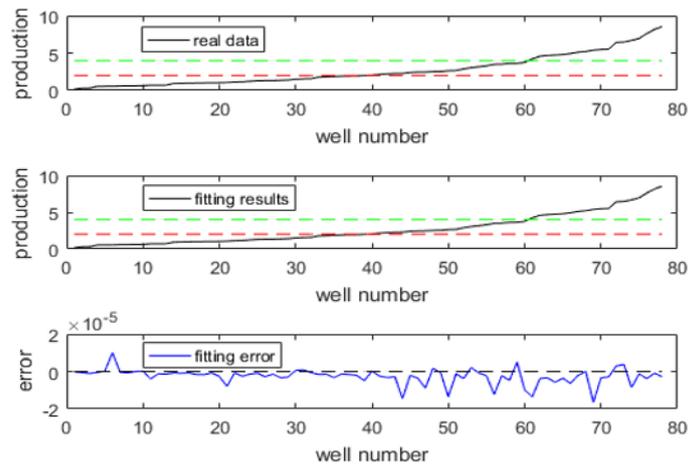
**Table 1.** Sorting results of grey relational degree

Rank	Parameter	Relevance Degree	Rank	Parameter	Relevance Degree
1	Oil Saturation	0.89871046	13	Number of Fracturing Intervals	0.7209038
2	Formation Pressure	0.890737075	14	RT	0.7132321
3	Reservoir Thickness	0.879845103	15	Maximum Displacement	0.7078482
4	Total liquid volume	0.877730578	16	Percentage of Preflux	0.6996847
5	Fracture Pressure	0.865928037	17	Static Filtration Loss Coefficient	0.6895624
6	Sand Allocation Relation	0.752540494	18	RI	0.6598642
7	Maximum Sand Content	0.745205188	19	Displacement Allocation Relation	0.6541495
8	Average Sand Content	0.744784008	20	Average Sand Ratio	0.6264023
9	Formation Permeability	0.734701519	21	Casing Pressure	0.6128984
10	Average Displacement	0.730000873	22	Pressure Gradient	0.6033414
11	Half Seam Length	0.726563294	23	Reservoir Depth	0.6008376
12	Formation Porosity	0.724954158	24	Formation Temperature	0.5545693

## 4 Discussion

As can be seen from the table above, the correlation degree ranked after the fifth is at a low level, so the first five factors are taken as the main control factors of oil well production. Firstly, we divide oil wells into three types: low production, medium production and high production. The production of low production wells is in

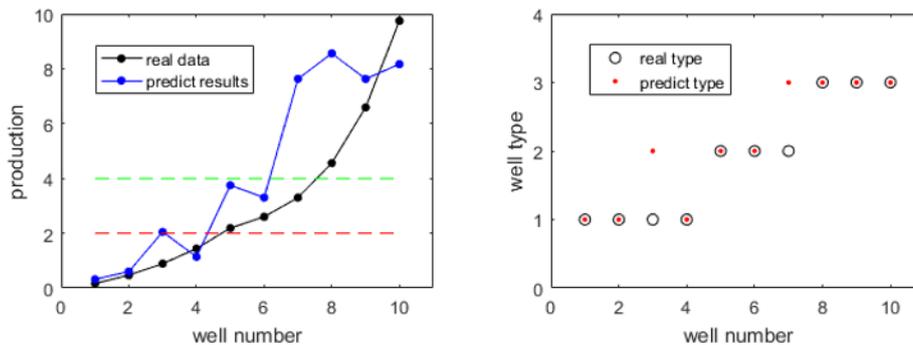
[0,2]; the production of medium production wells is in [2,4]; and the production of high production wells is greater than or equal to 4t/d. Then we randomly selected 78 data from 88 oil wells as training set and the remaining 10 data as prediction set. Firstly, the 78 data are used as training data of radial basis function network and fitted immediately. The fitting results are as follows:



**Figure 1.** Fitting results of training data set

The average relative error is used as the basis to measure the fitting effect, and the fitting accuracy of the model is 97.5%. The fitting effect is good, which proves that the model can be used in the next stage of prediction.

Ten wells from the prediction data set are substituted into the above radial basis function network model. The final prediction and classification results are as follows:



**Figure 2.** Radial Basis Function Network Prediction Result Diagram

As can be seen from the graph, 8 of the 10 wells have been correctly classified and only the third and the ninth wells have failed to predict. According to the classification conformity, the final prediction accuracy is 80%. The prediction effect is good and can meet the actual needs.

## 5 Conclusions

(1) Radial Basis Function (RBF) network model is an interpolation method in high- dimensional space and has great advantages. In this paper, this method is applied to well selection and reservoir selection to predict oil well production, and finally the feasibility of this method is verified.

(2) RBF Network has a strong fitting effect. The fitting accuracy of this paper is still 97.5% for the complicated field data. Using this method to predict and classify the production of oil wells, the final prediction accuracy is 80%.

## Acknowledgement

This paper was funded by the National Science and technology Major Project of China (No. 2017ZX05070).

## References

1. Figueroa, H. G., Lagreca, A., Gahan, B. C., Parker,

- R. A., Graves, R. M., & Batarseh, S., et al. (2002). Rock removal using high-power lasers for petroleum exploitation purposes. *Proc Spie*, 4760, 678-690.
2. Xingru, Z., & Zhenlie, W. (1995). Environmental protection during the oilfield exploitation in china. *China Oil & Gas* (2), 42-44.
3. Desroches, J., Detournay, E., Lenoach, B., Papanastasiou, P., Pearson, J. R. A., & Thiercelin, M., et al. (1994). The crack tip region in hydraulic fracturing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 447(1929), 39-48.
4. Gomm, J. B., & Yu, D. L. (2000). Selecting radial basis function network centers with recursive orthogonal least squares training. *IEEE Transactions on Neural Networks*, 11(2), 306-314.
5. Deng, A. (2012). The application of grey relevance degree in the structure analysis of grain production. *Henan Science*.
6. Ahmed, F., Gustafson, S. C., & Karim, M. A. (1996). Image interpolation with adaptive receptive field-based gaussian radial basis functions. *Microwave and Optical Technology Letters*, 13(4), 197-202.
7. Schwenker, F., Kestler, H. A., & Palm, G. (2001). Three learning phases for radial-basis-function networks. *Neural Networks*, 14.