

# Comparisons of Oil Production Predicting Models

Yishen Chen<sup>1,2</sup>, Xianfeng Ding<sup>3</sup>, Haohan Liu<sup>1,4\*</sup>, Yongqin Yan<sup>5</sup>

<sup>1</sup>School of Graduate of Southwest Petroleum University, Chengdu, China

<sup>2</sup>China Petroleum Engineering Construction Corporation, Beijing, China

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

<sup>4</sup>Sichuan College of Architectural Technology, Deyang, China

<sup>5</sup>School of Petroleum Engineering Management of Southwest Petroleum University, Chengdu, China

Email: \*[tsinghua616@163.com](mailto:tsinghua616@163.com)

Received June 11, 2013; revised July 11, 2013; accepted July 18, 2013

Copyright © 2013 Yishen Chen *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.

## ABSTRACT

Feasibility of oil production predicting results influences the annual planning and long-term field development plan of oil field, so the selection of predicting models plays a core role. In this paper, three different predicting models are introduced, they are neural network model, support vector machine model and GM (1, 1) model. By using these three different models to predict the oil production in XINJIANG oilfield in China, advantages and disadvantages of these models have been discussed. The predicting results show: the fitting accuracy by the neural network model or by the support vector machine model is higher than GM (1, 1) model, the prediction error is smaller than 10%, so neural network model and support vector machine model can be used to short-term forecast of oil production; predicting accuracy by GM (1, 1) model is not good, but the curve trend with GM (1, 1) model is consistent with the downward trend in oil production, so GM (1, 1) predicting model can be used to long-term prediction of oil production.

**Keywords:** Oil Field; Oil Production; Model; Predicting Accuracy

## 1. Introduction

Oil production prediction is very important in the oilfield development; study on oil production predicting method is a key topic of petroleum science. At present, there are four kinds of predicting methods [1-6] with physical meanings: Empirical formula such as Arps method, Hubbert model, water-flood decline curve method; Hydrodynamic model based on fluid mechanics model; Material balance equation model; Numerical reservoir simulation model. Besides the above mentioned four methods, there is a typical type of prediction model related to modern optimization, this model type is composed of GM (1, 1) model, neural network model, support vector machine model, etc. Oil field development system is a complex multi variables non-linear dynamical systems, different predicting model has different characteristics like predicting accuracy. Neural network model and support vector machine model are two effective methods to solve multi nonlinear mapping problem. At present they are used in many disciplines, even in the oil production pre-

diction. In this paper, neural network model and support vector machine model have been used to predict the oil production of XINJIANG oil field, and good predicting results have been achieved. Meanwhile, GM (1, 1) model has also been used to predict oil production. Predicting result shows that this model is adaptable to the case which lacks data and hard to establishes model with probabilistic method.

## 2. Predicting Model

### 2.1. Neural Network Predicting Model

At present, in the application of artificial neural networks (ANN) [7,8], most of them are back propagation (BP) ANN and their variations. It has been proved that BP neural network can approximate any multivariate continuous function. Kolmogorov rule guaranteed that any continuous function or mapping can be achieved by a 3-layer ANN.

3-layer BP ANN is used to establish the ANN model with prediction function. The first layer of BP-ANN is input layer, the second layer of it is middle layer, and the third layer of it is the output layer, see **Figure 1**.

\*Corresponding author.

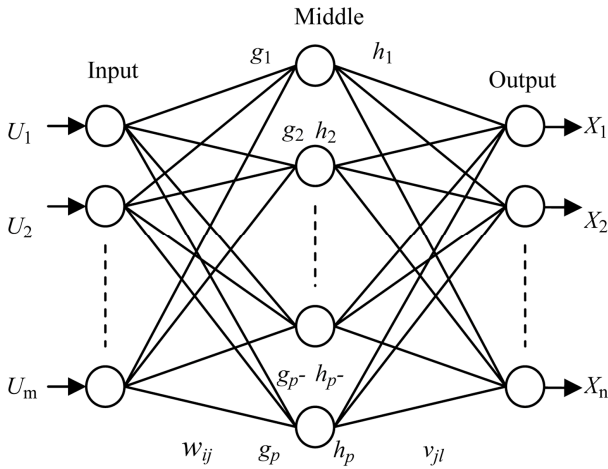


Figure 1. BP neural network structure.

**Network Simulation**

Assume  $N$ -samples (input-output data samples) are used during the training process:

The input variables are

$U(t) = (u_1(t), u_2(t), \dots, u_m(t))^T$ , the output variables are

$X(t) = (x_1(t), x_2(t), \dots, x_n(t))^T$ . For the  $k$ -sample

( $k = 1, 2, \dots, N$ ), let.  $U^k = (u_1^k, u_2^k, \dots, u_m^k)$  be input

mode vector;  $\bar{X}^k = (\bar{x}_1^k, \bar{x}_2^k, \dots, \bar{x}_n^k)$  be the expectation

output vector;  $G^k = (g_1^k, g_2^k, \dots, g_p^k)$  be the middle layer

input vector;  $H^k = (h_1^k, h_2^k, \dots, h_p^k)$  be the output unit

vector of middle layer;  $M^k = (m_1^k, m_2^k, \dots, m_n^k)$  be the

input vector of output layer;  $X^k = (x_1^k, x_2^k, \dots, x_n^k)$  be

the output vector of output layer;

$\{W_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, p\}$  be the connection weights from input layer to middle layer;

$\{V_{jl}, j = 1, 2, \dots, p; l = 1, 2, \dots, n\}$  be the connection weights from middle layer to output layer;

$\{\theta_j, j = 1, 2, \dots, p\}$  be the threshold value of middle layer;

$\{\gamma_l, l = 1, 2, \dots, n\}$  be the threshold value of output layer;

$\eta (0 < \eta < 1)$  is the learning rate.

Let the response function of ANN  $f(x)$  be Sigmoid-type function:  $f(x) = (1 + e^{-x})^{-1}$

The input and output values of each neural unit satisfy the following relationship:

Middle layer:

$$\begin{cases} \text{input: } g_j^k = \sum_{i=1}^m w_{ij} u_i^k - \theta_j \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, p) \\ \text{output: } h_j^k = f(g_j^k) \quad (k = 1, 2, \dots, N) \end{cases} \quad (1)$$

Output layer:

$$\begin{cases} \text{input: } m_l^k = \sum_{j=1}^p v_{jl} \cdot h_j^k - \gamma_l \quad (l = 1, 2, \dots, n) \\ \text{output: } x_l^k = f(m_l^k) \end{cases} \quad (2)$$

After training of  $N$  samples, the network error is:

$$E_{\text{total}} = \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^n (\bar{x}_l^k - x_l^k)^2.$$

Error of output layer unit:

$$d_l^k = (\bar{x}_l^k - x_l^k) x_l^k (1 - x_l^k) \quad (3)$$

Error of middle layer unit:

$$e_j^k = \left[ \sum_{l=1}^n d_l^k \cdot v_{jl} \right] h_j^k (1 - h_j^k) \quad (4)$$

The connection weights  $\{V_{jl}\}$  and the threshold value  $\{\gamma_l\}$  can be modified by the output layer error  $\{d_l^k\}$  and output value of middle layer unit  $\{h_j^k\}$ :

$$V_{jl}(q+1) = V_{jl}(q) + \eta \sum_{k=1}^N d_l^k h_j^k \quad (5)$$

$$\gamma_l(q+1) = \gamma_l(q) + \eta \sum_{k=1}^N d_l^k \quad (6)$$

The connection weights  $\{W_{ij}\}$  and the threshold value  $\{\theta_j\}$  can be modified by the middle layer error  $\{e_j^k\}$  and input value of input layer  $U^k = (u_1^k, u_2^k, \dots, u_m^k)$ :

$$W_{ij}(q+1) = W_{ij}(q) + \eta \sum_{k=1}^N e_j^k u_i^k \quad (7)$$

$$\theta_j(q+1) = \theta_j(q) + \eta \sum_{k=1}^N e_j^k \quad (8)$$

Repeat the above-mentioned learning mode, until the network converges to a given error range.

**2.2. Support Vector Regression (SVR) Predicting Model**

SVR [9-11] is an effective method to solve the regression problem. This regression problem can be described in mathematics:

Given training set  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (x, y)^l$ , where  $x_i \in x = R^n, y_i \in y = R, i = 1, \dots, l$ . The training set is composed of independent and identically distributed sample points following certain probability distribution  $p(x, y)$ . After giving loss function  $c(x, y, f)$ , the regression function  $f(x) = w^T \cdot \varphi(x) + b$  will be found to make the expected risk  $R[f] = \int c(x, y, f) dp(x, y)$  reach its minimum value, where  $\varphi(x)$  is the nonlinear mapping, it maps the data  $x$  into a high dimensional feature space;  $w$  and  $b$  are weight vector and bias value, separately.

If we denote the regression problem by minimum risk problem with the loss function

$c(x, y, f(x)) = \max\{0, |y - y(x)| - \varepsilon\}$  ( $\varepsilon \geq 0$ ), then the basic rule of SVR method is established: by introducing kernel function  $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$  into the non-linear mapping, the nonlinear regression problem in low-dimensional input space transformed into a linear regression problem in high dimensional feature space and the key is to solve the parameter  $w$  and  $b$  in the regression function. Hence, the basic rule of SVR in solving regression problem is to solve an optimal problem with the following

$$\min_{\alpha^* \in R^{2l}} \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_j^*)(\alpha_j^* - \alpha_i^*)K(x_i, x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i), \tag{9}$$

form:

$$\text{s.t. } \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, \\ 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, i = 1, 2, \dots, l$$

where  $\alpha_i, \alpha_i^*$  are Lagrange multipliers;  $C$  is a constant, called penalty factor;  $\varepsilon$  is a given positive value, it is a maximum error of regression.

By solving Equation (9) gives the optimal Lagrange multipliers  $\bar{a} = (\bar{a}_1, \bar{a}_1^*, \dots, \bar{a}_l, \bar{a}_l^*)^T$ , and then gives regression function when the expected risk  $R[f]$  gets its minimum:

$$f(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x) + \bar{b} \tag{10}$$

where the sample satisfied  $\bar{\alpha}_i^* - \bar{\alpha}_i \neq 0$  is the support vector;  $\bar{w} = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x)$ .

Select  $\bar{\alpha}_j$  or  $\bar{\alpha}_k^*$  in interval  $(0, \frac{C}{l})$ ;

If  $\bar{\alpha}_j$  is selected, then

$$\bar{b} = y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x_j) + \varepsilon$$

If  $\bar{\alpha}_k^*$  is selected, then

$$\bar{b} = y_k - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i, x_k) - \varepsilon$$

when solving regression problem, the proper kernel function can be selected to SVR training. Introducing sample factor vector  $x_i$  and predictor vector  $x$  into Equation (10) after training gives the prediction result.

### 2.3. GM (1,1) Predicting Model

The Let  $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$  be initial data series, its 1-AGO data series is

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

where  $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$ , denote the grey derivative of  $x^{(1)}$  by:

$$d(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$$

Let  $z^{(1)} = (x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n))$  be the generated data series of  $x^{(1)}$ , then

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)x^{(1)}(k-1)$$

Hence, denote the gray differential equation model of GM (1, 1) [12,13] by:

$$d(k) + az^{(1)}(k) = b$$

It is,

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{11}$$

In Equation (11)  $x^{(0)}(k)$  is the grey derivative,  $a$  is the developing coefficient,  $z^{(1)}(k)$  is the albino background value,  $b$  is the grey functional variable.

Introducing  $k = 2, 3, \dots, n$  into Equation (13) gives

$$\begin{cases} x^{(0)}(2) + az^{(1)}(2) = b \\ x^{(0)}(3) + az^{(1)}(3) = b \\ \dots \\ x^{(0)}(n) + az^{(1)}(n) = b \end{cases}$$

Let

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}; u = \begin{bmatrix} a \\ b \end{bmatrix}; B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

Then GM (1, 1) model can be expressed as  $Y = Bu$ , now the question comes down to find the value of  $a, b$ .

Introducing least square method to solve the estimating value:

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y.$$

In Equation (13), if  $x^{(0)}(k)$  is continuous variables when  $k = 2, 3, \dots, n$ , then  $x^{(1)}$  is a function of time  $t$ , it is  $x^{(1)}(t)$ , so the derivative of  $x^{(0)}(k)$  become  $\frac{dx^{(1)}(t)}{dt}$  and the derivative of albino background value  $z^{(1)}(k)$  becomes  $x^{(1)}(t)$ . Hence, the grey differential equation becomes:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{12}$$

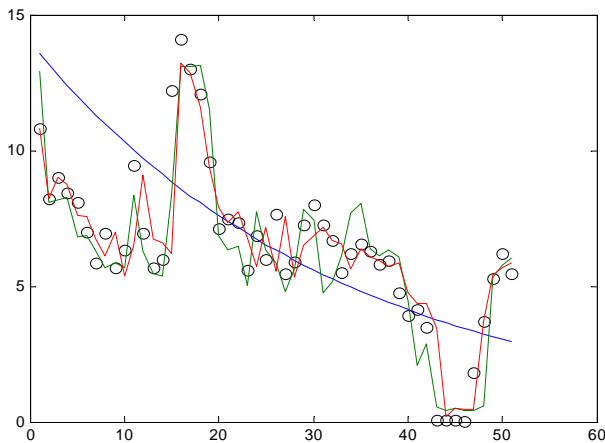
Equation (12) is the albino type of GM (1, 1) model. Given initial value  $x^{(1)}(t=1) = x^{(0)}(1)$ , the solution of Equation (12) becomes:

$$x^{(1)}(t) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} + \frac{b}{a}.$$

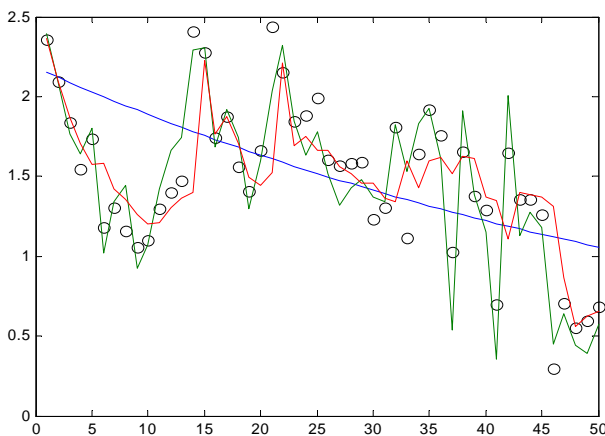
### 3. Applications and Discussions

Given the initial oil production data (from 1958 to 2012) of certain oilfield block in China, then the above-mentioned three method can be used to predict the oil production of different oilfield block (A1, A2, A3). After using the above-mentioned three different predicting model gives **Figures 2-4**.

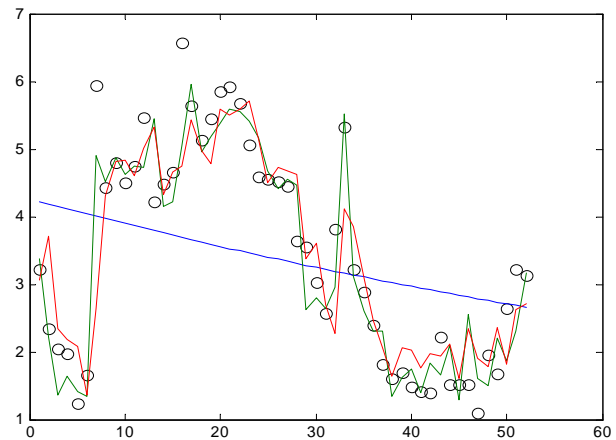
In **Figures 2-4**, the black circle means the real oil production, the green curve is the predicting curve with ANN model, the red curve is the predicting curve with SVR model, the blue curve is the predicting curve with GM (1, 1) model. **Figures 2-4** show the predicting accuracy with ANN model and SVR model is higher than



**Figure 2.** comparison of oil production prediction value with different method in block-A1.



**Figure 3.** comparison of oil production prediction value with different method in block-A2.



**Figure 3.** comparison of oil production prediction value with different method in block-A3.

the GM (1, 1) model, the maximum relative error is less than 10%, ANN model and SVR model can be used to short term prediction. However, the GM (1, 1) model predicts the overall trend in oil production decline; it can be used to middle-long term oil production prediction.

### 4. Conclusions

Prediction with ANN model and SVR model can comply with the actual oilfield production dynamics, the prediction errors of them are less than 10%. However, they are learning type of model; much data is needed to complete the prediction, so they are only suitable for the short term prediction;

Although the prediction accuracy with GM (1, 1) model is low, still the prediction result fits with the overall downward trend of oil production, so it can be used as a reference for long-term prediction.

### REFERENCES

- [1] B. Y. Ji, "Overview of Oilfield Development Indicators Forecasting Methods," *Petroleum Geology and Oilfield Development in Daqing*, Vol. 18, No. 2, 1999, pp. 19-22.
- [2] J. G. Hu, Y. Q. Chen and S. Z. Zhang, "A New Model Using for Prediction of Oil-Gas Field Production," *Acta Petrolei Sinica*, Vol. 16, No. 1, 1995, pp. 79-86.
- [3] G. Chen and X. L. Yang, "An Examination into Prediction Methods Based on Reservoir Performance Analyses," *China Offshore Oil and Gas (Geology)*, Vol. 16, No. 4, 2002, pp. 254-259.
- [4] Y. Q. Chen and Q. F. Zhao, "A Combined Method of Weibull Model and Water Drive Curve of B Type," *Xinjiang Petroleum Geology*, Vol. 21, No. 5, 2000, pp. 405-407.
- [5] Y. Q. Chen, J. G. Hu and D. J. Zhang, "Deduction of Logistic Model and Auto Regression Methods," *Xinjiang Petroleum Geology*, Vol. 17, No. 2, 1996, pp. 150-155.
- [6] P. Yue, X. F. Chen, L. G. Cui, et al., "Suggestion Solu-

- tion of Weng's Model and Logistic Model in Production Research," *Journal of Oil and Gas Technology*, Vol. 31, No. 4, 2009, pp. 277-279.
- [7] B. Yang, "Applied Neural Network in Oil Well Logging," Petroleum Industry Press, Beijing, 2005.
- [8] X. G. Wu and J. L. Ge, "Oil Production Forecast with ANN Method," *Petroleum Exploration and Development*, Vol. 21, No. 3, 1994, pp. 31-42.
- [9] C. H. Zhang, "Optimization Problems in Support Vector Machines," *China Agricultural University Doctoral Dissertation*, Vol. 5, 2004, pp. 36-39.
- [10] Z. H. Feng and J. M. Yang, "Practical Selection of Support Vector Machine Parameters for SVM Regression," *Mechanical Engineering and Automation*, Vol. 3, 2007, pp. 17-18.
- [11] K. Ito and R. Nakano, "Optimizing Support Vector Regression Hyperparameters Based on Cross-Validation," *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, 2003, pp. 2077-2082.
- [12] J. M. Yao, B. S. Yu, *et al.*, "Application of the Improved GM Model to Oil Production Forecasting of Tarim Basin," *Petroleum Geology and Oilfield Development in Daqing*, Vol. 26, No. 1, 2007, pp. 92-96.
- [13] M. F. Chen and Z. X. Lang, "Oil Production Forecasting with Modified Functional Simulation Model," *Xinjiang Petroleum Geology*, Vol. 24, No. 3, 2003, pp. 246-248.