

# The Application of Genetic Gray Model in Exploration Investment Prediction

Wang Xin<sup>1, a</sup>, Bian Min<sup>2, b</sup>, Wang Shichun<sup>3, c</sup>

<sup>1</sup>Computer Science and Technology ,China University of Petroleum, Beijing, 102249,China

<sup>2</sup>Computer Science and Technology ,China University of Petroleum, Beijing, 102249,China

<sup>3</sup>Computer Science and Technology ,China University of Petroleum, Beijing, 102249,China

<sup>a</sup>e-mail: xingwang@cup.edu.cn, <sup>b</sup>e-mail: bianminok@163.com, <sup>c</sup>e-mail: wsc132@163.com

**Keywords:** Genetic algorithm; GM (1,1) model; Data fitting

**Abstract.** This paper is a study of existing GM(1,1) model and manages to improve its inadequencies, such as the initial value, the background value of the model, the smoothness of the original data sequence, etc. By replacing the least square method with artificial intelligence method, the paper proposes a new model which combines genetic algorithm with the GM (1,1) model. The modified model can get more accurate data to improve the degree of data fitting and forecasting accuracy.

## Introduction

In the field of oil exploration, a wrong decision will often cause irreparable damage, so the judgments in oil exploration needs to be very correct. In the process of effective exploration, I hope we can not only save the cost of exploration as much as possible but also lower the investment in exploration. The most effective and direct way to make the right judgments of exploration investment is to study the influential factors of exploration investment and its inherent law so as to forecast exploration investment. In order to achieve a scientific and rational exploration investment, several factors need to be considered. First of all, on the basis of the available data, we preprocess the data to make the data fit to achieve high accuracy, so that it can reflect the overall trends of data providing a basis for forecasting exploration investment. But in reality, because the data of oil exploration investment doesn't have strong regularity, it is often very difficult to forecast. If we want to make valid predictions, we need to further explore the implication of internal variation in the data, to find a suitable and effective modeling method for data processing.

GM models have been widely used in the process of exploration and development investment since came into being<sup>[1-4]</sup>. people use GM model to predict the development of strategy and policy planning, and these applications also contributed to the development and progressment of gray GM model. Genetic algorithm is a common artificial intelligence algorithms. Using artificial intelligence algorithms to find the optimal solution is a common application<sup>[7]</sup>. So we want to integrate the genetic algorithm into GM(1,1) model to improve the degree of data fitting and forecasting accuracy<sup>[8]</sup>.

## GM(1,1) Principle and its Shortcomings

GM(1,1) forecasting model is the core of gray prediction, and its function is that the information can be used to predict the future based on uncertain information is known, its advantage is low demand for information.

Definition 2.1: set  $\mathbf{x}^0 = (x^0(1), x^0(2), \dots, x^0(n))$ ,  $\mathbf{x}^1 = (x^1(1), x^1(2), \dots, x^1(n))$

$$\mathbf{x}^0(k) + a\mathbf{x}^1(k) = \mathbf{b} \quad (2.1)$$

As the original form of GM (1,1) model,  $\mathbf{x}^1$  is 1-AGO sequence of  $\mathbf{x}^0$ . We can obtain the following definition by changing a little on the original form.

Definition 2.2 : set  $x^0$  and  $x^1$  As shown in definition 2.1, then the sequence  $z^1 = (z^1(2), z^1(3), \dots, z^1(n))$ , among  $z^1(k) = 0.5(x^1(k) + x^1(k-1))$ , called the  $x^0(k) + axz^1(k) = b$  (2.2)

as the basic form of GM(1, 1) model.  $z^1$  is a mean-generated sequence of  $x^1$ , called background value. The above aspect of the parameters a, b have the following definitions:

Definition 2.3:  $x^0 = (x^0(1), x^0(2), \dots, x^0(n)), x^0(k) \geq 0$ ,  $x^1$  is the 1-AGO sequence of  $x^0$ ,  $z^1$  is a mean-generated sequence of  $x^1$ . Matrix  $a = (a, b)^T$  is a parameter list, and

$$Y = \begin{bmatrix} x^0(2) \\ x^0(3) \\ \dots \\ x^0(n) \end{bmatrix}, B = \begin{bmatrix} -z^1(2)1 \\ -z^1(3)1 \\ \dots \\ -z^1(n)1 \end{bmatrix}$$

Least squares parameter estimation of  $x^0(k) + axz^1(k) = b$  meet  $a = (B^T B)^{-1} B^T Y$  (2.3)

Taking the estimated parameters obtained as described above into the standard form of GM (1,1) model. Then we can get the flowing definitions:

Definition 2.4:  $x^0$  is a non-negative sequence.  $x^1$  is the 1-AGO sequence of  $x^0$ .  $z^1$  is a mean-generated sequence of  $x^1$ .  $(a, b)^T = (B^T B)^{-1} B^T Y$  called

$$\frac{dx^1}{dt} + ax^1 = b \quad (2.4)$$

winterization equation for the grey differential equation of  $x^0(k) + axz^1(k) = b$ , also called shadow equation.

Definition 2.5:

(1) The solution of winterization equation  $\frac{dx^1}{dt} + ax^1 = b$  is  $x^1(t) = \left[ x^1(1) - \frac{b}{a} \right] e^{-at} + \frac{b}{a}$

which is called time corresponding function.

(2) Time corresponding sequence of grey differential equation  $x^0(k) + axz^1(k) = b$  is:

$$x^1(k+1) = \left[ x^1(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n$$

(3) Reducing value  $x^0(k+1) = a^1 x^{r1}(k+1) = x^{r1}(k+1) - x^{r1}(k), k = 1, 2, \dots, n$  (2.5)

Further elaborated the above definition, - a reflection of the development trend of  $x^{r1}$  and  $x^{r0}$ . The system action should be exogenous or front, but GM (1,1) model only uses the system action sequences and doesn't use outside the action sequence. The grey action excavates from the background values of data, reflecting the data changing. The existence of grey action not only distinguishes the grey system model and the traditional black-box model but also marks the difference between gray system model and gray box model. To discuss the application range of GM (1, 1).

Table 1 GM(1,1) Scope

$-a < 0.3$	Medium and long term forecast
$0.3 < -a < 0.5$	short-term forecast
$0.5 < -a < 0.8$	should be used with caution
$0.8 < -a < 1$	Can be used after correction
$-a > 1$	Unfavorable use

Precision of the model reflects the model prediction accuracy and practicability. We must go through a series of tests to prove whether the model is reasonably and available.

There are some problems about GM(1,1) model itself. Since the advent of GM(1, 1) model, many scholars go into GM (1,1) model and want to find the core factors to impact GM (1, 1)

prediction model in order to improve the prediction precision of GM(1,1) model. After years of constant efforts, the researchers found factors that affecting the accuracy of GM (1,1) model are: model of the initial value, the background value of the model, the smoothness of the original data sequence, the grey exponential law of modeling data sequence.

### **The Genetic Algorithm Improved GM(1,1) Model**

#### **a. The principle to improve**

By simply analyzing of the GM(1,1) model, we found some reasons which caused inaccurate results, including the background values and parameters. This main work is to find a way to improve the background values and parameters about GM(1,1) model.

The parameters of GM(1,1) are estimated from the least squares method. In the process of application of least square method, according to the definition of 2.3, there must be a set number of columns having the same importance between the before number and the after. But in practical applications, the importance of proportion is often difficult to determine which one is more important. According to the definition 2.3, we found that we can't lack this proportion, then how to modify the definition of 2.3 to ignore this proportion. Definition 2.3 is the process of applying the least squares method to estimate the GM(1,1) model parameters. The least-square method is using the method of minimum residual error to estimate. We can make a fitting about the original data to meet the definition of 2.2. Genetic algorithm can also achieve this fitting. Because the genetic algorithm adopts repeatedly iterative manners. In theory, the genetic algorithm can estimate a better result than the least squares method in an iterative process, if it does not fall into local optimal solution.

#### **b. Improved related processes**

How to use a genetic algorithm to estimate the parameters of GM(1,1) model. The general process of genetic algorithm is coding, selection, crossover and mutation. In this process, we need a fitness function to get the optimum offspring by this fitness function.

The problem is to find an appropriate fitness function .The result which genetic algorithm calculates is a fitting series that is consistent with the original series. But we last need two parameters instead of a set of data. The parameter is as the objects of coding, selection, crossover and mutation. Then we can get a series which is similar to the original data by means of parameters and related functions. First, we estimate a value range of parameters through the original sequence; Second, we randomly selected a set of parameters as a parent in this range; Third, we get a set number of columns through the relevant functions; Fourth, we the take the minimum of residuals which gets from the original series and obtained series as fitness function; After that, through selection, crossover, mutation, we get the minimum residuals, and then we obtain the required parameters based on the correlation function, the parameters of this group parameter is the GM(1,1) needed. Like fig.1

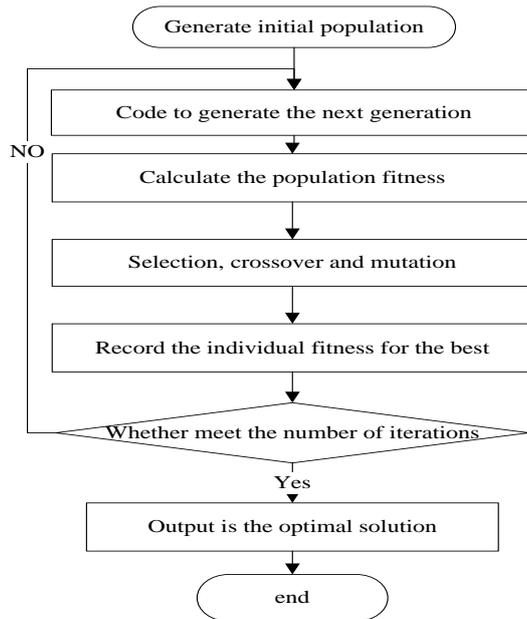


Fig.1 The flow chart of genetic algorithm

## Test Results

### a. GM(1,1)Predicting Outcomes

We use GM(1,1) to predict. In this paper, we could find out the GM(1,1) model parameters by processing the original sequence with GM(1,1). We set  $a=-0.094$ ,  $b=26.63$ , and take these two parameters and the original sequence back to the formula which described above in order to obtain its prediction formula.

$$x^{(1)}(k) = \left( 21.43 + \frac{26.63}{0.094} \right) e^{-0.094k} - \frac{26.63}{0.094}$$

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

When the  $k$  value is increasing, we can obtain the corresponding predicted values.

Take a look at the accuracy of the model. Generally there are two ways to verify the accuracy of the model, relative error and residual error. Here we compare the predicted values in the formula with the original sequence to estimate the precision of the model.

By calculating, the residual error of the model is 0.44, the standard deviation is 0.0043, the predicted values which is from GM(1,1) model are as follows: 21.43、30.26、33.23、36.49、40.08、44.02、48.34、53.10、58.32、64.06、70.37、77.30.

The difference of forecast data and the original data is expressed in curve line as shown in fig.2. The horizontal axis represents times and the ordinate axis represents the difference between the predicted data and the original data.

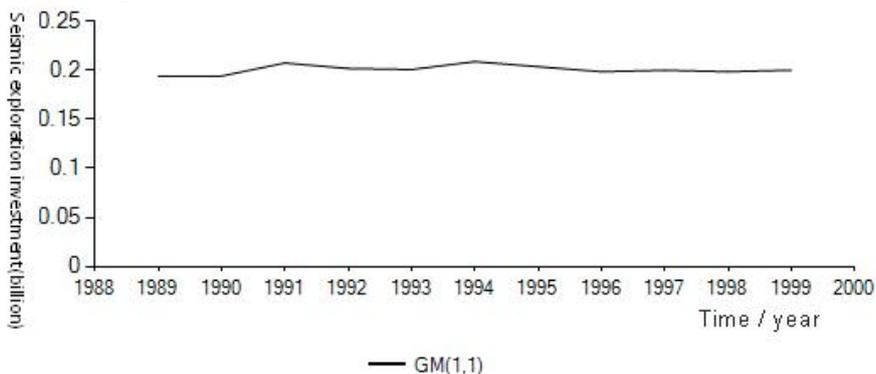


Fig.2 GM(1,1) forecast error

**b. The prediction results of GM(1,1) by genetic algorithm**

Firstly, we get the 1-AGO sequence from the original data, but we doesn't need to compute the mean generating sequence comparing with GM(1,1), and we use the genetic algorithm to estimate the parameters instead of the LMS algorithm. According to many similar applications of genetic algorithm for parameter values, the parameters setting of genetic algorithm are as follows:

The number of population equals to 500; Gene length equals to 14; Population Size equals to 120; Genetic algebra equals to 150; Crossover rate equals to 0.7; Mutation rate equals to 0.1; It can be obtained the parameters by using the genetic algorithm, a=-0.093、b=26.72。 Then, the grey prediction formula is:

$$x^{(1)}(k) = \left( 21.43 + \frac{26.72}{0.093} \right) e^{-0.094k} - \frac{26.72}{0.093}$$

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

When the k value is increasing, we can obtain the corresponding predicted values. In order to verify its accuracy, calculating the residual error is 0.046、 Standard deviation is 0.0011. The corresponding predicted value is as follows: 21.43、 30.14、 33.09、 36.34、 39.91、 43.83、 48.14、 52.88、 58.08、 63.80、 70.08、 76.98。

The difference of forecast data and the original data is expressed in curve line as shown in fig.2. The horizontal axis represents time and the ordinate axis represents the difference between the predicted data and the original data.

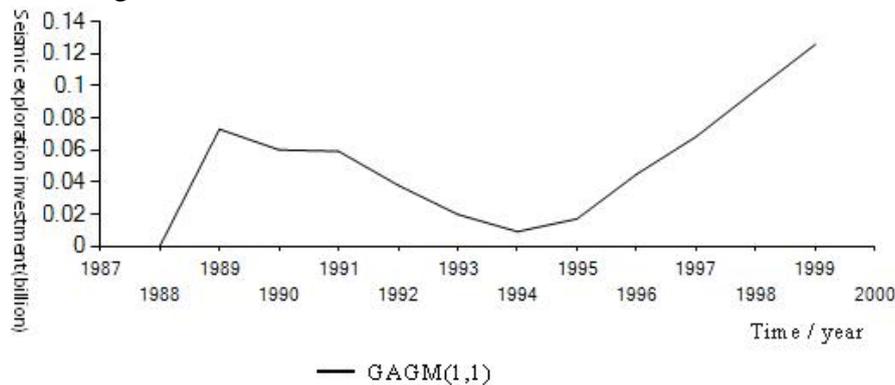


Fig.3 GAGM(1,1) forecast error

**Conclusion**

Through experimental data above, we can clearly see that the modified GM(1,1) model by using genetic algorithms is better than GM(1,1) model. But the speed of GM(1,1) model is faster.

Table 2 Algorithm comparison

	GM(1,1)	GAGM(1,1)
Residuals	0.44	0.046
Standard deviation	0.0043	0.0011

**Acknowledgement**

In this paper, the research was sponsored by the Research Group of Professor Wang.

**References**

- [1] Zhou Deqiang. Improved grey Verhulst model in the application of the medium and long-term load forecasting. The grid technology, 2009,10: 124-127.
- [2] Yang Hongping, Ma Dachuan. The application of grey forecasting theory in information

analysis and prediction. Intelligence method, 2004, 3:22-24.

[3] Zuo Xing, He Shiming, Li Wei, et al. The application of GM(1,1) model in oil engineering. Inner Mongolia Petrochemical Industry, 2007, 5:41-42.

[4] Ma Youping, Zhihua, Ai Xunru. Forecasting urbanization level by grey genetic algorithm. Hubei institute for nationalities, 2008, 3:68-74

[5] Liu xin. The improved research about Grey prediction GM(1,1) and GM(2,1) model, Inner Mongolia, Inner Mongolia university of technology, 2007.

[6] Liu Sifeng, Dang Yaoguo, et al. The grey system theory and its application of fifth edition, science press, 2010.

[7] Cheng Bin, Wang Zhihong. The grey prediction model based on improved genetic algorithm research. Computer and information technology, 2009, 8: 25-28.

[8] Dongxiao Niu, Wei Li, Zhuhua Han, et al. Power Load Forecasting based on Improved Genetic Algorithm-GM(1,1) Model Fourth International Conference on Natural Computation, 2008: 630-634.