

# An Optimal Combination Forecasting Model Based on Elite Genetic Algorithm: Evidence from China's Defense Expenditure

Kan Zhang, Bao-ping Liu and Dong Huang

*Dept. of Economics & Equipment Management, Naval University of Engineering,  
Wuhan, Hubei, China*

*zhangkanwh@163.com, liubpwh@yahoo.com.cn, 13397191239@189.cn*

## **Abstract**

*China's defense expenditure data usually have a small sample of non-linear characteristics. Aiming at this feature, the nonlinear residual G-Verhulst model is used to study the problem that the traditional measurement model is not enough for the residual information mining and the prediction accuracy is not high. On this basis, we choose elitegenetic algorithm (short for EGA) to establish G-Verhulst metering combination forecasting model, design the modeling algorithm logic flow and overall idea of nonlinear residuals G-Verhulst combination forecasting model to solve the multi-criteria objective optimization problems, compare the forecasting model results with other models. Empirical studies show that: the small sample nonlinear residual G-Verhulst metering combination forecasting model based on EGA converges fast, with good fitting results and more accurate predictions. The model can effectively solve the problem of China's defense expenditure forecast.*

**Keywords:** *combined forecasting, small sample, nonlinear residuals, G-Verhulst, EGA*

## **1. Introduction**

In the complex economic system, economic data by time sequence are vulnerable to noise outside, such as defense expenditure, while showing a non-stationary, randomness and other complex characteristics. Since 1970s, by the rapid development of modern econometrics, cointegration theory and volatility models [1] have appeared as a series of econometric model representative for accurate prediction, which provide effective means for accurately characterizing and predicting the changing characteristics of economic data. For the same forecasting problem, since the modeling mechanism and information presentation angles of various econometric models are different, the forecasting conclusion of each model can reflect only one aspect of data information, but cannot reflect all the information of forecasting problems. Especially for small sample economic data, the "poor information" characteristic is more typical, so easy to forecast the actual problem in larger errors by external noise interference for prediction accuracy. The literature researches show that there are two main ways to improve the accuracy of the model. First, use the data mining and analysis techniques to correct the residuals of the individual econometric model [2-5], and improve the adaptability and stability of model by collecting the useful information of the residuals. Second, the individual economic models are superimposed and integrated to form the combination forecasting model [6-9], which comprehensively considers the overall difference and influence of the prediction effect, improves the authority and systematical of the conclusion.

At present, the researches on the residual correction are mainly focus on the information mining of residual data. The gray model is usually used to study the characteristics of residual variation, and the residual information is used to improve the forecasting model. The literature [10] presented a method of forecasting fire occurrence based on residual correction GM(1,1) model. In the literature [11], the applicable

conditions and applicable ranges of generalized discrete GM(1,1) model were discussed in depth. The literature [12] showed that in order to effectively improve the prediction accuracy, the numerical integration trapezoidal formula can be used to solve GM(1,N) model. However, most of the economic data in the macroeconomic are non-stationary, and their residuals are typical nonlinear characteristics. The traditional GM(1,1) and GM(1,N) are both linear model in the form, and there is a problem that the modeling accuracy of the nonlinear data is not satisfactory in the residual improvement. Although the approximate whitening time response of the GM(1,N) model was given in literature [13], it was pointed out in literature [14,15] that there was still an inaccurate result in the response, the experimental error exceeded the tolerance threshold and the prediction was not satisfactory. On the other hand, the combination forecasting model embodies the scientific choice and overall application of comprehensive information, and avoids the prediction failure caused by inadequate and incomplete grasp of information, which is advantage that individual forecasting model does not have. The combination forecasting model has been widely used in many fields such as transportation [16], stock analysis [17], economic forecasting [18], and how to construct the most reasonable and efficient combination form has become a difficult problem in the field of combination forecasting.

In view of this, a small sample nonlinear residual G-Verhulst metering combination forecasting model is proposed to improve the prediction accuracy of residual time sequence of small samples. At the same time, an improved genetic algorithm with elitist strategy which is widely used in the fields of function combination optimization and artificial intelligence control is selected to dynamically weight the small sample nonlinear residual G-Verhulst metering combination forecasting model to further improve the overall prediction accuracy of the model. In order to test the effect of the model, this paper chooses the China's defense expenditures to carry on the empirical analysis. Based on the comparison of traditional modeling and forecasting methods, empirical researches and forecast conclusions are given.

## 2. The Establishment of Nonlinear Residual G-Verhulst Metering Combination Forecasting Model

The traditional Verhulst model [19] is a nonlinear model formed by the Malthusian linear model transformation with specific constraints. However, this model is more stringent for data distribution characteristics, and has certain limitations. The G-Verhulst model, which is based on the combination of gray theory and Verhulst model, greatly improves the stability of the model by relaxing the use conditions, and is widely used in many fields such as economy, society and military. Especially for non-stationary multivariable nonlinear system structure prediction, G-Verhulst model is more accurate than the traditional GM(1,1) model and GM(1,N) model to describe the data rule, the modeling error is small, the superiority is obvious. With the development of combination forecasting theory, the focus of prediction research is to shift to the field of residual analysis for the purpose of pursuing more accurate forecasting results. Based on this, the nonlinear residual G-Verhulst model which is highly consistent with the reality complex economic system is established to use system theory knowledge to fully tap residual hidden information, improve prediction accuracy and enhance model stability and ductility.

### 2.1. Model Specialized Discussion

**Discussion 1** Given  $Y = (y(1), y(2), \dots, y(n))$  is a non-stationary time sequence, denoted by  $Y = \{y(t), t = 1, 2, \dots, n\}, \forall y(t) \in y \Rightarrow t \in T = \{1, 2, \dots, n\}$ . If there is  $t^* \in T$  to

make  $y(t^*) > y(t)$ ,  $t \in \{1, \dots, t^* - 1\} \cup \{t^* + 1, \dots, n\}$ ,  $Y$  is called the non-stationary single peak sequence with  $y(t^*)$  as the peak.

**Discussion 2** Let  $Y^{(0)}$  be the original sequence observation value that meet the definition 1, denoted by  $Y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$ . Given  $\hat{Y}^{(0)}$  is prediction value of a econometric model, for the nonlinear residual sequence is  $\varepsilon^{(0)} = Y^{(0)} - \hat{Y}^{(0)}$ .  $\exists \varepsilon^{(0)}(t^*) < \varepsilon^{(0)}(t)$ ,  $t \in \{1, \dots, t^* - 1\} \cup \{t^* + 1, \dots, n\}$ , for  $\varepsilon_{\min}^{(0)} = \varepsilon^{(0)}(t^*)$ , denote  $\eta^{(0)}(t) = \varepsilon^{(0)}(t) + |\varepsilon_{\min}^{(0)}|$ . By gray theory knowledge, denote  $\eta^{(1)} = AGO\eta^{(0)}$ ,  $Z^{(1)} = MEAN\eta^{(1)}$ , which is  $\eta^{(0)} = (\eta^{(0)}(1), \eta^{(0)}(2), \dots, \eta^{(0)}(n))$ ,  $\eta^{(1)} = (\eta^{(1)}(1), \eta^{(1)}(2), \dots, \eta^{(1)}(n))$ ,  $Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$ . The model  $\eta^{(0)}(t) + az^{(1)}(t) = b(z^{(1)}(t))^\alpha$  is called the nonlinear residual grey power model, denoted by  $GM(1,1,\alpha)$  for short. Especially, for  $\alpha = 2$ , the model is called the nonlinear residual G-Verhulst model, in which  $a$  is called the development coefficient,  $b$  is called the gray effect.

## 2.2. Model Solving and Parameter Estimation

**Theorem 1** For the nonlinear residual G-Verhulst model  $\eta^{(0)}(t) + az^{(1)}(t) = b(z^{(1)}(t))^2$ , its whitening model is  $\frac{d\eta^{(1)}}{dt} + a\eta^{(1)} = b(\eta^{(1)})^2$ , the whitening response is

$$\bar{\eta}^{(1)}(t+1) = \frac{\frac{a}{b}}{1 + \left(\frac{a}{b\eta^{(0)}(1)} - 1\right)e^{at}}$$

**Proof** It is easy to get the relationship between the nonlinear residual G-Verhulst whitening model and the definition formula, which is based on the knowledge of gray theory and modeled by the relationship between the GM(1,1) whitening model and definition relation. The following is the key to prove the whitening response of nonlinear residual G-Verhulst model.

for 
$$\frac{d\eta^{(1)}}{dt} + a\eta^{(1)} = b(\eta^{(1)})^2 \Rightarrow \frac{d\eta^{(1)}}{b(\eta^{(1)})^2 - a\eta^{(1)}} = dt,$$

integral in both side

$$-\int_{\eta^{(1)}(t_0)}^{\eta^{(1)}(t)} \frac{d\eta^{(1)}(t)}{a\eta^{(1)}(t) - b(\eta^{(1)}(t))^2} = \int_{t_0}^t dt = t - t_0.$$

Decompose the left side of the equation

$$\begin{aligned} & -\int_{\eta^{(1)}(t_0)}^{\eta^{(1)}(t)} \frac{d\eta^{(1)}(t)}{a\eta^{(1)}(t) - b(\eta^{(1)}(t))^2} = -\frac{1}{a} \int_{\eta^{(1)}(t_0)}^{\eta^{(1)}(t)} \left( \frac{1}{\eta^{(1)}(t)} + \frac{b}{a - b\eta^{(1)}(t)} \right) d\eta^{(1)}(t) \\ & = -\frac{1}{a} \int_{\eta^{(1)}(t_0)}^{\eta^{(1)}(t)} \frac{d\eta^{(1)}(t)}{\eta^{(1)}(t)} - \frac{1}{a} \int_{\eta^{(1)}(t_0)}^{\eta^{(1)}(t)} \frac{d\left(\frac{a}{b} - \eta^{(1)}(t)\right)}{\left(\frac{a}{b} - \eta^{(1)}(t)\right)} = -\frac{1}{a} \left[ \ln \frac{\eta^{(1)}(t)}{\eta^{(1)}(t_0)} - \ln \frac{\frac{a}{b} - \eta^{(1)}(t)}{\frac{a}{b} - \eta^{(1)}(t_0)} \right] \\ & = -\frac{1}{a} \left[ \ln \frac{\eta^{(1)}(t)}{\eta^{(1)}(t_0)} \frac{\frac{a}{b} - \eta^{(1)}(t_0)}{\frac{a}{b} - \eta^{(1)}(t)} \right] = t - t_0 \end{aligned}$$

thus

$$\begin{aligned} \frac{\eta^{(1)}(t)}{\eta^{(1)}(t_0)} \frac{\frac{a}{b} - \eta^{(1)}(t_0)}{\frac{a}{b} - \eta^{(1)}(t)} & = e^{-a(t-t_0)} \Rightarrow \frac{a}{b} \eta^{(1)}(t) - \eta^{(1)}(t)\eta^{(1)}(t_0) = \left(\frac{a}{b} \eta^{(1)}(t_0) - \eta^{(1)}(t)\eta^{(1)}(t_0)\right) e^{-a(t-t_0)} \\ & \Rightarrow \eta^{(1)}(t) \left[ \frac{a}{b} - \eta^{(1)}(t_0) + \eta^{(1)}(t_0) e^{-a(t-t_0)} \right] = \left(\frac{a}{b}\right) \eta^{(1)}(t_0) e^{-a(t-t_0)} \end{aligned}$$

$\eta^{(1)}(t)$  is denoted as  $\bar{\eta}^{(1)}(t)$ , which is

$$\hat{\eta}^{(1)}(t) = \frac{\frac{a}{b}\eta^{(1)}(t_0)e^{-a(t-t_0)}}{\frac{a}{b}-\eta^{(1)}(t_0)+\eta^{(1)}(t_0)e^{-a(t-t_0)}} = \frac{\frac{a}{b}}{\frac{a}{b\eta^{(1)}(t_0)}-1+e^{-a(t-t_0)}} = \frac{\frac{a}{b}}{\left(\frac{a}{b\eta^{(1)}(t_0)}-1\right)e^{-a(t-t_0)}+1}$$

if  $t_0 = 1, t = t' + 1$ , that

$$\hat{\eta}^{(1)}(t'+1) = \frac{\frac{a}{b}}{1 + \left(\frac{a}{b\eta^{(1)}(1)} - 1\right)e^{at'}}$$

Proof is completion.

**Theorem 2** Make the normalized nonnegative residual sequence  $\eta^{(0)}$  as a non-stationary single peak sequence, which is  $\eta^{(1)} = AGO\eta^{(0)}$ ,  $z^{(1)} = MEAN\eta^{(1)}$ ,  $\eta^{(0)} = (\eta^{(0)}(1), \eta^{(0)}(2), \dots, \eta^{(0)}(n))$ ,  $\eta^{(1)} = (\eta^{(1)}(1), \eta^{(1)}(2), \dots, \eta^{(1)}(n))$ ,  $Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n))$ , thus parameter estimation vector  $P$  in G-Verhulst model  $\eta^{(0)}(t) + az^{(1)}(t) = b(z^{(1)}(t))^2$  is  $P = (a, b) = (A, B, C, D, E)$ . In this model,  $a = \frac{AE - DB}{CD - A^2}$ ,  $b = \frac{CE - AB}{CD - A^2}$ ,  $A = \sum_{t=2}^n z^{(1)}(t)^3$ ,

$$B = \sum_{t=2}^n z^{(1)}(t)\eta^{(0)}(t), \quad C = \sum_{t=2}^n z^{(1)}(t)^2, \quad D = \sum_{t=2}^n z^{(1)}(t)^4, \quad E = \sum_{t=2}^n z^{(1)}(t)^2\eta^{(0)}(t).$$

**Proof** For G-Verhulst model definition formula  $\eta^{(0)}(t) + az^{(1)}(t) = b(z^{(1)}(t))^2$ , set  $t = 2, 3, \dots, n$  and put it into the formula above

$$\begin{aligned} \eta^{(0)}(2) + az^{(1)}(2) &= b(z^{(1)}(2))^2, \\ \eta^{(0)}(3) + az^{(1)}(3) &= b(z^{(1)}(3))^2, \\ &\vdots \\ \eta^{(0)}(n) + az^{(1)}(n) &= b(z^{(1)}(n))^2, \end{aligned}$$

The formula above is expressed as a matrix

$$\eta^* = GP$$

$$G = \begin{bmatrix} -z^{(1)}(2) & z^{(1)}(2)^2 \\ -z^{(1)}(3) & z^{(1)}(3)^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & z^{(1)}(n)^2 \end{bmatrix}, \eta^* = \begin{bmatrix} \eta^{(0)}(2) \\ \eta^{(0)}(3) \\ \vdots \\ \eta^{(0)}(n) \end{bmatrix}, P = \begin{bmatrix} a \\ b \end{bmatrix}.$$

By the least squares criterion, we can get

$$P = (G^T G)^{-1} G^T \eta^*$$

$$G^T G = \begin{bmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \dots & -z^{(1)}(n) \\ z^{(1)}(2)^2 & z^{(1)}(3)^2 & \dots & z^{(1)}(n)^2 \end{bmatrix} \cdot \begin{bmatrix} -z^{(1)}(2) & z^{(1)}(2)^2 \\ -z^{(1)}(3) & z^{(1)}(3)^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & z^{(1)}(n)^2 \end{bmatrix} = \begin{bmatrix} \sum_{t=2}^n z^{(1)}(t)^2 & -\sum_{t=2}^n z^{(1)}(t)^3 \\ -\sum_{t=2}^n z^{(1)}(t)^3 & \sum_{t=2}^n z^{(1)}(t)^4 \end{bmatrix}$$

$$\therefore (G^T G)^{-1} = \begin{bmatrix} C & -A \\ -A & D \end{bmatrix}^{-1} = \frac{1}{CD - A^2} \begin{bmatrix} D & A \\ A & C \end{bmatrix}$$

also

$$G^T \eta^* = \begin{bmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \dots & -z^{(1)}(n) \\ z^{(1)}(2)^2 & z^{(1)}(3)^2 & \dots & z^{(1)}(n)^2 \end{bmatrix} \cdot \begin{bmatrix} \eta^{(0)}(2) \\ \eta^{(0)}(3) \\ \vdots \\ \eta^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -\sum_{t=2}^n z^{(1)}(t)\eta^{(0)}(t) \\ -\sum_{t=2}^n z^{(1)}(t)^2\eta^{(0)}(t) \end{bmatrix} = \begin{bmatrix} -B \\ E \end{bmatrix},$$

$$\therefore P = \begin{bmatrix} a \\ b \end{bmatrix} = (G^T G)^{-1} G^T \eta^* = \frac{1}{CD - A^2} \begin{bmatrix} D & A \\ A & C \end{bmatrix} \begin{bmatrix} -B \\ E \end{bmatrix} = \begin{bmatrix} \frac{AE - DB}{CD - A^2} \\ \frac{CE - AB}{CD - A^2} \end{bmatrix}.$$

Proof is completion.

### 3. Combination Forecasting and Elitist Genetic Algorithm

#### 3.1. Mathematical Description of Combination Forecasting

Suppose that a complex economic system can be predicted by  $m$  different econometric models. From discussion 1, we know that  $Y = \{y(t), t = 1, 2, \dots, n\}$  are actual observation values.  $f_i(t)$  is defined as the  $t$  th prediction value predicted by  $i$  th econometric model,  $\sigma_i$  is the weight coefficient of  $i$  th econometric model,  $E_i(t) = y(t) - f_i(t)$  is the residual of  $t$  th prediction value predicted by  $i$  th econometric model. The metering combination forecasting model composed of  $m$  econometric models is the residual of  $t$  th prediction value predicted by  $m$  econometric model.

$$y(t) = \sum_{i=1}^m \sigma_i f_i(t) + E(t) \quad (t = 1, 2, \dots, n; i = 1, 2, \dots, m)$$

In formula above,  $E(t)$  is the residual of  $t$  th prediction value predicted by metering combination forecasting model composed of  $m$  econometric models.

#### 3.2. Using EGA to Determine Combination Weight

Using the Elitist Genetic Algorithm (EGA) to obtain the optimal value of the combined weight coefficient  $\sigma_i$ , the optimal combination forecasting model can be obtained. Make  $\hat{\sigma}_i$  as the estimated value of the weight coefficient is the combined forecasting value,  $\hat{y}(t)$  is combination forecasting value. There is

$$\hat{y}(t) = \sum_{i=1}^m \hat{\sigma}_i f_i(t)$$

Usually, some combination forecasting criterion is used to determine the optimal combination weight coefficient  $\hat{\sigma}_i$ . The commonly criteria include:

**Criterion 1** The mean square error (MSE) minimum criterion, which is determining parameters  $\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_m$  to minimize the following function.

$$S_1 = \frac{1}{n} \sum_{t=1}^n [y(t) - \hat{y}(t)]^2 = \frac{1}{n} \sum_{t=1}^n [y(t) - (\hat{\sigma}_1 f_1(t) + \hat{\sigma}_2 f_2(t) + \dots + \hat{\sigma}_m f_m(t))]^2$$

**Criterion 2** The mean absolute error (MAE) minimum criterion, which is determining parameters  $\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_m$  to minimize the following function.

$$S_2 = \frac{1}{n} \sum_{t=1}^n |y(t) - \hat{y}(t)| = \frac{1}{n} \sum_{t=1}^n |y(t) - (\hat{\sigma}_1 f_1(t) + \hat{\sigma}_2 f_2(t) + \dots + \hat{\sigma}_m f_m(t))|$$

**Criterion 3** The mean absolute percent error (MAPE) minimum criterion, which is determining parameters  $\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_m$  to minimize the following function.

$$S_3 = \frac{1}{n} \sum_{t=1}^n \left| \frac{y(t) - \hat{y}(t)}{y(t)} \right| = \frac{1}{n} \sum_{t=1}^n \left| \frac{y(t) - (\hat{\sigma}_1 f_1(t) + \hat{\sigma}_2 f_2(t) + \dots + \hat{\sigma}_m f_m(t))}{y(t)} \right|$$

**Criterion 4** Theil coefficient minimum criterion, Theil coefficient is always between 0 and 1. The smaller the Theil coefficient is, the smaller the distance between the fitted value and the actual value is, the higher the prediction accuracy is, which is determining parameters  $\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_m$  to minimize the following function.

$$S_4 = \frac{\sqrt{n^{-1} \sum_{t=1}^n [y(t) - \hat{y}(t)]^2}}{\sqrt{n^{-1} \sum_{t=1}^n y(t)^2} \sqrt{n^{-1} \sum_{t=1}^n \hat{y}(t)^2}} = \frac{\sqrt{n^{-1} \sum_{t=1}^n [y(t) - (\hat{\sigma}_1 f_1(t) + \hat{\sigma}_2 f_2(t) + \dots + \hat{\sigma}_m f_m(t))]^2}}{\sqrt{n^{-1} \sum_{t=1}^n y(t)^2} \sqrt{n^{-1} \sum_{t=1}^n (\hat{\sigma}_1 f_1(t) + \hat{\sigma}_2 f_2(t) + \dots + \hat{\sigma}_m f_m(t))^2}}$$

In the actual decision-making, the weights of elite genetic algorithms are weighted by the combination of multiple criteria, which is a typical NP complete problem because of different criteria standard. The research on combination forecasting of NP complete problem can be transformed into the model below:

$$\begin{aligned} \min \phi &= \phi(S_1, S_2, \dots, S_4) \\ \text{s. t. } &\sum_{i=1}^m \sigma_i = 1 \\ &1 \geq \sigma_i \geq 0 \\ &i = 1, 2, \dots, m \end{aligned}$$

Where,  $\phi(S_1, S_2, \dots, S_4)$  is the objective function,  $S_1, S_2, \dots, S_4$  respectively are on behalf of the criteria 1-4. As the actual meaning of the negative weight coefficient is not clear enough, so the weight coefficient requirements must be positive.

### 3.3. EGA Optimization Algorithm Program Design

Elite genetic algorithm (EGA) is a new improved intelligent algorithm which combines standard genetic algorithm (SGA) and elite selection strategy [20,21]. Experimental results showed that the crossover operator and select operator of SGA has the irreversible destructive effect to high-order and long-range mode [22]. It is easy to make the best individual lose, the case will appear repeatedly with the evolution and iteration, will lead to SGA does not have global convergence, while EGA has global convergence, elite selection strategy can effectively overcome the slow convergence problem of computation due to the loss of the optimal individuals in SGA, and can prevent the evolution of individuals from being affected by the local optimal traps to occur the genetic premature phenomenon. EGA algorithm program is shown in Figure 1.

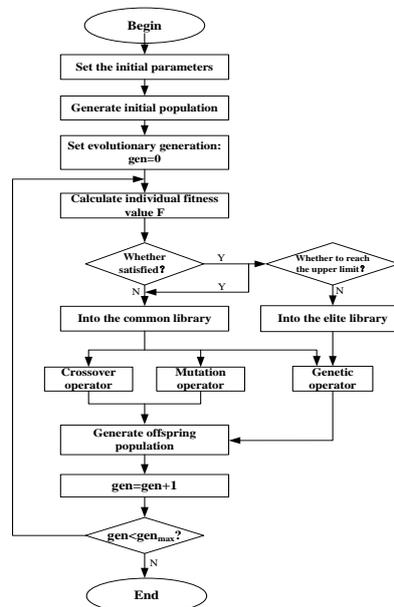


Figure 1. Complete Iterative Flow Chart of EGA

Step 1: Initialize the program. Set the main initial parameters are population size  $Q$ , elite population size  $Q^*$ , general population size  $Q - Q^*$ , population evolutionary generation

threshold  $gen_{max}$ , genetic probability  $p_r$ , crossover probability  $p_c$  and mutation probability  $p_n$ .

Step 2: Generate the initial population  $P(0)$ . Randomly select a point from  $[0,1]$ , denoted as  $\sigma_i(q,0), q=1,2,\dots,Q$ , and then randomly select  $m-1$  points from  $[0,1-\sigma_i(q,0)]$  to represent the combined initial weight coefficient, denoted as  $\sigma_i(q,0), i=1,2,\dots,m-1$ . Adding the  $m$ th point, set  $\sigma_m(q,0)=1-\sum_{i=1}^{m-1}\sigma_i(q,0)$ . The initial population  $P(0)$  composed of these points is a  $Q \times m$  matrix, means that each generation has  $Q$  chromosomes, each chromosome has  $m$  gene fragments, which is

$$P(0) = \begin{bmatrix} \sigma_1(1,0) & \sigma_2(1,0) & \dots & \sigma_m(1,0) \\ \sigma_1(2,0) & \sigma_2(2,0) & \dots & \sigma_m(2,0) \\ \vdots & \vdots & \dots & \vdots \\ \sigma_1(Q,0) & \sigma_2(Q,0) & \dots & \sigma_m(Q,0) \end{bmatrix}$$

Step 3: Population binary coding. Express  $\sigma_i(q,gen)$  as a binary string, in which  $gen$  represents evolutionary generation (initial order  $gen=0, gen \leq gen_{max}$ ), the encoding of  $\sigma_i(q,gen)$  may consist of  $x$  bit binary strings,  $x$  is determined by the accuracy of  $\sigma_i(q,gen)$ . The gene fragment on each chromosome in the population is represented by a binary string, totaling  $m$  binary strings.

Step 4: Find the fitness value  $F$ . The fitness formulas of different criteria are different. To simplify the process, only  $S_1$  is taken as an example. The fitness value of each chromosome  $\sigma_i(q,gen)$  in the population  $P(gen)$  is calculated, which is

$$F(\sigma_i(q,gen)) = \begin{cases} \phi(S_1)_{max} - f(\sigma_i(q,gen)) & 0 \leq \sigma_i(q,gen) \leq 1 \\ 0 & else \end{cases}$$

In the above formula,  $f(\sigma_i(q,gen)) = \sum_{t=1}^n (y(t) - \sum_{i=1}^m \sigma_i(q,gen) f_i(t))^2$ ,  $\phi(S_1)_{max}$  is the maximum of MSE under the  $S_1$  criterion in same generation.

Step 5: Select chromosome library. According to the individual fitness value  $F(\sigma_i(q,gen))$ , it is determined whether  $F(\sigma_i(q,gen))$  meets the requirements of the elite library, if not satisfied, put into the common library. If satisfied, it is further determined whether the number of elite individuals in the elite library reaches the upper limit of the population size, if not reached the upper limit, put into the elite library to produce elite chromosomes. Otherwise, put into the common library to produce common chromosomes.

Step 6: Genetic manipulation, cross and mutation operation. For the elite library, the excellent properties of the parent chromosome are maximizing preserved. The genetic probability is close to 1, allowing only a single point of variation with very low probability. For the common library, first calculate the survival probability of each chromosome in same generation, which is

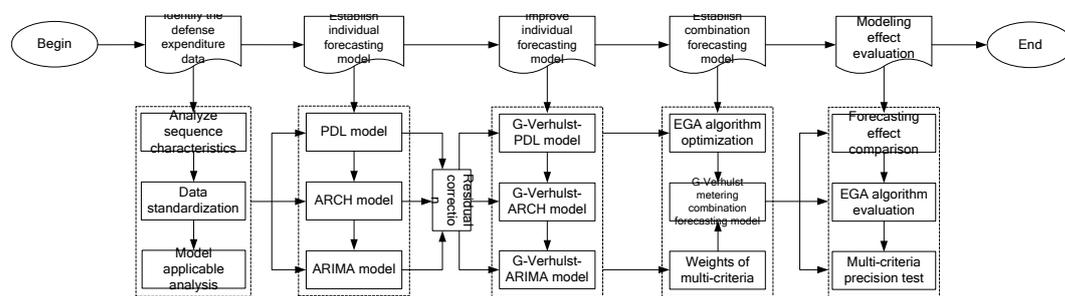
$$p_i^{(gen)} = \frac{F(\sigma_i(q,gen))}{\sum_{j=1}^Q F(\sigma_j(j,gen))}$$

Using roulette mode to select the survival chromosomes, and ensures that each individual is selected with probability  $p_i^{(gen)}$ . Then, the crossover operator  $p_c$  and the mutation operator  $p_n$  are used to calculate iteratively. Finally, the elite library and common library are combined to generate the offspring population  $P(gen+1)$ .

Step 7 : Iterative operation and evolution termination conditions. Repeated the implementation of Step 4~Step 6 program Steps, according to the fitness value  $F$  from large to small sort, update the elite library and common library. When the offspring population is generated, the population parameters are modified, and when the evolutionary generation reaches the prescribed threshold, the program terminates.

#### 4. Overall Modeling Design Ideas of China’s Defense Expenditure Combination Forecasting

Based on nonlinear residual G-Verhulst model and the combinative weight of EGA algorithm, the nonlinear residual feature of small sample is modified to give the whole idea of modeling in this paper. For the study subject of China’s defense expenditure, the complete modeling steps of the small sample nonlinear residual G-Verhulst metering combination forecasting model based on EGA algorithm is shown in Figure 2.



**Figure 2. Modeling Design of Small Sample Nonlinear Residual G-Verhulst Metering Combination Forecasting Model based on EGA**

Step 1: Identify the defense expenditure data. Select the appropriate small sample time sequence, use statistical methods to analyze its characteristics, including the trend, autocorrelation, heteroskedasticity

and so on. Complete the data standardization work, combined with the conclusion of the analysis to select the applicable model.

Step 2: Establish individual forecasting model. According to the multi-information characteristics of the sequence, the PDL model, ARCH model and ARIMA model are respectively established to describe the characteristics.

Step 3: Modification and improvement of individual forecasting model. Analyze the residuals of individual forecasting model to respectively establish the G-Verhulst-PDL model, the G-Verhulst-ARCH model and the G-Verhulst-ARIMA model for small sample nonlinear residuals. The basic forms are

$$\begin{cases} G\_Verhulst\_PDL = PDL + (\varepsilon_i\_PDL) \\ G\_Verhulst\_ARCH = ARCH + (\varepsilon_i\_ARCH) \\ G\_Verhulst\_ARIMA = ARIMA + (\varepsilon_i\_ARIMA) \end{cases}$$

In the above formulas,  $\varepsilon_i\_PDL, \varepsilon_i\_ARCH, \varepsilon_i\_ARIMA$  are corresponding residual models.

Step 4: Establish metering combination forecasting. The MSE criterion, MAE criterion, MAPE criterion and Theil criterion are used to design the objective function, and the G-Verhulst metering combination forecasting model is established. Select EGA to optimize the model weight, shorten the convergence time and further improve the prediction accuracy. The basic form is

$$\begin{cases} G\_Verhulst = \sigma_1 \cdot (G\_Verhulst\_PDL) + \sigma_2 \cdot (G\_Verhulst\_ARCH) + \sigma_3 \cdot (G\_Verhulst\_ARIMA) \\ \sigma_1 + \sigma_2 + \sigma_3 = 1 \\ \sigma_i \geq 0, i = 1, 2, 3 \end{cases}$$

Step 5: Modeling effect evaluation. Compare the prediction accuracy of different forecasting models, analyze the relationship among EGA parameter, convergence speed and calculation accuracy, and then verify the superiority of combination forecasting model.

## 5 Empirical Researches

### 5.1. Data Sources and Research Samples

Because China's reform and opening up year is 1978, the defense expenditure data before 1978 are not representative. In this paper, we only choose China's defense expenditure from 1978 to 2016 as research samples, all the data from the National Bureau of Statistics. In order to prevent the loss of data information and enhance the smoothness of data, the original data are processed by natural logarithm, 39 samples of processed data are obtained, which has typical small sample characteristic.

### 5.2. Empirical Analysis

The data of national defense expenditure are divided into two groups, 32 data from 1978 to 2011 are used as fitting samples and the other 5 data from 2012 to 2016 as prediction samples. The analysis of defense expenditure sample data, combined with the unit root test shows that national defense expenditure is non-stationary time sequence. Investigating the sequence characteristics of trend, autocorrelation, heteroskedasticity and volatility, we can respectively establish PDL model ARCH model and ARIMA model. Model expressions and related parameters are shown in Table 1.

**Table 1. Model Expressions and Related Parameters**

Model type	Parameter form	Model expression	AIC	SC	L-Statistic
PDL	$PDL(3,2)$	$\ln gff_t = -0.018 + 1.581 \ln gff_{t-1} - 0.700 \ln gff_{t-2} + 0.133 \ln gff_{t-3} + \varepsilon_t$	-2.984	-2.799	50.25
ARCH	$ARCH(1)$	$\begin{cases} \ln gff_t = 1.339 \ln gff_{t-1} - 0.327 \ln gff_{t-2} + \varepsilon_t \\ \sigma_t^2 = 0.002 + 0.399 \varepsilon_{t-1}^2 \end{cases}$	-2.714	-2.531	47.42
ARIMA	$ARIMA(2,1,2)$	$\begin{cases} (1 - 0.707L - 0.048L^2) d \ln gff_t = 0.147 + (1 - 0.871L - 0.104L^2) \varepsilon_t \\ d \ln gff_t = \ln gff_t - \ln gff_{t-1} \end{cases}$	-3.433	-3.201	58.21

The results show that the AIC value and the SC value of each model are small, and the L-Statistic value is large enough to prove that the model parameters are valid and the valuations are reasonable. The residual equation is established by linear hypothesis test of the residuals of each model. The test results are shown in Table 2.

**Table 2. Residual Equation and Linear Hypothesis Test**

Residual equation	$R^2$	$\bar{R}^2$	t-statistic		F-statistic
			Variable $c$	Variable $t$	
$\varepsilon_t\_PDL$	0.602%	-2.826%	-0.379 (0.707)	0.419 (0.678)	0.176 (0.678)
$\varepsilon_t\_ARCH$	13.251 %	10.359%	-2.427 (0.021)	2.141 (0.041)	4.583 (0.041)
$\varepsilon_t\_ARIMA$	0.004%	-3.44%	0.085 (0.933)	0.036 (0.972)	0.001 (0.972)

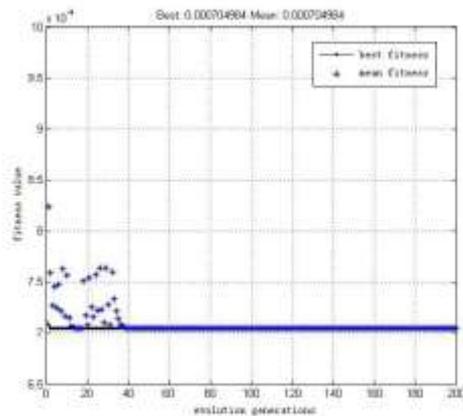
None of the residual equations pass by goodness of fit test, the determination coefficients  $R^2$  and  $\bar{R}^2$  are very small. For  $\varepsilon_t\_PDL$  and  $\varepsilon_t\_ARIMA$ , the basic parameters test and overall test conclusions are not significant.

Therefore, rejecting the null hypothesis, it is considered that the residuals of each individual forecasting model are nonlinear. Then, the nonlinear residual G-Verhulst model is established to correct the original forecasting model. The relevant conclusions are shown in Table 3.

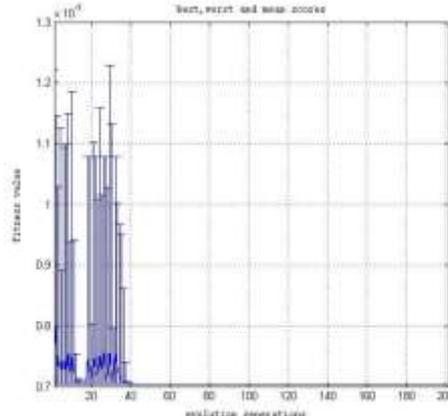
**Table 3. The Nonlinear Residual G-Verhulst Model Parameter Estimation and Error Correction**

Model type	First level parameters		Second level parameters					Total error before correction	Total error after correction
	<i>a</i>	<i>b</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>		
G-Verhulst-PDL	-0.129	-0.110	10.683	0.549	13.378	9.007	0.386	1.051	0.905
G-Verhulst-ARCH	-0.119	-0.080	34.042	0.941	30.978	38.921	0.912	1.404	1.343
G-Verhulst-ARIMA	-0.102	-0.089	4.457	0.351	7.357	2.882	0.196	0.838	0.500

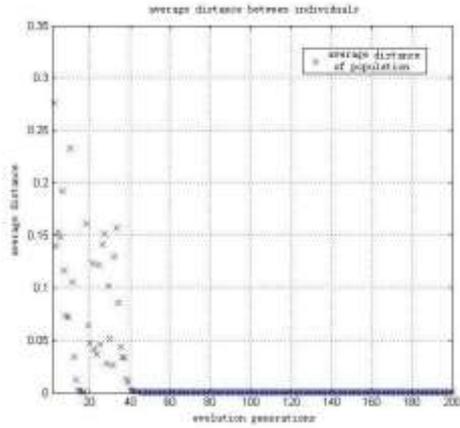
The data from 1978 to 2011 are analyzed. In Table 3, it shows the basic parameters of the whitened form prediction formula of G-Verhulst model. The total error after correction respectively is 0.905, 1.343 and 0.5, which is smaller than the corresponding original prediction model error. Based on the G-Verhulst-PDL model, the G-Verhulst-ARCH model and the G-Verhulst-ARIMA model, selecting the data from 1978 to 2011 as the training samples, the EGA optimization algorithm is used to establish the G-Verhulst metering combination forecasting model and optimize the model parameters. Rolling verification method are used to test the final forecast results by using the data from 2012 to 2016 as the test sample. The main parameters of the EGA are as follows: population size  $Q=100$ , elite population size  $Q^*=10$ , population evolution generation threshold  $gen_{max}=200$ , genetic probability  $p_r=0.99$ , crossover probability  $p_c=0.8$  and mutation probability  $p_n=0.2$ , fitness function value deviation threshold value  $1 \times 10^{-100}$ . EGA operating environment is as follows: MATLAB R2014a mathematical software, windows 7 operating system, Intel(R) Core CPU P8700, frequency 2.53G, RAM 2.99G. Under the above setting, when the objective function is  $\phi_1 = S_1$ , the best individual value  $\sigma_1=0.113, \sigma_2=0.231, \sigma_3=0.656$  and the best fitness value  $7.049 \times 10^{-4}$  are calculated. Figure 3 to Figure 6 show the change of relevant indicator data during the calculation.



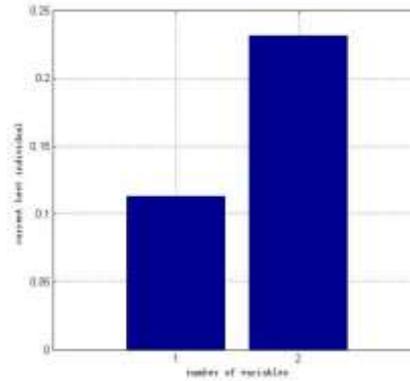
**Figure 3. EGA Convergence Scatter Plot in the Same Generation**



**Figure 4. Comparison of Fitness Values of the Best and the Worst Individual**



**Figure 5. Scatter Plot of Average Distance of Population**



**Figure 6. Best Individual Value (Evolution of 200 generations)**

It can be seen that the best fitness value of the population converges rapidly as the number of increases, and the individual spacing of population keeps decreasing, which indicates that the overall genetic status of the population is effectively improved. After about 40 generations, the results are satisfactory, the whole running time is short and the fitting result is quite accurate. Similarly, according to the weight distribution based on objective functions of  $\phi_2 = S_2$ ,  $\phi_3 = S_3$ ,  $\phi_4 = S_4$  and  $\phi_5 = S_1 + S_2 + S_3 + S_4$ , we can get a set of Pareto optimal solutions under multi-criteria objective decision, which belongs to the typical NP problem. The results are shown in Table 4.

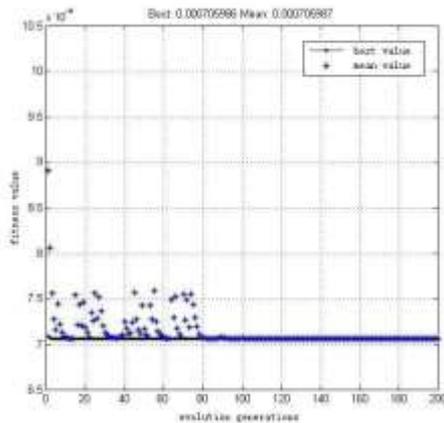
**Table 4. Pareto Optimal Solution under Multi-criteria Objective Decision**

Sequence number	Objective function	Optimal weight			Combination forecasting criterion			
		$\sigma_1$	$\sigma_2$	$\sigma_3$	MSE	MAE	MAPE	Theil
①	$\phi_1 = S_1$	0.113	0.231	0.656	$7.049 \times 10^{-4}$	$1.660 \times 10^{-2}$	$2.739 \times 10^{-3}$	$1.526 \times 10^{-3}$
②	$\phi_2 = S_2$	0.118	0.017	0.865	$7.723 \times 10^{-4}$	$1.516 \times 10^{-2}$	$2.505 \times 10^{-3}$	$1.598 \times 10^{-3}$
③	$\phi_3 = S_3$	0.128	—	0.872	$7.811 \times 10^{-4}$	$1.516 \times 10^{-2}$	$2.505 \times 10^{-3}$	$1.607 \times 10^{-3}$
④	$\phi_4 = S_4$	0.114	0.232	0.654	$7.050 \times 10^{-4}$	$1.661 \times 10^{-2}$	$2.741 \times 10^{-3}$	$1.526 \times 10^{-3}$
⑤	$\phi_5 = S_1 + S_2 + S_3 + S_4$	0.142	0.042	0.816	$7.531 \times 10^{-4}$	$1.517 \times 10^{-2}$	$2.513 \times 10^{-3}$	$1.578 \times 10^{-3}$

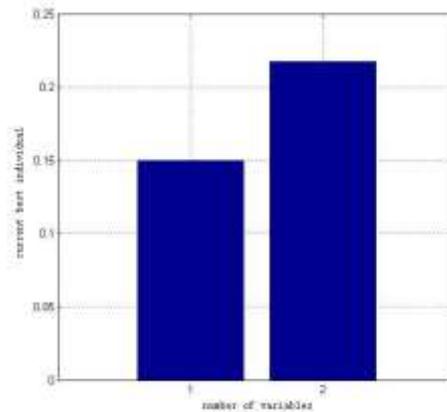
It is very easy to find that the error of each evaluation criterion under different objective functions is small, which means that the G-Verhulst metering combination forecasting model has satisfied the expectation requirement, the model has good fitting effect and good stability.

### 5.3. Comparison of Algorithms and Evaluation of Prediction

It is shown that the traditional SGA converges to the optimal value only after it has evolved to about the 80th generation, and the convergence rate of only half of the EGA. SGA's mean square error value is greater than EGA to  $7.060 \times 10^{-4}$ , indicating convergence performance is not as good as EGA. It can be seen, EGA's comprehensive performance is more superior.



**Figure 7. SGA Convergence Scatter Plot**



**Figure 8. Best Individual Value (Evolution of 200 Generations)**

Select the G-Verhulst metering combination forecasting model form ⑤, which has the best comprehensive performance, and compare the forecasting results with other forecasting models. I represents the PDL model, II represents the ARCH model, III represents the ARIMA model, IV represents the G-Verhulst-PDL model, V represents the G-Verhulst-ARCH model, VI represents the G-Verhulst-ARIMA model, VII represents the nonlinear residual G-Verhulst metering combination forecasting model based on EGA. The prediction samples are the logarithmically processed data for defense expenditure from 2012 to 2016, and the forecasting results are shown in Table 5.

**Table 5. Forecasting Results and Error Contrast of Each Forecasting Model**

Year	Actual value	I		II		III		IV		V		VI		VII	
		Forecast value	Error	Forecast value	Error	Forecast value	Error	Forecast value	Error	Forecast value	Error	Forecast value	Error	Forecast value	Error
2012	8.809	8.859	0.57%	8.854	0.51%	8.853	0.50%	8.834	0.28%	8.84	0.35%	8.825	0.18%	8.82	0.13%
2013	8.91	8.949	0.44%	9.014	1.17%	9.002	1.03%	8.925	0.17%	9.002	1.03%	8.974	0.72%	8.968	0.65%
2014	9.122	9.053	-0.76%	9.18	0.64%	9.151	0.32%	9.084	-0.42%	9.169	0.52%	9.122	0.00%	9.111	-0.12%
2015	9.236	9.173	-0.68%	9.35	1.23%	9.299	0.68%	9.152	-0.91%	9.339	1.12%	9.27	0.37%	9.256	0.22%
2016	9.363	9.214	-1.59%	9.523	1.71%	9.447	0.90%	9.232	-1.40%	9.514	1.61%	9.418	0.59%	9.391	0.30%

From the results in Table 5, we can see that the integrated forecast error of model VII is the smallest, and therefore this forecasting model is the best. At the same time, the models IV, V and VI take into account the residual effect, so the prediction effect is better than the basic prediction models I, II and III. This conclusion is consistent with the theory of combined prediction error correction.

## 6. Conclusions

In this paper, we propose a small sample nonlinear residual G-Verhulst combination forecasting model for the small sample nonlinear time sequence of China's defense expenditure, design the solution method and solution steps of multi-criteria objective decision problem based on EGA optimization. Using this model, the time sequence of national defense expenditure is tested empirically, and the basic parameters of the model are estimated. On the basis of this, the convergence rate and convergence performance of SGA and EGA are compared, which proves that EGA is more superior. Compared with other prediction methods, the small sample nonlinear residual G-Verhulst combination forecasting model based on EGA has smaller prediction error, higher prediction accuracy and better prediction effect.

## References

- [1] Z. S. Zhang and Z. Fan. "Cointegration theory and volatility models", Tsinghua University Press, Beijing, (2007).
- [2] L. Zhang and B. Liu, "Residuals statistics-based additive outlier detection algorithm for time series", *Application of Electronic Technique*, vol. 9, no. 43, (2015).
- [3] X. M. Wang and Z. M. Yao. "Dmitri L. Residual-based estimation of peer and link lifetimes in P2P networks", *Transactions on Networking*, vol. 3, no. 17, (2009).
- [4] C. H. Lin and D. Y. Shen. "Secret: a selective error correction framework for refresh energy reduction in drams", *ACM Transactions on Architecture and Code Optimization*, vol. 2, no. 12, (2015).
- [5] D. Bertozzi and P. Olivo. "Zambelli C. Flares: an aging aware algorithm to autonomously adapt the error correction capability in NAND flash memories", *ACM Transactions on Architecture and Code Optimization*, vol. 3, no. 11, (2014).
- [6] J. M. Bate and C. W. J. Granger, "Combination of forecasts", *Operations Research Quarterly*, vol. 4, no. 20, (1969).
- [7] L. Christiane, "Combinations of time series forecasts: when and why are they beneficial", Bournemouth University, Bournemouth, (2010).
- [8] H. A. Li, B. S. Kang and J. Zhang. "Combined forecasting model of cloud computing resources based on improved encompassing tests and IOWA operator", *Journal of Computational Information Systems*, vol. 7, no. 9, (2013).
- [9] N. Harvey and C. Harries. "Effects of judges' forecasting on their later combination of forecasts for the same outcomes", *International Journal of Forecasting*, vol. 3, no. 20, (2004).
- [10] L. Zhao and M. K. Tang, "Prediction of fire accidents based on GM(1,1) model with residual error correction", *Mathematics in Practice and Theory*, vol. 18, no. 42, (2012).
- [11] T. X. Yao, J. Forrest and Z. W. Gong, "Generalized discrete GM(1,1) model", *Grey Systems: Theory and Application*, vol. 1, no. 2, (2012).
- [12] M. X. He, "The new approach of establishing GM(1,N) predicting model", *System Sciences and Comprehensive Studies of Agriculture*, vol. 4, no. 13, (1997).
- [13] W. J. Qiu and S. F. Liu. "Dispersed structure solve of model GM(1,N)", *Systems Engineering and Electronics*, vol. 11, no. 28, (2006).
- [14] T. L. Tien. "The indirect measurement of tensile strength of material by the grey prediction model GMC(1,N)", *Measurement Science Technology*, vol. 4, no. 16, (2005).
- [15] T. L. Tien, "A research on the grey prediction model GM(1,N)", *Applied Mathematics and Computation*, no. 218, (2012).
- [16] F. Gao, "Study of combination forecast and application in traffic flow forecast", Shandong University of Science and Technology, Jinan, (2008).
- [17] Y. M. Ding, "The application and some discussion of combination forecast about chinese stock market", Huazhong University of Science and Technology, Wuhan, (2005).
- [18] M. X. He, "The application of combination forecast method in analysis and forecasting of tourism economy", *Mathematics in Practice and Theory*, vol. 3, no. 40, (2010).
- [19] J. L. Deng, "The basis of grey theory", Huazhong University of Science and Technology Press, Wuhan, (2002).
- [20] G. Soremekun, Z. Gurdal and R. T. Haftka, "Watson L T. Composite laminate design optimization by genetic algorithm with generalized elitist selection", no. 79, (2001).
- [21] J. L. Kim, "Permutation based elitist genetic algorithm using serial scheme for large-sized resource-constrained project scheduling", *Proceedings of the 2007 Winter Simulation Conference*, IEEE, USA, (2007).
- [22] H. T. Ba and C. Shi, "A route planning algorithm for aircraft based on elitist genetic algorithm", *Command Control & Simulation*, vol. 4, no. 32, (2010).

## Authors



**Kan Zhang**, he was born in 1985. He received his master degree of defence economic in Naval University in 2009. He is a PH.D. candidate in Electronic Engineering Academy and he is also a lecturer in Dept. of Economics & Equipment Management. His research interests include system engineering, defense economic, modeling and simulation of complex economic system.

Email: zhangkanwh@163.com.



**Bao-ping Liu**, he was born in 1963. He received his PH.D. in Huazhong University of Science and Technology in 2005. He is an associate professor in Dept. of Economics & Equipment Management. His research interests include the theory and application of nonlinear dynamics, economic analysis and forecasting.

Email: liubpwh@yahoo.com.cn.



**Dong Huang**, he was born in 1977. He received his PH.D. in Wuhan University in 2008. He is an associate professor in Dept. of Economics & Equipment Management. His His research interests include data mining and analysis of economy, defense economic.

Email: 13397191239@189.cn.