Multi-variable Grey Model based on Genetic Algorithm and its Application in Urban Water Consumption

Han Yan*, Xu Shiguo

School of Civil & Hydraulic Engineering, Dalian University of Technology, 2 Linggong, Ganjingzi, Dalian, 116024 China, hanyan78@163.com

ABSTRACT: Owing to the influence of economy, population, standard of living and so on, the urban water consumption possesses certain characteristics of grey. As the expansion and complement of grey system (GM(1,1)) model, the multi-variable grey model (MGM(1,n)) reveals the relationship of restriction and stimulation between variables. Genetic algorithm possesses the whole optimal and parallel characteristics. In this paper, through using the genetic algorithm, the parameter q of MGM(1,n) model has been optimized, and a multi-variable grey model (MGM(1,n,q)) based on genetic algorithm has been built. The model has been validated after examining the urban water consumption in Dalian city from 1990 to 2003. The result indicates that the multi-variable grey model (MGM(1,n,q)) based on genetic algorithm is better than MGM(1,n) model, and the MGM(1,n) model is better than GM(1,1). [Nature and Science. 2007;5(1):18-26].

KEYWORDS: Grey system; MGM(1,n,q); genetic algorithm; urban water consumption

1. INTRODUCTION

With the rapid development of economy, persistently increase of population and constantly improvement of standard living, the urban water demand has been constantly rising, but the amount of water supply is limited. The conflict between urban water demand and supply is gradually exacerbating, and the settlement of urban water shortage is the austere challenge of urbanization development. The forecast of urban water consumption is the premise and basic to plan and manage water resources. The results of forecast directly influence the reliability and practicability of assignment decision-making of water resources system, and also directly influence the sustainable consumption of urban water resources and sustainable development of social economy [1]. At present, there are many methods to predict the urban water consumption, such as the regress analytical method, the exponent smoothness method, the ration of water consumption method, the grey system forecast method, and the artificial neural network (ANN) method. ANN method needs long series of data, so it is difficult to predict due to the lack of historical data. The grey system theory takes uncertainty system as the study object, such as small sample and poor information. Through the creation and development of partial known information, the valuable information is picked out, so the operation behavior and evolvement rule are correctly described and effectively monitored and controlled. Practice has proved the grey system model needs less information but the precision of result is better, and can preferably reflect the practical condition of the system [2]. The grey system has been extensively applied in producing, engineering, science and technology. Owing to the influence of economy, population, living standard and so on, the urban water consumption has certain grey characteristics. Especially when the longer series of reliable data are unavailable, the grey system model is an available method to predict the urban water consumption [3].

2. MULTI-VARIABLE GREY MODEL MGM(1,n) [4]

Grey system model GM(1,1) is disposed through one accumulation of single variable time series $\{X_i\}\ (i=1,2,\cdots,n)$. Through first-order differential equation, the intrinsic rule of generating sequence can be revealed. The GM(1,1) can only be applied to modeling and predicting of single time series data.

$$\frac{dX^{(1)}}{dt}+aX^{(1)}=b \tag{1}$$

Only using the single time series data, the grey system model GM(1,1)can not reflect the influence and development between each other; however, the GM (1,n) model can mainly be applied to describing the correlation of variable between each other, not to predicting. The GM (1,1) model has been detailed described in reference [5]. In this paper, the Multi-variable Grey Model MGM(1,n) had been introduced. MGM(1,n) model is "n" variables first-order differential equations, it is the natural expansion of GM(1,1) model in "n" variables, not simple combination of GM(1,1) model. The first-order differential equation of MGM(1,n) model can be written as follows:

$$\begin{cases} \dfrac{dx_1^{(1)}}{dt}=a_{11}x_1^{(1)}+a_{12}x_2^{(1)}+\cdots+a_{1n}x_n^{(1)}+b_1 \\ \dfrac{dx_2^{(1)}}{dt}=a_{21}x_1^{(1)}+a_{22}x_2^{(1)}+\cdots+a_{2n}x_n^{(1)}+b_2 \\ \qquad\vdots \\ \dfrac{dx_n^{(1)}}{dt}=a_{n1}x_1^{(1)}+a_{n2}x_2^{(1)}+\cdots+a_{nn}x_n^{(1)}+b_n \end{cases} \tag{2}$$

The first-order differential equation of MGM(1,n) model is as follows:

$$\frac{dX^{(1)}}{dt}=AX^{(1)}+B \text{ or } \frac{dX^{(1)}}{dt}-AX^{(1)}=B \tag{3}$$

The sequence time response formula is

$$X_{(t)}^{(1)}=e^{A(t-1)}X_{(0)}^{(1)}+A^{-1}(e^{A(t-1)}-I)\bullet B \qquad (t=1,2,\cdots,n) \tag{4}$$

Where $e^{At}=I+At+\dfrac{A^2}{2!}t^2+\cdots=I+\sum_{k=1}^{\infty}\dfrac{A^k}{k!}t^k$

The parameter A and B can be estimated by least-square method:

$$\hat{a}_i=\begin{bmatrix}\hat{a}_{i1}\\ \hat{a}_{i2}\\ \vdots\\ \hat{a}_{in}\\ \hat{b}_i\end{bmatrix}=(L^TL)^{-1}L^TY_i \qquad (i=1,2,\cdots,n \quad) \tag{5}$$

$$\hat{A} = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ \hat{a}_{21} & \hat{a}_{22} & \cdots & \hat{a}_{2n} \\ \vdots & \vdots & & \vdots \\ \hat{a}_{n1} & \hat{a}_{n2} & \cdots & \hat{a}_{nn} \end{bmatrix}, \text{ and } \quad \hat{B} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_n \end{bmatrix}$$

where

$$L = \begin{bmatrix} \frac{1}{2}(x_1^{(1)}(2)+x_1^{(1)}(1)) & \frac{1}{2}(x_2^{(1)}(2)+x_2^{(1)}(1)) & \cdots & \frac{1}{2}(x_n^{(1)}(2)+x_n^{(1)}(1)) & 1 \\ \frac{1}{2}(x_1^{(1)}(3)+x_1^{(1)}(2)) & \frac{1}{2}(x_2^{(1)}(3)+x_2^{(1)}(2)) & \cdots & \frac{1}{2}(x_n^{(1)}(3)+x_n^{(1)}(2)) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{1}{2}(x_1^{(1)}(m)+x_1^{(1)}(m-1)) & \frac{1}{2}(x_2^{(1)}(m)+x_1^{(1)}(m-1)) & & \frac{1}{2}(x_n^{(1)}(m)+x_n^{(1)}(m-1)) & 1 \end{bmatrix}$$

and $\quad Y_i = (x_i^{(0)}(2), x_i^{(0)}(3), \cdots, x_i^{(0)}(m))^T$

3. THE ESTABLISHMENT OF MGM(1,n,q) MODEL

Since original data are mainly time-series data in practice, the derivative can be transformed into a forward difference equation [6], such as

$$\frac{X_{t_1}^{(1)} - X_{t_2}^{(1)}}{t_1 - t_2} - AX_{t_2}^{(1)} = B$$

In fact, $t_1 - t_2 = 1$, so we can get

$$X_{t+1}^{(1)} - X_t^{(1)} - AX_t^{(1)} = B \tag{6}$$

in the same way, a backward difference equation is

$$X_t^{(1)} - X_{t-1}^{(1)} - AX_t^{(1)} = B \text{ or } X_{t+1}^{(1)} - X_t^{(1)} - AX_{t+1}^{(1)} = B \tag{7}$$

Different sequence satisfies different equation. Some satisfy the Eq.(6), while others satisfy the Eq.(7). The common difference equation is

$$X_{t+1}^{(1)} - X_t^{(1)} = B + qAX_t^{(1)} + (1-q)AX_{t+1}^{(1)} \tag{8}$$

The Eq.(8) is the common difference equation of MGM(1,n). The building model is MGM(1,n,q) by Eq.(8). When q=0.5, the MGM(1,n,q) model turns to the GM(1,1) model. For the any q_0 the background matrix is

$$L_0 = \begin{bmatrix} (q_0 x_1^{(1)}(2)+(1-q_0)x_1^{(1)}(1)) & (q_0 x_2^{(1)}(2)+(1-q_0)x_2^{(1)}(1)) & \cdots \\ (q_0 x_1^{(1)}(3)+(1-q_0)x_1^{(1)}(2)) & (q_0 x_2^{(1)}(3)+(1-q_0)x_2^{(1)}(2)) & \cdots \\ \vdots & \vdots & \\ (q_0 x_1^{(1)}(m)+(1-q_0)x_1^{(1)}(m-1)) & (q_0 x_2^{(1)}(m)+(1-q_0)x_1^{(1)}(m-1)) & \cdots \end{bmatrix}$$

$$\begin{matrix} (q_0 x_n^{(1)}(2)+(1-q_0)x_n^{(1)}(1)) & 1 \\ (q_0 x_n^{(1)}(3)+(1-q_0)x_n^{(1)}(2)) & 1 \\ \vdots & 1 \\ (q_0 x_n^{(1)}(m)+(1-q_0)x_n^{(1)}(m-1)) & 1 \end{bmatrix} \quad (9)$$

$$\hat{a}_i = \begin{bmatrix} \hat{a}_{i1} \\ \hat{a}_{i2} \\ \vdots \\ \hat{a}_{in} \\ \hat{b}_i \end{bmatrix} = (L_0^T L_0)^{-1} L_0^T Y_i ,\quad i = 1,2,\cdots,n \quad (10)$$

According to the above, if only given q_0, the $\hat{A}$ and $\hat{B}$ may be obtained by Eq.(9) and Eq.(10), then the $X_t^{(1)}$ may be obtained by Eq.(4). When the original series $X_i^{(0)}$ are given, the parameter q is the only factor that influences the precision of MGM(1,n) model. The relationship is very non-linear between q and errors. If the genetic algorithm is adopted, an ideal value about the parameter q may be obtained. In this paper, the MGM (1,n) model is combined with the genetic algorithm (GA), which is called MGM (1,n,q) model.

4. THE SOLUTION OF MGM(1,n,q) MODEL WITH GA

4.1 The genetic algorithm

The genetic algorithm(GA)[7,8] is an adaptive whole search and probability optimization arithmetic, which simulates the heredity and evolution of biology in environments. This arithmetic was initially presented by John Holland, professor of Michigan University U.S. in 1975. Through genetic operation of selection, and mutation to current population, the new generation is created and gradually evolves to optimal state. Only the evaluation function is used during seeking optimization and the differentiability of objective function is not required, can the genetic algorithm be whole, parallel, speed, adaptable and robust, so it has been extensively applied in the field of function optimization, production control, automation control, image disposal, artificial life and so on.

4.2 The basic steps of GA

1) Encoding. The $q \in [0,1]$ can be expressed by a binary cluster. The length of q (chromosome) can be determined by the precision of q .

2) Initialization of the population. N numbers selected from 0 to 1 at random are regarded as initial population. (N is the number of population).

3) The fitness evaluation of individual. The fitness function indicates the degree of adaptation capability to the environment, which is related with objective function. The fitness value of the No. i individual

$Fit(q(i,k))$ (k is the iterative times) may be calculated by using the following formula.

$$Fit(q(i,k)) = \begin{cases} C_{\max} - fit(q(i,k)) & , fit(q(i,k)) \prec C_{\max} \\ 0 & , fit(q(i,k)) \geq C_{\max} \end{cases}$$

The objective function can be expressed by the square sum of error; $fit(q(i,k)) = \sum_{i=1}^{N} (\hat{x}_i^{(0)} - x_i^{(0)})^2$ .where $C_{\max}$ may be the import parameter, the maximum of $fit(q(i,k))$ until now or the maximum of $fit(q(i,k))$ at current population or in latest several generation populations.

4) Selection. The survival probability of the i individual in the k generation is

$$p_i^{(k)} = \frac{Fit(q(i,k))}{\sum_{j=1}^{N} Fit(q(j,k))}$$

We can choose a strategy (such as roulette), so the selected probability of the i individual is $p_i^{(k)}$. The higher fitness of individual, the more chance of the individual can be selected as the new individual. The lower fitness of the individual, the fewer chance of individual can be selected as the new individual, and the more probability of being eliminated.

5) Crossover. A pair of individuals for crossover is selected stochastically. The simplest method of crossover is to select a truncation point stochastically, split each gene chain into two sections at this point, and then exchange their tails, for instance:

$$\begin{array}{l|l} 1100110011 & 10001 \\ 1011100101 & 01010 \end{array} \rightarrow \begin{array}{l} 1100110011\ 01010 \\ 1011100101\ 10001 \end{array}$$

The crossover embodies the process of exchanging information in course of biology heredity.

6) Mutation. Several individuals are selected from the population with probability (p_m). For the selected individual, a bit is selected randomly for mutation, namely turns 1 to 0 (or 0 to 1). For example 10110[1]011 is changed to 101100011. The mutation embodies the accidental gene mutation in the course of biology heredity.

7) Evolutionary iteration. The filial generation from the previous is regarded as the new one, then repeat the 3) step to 6) step until an acceptable solution will be found or the reserved iteration times fulfilled.

Through using the genetic algorithm, the ideal q_0 can be obtained; the matrix L_0 can be obtained by putting the q_0 into Eq.(9); then the value of $\hat{A}$ and $\hat{B}$ can also be obtained by the Eq.(10).

The calculation process of MGM(1,n,q) model is

$$\hat{X}^{(1)}(t) = e^{\hat{A}(t-1)} X^{(1)}(1) + \hat{A}^{-1}(e^{\hat{A}(t-1)} - I) \bullet \hat{B} \qquad (t = 1,2,\cdots,n)$$

$$\hat{X}^{(0)}(1) = X^{(0)}(1)$$

$$\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k) - \hat{X}^{(1)}(k-1) \qquad (k = 2,3,\cdots n)$$

The relative error and the square sum of errors can be used to evaluate the MGM(1,n,q) model forecasting. The relative error is $\frac{\left|\hat{x}_i^{(0)} - x_i^{(0)}\right|}{x_i^{(0)}}$; the square sum of errors is $\sum_{i=1}^{n} (\hat{x}_i^{(0)} - x_i^{(0)})^2$.

5. CASE STUDY

Owing to the influence of economy, population, standard of living and so on, the urban water consumption possesses certain characteristics of grey. Due to the need of less information and higher precision, the grey system can preferably reflects system practical condition. The following is an analysis and calculation of urban water consumption for several years in Dalian city. Due to the available data of water consumption are scarce in Dalian city, only the data from 1990 to 2003 can be analyzed

Table 1. The Statistic of urban water consumption from 1990 to 2003 in Dalian

Year	Urban water consumption ($10^4 m^3$)	Urban population (10^4 people)
1990	20826	239.64
1991	19305	241.57
1992	25107	244.94
1993	25416	248.67
1994	27485	252.35
1995	30354	254.74
1996	30910	257.23
1997	32784	259.71
1998	27105	262.40
1999	33142	264.17
2000	32249	267.78
2001	31868	270.68
2002	33262	273.23
2003	35460	274.78

The data from 1990 to 2000 year are regarded as the basic, and the data from 2001 to 2003 as the test. According to the above theory, we perform a simulation forecast. In course of genetic algorithm, the binary coding has been adopted, the number of population is 10, the length of coding is 20, the probability of crossover is 0.95, and the probability of mutation is 0.08. After optimization of genetic algorithm, the parameter q=0.456534. At last, we compare the simulation value and the errors of GM(1,1), MGM(1,n) and MGM(1,n,q). The result is shown in table 2.

Table 2. The analysis of simulation value and errors about three models

year	water consumption (10^4m^3)	GM(1,1)		MGM(1,n)		GM(1,n,q)	
		Simulation value (10^4m^3)	Relative errors (%)	Simulation value (10^4m^3)	Relative errors (%)	Simulation value (10^4m^3)	Relative errors (%)
1990	20826	20826.00	0	20826.00	0	20826.00	0
1991	19305	23538.98	21.93	19847.87	2.81	19741.59	2.26
1992	25107	24506.24	2.39	23939.15	4.65	23866.78	4.94
1993	25416	25513.24	0.38	26536.8	4.41	26495.03	4.24
1994	27485	26561.62	3.36	28228.06	2.70	28210.75	2.64
1995	30354	27653.08	8.90	29369.68	3.24	29370.61	3.24
1996	30910	28789.39	6.86	30178.46	2.37	30192.29	2.32
1997	32784	29972.40	8.58	30786.09	6.09	30808.66	6.02
1998	27105	31204.01	15.12	31272.61	15.38	31300.80	15.48
1999	33142	32486.24	1.98	31686.65	4.39	31718.21	4.29
2000	32249	33821.15	4.88	32057.78	0.59	32091.12	0.49
2001	31868	35210.92	10.49	32403.97	1.68	32437.97	1.79
2002	33262	36657.79	10.21	32736.13	1.58	32770.04	1.48
2003	35460	38164.12	7.63	33060.92	6.77	33094.21	6.67
Mean relative error(%)		7.33		4.05		3.99	

From the result in table 2, we can draw the conclusion that the precision of MGM (1,n) is much higher than the GM (1,1) due to the restriction and stimulation between variables. The MGM (1,n,q) is higher than the MGM (1,n) due to the parameter q being wholly optimized. The mean relative error is enhanced from 7.33% to 3.99%. The curves of modeling and forecast of three models are shown in Figure 1.

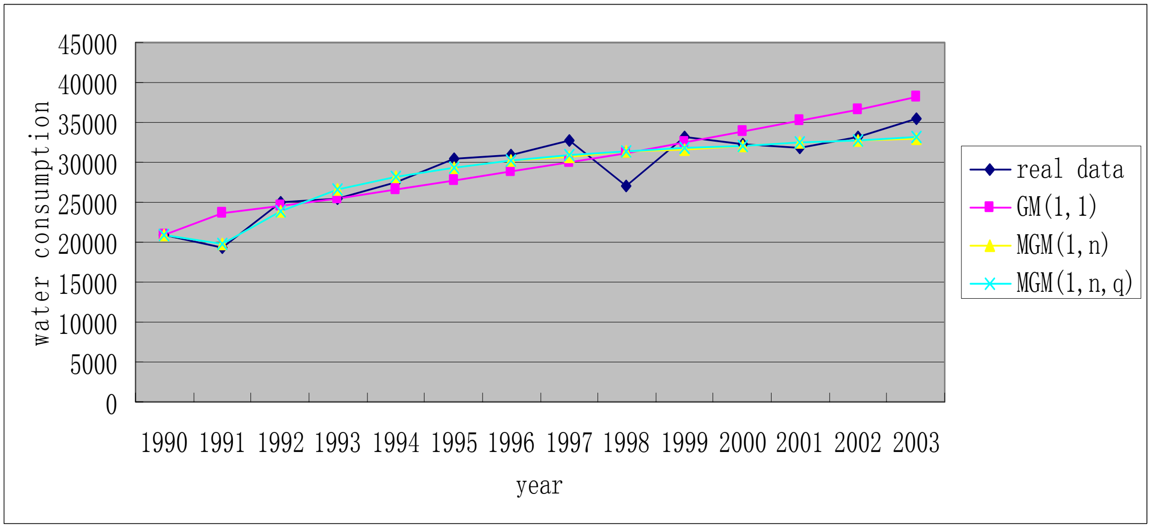

Figure 1. The curves of simulation and forecast of models

6. CONCLUSIONS

As the expansion and complement of GM (1,1) model, the MGM (1,n) can reflect the relationship of restriction and stimulation between variables. According to the characteristics of whole search and connotative parallel calculation, the genetic algorithm is adopted and combined with the multi-variable grey model (MGM (1,n)) well, and a Multi-variable model (MGM (1,n,q)) based on Genetic Algorithm has been established. Owing to the influence of economy, population, standard of living and so on, the urban water consumption possesses certain characteristics of grey. The grey system model is an available method to predict the urban water consumption. Taking the data of urban water consumption in Dalian city from 1990 to 2003 as example, the model is proved. The result indicates that the MGM(1,n,q) model is better than MGM(1,n) model, and the MGM(1,n) model is better than GM(1,1). At the same time, in the course of modeling of multi-variable grey model, if the choice of variable is improper, the morbidity of matrix or inverse matrix would occur. How to choose the proper variable is difficult in the course of applying, which needs to be further researched.

Correspondence to:
Han Yan
School of Civil & Hydraulic Engineering
Dalian University of Technology
Linggong, Ganjingzi, Dalian, Liaoning 116024
China
Telephone: 086-0411-84707340
Email: hanyan78@163.com

Received: March 3, 2007

REFERENCES

1. LI Lin, ZUO Qi-ting. Comparative research on predicting method and application for city water consumption [J]. Journal of Water Resources & Water Engineering, 2005, 16(3):6-10
2. LIU Si-feng, DANG Yao-guo, etc, FANG Zhi-geng. Theory and application of grey system [M]. BeijingScience Press, 2005.
3. LOU Yu, ZHANG Qing, LIU Guo-hua, etc. Application of grey multi-variable forecasting model in predicting urban water consumption [J]. Water Resources Protection, 2005, 21(1):11-13.
4. ZHAI Jun, SHENG Jian-ming, FENG Ying-jun. The grey Model MGM(1,n) and Its Application [J]. Systems Engineering——Theory & Practice, 1997(5):109-113.
5. DENG Ju-long. Grey System Theory [M]. Wuhan: Huazhong University of Science and Technology Press, 1990
6. XIE Kai-gui, LI Chun-yan, ZHOU Jia-qi. Grey model (GM(1,1,λ) based on genetic algorithm [J]. Journal of Systems Engineering, 2000, 15(2):168-172
7. Holland J H. Genetic algorithm[J]. Scientific American,1992(4):40-45
8. YAO Wen-jun. Study on back propagation network optimization with genetic algorithms [J]. Journal of Wuhan Institute of Chemical Technology, 2004, 26(3):72-75