Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# Utility water supply forecast via a GM (1,1) weighted Markov chain*

Yi-mei TIAN[†1], Hai-liang SHEN[†‡2], Li ZHANG[1], Xiang-rui LV[1]

*([1]College of Environmental Science & Engineering, Tianjin University, Tianjin 300072, China)*

*([2]School of Engineering, University of Guelph, Guelph, Ontario, N1G 2W1, Canada)*

[†]E-mail: ymtian_2000@yahoo.com.cn; shenh@uoguelph.ca

Received Dec. 10, 2009; Revision accepted June 1, 2010; Crosschecked Aug. 11, 2010

**Abstract:** This paper describes the procedure of using the GM (1,1) weighted Markov chain (GMWMC) to forecast the utility water supply, a quantity that usually has significant temporal variability. The GMWMC is formulated into five steps: (1) use GM (1,1) to fit the trend of the data, and obtain the relative error of the fitted values; (2) divide the relative error into 'state' data based on pre-set intervals; (3) calibrate the weighted Markov chain model: herein the parameters are the pre-set interval and the step of transition matrix (TM); (4) by using auto-correlation coefficient as the weight, the Markov chain provides the prediction interval. Then the mid-value of the interval is selected as the relative error for the data. Upon combining the data and its relative error, the predicted magnitude in a specific time period is obtained; and, (5) validate the model. Commonly, static intervals are used in both model calibration and validation stages, usually causing large errors. Thus, a dynamic adjustment interval (DAI) is proposed for a better performance. The proposed procedure is described and demonstrated through a case study, which shows that the DAI can usually achieve a better performance than the static interval, and the best TM may exist for certain data.

**Key words:** Dynamic adjustment interval (DAI), Forecast, GM (1,1), Markov chain, Water supply
**doi:**10.1631/jzus.A0900754 **Document code:** A **CLC number:** TU991.31

## 1 Introduction

The amount of urban water supply is the basic scheduling data for water utilities. Its accuracy in prediction allows the water supply operation to be more economic and effective. The hourly water supply data are time series and have been studied by such methods as trigonometric function (Liu and Zhang, 2002) and neural networks (Zhou *et al.*, 2004). The trigonometric function accomplishes parameter estimation with multivariate statistical theory, but there may be multi-collinearity in actual application. For neural networks, over-fitting usually exists, which causes weak extrapolation capability; in addition, the number of hidden layers is often difficult to determine (Liao and Tsao, 2004).

The grey system theory was proposed by Deng (1982). It is widely used in prediction. The grey system modeling is focused on a small dataset with the objective being to build a grey differential equation (Deng, 2005). GM (1,1) is the most commonly used prediction model of grey models, and can be applied in prediction when the original data sequence is smooth. It does not, however, work well when dealing with data with large variability. The prediction theory of Markov chain is Markov process. Markov chain is applied when data have large variability, using a one step or multi-step transition matrix (TM) to find the internal regularity of the data. Therefore, the combination of GM (1,1) and Markov chain can achieve a better predictive capability. Currently, the GM (1,1) weighted Markov chain (GMWMC) is used in such areas as highway freight prediction (Gai and Pei, 2003), network traffic prediction (Yao *et al.*, 2007), bridge technical conditions (Geng *et al.*, 2007), elec-

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 50778121), the National Basic Research Program of China (No. 2007CB407306-1), and the National Water Pollution Control and Management of Science and Technology Project of China (No. 2008ZX07317-005),

tric power requirement forecasts (He and Huang, 2005), prediction of the number of international airlines (Li *et al.*, 2007), and prediction of machining accuracy (Tien, 2005). Currently, after obtaining the sequence for building the TM, a static interval is applied for assigning the discrete state. In this study, GMWMC is applied due to the large variability of water supply data, and an improved method to static interval, a dynamic adjustment interval (DAI) method, is proposed to improve the prediction performance.

## 2 Methodology

Herein, GM (1,1) is applied as the base model to extract the trend of the time series, and the relative error of the fitted value is employed to construct the weighted Markov chain model.

### 2.1 Steps in establishing the GM (1,1) model

1. Generate the first order cumulative sequence

The original water supply time series $X^{(0)}=\left(x_{(1)}^{(0)},x_{(2)}^{(0)},\ldots,x_{(n)}^{(0)}\right)$ is cumulated to reduce fluctuation in the dataset. A new sequence $X^{(1)}$ is generated by the first accumulated generating operation (1-AGO):

$$X^{(1)}=\left(x_{(1)}^{(1)},x_{(2)}^{(1)},\ldots,x_{(n)}^{(1)}\right),$$

where $x_{(k)}^{(1)}=\sum_{i=1}^{k}x_i^{(0)}$, $k$=1,2,…,$n$.

2. Establish the GM (1,1) model

$$\frac{\mathrm{d}x}{\mathrm{d}t}+ax=b, \tag{1}$$

where $a$ and $b$ are parameters required to be estimated.

3. Solve the model

The time response equation is presented as follows:

$$X_{(k+1)}^{(1)}=\left[X_{(1)}^{(0)}-\frac{b}{a}\right]\mathrm{e}^{-ak}+\frac{b}{a}. \tag{2}$$

4. Reduce data

$$X_{(k+1)}^{(0)}=X_{(k+1)}^{(1)}-X_{(k)}^{(1)}=(1-\mathrm{e}^{a})\left[X_{(1)}^{(0)}-\frac{b}{a}\right]\mathrm{e}^{-ak}, \tag{3}$$

where $x_{(k+1)}^{(0)}$ is the ($k$+1)th prediction.

### 2.2 Weighted Markov chain

The weights are applied to estimate the relative importance of each TM.

The random sequence defined in the probability space ($\Omega$,$F$,$P$) is $\{X(t), t\in T\}$ (where $T=\{t_0,t_1,\ldots,t_{k+1}\}$), and has the state space of $I=\{0,1,\ldots,k+1\}$. If for any positive integer $k$, any $t_i\in T$, $t_i<t_{i+1}$, $i$=0,1,…,$k$+1, and any non-negative integer, there is

$$\begin{aligned}&P\left\{X_{t_{k+1}}=i_{k+1}\,\middle|\,X_{t_0}=i_0,X_{t_1}=i_1,\cdots,X_{t_k}=i_k\right\}\\&=P\left\{X_{t_{k+1}}=i_{k+1}\,\middle|\,X_{t_k}=i_k\right\},\end{aligned} \tag{4}$$

where $X(t)$ is called discrete time Markov chain.

In practice, a homogeneous Markov chain is generally used, which means $p=p_{(n)}^{(k)}=p\{X_{n+k}=j|X_n=i\}$, where the change of $n$ will not change the probability $p$.

### 2.3 Steps in establishing the Markov chain model

1. Obtain discrete states

The range between the maximum and the minimum of the relative errors of the fitted value obtained from GM (1,1) is continuous. To obtain discrete states, usually the range is split evenly to assign a state value to each interval. The proposed DAI procedure is described below.

2. Determine the TM

An element $P_{ij}$ of TM is the probability of transition from states $i$ to $j$, which is calculated as

$$P_{(i,j)}^{k}=\frac{M_{ij}^{(k)}}{M_i}, \tag{5}$$

where $M_i$ is the number of state $i$ in the Markov chain, and $M_{ij}^{(k)}$ is the number of occurrence that transiting from states $i$ to $j$ with $k$ steps.

3. Quantify the weight

The relative importance of each TM is quantified by an auto-correlation coefficient defined as

$$r_k = \frac{\sum_{l=1}^{n-k}(x_l - \bar{x})(x_{l+k} - \bar{x})}{\sum_{l=1}^{n}(x_l - \bar{x})^2}. \tag{6}$$

Then $r_k$ is normalized to get

$$w_k = \frac{|r_k|}{\sum_{l=1}^{m}|r_k|}. \tag{7}$$

The normalized weight vector of each TM is $\bar{\boldsymbol{w}} = [w_1, w_2, \cdots, w_n]$.

4. Construct the forecast matrix

The states at times $t, t-1, \ldots, t-\text{m}$ are $i_t, i_{t-1}, \ldots, i_{t-\text{m}}$. The forecast matrix is constructed:

$$\mathbf{PY} = \begin{bmatrix} \mathbf{PZ}^{(1)}_{(i_t,:)} \\ \mathbf{PZ}^{(2)}_{(i_{t-1},:)} \\ \vdots \\ \mathbf{PZ}^{(n)}_{(i_{t-n+1},:)} \end{bmatrix}, \tag{8}$$

where $\mathbf{PZ}^{(k)}_{(i_{t-k+1},:)}$ is all elements in $i_{t-1}$ row of the $k$-step TM, likewise for others; $n$ is the total TM steps considered.

5. Solve the judgment vector

The judgment vector is obtained:

$$\boldsymbol{wp} = \bar{\boldsymbol{w}} \times \mathbf{PY}. \tag{9}$$

According to the principle of the maximum likelihood, the prediction of the state $i$ at time $t$+1 is the one corresponding to the largest number in **wp**. The prediction of the relative error is the midway of the high value and the low value of the interval.

On combining the fitted value in GM (1,1) and the prediction of relative error in Markov chain, a new prediction data in time $t$+1 is obtained.

### 2.4 Procedure of dynamic adjustment interval

Step 1: The relative error of the fitted value range is divided with equal interval, whose length is $L(0)$. The predicted state in time $t$+1 is calculated as $i_{t+1}$.

Step 2: Decrease the high value of the interval of state $i_{t+1}$, and calculate the new predicted state $j$. If state $j$ is the same as $i_{t+1}$, this adjustment is successful, record the new interval length, and repeat Step 2; if state $j$ is different from $i_{t+1}$, this adjustment is failed, go to Step 3.

Step 3: Increase the low value and calculate the new predicted state $j$. If state $j$ is the same as $i_{t+1}$, this adjustment is successful, record the new interval length; repeat Step 3; if state $j$ is different from $i_{t+1}$, this adjustment is failed, go to Step 4.

Step 4: Terminate the algorithm.

## 3 Case study

Fig. 1 presents the water supply data within continuous 60 h (where time hour 0 is 0:00 am) from a water utility of a northern city in China. It shows that the water supply data for 5 h at night (from 0:00 am through 5:00 am each day) is relatively stable, while others have greater fluctuation. The data of 0–48 h are used to build the GM (1,1), i.e., estimate parameters $a$ and $b$ in Eq. (1). The data of 49–55 h are used to build the Markov chain model, where the DAI procedure is applied and the best TM step is estimated. The data of 56–60 h are used to validate the GMWMC model.

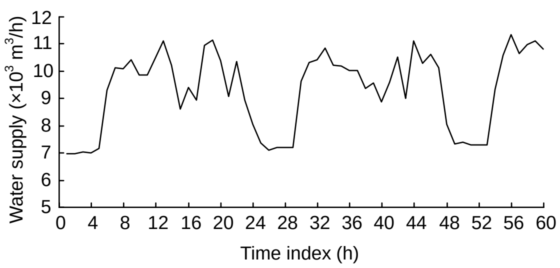


**Fig. 1 Utility hourly water supply data**

Fig. 2 shows the first 48 h water supply data fitted values of GM (1,1). With GM (1,1), the relative errors of the fitted values in 0–48 h are calculated. These relative errors are the data to establish the Markov chain model. The water supply data of the 49th hour are calculated by GM (1,1), and its relative error is added to the Markov chain, whose data amount is 49; likewise until the number of the data arrives 55.

For the purpose of determining the best TM step, TM steps four-step, five-step, six-step, and seven-step are tested. Table 1 lists the relative errors and mean absolute percentage errors (MAPEs) of the fitting values by static interval and DAI using four through seven TM steps.

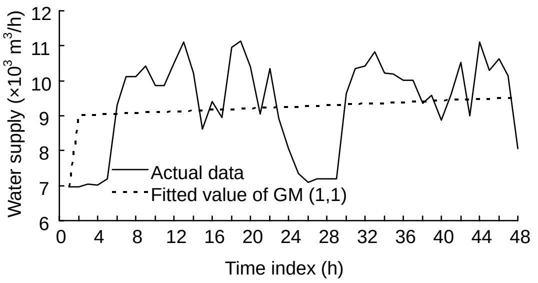


**Fig. 2 Fitted value of GM(1,1)**

From Table 1, the MAPE of the fitting values of DAI with six-step TM is the smallest. As a result, the DAI with the six-step TM can be used to establish a model for forecasting. The six-step TM has the smallest MAPE, thus the best calibration parameter. To validate the calibrated model, the static interval (SI) and DAI with the six-step TM are applied in prediction, with relative error, maximum relative error and MAPE presented in Table 2.

## 4 Results and discussion

From Table 1, the MAPE of the fitted value from GM (1,1) is 24.12%, and the combination of GM (1,1) with Markov chain by using TM steps 4 through 7 can greatly reduce the fitting error. This is because the GM (1,1) can only extract the trend of hourly water supply data, and the relative error of the fitted value also has large variability. The Markov chain is capable of fitting data sequences with large variability.

Table 1 shows that with increasing steps of TM, the MAPE is reduced, except for the seven-step TM, which becomes larger. This variation trend indicates that for specific data amount, there may be one best TM step.

From Tables 1 and 2, by comparing the DAI with the corresponding SI with 4–7 steps TM, the MAPE decrease by 0.10%–6.57%, with the largest reduction of 27.30%, and the MAPE of forecast decreased by 0.03%–1.09%, with the largest reduction of 30%.

The impact of the TM step on model calibration and validation are basically consistent. With the six-step TM, the MAPE of the calibrating model is the lowest; while with the five-step or six-step, the

**Table 1 Relative errors of fitting of each method (%)**

| Time index (h) | GM (1,1) | Four-step TM | | Five-step TM | | Six-step TM | | Seven-step TM | |
|---|---|---|---|---|---|---|---|---|---|
| | | DAI | SI | DAI | SI | DAI | SI | DAI | SI |
| 49 | 35.25 | 6.38 | 7.39 | 5.53 | 6.32 | 5.46 | 5.63 | 40.49 | 35.41 |
| 50 | 31.33 | 0.24 | 4.20 | 0.23 | 3.17 | 1.61 | 2.50 | −2.46 | 41.34 |
| 51 | 30.72 | 0.55 | 4.77 | 0.54 | 3.76 | 2.23 | 3.10 | −2.41 | 2.63 |
| 52 | 28.41 | −1.10 | 2.60 | −1.11 | 1.62 | 0.12 | 0.97 | −4.47 | 28.43 |
| 53 | 26.30 | −2.74 | 0.19 | −2.76 | −0.78 | −2.27 | −1.43 | −6.86 | −1.88 |
| 54 | −2.80 | −24.39 | −23.35 | −25.34 | −24.12 | 0.66 | 2.45 | −28.69 | −24.97 |
| 55 | −14.05 | −1.04 | −2.96 | −0.97 | −1.55 | −9.05 | −9.46 | −36.92 | −33.60 |
| MAPE | 24.12 | 5.20 | 6.49 | 5.21 | 5.90 | 3.06 | 3.65 | 17.47 | 24.04 |

DAI: dynamic adjustment interval; SI: static interval; MAPE: mean absolute percentage error

**Table 2 Predictive relative errors of each method (%)**

| Time index (h) | Four-step TM | | Five-step TM | | Six-step TM | | Seven-step TM | |
|---|---|---|---|---|---|---|---|---|
| | DAI | SI | DAI | SI | DAI | SI | DAI | SI |
| 56 | −6.34 | −8.60 | −6.28 | −7.30 | −5.94 | −6.41 | −5.39 | −13.00 |
| 57 | 1.40 | −1.00 | 1.45 | 0.38 | 1.80 | 1.32 | 2.01 | 2.01 |
| 58 | −0.98 | −2.81 | −0.42 | −1.47 | −0.10 | −0.55 | −6.61 | 0.12 |
| 59 | −1.28 | −2.56 | −0.16 | −1.21 | −0.15 | −0.29 | −6.40 | −7.08 |
| 60 | 2.86 | −3.33 | 4.05 | 2.93 | 4.39 | 3.92 | −2.70 | −3.33 |
| Maximum relative error | −6.34 | −8.60 | −6.28 | −7.30 | −5.94 | −6.41 | −6.61 | −13.00 |
| MAPE | 2.57 | 3.66 | 2.47 | 2.66 | 2.47 | 2.50 | 4.62 | 5.11 |

DAI: dynamic adjustment interval; SI: static interval; MAPE: mean absolute percentage error validation is the best. The establishment of forecast model should be based on not only the lowest fitting error, but the forecast ability. And the maximum relative error with the six-step TM is smaller than the one of the five-step TM. As a result, the GMWMC with the six-step TM by DAI should be used to establish a water supply forecast model with real-time scheduling.

To compare the GMWMC with the six-step TM by DAI with trigonometric function model and neural network, the same data were used to establish a trigonometric function model and a neural network model. Forecast results of each method are shown in Fig. 3 and Table 3.

Table 3 shows that the proposed model with DAI has the best forecast results. Comparing with the trigonometric function and neural network, either the largest scope of errors or MAPE is improved greatly.

Generally, a neural network is suitable for nonlinear prediction, but needs larger sample size. With a small sample size, the neural network may also be fitted well by increasing nodes number in hidden layer, however, it will lead to over fit and accordingly poor predication ability. This study only uses two periods of sample water supply data to train the network model, so it is difficult to obtain high prediction performance without over fitting the neural network. If the sample size is increased to seven or eight period, when the variation characteristics of water supply are fully described, the prediction accuracy might be greatly improved.

Trigonometric function prediction establishes the forecasting regression model of sample data to obtain better fitting of the data's variation trend. If it is interfered with by other random factors, the prediction accuracy will be reduced. So the trigonometric function results in better fitting, but worse prediction results. It can be observed from Fig. 4 that the water supply data within the first 48 h are fitted well, but the later 12 h possess a certain lag comparing with observed data.

The DAI method proposed in this study increases computational complexity, and prediction accuracy may be affected when the data possess large variability. As well, uncertainties related to the prediction values are not characterized in this study, which provides more information in utilities decision-making process in terms of water supply.

## 5 Conclusions

1. GMWMC is applied for real-time water supply amount prediction. A DAI method is proposed to shorten the prediction interval to generate a higher

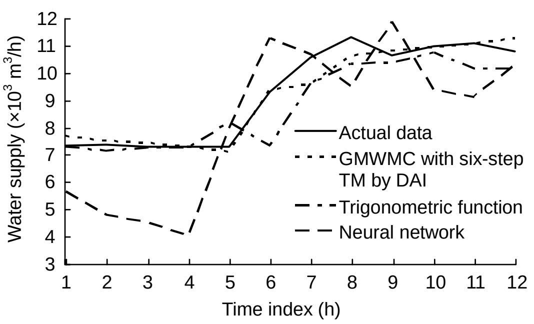


**Fig. 3 Forecast results of the GMWMC, trigonometric function and neural network**

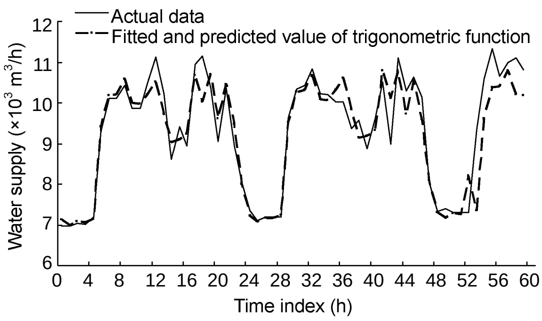


**Fig. 4 Comparison of actual data and fitted and predicted value of trigonometric function model**

**Table 3 Comparison of three forecast results of the GMWMC, trigonometric function and neural network**

| Parameter | GMWMC with six-step TM by DAI | Trigonometric function | Neural network |
|---|---|---|---|
| Maximum relative error (%) | −9.05 | −21.4 | −44.59 |
| Minimum relative error (%) | −0.10 | −0.29 | 1.14 |
| MAPE (%) | 2.82 | 6.15 | 19.62 |
| Percent of data of relative error >5% in all data (%) | 25 | 50 | 83 |

GMWMC: GM (1,1) weighted Markov chain; DAI: dynamic adjustment interval; MAPE: mean absolute percentage error prediction accuracy. As demonstrated by a case study, we can obtain: (1) there is an optimal TM step for a certain sample size; (2) the described DAI can obtain better performance both in calibration and validation stages; and (3) on comparing with neural network and trigonometric function, the GMWMC performs better when sample size is small.

2. By including DAI, the increment of forecast accuracy is related to the initial interval division method and the interval adjustment method. Equal interval division is used, and a further interval division method may be explored. In addition, the interval adjustment method also requires improvements. Other adjustment methods, such as neural network and genetic algorithms may be applied (Gong, 2006).

## References


Deng, J., 2005. The Primary Methods of Gray System Theory. Huazhong University of Science and Technology Press, Wuhan, China (in Chinese).

Gai, C., Pei, Y., 2003. Study of the GM (1,1)-Markov chain model on highway freight forecast. *China Journal of Highway and Transport*, **16**(3):113-116 (in Chinese).

Geng, B., Wang, J., Zhang, X., 2007. GM (1,1)-Markov model for prediction of bridge technical condition. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, **31**(1):107-110 (in Chinese).

Gong, L., 2006. The NN model based on annealing arithmetic of genetic simulation under the application of water-demand forecast in Shaanxi Province. *Underground Water*, **28**(5):10-13, 20 (in Chinese).

He, Y., Huang, M., 2005. A grey-Markov forecasting model for the electric power requirement in China. *LNCS*, **3789**:574-582. [doi:10.1007/11579427_58]

Li, G., Yamaguchi, D., Nagai, M., 2007. A GM (1,1)-Markov chain combined model with an application to predict the number of Chinese international airlines. *Technological Forecasting & Social Change*, **74**(8):1465-1481. [doi:10.1016/j.techfore.2006.07.010]

Liao, G., Tsao, T., 2004. Application of fuzzy neural networks and artificial intelligence for load forecasting. *Electric Power System Research*, **70**(3):237-244. [doi:10.1016/j.epsr.2003.12.012]

Liu, H., Zhang, H., 2002. Comparison of the city water consumption short-term forecasting methods. *Transactions of Tianjin University*, **8**(3):211-215 (in Chinese).

Tien, T., 2005. A research on the prediction of machining accuracy by the deterministic grey dynamic model DGDM(1, 1, 1). *Applied Mathematics and Computation*, **161**(3):923-945. [doi:10.1016/j.amc.2003.12.061]

Yao, Q., Li, C., Ma, H., Zhang, S., 2007. Novel network traffic forecasting algorithm based on grey model and Markov chain. *Journal of Zhejiang University (Science Edition)*, **34**(4):396-400 (in Chinese).

Zhou, G., Wang, H., Wang, D., Zhang, G., 2004. Urban water consumption forecast based on neural network model. *Information and Control*, **33**(3):364-368 (in Chinese).