Desalination and Water Treatment
www.deswater.com
doi:10.5004/dwt.2018.22358

114 (2018) 51–60
May

Regional annual water consumption forecast model

Xianqi Zhang[a,b], Mingliang Yue[a], Yuan Yao[a,*], Hao Li[a]

[a]*College of Water Resources, North China University of Water Resources and Electric Power, Zhengzhou 450045, China, Tel. +86-130-2775-6482; email: 602174930@qq.com (Y. Yao), Tel. +86-158-3719-7937; email: zxqi@163.com (X. Zhang), Tel. +86-181-3788-7641; email: 1725996950@qq.com (M. Yue), Tel. +86-158-3811-4788; email: 553935627@qq.com (H. Li)*
[b]*Henan Key Laboratory of Water Environment Simulation and Treatment, Zhengzhou 45002, China*

Received 23 September 2017; Accepted 8 April

ABSTRACT

Annual water consumption forecasting is critical to the optimal operation and reasonable allocation of water resource. Based on the summary of annual water consumption studies at home and abroad, with the principle of maximal precision and minimal error, and using data of Chinese water consumption in the past 10 years, this paper analyzed the feature and property of time series prediction method in water consumption study, predicted China's water consumption in 2015, 2020 and 2030 and evaluated its accuracy and error through power function model, linear function model, logarithm function model and parabolic function model, and its prediction accuracy and error were analyzed. The results demonstrated that the parabolic curve fitting correlation coefficient was the biggest and mean absolute percent error was the smallest, and the fitting effect was the best when conducting the curve fitting of water use in 2015, 2020 and 2030. Research results were basically consistent with those of the actual annual water use, which will provide references and basis for the prediction of the regional annual water consumption.

Keywords: Water consumption; Time series method (TSM); Model; Curve fitting; Forecast

1. Introduction

With the rapid development in technology, economy and urbanization as well as the consequent increase in population, buildings and all kinds of facilities, aggregate water consumption augments by leaps and bounds. In order to relieve water supply shortage, many countries are devoted to researches seeking efficient solutions, among which water consumption forecast is one of the significant studies in water resource management. For example, California State of America launched the water consumption forecast study as early as 1956 and made revision on the result according to real water consumption conditions. Japan made water consumption forecast an important index for the decennial land planning since 1960. Ever since a variety of countries began to use water consumption forecast as a means of fulfilling water management policy. The water conservancy department in China initiated nation-wide water resource assessment work in 1979, after 7 years, completing the mission in 1986 with a research report named *China Water Consumption Utilization*, water supply and demand listed as an independent chapter in the report.

Domestic and foreign scholars and experts made plentiful studies in the methodology of water consumption forecast. Shvars and Feldman [1] established pattern recognition model to have short-term water consumption forecast. Applied in the domestic and industrial water consumption forecast in cities like Madrid, this model has proved to be effective. Australian scholars such as Zhou et al. [2] constructed the time series prediction model, applying it to day water consumption prediction in Melbourne which worked and the result is very good. Using annual precipitation, residents' average income, population density and water price as correlation factors, Mays [3] and other scholars built mid- and

* Corresponding author.

long-term logarithm and semi-logarithm regression model on water consumption and its correlation factors and applied it to the mid- and long-term water consumption prediction in Texas. Jain et al. [4] evaluated eight regression models, respectively, forecast short-term water consumption and tested the models with the highest monthly water consumption, water consumption on the maximum temperature day and water consumption on rainy day. Day and Howe [5] analyzed the nonweather factors influencing the highest day water consumption and made the forecast. Brekke et al. [6] made water consumption forecast through stepwise regression and tested its applicability. Scholars in China also realized substantial accomplishment in water consumption forecast methodology. They proposed a wide range of approaches including time series method (TSM), regression analysis, back propagation (BP) neural network, grey model, set pair analysis and genetic algorithm. For instance, Li et al. [7] had detailed analysis on urban water consumption based on autocorrelation analysis theory and optimized water consumption forecast models. Yang et al. [8] conducted forecasts on a city's day water consumption with exponential smoothing, BP neural network and grey prediction models, respectively. By comparison, they drew the conclusion that BP neural network method has more accuracy. Aiming at the increasing imbalance between water supply and demand, Zhao [9] made forecasts on per person water consumption using multilinear regression and BP neural network models, respectively.

Generally, water consumption model can be classified into two categories. The first one is mathematical models based on water consumption and its correlation factors. However, since this method is built upon regression analysis which uses correlation parameters as independent variables in forecasting water consumption, the requirement of large amount of information and the difficulty in accurate observation of independent variables complicate the forecast and reduce its effectiveness [10]. The second category is mathematical models established on the autocorrelation between past and present water consumption without considering other parameters such as the weather, that is through TSM. The merit of this model is that the change in water resource conditions leads to change in water consumption. But change in water consumption is relatively slow therefore whose effects are not obvious. Besides even if the change does have some effects, these effects can be demonstrated from historical data. In this paper, the author chose the second approach to conduct water consumption forecast study. Using original data of water consumption for China's past 10 years and holding principle of maximum accuracy and minimum error, it explored the extent of accuracy and error in trend extrapolation and came up with the best annual water consumption, model and providing evidence for the choice of methods in regional annual water consumption forecast study.

2. Forecast methods and applicability

2.1. Forecast methods categories

The complication of water system makes it difficult to describe the system by a fixed model. Therefore, the majority of water consumption forecast studies are based on the analysis of historical data. However, these studies differ in their employing various data processing methods, according to which we classify them into TSM, morphological analysis and system approach [11]. The detailed categories are shown in Table 1.

TSM is established on the study of changes in water consumption over the years so as to make judgement on future situations. It usually requires complete and abundant data. Morphological analysis demands not only time factors but also internal connection between water consumption and its correlation factors and the rule of change. By choosing a number of affecting factors, this method estimates the correlation between them and water consumption. Then, model equations of unknown parameters are listed, and real data are substituted in to determine each parameter and make forecast. In system approach, water consumption is seen as complicated system engineering. Through system objective analysis, system element analysis, system environment analysis, system resource analysis and system management analysis problems are clearly defined, origin of water consumption issue revealed and at last requirement met and solutions proposed for water consumption forecast [16,17].

2.2. Applicability of forecast methods

With the rapid advancement of technology, water consumption forecast methods are also multiplied. The current methods of BP neural network, multilinear regression, have more flexibility and accuracy compared with the previous methods such as index method [18]. Today, there are hundreds of water consumption forecast methods, differing one another to some extent:

- Categories by predictive period

According to the predictive period, water consumption forecast is divided into long-term forecast and short-term forecast, each matches different forecast methods. Morphological analysis, grey prediction in system analysis and system dynamics method are fit into mid- and long-term forecast, while autoregressive integrated moving average model (ARMA) in TSM and artificial neural network (ANN) method, etc., are suitable for short-term forecast.

- Categories by scale of forecast

According to the scope of forecast, large-scale forecast may use methods having links with macro factors based on comprehensive data analysis, such as the tendency analysis in TSM, exponential smoothing, morphological analysis and grey prediction in system approach. By contrast, the easy obtaining of required base data in small-scale forecast leads to applicability of micro-factors-related methods (TSM, ARMA, Markov, etc.), conducted on the basis of diverse detailed data analysis.

- Categories by the degree of completion of data

The lack of effective long-term historical data causes certain limitation on the choice of water-demand forecast method. In this situation, we can use exponential smoothing of TSM, elastic coefficient method in morphological analysis, analytical approach, comparison method, statistical

Table 1
Summary sheet of main mathematical forecast methods

Time series method	Deterministic	Moving average method	Simple average method	
			Simple moving average method	
			Weighted moving average method	
		Exponential smoothing	Single-index moving method	
			Double exponential smoothing	Brown single parameter exponential smoothing method
				Holt's double exponential smoothing
			Cubic exponential smoothing method	Brown single parameter exponential smoothing method
				Winter linear seasonal exponential smoothing
		Trend extrapolation	Polynomial model	
			Exponential curve model	
			Logarithm function model	
			Growth curve model	
		Seasons change method [12]	Seasonal level model	
			Seasonal mutual tendency model	
			Seasonal trend model superposition	
		Grey prediction models	Grey correlation analysis	
			Grey sequence forecast	
			Grey index prediction	
			Calamities grey prediction	
			Grey topological predicting	
	Random probability	Markov models [13]	A heavy chain-related prediction	
			Model prediction	
		Box-Jenkins models [14]	Autoregression (AR) model	
			Moving average (MA) model	
			Autoregressive integrated moving average (ARMA) model	
The structure analysis method	Regression analysis		Unitary linear recursive analysis	
			Multiple linear regression analysis	
			Nonlinear regression analysis	
	Industrial water elastic coefficient forecast method			
	Index analysis method			
	Artificial neural network (ANN) method [15]			
System analysis method	System dynamics method			

information analysis and grey prediction of system approach. Nevertheless, areas with sufficient historical data can apply tendency method in TSM, ARMA and Markov, regression and empirical formula in morphological method, ANN method, system dynamics method, etc.

3. Forecast method selection

Currently, the aggregate water consumption guidepost has been allocated to each region (administrative region) in China. The following step is to make annual water consumption plans from 2015 to 2030. Despite that the qualified indicators of total water consumption in 2015, 2020 and 2030 are given, those of node years such as 2016–2019 and 2021–2029 are not known yet, in need of forecast. The common practice for water management administrators in each region is to have linear interpolation from indicators in 2015, 2020 and 2030 which is sometimes so unscientific as to make errors. Thus, more reasonable forecast with fixed criterion is necessary in order to supervise the performance in annual water consumption and implement strict water resource management with efficient audit, management and evaluation.

3.1. Criterion of models selection

Annual water consumption is important parameter in the total quantity control and management of water resource. It is affected by various aspects, such as the structure of water consumption, industrial structure, population scale, educational status and so on. The key element in its forecast is to choose valid models targeting at water consumption in different industries to increase accuracy. With population growth, socioeconomic development and the adjustment of industrial structure, it is certain that there will be an increase in annual water consumption. However, there are many factors leading to its growth, some of them are difficult to obtain accurately and the mechanism of influence is ambiguously fuzzy and uncertain. So a variety of time series fitting methods is used to predict the annual water consumption. Their main features are the use of mere historical data in the forecast objective without taking its correlation factors into consideration. It possesses a certain precision as a mid- and long-term forecast model.

Studies show that the forecast precision is not only related to the chosen model but also to the series of original data. At present, the common practice is to choose models potential for maximum fitting precision and maximum interpolation forecast accuracy. Based on these conclusions, this project put forward a method with the highest fitting precision and lowest error [19]. By this method, we make forecast with reserved data in recent several years (assume it L) and other required data in the construction of parameter model. We then made comparison between the forecast and measured value of their corresponding years to get their correlation coefficient and absolute percent error and draw consequent mean of absolute percent error in the forecast value of this model. Similarly, the mean absolute percent error (MAPE) in other candidate models for the recent M years is obtained. The model with the biggest fitted curve coefficient and the smallest MAPE is the designated model.

3.2. Mathematical models

Suppose the number of candidate models is m, $y = f_j(x)$, in the equation y is water consumption, 100,000,000 m^3; x is time duration, year; j is model's serial number, j = 1,2, …, m; n is series length of data, that is x_i, y_i, of which i is the serial number of water consumption, i = 1,2, …,n; the forecast period is L years. Curve fitting is conducted to maximize the correlation coefficient R in this series and minimize its MAPE.

In the TSM, power function model, linear function model, exponential function model, logarithm function model and parabolic function model (second-order polynomial function) are the most frequently used ones [20]. According to the analysis of fundamental features in China's water consumption change and its historical data, this research adopted linear function model, logarithm function model, parabolic function model and power function model to test their effectiveness. Corresponding expression is listed below:

$$y = ax + b \tag{1}$$

$$y = a\ln(x) + b \tag{2}$$

$$y = ax^2 + bx + c \tag{3}$$

$$y = ax^b \tag{4}$$

3.3. Case study

Based on the official report on water resource of People's Republic of China, this study focuses on the analysis of data in real water consumption from 2003 to 2011 and government-authorized limit water consumption for 2015, 2020 and 2030. In 2003, the national aggregate water consumption is 532 billion m^3; the number is 554.8 in 2004, 563.3 in 2005, 579.5 in 2006, 581.9 in 2007, 591.0 in 2008, 596.5 in 2009, 602.2 in 2010, 608 in 2011, 635 in 2015, 670 in 2020 and 700 in 2030, respectively.

- Sampling on water consumption in 2003–2011, 2015, 2020 and 2030.

The water consumption data in 2003–2011, 2015, 2020 and 2030 are applied in a certain predictive period under the principle of minimum error and maximum correlation coefficient to test the data samples in 2015, 2020 and 2030. After conducting linear fitting, logarithm curve fitting, polynomial curve fitting and power fitting, an annual water consumption fitting curve is obtained, as shown in Figs. 1–4. The correlation coefficient and MAPE are seen in Table 2.

As is seen in Table 2, based on the water consumption data in 2003–2011, 2015, 2020 and 2030, the correlation coefficients in conducting curve fitting are relatively big, while MAPE small. Moreover, the correlation coefficient is the biggest and MAPE the smallest in the polynomial function model test, which is 0.9897, and 0.3818, respectively.

- Sampling on data in 2003–2011

In this case, real water consumption in 2003–2011 is used as samples and water consumption in 2015, 2020 and 2030 as test samples. Conducting linear fitting, logarithm fitting, polynomial fitting, power function curve fitting, an annual water consumption fitting curve is acquired, as shown in Figs. 5–8. The forecast model's correlation coefficient and MAPE are shown in Table 3.

As is seen in Table 3, conducting test on data samples of 2003–2011 through curve fitting, the correlation coefficients are big while MAPE also big. Among which, the correlation

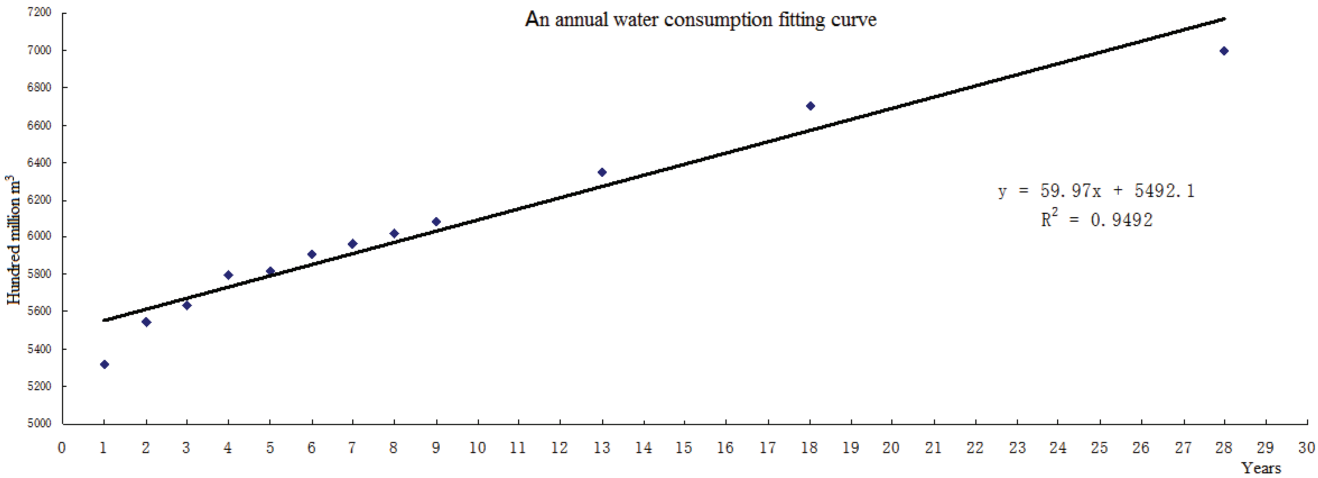

Fig. 1. The annual water consumption is depicted using the linear curve fitting.

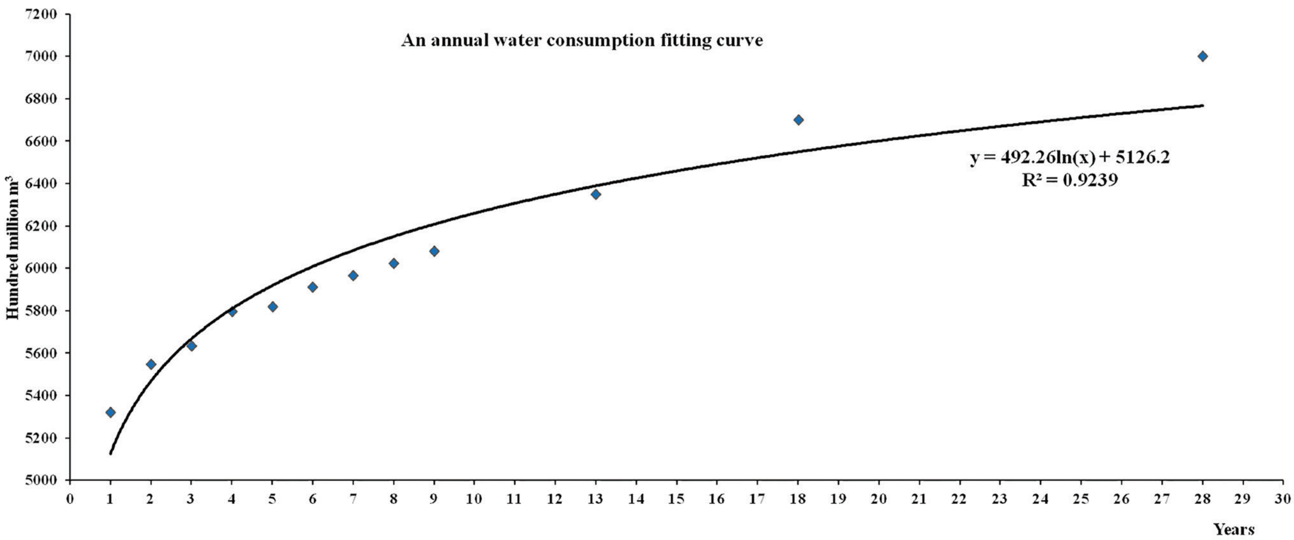

Fig. 2. The annual water consumption is depicted using the logarithmic curve fitting.

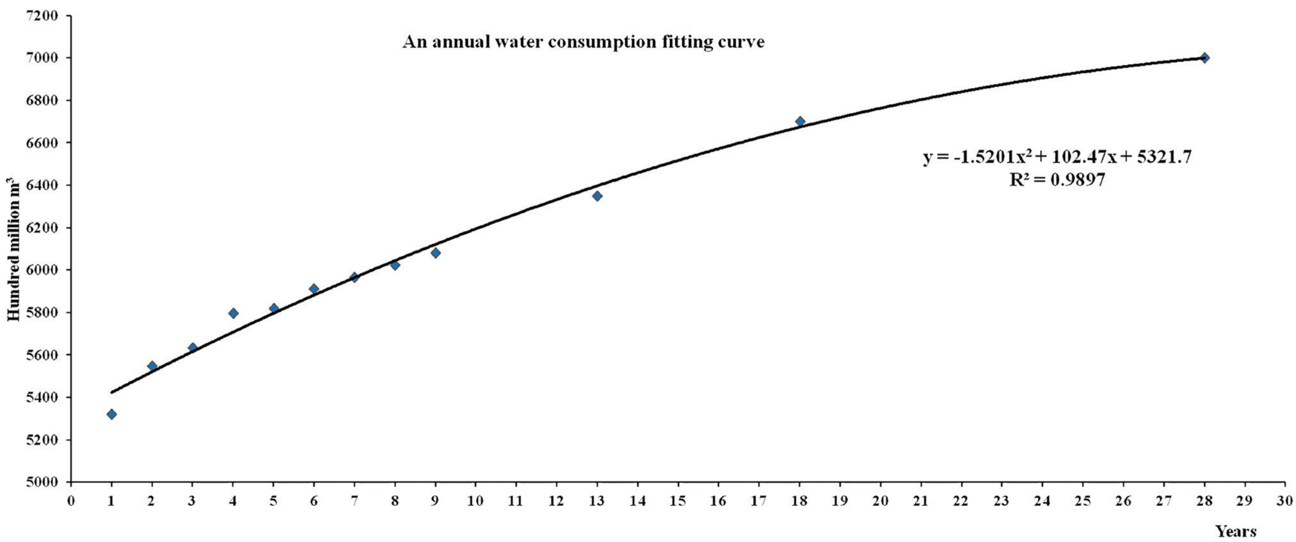

Fig. 3. The annual water consumption is depicted using the polynomial curve fitting.

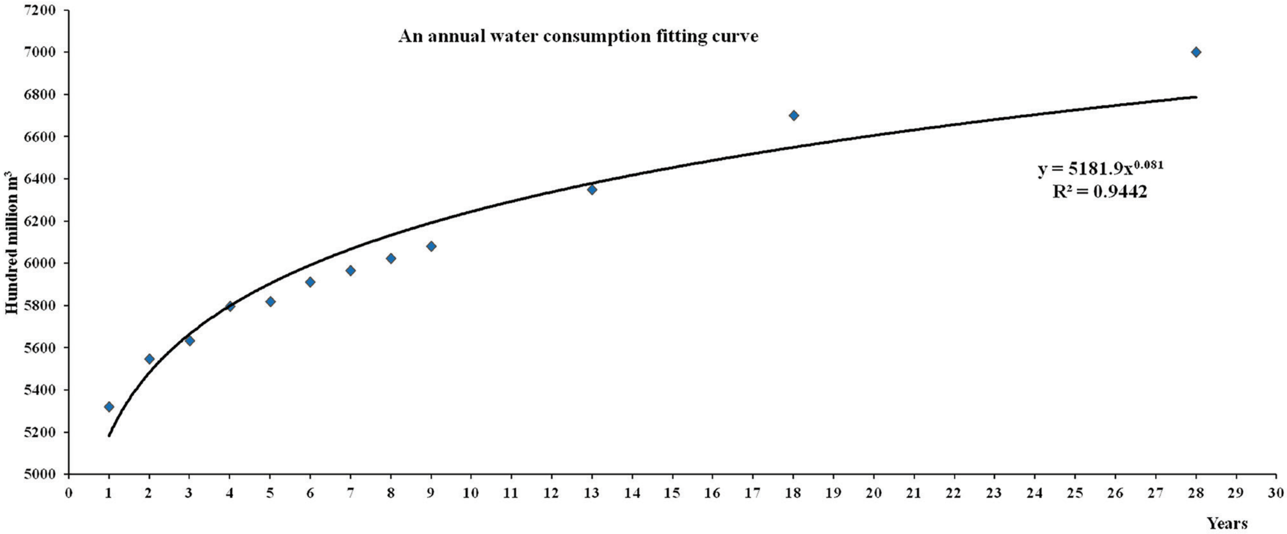

Fig. 4. The annual water consumption is depicted using the power curve fitting.

Table 2
Correlation coefficient of each curve fitting and their forecasting MAPE

Model	Predictive value			Correlation coefficient	MAPE
	In 2015	In 2020	In 2030		
Linear function	6,271.7	6,571.6	7,171.3	0.9492	1.8655
Logarithmic function	6,388.8	6,549.0	6,766.5	0.9239	2.0668
Polynomial function	6,396.9	6,673.6	6,999.1	0.9897	0.3818
Power function	6,378.5	6,548.9	6,787.5	0.9442	1.9133

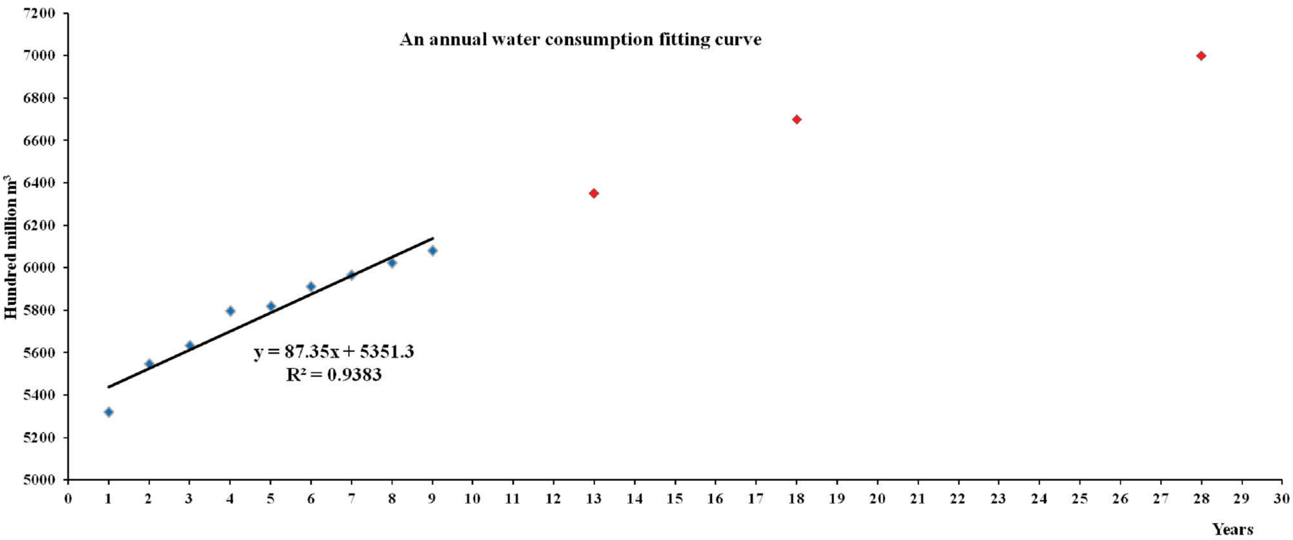

Fig. 5. The annual water consumption is depicted using the linear curve fitting.

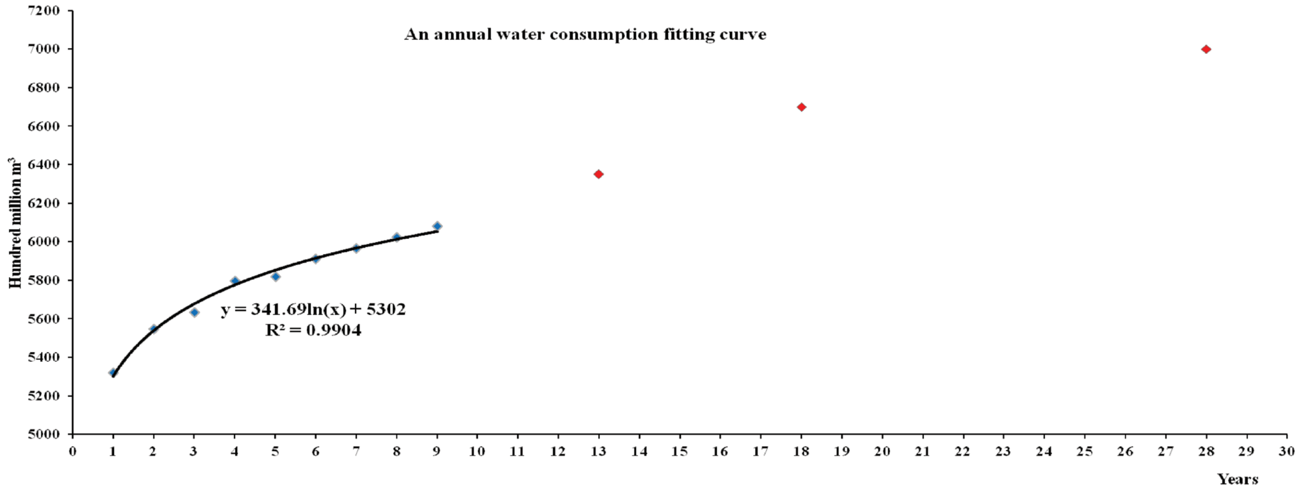

Fig. 6. The annual water consumption is depicted using the logarithmic curve fitting.

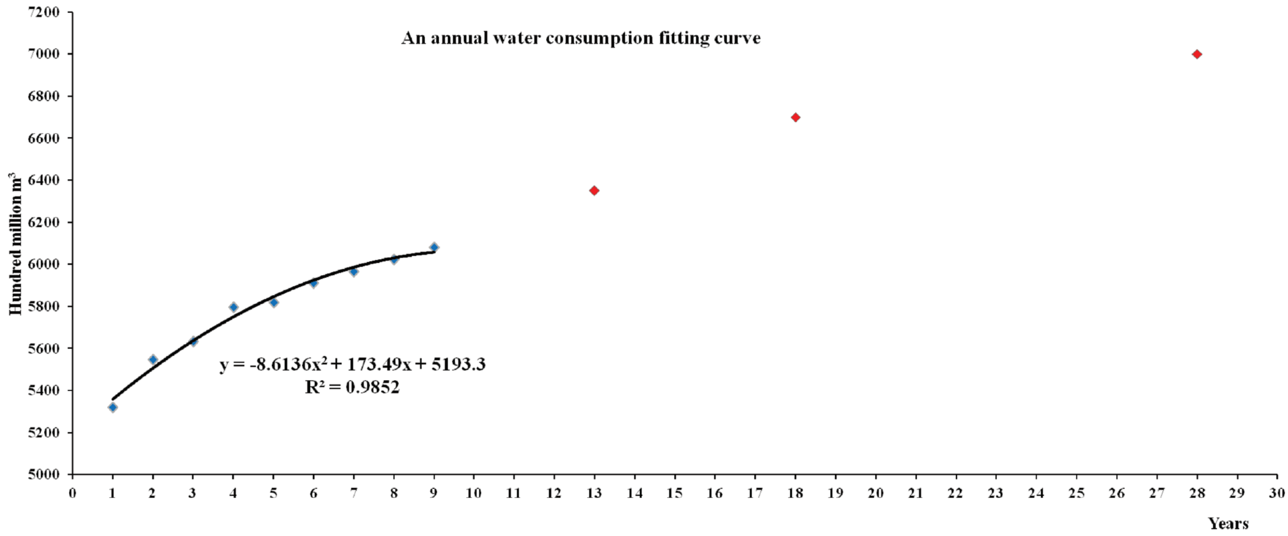

Fig. 7. The annual water consumption is depicted using the polynomial curve fitting.

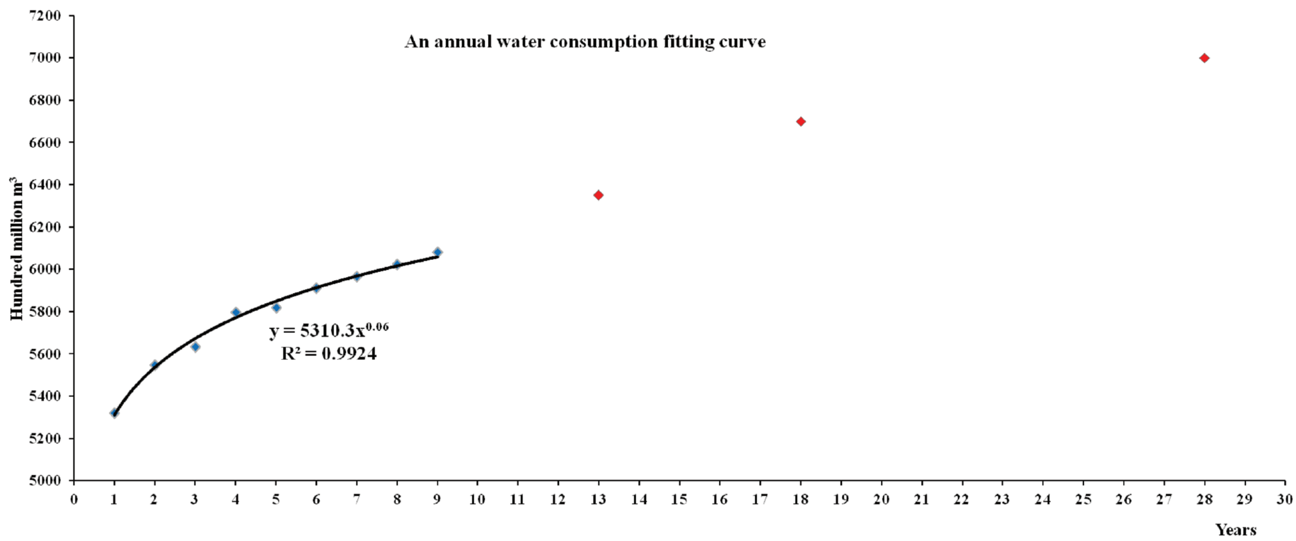

Fig. 8. The annual water consumption is depicted using the power curve fitting.

Table 3
Correlation coefficient of each curve fitting and their forecasting MAPE

Model	Predictive value			Correlation coefficient	MAPE
	In 2015	In 2020	In 2030		
Linear function	6,486.9	6,923.6	7,797.1	0.9383	5.6268
Logarithmic function	6,178.4	6,289.6	6,440.6	0.9904	5.6064
Polynomial function	5,993.0	5,525.3	3,298.0	0.9852	25.3469
Power function	6,193.8	6,315.9	6,485.6	0.9924	5.1804

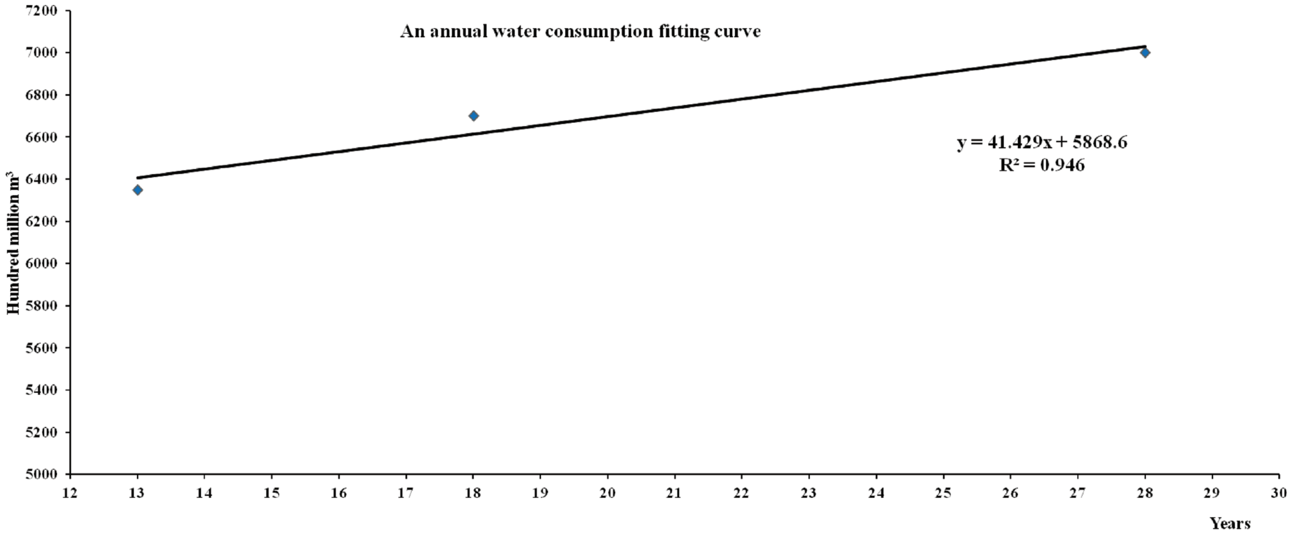

Fig. 9. The annual water consumption is depicted using the linear curve fitting.

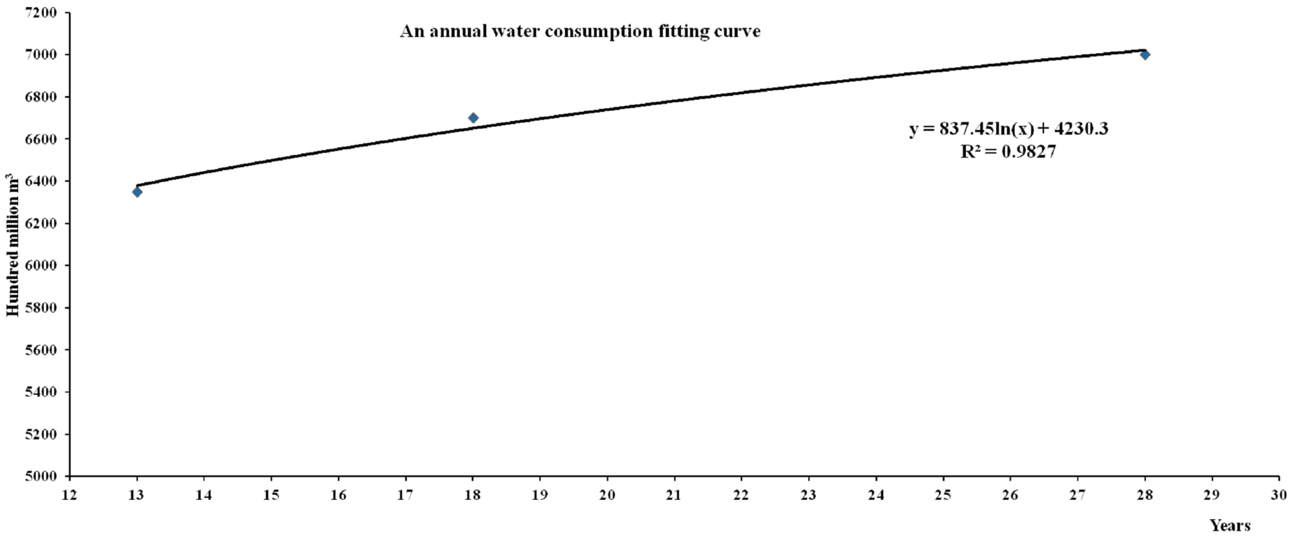

Fig. 10. The annual water consumption is depicted using the logarithmic curve fitting.

coefficient is the biggest and MAPE the smallest in power function curve fitting, 0.9924 and 5.1804, respectively.

- Sampling on water consumption data of 2015, 2020 and 2030.

Curve fittings are conducted to test the data of government-authorized water consumption for 2015, 2020 and 2030 and make subsequent water consumption forecast between 2015 and 2030. Linear fitting, logarithm fitting, polynomial fitting and power function curve fitting are employed.

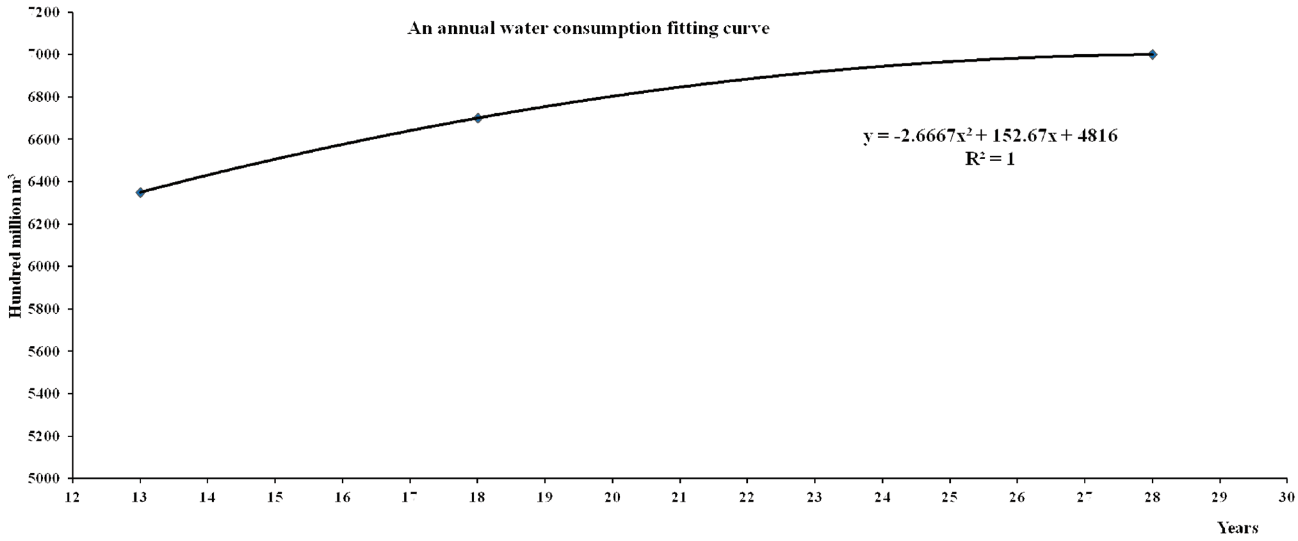

Fig. 11. The annual water consumption is depicted using the polynomial curve fitting.

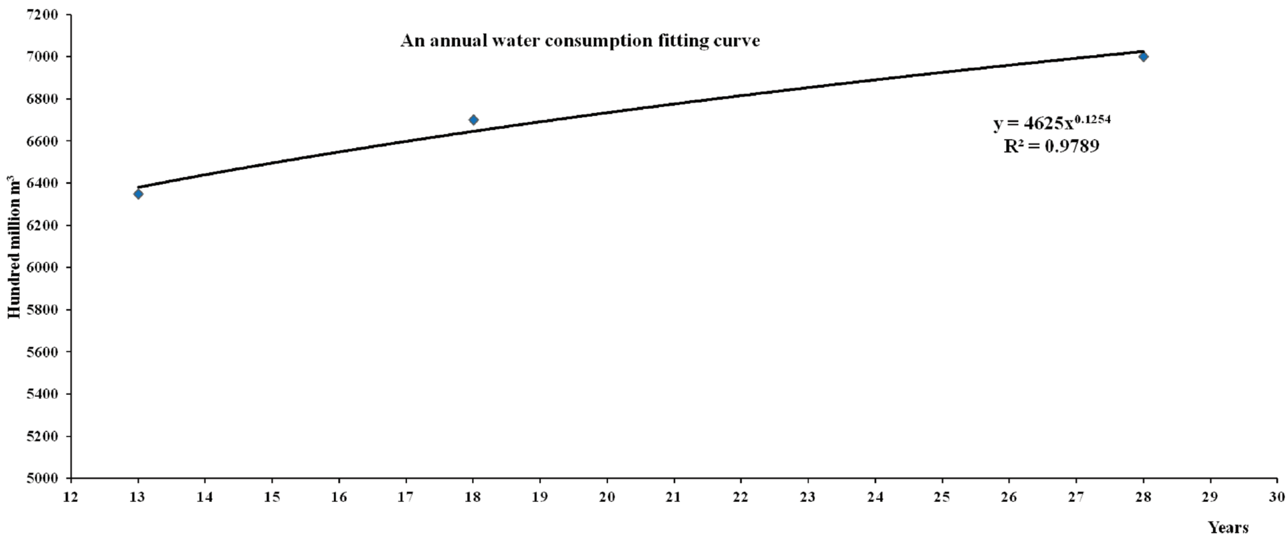

Fig. 12. The annual water consumption is depicted using the power curve fitting.

Table 4
Correlation coefficient of each curve fitting and their forecasting MAPE

Model	Predictive value			Correlation coefficient	MAPE
	In 2015	In 2020	In 2030		
Linear function	6,407.2	6,614.3	7,028.6	0.9460	0.8628
Logarithmic function	6,378.3	6,650.8	7,020.9	0.9827	0.4929
Polynomial function	6,350.0	6,700.0	7,000.1	1.0000	0.0005
Power function	6,379.7	6,645.4	7,024.0	0.9789	0.5418

The resulting water consumption fitting curve is shown in Figs. 9–12. Correlation coefficient and MAPE are shown in Table 4.

As seen in Table 4, the correlation coefficients are relatively big and the MAPE small in the curve fitting test on the data samples from 2015, 2020 and 2030 with these four functions, of which, the correlation coefficient is the biggest (1.0) and MAPE the smallest (0.0005) with the polynomial model, an indication that the annual water consumption data series in our country conform to the fundamental characteristics of polynomial function. Therefore, polynomial function is recommended to conduct time series curve fitting in China's annual water consumption forecast.

4. Conclusions

- Regional water consumption forecast is fundamental and prerequisite for the optimization of water resource usage and allocation. The choosing of forecast methods imposes significant influence on the scientific and precise forecast of regional water consumption, therefore having impact on the efficient allocation of water resource. In this research, the author analysed water consumption forecast methods at home and abroad, classified them into different categories and expounded on the applicable situations of each method.
- Through case study of China's water consumption in 2003–2011 and the government-authorized limit water consumption for 2015, 2020 and 2030, this research employed data in 2003–2011, 2015, 2020 and 2030, data in 2003–2011 and data in 2015, 2020 and 2030 as samples, respectively. Under the guidance of maximum precision and minimum error, it analyzed the annual water consumption forecast effectiveness with power function model, linear function model, logarithm function model and polynomial function model accordingly.
- The results show that correlation coefficient is the biggest and MAPE the smallest in the curve fitting by polynomial function model. Thus, the data series in China's regional annual water consumption fits into the core features of polynomial function, which is then recommended in conducting time series curve fitting in future study.
- The forecast of regional annual water consumption is affected by various factors. For instance, water consumption fluctuates with the changes in water exploitation conditions and the structure of water resource utilization. Henceforth, more researches need to be done to study the regional applicability of different forecast models.

Acknowledgements

This work is financially supported by Water Environment Governance and Ecological Restoration Academician Workstation of Henan Province, Henan province science and technology research project (no. 152102410052), Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Henan Province, Program for Science & Technology Innovation Talents in Universities of Henan Province (no. 15HASTIT049). Our gratitude is also extended to reviewers for their efforts in reviewing the manuscript and their very encouraging, insightful and constructive comments.

References

[1] L. Shvars, M. Feldman, Forecasting hourly water demand of pattern approach, Water Supply, 4 (2003) 168–172.
[2] S.L. Zhou, T.A. McMahon, A. Walton, J. Lewis, Forecasting daily urban water demand: a case study of Melbourne, Hydrology, 3 (2000) 153–164.
[3] L.W. Mays, Water demand forecasting, Hydrol. Syst. Eng. Manage., 12 (1992) 24–32.
[4] A. Jain, L.E. Ormsbee, Short-term water demand forecast modeling techniques-conventional methods versus AI, J. Am. Water Works Assn., 94 (2002) 64–72.
[5] D. Day, C. Howe, Forecasting peak demand-what do we need to know, Water Sci. Technol. Water Supply, 3 (2002) 33–39.
[6] L. Brekke, M.D. Larsen, M. Ausburn, L. Takaichi, Suburban water demand modeling using stepwise regression, J. Am. Water Works Assn., 94 (2002) 108–112.
[7] H.Y. Li, J.G. Cui, X.Q. Zhang, Optimization of urban water consumption forecast model research, China, Water Wastewater, 20 (2004) 41–44 (in Chinese).
[8] D.Y. Yang, M. Liu, Z.J. Tan, Contrastive research on three forecasting models of daily urban water consumption, J. Qingdao Technol. Univ., 32 (2011) 74–79 (in Chinese).
[9] L.M. Zhao, Y. Zhao, Multivariate linear regression and BP neural network model for per capita water consumption forecasting model research in Shandong Province, J. Chongqing Univ. Technol. (Nat. Sci.), 28 (2014) 131–139 (in Chinese).
[10] W.J. He, J.X. Wang, H.B. Zhao, H.D. Han, Urban water simulation research in Tianjin, Water Wastewater Eng., 10 (2001) 43–44 (in Chinese).
[11] J.Y. Zhang, Q.S. Liu, Water demand prediction method of evaluation and merit, China, Water Wastewater, 17 (2001) 20–27 (in Chinese).
[12] J. Caiado, Performance of combined double seasonal univariate time series models for forecasting water demand, J. Hydrol. Eng., 15 (2010) 215–222.
[13] L. Shvartser, U. Shamir, M. Feldman, Forecasting hourly water demands by pattern recognition approach, J. Water Resour. Plann. Manage., 119 (1993) 611–627.
[14] S. BuHamra, N. Smaoui, M. Gabr, The Box-Jenkins analysis and neural networks: prediction and time series modeling, Appl. Math. Modell., 27 (2003) 805–815.
[15] G. Ghiassi, D.K. Zimbra, H. Saidane, Urban water demand forecasting with a dynamic artificial neural network model, J. Water Resour. Plann. Manage., 134 (2008) 138–146.
[16] P. Wang, S.R. Xu, G.Q. Tang, Y. Hu, Model for forecast of urban water demanded based on set pair analysis and its application, Resour. Dev. Market, 33 (2017) 408–410.
[17] J. Liu, H.H.G. Savenije, J. Xu, Forecast of water demand in Weinan City in China using WDF-ANN model, Phys. Chem. Earth., 28 (2003) 219–224.
[18] Z.H. He, The Research for Forecasting Urban Water Demand, Anhui University of Science & Technology, Huainan, 2009 (in Chinese).
[19] R.Q. Fan, L. Zhang, Y.C. Zhang, Y.R. Wang, Research of water consumption forecasting model selection method, Sci. Technol. Innovation Herald, 18 (2012) 1–2 (in Chinese).
[20] S.L. Cruz, V. Yevjevich, Stochastic structure of water use time series, Colorado State Univ., 52 (1972) 24–32.