Pol. J. Environ. Stud. Vol. 23, No. 4 (2014), 1165-1174

Original Research

An Improved Coupling Model of Grey-System and Multivariate Linear Regression for Water Consumption Forecasting

Fang He[1, 2]*, Tao Tao[2]**

[1]College of Urban Construction, Wuhan University of Science and Technology, Wuhan, China
[2]School of Environmental Science and Engineering,
Huazhong University of Science and Technology, Wuhan, China

Received: 10 February 2014
Accepted: 15 April 2014

Abstract

Water prediction is the basis for water resource planning and management. However, water resource systems are complex. Water consumption is influenced by various factors whose relations are also complicated. The degree of influence is always different for the same factor in different areas. The effective factors of water consumption are analyzed thoroughly. The influencing factors of high degree are selected to establish an improved coupling model of grey system and multiple regressions to predict water consumption in Wuhan. The coupling model is clear in concept, simple in structure, and convenient in use. The complex relationship between water consumption and its main influencing factors is reflected. The model has the potential advantage for predicting annual water consumption. The applied research in Wuhan showed that the forecast effect of improved coupled model is good with relative error less than 1%. The model is used to predict water consumption of 2015 in Wuhan as 4.1430424 billion tons.

Keywords: water consumption forecasting, grey relational analysis, grey model, multiple linear regression, coupling model of grey-system, and multiple linear regression

Introduction

Water prediction is the basis of water resource planning and management. It is also an important part of water supply system optimization scheduling management. It can provide the water data needed in future social and economic development for government so as to provide the foundation for future urban development and to predict and deal with all kinds of water problems in advance. Therefore, it has important theoretical significance and practical value to establish an effective prediction model and improve the predictive accuracy based on the influencing factors of water consumption.

*e-mail: hefang@ wust.edu.cn
**e-mail: taocaiye@public.wh.hb.cn

To urban water consumption it can be divided into long-term, medium-term, and short-term water consumption forecasting according to the prediction time horizon and the requirements of a water supply system. Short-term prediction of water consumption plays an important role in guiding the operation of water plants and the debugging of a water supply system. It can provide real-time information for the optimal scheduling of water distribution system so as to ensure the management of water supply and the quality of service. The medium and long-term forecasting of water consumption can correspond to urban planning, so as to plan water supply, water consumption, and water savings for urban developments and guide planning and construction of water systems.

On one hand there are certain changes in the urban water consumption index itself. On the other hand, the change of the water is affected by various complex objective factors. So comprehensive analysis can grasp the main issue and eliminate interference factors, making accurate predictions for water consumption. A suitable model can be made to reflect the variation of water consumption itself. The objective factors influencing water consumption generally involve the population, gross domestic product (GDP), per capita disposable income, water resources, water price, industrial production, repeating utilization factor, energy consumption per unit of GDP, average temperature, water savings, and water supply policies, etc. It is unrealistic and unnecessary to analyze so many factors. As for the same influencing factor, it has different degrees for water consumption in different places. So the big extent and quantitative factors should be selected for modeling.

This study has two objectives:

(1) analyzing the effective factors of water consumption
(2) selecting the high degree influencing factors to propose an improved coupling model of grey system and multiple regression model to reflect the intrinsic relationship between water consumption and influencing factors so that an out-of-sample prediction may be used for medium and long-term water consumption forecasting in a fast-growing urban setting.

Literature Review

Many foreign and domestic scholars have explored the medium and long-term forecasting of water consumption for a long time. Meanwhile, they also have formed a series of effective prediction methods. However, a water resource system is very complex. Water consumption is influenced by various factors whose relations are also complicated. The degree of influence is always different for the same factor in different areas. So the applicable models are mixed in different cities. To summarize, there are several methods: time series analysis, multivariate statistical analysis, grey system prediction method, expert system method, artificial intelligence method (such as neural network, genetic algorithm, etc.), the combination model, etc.

Fernando Arbués et al. analyzed several tariff types and their objectives of the literature on residential water demand. In the research water price, income and household composition were crucial determinants of residential consumption [1]. Fox C. et al. established a property characteristic-based approach to forecast water demand. It is analyzed under a univariate classification of property type showing significant differences for properties of different size (number of bedrooms), architectural type (e.g. flats vs. terraced), and garden presence (but not for age or for garden aspect) [2]. Mehmet Ali Yurdusev and Mahmut Firat investigated an adaptive neuro fuzzy inference system (ANFIS) to forecast monthly water consumption of Izmir, Turkey. Several socio-economic and climatic factors were considered, including average monthly water bills, population, number of households, gross national product, monthly average temperature observed, monthly total rainfall, monthly average humidity observed, and inflation rate [3]. John H. Lowry Jr. et al. presented an innovative approach to estimating residential irrigation water demand for a large metropolitan area using GIS data, weather station data, and a water budget modeling approach commonly used by plant scientists and landscape management professionals. As a result, irrigation water demand was governed largely by the areal extent of irrigated landscapes, and increased urban forest abundance over time within a fixed area may decrease demand [4]. Ibrahim Almutaz, et al. presented a probabilistic-based methodology forecasting water demand of the tourist city, Mecca.

It is quantified uncertainties for the number of visitors and local population growth as well as for the selected explanatory variables of temperature, household size, and average household income are considered [5]. Christopher Bennett et al. developed the residential water end-use demand forecasting models, including two feed-forward back propagation networks and one radial basis function network. The study demonstrated that applying ANN-based modeling methodology is a feasible means of producing residential water demand end use forecasting models [6]. Lucas Beck and Thomas Bernauer combined a comprehensive set of water demand scenarios and climate change projections with a hydrological model to estimate future water availability in key parts of the Zambezi river basin (ZRB) until 2050. The results implied that an effective governance mechanism was necessary for water allocation and for dealing with flow variability [7]. Shirley Gato et al. made a new daily demand model incorporating base use values calculated using temperature and rainfall thresholds for East Doncaster, Victoria. The prediction accuracy of the new model is high. Forecasting results revealed these base values to be climate-independent but are affected by weekends and weekdays [8]. Inmaculada Pulido-Calvo, et al. developed a combination model of linear multiple regressions and feed forward computational neural networks (CNNs) trained with the Levenberg-Marquardt algorithm for the purpose of irrigation demand modeling [9]. H. Chen and Z.F. Yang constructed a model based on extended linear expenditure system (ELES) to simulate the relationship between block water price and water demand, which provide theoretical support for the decision-makers. It is used to simulate residential water demand under block rate pricing in Beijing [10]. Mohamed M. Mohamed et al. made the constant rate model of the IWR-MAIN software to forecast the water demand in the Emirate of Umm AlQuwain (UAE) for the next 20 years. The results will be taken to decide on building new desalination plants [11].

Cheng Qi and Ni-Bin Chang agreed that water demand was associated with climate changes, economic development, population growth and migration, and even consumer behavioral patterns. A new system dynamics model was proposed to reflect the intrinsic relationship between water demand and macroeconomic environment using out-of-sample estimation for long-term municipal water demand forecasts in a fast growing urban region [12]. Mahmut Firat et al. developed two types of fuzzy inference systems (FIS) to predict municipal water consumption time series including an adaptive neuro-fuzzy inference system (ANFIS) and a Mamdani fuzzy inference systems (MFIS). The forecasting results demonstrated that the ANFIS model is superior to MFIS models and can be successfully applied for prediction of water consumption time series. A series of artificial neural network (ANN) techniques was compared, including generalized regression neural networks (GRNN), Cascade correlation neural network (CCNN), and feed forward neural networks (FFNN) in monthly water consumption time series predicting. The results showed that ANN techniques can be successfully applied to establish accurate and reliable water consumption prediction models that will further be used for future water demand [13, 14]. Combination models developed better with high accuracy. L. Suganthi, et al. reviewed energy demand forecasting models for commercial and renewable energy. It is found that ARIMA models are linked with neural networks and other soft computing techniques to improve the accuracy of energy demand forecasting.

Grey prediction is yet another technique being tried successfully for energy demand analysis. Genetic algorithms, fuzzy logic, SVR, AGO, and PSO are emerging techniques in forecasting commercial and renewable energy sources. The econometric models have indicated that GNP, energy price, gross output, and population are being linked to energy demand [15]. G. Peter Zhang proposed a hybrid methodology that combines autoregressive integrated moving average (ARIMA) and artificial neural network (ANNs) models. The combination model took advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. Results indicated that the combined model can be an effective way to improve forecasting accuracy achieved by either of the models used separately [16]. Inmaculada Pulido-Calvo et al. developed a hybrid methodology combining feed forward computational neural networks (CNNs), fuzzy logic, and genetic algorithm to forecast one-day-ahead daily water demands.

To compare the combination model with other single models, results showed that the hybrid model performed significantly better than variable and multivariate autoregressive CNNs [17]. Xiaoling Wang et al. described an optimization model with the HGSAA solution to forecast the long-term water demands. This has been applied to optimize water resources in the Haihe river basin. The results demonstrated that the mean relative errors of BP and polynomial models are 2.3% and 4.9%, but that of the combined forecasting method is only 1.93%, respectively. It is indicated that the combined forecasting method can improve forecast precision [18].

Mohsen Nasseria et al. developed a hybrid model that combines extended kalman filter (EKF) and genetic programming (GP) for forecasting of water demand in Tehran. In the combination model, the EKF was applied to infer latent variables in order to forecast based on GP results of water demand. This is an important solution in the valid forecasting of complicated temporal phenomena. The results can help decision makers of water resources to reduce their risks of online water demand forecasting and optimal operation of urban water systems [19].

Methodology

Grey Relationship Analysis Method

The grey relationship analysis method is a kind of factor analysis method. The principle is to measure correlation degree based on similar or dissimilar degree of development between factors. The measurement of the correlation between two factors is called relational degree. It describes the relative changes between factors in system development, such as the change value, direction, and speed [20]. As to grey relationship analysis, it is not strict with sample size and has no restrictions on the distribution of original data. So it's widely used with simple calculation. The calculating process of grey correlation analysis is as follows: qualitative analysis, raw data processing, correlation coefficient calculating, correlation degree solving, and correlation degree sorting.

Step 1. Qualitative analysis

The reference factor sequence is $X_0 = \{x_0(1), x_0(2),..., x_0(n-1), x_0(n)\}$ and the compare factors sequence is $X_i = \{x_i(1), x_i(2),..., x_i(n-1), x_i(n)\}$, i=1,2,...,m.

Step 2. Raw data processing

Applied dimensionless method to the data, then the comparable dimensionless sequence is obtained. The purpose is to remove the effects on the evaluation results by different units and dimensions and increase the comparability between different affecting factors. The common approach is the mean method:

$$x_i(k)' = \frac{x_i(k)}{\frac{1}{n}\sum_{i=1}^{n} x_i(k)} \quad (1)$$

Step 3. Correlation coefficient calculating

The grey relational coefficient of reference sequence and compare sequence in the first k point is:

$$r(x_0(k), x_i(k)) = \frac{\min_i \min_k \Delta_{0i}(k) + \zeta \max_i \max_k \Delta_{0i}(k)}{\Delta_{0i}(k) + \zeta \max_i \max_k \Delta_{0i}(k)} \quad (2)$$

...where $\Delta_{0i}(k)$ is the absolute value of the difference between the reference sequence and the compare sequence in each moment.

$$\Delta_{0i}(k) = |x_0(k) - x_i(k)| \quad (3)$$

ζ – Distinguishing coefficient. $\zeta \in [0,1]$.

The effect of ζ is to improve the significant difference between the correlation coefficients, generally ζ=0.5.

Step 4. Correlation degree solving

The grey relational degree of the reference sequence X_0 and the compare sequence X_i is: $r_{0i} = \frac{1}{n}\sum_{k=1}^{n} \xi_{0i}(k)$,

$i = 1,2,...,m$ (4)

Step 5. Correlation degree sorting

According to the results of formula (4), the correlation degrees are ordered and the influence extent of various factors is obtained. The closer the correlation is to l, the bigger the correlation degree between the reference sequence and compared sequence, the greater the effect of the compared sequence to the reference sequence.

Multiple Regression Models

There are linear relationships between the predicted value y_t and the variables $x_{1t}, x_{2t},..., x_{nt}$ based on the multiple linear regression prediction method. The curve fitting is done between y_t and $x_{1t}, x_{2t},..., x_{nt}$. Extending the curve to future, the predicted dependent variable is obtained as the independent variables are given. The expression is:

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_n x_{nt} + \varepsilon_t \quad (5)$$

...where $\beta_0, \beta_1,..., \beta_n$ are parameters to be determined. ε is residual, obeying normal distribution, $NID(0, \sigma_s^2)$.

After y_t and $x_{1t}, x_{2t},..., x_{nt}$ are smoothed (taking the observation value of x_i identically equal to 1), the equivalent expression of formula (5) is obtained as follows:

$$Y_t = \beta_1 x_{1t} + ... + \beta_n x_{nt} + \varepsilon \quad (6)$$

...where $Y_t = y_t - \overline{y}$

The formula (6) can be expressed in the form of matrix as follows:

$$Y = BX + \varepsilon \quad (7)$$

...where $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}, \quad X = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1N} & X_{2N} & \cdots & X_{nN} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$

The parameter estimate of β is obtained by least square method.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (8)$$

The predicted value of Y is obtained by the estimate value of β.

The following six assumptions are usually met to make a multiple linear regression analysis.

(1) The error term ε is a random variable obeying normal distribution, that is to say, $\varepsilon \sim NID(0, \sigma^2)$.

(2) The mathematical expectation of error term ε is zero, $E(\varepsilon) = 0$. That is the unbiased hypothesis of ε.

(3) The variances of the error term ε is a constant independent with i, $V(\varepsilon) = \sigma^2$. That is the homoscedastic hypothesis.

(4) To ε_i and ε_j, the covariance is zero. There is no serial correlation assumption.

(5) There is no strict linear correlation and multicollinearity between any independent variables.

(6) Any independent variable is irrelevant to any error term. It's generally assumed that the independent variable is a non-random variable. So the assumption is tenable.

The correlation coefficient test and significance test are often used to test models.

Grey Model

GM (n, h) model is the basic model of the grey system theory. The first number n is the order of differential equation. The second number h is the number of variables in system. Based on time series of variables, the model is established by differential equation fitting method. GM (1, 1) model is the most common basic grey model [21]. It is modeled by a first-order differential equation containing only a single variable. It is in the special case of GM (1, n) model as $n = 1$ and is an effective model of water prediction.

There are original time series of no obvious rule from observed data, $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2),..., x^{(0)}(n)\}$

It is accumulated to generate the regular $x^{(0)}$-AGO series $x^{(1)}$.

$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2),..., x^{(1)}(n)\}$,

...where $x^{(1)}(k) = \sum_{i=1}^{n} x^{(0)}(i)$

Then the differential equation can be fitted by accumulated sequence. The sequence $x^{(1)}$ shows regular with approximate exponential growth. The solution of first-order differential equation is exactly in the exponent. So the sequence $x^{(1)}$ can be thought of as meeting linear first-order differential equation as follows:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (9)$$

...where a is the development coefficient reflecting the development trend of $x^{(0)}$ and $x^{(1)}$, b is grey action, and a and b are the parameters of the differential equation.

After determining the parameters a and b in equation (9), the fluctuation law of sequence $x^{(1)}$ is obtained with time by solving differential equations.

According to the new sequence $\{x^{(1)}(k)\}$, the white background value $z^{(1)}(k)$ is calculated in the grey model as follows:

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad (10)$$

The grey differential equation is $x^{(0)}(k) + az^{(1)}(k) = b$, (11)

Made $A=(a, b)^T$, the matrix form of equation (11) is as follows:

$$\begin{vmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{vmatrix} = \begin{vmatrix} -0.5[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -0.5[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -0.5[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{vmatrix} \bullet \begin{vmatrix} a \\ b \end{vmatrix} \quad (12)$$

That is: $Y = BA$

...where $B = \begin{vmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \ldots & \ldots \\ -z^{(1)}(n) & 1 \end{vmatrix}, \quad Y = \begin{vmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \ldots \\ x^{(0)}(n) \end{vmatrix}$

The approximate solution can be obtained by the least square method.

$$A = (B^TB)^{-1} B^TY \quad (13)$$

The response equation of the white GM(1,1) equation is the solution of differential equation in the initial conditions as $x^{(0)}(1) = x^{(1)}(1) = \hat{x}^{(1)}(1)$. Then the forecasting model of generate data sequence $x^{(1)}$ is:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (14)$$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (15)$$

...where $x^{(0)}(k)$ is the element value of the first k years in the original sequence-$x^{(0)}$. $x^{(1)}(k)$ is the element value of the first k years in the accumulated sequence-$x^{(1)}$. ($k = 1, 2, ..., n$). $\hat{x}^{(0)}(k)$ and $\hat{x}^{(1)}(k)$ are the predicted values at the time of k in the original sequence and the accumulated sequence, respectively.

Equations (14) and (15) are time response function models of the basic grey prediction model. They are the specific function equations to predict future data.

Coupling Model of Grey-System and Multivariate Linear Regression

A multiple linear regression model is built between predicted value and the influencing factors. The grey prediction method is used to forecast the independent variables. The corresponding GM (1, 1) prediction model is established to forecast the independent variables. The dependent variable can be predicted by putting the predicted independent variables to the multiple regression model. Based on the multivariate linear regression model and the GM (1, 1) model, the combination forecast model is called coupling model of grey-system and multivariate linear regression.

The general thought of the coupled prediction method is as follows:

Step 1. Taking the main influencing factors calculated by grey correlation degree as independent variables, the multivariate linear regression forecasting model is set up.

Step 2. Calculating the undetermined parameters by the least squares method, the regression equation is obtained.

Step 3. Calculating the correlation coefficient r and F, the fitting optimization level and significance of the regression are tested. Then the optimal regression prediction equation is determined.

Step 4. GM (1, 1) model is set up to predict the selected independent variables to each year.

Step 5. Plugging the predicted independent variables into the optimal equation, the predicted value of the dependent variable is obtained.

Improved Coupling Model of Grey-System and Multivariate Linear Regression

In order to eliminate the random fluctuations or error of raw data, grey-regression coupling model is improved to increase prediction accuracy. GM (1, 1) models are set up based on the raw data of the variables. Then the predicted variables are taken as the basic sequence of coupling model.

The procedures of the improved coupling model of grey-system and multivariate linear regression are as follows:

Step 1. Taking the main influencing factors calculated by grey correlation degree as independent variables, GM (1, 1) models are set up to forecast values of independent variables and dependent variable.

Step 2. Based on the predicted values of GM (1, 1) models, a multivariate linear regression forecasting model is set up.

Step 3. Calculating the undetermined parameters by the least squares method, the regression equation is obtained.

Step 4. Calculating the correlation coefficient r and F, the fitting optimization level and significance of the regression are tested. The regression model is checked.

Step 5. Plugging the predicted independent variables into the demonstrated equation, the predicted value of the dependent variable is obtained.

Case Analysis

Based on time series data from 2004 to 2011 in Wuhan, medium-term water consumption was forecasted in this paper. As the biggest city in the Midwest of China, water consumption in Wuhan is increasing with the rapid development of economy in recent years. To ensure the steady development of urbanization, water consumption must be controlled in the case of dwindling water resources. Water consumption forecasting is an important solution to the issue. Based on the analysis of influencing factors on water consumption, the big extent influencing factors are selected to establish improved grey-regression coupled model for predicting middle water consumption in Wuhan.

Grey Relationship Analysis

Step 1. Qualitative analysis

Given that the records of repeating utilization factor in Wuhan city began in 2004, time series data was selected from 2004 to 2011 to be analyzed by grey relational analysis. In this paper raw data comes from the Wuhan Statistical Yearbook of 2012 and the Wuhan Water Resources Bulletin [22, 23]. The sequence of water consumption (X_0, 10^4 m^3) changing with years is the reference sequence. The compared sequences are annual average temperature (X_1, °C), population (X_2, 10^4 people), GDP (X_3, 10^8 yuan), per capita disposable income (X_4, Yuan), repeating utilization factor (X_5, %), and energy consumption per unit of GDP (X_6, ton SCE/10^4 yuan). The raw data was shown in Table 1.

Step 2. Processing raw data

Mean method is used in processing raw data based on formula (1). The processed data is shown in Table 2.

Table 1. Statistical data of factors.

Year	X_0	X_1	X_2	X_3	X_4	X_5	X_6
2004	415438	18.3	785.9	1882.24	9564.05	73.5	1.37
2005	367204	17.8	801.36	2261.17	10849.72	75	1.36
2006	369355	18.3	818.84	2679.33	12359.98	78	1.32
2007	363883	18.6	828.21	3209.47	14357.64	84.6	1.26
2008	360467	17.7	833.24	4115.51	16712.44	86.69	1.18
2009	377944	17.9	835.55	4620.86	18385.02	87.1	1.11
2010	379346	16.6	836.73	5565.93	20806.32	87.4	1.06
2011	397345	16.3	827.24	6762.2	23738.09	88.1	0.82

Table 2. Sequence of equalization.

Year	X_0	X_1	X_2	X_3	X_4	X_5	X_6
2004	1.097	1.035	0.957	0.484	0.604	0.89	1.156
2005	0.969	1.006	0.976	0.582	0.685	0.909	1.148
2006	0.975	1.035	0.998	0.689	0.78	0.945	1.114
2007	0.96	1.052	1.009	0.826	0.906	1.025	1.063
2008	0.951	1.001	1.015	1.059	1.055	1.05	0.996
2009	0.998	1.012	1.018	1.189	1.16	1.055	0.937
2010	1.001	0.939	1.019	1.432	1.313	1.059	0.895
2011	1.049	0.922	1.008	1.74	1.498	1.067	0.692

Table 3. Sequence of absolute differences.

Year	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6
2004	0.062	0.1391	0.612	0.493	0.206	0.121
2005	0.037	0.007	0.387	0.285	0.061	0.141
2006	0.06	0.0225	0.286	0.195	0.03	0.079
2007	0.091	0.0485	0.135	0.054	0.064	0.012
2008	0.049	0.0636	0.107	0.103	0.099	0.005
2009	0.014	0.0203	0.191	0.163	0.058	0.075
2010	0.063	0.0182	0.431	0.312	0.058	0.044
2011	0.127	0.041	0.691	0.449	0.018	0.230

Step 3. Calculating correlation coefficient

The absolute difference values of equalization sequences in Table 2 are calculated on the basis of formula (3).

According to the calculation steps of grey correlation analysis and formula (2), the grey correlation degrees of the reference sequence X_0 and compared sequence X_1, X_2, X_3, X_4, X_5, X_6 are calculated. The results of grey relational coefficients are shown in Table 4.

Step 4. Solving correlation degree

Based on formula (4), the grey correlation degrees of six influencing factors are, in order, 0.653, 0.73, 0.701, 0.652, 0.733, and 0.659.

Step 5. Correlation degree sorting

The six impact factors of water consumption in Wuhan are determined from big to small in turn as repeating uti-

Table 4. Grey relational coefficients table.

Year	ζ_{01}	ζ_{02}	ζ_{03}	ζ_{04}	ζ_{05}	ζ_{06}
2004	0.618	0.367	0.472	0.406	0.391	0.507
2005	0.77	1	0.617	0.566	0.739	0.468
2006	0.629	0.831	0.717	0.681	0.909	0.618
2007	0.501	0.648	0.942	0.999	0.723	0.947
2008	0.687	0.575	0.999	0.859	0.6	1.001
2009	0.994	0.852	0.843	0.734	0.754	0.631
2010	0.615	0.873	0.583	0.538	0.753	0.755
2011	0.406	0.692	0.437	0.432	0.996	0.348
Average grey relational coefficients	0.653	0.73	0.701	0.652	0.733	0.659

Table 5. Predicted value of water consumption and different affecting factors in GM (1,1).

Year	Water consumption	Population x_2	GDP x_3	Repeating utilization factor x_5
2005	367204	801.36	2261.17	75
2006	360108.2	825.01	2719.71	81.2
2007	365826.9	826.99	3257.4	82.8
2008	371636.5	828.97	3901.39	84.44
2009	377538.3	830.95	4672.69	86.1
2010	383533.8	832.94	5596.49	87.8
2011	389624.5	834.94	6702.91	89.53
2012	396812	836.94	8028.08	91.3
2013	402097.7	838.95	9615.24	93.1
2014	408483.3	840.96	11516.2	94.94
2015	414970.2	842.97	13792.9	96.81

lization factor, population, GDP, energy consumption per unit of GDP, average temperature, and per capita disposable income.

Multiple Regression Models

The three influencing factors of repeating utilization factor, population and GDP were selected as independent variables. With water consumption as the dependent variable, a multiple linear regression the model is established. As data in 2004 is abnormal, the model is set up based on the data from 2005 to 2011 as follows.

$$y = 484701.53+11.867x_2+11.796x_3 - 2028.626x_5 \quad (16)$$

Correlation index $R^2 = 0.886$
F statistical observation value F_0=7.746
Probability of Statistic F p=0.063

The F statistical observation value 7.746 is more than $F_{0.90}(3,3) = 5.39$. The probability of statistic F is p = 0.063, less than 1. So the model can be thought of as high correlation and significant regression effect. It can be used in regression forecasting.

GM (1,1) Model

Based on raw data from 2005 to 2011, GM (1,1) is set up to forecast water consumption, population x_2, GDP x_3, and repeating utilization factor x_5 in Wuhan.

(1) Building GM (1,1) of water consumption
Estimates of the parameters a and b are as follows:

$$\begin{cases} a = -0.015756 \\ b = 351493.13 \end{cases}$$

The GM (1,1) is as follows:

Table 6. Predicted value of coupling model of gray-multiple linear regression model.

Year	2005	2006	2007	2008
Predicted value (10^4 m^3)	368737.08	361849.19	364969.48	369262.53
Year	2009	2010	2011	2012
Predicted value (10^4 m^3)	375016.77	382488.86	392054.4	404119.17

Table 7. Predicted value of improved coupling model of gray-multiple linear regression model.

Year	2005	2006	2007	2008
Predicted value (10^4 m^3)	367194.98	360115.89	365788.52	371637.05
Year	2009	2010	2011	2012
Predicted value (10^4 m^3)	377526.3	383543.18	389607.62	395776.33

$$\hat{x}^{(0)}(1) = x^{(0)}(1) = 367204$$

$$\hat{x}^{(1)}(k+1) = 22675991.33e^{0.015756k} - 22308787.33$$

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

The predicted values of water consumption from 2005 to 2015 in Wuhan are shown in Table 5.

(2) Building GM (1,1) of population x_2

Estimates of the parameters a and b are as follows:

$$\begin{cases} a = -0.002393 \\ b = 822.1057 \end{cases}$$

The GM (1,1) is as follows:

$$\hat{x}^{(0)}(1) = x^{(0)}(1) = 801.36$$

$$\hat{x}^{(1)}(k+1) = 344347.41e^{0.002393k} - 343546.05$$

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

The predicted values of population from 2005 to 2015 in Wuhan are shown in Table 5.

(3) Building GM (1,1) of GDP x_3

Estimates of the parameters a and b are as follows:

$$\begin{cases} a = -0.1804 \\ b = 2073.838 \end{cases}$$

The GM (1,1) is as follows:

$$\hat{x}^{(0)}(1) = x^{(0)}(1) = 2261.17$$

$$\hat{x}^{(1)}(k+1) = 13756.736e^{0.1804k} - 11495.566$$

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

The predicted values of GDP from 2005 to 2015 in Wuhan are shown in Table 5.

(4) Building GM (1,1) of repeating utilization factor x_5

Estimates of the parameters a and b are as follows:

$$\begin{cases} a = -0.01954 \\ b = 78.9455 \end{cases}$$

The GM (1,1) is as follows:

$$\hat{x}^{(0)}(1) = x^{(0)}(1) = 75$$

$$\hat{x}^{(1)}(k+1) = 4115.572e^{0.01954k} - 4040.572$$

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

The predicted values of repeating utilization factor from 2005 to 2015 in Wuhan are shown in Table 5.

Grey-Regression Coupled Model

Water consumption in Wuhan predicted by grey-regression coupled model is obtained after plugging the predicted different influence factors in Table 5 into formula (16). The results are shown in Table 6.

Improved Grey-Regression Coupled Model

Based on the predicted water consumption and different influence factors from 2005 to 2011 in Table 5 by GM (1,1), the multiple linear regression model is established between water consumption and three influence factors as population, GDP, and repeating utilization factor.

$$y' = 1421325 - 1883.627x_2 - 0.637x_3 + 6090.449x_5 \quad (17)$$

Correlation index $R^2 = 0.99999$

F statistical observation value $F_0 = 401834.8$

Probability of statistic F $p = 0.666\times10^{-8}$

Table 8. Predicted value of different models.

Year	Actual value	GM (1,1)	Residual	Regression in formula (16)	Residual	Coupled model in formula (16)	Residual	Coupled model in formula (17)	Residual
2005	367204	367204	0	368737.08	0.42	368737.08	0.42	367194.98	0
2006	369355	360108.2	-2.5	367791.25	-0.4	361849.19	-2	360115.89	-2.5
2007	363883	365826.9	0.53	360767.05	-0.9	364969.48	0.3	365788.52	0.52
2008	360467	371636.5	3.1	367274.56	1.89	369262.53	2.44	371637.05	3.1
2009	377944	377538.3	-0.1	372431.34	-1.5	375016.77	-0.8	377526.3	-0.1
2010	379346	383533.8	1.1	382984.8	0.96	382488.86	0.83	383543.18	1.11
2011	397345	389624.5	-1.9	395563.35	-0.4	392054.4	-1.3	389607.62	-1.9
2012	395713	396812	0.28	409474.43	3.48	404119.17	2.12	395776.33	0.02

The units of the predicted results in Table 8 are all 10^4 m^3, and the unit of residual is %.

Table 9. Predicted value of water consumption in Wuhan from 2012 to 2016.

Year	2012	2013	2014	2015
Predicted value (10^4 m^3)	395776.33	401942.02	408151.45	414304.24

Statistical analysis showed that the improved model has higher correlation, almost equal to 1. The F statistical observation value is bigger than $F_{0.99}(3,3) = 29.5$ in significance test. The probability of statistic F is almost zero. So the model can be thought of as the most significant regression effect. Formula (17) can be used to model. The predicted results of water consumption from 2005 to 2012 are shown in Table 7.

Comparison of Predicted Results

Comparing the actual value with the predicted values by GM (1, 1) model, multiple linear regression model, and two kinds of grey-regression coupled models shows results in Table 8.

The predicted results are tested by the data from 2005 to 2011 in Table 8. The prediction accuracy is higher in the four models with the average accuracy above 98%. The residual of improved grey-regression model is less overall and no more than 3%. Tested the predicted results with actual data in 2012, the residual of improved grey-regression model is -0.02%, smaller than other models. So compared to other models, the improved grey-regression model increased the accuracy and reliability of predicted results with smaller residual. So the improved grey-regression model in formula (17) is built as the water consumption prediction model in Wuhan. Plugging the predicted data in Table 5 into formula (17), water consumption in Wuhan from 2012 to 2015 is predicted in Table 9.

Conclusions

The coupling model of grey-system and multivariate linear regression is clear in concept, simple in structure, and convenient in use. The complex relationship between water consumption and its main influencing factors is reflected. The model has the potential advantage to predict annual water consumption. Based on the feature of weakening randomness and mining system evolution rules in grey model and the characteristics of strong fusion strength and penetrability in multiple regression model, the coupled model is established and improved to raise prediction accuracy. The applied research in Wuhan showed that the forecast effect of improved coupled model is good with relative error less than 1 percent. The model is used to predict water consumption (2015) in Wuhan as 4.1430424 billion tons. The results are detailed in Table 9.

Middle water consumption in Wuhan showed a little increasing trend with micro decreasing in individual years from Table 9. There are two reasons. On one hand, water consumption is inevitably increasing with the economic and rapid social development in Wuhan. On the other hand, water conservation is promoted with a number of water conservation planning and regulations to contain growth rate to a certain extent in Wuhan. They restrict each other and lead to ups and downs of water consumption.

Considering the predicted water consumption and the present condition of water resources in Wuhan, strict control on total volume must be carried out and water-saving must be increased to strengthen the construction of “two-oriented society.” One is reasonably readjusting industrial structure, changing economic development mode, and being strict with market access with rapid development. The other is to improve utilization efficiency of regional water resources by water management, technological transformation of water savings, unconventional water use, and so on.

Acknowledgements

The authors are grateful for the comments and historical data provided by the Wuhan Water Affairs Authority.

References

1. ARBUÉS F., GARCIA-VALIÑAS M. Á., MARTINEZ-ESPIÑEIRA R. Estimation of residential water demand: a state-of-the-art review. Journal of Socio-Economics, **32**, (1), 81, **2003**.
2. FOX C., MCINTOSH B. S., JEFFREY P. Classifying households for water demand forecasting using physical property characteristics. Land Use Policy, **26**, (3), 558, **2009**.
3. ALI YURDUSEV M., FIRAT M. Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: An application to Izmir, Turkey. J. Hydrol., **365**, (3), 225, **2009**.
4. LOWRY J. H. JR., RAMSEY R. D., KJELGREN R. K. Predicting urban forest growth and its impact on residential landscape water demand in a semiarid urban environment. Urban Forestry & Urban Greening, **10**, (3), 193, **2011**.
5. ALMUTAZ I., AJBAR A., KHALID Y., ALI E. A probabilistic forecast of water demand for a tourist and desalination dependent city: Case of Mecca, Saudi Arabia. Desalination, **294**, (1), 53, **2012**.
6. BENNETT C., STEWART R. A., BEAL C. D. ANN-based residential water end-use demand forecasting model. Expert Syst. Appl., **40**, (4), 1014, **2013**.
7. BECK L., BERNAUER T. How will combined changes in water demand and climate affect water availability in the Zambezi river basin?. Global Environ. Chang., **21**, (3), 1061, **2011**.
8. GATO S., JAYASURIYA N., ROBERTS P. Temperature and rainfall thresholds for base use urban water demand modelling. J. Hydrol., **337**, (3), 364, **2007**.
9. PULIDO-CALVO I., MONTESINOS P., ROLDAN J., RUIZ-NAVARRO F. Linear regressions and neural approaches towater demand forecasting in irrigation districts with telemetry systems. Biosystems Engineering, **97**, (1), 283, **2007**.
10. CHEN H., YANG Z. F. Residential water demand model under block rate pricing: A case study of Beijing, China. Communications in Nonlinear Science and Numerical Simulation, **14**, (5), 2462, **2009**.
11. MOHAMED M., AYSHA A., AL-MUALLA. Water demand forecasting in Umm Al-Quwain using the constant rate model. Desalination, **259**, (3), 161, **2010**.
12. QI C., CHANG N.-B. System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. J. Environ. Manage., **92**, (6), 1628, **2011**.
13. FIRAT M., TURAN M. E., ALI YURDUSEV M. Comparative analysis of fuzzy inference systems for water consumption time series prediction. J. Hydrol., **374**, (3), 235, **2009**.
14. FIRAT M., TURAN M. E., ALI YURDUSEV M. Comparative analysis of neural network techniques for predicting water consumption time series. J. Hydrol., **384**, (2), 46, **2010**.
15. SUGANTHI L., ANAND A. S. Energy models for demand forecasting – A review. Renew. Sust. Energ. Rev., **16**, (2), 1223, **2012**.
16. ZHANG G. P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, **50**, (1), 159, **2003**.
17. PULIDO-CALVO I., GUTIERREZ-ESTRADA J. C. Improved irrigation water demand forecasting using a soft-computing hybrid model. Biosystems Engineering., **102**, (2), 202, **2009**.
18. WANG X., SUN Y., SONG L., MEI C. An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. Part II. Model application and results. J. Environ. Manage., **90**, (8), 2612, **2009**.
19. NASSERI M., MOEINI A., TABESH M. Forecasting monthly urban water demand using Extended Kalman Filter and Genetic Programming. Expert Syst. Appl., **38**, (6), 7387, **2011**.
20. DENG J. Introduction to grey mathematical resource science. Huazhong University of Science and Technology Press, Wuhan. **2007**.
21. DANG Y. Research of Grey Forecasting and Decision Model. Science Press, Beijing. **2009**.
22. Wuhan Statistical Yearbook. Wuhan bureau of statistics. Beijing: China Statistics Press, 2012.9., No. 24, **2012**.
23. Wuhan water resources bulletin (**2004-2011**), Wuhan Water Affairs Bureau.