

# Forecast-based analysis for regional water supply and demand relationship by hybrid Markov chain models: a case study of Urumqi, China

B. Wang, L. Liu, G. H. Huang, W. Li and Y. L. Xie

## ABSTRACT

A clear understanding of regional water supply and demand trend is crucial for proper water resources planning and management in water-deficient areas, especially for Northwest China. In this study, three hybrid stochastic models (Markov chain model, unbiased Grey-Markov model and Markov model based on quadratic programming) were developed separately for predicating the available water resources, water demand, and water utilization structure in Urumqi. The novelty of this study arises from the following aspects: (1) compared with other models, the developed models would provide ideal forecasting results with small samples and poor information; (2) this study synthetically took into account water supply and demand, water utilization structure trend; (3) the prediction results were expressed as interval values for reducing the forecasting risk when carrying out water resources system planning and operational decisions. Analysis of water supply and demand in Urumqi under different reuse ratios was also conducted based on the forecasting results. The results would help managers and policy-makers to have a clear understanding of regional water supply and demand trend as well as the water utilization structure in the future.

**Key words** | forecast, Markov chain model, water supply and demand, water utilization structure

### B. Wang

Resources and Environmental Research Academy,  
China-Canada Energy and Environmental  
Research Center,  
North China Electric Power University,  
Beijing 102206,  
China

L. Liu (corresponding author)

### G. H. Huang

### W. Li

MOE Key Laboratory of Regional Energy and  
Environmental Systems Optimization,  
Resources and Environmental Research  
Academy,  
North China Electric Power University,  
Beijing 102206,  
China  
E-mail: lei.liu@dal.ca

### Y. L. Xie

School of Mechanical Engineering,  
University of Science and Technology Beijing,  
Beijing 100083,  
China

## INTRODUCTION

Water resources are increasingly being viewed as a severely stressed and scarce natural resource, especially for arid and semi-arid regions in China (Babel & Shinde 2011; Shen *et al.* 2013). Moreover, the regional population growth and economic development has so far led to high water consumption. This would put more and more pressure on regional water supply. Taking Urumqi City, a core city in an arid area of China as an example, the daily capacity of water supply has increased from  $0.74 \times 10^6 \text{ m}^3$  in 2011 to  $1.50 \times 10^6 \text{ m}^3$  in 2015, resulting in the continuous decline of groundwater level. In addition, the water utilization structure has been significantly changed since water saving policies were implemented in recent years. Given these complex and dynamic situations, efforts should be made to help policy-makers make a reliable estimate of future water

supply and demand in order to identify desired management alternatives.

Mathematical models can be a useful technique to predict and analyze regional water supply and demand problems, and previous studies made valuable attempts in developing forecasting models, such as regression analysis, artificial neural networks, time series analysis, and system dynamics models (Miaou 1990; Jain *et al.* 2001; Altunkaynak *et al.* 2005; Mohamed & Al-Mualla 2010; Tiwari & Adamowski 2013; Kim & Seo 2015). For example, Babel *et al.* (2007) developed a multivariate econometric approach to forecast and manage the domestic water use/demand of Kathmandu Valley, Nepal. Cheng & Chang (2011) proposed a novel system dynamics model to reflect the interactive relationship between water demand and macroeconomic environment

and forecast long-term municipal water demand in Manatee County. Zhang *et al.* (2013) adopted the Cobb–Douglas model to predict regional water demand and calculated the contribution rates of the regional water demand influencing factors. Nourani *et al.* (2014) applied genetic fuzzy system (GFS) model and multivariate wavelet-GFS (WGFS) model to predict the runoff discharge of two distinct watersheds. The obtained results showed that the runoff could be better forecast through the proposed WGFS model.

In general, most of the aforementioned models have been successfully applied to water demand predication and other water resources management issues; however, these models generally need large numbers of historical data and complicated input factors to make reasonable predictions. In fact, regional water supply and demand is affected by various climatic elements, socio-economic factors, and related government policies. Under the influence of water saving policies and climate changes, the developing trends of regional water demand and water utilization structure would significantly change and the available observed data may not satisfy the requirements during those modeling processes. Moreover, the forecasted results from the traditional models are typically expressed as a deterministic value rather than interval value. This would result in a false or excessive confidence in the model and inaccuracy of the forecasting results. In contrast, an interval forecasting model of water resources management has the advantage of taking the variability and uncertainties into consideration in order to reduce the forecasting errors and risk when making water resources planning and operational decisions (Alvisi & Franchini 2012; Alvisi *et al.* 2012; Xiong *et al.* 2014). Besides, although various forecasting models have been successfully adopted for the predication of water demand and annual streamflow, the integrated predication and analysis for regional water supply and demand, the water utilization structure still needs to be investigated.

Therefore, this study attempts to develop hybrid models for forecasting of streamflow, water demand, and water utilization structure. The traditional Markov chain model was selected to forecast the annual streamflow and related available water resources while the unbiased Grey-Markov model (UGMM) was applied for regional water demand prediction. As an extension of the traditional Grey model (1, 1) and Markov chain model, the UGMM can be used to forecast

the water demand with large random fluctuations. The forecasted results would be expressed as interval values with corresponding probability aiming to reduce the forecasting risk. In addition, the Markov approach based on quadratic programming model (QP-Markov model) was selected for water utilization structure prediction. The QP-Markov model minimizes the total difference between predicted value and actual value and provides a more precise one-step transition probability than the Markov chain model. The proposed models are applicable to water resources planning and management in Urumqi, China. Analysis of water supply and demand in Urumqi under different reuse ratios was conducted based on the forecasting results. The results would help managers and policy-makers obtain a clear understanding of regional water supply and demand trend as well as the water utilization structure in the future.

## METHODOLOGY

In this section, we discuss the theories of three time series models based on the Markov approach that have been identified and applied to forecast regional streamflow, water demand levels, and water utilization structure in Urumqi. The traditional Markov chain model was selected to forecast the annual streamflow and available water resources with dynamic and stochastic characteristics. The UGMM was used to forecast the regional water demand while the QP-Markov model was applied for water utilization structure prediction. The schematic of the proposed models is shown in Figure 1. Details of each model are separately described below.

### Markov chain model

Markov chain is a particular type of stochastic process, and it has been used extensively to establish stochastic models and predict future data by occurred events (Mao & Sun 2011). Regional streamflow and related available water resources are dynamic and stochastic variables due to changes in average climatic parameters, particularly temperature and rainfall. The Markov chain forecasting model can be used to forecast the regional available water

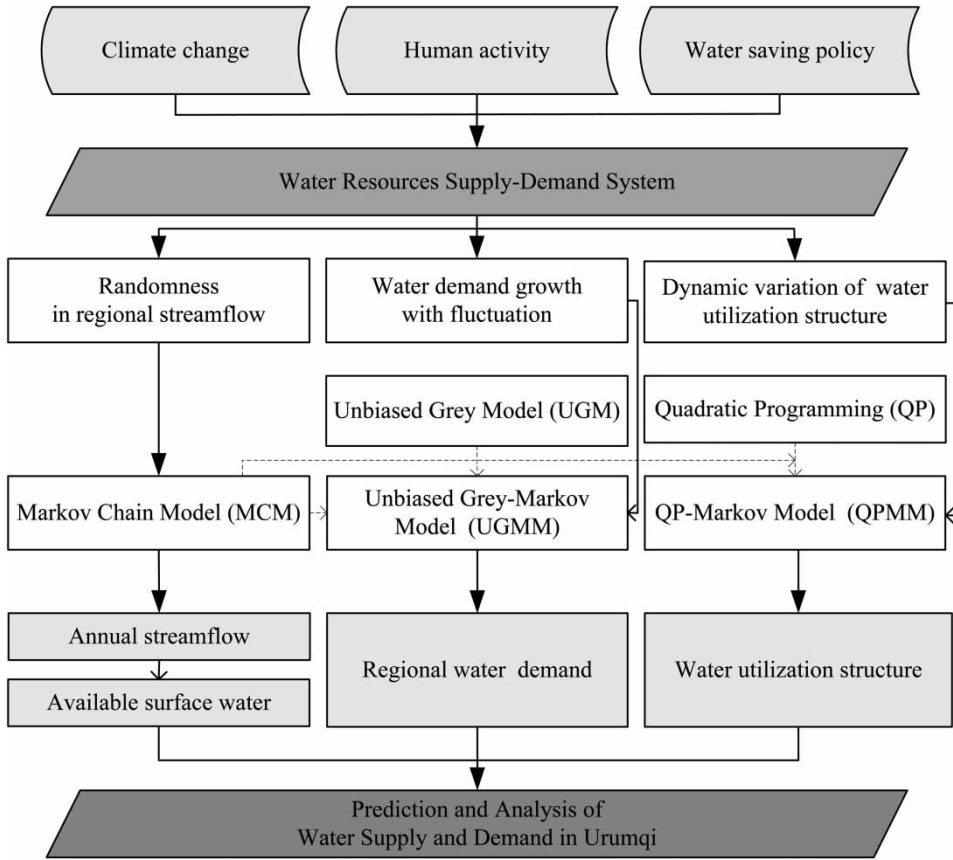


Figure 1 | Schematic of the proposed models.

resources and the associated probability distributions. Procedure of the special model is summarized as follows.

Step 1: Calculate the average value ( $\bar{X}$ ) of annual streamflow  $X(t)$  and the relative anomaly ( $e(t)$ ) as

$$e(t) = \frac{X(t) - \bar{X}(t)}{X(t)} \times 100\%, \quad t = 1, 2, 3, \dots \quad (1a)$$

Step 2: The relative anomaly is classified into three continuous interval values, which correspond to different water inflow levels (low, normal, and high flow years) according to the standard for hydrological information and hydrological forecasting (GB/T 22482-2008).

Step 3: Calculate the transition probability as

$$p_{ij} = \frac{N_{ij}}{N_i} \quad (1b)$$

$$\sum_{j=1}^{j=5} p_{ij} = 1 \quad (1c)$$

where  $p_{ij}$  denotes the transition probability that the proportion of the system belongs to state  $i$  at time  $t-1$  is transferred to state  $j$  at time  $t$ .  $N_{ij}$  is the number of observation pairs  $X(t)$  and  $X(t+1)$  with  $X(t)$  in state  $i$  and  $X(t+1)$  in state  $j$ ;  $N_i$  is the number of  $X(t)$  in state  $i$ .

Step 4: Define the transition probability matrix between different kinds of  $e(t)$  as

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \quad (1d)$$

Step 5: Calculate the probability that the state of system at time  $t+k$ :

$$p(t+k) = p(t) * P^k \quad (1e)$$

where  $p(t)$  is the initial probability distribution of the state at time  $t$ . Once the probability distribution of different states is obtained, the associated streamflow and the available water resources at time  $t + k$  can be calculated through Equation (1e).

### An unbiased Grey-Markov model

The unbiased Grey model (UGM) (1, 1) is a variant of GM (1, 1) that adopts the essential part of Grey system theory (Ji et al. 2000; Mu 2003). This model eliminates the deviations of the conventional GM (1, 1) by optimizing the background value of Grey function (Ji et al. 2001). The UGM (1, 1) can be used to forecast the developing trends of regional water demand with relatively few data ( $n \geq 4$ ). The developed model is summarized as follows:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)) \tag{2a}$$

where  $x^{(0)}$  is the number sequence of historical data and  $n$  is the sample size of the data.  $x^{(1)}$ , the accumulated sequence, is obtained by 1-AGO (one time accumulated generating operation). It is obvious that accumulated sequence is monotonically increasing.

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)) \tag{2b}$$

where

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k = 1, 2, 3, \dots, n \tag{2c}$$

The first order Grey differential equation of GM (1, 1) and its whitening equation can be expressed as:

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \dots, n \tag{2d}$$

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(k) = b \tag{2e}$$

In Equation (2e),  $a$  and  $b$  are the coefficients obtained by using the least square method,  $z^{(1)}(k)$  is the background

value of grey derivate, generally expressed as  $\frac{1}{2} [x^{(1)}(k) + x^{(1)}(k - 1)]$ ,  $k = 2, 3, \dots$  (Kumar & Jain 2010).

The least squares method is described below:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \tag{2f}$$

where

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(1)) & 1 \\ -\frac{1}{2}(x^{(1)}(3) + x^{(1)}(2)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(n) + x^{(1)}(n - 1)) & 1 \end{bmatrix}; \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} \tag{2g}$$

The associated coefficients in UGM (1, 1) can be expressed as:

$$a' = \ln \frac{2 - a}{2 + a} \tag{2h}$$

$$b' = \frac{2b}{2 + a} \tag{2i}$$

According to the former equations, the data of  $\tilde{x}^{(0)}(k)$  at time  $k$ :

$$\tilde{x}^{(0)}(k) = \begin{cases} x^{(0)}(1), & k = 1; \\ b'e^{a'(k-1)}, & k = 2, 3, \dots, n; \end{cases} \tag{2j}$$

Generally, the relative error  $e(k)$  and the mean absolute percentage error (MAPE) are used to examine and check the predication accuracy of the forecasting model. In this study, the following equations are utilized to calculate the relative error and MAPE:

$$e(k) = \frac{x^{(0)}(k) - \tilde{x}^{(0)}(k)}{x^{(0)}(k)} \times 100\%, \quad k = 2, 3, 4, \dots, n \tag{2k}$$

$$MAPE(\%) = \frac{1}{n} \sum_{k=1}^n \left| \frac{x^{(0)}(k) - \tilde{x}^{(0)}(k)}{x^{(0)}(k)} \right| \tag{2l}$$

Although the UGM (1, 1) is a simple and relatively accurate method to forecast the regional water demand with an

exponential increasing trend, the accuracy may decrease when the historical data sequence fluctuates with time. The Markov chain forecasting model can also be used to forecast the water demand with randomly varying time series. Therefore, an UGMM can be obtained through the incorporation of UGM (1, 1) and Markov chain model. The relative error series obtained through the former UGM (1, 1) can be used to predict the relative error in future by the Markov chain model. Thus, the revised water demand can be obtained through the integration of UGM and Markov chain model. The steps of the Markov chain model in UGMM are similar to the former steps of a typical Markov model. In order to maintain the integrity of the UGMM,  $\otimes$  is introduced and the specific process is described below.

Step 1: Establish a Markov model based on relative error series  $e(k)$ : divide the relative error series into  $n$  states. Any state of relative error can be denoted as an interval:

$$\begin{aligned} \otimes_j &= [\otimes_{j-}, \otimes_{j+}]; \quad \otimes_{j-} = \tilde{x}^{(0)}(j) + A_j; \\ \otimes_{j+} &= \tilde{x}^{(0)}(j) + B_j \end{aligned} \tag{2m}$$

Step 2: Assume  $n_i$  is the data number of historical sequence, the transition probability from  $\otimes_i$  to  $\otimes_j$  can be established:

$$p_{ij}(m) = \frac{n_{ij}(m)}{n_i} \quad i, j = 1, 2, 3, \dots \tag{2n}$$

where  $p_{ij}(m)$  is the transition probability of state  $\otimes_j$  transferred from state  $\otimes_i$  for  $m$  steps,  $m$  is the number of transition steps each time,  $n_i$  is the number of data in state  $\otimes_i$ ,  $n_{ij}(m)$  is the number of historical data of state  $\otimes_j$  transferred from state  $\otimes_i$  for  $m$  steps, its transition probability matrix can be expressed as follows:

$$P(m) = \begin{bmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \dots & p_{1n}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \dots & p_{2n}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}^{(m)} & p_{n2}^{(m)} & \dots & p_{nn}^{(m)} \end{bmatrix} \tag{2o}$$

Generally, it is necessary to observe one-step transition matrix,  $P(1)$ . Suppose the object to be forecasted is in state  $\otimes_r$ ,  $1 \leq r \leq n$ , row  $r$  in matrix  $P(1)$  should be considered. If  $\max_j P_{rj}(1) = P_{r1}(1)$ , then what will most probably

happen in the next moment is the transition from state  $\otimes_r$  to state  $\otimes_1$ .

Step 3: Calculate the predicted probability of the future relative error. The relative error zone  $[\otimes_{j-}, \otimes_{j+}]$  in the future is predicted by studying the transition probability matrix  $P(m)$ . Therefore, the forecasting result of water demand can be presented as a single value through the medium of relative error zone or an interval according to Equation (2k).

### Markov approach based on quadratic programming model

Water resources consumption is generally composed of agriculture, industry, municipality, and eco-environment. In Urumqi, recent years have seen not only increasing water consumption but also the varying structure of water consumption. Under the influence of a series of policies, the proportion of agricultural water consumption is decreasing while that of municipality and environment is increasing year by year. The water utilization structure in the future depends mainly on the status in recent years rather than in past states. Therefore, a novel stochastic model, incorporating the quadratic programming model into Markov approach, was proposed to predict the utilization structure of water resources in Urumqi. The developed model is summarized as follows.

Step 1: Collect the annual data of water utilization structure  $w_i(t)$  ( $i = 1, 2, 3, 4; t = 1, 2, \dots, n$ ).

Step 2: Define the transition probability matrix between different water users as

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \tag{3a}$$

where  $p_{ij}$  means one-step transition probability that the proportion belongs to user  $i$  at  $t-1$  period is transferred to user  $j$  at  $t$  period.

Step 3: Calculate the one-step transition probability using the quadratic programming model. The predicted proportions of each user at time  $t$  is expressed as

$$\bar{w}_j(t) = \sum_{i=1}^4 \bar{w}_i(t-1)P_{ij} \tag{3b}$$

To minimize the total difference between the real value  $w_j(t)$  and simulated value of  $\bar{w}_j(t)$ , the quadratic programming can be established as follows:

$$\min \sum_{t=1}^n \sum_{j=1}^4 (w_j(t) - \sum_{i=1}^4 \bar{w}_i(t-1)P_{ij})^2 \quad (3c)$$

$$\text{st:} \begin{cases} \sum_{i=1}^4 P_{ij} = 1 \\ P_{ij} \geq 0 \end{cases} \quad (3d)$$

Then we can get the transition probability value.

Step 4: The absolute weighted error (*AE*) of each year and mean absolute weighted error (*MAE*) is defined as indicators for accuracy assessment.

$$AE(k) = \frac{1}{4} \sum_{i=1}^4 |w_i(k) - \bar{w}_i(k)| \quad (3e)$$

$$MAE(k) = \frac{1}{n} \sum_{t=1}^n AE(k) \quad (3f)$$

The real water utilization structure can be forecasted based on the solved transition probability matrix and the water utilization structure in the initial year.

## MODEL APPLICATION

### Study area description

Urumqi City (42°45'–44°08'N; 86°37'–88°58'E), one of the core cities in the Silk Road economic belt, is the capital of Xinjiang Uygur Autonomous Region. It is located in the northwest of China, overlooking the Junggar Basin to the North and the Tianshan Mountain to the South (Figure 2). It covers an area of  $14.20 \times 10^3 \text{ km}^2$  with approximately 50.00% of the total area being mountainous. The city is situated in the temperate continental drought climatic zone with large diurnal temperature differences and four distinctive seasons. The annual average precipitation is roughly 236 mm, however, the annual evaporation exceeds 2,000 mm. There are 46 rivers in Urumqi, connected to the following large

water bodies: Urumqi River, Toutun River, Ala River, Baiyang River, and Chaiwopu Lake. These are the main water sources for supporting the basic water requirements of production and human life in Urumqi.

In recent years, Urumqi has entered into the fast lane of economic growth since the strategies of the Great Western Development and Silk Road Economic Belt were promulgated. The process of industrialization and urbanization is accompanied by massive water consumption. For example, the annual average available water resources are  $11.20 \times 10^8 \text{ m}^3$ ; however, in 2015 the estimated water demand reached  $12.20 \times 10^8 \text{ m}^3$ . Therefore, water shortage becomes an important restrictive factor for economic and social sustainable development. In Urumqi, agricultural water demand represent about 60% of the total water consumption, followed by regional domestic and industrial sectors. It is unsustainable in the long term due to the limited water resources and irrational water utilization structure. Therefore, effective measures are needed for the sustainable utilization of all the water sources.

### Data

The yearly time series data of regional streamflow, total water consumption, and proportions of water utilization structure in Urumqi were obtained from the Urumqi Water Resources Bulletin. The data sequence of annual streamflow is available for the period 1956–2000. Figure 3 presents the water consumption of Urumqi for the period 1995–2014. It is obvious that the water consumption in Urumqi has kept increasing with large random fluctuations during the past two decades. There is huge pressure on water resources supply due to the increasing water demand. Faced with water scarcity, a multitude of water-saving policies and measures has been applied in Urumqi and the growth rate of water consumption has reduced in recent years. Figure 4 depicts the variations of water utilization structure in Urumqi from 1999 to 2013. Since 1999, the proportion of agriculture has kept at approximately 60.00% of the total water consumption in Urumqi. In recent years, the consumption ratio of the agriculture sector dropped from 64.37% in 2010 to 56.30% in 2013, while the proportions of industry, municipality, and environment increased slightly.

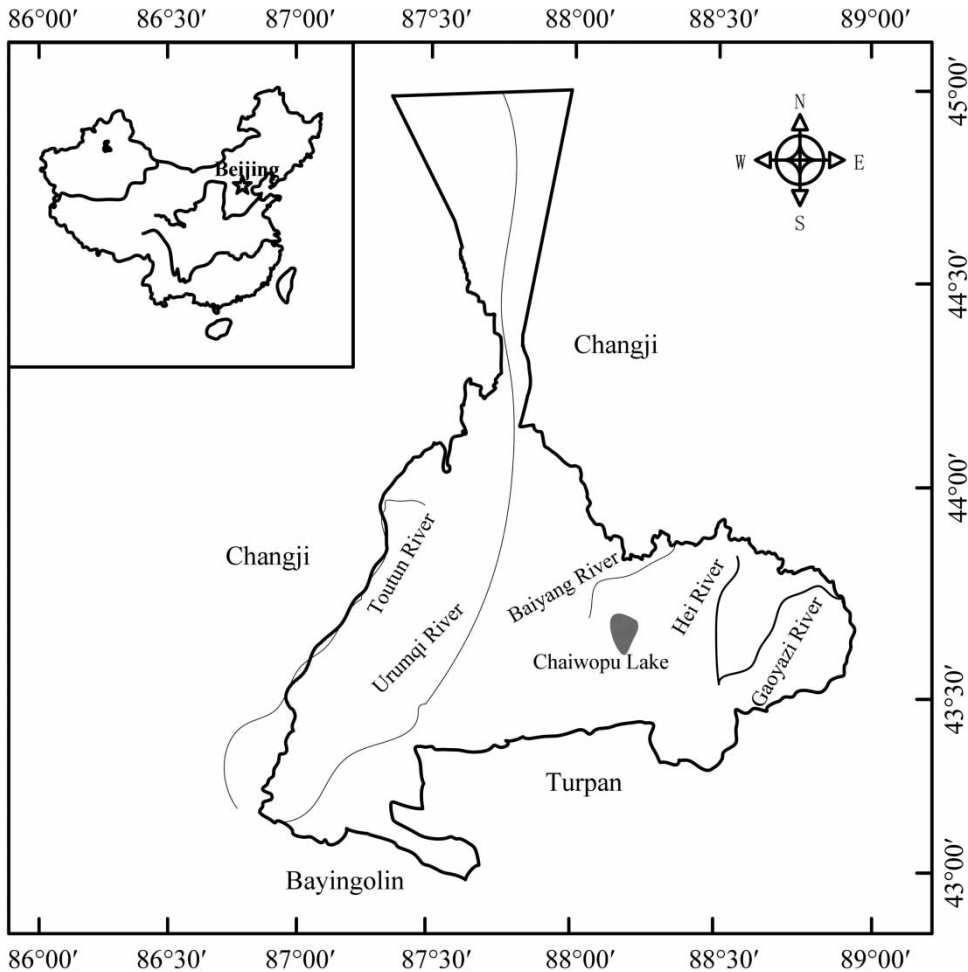


Figure 2 | Location of the study area.

## RESULTS AND DISCUSSION

### Available water resources prediction during 2015–2020

Following the basic procedure of Markov chain model defined in the section Methodology, we collected the time series data of annual streamflow from 1956 to 2000 and established the Markov chain model (Figure 5). The average annual streamflow ( $\bar{X}$ ) is  $10.20 \times 10^8 \text{ m}^3$  and the interval of relative anomaly ( $e(k)$ ) is  $[-28.23, 33.68]\%$ . According to the standard for hydrological information and hydrological forecasting (GB/T 22482-2008), partitions of states were done by forming three contiguous intervals, which correspond to high-, normal-, and low-flow level years, respectively. Table 1 shows the relative anomaly, the related

streamflow, and available water resources under different water inflow-level years. The available water resources were obtained through deductions of unavailable water resources from the annual streamflow.

A visual inspection of Figure 5 reveals that the number of historical data in each state and one-step transition probability for every state can be calculated by Equation (1a). The one-step transition probability matrix is presented as follows:

$$P = \begin{bmatrix} 0.2727 & 0.6363 & 0.0909 \\ 0.2400 & 0.5600 & 0.2000 \\ 0.1250 & 0.5000 & 0.3750 \end{bmatrix}$$

It is obvious that the observed data for 2000 lie in state 3. The state in 2001 and 2002 can be calculated through

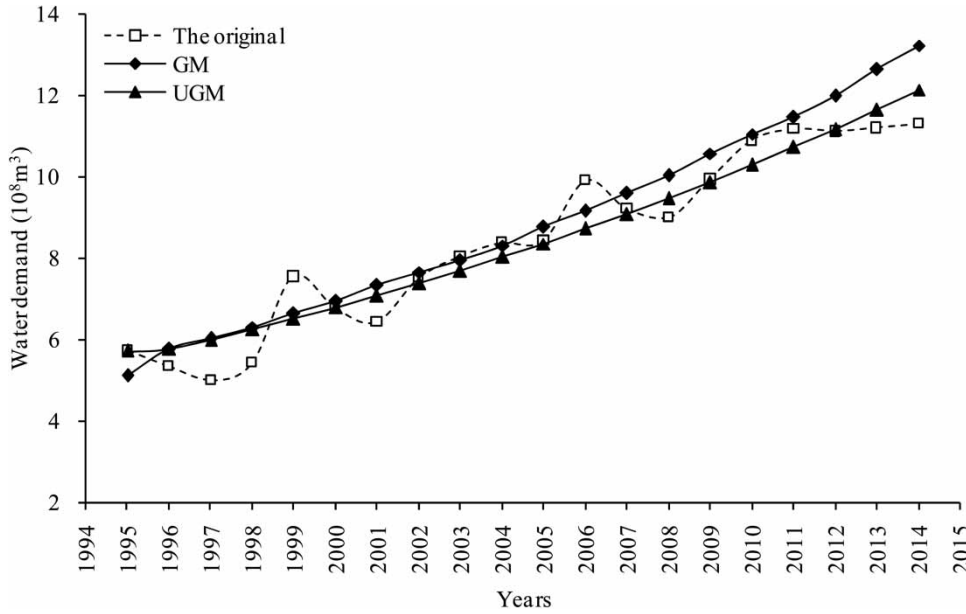


Figure 3 | Actual and forecasted water demands in Urumqi.

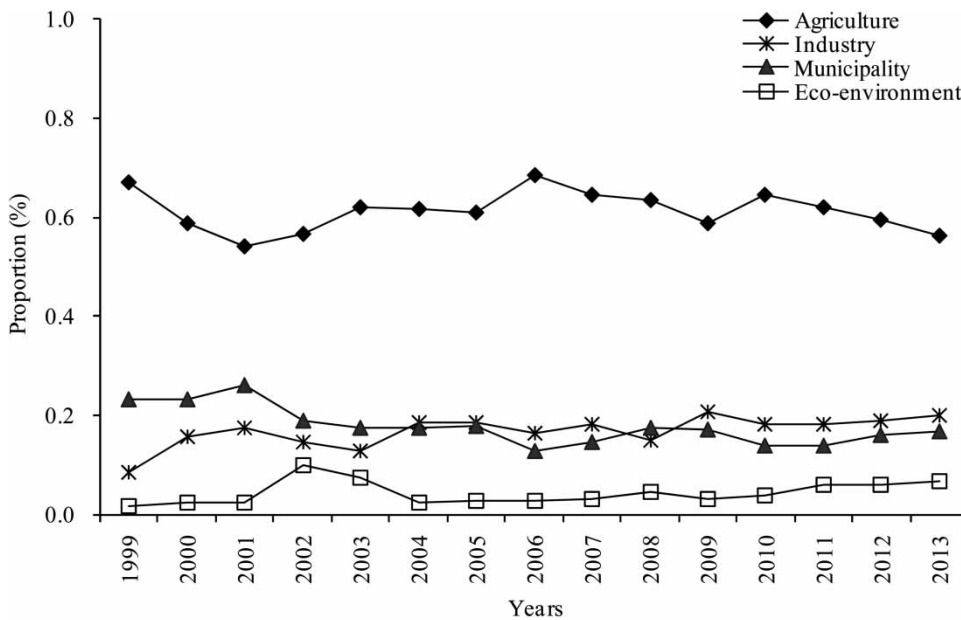


Figure 4 | Variations of water utilization structure in Urumqi.

Equation (1d). The forecasting state of 2001 and 2002 would be normal water inflow-level year and the associated annual streamflow would be  $[9.18, 11.22] \times 10^8 \text{m}^3$ . The actual streamflow in 2001 and 2002 were  $10.50 \times 10^8 \text{m}^3$  and  $9.20 \times 10^8 \text{m}^3$ , respectively. Results indicated that the Markov chain model can be used for forecasting the

annual streamflow and the associated probabilities. The streamflow in 2012 was selected as initial point and the probabilities of water inflow states for 2015–2020 are represented in Table 2. The corresponding available water resources would be obtained when the water inflow state is recognized.



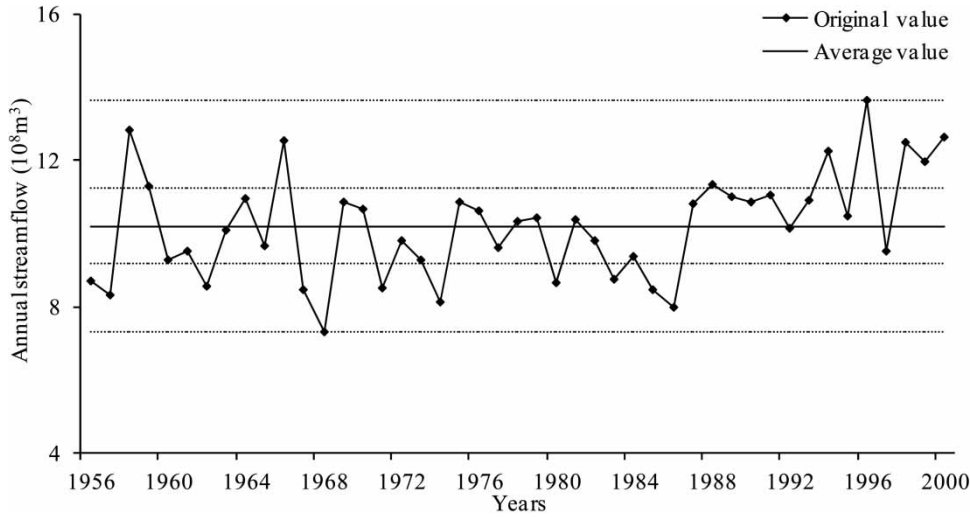


Figure 5 | Time series variations of streamflow for Urumqi.

Table 1 | Relative anomaly and the available water resources under different water inflow years

States	Relative anomaly (%)	Streamflow value (10 <sup>8</sup> m <sup>3</sup> )	Available surface water (10 <sup>8</sup> m <sup>3</sup> )
Low flow	[-28.23, -10.00]	[7.33, 9.18]	[3.50, 5.54]
Normal flow	[-10.00, 10.00]	[9.18, 11.22]	[5.54, 7.58]
High flow	[10.00, 33.68]	[11.22, 13.64]	[7.58, 10.13]

Table 2 | Forecasted probabilities of low, normal and high flow years in 2015–2020

Years	Low flow	Normal flow	High flow
2015	0.2234	0.5646	0.2120
2016	0.2229	0.5643	0.2128
2017	0.2228	0.5643	0.2129
2018	0.2228	0.5642	0.2130
2019	0.2228	0.5642	0.2130
2020	0.2228	0.5642	0.2130

### Annual water demand forecast during 2015–2020

Based on Equations (2a)–(2j) and the historical data of water consumption from 1994 to 2014, the associated coefficients in GM (1, 1) and UGM (1, 1) were obtained as follows:

$$a = -0.0415, b = 5.4004$$

$$a' = 0.0415, b' = 5.5148$$

Therefore, the GM (1, 1) and UGM (1, 1) are established according to Equation (2j) as follows:

$$x^{(0)}(k + 1) = 5.5218e^{0.0415k}, k = 0, 1, 2, \dots, n;$$

$$\tilde{x}^{(0)}(k) = \begin{cases} 5.7050, & k = 1; \\ 5.5148e^{0.0415(k-1)}, & k = 2, 3, \dots, n; \end{cases}$$

The forecasting trend curves for water demand were built by GM (1, 1) and the associated variant UGM (1, 1). These curves are shown as solid lines in Figure 3. It should be noted that the time series of water consumption in Urumqi vacillated around the trend curves. The fluctuations in actual water consumption would reduce the forecasting accuracy and the associated variant. Following Equation (2l), the values of MAPE in GM (1, 1) and UGM (1, 1) are 8.12% and 5.97%, respectively. The relative errors of prediction from GM (1, 1) are [-20.46, 11.91] %, and there have been 12 results whose relative errors exceed the range [-5.00, 5.00] %. The relative errors from UGM (1, 1) are [-19.85, 13.65] % with nine results beyond the range [-5.00, 5.00] %. The results indicated that neither of the two forecasting models have performed well for random time series prediction.

A hybrid model, incorporating the UGM (1, 1) and Markov chain model into a modeling framework, is applied for regional water demand prediction. As mentioned in the section An unbiased Grey-Markov mode, the ranges of relative error series from the UGM (1, 1) were divided into three contiguous intervals. The three interval values are  $[-20, -5]$  %,  $[-5, 5]$  %,  $[5, 15]$  %, respectively. The number of related error in each state ( $n_i$ ) is as follows:

$$n_1 = 5, n_2 = 10, n_3 = 3$$

Next, the numbers of one step transiting between each state ( $n_{ij}$ ) can be calculated and the one-step transition probability for each state can be calculated by using Equation (2n). Therefore, the calculated probability values have been presented in the one-step transition matrix as follows:

$$P = \begin{bmatrix} 0.40 & 0.40 & 0.20 \\ 0.30 & 0.50 & 0.20 \\ 0 & 1 & 0 \end{bmatrix}$$

The relative error zone in the future would be predicated using the one-step or  $m$ -steps transition matrix. Correspondingly, the water demand and the related probabilities would be obtained. In this study, the water consumption in 2014 was chosen as the historical point. The forecasting results of water demand during 2015–2020 are presented in Table 3. It is obvious that the forecasted values of GM (1, 1) would always lie in the intervals of the highest water demand with lowest probabilities. However, the results of UGM (1, 1) and the modified values would remain at the medium demand level with maximum probability. A comparison of the predicted values form GM (1, 1) and UGM (1, 1) indicated that the forecasted results of the hybrid model would be presented as a single value or interval values. The interval values of water demand, taking the variability or uncertainties into account, have the advantage of reduction in risk when making water resources planning and operation decisions. The results obtained in this study can also be compared with the predicated values in urban comprehensive planning. The forecasted water demand of Urumqi in 2015 and 2020 would be  $12.20 \times 10^8 \text{ m}^3$  and  $15.00 \times 10^8 \text{ m}^3$ , respectively. It is clear that the values lie in

**Table 3** | Forecasted values of water demand and associated probability in 2015–2020 ( $10^8 \text{ m}^3$ )

Years	GM	UGM	UGMM		
			Modified value	Interval value	Probability (%)
2015	13.82	12.65	12.28	[10.54, 12.05]	40.00
				[12.05, 13.31]	40.00
				[13.31, 15.81]	20.00
2016	14.46	13.18	12.94	[10.99, 12.56]	28.00
				[12.56, 13.88]	56.00
				[13.88, 16.48]	16.00
2017	15.24	13.74	13.50	[11.45, 13.09]	28.00
				[13.09, 14.47]	55.20
				[14.47, 17.18]	16.80
2018	15.94	14.33	14.07	[11.94, 13.64]	27.76
				[13.64, 15.08]	55.60
				[15.08, 17.91]	16.64
2019	16.54	14.93	14.67	[12.44, 14.22]	27.78
				[14.22, 15.72]	55.55
				[15.72, 18.67]	16.67
2020	17.30	15.57	15.29	[12.97, 14.82]	27.78
				[14.82, 16.39]	55.56
				[16.39, 19.46]	16.66

the forecasted intervals with maximum probability. The comparison clearly points to the enormous potential that the model possesses in water demand forecasting and can be considered as a viable alternative.

### Simulation and forecasting of the water utilization structure

Following the process of the section Markov approach based on quadratic programming model, the annual sequence data of water utilization structure during 1999–2008 were collected and the QP-Markov model was established. To protect the water demands of municipality and environment, a few of the transition probabilities from municipality and environment to agriculture and industry are specified to be 0, as  $P_{31} = P_{32} = P_{41} = P_{42} = 0$ . By solving the developed model, the transition probability would be obtained, as shown in Figure 6. The results show that the transition of eco-environment to municipality would be the largest component in the entire transitions. It indicates that the water resources belonging to the eco-environment would be occupied by the growing demand of domestic life, leading to the deterioration of eco-environment in

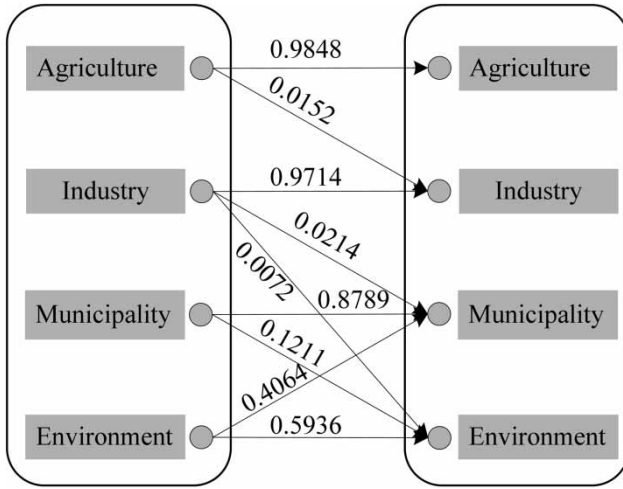


Figure 6 | Transition schematic of different water users.

Urumqi. According to the water utilization structure in 2008 and the transition probability in Figure 6, the proportions of different water users in 2009–2013 would be simulated. The historical data and simulated results are shown in Table 4. The mean absolute weighted error is 2.08%, which is an acceptable forecasting accuracy that the developed model can be used to forecast the water utilization structures during 2015–2020 (as shown in Figure 7). The results indicate that the water consumption proportions for agriculture will continue to decrease. For example, the values would reduce to 58.71% in 2016 and 55.22% in 2020. However, the water consumption proportions of industry, municipality, and eco-environment would slowly increase.

Table 4 | Simulated data of water utilization structure in 2008–2013 (%)

Year	Historical data				Forecasting results			
	Agriculture	Industry	Municipality	Eco-environment	Agriculture	Industry	Municipality	Eco-environment
2009	58.92	20.64	17.22	3.22	62.38	15.43	17.35	4.84
2010	64.37	18.00	13.74	3.89	61.42	15.93	17.56	5.09
2011	62.07	18.35	13.75	5.83	60.49	16.42	17.83	5.26
2012	59.41	18.90	15.84	5.85	59.57	16.86	18.17	5.40
2013	56.30	20.02	16.89	6.79	58.66	17.28	18.53	5.53
							MAE	2.08

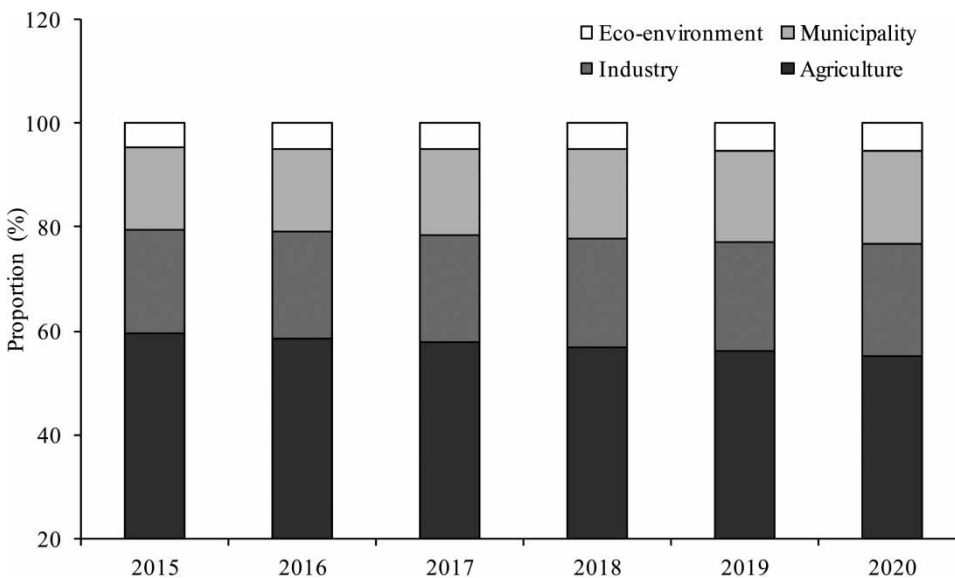


Figure 7 | Forecasted results of water utilization structure in 2015–2020.

### Water supply and demand analysis under different scenarios

According to the master planning of Urumqi, three reuse ratios of domestic water were defined as 20% (of the baseline scenario), 25% and 30% during 2016 to 2020. Figure 8 presents the regional water demand and supply under different scenarios. WS-L, WS-M, WS-H denote the low, medium, and high level of water supply. WD-L, WD-M, and WD-H represent the low, medium, and high level of water demand, respectively. It is apparent that water deficit

in Urumqi would greatly decrease when the water supply level increased. There is even surplus water under high supply level. However, an increase in the reuse ratio would alleviate the water shortage to some extent. For example, in 2020, under the medium water supply level, the amounts of available water resources would be [12.16, 14.61], [12.27, 14.75], and [12.39, 14.88]  $\times 10^8 \text{ m}^3$  when the reuse ratios are 20%, 25%, and 30%, respectively. The related water demand would be [12.97, 14.82]  $\times 10^8 \text{ m}^3$ . The water deficit would slightly decrease with the increase of reuse ratios. Therefore, it can be concluded that water

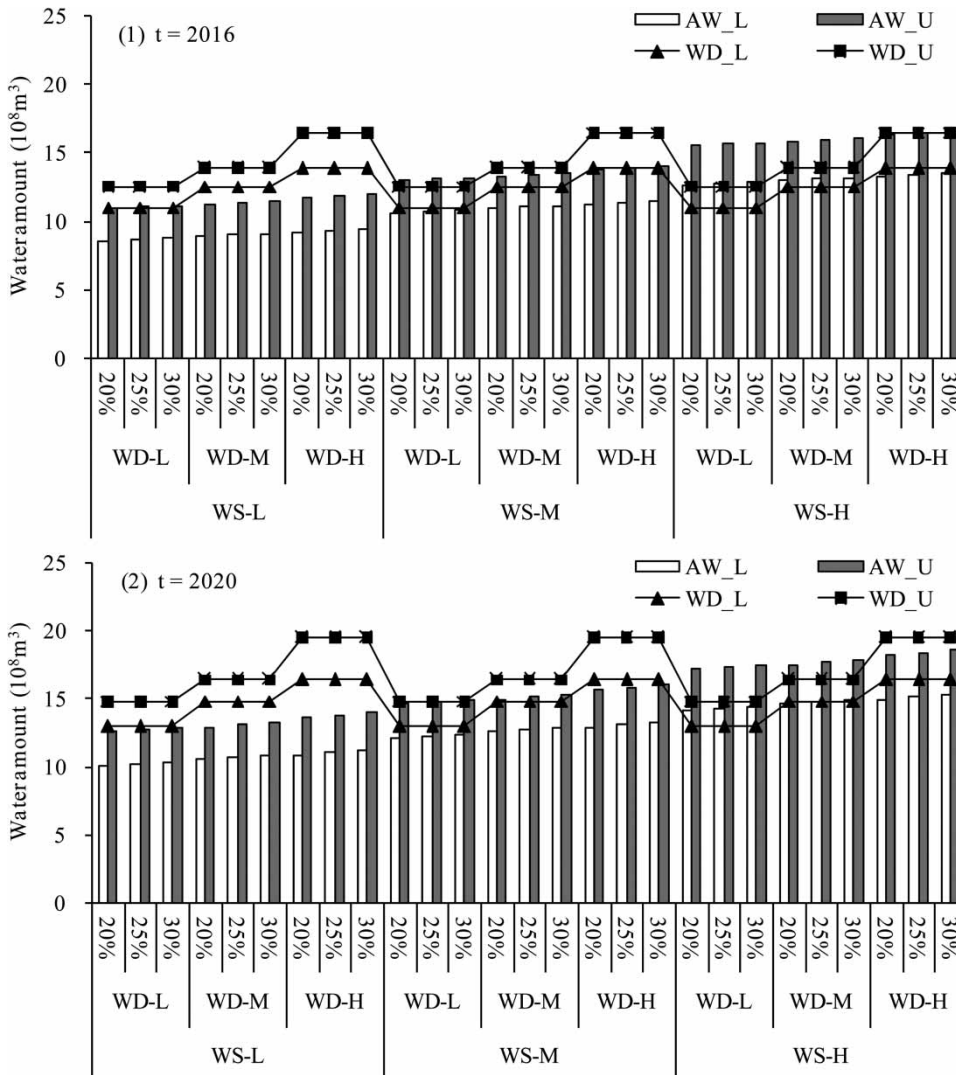


Figure 8 | Water supply and demand analysis under different scenarios.

shortage would not highly decrease unless the reuse ratio increased markedly in the future. Under the low water supply level, the total amounts of water demand in Urumqi would always be larger than regional available water resources. For example, in 2016, the total water demand would be [10.99, 12.56], [12.56, 13.88], and [13.88, 16.84]  $\times 10^8 \text{m}^3$  under the three demand levels, respectively. However, the available water resources in Urumqi would be [8.77, 11.14], [9.10, 11.42], and [9.38, 11.97]  $\times 10^8 \text{m}^3$  when the reuse ratio is 30%. The results mean that Urumqi would be confronted with severe water shortage despite the reuse ratio increases. Under the low supply level, improvement in the reuse ratio would have no significant effect on relieving water shortage. Therefore, increasing diverted water amounts would be an applicable measure for alleviating the water shortage in Urumqi.

## CONCLUSIONS

In this study, three hybrid Markov chain models have been proposed for predicting regional streamflow, water demand, and water utilization structure in Urumqi. The hybrid models required less computational data than other traditional models and took into account water supply and demand, and water utilization structure trend. Results show that 2015–2020 would correspond to the normal flow year and the associated streamflows were [9.18, 11.22]  $\times 10^8 \text{m}^3$  with maximized probability. The water demand of Urumqi during 2015–2020 would keep increasing slightly. For the water utilization structure, the proportions of agriculture would continue to decrease while the proportions of industry, municipality, and environment would increase slowly. Analyses of water supply and demand in Urumqi under different reuse ratios were conducted, and results indicated that water deficit would not be dramatically relieved despite the reuse ratio increases. The reduction of water consumption and an increase of diverted water would be the applicable measures for alleviating the water shortage in Urumqi. In general, the results presented in this paper would help managers and policy-makers to have a clear understanding of the regional water supply and demand trend as well as the water utilization structure.

## ACKNOWLEDGEMENTS

This research was supported by the Fundamental Research Funds for the Central Universities (2015XS99), National Natural Science Foundation of China (61471171). The authors are extremely grateful to the editor and the anonymous reviewers for their insightful comments and suggestions.

## REFERENCES

- Altunkaynak, A., Özger, M. & Çakmakc, M. 2005 Water consumption prediction of Istanbul city by using fuzzy logic approach. *Water Resour. Manage.* **19** (5), 641–654.
- Alvisi, S. & Franchini, M. 2012 Grey neural networks for river stage forecasting with uncertainty. *J. Phys. Chem. Earth* **42** (208), 108–118.
- Alvisi, S., Creaco, E. & Franchini, M. 2012 Crisp discharge forecasts and grey uncertainty bands using data-driven models. *Hydrol. Res.* **43** (5), 589–602.
- Babel, M. S., Gupta, A. D. & Pradhan, P. 2007 A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. *Water Resour. Manage.* **21** (3), 573–589.
- Babel, M. S. & Shinde, V. R. 2011 Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. *Water Resour. Manage.* **25** (6), 1653–1676.
- Cheng, Q. & Chang, N. B. 2011 System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. *J. Environ. Manage.* **92** (6), 1628–1641.
- Jain, A., Joshi, U. & Varshney, A. 2001 Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resour. Manage.* **15** (5), 299–321.
- Ji, P. R., Huang, W. S. & Hu, X. Y. 2000 An unbiased grey forecasting model. *Syst. Eng. Elec.* **22** (6), 6–7.
- Ji, P. R., Huang, W. S. & Hu, X. Y. 2001 A study on the properties of grey forecasting model. *Syst. Eng. Theory Pract.* **21** (9), 105–108.
- Kim, S. E. & Seo, W. 2015 Artificial neural network ensemble modeling with exploratory factor analysis for stream flow forecasting. *J. Hydroinform.* **17** (4), 614–639.
- Kumar, U. & Jain, V. K. 2010 Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **35** (4), 1709–1716.
- Mao, Z. L. & Sun, J. H. 2011 Application of Grey-Markov model in forecasting fire accidents. *Proc. Eng.* **11**, 314–318.
- Miaou, S. P. 1990 A stepwise time series regression procedure for water demand model identification. *Water Resour. Res.* **26** (9), 1887–1897.

- Mohamed, M. M. & Al-Mualla, A. A. 2010 Water demand forecasting in Umm Al-Quwain (UAE) using the IWR-MAIN specify forecasting model. *Water Resour. Manage.* **24** (14), 4093–4120.
- Mu, Y. 2003 A direct modeling method of the unbiased GM(1,1). *Syst. Eng. Elec.* **25** (9), 13–16.
- Nourani, V., Tahershamsi, A., Abbaszadeh, P., Shahrabi, J. & Hadavandi, E. 2014 A new hybrid algorithm for rainfall-runoff process modeling based on the wavelet transform and genetic fuzzy system. *J. Hydroinform.* **16** (5), 1004–1024.
- Shen, Y., Li, S., Chen, Y., Qi, Y. & Zhang, S. 2013 Estimation of regional irrigation water requirement and water supply risk in the arid region of Northwestern China 1989–2010. *Agr. Water Manage.* **128** (10), 55–64.
- Tiwari, M. K. & Adamowski, J. 2013 Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resour. Res.* **49** (10), 6486–6507.
- Xiong, T., Bao, Y. & Hu, Z. 2014 Interval forecasting of electricity demand: a novel bivariate EMD-based support vector regression modeling framework. *Int. J. Elec. Power Energ. Syst.* **63** (12), 353–362.
- Zhang, Q., Diao, Y. & Dong, J. 2013 Regional water demand prediction and analysis based on Cobb-Douglas model. *Water Resour. Manage.* **27** (8), 3103–3113.

First received 5 October 2015; accepted in revised form 30 January 2016. Available online 11 March 2016