Integrated RAGA-PP water demand forecast model
(case study: Shaanxi Province, China)
Jun Yang, Yanning Mao, Yuqi Ma, Wei Wu and Yuan Bai

ABSTRACT
The demand for water resources in Shaanxi Province increases greatly due to the continuous growth of its population and the rapid development of the social economy. Water demand forecasting is a significant issue in the designing, maintaining and operating of a reliable and economical water supply system. An explicit mathematical method was presented in this study, based on the indicators of industrial output value, such as the gross output value of agriculture, forestry, animal husbandry and fishery. The impact of total retail sales and year trends in the domestic or industrial water demands, can accurately forecast the water demand fluctuations for a municipality. Adopt RAGA-PP optimal selection model through a grouping method of data handling for water demand management to test for the case study Shaanxi, China. Results showed that the prediction effect of multivariate logarithmic model accuracy can reach 99.50%, and it is estimated that the demand for water resources in Shaanxi would exceed 10 billion m³ by 2030. The average relative error of the water consumption from 2015 to 2017 is 3.05% for the model of multiple linear and 0.50% for the model of multivariate logarithm model. Our framework can assist in developing sustainable solutions.

Key words | accelerated genetic algorithm, projection pursuit model, regression analysis, water demand forecasting

HIGHLIGHTS
● Using the RAGA-PP model to realize the optimal selection of water resources demand indicators can improve model accuracy compared with model solving.
● Adopt a curve fitting method and regression analysis method to establish the water demand model and solve it.
● Due to the huge consumption of agricultural water, this paper puts forward corresponding agricultural water-saving measures.
INTRODUCTION

There is a great demand for freshwater because of the rapid development of the social economy and its growing urban regions in Shaanxi Province. According to statistics, the total water reserve of Shaanxi from 2009 to 2018 was lower than that of China, only 40.498 billion m³, and the average water resources per capita was only half the national average level. The potable water supply for Shaanxi depends on its rivers and groundwater sources. In recent years, due to the over-exploitation of groundwater in Shaanxi, its level in some areas has dropped sharply, and some areas are even threatened with groundwater depletion. The water resource shortage in Shaanxi is becoming serious. Reasonable control of the water demand among the residents, businesses and other customers in the community through forecasting is a key measurement in creating a reliable and economic system.

Researchers have successfully applied time series modeling based on stochastic modeling (linear) and/or artificial intelligence methods (non-linear) for water demand forecasting. Zubaidi et al. (2020) adopted a new methodology combining discrete wavelet transform (DWT) with an adaptive neuro-fuzzy inference system (ANFIS) which was proposed to predict monthly urban water demand based on several intervals of historical water consumption. Lu et al. (2020) adopted a hybrid model to improve the accuracy of predicting water demand. The model combined a simulated annealing (SA) algorithm, genetic cross factor (GCF), fruit fly optimization algorithm (FFOA), and support vector machine (SVM) to predict monthly water demand. Yuan et al. (2019) employed a grey power model [GPM (1, 1) model] which was a good prediction method for predicting urban water consumption. The traditional GPM (1, 1) model generated its grey sequence by a first-order accumulating generation operator (1-AGO) and produced the predicted results by a first-order inverse accumulating generation operator (1-IAGO). Smolak et al. (2020) compared classical and adapted machine learning algorithms used for water usage predictions including autoregressive integrated moving average (ARIMA), support vector regression, random forests and extremely randomized trees. Chen et al. (2017) adopted a multiple random forests model, integrated wavelet transform and random forests regression (W-RFR) which were proposed for the prediction of daily urban water consumption in southwest of China. Jain & Ormsbee (2002) adopted conventional regression and time series analysis methods which all were artificial neural network models to model and forecasted short-term water demand. They considered the rainfall, the maximum air temperature and the past water demands as effective parameters for water demand forecast model. Herrera et al. (2010) employed different learning machine methods to forecast hourly urban water demand. Wu & Yan (2010) utilized two genetic programming (GP) approaches to find a model for water demand forecast for a real water system, which showed that the average temperature had a significant impact on water demand. Gharabaghi et al. (2019) introduced a linear transfer function forecasting model based on the ARIMA to forecast monthly water consumption and considered several climatic, economic and flow rates as inputs for the model development in El Paso city in Texas, USA. As a result, changes in climatic and economic parameters were compared to changes in water rates parameters, which had a greater impact on consumption per customer. Lopez et al. (2018) applied a multi-model predictor
architecture for water demand forecasting in water distribution networks. Yin et al. (2018) developed a feed-forward artificial neural network model to forecast urban water resources and energy demand. Ghalekhondabi et al. (2017) reviewed soft computing techniques for water demand forecast. More recently, Bonakdari et al. (2018) introduced a new insight in time series modeling by detailing three linear/nonlinear/hybrid methods. Zeynoddin et al. (2018) proposed a new methodology for modeling of a time series. This methodology was based on a combination of stationarization techniques, and normality transforms into three different scenarios. Zeynoddin et al. (2018) and Ebtihaj et al. (2019) highlighted the impact of appropriate pre-processing techniques for elimination of deterministic term in time series, leading to a reliable linear approach with higher accuracy compared to non-linear approaches. Sanchez et al. (2020) used a Geographically Weighted Regression model informed by socio-economic, environmental and landscape pattern metrics to evaluate how future scenarios of population densities and climate warming would jointly affect water demand across two rapidly growing US States (North Carolina and South Carolina). Results showed that future water demand would increase under rising temperatures, but could be ameliorated by policies that promoted higher density development and urban infill. These water-efficient land-adopted policies led to a 5% regional reduction in water demand and up to 25% reduction locally for counties with the highest expected population growth by 2065. For rural counties experiencing depopulation, we considered that the land policies carried out were insufficient to significantly reduce water demand.

Balha et al. (2020) analyzed the impact of urban land-use change on Delhi’s water resources during time period 2005–2016 and for 2031. Using Monte-Carlo simulations in Land Change Modeler (LCM), they predicted the future LULC for year 2031 which exhibited a similar pattern of LULC change as observed during 2005–2016. An increase of 36.49% in urban area was observed during time period 2005–2016 and an increase of 14.05% was predicted for 2016–2031. Adopting in-situ groundwater measurements, the groundwater abstract (−18.75%) was found to be greater than the amount of total groundwater recharge (14.67%).

In this paper, we aimed to propose a model with an explicit equation with high accuracy for forecasting the water demand. The projection pursuit (PP) model optimized by the accelerated genetic algorithm (RAGA) was used to study the contribution degree of each evaluation index of water consumption.

DATA DESCRIPTION

Shaanxi Province is located in the northwest of China. It spans the middle of the Yellow River and the Yangtze River, connecting the eastern and central regions of China with important hubs in the northwest and southwest. The values of total water resources in Shaanxi from 2009 to 2018 are shown on Figure 1. The total amount of water resources in Shaanxi increased from 2009 to 2011 and has increased since 2017, but it is slightly lower than that of a decade ago. According to the Statistical Yearbook of Shaanxi Province, the total amount of water resources in Shaanxi Province was closely related to the amount of precipitation. In 2011, the amount of precipitation was the greatest, which made the total amount of water resources in that year the largest.

The water demand of Shaanxi mainly includes four parts: industrial water demand, agricultural water demand, domestic water demand, and ecological water demand. According to the average water consumption percentage of Shaanxi Province from 2004 to 2014, the highest agricultural water consumption was 67.95%, while the proportion of industrial water consumption, domestic water consumption and the ecological water consumption were 15.39%,

Figure 1 | Annual total amount of water resources level in Shaanxi during 2009–2018.
14.57%, and 2.08% respectively. The results of annual average water consumption ratio of four parts in Shaanxi during 2004–2014 are shown on Figure 2. On the premise of the model accuracy less affected, the ecological water consumption with the least share could be ignored in order to simplify the influencing factors of water resources demand in Shaanxi Province. It was reasonable to ignore ecological water consumption in terms of the accuracy of prediction results. The following analysis mainly considered industrial water demand, agricultural water demand and domestic water demand.

**METHODS**

**Projection pursuit model based on accelerated genetic algorithm**

Projection pursuit was proposed by Friedman and Tukey (Liu et al. 2019). It is mainly used to project high-dimensional data into low-dimensional subspace. The best projection direction and projection value can be obtained after conversion.

In this paper, the contribution rate of each evaluation index and the change trend of water resources demand in Shaanxi Province can be judged respectively.

A PP model optimized by accelerated genetic algorithm was used to study the contribution of water consumption evaluation indexes in Shaanxi Province, and to determine the evaluation indexes of water resource demand model.

**Normalization of evaluation indicators**

Let the sample set to be evaluated as \( \{x_{ij} \mid i = 1, 2, \ldots, n; j = 1, 2, \ldots, m\} \), where \( x_{ij} \) is the \( j \)th index of the \( i \)th sample. In order to unify the variation range of each index value and eliminate the influence of each index dimension, it is necessary to normalize each index.

For positive indicators:

\[
x'_{ij} = \frac{x_{ij} - x_{\text{min}(j)}}{x_{\text{max}(j)} - x_{\text{min}(j)}}
\]

For antidromic indicators:

\[
x'_{ij} = \frac{x_{\text{max}(j)} - x_{ij}}{x_{\text{max}(j)} - x_{\text{min}(j)}}
\]

where, \( x'_{ij} \), \( x_{\text{max}(j)} \) and \( x_{\text{min}(j)} \) are the normalized values of each index, the maximum and minimum values of the \( j \)th index, respectively.

**Constructive projection function**

The PP method is used to process \( m \) dimension data \( \{x'_{ij} \mid j = 1, 2, \ldots, m\} \) to \( a = \{a_{(1)}, a_{(2)}, \ldots, a_{(m)}\} \) direction projection to obtain a one-dimensional projection of \( m \)-dimensional data \( z_{(i)} \), equation as follows:

\[
z_{(i)} = \sum_{j=1}^{p} a_{(j)} \times x'_{ij}
\]

The projection function can be constructed as follows:

\[
Q(a) = S_z \times D_z
\]

where, \( S_z \) and \( D_z \) are the standard deviation and local density of \( z_{(i)} \), respectively.

**Based on RAGA optimized projection function**

When the sample index is given, the projection function \( Q(a) \) is only related to the projection direction \( a = \{a_{(1)}, a_{(2)}, \ldots, a_{(m)}\} \), so the optimal projection direction can be solved by solving
the maximum projection function $a$, equation as follows:

$$\max Q(a) = S_z \times D_z$$ (5)

subject to $\sum_{j=1}^{p} a^2_{ij} = 1$ (6)

Step 8: Accelerating the cycle: use the excellent individuals generated by each iteration to adjust the variable interval to make it closer to the optimal point.

### Selection of assessment indicators

Many indicators can reflect the demand for water resources in Shaanxi Province, such as the quota of 10,000 Yuan (100 Yuan = 15 US dollars) output value, industrial added value, water consumption for industrial increase, industrial output value, irrigation area, total output value of agriculture, forestry, animal husbandry and fishery, total sowing area, population quantity, per capita disposable income of urban residents, total retail sales of consumer goods, water consumption quota, etc. Regional water demand assessment indicators are shown in Table 1.

The PP model parameters of accelerated genetic algorithm were set as follows: population size $N$ was 400, crossover probability $Pc$ was 0.8, mutation probability $Pm$ was 0.2, random number required for variation direction $M$ was 10, acceleration times $Ci$ was 7, number of variables $n$ was 9, and limit number set for acceleration once after two evolution was $D2$. Optimal projection direction and Table 1

<table>
<thead>
<tr>
<th>Target layer</th>
<th>System layer</th>
<th>Indicator layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water resources demand in Shaanxi A</td>
<td>Industrial water demand B1</td>
<td>10,000 Yuan output value quota C1 ($\times 10^8$ Yuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial value added C2 ($\times 10^8$ Yuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial output value C3 ($\times 10^8$ Yuan)</td>
</tr>
<tr>
<td></td>
<td>Agricultural water demand B2</td>
<td>Irrigation area C4 (1,000 Ha)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gross output value of agriculture, forestry, animal husbandry and fishery C5 ($\times 10^8$ Yuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gross sowing area C6 (1,000 Ha)</td>
</tr>
<tr>
<td></td>
<td>Living water demand B3</td>
<td>Population C7 (10,000) C8 (Yuan per capita disposable income of urban residents)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gross retail sales of consumer goods C9 ($\times 10^8$ Yuan)</td>
</tr>
</tbody>
</table>
projection values of the nine indexes C1–C9 the obtained index layer are shown in Tables 2 and 3. The evaluation index of water resource demand selects the maximum of the best projection direction from each of the three system layers. That was to choose C3, C5 and C9 as the evaluation index of water resources demand.

Figure 3 shows the demand for water resources in Shaanxi. From the overall trend of change, the trend of projection value was basically in line with that of water resources demand. This can indicate that the selection of C1-C9 index was reasonable, and was appropriate to use the RAGA-PP model to select the optimal index of water resources demand prediction in Shaanxi.

### Establishment and solution of water resource demand forecasting model

#### Model establishment

The total output value of agriculture, forestry, animal husbandry and fishery ($X_1$), industrial output value ($X_2$), total retail sales of consumer goods ($X_3$) and water consumption ($Y$) of the year were fitted to the curves in Table 4, showing that the best fit between the three indexes and water consumption was logarithmic function. Many fit results showed that the Log3P1 and Bradley functions fitted best. All correlation coefficient Rs were above 0.9, and the fit results are shown in Figure 4.

Log3P1 function:

$$y = a - b \times ln(x + c) \tag{9}$$

### Table 2 | Optimal projection direction of index

<table>
<thead>
<tr>
<th>Index</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
</tr>
<tr>
<td>Optimal projection direction</td>
<td>0.33</td>
<td>0.36</td>
<td>0.37</td>
</tr>
</tbody>
</table>

### Table 3 | Projection values during 2004–2014

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Projection value</td>
<td>0.36</td>
<td>0.49</td>
<td>0.53</td>
<td>0.73</td>
<td>1.04</td>
<td>1.15</td>
<td>1.46</td>
<td>1.84</td>
<td>2.15</td>
<td>2.43</td>
<td>2.67</td>
</tr>
<tr>
<td>Water consumption ($\times 10^8$ m$^3$)</td>
<td>75.53</td>
<td>78.76</td>
<td>84.07</td>
<td>81.54</td>
<td>85.46</td>
<td>84.34</td>
<td>83.40</td>
<td>87.76</td>
<td>88.04</td>
<td>89.21</td>
<td>89.81</td>
</tr>
</tbody>
</table>
Bradley function:
\[ y = a \times \ln \left( \frac{C_0}{x} \right) \]  \hspace{1cm} (10)

According to the fitting of three indexes and water consumption, this thesis builds a multivariate logarithmic model to predict water resources demand in Shaanxi, and establishes a multivariate linear model to make a comparison.

Multiple linear model:
\[ Y_1 = a_1X_1 + a_2X_2 + a_3X_3 + A_0 \]  \hspace{1cm} (11)

Multivariate logarithmic model:
\[ Y_2 = \ln \left( \frac{X_1}{C_0} \right) + \ln \left( \frac{X_2}{C_0} \right) + \ln \left( \frac{X_3}{C_0} \right) + A_0 \]  \hspace{1cm} (12)

where, \( a_i b_i \) are the coefficient variable of the model; \( A_0 \) is a constant variable.

Model solution

Regression analysis of multivariate linear and multivariate logarithmic models using R software, and relevant output results were obtained by stepwise analysis, as shown in Tables 5 and 6.

The results of solving the multivariate linear and multivariate logarithmic models are shown in the equation (13) and equation (14), respectively.

Multiple linear model:
\[ Y_1 = -0.000448X_1 + 0.015123X_2 - 0.002712X_3 + 74.279 \]  \hspace{1cm} (13)

Table 5 | Comparison of two predictive regression models

<table>
<thead>
<tr>
<th>Models</th>
<th>( R^2 )</th>
<th>Errors in standard estimates</th>
<th>Errors in Durbin-Watson statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate linear models</td>
<td>0.891</td>
<td>2.411</td>
<td>1.593</td>
</tr>
<tr>
<td>Multivariate logarithmic models</td>
<td>0.951</td>
<td>1.650</td>
<td>2.559</td>
</tr>
</tbody>
</table>
Multivariate logarithmic model:

\[ Y_2 = \ln \left( \frac{X_1}{C_0^{611.8071}} \right)^{7.6515} \ln \left( \frac{X_2}{2996.1491} \right)^{3.1250} \ln \left( \frac{X_3}{1085.7016} \right)^{0.1215} + 38.5806 \]  

(14)

### VERIFICATION INDICES

The total output value of agriculture, forestry, animal husbandry and fishery \((X_1)\), industrial output value \((X_2)\) and total retail sales of consumer goods \((X_3)\) in Table 7 were substituted into multivariate linear and multivariate logarithmic models to obtain the \(Y_1\) and \(Y_2\) of water consumption in 2015 and 2017, respectively. Table 7 shows the actual water consumption and three index values during 2015–2017, while Table 8 shows the prediction results and relative errors of two models.

In Table 8, the average relative errors of water consumption during 2015–2017 were 3.05% and 0.50% respectively. The multivariate logarithmic model was used to predict the water resources demand in Shaanxi Province with high accuracy and better prediction results.

### RESULTS AND DISCUSSION

The results of water demand forecast in Shaanxi

In order to predict the water resources demand in Shaanxi Province in the future, it was necessary to establish a functional relationship between the year and each index, and the index relation with the year obtained on Figure 5. With the exponential function increasing rapidly, the precision of the coefficient goes higher. So it was necessary to logarithmic the index.
All the correlation coefficients of year and index curve $R^2$ were close to 1. The functions are listed as follows:

\[ \ln X_1 = -255.63628 + 0.13083x \]  \hspace{1cm} (15)

\[ \ln X_2 = -311.63124 + 0.15964x \]  \hspace{1cm} (16)

\[ \ln X_3 = -283.87489 + 0.14523x \]  \hspace{1cm} (17)

where, the $X_1$, $X_2$, $X_3$ stands the total output value of agriculture, forestry, animal husbandry and fishery, the industrial output value, and the total retail sales of consumer goods, the $X$ indicates the year, respectively.

By substituting equations (15–17) into equation (14), the functional relationship between year and water resources demand in Shaanxi can be obtained. The water resources demand of Shaanxi in 2020, 2025 and 2030 was shown in Table 9.
According to the data above, Shaanxi Province will need 9.578 billion m³ of water in 2020 and 9.979 billion m³ in 2025 which is 0.401 billion m³ more than that of 2020. Compared with 2025, the demand for water resources will be 10.381 billion m³ in 2030, which is 0.402 billion m³ higher.

Thereafter, this paper concludes that the demand for water resources in Shaanxi is increasing year by year. However, on Figure 1, the total amount of water resources has decreased in the past ten years which would cause many water resources shortage problems.

Various water-saving measures

Figure 2 shows that the agricultural water demand accounts for the largest proportion, followed by industrial water demand, domestic water demand, and ecological water demand. So, this paper proposes the following suggestions:

a. Adopt water-saving irrigation technology, such as sprinkler irrigation, drip irrigation, seepage irrigation. As one of the most advanced water utilization technologies, this kind of irrigation technology has no surface evaporation and deep leakage loss.

b. Use soil water storage and moisture conservation technology. Increase soil water storage capacity by fertilizer.


In order to, increase runoff rate and reduce surface storage and infiltration capacity, water surface seepage prevention treatment is an effective technical means.

d. Adopt solid water planting technology.

CONCLUSIONS

(1) The RAGA-PP model to optimize the selection of water resources demand indicators is reasonable.

(2) The average relative error of water resources demand in Shaanxi during 2015–2017 predicted by multivariate logarithmic model was only 0.50%, and the accuracy of prediction results was six times than that of the multivariate linear model.

(3) Demand for water resources in Shaanxi will continue to increase in the future, and it is expected to reach close to 10 billion m³ in 2025 and 12 billion m³ in 2030. With the decrease of total amount of water resources year by year, it would face a series of problems of water resources shortage, which means that various water-saving measures should be considered.

FUNDING

This research is supported by the fund of Industrial Support Plan Project of Gansu Province colleges and universities, No. 2020C-38.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES


First received 24 October 2020; accepted in revised form 19 January 2021. Available online 2 February 2021.