

Prediction of annual water consumption in Guangdong Province based on Bayesian neural network

Tao Tian ¹ and Huifeng Xue ¹

¹ China Aerospace Academy of Systems Science and Engineering, No 16, Fucheng road, Beijing 100048, China.

tiantaojam@163.com

Abstract. In the context of the implementation of the most stringent water resources management system, the role of water demand forecasting for regional water resources management is becoming increasingly significant. Based on the analysis of the influencing factors of water consumption in Guangdong Province, we made the forecast index system of annual water consumption, and constructed the forecast model of annual water consumption of BP neural network, then optimized the regularization BP neural network in utilization rate of water. The results showed that the average absolute percentage error of Bayesian neural network prediction model and BP neural network prediction model is 0.70% and 0.46% respectively. BP neural network model by Bayesian regularization is more ability to improve the accuracy of about 0.24%, more in line with the regional annual water demand forecast high precision requirements. Take the planning index value of Guangdong Province's thirteen five plan into Bayesian neural network forecasting model, and its forecast value is 45.432 billion cubic meters, which will reach 456.04 billion cubic meters of red water in Guangdong Province in 2020.

1. Introduction

Water is the source of life. Water resources are not only the basic natural resources of the countries in the world, but also the important economic resources, the balance between the supply of water resources and the actual demand for the sustainable development of the regional social, economic and ecological environment. China is a relatively poor water resources of the country. In 2011, the central document clearly put forward to implement the most stringent water management system, and in the following year to establish the total amount of water, water efficiency and water functional area to limit the three red lines. The total amount of water red line as the "three red line" in the most important indicators of water resources control, water administrative departments of the administrative region is responsible for the assessment of key indicators. On the basis of ensuring the sustainable development of regional economic society, the paper that uses the forecasting method with high precision to tap the regional water use standard, formulate a more scientific water basin water allocation scheme and annual water plan of each administrative region to realize the efficient and rational utilization of regional water resources and effective management has become a hot research issue.

According to the annual water consumption forecast, Liu Junliang and other factors analyzed the impact of urban water and quantified the interaction among the various factors, used system dynamics to establish a prediction model of urban water [1]. Guo Huifang et al. used the regression prediction method based on partial least squares to calculate the regression model of partial least squares by calculating the main influencing factors of water consumption by calculating the regression coefficient



between water consumption and influencing factors [2]. In this paper, the Bayesian regularized BP neural network method is used to predict the annual water consumption of Guangdong Province, and the results are used to help to enrich the methods of water resources management.

2. Present Situation of Water Resources in Guangdong Province

2.1 Status of water resources

In 2015, the average annual rainfall of 1875.7mm, compared with the previous year and perennial respectively 10.9% and 5.9%, was a partial water years. The spatial and temporal distribution of precipitation in the province was prominent. In the time distribution, mainly concentrated in the flood season (April to September), the precipitation was 1384.2mm (accounting for 73.8% of the total annual) which 1.5% less than normal; the pre-flood season (April to June) precipitation was 736.5mm (Accounting for 39.3% of the year) which 4.5% less than normal, after the flood season (July to September) the precipitation is 647.7mm (34.5% of the year) which 2.3% more than normal.

2.2 Situation of Water Resources Development and Utilization

The total water supplied was 44.31 billion m³ (excluding Hong Kong, Macao water supply 860 million m³), increased 0.6 million m³ over the previous year. Among them, the surface water supply 42.6 billion m³, accounting for 96.2% of the total water supply; groundwater supply was 1.53 billion m³; other water supply was 170 million m³, accounting for 0.4%. The direct use of seawater was 31.73 billion m³.

The total water consumption was 44.31 billion m³. Among them: agricultural water consumption was 22.7 billion m³ (up 270 million m³), accounting for 51.2% of total water consumption; industrial water was 11.25 billion m³ (450 million m³ less than last year), accounting for 25.4%; living water was 9.83 billion m³ increased 220 million m³ over the previous year), accounting for 22.2%; ecological environment replenishment was 530 million m³ (0.1 billion m³ increase over the previous year), accounting for 1.2%. According to the production, living and ecological division which the production of water 36.83 billion m³, accounting for 83.1% of total water consumption; life water was 6.95 billion m³, accounting for 15.7%; ecological replenishment was 530 million m³. The province's sewage treatment and reuse, rainwater utilization, desalination and other unconventional water use capacity was 175 million m³.

2.3 Analysis of water use

Per capita comprehensive water consumption was 411 m³, million yuan GDP water was 61 m³, million industrial added value water consumption was 37 m³, farmland irrigation per mu water consumption was 753 m³, urban residents per capita living water consumption was 193L/D, rural residents per capita living water consumption was 136L/D. With the previous year, per capita, million GDP, million industrial added value of water consumption have declined.

3. Construction of Regional Water Consumption Forecasting Model Based on Bayesian Neural Network

Since the number of hidden layers of neural network is sufficient to fit any nonlinear problem, the number of hidden layers is chosen as 1, and a neural network prediction model of single hidden layer is constructed. For the number of neurons in the hidden layer, we need make it according to the specific data at the time of the study. Therefore, the prediction model based on Bayesian neural network is mainly to determine the input and output neurons of the network. For the construction of regional water demand forecasting model, according to the input of neural network, can be divided into explanatory prediction [3]. In the explanatory predictive model, it is necessary to analyze the factors influencing the water consumption, and then determine the forecasting index system according to the principle of index selection.

3.1 Analysis of influencing factors

From the perspective of the influencing factors of regional water use, water consumption is not only affected by the distribution of local water resources, but also affected by factors such as local economic development and population [4]. In addition, the water resources related legal system and access to water-related facilities and technology breakthroughs and input level will bring a significant impact on water consumption. If the amount of water resources available in the region is less, it will restrict the development of the local economy, and the economic growth will be accompanied by the increase of the total amount of water. The increase of the population will lead to the increase of the total water consumption in the area water. When the water treatment and water management equipment and technology means more advanced, the degree of waste of water will reduce. When utilization rate of water is high, the total amount of water will decrease. The deepening of the contradiction between supply and demand of water resources will lead the government to make reasonable use of the corresponding water security system, such as the water abstraction permit system, the water right transaction system, the most stringent water resources management system, the water price system and so on, and the introduction of these systems will reduce the total amount of water, while promoting the development of water resources-related technology, and thus promote economic development, impact the region's total water use by the interaction between factors [5].

From the perspective of water type, water consumption can be divided into industrial water, agricultural water, living water and ecological environment water. The main influencing factors of industrial water are industrial output value. Industrial output value and the increase in industrial added value will lead to increased water consumption to varying degrees; if industrial water recycling rate is high, the industrial needs of the fresh water can be reduced; the level of industrial water prices directly affect the enterprise water consumption. If the size of the industrial water prices are high, the enterprise generally use to reduce the waste of unnecessary water, while industrial water prices and the above two economic indicators of the trend between the differences will lead to changes in water consumption, if the enterprise profits lower water prices rise. The impact of other systems and technologies on industrial water is mainly reflected in the reuse rate of industrial water [6]. The progress of science and technology will increase the reuse rate of industrial water. The mandatory requirement of water resources or environment-related system will encourage enterprises to increase the treatment of sewage wastewater efforts to improve the reuse rate of industrial water.

Living water not only includes residents living water, but also the tertiary industry and construction industry and other public service water [7]. The main influencing factors of living water consumption are the number of regional population, the price of living water, the per capita income of urban area, the amount of regional water resources, habit of residents' water consumption, the total output value of tertiary industry and the construction area. For different regions, the residents of different water habits. The north due to dry weather and the average temperature is lower than the south, coupled with the lack of water resources, resulting in significant differences in water usage habits between residents, and then per capita water consumption is different. The increase in water use will lead to an increase in the amount of water used in the past. There is even some water forecasting methods or water planning to estimate the annual water consumption of the area directly from the population. Similar to the amount of industrial water consumption, the difference between water price and the per capita income of the residents will have an impact on the residents' water use habits to varying degrees [8]. Tertiary industry water, the main influencing factors are population, the level of consumption and the tertiary industry output value. As the construction process and materials are similar, the unit building area of water consumption fluctuations, therefore, the main factor affecting the construction of water is the construction area.

Agricultural water included agriculture, forestry, animal husbandry and fishing water. The main influencing factors of agricultural water use are the added value of agriculture, forestry, animal husbandry and fishery, the amount of precipitation, the effective irrigation area and the effective utilization coefficient of farmland irrigation. Precipitation will be some extent to the agricultural water for some added. If there is more precipitation, the less demand for fresh water in agriculture. The irrigated area of farmland is the most important factor of agricultural water. In the case of crop

determination, the water requirement at each stage of the crop life cycle remains constant. However, the main influencing factor of the crop's water consumption is the effective irrigation area. In addition, the effective use of farmland irrigation coefficient also directly affects the agricultural water consumption. It also indirectly reflects the water-saving technology on the impact of water consumption. So it's an important agricultural water assessment indicator.

3.2 Index selection principle

For the explanatory predictions of water consumption, the selection of the influencing factors in the forecasting index system must have the following properties:

- Scientific. The predictive index system can reflect the inherent law of water consumption as objectively as possible.
- Completeness. The forecasting system should take into account the various factors that affect water consumption.
- Operability. The choice of each water supply needs to be simple and easy to quantify.
- Independence. The information expression between the influencing factors does not overlap as much as possible.

In addition to the above principles, the selected indicators need to have a higher accuracy of the forecasting method or a mandatory man-made.

3.3 Forecast target selection

In this paper, based on the principle of selection of indicators, combined with the previous research basis, in the impact of various types of water analysis, and then integrated selection of indicators, delete redundant indicators, finally selected as shown in Table 1 as the annual water intake Forecast indicators:

Table 1. The total amount of water prediction index system of year

Total water consumption	Water supply	Total water resources
		Precipitation
	Water use	GDP
		Industrial added value
		Industrial water reuse rate
		Permanent residents
		The added value of the tertiary industry
		Construction area
		Effective irrigated area
		Effective Utilization Coefficient of Farmland Irrigation
		Value added of agriculture, forestry, animal husbandry and fishery
		Urban green area

The above indicators are used as the 12 inputs of the network. Therefore, the regional annual water consumption explanatory model based on Bayesian neural network is shown in Fig. 1

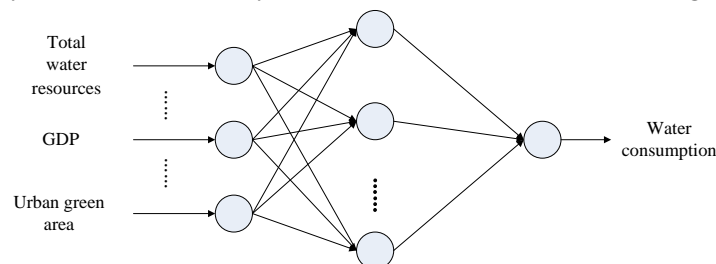


Figure 1. Bayesian Neural Networks in the area of water interpretative prediction mode

4. Explanatory Prediction of Annual Total Water Consumption in Guangdong Province

The input and output of the explanatory water total forecast have been fixed after the selection of the indicator. The prediction model outputs 1 result when input 12 neurons. The hidden layer is still selected as 1 layer. The number of neurons in the hidden layer can be obtained the best value by substituting the input data. The maximum number of training network failure is set 5, the target error is set 0.005 and the maximum number of training is set 3000 times. From 2001 to 2015 a total of 15 samples, the first 13 samples are as a training sample, the rest two samples are as the test sample.

4.1 Data normalization

Before it input into the network, the inputting variable data needs to be normalized so that the data falls between $[-1, 1]$. Data normalization formula:

$$Xp = 2 \frac{X - X_{\min}}{X_{\max} - X_{\min}} - 1. \quad (1)$$

4.2 Determination of the number of hidden neurons in the network

In the selection of the number of neurons on the hidden layer, we generally use the test method to select the more appropriate number. When the number of neurons is small, the network may not get the satisfactory fitting model because of the complexity. However, when the number of hidden neurons is large, it is easy to make the network more complicated, which reduces the budget rate of the network, but did not make the accuracy of the neural network model greatly improved. According to Kolmogorov theorem, when the input node of the network is k , the number of hidden layer nodes in the network can get the ideal simulation effect in the vicinity of $2k+1$. Therefore, this paper tests the number of neurons at $2k+1$ and its vicinity, and then tries to approach the model with high accuracy on the basis of experience. Finally, we try to find the value of the hidden neurons in the highest precision. The toolbox of MATLAB is called to create a BP network model. The prediction model based on Bayesian neural network is constructed for the three models of 12-19-1, 12-23-1 and 12-27-1, respectively. Finally, the predicted results show that the hidden layer of neurons for the 19 is the best results, when the network is 19, 23, 27.

And then narrow the range, select the network prediction results when the number of hidden layer neurons in the network was 17, 19, 21, and so on, the finally selected 20 as the number of neurons in the hidden layer of network, the neural network model is determined to be 12-20-1.

4.3 Model Accuracy Evaluation

In this paper, the average absolute percentage error MAPE (Mean Absolute Percentage Error) and RMSE (Root Mean Square Error) were used to evaluate the prediction effect of the two models.

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{p}_i - p_i}{p_i} \right| \quad (2)$$

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{p}_i - p_i)^2} \quad (3)$$

When the number of hidden neurons in the BP neural network prediction model is determined, we will use the total amount of water to predict the sample into the network for training, and use Bayesian regularization to optimize. Table 2 shows the total amount of water for 2001-2015 and Table 3 shows the values for each indicator from 2001 to 2015. All the Data are from the National Statistical Yearbook, Guangdong Province Water Resources Bulletin and Guangdong Province Environmental Statistical Yearbook. Fig.2 and Fig.3 show the fitting effect of BP neural network and Bayesian neural network on training samples. Table 4 and Table 5 show the two models of the forecast results and performance comparison. Table 6 and Table 7 show the first five neurons connection weight after training in network input layer (Input) and the hidden layer (La)

Table 2. Water consumption in the year 2001 to 2015

No	Year	Total living water (Billion m ³)	No	Year	Total living water (Billion m ³)
1	2001	429.8	9	2009	461.53
2	2002	446.4	10	2010	463.41
3	2003	447	11	2011	469.01
4	2004	457.64	12	2012	464.22
5	2005	464.8	13	2013	450.98
6	2006	458.95	14	2014	443.16
7	2007	459.4	15	2015	442.54
8	2008	462.51			

Table 3. Explanatory prediction of each index value in the years 2001 to 2015

	2001	2002	..	2012	2013	2014	2015
Total water resources (B m ³)	1816.32	2129.73		2026.55	2263.17	1718.45	1875.7
GDP(M)	10741.2	12039.2		53210.2	57067.9	62474.7	67809.8
	5	5		8	2	9	5
Industrial added value (B)	4463.06	4941.15		24649.6	25810.0	26894.5	29144.1
					7	4	5
Industrial water reuse rate	83.11	83.60		90.01	91.94	92.07	93.82
Permanent residents (M)	8650	8733		10505	10594	10644	10724
The added value of the tertiary industry (B)	4755.42	5544.35		24097.7	26519.6	30503.4	33223.2
				0	9	4	8
Construction area (M m ²)	15168.7	17516.3		37876.8	42431.7	52397.2	53443.2
	0	0		0	4	1	1
Effective irrigated area (Th)	1478.51	1447.1		1873.16	1874.44	1770.76	1770.9
Precipitation (mm)	1762.7	2314.5		1261.23	1920.81	1839.53	1654.98
Effective Utilization Coefficient of Farmland Irrigation	0.28	0.30		0.41	0.43	0.45	0.48
Value added of agriculture, forestry, animal husbandry and fishery (B)	982.5	939.6		2286.98	2665.2	2847.3	3047.5
Urban green area (M m ²)	19.89	21.03		41.06	40.17	41.20	42.19

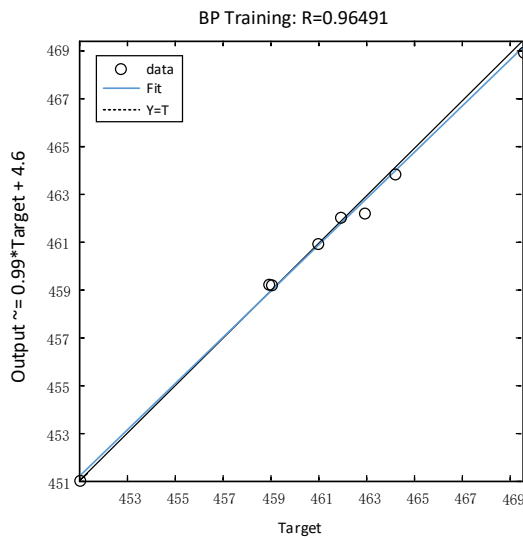


Figure 2. BP Neural Network fitting effect of training samples

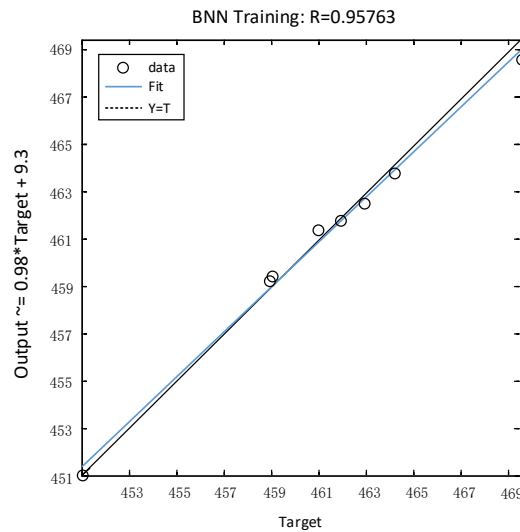


Figure 3. Bayesian Neural Network fitting effect of training samples

The predictive effects of the two predictive models are shown in Table 4 and Table 5:

Table 4. The prediction error comparing total water forecasting model based on Bayesian Neural Network and BP Neural Network

Year	Actual value	Predictive value (B m ³)		Relative error (%)	
		BP Neural Networks	Bayesian Neural Networks	BP Neural Networks	Bayesian Neural Networks
2014	443.16	447.98	445.7	-0.34	-0.53
2015	442.54	443.9	444.1	0.43	-0.37

Table 5. Performance parameters compared BP Neural Network Model and Bayesian Neural Network Prediction Model

Forecasting model	MAPE (%)	RMSE(B m ³)
BP Neural Networks	0.70	3.54
Bayesian Neural Networks	0.46	2.11

The results show that the mean square error of the BP neural network prediction model is 354 million m³, while the Bayesian neural network prediction model is 211 million m³, and the average absolute percentage error of the BP neural network prediction model is 0.70%. The neural network prediction model is 0.46%, and the accuracy of Bayesian regularization increases by 0.24%.

Table 6. Train connection between the value of BP Neural Network input layer and the hidden layer

	Input1	Input2	Input3	Input4	Input5	Input6	Input7	Input8	Input9	Input10	Input11	Input12
La1	-0.264	0.773	-0.819	-0.687	-0.110	0.132	0.970	0.445	-0.997	0.667	0.015	0.358
La2	-0.133	-0.572	-0.054	0.377	0.259	0.120	-0.064	-0.324	0.723	-0.277	-0.033	0.213
La3	-0.169	0.306	-0.648	0.250	-0.091	0.124	0.110	-0.217	-0.347	0.117	0.392	0.169
La4	-0.285	-0.535	0.597	-0.248	-0.052	-0.316	-0.063	0.327	0.212	0.413	-0.262	-0.327
La5	-0.176	-0.342	-0.053	-0.398	-0.564	-0.439	0.297	1.070	0.423	0.331	0.113	0.145

Table 7. Train connection between the value of Bayesian Neural Network input layer and the hidden layer

	Input1	Input2	Input3	Input4	Input5	Input6	Input7	Input8	Input9	Input10	Input11	Input12
La1	0.653	-0.366	0.001	-0.564	-0.001	0.171	0.376	-0.001	-0.185	-0.322	0.001	0.217
La2	0.000	-0.259	0.207	-0.787	0.177	0.225	0.504	0.950	0.070	0.088	0.001	-0.048
La3	0.135	-0.039	-0.001	-0.001	-0.150	-0.132	0.001	0.037	0.001	0.068	-0.094	-0.124
La4	0.399	-0.091	0.005	0.381	0.464	-0.370	-0.541	0.056	0.435	0.131	-0.490	-0.392
La5	-0.661	-0.000	-0.587	0.196	-0.252	0.070	0.638	0.373	0.267	-0.001	-0.497	-0.486

From contrast result of Table 6 and Table 7 can obtain that BP neural network connection weight is basically distributed between (0.1, 1). There are more connection weights less than 0.1 in Bayesian neural network, and there are 13 connection values less than 0.001, it can be approximated to 0. This indicated that Bayesian regularization can improve the complexity of the network by adjusting the connection value of the network, and enhance the generalization ability of the network and improve the prediction accuracy of the network.

December 2015 formally promulgated the "Guangdong Province thirteenth five-year planning proposal" clearly put forward that the region's GDP of Guangdong Province will be about 11 trillion yuan in 2020, industrial added value will be about 4.5 trillion yuan, the third Industrial added value will be over 5 trillion yuan, agriculture, forestry, animal husbandry and animal husbandry added value will be about 500 billion yuan, the end of the resident population will be about 114 million people, farmland irrigation utilization coefficient will be 0.55, urban public green area will be 500,000 square meters, construction area will be 65000 million m³, the total amount of water resources and precipitation and effective irrigation area with reference to the views of the Ministry of Water Resources experts, the final set for 20.87 billion cubic meters, 1854 mm, 1900 thousand hectares. The results of each index were calculated into the explanatory model of the total amount of water used in the training, and the output of the network was 45.432 billion m³. The total amount of water for Guangdong Province for the 45.604 billion m³ in 2020, and this paper forecasts the total water consumption in Guangdong Province is 45.432 billion m³ in 2020, by 2020 the total water consumption in Guangdong Province has been on the brink of red line.

5. Conclusions

Based on the Bayesian neural network and BP neural network algorithm, the absolute percentage error of the predicted value is 0.70% and 0.46% respectively, which indicates that the effective selection of the index is effective. The results show that the Bayesian regularization energy can eliminate the redundancy of the network structure and reduce the network complexity under the condition of satisfying the fitting precision of the training samples. After the two models trained, the fitting degree and the connection weight are compared. The accuracy of the Bayesian neural network model is 0.24% higher than that of the BP neural network model, which proves that the Bayesian regularization method can effectively improve the generalization ability of the network. The model can provide reference for the forecast of annual water consumption in Guangdong Province. That provides the accurate data for the total water controlling and support to promote the rational use of water resources in Guangdong Province in 2020.

Acknowledgements

This research was financially supported by the National Natural Science Fund of China (NSFC) through the project of Research on Multi-variable Data Fusion and Modeling Method for Water Resources in Smart City (No. U1501253) and application of special funds of Guangdong Province

through the project of Water Resources Data comprehensive application platform development and industrialization (No.2016B010127005)

References

- [1] LIU Junliang 2012 *Forecasting Model of Urban Water Consumption BP Network* (Harbin: Journal of Harbin University of Civil Engineering)
- [2] GUO Huifeng 2011 *Artificial Neural Network Method for Forecasting Water Consumption* (Tianjin: China water supply and drainage)
- [3] WANG Zhi-wen 2011 *Research on the application of bayes neural network and FTIR in quantitative analysis of multi-gases* (Taiyuan: North University of China)
- [4] FANG Cai-ting 2006 *Study of image recognition and tracking based on BP neural network* (Xi'an: Xidian University)
- [5] TIAN Zhi-wei 2011 *Application of Bayesian neural network on stock forecasting* (Wuxi: Jiangnan University)
- [6] MAIDMENT D, PAZREN E 1984 *Time patterns of water use in six Texas cities*. Journal of Water Resources Planning and Management ASCE 110(1): 90-106
- [7] CHEN Jia 2012 *Rolling bearing fault diagnosis based on rough set theory and artificial neural network* (Chengdu: Southwest Jiaotong University)
- [8] HE Zhong-hua 2009 *The research for forecasting urban water demand* (Hefei: Anhui University of Science and Technology)