

Application of Principal Component Analysis in Measuring Water Supply Capacity

Dejie Zhao^{1, a}

¹North China Electric Power University, Baoding, 071000, China

^a877040094@qq.com

Abstract: Measuring the water supply capacity is very complex with many impact factors. It is difficult to find the main factors. After constant comparing and calculating, I finally determine 12 factors have greater impact on sustainable development. Due to the influence factors is very complex and linked between them, I adopt the way of principal component analysis (PCA) to reduce the factors as much as possible. I define four indicators, which are Probability of Water Supply (IP), Reliability of Water Supply (IR), Water Utilization Rate (IW) and Efficiency of Water Utilization (IE). The grades (>, =, <) of them respectively represent the indicators are high, middle and low.

Keywords: Water Supply System Evaluation , PCA, GEM.

1 Basic Model and Thought

The paper use a special method to formulate the evaluation model. One is the Principal component analysis (PCA) ,which was invented in 1901 by Karl Pearson[1], mostly used as a tool in exploratory data analysis and for making predictive models.

In this paper, I use PCA to deal with the large amount of factors. After calculating the four indicators of an area, I set the standard rating of the four indicators .And our model is a reliable way in assessing a country's water supply capacity.

2 Symbols and Definitions

The paper use some symbols for constructing the model as follow.

Table 1. Symbols have been used

Symbols	Definitions
A	Total annual renewable freshwater supply
D	Total annual water demand.
U	Actual water on average.
W	Total annual freshwater withdraw
α	Effective utilization coefficient of agricultural water
β	Recycling rate of industrial water
P_A	The ratio of agricultural water in total water
P_I	The ratio of industrial water in total water

3 Principal Component Analysis

3.1 Select factors

Measuring the ability of a region to provide clean water is very complex with impact factors. It is difficult to find the main factors. So I search a great deal of literature and data, and preliminarily select some factors.

After constant comparison and calculation, I finally determine 12 factors with greater impact on the region's ability, they are daily per capita water consumption, water consumption per unit of GDP, farmland irrigation water per square kilometer, available water per capita, population growth, population density, average annual rainfall, pollution rate, sewage treatment daily per capita, the proportion of investment in water conservancy facilities and water consumption per unit of GDP.

In order to assess a region's ability to provide clean water, I also select six as samples. They are Spain South, Africa, UK, Singapore, Argentina and New Zealand.

According to the data, I obtain the table as follow :

Table 2. Twelve factors of six regions' data

	A	B	C	D	E	F	G	H	I	J	K	L
Spain	485.9	450.7	143.2	0.654	93	143	0.63	198	6434	439	6.21	35.8
South Africa	295.6	324.6	95.3	-0.41	40	153	1.67	168	15864	1456	8.03	7.6
UK	756.8	849.7	246.5	0.553	256	257	0.51	286	3828	592	1.34	74.2
Singapore	211	1954.3	42.2	1.933	7634	164	0.93	258	1848	396	1.02	21.5
Argentina	269.9	2356.5	30.2	0.997	15	85	2.36	149	1835	2293	8.62	10.9
new Zealand	682.6	676.4	243.2	0.863	16	193	1.46	259	3456	623	0.67	31.5

(A, B, C, ... L represent the twelve factors, they are given in Figure 1)

3.2 Use PCA to put the 12 factors into four indicators

The Principal Component Analysis (PCA) is utilized to analyze the twelve factors that are selected from various dynamic factors, such as rainfall and population growth[2].

I use the data to reduce dimensions through Principal Component Analysis. Then, I calculate the coefficient of each factor in the corresponding principal component.

Firstly, I do Min-Max standardization for the data and calculate correlation coefficient matrix. Based on the result, I use PCA, then I get the eigenvalues of correlation coefficient matrix and contribution rate of each principal component.

Secondly, I determine to choose four most influential principal components to do comprehensive evaluation.

Finally, I get the classification of total factors from principal component coefficient.

The result is shown in the Figure 1.

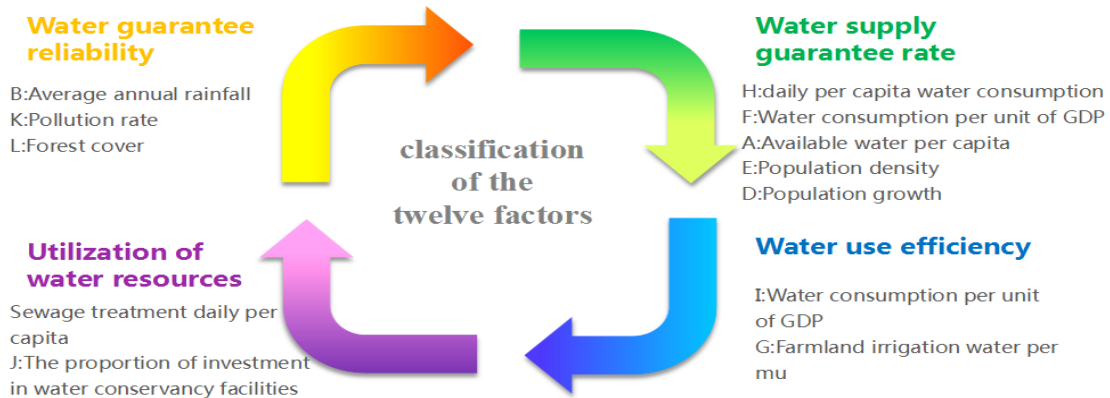


Figure 1. Classification of the twelve factors

By calculation, I obtain four principal components, they are water supply guarantee rate, water guarantee reliability, utilization of water resources and water use efficiency.

4 Evaluation and decision-making model

4.1 Index design for measuring regional capacity

According to PCA, I choose water supply guarantee rate, water guarantee reliability, utilization of water resources, water use efficiency as four indicators. Then I build assessment and decision-making model to measure the regional capacity to provide clean water. This method is easy to understand and evaluate, at the same time, it's also suitable for decision-making.

The four indicators can reflect the regional capacity to provide clean water and the ability to adapt to water scarcity. They can evaluate regional capacity to provide clean water and discover the characteristics of regional water shortages. Thus it's convenient for us to develop an intervention plan.

4.2 Index calculation[3]

- **Probability of water supply**

Water supply guarantee rate is to describe the protection level under the multi-year average state. The formula is as follows:

$$I_p = \frac{A}{D} \quad (1)$$

- **Reliability of water supply**

Water guarantee reliability describes whether the water demand can be met under certain guaranteed rate. I have the equation as follows:

$$I_r = \frac{A_r}{D_r} \quad (2)$$

- **Water utilization rate**

Utilization of water resources represents the potential for water development and utilization. The formula is as follows:

$$I_U = \frac{U}{W} \quad (3)$$

- **Efficiency of water utilization**

Water use efficiency represents the potential for water conservation. Water consumption can be divided into the consumption of agriculture, industry, life and the environment consumption. Among them, agriculture and industry water consumption usually use most of the water. Meanwhile, taking the availability of the data into account, I choose agricultural and industrial water levels to represent the overall level. Then I have the equation as follows:

$$I_E = \frac{\alpha \cdot P_A + \beta \cdot P_I}{P_A + P_I} \quad (4)$$

4.3 Index rating standard

After the measure of the known area water supply capacity evaluation, I set the standard rating of the four indicators. To calculate the four indicators of an area, compare them in the table below then I can measure the abilities of the region's water supplies.

Table3. Rating standards for the indicators

Grade	Is>/ Ir>	Is=/ Ir=	Is</ Ir<	Iu>	Iu=	Iu<	Ie>	Ie>	Ie<
Interval	>1.2	0.7~1.2	<0.7	>0.8	0.6~0.8	<0.6	>0.8	0.6-0.8	<0.6

5. References

- [1] Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine* 2 (11): 559–572. doi: 10.1080/14786440109462720.
- [2] Ohlsson L, Appelgren B. *Water and Social Resource Scarcity*. Rome Italy: FAO Issue Paper, 1998.
- [3] Jiuyi Li, Lijuan Li, Yumei Liu, *Regional Scale Water Shortage Risk Assessment and Decision-making System [J]*. *Progress In Geography*. 2010.