Analyzing Water Resources Utilization Efficiency in Various Areas of China Based on Density Clustering

Puhong Xue[1, a, *], Yugui Cao[1, b]

[1]School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou, 450046, China

[a]xuepuhong3@163.com, [b]cyg@ncwu.edu.cn

Abstract

Improving the utilization efficiency of water resources is one of the effective ways to alleviate the problems of water shortage, water waste and water pollution in China. In this paper, BC^2 data envelopment analysis (DEA) is used to comprehensively evaluate and analyze the water resources utilization efficiency of 31 areas in China from 2012 to 2020, taking agricultural water, production water, domestic water, ecological water, total population, fixed asset and COD discharge as input and gross regional product as output. The pure technical efficiency index and scale efficiency index of water resources utilization in each area are clustered by python. The 31 areas are divided into five types: comprehensive high-efficiency type, slightly low-technology type, low-technology type, slightly low-scale type and low-scale type. By analyzing the reasons for the differences in water resources utilization efficiency in various areas, suggestions are put forward for the development and utilization of water resources in five types of areas, so as to provide help and reference for improving water resources utilization efficiency in various areas.

Keywords

BC^2-DEA; Density peaks clustering; Water resources utilization efficiency.

1. INTRODUCTION

Water is a basic and strategic resource for the survival and development of human society. In 2020, the total amount of water resources in China will be 3.16052 trillion cubic meters. Due to China's large population, per capita water resources are only one fourth of the world average. From the distribution of provinces, the per capita water resources of eleven provinces are below the severe water shortage line, two provinces are in the severe water shortage range, seven provinces are in the moderate water shortage range, and two provinces are in the mild water shortage range. It can be seen that the shortage of water resources in China is relatively severe. Most economically developed provinces belong to water shortage areas, showing a trend of uneven regional distribution. Therefore, analyzing the reasons for the differences in water resources utilization efficiency among different areas in China and improving water resources utilization efficiency are effective ways to alleviate the problems of water resources shortage, water resources waste and water resources pollution in China, which can provide help and reference for improving water resources utilization efficiency in various areas.

According to the data collected at present, most of the methods used by domestic scholars to study the utilization efficiency of water resources are data envelopment analysis [1], and other methods include super efficiency SBM model [2], projection pursuit model [3], BP neural network [4]. On the whole, there is a lack of research on the specific reasons for the differences in water resources utilization efficiency in various areas. In 2015, Wu Qiong and Chang Hao [5]

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

combined data envelopment analysis and K-means clustering to conduct Q-type clustering on the pure technical efficiency index and scale efficiency index of water resources utilization in various areas, so as to explore the reasons for the differences in water resources utilization efficiency in various areas and their improvement direction. Because of pure technical efficiency and scale efficiency index data set more belong to the non-convex data sets, K-means clustering method is difficult to converge for non-convex data sets, however the clustering algorithm based on density not dependent on data distribution, which can effectively find any shape, any arbitrary density and size of the dataset class cluster center, and reasonable distribution of sample points to the corresponding class cluster. Therefore, by combining density peak clustering method and data envelopment analysis method, this paper discusses the reasons for the differences of water resource use efficiency in various areas, and puts forward improvement directions and suggestions.

2. RELATED WORKS

2.1. Data Envelopment Analysis

Data envelopment analysis is a nonparametric statistical method based on the concept of relative efficiency, which is used to evaluate whether the decision-making units with the same type of multi input and multi output are technically effective. The basic principle is to keep the input or output of DMU unchanged to determine the relatively effective production front surface with the help of linear programming and statistical data, and judge the relative effectiveness by comparing the degree of DMU deviation from the front surface. The most representative models in this method are C^2R model and BC^2 model.

The C^2R model was proposed by American operational research scientists A.Charnes and W.W.Cooper [1] in 1978. The model calculates the relative efficiency of each DMU under the assumption of constant return to scale. At the same time, the model can be used to measure the overall comprehensive efficiency, also known as technical efficiency (TE). However, when the inefficiency is, the C2R model cannot analyze whether the result is caused by pure technology or scale factors.

In 1984, R.D.Banker added a convexity assumption $\sum \lambda_j{}^* = 1$ to the C2R model, and obtained BC2 model, which decomposed technical efficiency (TE) into pure technical efficiency (PTE) index and scale efficiency (SE) index. PTE refers to whether DMU is engaged in production with the combination of inputs at the lowest cost in the short term, reflecting the production efficiency of the enterprise affected by management and technology. SE reflects the difference between the actual scale and the optimal production scale. Through PTE and SE indicators, we can better analyze whether the water use efficiency of each area is caused by pure technology or scale factors when the efficiency is ineffective.

Among them, the BC^2 model is:

$$
s.t.\begin{cases}
\min[\theta - \varepsilon\left(\sum_{r=1}^{t} s_r^+ + \sum_{i=1}^{m} s_i^-\right)] \\
\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rj0} \\
\sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- - \theta x_{ij0} = 0 \\
\lambda_j,\ s_i^-,\ s_r^+ \geq 0,\ (j = 1,2\cdots,n) \\
\sum_{j=1}^{n} \lambda_j{}^* = 1
\end{cases}
$$

λ_j is the weight of the j-th input resource or output item; n is the number of DMU; m is the number of input resource items of each DMU; t is the number of output items of DMU in each decision making unit; x_{ij} represents the input resources of item i of the j-th decision-making unit; y_{rj} represents the r-th output of the j-th decision-making unit; s_i^- and s_r^+ are input redundancy and output deficiency respectively; ε is Archimedean infinitesimal; θ represents the efficiency value of the j-th DMU.

(1) When $\theta = 1$, and $s_i^- = s_r^+ = 0$, the decision-making unit DEA is said to be valid, otherwise, the decision-making unit DEA is said to be invalid.

(2) That $s = \theta/\sigma$, s is scale efficiency. When $s = 1$, the scale is effective; When $s < 1$, the scale is invalid. When $\sigma = 1$, pure technology is effective; When $\sigma < 1$, then pure technology is invalid.

2.2. Density Peaks Clustering

The commonly used clustering methods include K-means clustering, hierarchical clustering, SOM neural network clustering and density peak clustering. Density peaks clustering can deal with non-convex data sets better. The algorithm has two important parameters: local density ρi and distance to higher local density points δi. They are defined as:

$$\rho_i = \sum_{j \neq i} \chi(d_{ij} - d_c)$$

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij})$$

d_{ij} is the distance between x_i and x_j; d_c is the cutoff distance; $\chi(x)$ is a logical judgment function, $x < 0$, $\chi(x) = 1$, otherwise $\chi(x) = 0$.

3. DATA PROCESSING AND ANALYSIS

3.1. Data Source and Index Selection

The data used in this paper are from China Statistical Yearbook, China water resources bulletin and provincial statistical yearbooks from 2012 to 2020, and take 7 items of agricultural water consumption, production water consumption, domestic water consumption, ecological water consumption, total population, fixed asset investment and COD (chemical oxygen demand) emission in 31 areas of the country as input indicators, take the gross regional product of each place as output index.

3.2. Efficiency Calculation

Import the data into Deap2.1 software for data envelopment analysis, and use BC^2 model. Limited by space, table 1 gives the average value of water resource utilization efficiency from 2012 to 2020, and table 2 gives the basic statistical characteristics of national water resource utilization efficiency.

(1) The national average TE from 2012 to 2020 is 0.74, and the overall utilization rate of water resources in China is not high, which still needs to be improved.

(2) Generally speaking, the average SE in most parts of the country is higher than 0.9 and the average PTE is 0.81. It can be seen that the performance of scale efficiency in all areas of China is better than that in pure technical efficiency.

(3) From the standard deviation, the standard deviations of PTE and SE are 0.19 and 0.11 respectively. It can be seen that the difference in pure technical efficiency among areas in China is greater than that in scale efficiency.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

(4) In terms of technical efficiency, the areas with DEA effective water resource utilization efficiency include Beijing, Tianjin, Shanghai and Guangdong. The average of TE in these areas reached 1, accounting for 12.9% of the total, indicating that the resources invested in these areas in the past nine years were effectively utilized and both input and output reached the best state. The remaining 27 areas are non DEA effective, and their input resources have not been effectively utilized. Among them, the pure technical efficiency of Jiangsu, Zhejiang, Shandong, Tibet and Qinghai is 1, which reaches the pure technical efficiency. However, due to the invalid scale, the technology is invalid.

Table 1. Average value of water use efficiency in various areas of China from 2012 to 2020

Area	Average of TE	Average of PTE	Average of SE
Beijing	1.00	1.00	1.00
Tianjin	1.00	1.00	1.00
Hebei	0.69	0.97	0.71
Shanxi	0.74	0.96	0.77
Inner Mongolia	0.82	0.98	0.84
Liaoning	0.86	0.92	0.93
Ji Lin	0.57	0.94	0.60
Heilongjiang	0.83	0.99	0.84
Shanghai	1.00	1.00	1.00
Jiangsu	0.98	0.98	1.00
Zhejiang	0.93	0.93	1.00
Anhui	0.50	0.99	0.51
Fujian	0.71	0.97	0.73
Jiangxi	0.51	0.98	0.52
Shandong	0.98	0.98	1.00
Henan	0.61	0.93	0.65
Hubei	0.86	0.98	0.88
Hunan	0.59	0.99	0.60
Guangdong	1.00	1.00	1.00
Guangxi	0.46	0.97	0.48
Hainan	0.92	0.94	0.98
Chongqing	0.89	0.96	0.93
Sichuan	0.70	0.98	0.72
Guizhou	0.72	0.96	0.75
Yunnan	0.62	0.96	0.65
Tibet	0.79	0.79	1.00
Shaanxi	0.82	0.97	0.84
Gansu	0.47	0.84	0.57
Qinghai	0.55	0.55	1.00
Ningxia	0.58	0.59	0.98
Xinjiang	0.44	0.91	0.49

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

Table 2. Statistical characteristics of national water use efficiency from 2012 to 2020

Efficiency type	Maximum	Minimum	Mean value	Standard deviation	Number of effective areas
TE	1	0.44	0.75	0.18	4
PTE	1	0.48	0.81	0.19	9
SE	1	0.55	0.93	0.11	4

3.3. Analysis of Density Peak Clustering Results

Using Python language programming, the density peak clustering analysis the average of TE and the average of SE of national water resources from 2012 to 2020 is carried out, and the clustering center is shown in Table 3.

Table 3. Final clustering center

Variable	Class 1	Class 2	Class 3	Class 4	Class 5
Average of PTE	1	0.85	0.62	0.98	0.99
Average of SE	1	0.98	0.95	0.93	0.57

It can be seen from Table 3 that the mean values and SE values of all types of PTE are significantly different. By referring to the classification model in literature [5], the areas in class 1 are named as comprehensive high-efficiency areas, the areas in class 2 and class 3 are named as slightly low-technology areas and low-technology areas respectively, and the areas in class 4 and class 5 are named as slightly low-scale areas and low-scale areas respectively. The attribution of each area is shown in Table 4.

Table 4. Classification of water resources utilization efficiency in various areas of China from 2012 to 2020

Type	Area	Number
Comprehensive high-efficiency type	Beijing, Tianjin, Shanghai, Guangdong	4
Slightly low-technology type	Inner Mongolia, Heilongjiang, Hubei, Shaanxi	4
Low-technology type	Hebei, Shanxi, Jilin, Anhui, Fujian, Jiangxi, Henan, Hunan, Guangxi, Sichuan, Guizhou, Yunnan, Gansu, Xinjiang	14
Slightly low-scale type	Liaoning, Jiangsu, Zhejiang, Shandong, Hainan, Chongqing, Tibet	7

(1) The scale and the pure technology of the comprehensive high-efficiency areas are effective, which shows that the resources invested by the four areas over the past nine years have been effectively utilized, and the input and output have reached the best state.

(2) The scale and pure technology of slightly low-technology areas and low-technology areas are invalid. From 2012 to 2020, the scale efficiency index of these two types of areas performed well, but the pure technical efficiency index of slightly low-technology areas performed better than that of low-technology areas. Therefore, slightly low-technology areas performed better in water resource utilization efficiency than low-technology areas. Generally speaking, the PTE value of these two types of areas is lower than the SE value, indicating that the resource allocation of these two types of areas is unreasonable. By referring to the redundancy of each

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

input resource, the structure of input elements can be continuously optimized to achieve the optimal combination of resources.

(3) The scale of slightly low-scale areas and low-scale areas was invalid, and the pure technology of other areas was invalid except Jiangsu, Zhejiang, Shandong, Xizang and Qinghai. From 2012 to 2020, the pure technical efficiency index of the two types of areas performed well, but the scale efficiency index of the slightly low-scale areas performed better than the scale efficiency index of the low-scale areas, so the slightly low-scale areas performed better than the low-scale areas in the water resource utilization efficiency. On the whole, SE value of these two types of areas is lower than PTE value, indicating that the actual scale of water resources utilization in these two types of areas does not match the input resources, so it is necessary to expand the investment scale, adjust and improve the scale structure of water resources input, so that the input resources can be effectively utilized.

3.4. Analysis on the Change of Water Resources Utilization Efficiency in Various Areas

From 2012 to 2020, the water resource utilization efficiency of comprehensive high-efficiency areas remains at a high level, and the PTE value and SE value do not change on the whole. Therefore, comprehensive high-efficiency areas should continue to maintain the current water resource utilization level, strengthen exchanges with other areas, and teach more experience to areas with low water resource utilization efficiency.

The scale efficiency index of slightly low-technology areas and low-technology areas performs well. Therefore, this paper analyzes the reasons for the differences in water resource utilization efficiency in these areas by observing the changes of PTE values in these areas from 2012 to 2020. Figure 1 shows the change trend of PTE value in slightly low-technology areas from 2012 to 2020. Figure 2 shows the change trend of PTE value in low-technology areas from 2012 to 2020.

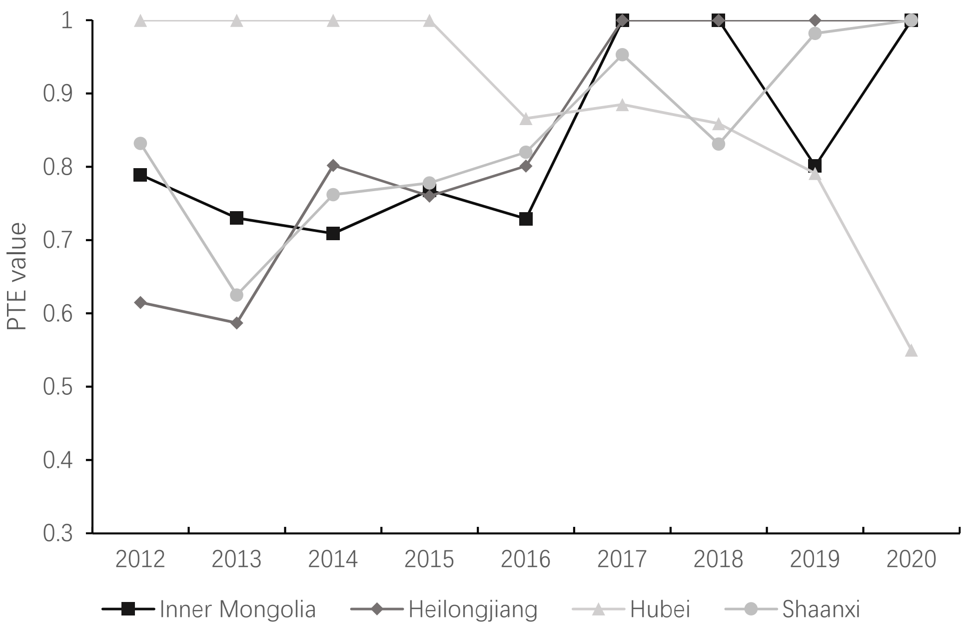

Figure 1. Change trend of PTE in slightly low-technology areas from 2012 to 2020

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

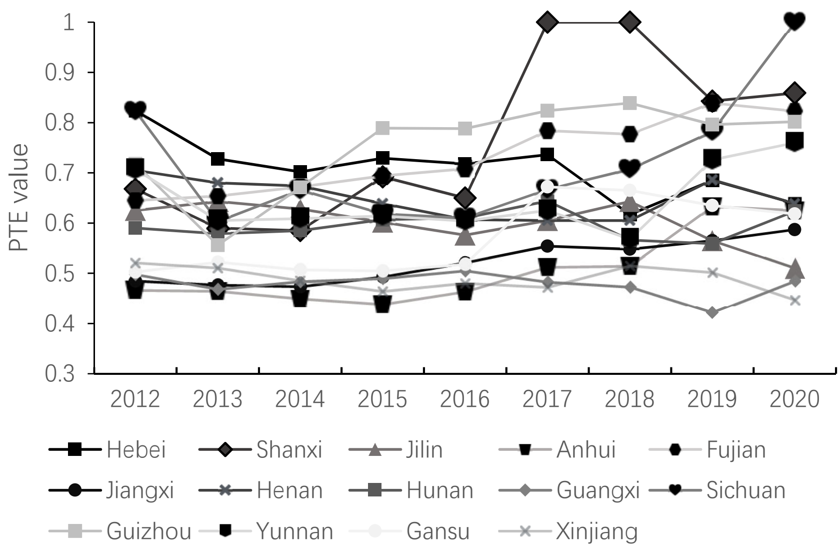

Figure 2. Change trend of PTE in low-technology areas from 2012 to 2020

As can be seen from Figure 1, the PTE value of Inner Mongolia, Heilongjiang and Shanxi shows an overall upward trend and has risen to 1 by 2020. The PTE value of Hubei Province generally shows a downward trend, which is still at a good level from 2012 to 2015, but it began to decline after 2015 and decreased by 0.45 in 2020. To sum up, the gap between the PTE value of the slightly low-technology areas and the low-technology areas is slightly smaller. By 2020, the water resource utilization efficiency of Inner Mongolia, Heilongjiang and Shanxi will reach DEA effective. These three areas should continue to maintain the current water resource utilization level. Hubei should find out the reasons for the decline of PTE value from 2015, and pay more attention to improving the level of resource allocation and technical management. By referring to the redundancy of various input resources, Hubei should continuously optimize the structure of input elements to achieve the optimal combination of resources.

As can be seen from Figure 2, PTE values in most areas have not changed significantly, and the overall fluctuation is within 0.2. The PTE values of Hebei, Jilin, Henan, Guangxi and Xinjiang showed a downward trend as a whole. The PTE value of Hebei decreased by 0.19 by 2020, that of Jilin decreased by 0.12 by 2020, and that of Henan, Guangxi and Xinjiang decreased by less than 0.1. The PTE value of Shanxi, Anhui, Fujian, Jiangxi, Sichuan and Gansu shows an upward trend on the whole, but there is still a big gap between the PTE value and the comprehensive high-efficiency areas by 2020. The PTE values of Hunan, Guizhou and Yunnan also showed an upward trend, but there was no significant increase. To sum up, there is a big gap between the PTE value of low-technology areas and that of comprehensive high-efficiency areas, which also shows that there is a lot of room for progress in the PTE value of low-technology areas. Low-technology areas should identify their own benchmark provinces and cities, learn their water use methods, water-saving skills and resource allocation, optimize investment in fixed assets, provide water-saving guidance to water users, and reduce ecological pollution in the process of water resource utilization.

The pure technical efficiency index of slightly low-scale areas and low-scale areas is well realized. By observing the changes of SE values in these areas from 2012 to 2020, the reasons for the differences in water resource utilization efficiency in these areas are analyzed. Figure 3 shows the change trend of SE value in slightly low-scale areas from 2012 to 2020. Figure 4 shows the change trend of SE in low-scale areas from 2012 to 2020.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

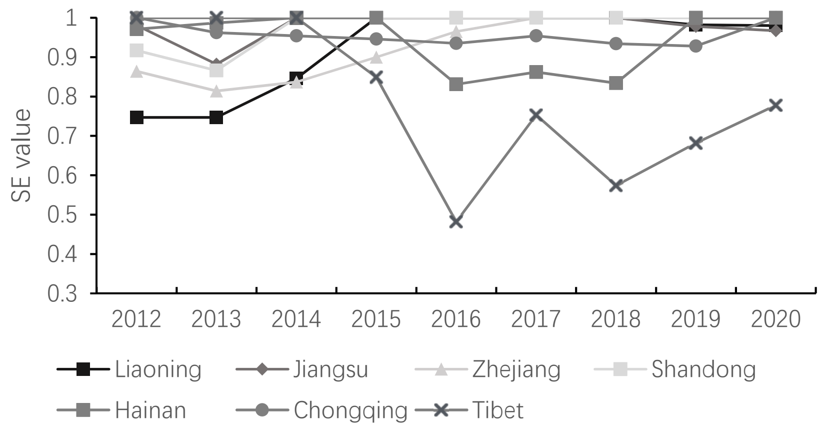

Figure 3. Change trend of PTE in low-technology areas from 2012 to 2020

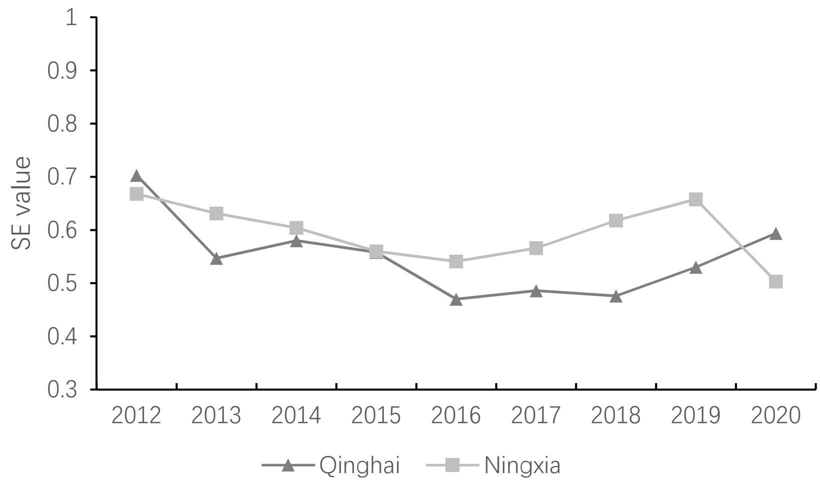

Figure 4. Change trend of SE in low-scale areas from 2012 to 2020

As can be seen from Figure 3, the SE value of Zhejiang, Shandong, Hainan and Chongqing shows an overall upward trend and has risen to 1 by 2020. The overall SE value of Liaoning also shows an upward trend and increases greatly. By 2015, the water resource utilization efficiency has reached DEA effective. However, in 2019 and 2020, the SE value decreased slightly. The SE value of Jiangsu and Tibet shows a downward trend on the whole, and the SE value of Jiangsu will drop to 0.97 by 2020. The SE value of Tibet will drop to 0.78 by 2020. To sum up, the gap between the SE value of slightly low-scale areas and that of comprehensive high-efficiency areas is slightly small. Zhejiang, Shandong, Hainan and Chongqing should continue to maintain the current level of water resources utilization. Liaoning, Jiangsu and Tibet should find the reasons for the fluctuation of SE value, timely adjust and improve the scale structure of water resources investment, expand the scale of investment, and make effective use of their invested resources.

As can be seen from Figure 4, the SE value of Qinghai and Ningxia shows a downward trend, and the SE value is relatively low. The SE value of Ningxia will drop to 0.5 by 2020. The SE value of Qinghai will drop to 0.594 by 2020. To sum up, there is a big gap between the SE value of low-scale areas and that of comprehensive high-efficiency areas, which also shows that there is a lot of room for progress in the SE value of low-scale areas. Qinghai and Ningxia should find their own benchmark provinces and cities. At the same time, they need to adjust the matching degree between the actual scale of water resources utilization and the invested resources, expand the scale of water resources investment, adjust and improve the scale structure of water resources investment, so that the invested resources can be effectively utilized.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202205_8(5).0069

4. CONCLUSIONS AND SUGGESTIONS

Through BC^2 data envelope analysis and density peak clustering, the water resources utilization efficiency of 31 areas in China from 2012 to 2020 is comprehensively evaluated and analyzed. Finally, the 31 areas are divided into five types: comprehensive high-efficiency type, slightly low-technology type, low-technology type, slightly low-scale type and low-scale type.

(1) For comprehensive high-efficiency areas, Beijing, Tianjin, Shanghai and Guangdong should continue to maintain the current level of water resources utilization, strengthen exchanges with other areas, and transfer some experience to areas with low water resources utilization efficiency.

(2) For slightly low-technology areas, they have a good performance in the actual scale of water resources utilization. The invalidity of DEA is mainly caused by the invalidity of pure technology. Therefore, the slightly low-technology areas need to be adjusted in water resources management and updated and upgraded in water resources utilization technology.

(3) For low-technology areas, DEA is invalid because pure technology is invalid, and the gap between pure technical efficiency index and pure technical efficiency index of comprehensive high-efficiency areas is too large. Therefore, low-technology areas need to optimize the treatment of resource allocation and strengthen the improvement of technology management. At the same time, it can conduct technical exchanges with comprehensive high-efficiency areas. We should also strive to find our own benchmark provinces and cities, learn the water use methods, water-saving skills and resource allocation in the benchmark areas, and choose a suitable local way to improve the utilization efficiency of water resources.

(4) For slightly low-scale areas and low-scale areas, the reason for the invalidity of DEA is that its scale is invalid and the scale does not match the input-output. Therefore, these two types of areas need to adjust and improve the scale structure of water resources investment, expand the scale of investment, make effective use of their invested resources and improve the utilization rate of water resources.

REFERENCES

[1] Charnes A, Cooper W W, Rhodes E. Measuring the efficiency of decision-making units. European Journal of Operational Research, 1979, 3(4) :339.

[2] WU Xiangdong, XU Xinfa, CHENG Jingqing, WEN Tianfu, LIU Zhangjun, WU Jiaqi. Spatio-temporal evolution and influencing factors of water resources utilization efficiency in Jiangxi Province, Yangtze River, 2021, 52(12) : 92-98+121.

[3] CAO Lei, ZHOU Weibo, ZHUANG Yan. Analysis of use efficiency of water resources based on model of genetic projection pursuit in Yanan, Journal of Water Resources and Water, 2015, 26(02) : 126-128+134.

[4] LIU Zeng-jin, ZHANG Min, WANG Zhen-yu, LI Xiao-yu. A Comprehensive Assessment of the Sustainable Utilization of Water Resources in Zhengzhou City Based on the ANN Evaluation Method, China Rural Water and Hydropower, 2008(12) : 55-58+62.

[5] WU Qiong, CHANG Haojuan, LIU Zhao. Analyzing water resources utilization efficiency in various regions of China based on clustering method, Yangtze River, 2018,49(14):55-60.