Assessment Model Based on Factor Analysis and TOPSIS Method

Lin Lu *

Jinan University – University of Birmingham Joint Institute at Jinan University Jinan University Guangdong, China

* Corresponding Author Email: lulin020104@163.com

Abstract. In this paper, a comprehensive evaluation index system was established based on the data set of Seyhanba, and the mutual influence was considered from two perspectives: its own and the surrounding influencing factors. Factor analysis method was used to merge the collected secondary indicators into primary indicators. Finally, on the basis of the first-level indicators, the TOPSIS entropy weighting method was used to calculate the comprehensive indicator scores. The results showed that the score was 0.2166 in 1991 and 0.5011 in 2010, which was about twofold increase. This data was studied and analyzed by factor analysis and TOPSIS entropy weighting, and the results obtained were good, with a view to providing some reference significance for other fields.

Keywords: Assessment model, Topsis, Factor analysis.

1. Introduction

Factor analysis is based on the idea of dimensionality reduction, in the case of losing as little or as much information as possible from the original data, the intricate numerous variables are aggregated into a few independent common factors, which can reflect the main information of the original numerous variables, while reducing the number of variables and at the same time reflecting the intrinsic connection between variables[1]. Usually, factor analysis has three roles: first, it is used for factor dimensionality reduction, second, it calculates factor weights, and third, it calculates weighted factor summary composite scores[2].

TOPSIS method is a solution evaluation and analysis method for the existence of multiple indicators and multiple solutions, that is, to determine the advantages and disadvantages of each solution in the data based on the existence of a data[3]. The central idea is to first determine the optimal ideal value (positive ideal value) and the worst ideal value (negative ideal solution) of each indicator, the so-called positive ideal value is a conceived best value (program), the value of its various attributes are up to the best value of each candidate program, while the negative ideal solution is another conceived worst value (program), and then find out the weighted Euclidean distance between each program and the positive ideal value and negative ideal value, thus the closeness of each solution to the optimal solution is obtained as the criterion for evaluating the superiority and inferiority of the solutions, and finally the superiority and inferiority values of each solution are obtained[4].

2. Data Collection

2.1. Data preprocessing

We requires quantitative analysis of the impacts of Saihanba[5]. Since the forest itself will not only affects the local environment, but also may has some effects on other cities, the data should include two aspects:

Indicators reflecting the changes of local ecological environment in Saihanba. Through searching a large number of references and visiting observatory websites, due to the availability and reliability of data, we finally select four first-level indicators, that is, air, microclimate, carbon sequestration and

water conservation for analysis. The benefit score of Saihanba is obtained by integrating the first-level indexes [6].

Indicators reflecting the impacts of Saihanba on the ecological environment of surrounding cities. After checking the map to learn about the location of Saihanba and reading different articles, it is found that Saihanba made a great contribution to the ecological environment of Beijing-Tianjin-Hebei region. However, there are obvious differences in the ways of data collection and data closure of the yearbook in each city or province. After ensuring the same first-level indicators in selected cities, the second-level indicators are selected according to local conditions and the amount of data should be collected as much as possible. Therefore, microclimate, air quality and water conservation are chosen as the first-level indicators. The ecological benefit score of Beijing, Tianjin and Hebei are obtained by integration respectively [7].

Since Saihanba National Forest Park was established in 1993, 1993 is regarded as a "turning point" [8], i.e., the time when the construction of Saihanba was completed. There would be significant differences in ecological environment data before and after that time. However, the data is difficult to obtain. Therefore, the data of certain three years before 1993 and after 1993 are collected for analysis to keep the errors as small as possible.

2.2. Missing values processing

After checking the acquired data, it is found that the data of Saihanba, Beijing and Hebei are enough to analyze, but the data of Tianjin only have partial values (data between 2001 and 2019) [9]. It lacks the water conservation data before 1993. Therefore, it is necessary to make data interpolation. Regression method, fixed value, nearest neighbor interpolation and others are commonly used to create data. Using the existing data and after analysis, it is shown that the trend of the water conservation data is growing and regression method is the best fit model for interpolation. Data preprocessing of Tianjin is shown as figure1.

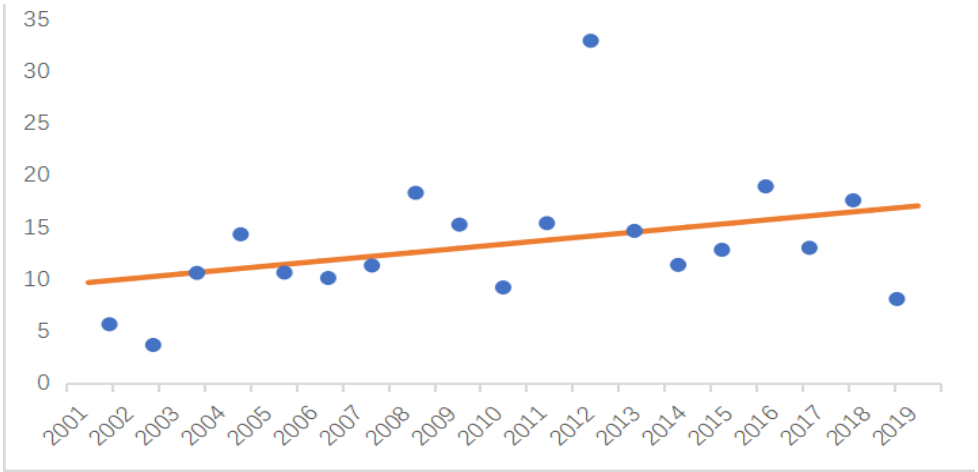


Figure 1 Data preprocessing of Tianjin

Therefore, the water conservation data of Tianjin are complemented. It is calculated by regression that the estimated values of water conservation from 1991 to 1993 are 5.586, 5.995 and 6.404 respectively[11].

2.3. Data normalization

Due to different features, second-level indicators chosen in this model can be divided into four types: Benefit type indicator, Cost type indicator, Fixed value type indicator and Interval type indicator.

Table 1: The type of second-level indicators

Indicator Type	Second-level indicators
Benefit type indicator	Days of rain(day)
	Water conservation
	The number of days with air quality at or above grade II
	Carbon sequestration of wood
	Soil carbon sequestration
Cost type indicator	Average wind speed
	Days of strong wind(day)
	Annual average concentration of PM10
	Annual average concentration of SO_2
	Annual average concentration of NO_2
Fixed value type indicator	Extreme maximum temperature
	Extreme minimum temperature
Interval type indicator	Annual total precipitation
	Annual average temperature

Since the dimensions are greatly different between indicators, by normalization we normalize the data to eliminate the effect, which means that the values are between 0 and 1 without unit and all data are changed into the same trend.

3. Ecological environment assessment model based on factor analysis and Entropy Weight TOPSIS method

This essay need to quantify the influence of Saihanba Forest farm on its surrounding ecological environment. By research a branch of literature and data bases, we finally determine to establish three first-level indicators without loss of rationality and comprehensiveness. They are microclimate, water source and air quality. In order to evaluate the effect of eco-environmental impact index accurately, we establish a number of secondary indexes under each first-level indicator. The reason of choosing above three indicators as well as the details of secondary indexes are as follows:

- Microclimate: Microclimate regulation of forest farm includes increasing precipitation, increasing frost-free period, reducing gale days, decreasing summer temperature and increasing atmospheric humidity. Therefore, there are several second-level indicators: Annual Average Temperature, Highest and Lowest Temperature, Average Relative Humidity, Annual Average Precipitation.[2]
- Water sources: Forest farm has the function of water conservation and water purification. Forest farm can not only change the distribution of precipitation, but also intercept and absorb precipitation by the forest canopy and soil layer and other levels, so as to store water. Such water storage function is related to the total amount and intensity of local precipitation respectively.Therefore, there is second-level indicators which is Local Water Conservation.[1]
- Air quality: Forest ecosystem has the function of dust removal and air purification, which can reduce and control the pollution, To be exact, forests can absorb harmful substances such as sulfur dioxide, fluoride, nitrogen oxide and dust retention etc. Thus when quantify the influence of forest on air quality, some secondary indexes need to be established. And in this essay,we consider it as Number of days with air quality at or above Grade II, Annual average concentration of PM10, Annual average concentration of SO2 and Annual average concentration of NO2.[2]

Ecosystem is a very complex system, which affects different areas in different ways.According to the characteristics of different regions, we select different secondary indicators and collect corresponding data to establish a complete ecosystem impact assessment system according to local conditions. According to the problems, the ecosystem impact assessment system for Saihanba Forest

farm, Tianjin and Beijing was established which can be easily seen in the following charts.The Ecological Efficiency System of Saihanba is shown as figure2.The Ecological Efficiency System of Beijing is shown as figure3.The Ecological Efficiency System of Hebei is shown as figure4.

The Ecological Efficiency System of Tianjin is shown as figure5.

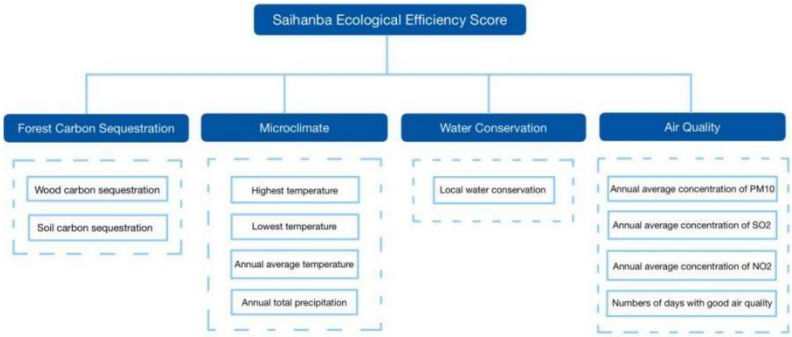


Figure 2 The Ecological Efficiency System of Saihanba

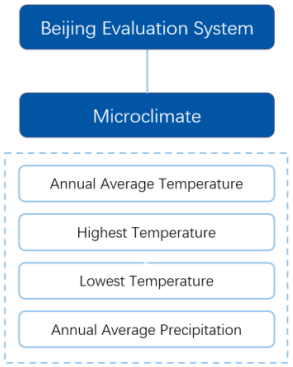


Figure 3 The Ecological Efficiency System of Beijing

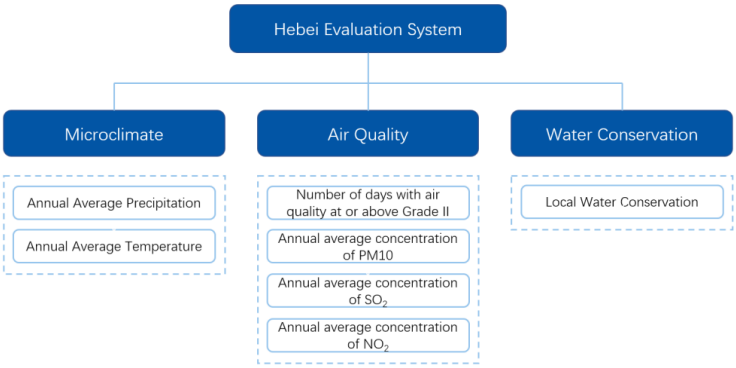


Figure 4 The Ecological Efficiency System of Hebei

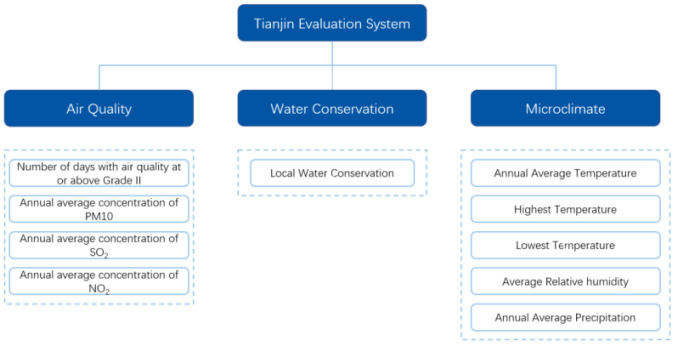


Figure 5 The Ecological Efficiency System of Tianjin

4. Establishment of ecological environment comprehensive evaluation index system

4.1. The combination of secondary indicators on the basis of factor analysis

(1) Standardized processing of raw data

In this paper, there are n secondary index variables for factor analysis and for each secondary index variables have m data, that is there are m evaluated object. The value of the j-th second-level index of the i-th evaluation object under a certain first-level index is X_{ij}. Then convert each index value into a standardized scalar $\widehat{x_{ij}}$ by the following formula:

$$x_{ij} = \frac{x_{ij} - \overline{x_j}}{S_j}, (i = 1,2,\dots,n;\ j = 1,2,\dots,m) \tag{1}$$

Where $\overline{x_j}$ is the mean of the j-th second-level index sample which is

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{2}$$

And S_j is the standard variance of the j-th second- level index sample which is:

$$S_j = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 \tag{3}$$

Apart from that, correlation coefficient matrix can also be obtained that is $R = (r_{ij})_{m\times m}$, among which

$$r_{ij} = \frac{1}{n-1}\sum_{i=1}^{n} x_{ki} \cdot x_{ki}, (i,j=1,2,...,m) \tag{4}$$

In equation (4), we know that $r_{ij} = 1, r_{ij} = r_{ji}$, and r_{ij} is the correlation coefficient of i-th second-level index and j-th second-level index.

(2) Calculate the elementary load matrix

Calculate the eigenvalues λ_i of the correlation coefficient matrix R, where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_m \geq 0$ and calculate the corresponding eigenvectors which are denoted by $u_1, u_2, \dots, u_m$ where $u_j = (u_{1j}, u_{2j}, \dots, u_{nj})^T$.

(3) Take all the factors and rotate them

According to the elementary load matrix, the contribution rate of each common factor can be calculated. Since the size of the sample data is small, we adopt all the common factor on the principle that the cumulative variance contribution rate reaches 100% in order not to lose the precise and rare sample data. Then we rotate the extracted factor load matrix to obtain the matrix $B = A_2 \cdot T$, where A_2 is the first two column of A, and T is the orthogonal matrix, then construction factor model as follows:

$$\begin{cases} x_1 = a_{1,1}F_1 + a_{1,2}F_2 \\ \dots \\ x_m = a_{m,1}F_1 + a_{m,2}F_2 \end{cases} \tag{5}$$

In order to explain the practical meaning of each common factor and reduce the subjectivity of interpretation, we can use factor rotation to achieve the purpose.

calculate the score of the factors

In this paper, the single factor score function is obtained by regression method, and we can set the formula as the following:

$$\begin{cases} F_1 = b_{1,1}x_1 + b_{1,2}x_2 \\ \dots \\ F_n = b_{n,1}x_1 + b_{n,2}x_m \end{cases} \tag{6}$$

Take the i-th sample point versus the j-th factor, then we can obtain the estimate score of F

$$F_{ij} = b_{j,1}x_{i,1} + b_{j,2}x_{i,2} + \ldots + b_{j,p}x_{i,p} \quad (7)$$

$$(i = 1,2,\ldots,m, j = 1,2)$$

Then we have the following equation,

$$\begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ \vdots & \vdots \\ b_{1,m} & b_{2,m} \end{bmatrix} = R^{-1}B \quad (8)$$

Then $F = (F_{ij})_{n\times m} = X_0 \cdot R^{-1} \cdot B$, where X_0 is $n \times m$ is the original data matrix, R is the correlation coefficient matrix, B is the load matrix obtained from step (4)

4.2. Calculation of the ecological benefit score on the basis of entropy weight TOPSIS

TOPSIS method is a multi-attribute decision-making method with finite options. In essence, entropy weight TOPSIS is an improvement of traditional TOPSIS method. Its main feature is to determine the weight of each evaluation index based on entropy weight method, and then determine the ranking of evaluation objects by using the technology of approximating the ideal solution. The basic idea of the method is as follows: firstly, determine the ideal solution (negative ideal solution), that is, the value of each attribute has reached the optimal (inferior) value in the alternative scheme; Then, the level of superiority and inferiority is measured by the relative distance between each evaluation object with the optimal solution and the worst solution. If the evaluation object is closest to the optimal solution and furthest away from the worst solution, it is optimal.Otherwise, it is not optimal.

Determine the indicator type

Depending on different features, indicators can be divided into four types: Benefit Type Indicator which is the larger, the performance becoming better, Cost Type indicator which is the smaller, the performance becoming the better, Intermediate indicator which is the closer to a defined value, the better and Interval indicator which is the more lies in the interval, the performance getting better. In order to calculate easier. We change all data into the same trend. In this case, we change all indicators into benefit type indicators.

Data Standardization

Since the data are all Benefit Type Indicators, we standardize the data to eliminate the effect, The standardization equation should be:

$$z_{ij} = \begin{cases} \dfrac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} & x_{ij} \geq 0 \\ \dfrac{x_{ij} - min\{x_{1j}, x_{1j}, \ldots, x_{1n}\}}{max\{x_{1j}, x_{1j}, \ldots, x_{1n}\} - min\{x_{1j}, x_{1j}, \ldots, x_{1n}\}}, & \\ x_{ij} < 0 & \end{cases} \quad (9)$$

Then we obtain the normalized matrix Z:

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & Z_{23} & \cdots & Z_{2n} \\ Z_{31} & Z_{32} & Z_{33} & \cdots & Z_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & Z_{m3} & \cdots & Z_{mn} \end{bmatrix} \quad (10)$$

Determine the ideal solution

The normalized matrix was examined to determine the maximum values of the three indicators in the two countries and set them as the maximum reference sample (positive ideal point).Then, find the minimum value of the three indicators, and determine the minimum reference sample (negative ideal point).The specific formula are as follows:

Determine the weight of indicators.

The entropy weight is calculated on the basis of each evaluation index value after the set of evaluation objects is given. The smaller the index entropy is, the larger the entropy weight is which means that the index is more important. Conversely, index is less important. Entropy weight method can effectively eliminate subjective error, and the specific operation methods are as follows:

First, set

$$E_j = -k\sum_{i=1}^{m} f_{ij} \ln f_{ij} \tag{11}$$

Among which,

$$f_{ij} = \frac{z_{ij}}{\sum_{i=1}^{m} z_{ij}}, k = \frac{1}{\ln m} \tag{12}$$

Then the coefficient of difference was calculated which is$G_j = 1 - E_j$. The greater the coefficient G_j is, the bigger the difference as well as the weight are. Finally, determine the weight of indicators:

$$w_j = \frac{G_j}{\sum_{j=1}^{n} G_j} \tag{13}$$

Where $W_j \in [0,1]$ and $\sum_{j=1}^{n} W_j = 1$.

Calculate the Euclidean distance

With data above, we find out the Euclidean distances between the sample index and the positive / negative ideal solution can be calculated as follows:

$$D_i^- = \sqrt{\sum_{j=1}^{m} w_j (Z_j^- - z_{ij})^2} \tag{14}$$

$$D_i^+ = \sqrt{\sum_{j=1}^{m} w_j (Z_j^+ - z_{ij})^2} \tag{15}$$

Where W_j is the weight of the j-th indicator.

It should be noted that if each index in the sample reaches the maximum value, the distance between the sample and the positive ideal solution$D^+$is 0, and the distance between the sample and the negative ideal solution $D^-$is 1, indicating that the Eco-environmental system of the sample has reached the optimal state. On the contrary, it indicates that the situation of sample Eco-environmental system is poor.

Calculate the degree of compactness

Use the following formula to determine the degree of relative closeness$C_i$between the evaluated object and the positive ideal solution:

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{16}$$

If c_i is approaching to 1, the performance of the evaluated object is becoming better.

Thus the ecological environment assessment model has been built, and C_i is the ecological benefit score.

5. Result of the score

The annual score of the first-level indicator was obtained through factor analysis of the second-level index of Saihanba, and the scientific and objective weight of the first-level index was obtained through entropy weight TOPSIS method (as shown in the pie chart below). Besides, the score of

Saihanba in 1983, 1991, 2000 and 2010 can also be obtained. The scores of first-level indicator of Saihanba in different year is shown as table2. The scores trend of first-level indicator of Saihanba in different year is shown as figure6.

Table 2 The scores of first-level indicators of Saihanba in different year

Indicators / Year	Forest Carbon Sequestration	Conservation of Water	Microclimate	Air quality
1983	-1.0509	0.811594203	0.1909	0.766781216
1991	-0.5058	0.862318841	0.1766	0.843541971
2000	0.5095	1	0.3594	0.982142769
2010	1.0472	0.905797101	0.273	1

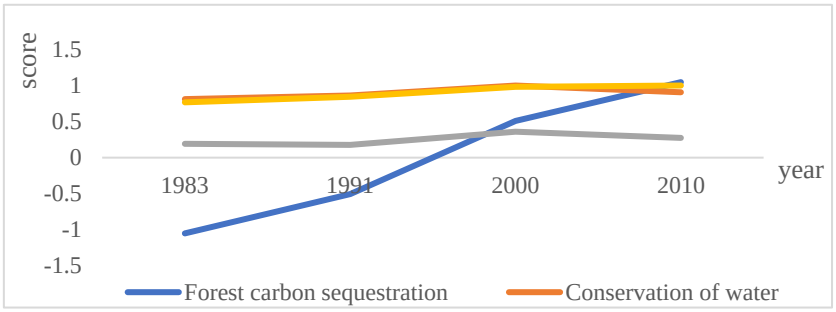


Figure 6 The scores trend of first-level indicator of Saihanba in different year、

We can conclude from the figure above that the four first-level indicators of Saihanba Forest Farm show an overall upward trend. Among them, the forest sequestration score increased the most, and the left showed a steady upward trend from 1991 to 2010.

Since we have obtained the scores of the first-level indicator of Saihanba Tree Farm, then we use Entropy Weight TOPSIS to obtain the weight of the first-level indicator and the ecological benefit score. The weight of the first-level indicators of Saihanba is shown as table2. The scores of ecological benefits of Saihanba in different year is shown as figure7.

Table 3 The weight of the first-level indicators of Saihanba

Indicator	Forest Carbon Sequestration	Conservation of Water	Microclimate	Air quality
Weight	0.87945395	0.007066542	0.099423301	0.014056207

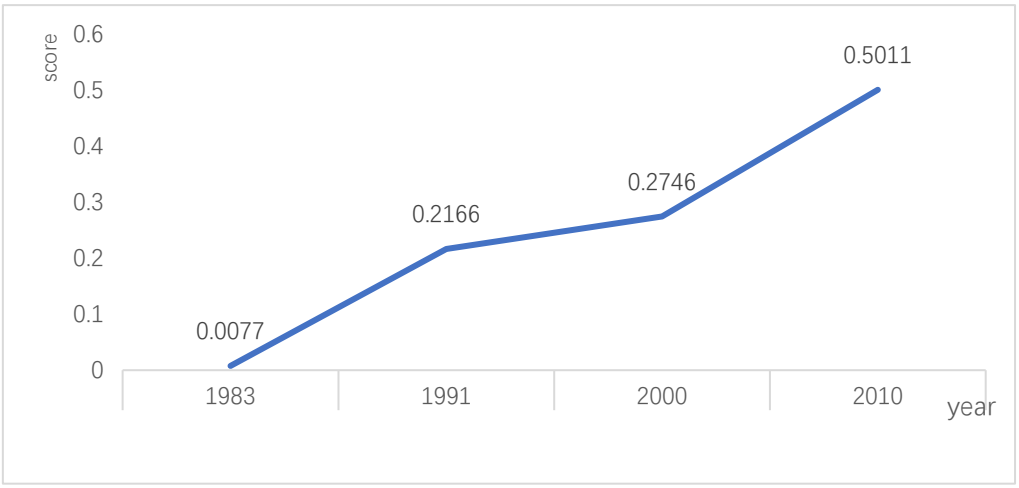


Figure 7 The scores of ecological benefits of Saihanba in different year

The evaluation of ecological benefit shows an increasing trend year by year. In particular, there was a big breakthrough from 1983 to 1991 that is the ecological efficiency score rose from 0.0077 to 0.2166, an increase of about 28.19 times. Additionally, the ecological efficiency score have doubled from 2000 to 2010 which increase from 0.2746 to 0.5011. According to their own ecological efficiency system, the scores of Beijing, Hebei and Tianjing which are the main surrouding cities with Saihanba saw an increasing trend in recent decades. Specifically, The score of ecological benefit of Tianjin city increased from 0.0231 in 1991 to 0.1671 in 2018 which means the establishment of the Saihanba ecological zone affects its surrounding areas postively.

References

[1] Zhang Huiyuan,Wei Guiwu,Wei Cun. TOPSIS method for spherical fuzzy MAGDM based on cumulative prospect theory and combined weights and its application to residential location[J]. Journal of Intelligent & Fuzzy Systems,2022,42(3).

[2] Kustiyahningsih Y,Rahmanita E,Anamisa D. R.. Hybrid FAHP and TOPSIS to determine recommendation for improving SMEs facing Covid-19 pandemic[J]. Journal of Physics: Conference Series,2022,2193(1).

[3] Lam Weng Siew,Lam Weng Hoe,Mohd Abidin Bakar,Lee Pei Fun. Data Driven Decision Analysis on the Performance of Electronic Companies with TOPSIS Model[J]. JAIT,2022,13(1).

[4] Jiang Jirong,Ren Min,Wang Jiqiang. Interval number multi-attribute decision-making method based on TOPSIS[J]. Alexandria Engineering Journal,2022,61(7).

[5] Li, Qiang,Sui, Wanghua,Sun, Bangtao,Li, Dongli,Yu, Shibo. Application of TOPSIS water abundance comprehensive evaluation method for karst aquifers in a lead zinc mine, China[J]. Earth Science Informatics,2022(prepublish).

[6] Chakraborty Subrata. TOPSIS and Modified TOPSIS: A comparative analysis[J]. Decision Analytics Journal,2022,2.

[7] Wang Huadong,Xu Yunhong,Suryati Sulong Rini,Ma HuiLing,Wu Lifeng. Comprehensive Evaluation of Water Carrying Capacity in Hebei Province, China on Principal Component Analysis[J]. Frontiers in Environmental Science,2021,9.

[8] Agrawal, A.,Arora, D.,Kumar, A.,Jain, A.,Doshi, C.. RAPID EYE: an Aid Memoir for Comprehensive Evaluation of Patient in Post-cardiac Surgery Recovery Unit[J]. Indian Journal of Surgery,2021(prepublish).

[9] Tan Yuanqiang,Zhang Jiangtao,Li Xiang,Xu Yangli,Wu Chuan-Yu. Comprehensive evaluation of powder flowability for additive manufacturing using principal component analysis[J]. Powder Technology,2021,393.

[10] Xiao-xing Li,Jun-kai Du,Yao Fu,Jing-Jun Liu,Ya-xiao Di,Bo Pang. Post-Project Analysis Modeling of Environmental Impacts of Coal Mining Based on Structural Entropy Weight-Fuzzy Comprehensive Evaluation[J]. IOP Conference Series: Earth and Environmental Science,2021,826(1).