Environment, Resource and Ecology Journal (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/erej.2023.070501
ISSN 2616-3756 Vol. 7 Num. 5

Study on the statistical measure of green development in China based on entropy weight method and TOPSIS method

Chen Wang[1]

[1]*Nanjing University of Aeronautics and Astronautics, Nanjing, China*

Keywords: Entropy weight method, TOPSIS method, Pearson correlation, Hierarchical clustering analysis, Unary linear regression model

Abstract: The party's 20th Congress pointed out that we should promote the building of a beautiful China, adhere to the integrated protection and systematic management of mountains, rivers, forests, farmland, lakes, grass, and sand, coordinate industrial restructuring, pollution control, ecological protection, climate change, promote carbon reduction, pollution reduction, green expansion and growth, and promote ecological priority, economical, intensive, green and low-carbon development. Modern China has made remarkable achievements in green development. By applying entropy weight method and TOPSIS method, we select the appropriate evaluation index to evaluate the influence of ecological environment construction on ecological environment. First, appropriate indicators were selected and relevant data were collected to establish an evaluation model for the impact of ecological environment construction on ecological environment in Hebei Province. Then, appropriate indicators were selected, pertinent data were collected, and a statistical model was established to evaluate the impact of Saihan Dam on sandstorm resistance in North China, and the role of Saihan Dam in sandstorm resistance in North China was quantitatively evaluated. Finally, through the analysis of the carbon emission and carbon absorption rate of each province, calculate the carbon absorption rate of each province, predict the change trend of the carbon emission, carbon absorption rate and carbon absorption rate of each province in the coming years, and evaluate the progress made in carbon neutrality.

1. Introduction

Since the reform and opening up, China's economic and social development has made remarkable achievements. The rapid economic development is accompanied by the rapid advancement of industrialization and urbanization process. However, under the long-term extensive development model, the increasing ecological vulnerability, ecological deterioration and environmental pollution have constrained the modernization process, restricted economic and social development, and affected the quality of people's lives. To promote ecological progress is to address the efficient use of resources, reduce environmental pollution, and enhance the country's capacity for long-term sustainable development. In the huge systematic project of promoting ecological civilization construction, the most prominent weakness in China is the lack of systematic evaluation and overall cooperative governance mechanism, which will seriously affect the effectiveness of China's ecological civilization construction. The party's 20th congress pointed out that promoting green development and promoting harmonious coexistence between man and nature. Based on China's national conditions, it is a key and urgent task to vigorously promote the national ecological governance system and governance capacity.

Since the 18th National Congress, china has been practicing the concept of green water and green mountains is the silver mountain of gold, unswervingly taking the road of ecological priority and green development, building a modern society where people and nature live in harmony, and creating another remarkable ecological miracle and green development miracle.

China's green development has made an indelible contribution to the ecological construction and environmental protection of the world, benefiting both the Chinese people and the people of the world.

2. Processing and description of the data

Three categories of data on forest resources and meteorological conditions in the Beijing-Tianjin-Hebei region and the energy consumption of forest resources in China were collected through relevant departments. Data processing and chart production using Excel, Python. We can see the index for each data set, and query the number of missing values. The indicators with more missing values are discarded; those with fewer missing values are filled with the average value.

3. Model establishment and analysis

3.1 The construction of the evaluation system for the impact of green development on the ecological environment—Take Hebei province as an example

3.1.1 Selection of evaluation indicators

Table 1: The impact assessment system on the ecological environment in Hebei Province

Level 1 indicators	Level 2 indicators	Level 3 indicators	Indicator type
Comprehensive index of ecological environmental impact of Hebei Province A	environmental protection B1	Forest area C1	Positive
		Forest cover C2	Positive
		Forest stock C3	Positive
		Total area afforested C4	Positive
		Carbon dioxide absorption C5	Positive
	Maintain ecological balance B2	Area of nature reserve C6	Positive
		Number of nature reserves C7	Positive
		Area of national nature reserves C8	Positive
		Area of forest disease occurrence C9	Negative
		Area of forest pest infestation C10	Negative
	Changes to the local climate B3	Annual precipitation C11	middle
		Average annual temperature C12	middle
		Average annual maximum temperature C13	middle
		Average annual minimum temperature C14	middle
		Annual average wind speed C15	Negative
		Gale days C16	Negative

Reasonable construction of index system is the basis for the evaluation of the impact of green development on the ecological environment. The selection of index system in this paper follows the following principles. Comprehensiveness, Simplicity, Operability, Objectivity. The final evaluation system of impact on ecological environment in Hebei Province is shown in Table 1.

3.1.2 Normalization and forward of data

Because the units between the indicators are not unified, the parameter order of magnitude difference between the indicators is large. Therefore, the data need to be normalized. The most value method is normalized as follows:

$$x' = \frac{\bar{x} - x_{min}}{\mathrm{R}} \tag{1}$$

In the above equation, x' is the normalized data, x_{min} is the minimum value of the sample, and R is the extreme data, that is, $R = x_{max}-x_{min}$.

3.1.3 To solve the weight

After normalization the data, the proportion of the i item under the j index is calculated.

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{2}$$

The entropy values and weights of the j-th index were further calculated.

$$e_j = \frac{\left(\sum_{i=1}^{m} p_{ij}\right) \ln p_{ij}}{\ln m} \tag{3}$$

$$w_j = \frac{1 - e_j}{\sum_{i=1}^{n}\left(1 - e_j\right)} \tag{4}$$

3.1.4 The calendar year TOPSIS score was calculated

Let the data set matrix after normalization be $B = (b_{ij})_{m\times n}$ and the weighted matrix $C = (c_{ij})_{m\times n}$, using the weight vector $\vec{w} = (w_1, w_2, ..., w_n)$, then $c(ij) = w_j b_{ij}$. Where, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$.

The distance of the i-th evaluation object from the maximum and minimum values is shown in Equation (5). [1]

$$D_i^{\pm} = \sqrt{\sum_{j=1}^{n}\left(c_{ij} - c_j^{\pm}\right)^2} \tag{5}$$

The score of TOPSIS is shown in Equation (6).

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{6}$$

The plot of the TOPSIS score and the influence of the ecological environment in Hebei Province from 2004 to 2021 is shown in Figure 1.

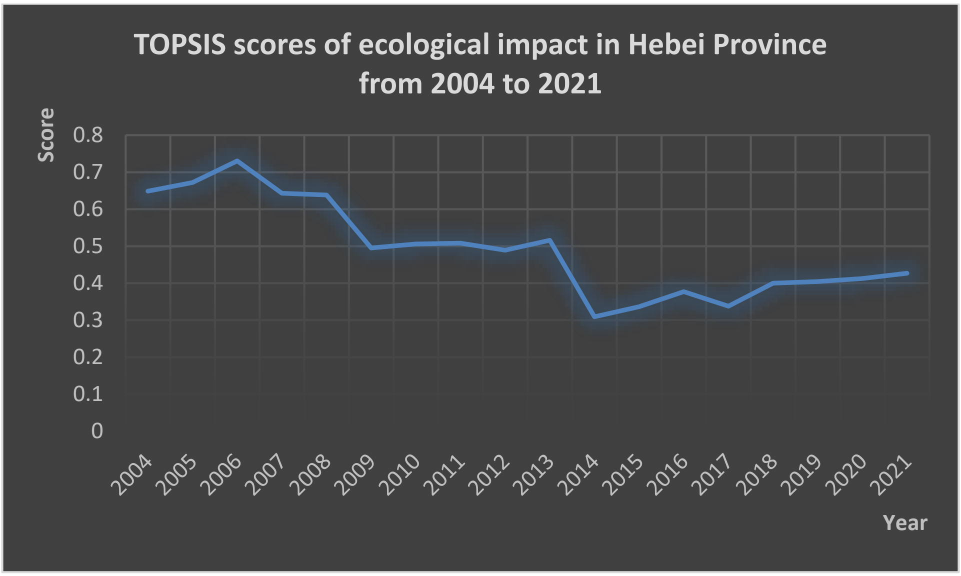

Figure 1: TOPSIS score for ecological environment impact in Hebei Province from 2004 to 2021

In the TOPSIS method, a lower score indicates the better the scheme. Because in the TOPSIS method, the scheme with a low score has a smaller distance from the positive ideal solution (the distance from the positive ideal solution), while the small distance is closer to the positive ideal solution, that is, the better. Therefore, the lower the score, the more ranked in the TOPSIS ranking (the smaller the score). From the graph, TOPSIS scores showed a downward trend from 2004 to 2021. This shows that the overall influence on the environment in Hebei province shows an upward trend, which is related to China's emphasis on green development in China.

3.2 The influence of Saihan Dam on the dust weather in North China—Take Beijing as an example

3.2.1 The causes and indicators of dust weather

Dust weather refers to a very harmful natural disastrous weather phenomenon where strong winds lift a large amount of dust on the ground, making the air particularly cloudy and the horizontal visibility is lower than 1km. Sandstorms mainly occur in the middle latitude zone in the northern hemisphere, and their source areas are located in arid and semi-arid areas. Related studies show that dust weather is related to gale days, NDVI (normalized vegetation index) and precipitation. Based on the above analysis, this paper will take Beijing in the south of Saihan Dam as an example to select forest area, actual forest coverage rate, air quality index, AQI), number of days of pollution, gale days, average wind speed and average annual rainfall to study the impact of Saihan Dam on dust weather in Beijing.

3.2.2 Correlation analysis

Pearson's correlation coefficient, also known as Pearson's product moment correlation coefficient and simple correlation coefficient, describes the closeness between two fixed distance variables. It is used to measure the correlation (linear correlation) between two variables x and y. Its value is [-1, 1], which is generally expressed by r. The calculation formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (7)$$

Generally, the correlation strength of variables is judged by the following value range, as shown in Table 2.

Table 2: Correlation coefficient and correlation control table

Correlation coefficient r	Relevance
0.8-1.0	Extremely strongly correlated
0.6-0.8	Strong correlation
0.4-0.6	Moderately relevant
0.2-0.4	Weakly correlated
0.0-0.2	Very weak or no correlation

The Pearson correlation coefficients of the forest area, actual forest cover and Beijing air quality index, pollution days, windy days, average wind speed and precipitation in Saihan Dam were calculated as shown in Figure 2.

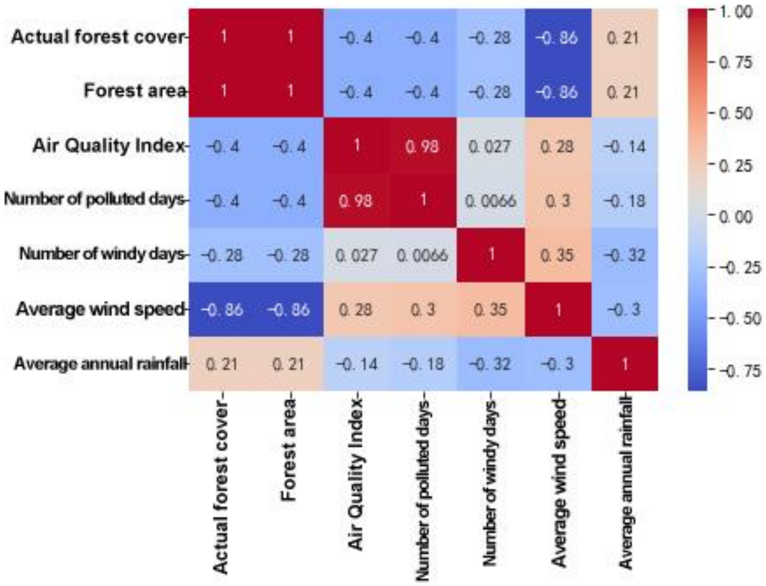

Figure 2: Thermal map of the correlation coefficient between the indicators

According to the above analysis, the correlation coefficient between the forest area and actual forest coverage rate of Saihan Dam and the air quality index, number of pollution days and average wind speed in Beijing is r 0.4, indicating that the forest area and actual forest coverage rate of Saihan Dam are negatively correlated with the air quality index, number of pollution days and average wind speed in Beijing.$r \geq 0.4$ indicates that the correlation is strong, and the lower the air quality index, pollution days and the average wind speed, the better the air quality, and the less likely to occur. Therefore, the increase of forest area and forest coverage rate in Saihan Dam can effectively improve the air quality of Beijing and reduce the occurrence of dust weather.

3.3 Evaluation of green development and provincial carbon-neutral targets

3.2.3 Statistics and calculation of data on carbon emissions and carbon absorption in the provinces

Using the data of provincial fossil energy consumption and forest area given by the National

Bureau of Statistics, the consumption of the i-th fossil energy is c_i, the carbon emission coefficient of the fossil energy is μ_i. The carbon emission factors corresponding to fossil energy using the estimation method in the 2006 IPCC Guidelines for National Greenhouse Gas Inventories are shown in Equation (8).[2-3]

$$\mu_i = NCV_i \times CC_i \times COF_i \times \frac{Mr(CO_2)}{Ar(C)} \tag{8}$$

Among them, NCV_i is the average low calorific value of the i energy, CC_i is the unit calorific value of the i energy, COF_i is the carbon oxidation ratio of the i energy (complete oxidation at $COF_i = 1$), $Mr(CO_2)$ is the relative molecular mass of carbon dioxide (usually $Mr(CO_2) = 44$), and $Ar(C)$ is the relative atomic mass of carbon atoms (usually $Ar(C) = 12$).

3.2.4 Hierarchical clustering analysis

Given a sample set X, X is the set of points in the m dimensional real vector space R^m, where, x_i, $x_j \in X$, $xi = (x1i, x2i, ..., x_{mi})^T$, $x_j = (x_{1j}, x_{2j}, ..., x_{mj})\ T$, x_i and x_j Euclidean distances are defined as:

$$d_{ij} = \left(\sum_{k=1}^{m} \left| x_{ki} - {x_{kj}}^2 \right| \right)^{\frac{1}{2}} \tag{9}$$

For n samples, the Euclidean distance between pairs was calculated, and the distance matrix $D = [d_{ij}]_{n \times n}$. Construct n classes, with only 1 sample in each class. Combine the two classes with the smallest class spacing, define the inter-class distance as the shortest distance, and construct a new class. Calculate the distance between the new class and the current class, until the number of classes is only 1 [4].

Every four years from 2006 to 2021 (2006, 2010, 2014 and 2018), a hierarchical clustering map, as shown in Fig. The provincial carbon absorption rate levels were divided into three categories, as shown in Table 3.

Table 3: Clustering table of carbon absorption rate by province

Year	Provinces with a carbon uptake rate of 1($\eta = 1$)	Provinces with high carbon uptake rates($\eta \in (0.5,1)$)	Provinces with low carbon uptake rates($\eta \in [0,0.5]$)
2006	Other provinces	Jiangsu Province	Tianjin, Shanghai
2010	Other provinces	Jiangsu Province, Shandong Province	Tianjin, Shanghai
2014	Other provinces	Jiangsu Province, Shandong Province	Tianjin, Shanghai
2018	Other provinces	Jiangsu Province, Shandong Province	Tianjin, Shanghai

We can conclude that the carbon absorption rate of Tianjin and Shanghai is lower than that of other provinces. The main reason is the high level of economic development in these areas, the high degree of urbanization, land use is mostly urbanization land or industrial land, and a large number of human activities will lead to the destruction of the ecological environment, affecting the vegetation growth and carbon absorption capacity. Secondly, the land use methods in these areas are mostly agricultural, urban land or industrial land, especially the regular large-scale real estate construction in urban areas, which leads to the logging and occupation of a large number of original green space, leading to the destruction of the ecosystem and the decline of carbon absorption capacity. For these reasons, the following measures can be taken:

(1) Increase the green space

(2) Optimal land utilization
(3) Develop an ecological economy
(4) We will promote environmental education
(5) We will support scientific and technological innovation

3.2.5 Prediction of carbon absorption rate based on a unitary linear regression model

The unary linear regression model represents the relationship of the dependent variable y and the independent variable x as

$$y = kx + b + \epsilon \tag{10}$$

Where k, x are the regression coefficient, and ϵ is the error term of the model, aiming to balance the values on both sides of the equation. The key to determining the regression coefficient is to measure the error term ϵ, and the mean squared error is the most commonly used performance measure of the regression models, in order to minimize the mean squared error [5][6].

$$F(k,b) = \sum_{i=1}^{n} \epsilon^2 = \sum_{i=1}^{n} (y_i - kx_i - b)^2 \tag{11}$$

Eventually, the values of k, b can be obtained

$$\begin{cases} k = \dfrac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n} x_i^2 - \dfrac{(\sum_{i=1}^{n} x_i)^2}{n}} \\ b = \bar{y} - k\bar{x} \end{cases} \tag{12}$$

We took Tianjin as an example to divide the data set into training set and test set in a ratio of 8:2. The linear regression equation is trained from the training set, and the actual data and the data predicted by the linear regression model are compared and tested for [7]. Using the mean squared error (MSE) test, the value of the MSE is 0.0015314534417283417. The value range of MSE is between 0 and positive infinity, and the smaller the value, the better the model predicts. Ideally, the value of the MSE should be 0, indicating that the model perfectly predicts all the true values. While the scatter plot and linear regression equation images of the real data are shown in Figure 3.

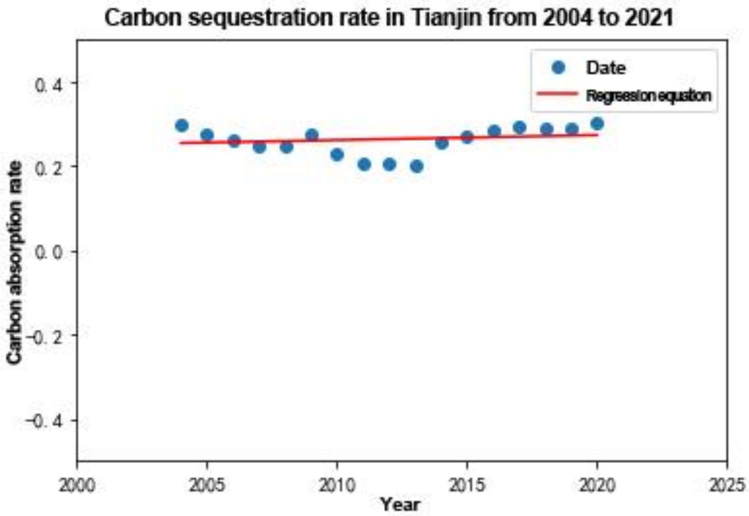

Figure 3: Actual data and predicted curve of carbon uptake rates from 2004 to 2021

We can see from Figure 3 that the linear regression equation is well fitted to the real data. However, due to the small amount of data here, there may still be a large error. The prediction ability and stability of the model can be improved by increasing the sample size and improving the modeling method.

4. Conclusions

On how to continue to promote green development, it is necessary to formulate targeted policies and plans and support measures according to the actual situation and development bottlenecks of the provinces. For example, provinces with high carbon emissions can strengthen environmental protection and promote green energy and clean production, and provinces with low carbon emissions can strengthen environmental protection industries and promote a low-carbon economy. In the process of promoting green development, we need to strengthen policy guidance, improve laws and regulations, increase investment support, encourage innovation and strengthen cooperation, so as to provide solid support for the national goal of carbon neutrality.

Entroropy weight method, TOPSIS method, correlation analysis, cluster analysis and regression model can be applied in the green development evaluation and decision-making process of governments at all levels, enterprises and institutions, social organizations and the public. For example, at the government level, the above model can be used to comprehensively evaluate and compare the progress and effect of green development in various regions and industries, so as to formulate corresponding policies and measures to promote green development. At the enterprise level, the above model can be used to evaluate the environmental protection level and green development achievements of enterprises, encourage enterprises to strengthen environmental protection investment and technological innovation, and promote sustainable development with green and low-carbon as the core. At the level of public participation, community environmental protection activities and publicity and education will be conducted to enhance public awareness of environmental protection and participation enthusiasm, and promote green consumption and lifestyle transformation.

References

[1] Yi Zechuan, Tan Xijun, He Runhui. Ecological and environmental impact assessment of Saihan Dam ecological model based on TOPSIS [J]. Energy and Energy Conservation, 2022.

[2] GB / T 2589-2020, the general rules of comprehensive energy consumption calculation [S].

[3] Zheng S , Zhang C , Zhou Y . Study on the evaluation system of green development in Tuojiang River Basin based on entropy weight method and grey relation analysis[J].IOP Conference Series: Earth and Environmental Science, 2020, 512(1):012046 (6pp).DOI:10.1088/1755-1315/512/1/012046.

[4] Li Hang. Machine learning methods [M]. Beijing: Tsinghua University Press, 2022.

[5] Zhou Zhihua. Machine learning [M]. Beijing: Tsinghua University Press, 2016.

[6] Jun-Min Z, Bo P. Study on Comprehensive Development of Zhoushan Port in Ningbo based on Entropy Weight Method and TOPSIS Method[J]. Special Zone Economy, 2018.

[7] Xue Wei. Statistical Analysis and Application of SPSS [M], Beijing: China Renmin University Press, 2021.