Environment, Resource and Ecology Journal (2021) 5: 55-59
Clausius Scientific Press, Canada

DOI: 10.23977/erej.2021.050411
ISSN 2616-3756

Evaluation and Prediction of Ecological Environment based on Principal Component Analysis

Kaige Chen[1], Junhong Long[2], Huizhen Zheng[3, *]

[1]*School of Mechanical and Transportation, Taiyuan University of Technology Engineering, Taiyuan, Shanxi, 030002, China*

[2]*School of computer science and Engineering, Jiangsu university of science and technology, Zhenjiang, Jiangsu, 212003, China*

[3]*School of Public Administration, China University of Geosciences, Wuhan, Hubei, 430074, China*

*Corresponding author

Keywords: HSB model, Principal component analysis, Linear fitting

Abstract: Based on HSB model, principal component analysis and linear fitting are used in this paper. First of all, 11 factors that may affect the ecological environment are collected and quantified. Secondly, the dimension of the influencing factors is reduced based on principal component analysis. Finally, the correlation coefficient between the main influencing factors and the ecological environment is obtained based on the linear model. The three factors that have the greatest impact on the environment of Saihanba are: climate resources factor, forest resources factor and water resources factor.

1. Introduction

Understanding the changes in forest ecosystems through the evaluation of ecosystem service values is getting more and more attention and importance from scholars worldwide. Sehamba is located at the southern edge of Hunsandak sandy area, which is the transition zone from Inner Mongolia plateau to North China Mountains, and has the important mission of blocking sand source, preserving water source and maintaining ecological security for Beijing and Tianjin.

2. HSB model building

In this paper, to address the standardization and diversification of factors affecting coal, we collected 11 factors affecting ecological environment in 13 years from 2006 to 2020, cleaned, interpolated and standardized the data, and established HSB model based on principal component analysis to solve the main factors affecting ecological environment [1].

(1) Vegetation cover

Table 1: Area share of vegetation cover in 3 periods in Sehanba mechanical forestry

Vegetation cover class	Area share /%		
	1980year	2000year	2019year
Low coverage	14	9	7
Lower coverage	7	8	4
Medium coverage	16	23	16
Higher coverage	29	30	32
High coverage	34	30	42

Table 2: Changes in vegetation cover landscape index from 1980 to 2019 in Seyhanba Mechanical Forestry

Year	NP	LPI	AI	CONTAG	SHEI	SHDI
1980	8665	16.4647	87.3461	39.1058	0.9436	1.5187
2000	42234	87.3461	71.2133	32.1761	0.8763	1.4103
2019	39058	74.7048	74.7048	36.8382	0.8102	1.3039

(2) Biological abundance index

The biological abundance index, which refers to the difference in the number of biological species [2] per unit area of different ecosystem types, indirectly reflects the degree of biological abundance and poverty in the evaluated area.

Biological Abundance Index = A_{bio} × (0.35 × forest land + 0.21 × grassland + 0.28 × water wetland + 0.11 × arable land + 0.04 × construction land + 0.01 × unused land)/area

$$A_{bio} = 100 / A_{\max} \tag{1}$$

Where, A_{bio} normalized index of biological abundance, $A_{\max}$ maximum value of biological abundance index before normalization treatment. Where the weights of each land cover type were assigned as shown in Table 2.

(3) Precipitation change

3. HSB model solution

3.1 Data pre-processing

In this paper, a large amount of data of model input variables are collected, and firstly, the missing values [3] are filled by interpolation fitting, followed by singular value processing and data cleaning.

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

where X_{norm} is the normalized data, X is the original data, $X_{\max}$ is the maximum value of the original data set, and $X_{\min}$ is the minimum value of the original data set. After the normalization process, we can further process the data [4].

3.2 Applicability test

Before performing principal component analysis on the data, it is necessary to test the data for suitability in order to ensure the suitability and reasonableness of the data used. Checking the applicability of principal component analysis.

Assuming that the correlation coefficient of X_i and X_j is r_{ij}, we can determine the correlation coefficient between the variables based on the following calculation [5].

$$r_{ij}=\frac{\operatorname{cov}\left(X_i,X_j\right)}{\sqrt{Var\left[X_i\right]Var\left[X_j\right]}} \tag{3}$$

Where $\operatorname{cov}\left(X_i,X_j\right)$ is the covariance of X_i and X_j, and $Var\left[X_i\right]$ and $Var\left[X_j\right]$ are the variances of X_i and X_j, respectively. If most of the correlation coefficients are greater than or equal to 0.3, then we will be able to judge the preliminary applicability of the principal component analysis method.

After the calculation, the obtained data overwhelmingly satisfy the above conditions, so it is initially determined that the principal component analysis method can be applied. Next, we conduct further tests.

4. Principal component analysis

Calculate the correlation coefficient matrix. The calculated matrix has the form.

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \tag{4}$$

$r_{ij}\left(i,j=1,2,\cdots,p\right)$ is the correlation coefficient of X_i and X_j, which is calculated as

$$r_{ij}=\frac{\Sigma_{k-1}^{n}\left(x_{ki}-\bar{x}_i\right)\left(x_{kj}-\bar{x}_j\right)}{\sqrt{\Sigma_{k-1}^{n}\left(x_{ki}-\bar{x}_i\right)^2\Sigma\left(x_{kj}-\bar{x}_j\right)^2}} \tag{5}$$

Considering that R is a real symmetric matrix, it is sufficient to calculate only the upper or lower triangular elements.

First solve the eigenequation to find the eigenvalues $l_i=\left(i=1,2,\cdots,n\right)$ and put them in order of magnitude. Then the eigenvectors $e_i=\left(i=1,2,\cdots,n\right)$ corresponding to the eigenvalues are found separately. The eigenvalues are calculated from the correlation coefficient matrix, as well as the contribution rate and cumulative contribution rate of each principal component.

The Contribution Rate (CR) is

$$CR=\frac{\lambda_i}{\Sigma_{k=1}^{p}\lambda_k}\left(i=1,2,\cdots,p\right) \tag{6}$$

The Cumulative Contribution Rate (CCR) is

$$CCR=\frac{\Sigma_{k=1}^{i}\lambda_k}{\Sigma_{k=1}^{p}\lambda_k}\left(i=1,2,\cdots,p\right) \tag{7}$$

The calculation results are as follows:

Table 3: Results of the component matrix

Component factor	Component 1	Component 2	Component 3
AQI	0.1409	-0.3346	0.3294
TSP	0.2018	0.1437	-0.1344
$PM_{2.5}$	0.2467	-0.2851	-0.3104
PM_{10}	-0.0042	0.0227	-0.5829
f_{area}	0.1441	0.2170	0.4038
$l\ use$	-0.2273	0.4944	0.2896
$f_{\ Coverage}$	-0.4376	0.0321	-0.2994
T_{number}	-0.3953	0.1207	0.1680
w	-0.0623	0.2381	-0.3933
S_{rich}	-0.1894	-0.7482	0.1770
CO_2	0.6529	-0.1404	-0.1802
CR	0.6814	0.2460	0.0606
ACR	68.14%	92.74%	98.80%

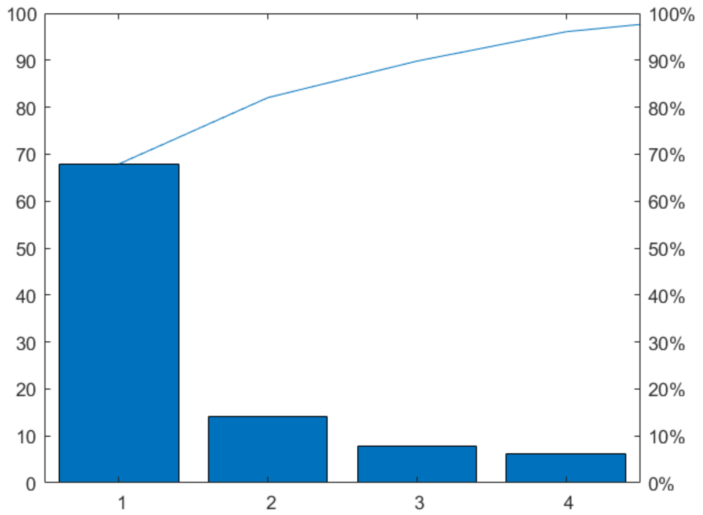

Figure 1: Contribution of principal components

For Example:

$$Y_1 = 0.1409AQI + 0.2018TSP + 0.2467PM_{2.5} - 0.0042PM_{10} + 0.1441$$
$$f_{area} - 0.2273l_{use} - 0.4376f_{coverage} - 0.3953T_{number} - 0.0623w - 0.1894S_{rich} \quad (8)$$
$$+0.6529CO_2$$

In the expression of the first principal component, it can be seen that CO_2, AQI, and TSP have high loadings, and their correlations are strong, so this principal component is defined as the climate resource impact component. In the expression of the second principal component, it can be seen that l_{use} occupy higher loadings, so this principal component is defined as a forest resource influence component. In the expression of the third principal component, it can be seen that water resources w

has a higher influence on this principal component, so this principal component is defined as the water resources influence component.

By analyzing the graph of the total variance explained by the principal components, it can be seen that the cumulative contribution percentage of the screened three principal components to the initial 11 variables is 96.80%, which is much more than 85% and has a strong representation, so this paper classifies the dominant influencing factors of ecological environment into three categories: climate resource influencing component, forest resource influencing component, and water resource influencing component.

$$Y = \frac{5}{12}Y_1 + \frac{1}{3}Y_2 + \frac{1}{4}Y_3 \tag{9}$$

Then the relationship between Y (ecological environment index) and climate resource influence component, forest resource influence component, and water resource influence component is derived.

5. Conclusion

Around the ecological environment of Saihanba, based on principal component analysis and linear fitting method, firstly, 11 factors that may affect the ecological environment are collected and quantified. Secondly, the dimension of the influencing factors is reduced based on principal component analysis. Finally, based on the linear model to solve the correlation coefficient between the main influencing factors and the ecological environment, it is concluded that the factors affecting the Saihamba environment from high to low are climate resources factor, forest resources factor and water resources factor. To contribute to the research on the influencing factors of ecological environment.

References

[1] Xing Meihua, Huang Guangti, Zhang Junbiao. A review of theoretical methods and empirical studies on forest resource valuation [J]. Journal of Northwest Agriculture and For-estry University: Social Science Edition, 2007

[2] Tian Jun, Liu Haiying, Cheng Shun, et al. Rumination on the climate and atmos-pheric contribution of Sehanba forest [J]. Hebei Forestry and Fruit Research, 2012

[3] Shen Songyu, Chen Weilin. Changes in vegetation cover in the source area of sand-storms--Beijing as an example, China Science and Technology Information, 2015

[4] Jiang Hanying, Duan Yiran, Zhang Zhe et al. Statistically based study of C02 emis-sion peaking in typical large cities in China. Advances in Climate Change Research, 2021

[5] Qiu Shengrong, Zhang Ximing, Bai Ling et al. Study on the construction of nature reserve planning system in China, World Forestry Research, 2021.