Regional Light Pollution Risk Level Measurement Study Based on PCA-EWM Model

Yizhe Tian [1], Ziyi Wang [*, 2], Meilei Zhao [2]

[1] Quality and Technical Supervision College, Hebei University, Baoding, Hebei, China 071032

[2] School of Economics, Hebei University, Baoding, Hebei, China 071032

* Corresponding Author Email: ziyiw15@gmail.com

Abstract. Light pollution exerts an influence on environment and poses a great threat to health and safety. In order to measure the light pollution risk level of a certain area, this paper constructs the evaluation index system of light pollution from five aspects: urbanization, industrialization, biology, atmospheric environment and geographical location. Subsequently, PCA-EWM Model is established to calculates the light pollution score of a location. According to the score, three light pollution risk levels are established: high light pollution, moderate light pollution and non-light pollution. Based on the above model, this paper collects sample data of urban community, suburban community, rural community and protected land in the world, and calculate light pollution score in four diverse types of locations. Finally, this paper makes a judgment on the risk level of different locations.

Keywords: PCA-EWM Model, Light Pollution Risk Level, Index System.

1. Introduction

Light pollution is a form of environmental pollution that is caused by excessive or misdirected artificial light. The effects of light pollution are numerous. For one, it can disrupt the natural cycles of many plants and animals, It can also affect human health. Therefore, it is necessary to analyze and study light pollution. In order to measure the level of light pollution risk in a certain area, scholars measured the level of light pollution through different methods, such as satellite images of the night sky [1] and reflection of light [2]-[4].There are also scholars who have carried out research on a certain aspect of light pollution, such as the problem of light pollution problems of overpass lighting [5] and residential areas [6].In addition, the negative effects of light pollution are studied, including the harm of light pollution to human health, the negative effects of traffic electronic monitoring flashes [7]. The existing research on light pollution is relatively rich, but few scholars comprehensively evaluate the light pollution level of a certain area. Therefore, this paper constructs a light pollution evaluation index system, and establishes a principal PCA-EWM model to measure the comprehensive light pollution level of a certain area, so as to enrich the research content of light pollution.

2. Establishment of Indicator System

2.1. Index selection

Light pollution refers to the excessive and wrong use of artificial light, which leads to over-illumination of the night sky. Although artificial lighting is an integral part of modern life, increased light pollution has a series of negative effects on the environment, wildlife and human health [8]. Light pollution can destroy natural ecosystems, affect human health in various ways, and cause waste of energy and resources. In order to determine the risk level of light pollution at any point, this paper selects 31 cities on the earth and uses five first-level indicators and 12 second-level indicators to comprehensively evaluate. The indicator system is shown in Figure 1.

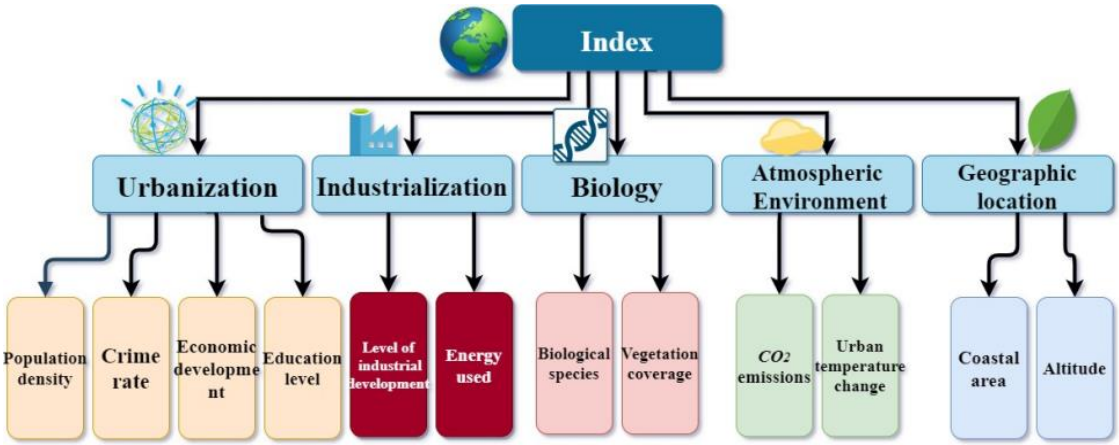


Figure 1 Evaluation index system of light pollution

2.2. Data preprocessing

1. Data cleaning: This article found 36 sets of sample data through the Internet. We chose to delete the sample data with too many missing values, repair and fill some features of the samples with few missing values, and finally retained 31 sets of data. Data integration: In this paper, data from multiple data sources are combined and stored uniformly, thereby establishing a data warehouse.

2. Data change: Different indicators have different weight impacts: the difference in magnitude will not only lead to the dominance of attributes with larger magnitudes, but also slow down the iterative convergence speed.

In order to eliminate this effect, this paper performs min-max standardization on the data set. For each index, set Amin and Amax as the minimum and maximum values of index A, respectively, map the original value X of A into the interval X' in [0,1], through the formula:

When the indicator A is a positive indicator,

$$X' = \frac{X - minA}{maxA - minA} \tag{1}$$

When indicator A is a negative indicator,

$$X' = \frac{maxA - X}{maxA - minA} \tag{2}$$

2.3. Index rationality verification

In order to verify that the indicators selected in this paper have wide applicability, this paper constructs a coupling coordination model to reflect the degree of interdependence and mutual restriction between indicators. The coupling coordination degree model is used to analyze the coordinated development level of things. Coupling degree refers to the interaction and influence between two or more systems, and realizes the dynamic correlation relationship of coordinated development, which can reflect the degree of interdependence and mutual restriction between systems [9].

First, the weight of each indicator is calculated by the entropy weight method:

$$u_i = \sum_{i}^{m} w_{ij} u_{ij} \tag{3}$$

$$\sum_{i}^{m} w_{ij} = 1 \tag{4}$$

Among them, u_i is the ordered contribution of the subsystem i to the total subsystem; u_{ij} is the normalized value of the index j of the subsystem i.

The interaction coupling degree model of multiple indicators can be expressed by the following model:

$$C_n = \left\{ \frac{(u_1 u_2 \cdots\cdots u_n)}{\prod (u_1 + u_2)} \right\}^{1/n} \tag{5}$$

C_n is the coupling degree of the n-element system; $u_1 u_2 \cdots\cdots u_n$ are the contribution of the first subsystem to the n^{th} subsystem to the total system order degree by the formula:

$$D = (C \cdot T)^{1/2} \tag{6}$$

$$Y = \beta_1 U_1 + \beta_2 U_2 + \beta_3 U_3 + \cdots \tag{7}$$

Among them, $\beta_1, \beta_2, \beta_3$ represent the weight.

Calculate the coordination index *T* value and coupling coordination degree *D* value of each region, as shown in Table 1:

Table 1 coordination analysis results

index item	C value	T value	D value	Coordination level	Degree of coupling coordination
New York	0.465	0.512	0.488	5	On the verge of disorder
Paris	0.669	0.520	0.590	6	Barely coordinated
Montpellier	0.718	0.405	0.540	6	Barely coordinated
Vancouver	0.587	0.356	0.457	5	On the verge of disorder
Beijing	0.560	0.351	0.443	5	On the verge of disorder
Shandong	0.741	0.458	0.582	6	Barely coordinated
Shanghai	0.563	0.381	0.463	5	On the verge of disorder
Tokyo	0.415	0.276	0.338	4	mild disorder
Seoul	0.711	0.600	0.653	7	Junior Coordinator
Mumbai	0.730	0.591	0.657	7	Junior Coordinator
New Delhi	0.753	0.464	0.591	6	Barely coordinated
Mandalay	0.774	0.371	0.536	6	Barely coordinated
Kuala Lumpur	0.641	0.420	0.519	6	Barely coordinated
Yangon	0.687	0.368	0.503	6	Barely coordinated
Moscow	0.803	0.491	0.628	7	Junior Coordinator
Helsinki	0.635	0.381	0.491	5	On the verge of disorder
Gothenburg	0.645	0.328	0.460	5	On the verge of disorder
Kyiv	0.446	0.345	0.392	4	mild disorder
Copenhagen	0.445	0.413	0.429	5	On the verge of disorder
Yienna	0.726	0.290	0.459	5	On the verge of disorder
Toronto	0.442	0.363	0.400	5	On the verge of disorder
London	0.604	0.471	0.533	6	Barely coordinated
Boston	0.628	0.356	0.473	5	On the verge of disorder
Brighton	0.664	0.315	0.458	5	On the verge of disorder
Chiang rai	0.726	0.329	0.489	5	On the verge of disorder
cape town	0.337	0.382	0.359	4	mild disorder
Cairo	0.549	0.398	0.467	5	On the verge of disorder
Beirut	0.435	0.379	0.406	5	On the verge of disorder
Melbourne	0.453	0.487	0.470	5	On the verge of disorder
Berlin	0.492	0.415	0.452	5	On the verge of disorder
Hamburger	0.558	0.434	0.492	5	On the verge of disorder

It can be seen from the table that more than 80% of the cities' indicators are out of coordination, and the correlation between each indicator is low. If the correlation is high, it will lead to a certain amount of information overlap, which will affect the rationality of the evaluation results. This table shows that the indicators selected in this paper can reflect the risk level of light pollution from various aspects, thus verifying that the selected indicators are widely applicable.

3. PCA-EWM Model

Principal component analysis uses the idea of dimensionality reduction to convert multiple indicators into a few comprehensive indicators, which retains most of the original variable information. Each principal component is a linear combination of the original indicators, and each one is not correlated with each other. In short, principal component analysis simplifies the complexity of the problem and captures the main contradiction. The basic idea of entropy weighting method is to determine the objective weight according to the size of the variability of indicators [10]. Therefore, the entropy weight method is a suitable method for determining the weights of the principal components.

3.1. Steps of PCA-EWM Model

1. Determine the evaluation index set. Suppose that *m* indicators are given as $X_1, X_2, \ldots, X_m$,where $X_i = \{x_1, x_2, \ldots, x_m\}$. Assume that the value after normalizing the data of each indicator is $Y_1, Y_2, \ldots, Y_m$.

2. Extraction of principle components. Assuming that there are n regions, and the index set is *X*, then the observation matrix is

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{8}$$

Suppose the extracted principle components are $y_1, y_2, \cdots, y_m, s \le m$, and the expressions of the principle component are

$$\begin{cases} y_1 = a_{11}x_{11} + a_{12}x_{12} + \cdots + a_{1m}x_{1m} \\ y_2 = a_{21}x_{21} + a_{22}x_{22} + \cdots + a_{2m}x_{2m} \\ \cdots\cdots \\ y_s = a_{s1}x_{s1} + a_{s2}x_{s2} + \cdots + a_{sm}x_{sm} \end{cases} \tag{9}$$

$y_i \left(i = 1, 2, \cdots, s\right)$ have to properties:

$$a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2 = 1, i = 1, 2, \cdots, n; \tag{10}$$

$$\operatorname{cov}\left(y_i, y_j\right) = \begin{cases} 0, i \neq j \\ \lambda_i, i = j \end{cases} \tag{11}$$

Among them, $\lambda_1, \lambda_2, \cdots, \lambda_p$ are the eigenvalues of the covariance matrix of *X*, $a_{i1}, a_{i2}, \cdots, a_{ip}$ are eigenvectors.

According to the correlation coefficient matrix of the evaluation index *X*, the eigenvalues and eigenvectors are solved. Then calculate the variance contribution rate and cumulative variance contribution rate of new variables.

The variance contribution rate is expressed as

$$\frac{\lambda_i}{\sum_{i=1}^{m}\lambda_i} = \frac{\lambda_i}{m} \tag{12}$$

The cumulative variance contribution rate id expressed as

$$\sum_{i=1}^{k}\frac{\lambda_i}{\sum_{i=1}^{m}\lambda_i} = \sum_{i=1}^{k}\frac{\lambda_i}{m} \tag{13}$$

According to the standard of cumulative variance contribution rate, the principal components are extracted.

The criterion for extracting principal components is $\sum_{i=1}^{k}\frac{\lambda_i}{m} \geq 80\%$.

3.Calculating principal component weights using entropy weight method.

First define the standardization of the indicators.

$$f_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m} Y_{ij}} \tag{14}$$

Calculate the entropy value and of each principal component.

$$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} f_{ij} \ln f_{ij} \tag{15}$$

Calculate the information utility value of each indicator.

$$d_j = 1 - e_j \tag{16}$$

Determine the entropy weight of each principal component $W_1, W_2, \ldots, W_j$.

$$W_j = \frac{d_j}{\sum_{i=1}^{p} d_j} = \frac{1-e_j}{p-\sum_{i=1}^{p} e_j} \tag{17}$$

Among them, p represents the number of new indicators.

3.2. Calculation results by PCA-EWM model

According to the principal component selection principle in the principal component analysis method, five principal components were finally selected. Their cumulative variance contribution rate is 80.25%, which basically contains the main information of the original indicators.

According to the principal component loading matrix loading matrix, Costal Aera, Industrialization Level and Electricity Consumption have larger loadings in the first principal component; Technology Level, Economic Level, Altitude, Crime Rate have larger loadings in the second principal component; Biodiversity and Population Density have larger loadings in the third principal component; Vegetation Coverage has larger loadings in the fourth principal component. CO2 Emissions has larger loadings in the fifth principal component. Therefore, we define the names of the principal components. The principal component names are shown in Table 2.

Table 2 Name of principal component

Principle Component	Name
F_1	Industrialization Development
F_2	Urban Development
F_3	Coordinated Development
F_4	Forest Resources
F_5	Atmosphere Environment

The weights of each principal component calculated by EWM are shown below.

$$\mathbf{W}=\begin{pmatrix}12.204 & 25.604 & 36.750 & 8.011 & 17.432\end{pmatrix}^T \tag{18}$$

From the formula (8), we know that Urban Development and Coordinated Development play important roles with 36.75% and 25.604%. Industrialization Development and Atmospheric Environment have medium weight, and their shares are 17.432% and 12.204%. Forest Resources have the least weight.

3.3. PCA-EWM Model of Light Pollution

Based on the above discussion, we obtained the final evaluation scores for each area for the five dimensions, which are expressed as

$$S_i=12.204X_{i1}+25.604X_{i2}+36.750X_{i3}+8.011X_{i4}+17.432X_{i5} \tag{19}$$

Calculate the scores of 31 regions by formula, the higher the score, the more serious the pollution. Figure 2 is the division of light pollution risk level based on sample data, different score ranges correspond to different risk levels. Score range from 7.855 to 43.648 is at the level of non-light pollution, with a total of 13 cities; Scores from 43.648 to 80.158 are moderate levels of light pollution, with a total of 9 cities; There are a total of 9 cities with high light pollution, with scores ranging from 80.158 to 120.393.

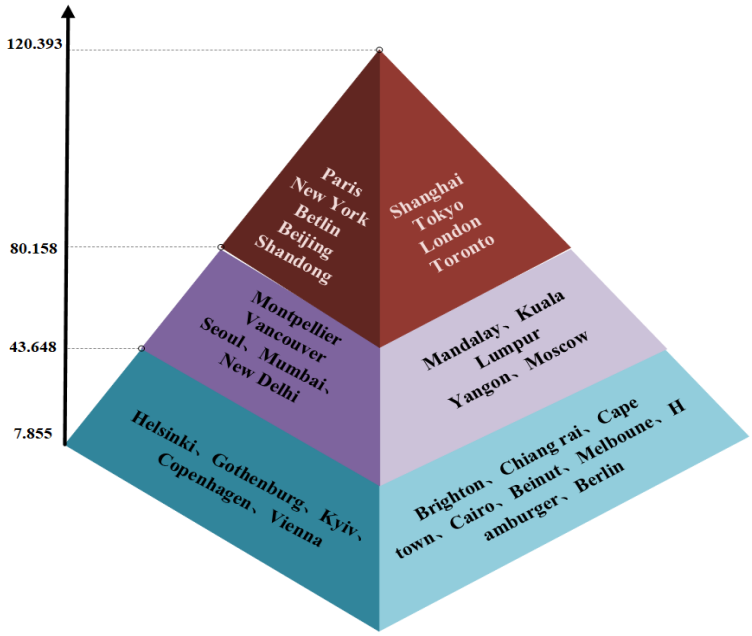


Figure 2 Risk level results

4. Light pollution risk level in four diverse types of locations

This paper found relevant data on urban community, suburban community, rural community, and protected land in Beijing. Then apply the PCA-EWM Model of Light Pollution established above to analyze the heterogeneity in light pollution risk levels.

Assume that the standardized data matrix of 12 indicators in four diverse types of locations is $\mathbf{LP}_{4\times12}$, the coefficient matrix for 12 indicators in the five principal component expressions is $\mathbf{Y}_{12\times5}$.

We obtain the light pollution level score expression of four different regions expressed as

$$\mathbf{S}=\mathbf{LP}\times\mathbf{Y}\times\mathbf{W} \tag{20}$$

The final light pollution risk level in in four diverse types of locations are shown in Table 3.

Table 3 Risk level of light pollution in each location

Location	Score	Risk Level
UD	81.512	High
CD	57.454	Moderate
ID	36.393	Non
AE	30.453	Non

The results showed that there was high light pollution in urban community, moderate light pollution in suburban community, and light pollution or non-light pollution in rural community and protected land.

5. Conclusions

In this paper, the light pollution level evaluation index system is constructed from five aspects, the PCA-EWM model is established to measure it, and three light pollution risk levels are determined according to the score: high light pollution, moderate light pollution and non-light pollution. Subsequently, this paper applies the PCA-EWM Model to calculate light pollution score in four diverse types of locations. The results show that the score of non-light pollution level varied from 7.855 to 43.648, the score of moderate light pollution level varied from 43.648 to 80.158, and the score of high light pollution varied from 80.158 to 120.393; there is high light pollution in urban community, moderate light pollution in suburban community, and light pollution or non-light pollution in rural community and protected land.

References

[1] Su Xiaoming, Hao Zhanguo, Zhang Mingyu. Study on the characteristics of light pollution in the night sky of Hohhot city [J]. Journal of Lighting Engineering, 2015, 26(04): 124-128.

[2] Yang Zhen. Research on design strategy of reflective glare prevention and control of glass curtain wall building [D]. Tianjin University, 2017.

[3] Yang Zhen, Zhang Mingyu. Analysis of influencing factors of glare reflection in glass curtain wall building [J]. Journal of Lighting Engineering, 2017, 28(03): 51-56.

[4] Mao Hequn, Wang Fan, Xu Guiru, et al. Influence of reflected light of architectural glass curtain wall on urban roads [J]. Journal of Shanghai Institute of Shipping and Transportation Science, 2019, 42(02): 74-78.

[5] Li Zhuo, Liu Gang, Yao Xin, et al. Discussion and analysis of the current situation of outdoor light environment at night in residential area——Taking Zhengzhou City as an example [J]. Science and Technology Information, 2010, No.342(22): 28-29.

[6] Niu Shengnan, Wang Lixiong, Yang Xianguo. Application of green lighting in the prevention and control of light pollution of urban overpass [J]. China Lighting Appliances,2008, No.278(05): 7-11.

[7] Wang Lixiong, Su Xiaoming, Dang Rui, et al. Research on detection and evaluation of light environment in residential area [J]. Building Science, 2013, 29(08): 19-21+33.

[8] Zhou Zutong. Light Pollution in the Environment and Its Prevention and Control Countermeasures [J]. Legal Expo, 2021, (17): 174-175.

[9] Zhang Ping, Wang Xinyu. Research on coupling coordination and dynamic relationship between new urbanization and rural revitalization: A case study of Zhejiang Province [J]. Statistics and Management, 2022, 37(11): 48-56.

[10] Wang Qiong, Lu Cong, LI Fayun, et al. A method of evaluation of river habitat quality based on principal component analysis and entropy weight method: A case study of Qinghe [J]. Ecological Science, 2017, 36(04): 185-193.