Research on the classification of light pollution risk in different regions based on K-means clustering and TOPSIS--Taking Hubei Province as an example

Haosen Dong*, Xiaoxuan Chen

School of Economics and Management, North China Electric Power University, Baoding, Hebei, China, 071003

*Corresponding author: m15603591650@163.com

Abstract. With the continuous development of the economy and technology, light pollution has become a new form of environmental pollution that affects human health and the ecosystem. This article aims to evaluate the level of local light pollution risk by using Principal Component Analysis (PCA) to identify 14 factors that affect the risk of light pollution, which are grouped into five categories. The Analytic Hierarchy Process (AHP) and entropy weight method are then applied to assign weights to each factor, and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is used to obtain the evaluation results. Our collected data from Hubei province and used the K-means clustering method to categorize the results of the light pollution risk assessment into three levels: high, average, and low.

Keywords: K-means clustering method, TOPSIS, Principal Component Analysis, Light pollution.

1. Introduction

The neon lights that give cities their vibrant colors at night also have a detrimental effect – light pollution. The excessive use of artificial lighting causes light pollution, which often emanates from nighttime illumination. Common light pollution phenomena include light intrusion, over-illumination, and light clutter, which now extend beyond urban areas to remote locations due to the widespread use of electricity. The degree of light pollution in a given area is influenced by both human factors, such as the level of urbanization, and non-human factors, such as geography and climate.

Light pollution not only adversely affects human health by disrupting sleep quality and circadian rhythms, which can cause eye diseases[1][2], but also poses a serious threat to ecological balance and biodiversity[3]. The negative consequences of light pollution are manifold. Several countries and regions have established relevant light pollution standards to mitigate its impact. Officials must therefore devise practical measures to manage the use and construction of lighting in communities, and evaluate the level of light pollution risk in a given region. A comprehensive literature review can aid in developing a light pollution risk assessment model that is applicable to diverse areas.

So far, a large number of researchers have made part of the research on the hazards of light pollution and related intervention policies. In terms of the harm of light pollution, Cinzano et al. pointed out that artificial light sources would have a negative impact on astronomical observation, ecosystem, human health and other aspects [4]. Rich and Longcore discuss the effects of artificial nighttime pollution on wildlife, plants, and ecosystems, noting the effects of light pollution on animal biological clocks, seasonal behavior, migration, plant production, flowering, and fruiting[5]. Falchi et al. discussed the effects of artificial light pollution at night on human health, the environment, and astronomical observations, showing that long-term exposure to light pollution can affect the circadian clock, sleep quality, immune system, and other health problems [6]. In terms of relevant interventions, Gaston et al. evaluated the currently available light pollution interventions and noted that effective interventions should be context-specific and demand based, with greater emphasis on human health and ecosystem needs[7]. Falchi et al. proposed a number of light pollution interventions based on scientific evidence to protect ecosystems and astronomical observations [8]. Aubé explores the impact of light pollution on ecosystems and how to design more sustainable lighting systems to mitigate its

impact [9]. Luginbuhl et al. provide referential experience by listing some successful light pollution interventions [10].

To sum up, there has been a lot of research on the harm of light pollution, but there is still a lack of research on how to comprehensively evaluate light pollution in an area and how to quantify the degree of light pollution. To build a light pollution risk level assessment model that can be widely applied to various types of areas. We established a model that integrates the factors of resident health, economy, environment and urban planning, called HALOEN model.

2. Establishment of model

2.1. Selection of index

When we evaluate the level of light pollution risk about an area, a lot of factors should be considered. There are some complicated relationships about these factors which may enhance the difficulty of analyzing the problem. In view of this, we collected 37 indicators for light pollution evaluation by reviewing the literature. Then, the 37 selected indicators were reduced by the PCA method, and 14 indicators were selected according to the principle of PCA to constitute the basic elements of the index system. The results are shown in the **Table 1**.

Table 1. Total variance interpretation

Total variance interpretation						
Composition	Initial eigenvalue			Extract the sum of loads squared		
	Total	Percentage of variance	Accumulation %	Total	Percentage of variance	Accumulation %
1	9.169	15.282	15.282	9.169	15.282	15.282
2	7.190	11.984	27.266	7.190	11.984	27.266
3	6.042	10.070	37.336	6.042	10.070	37.336
4	5.118	8.530	45.866	5.118	8.530	45.866
5	4.529	7.548	53.414	4.529	7.548	53.414
6	4.446	7.411	60.825	4.446	7.411	60.825
7	4.234	7.057	67.882	4.234	7.057	67.882
8	3.443	5.738	73.620	3.443	5.738	73.620
9	2.997	4.995	78.615	2.997	4.995	78.615
10	2.639	4.398	83.013	2.639	4.398	83.013
11	2.297	3.829	86.842	2.297	3.829	86.842
12	2.041	3.401	90.243	2.041	3.401	90.243
13	1.793	2.988	93.231	1.793	2.988	93.231
14	1.697	2.828	96.059	1.697	2.828	96.059

Based on the 14 indicators selected, they are divided into five areas that do not affect each other. These five areas are economic energy consumption, health status of residents, natural environment, urban planning and optical index. The index system is shown in the **Figure 1**.

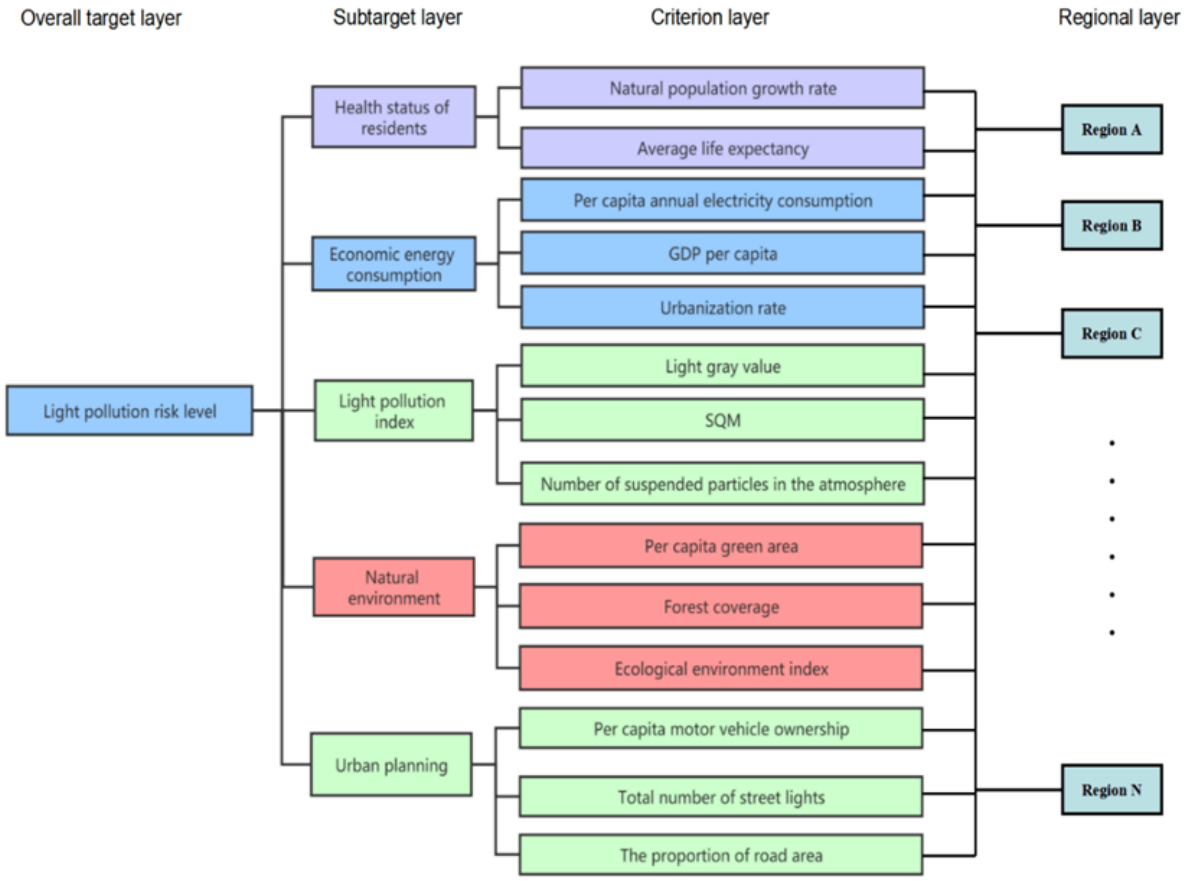


Figure 1. Overview of risk level assessment indicators of light pollution

2.2. Determination of weights

As for the determination of index weight, the evaluation problem based on the degree of light pollution risk can neither ignore the subjective preference of decision-makers nor ignore the data information of the evaluation object itself. Therefore, AHP+ EWM in the combined weight method can effectively avoid the arbitrariness of the subjective method and the data sensitivity of the objective method, making the evaluation results more objective.

2.2.1 EWM

The entropy weight method is an objective weighting method. It measures the degree of the chaos of data by calculating the information entropy of each indicator. The larger the information entropy is, the less information the indicator contains, and the smaller the corresponding weight should be. The specific steps are as follows:

1. Standardized processing of indicator data. Suppose there are a total of m indicators:$X_1, X_2 \dots \dots X_m$the j-th sample data are respectively: $X_j = \{x_1, x_2 \dots \dots x_n\}$the data after standardized processing is $Y_1, Y_2, \dots \dots, Y_k$.Among them:

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \tag{1}$$

2. Calculate the ratio of each index under each scheme. The formula is as follows:

$$p_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}} (i = 1,2, \dots \dots, n; j = 1,2, \dots \dots, m) \tag{2}$$

3. Solve the information entropy of each index

According to the definition of information entropy in information theory, the information entropy of a group of data is:

$$E_j = -ln(n)^{-1} \sum_{i=1}^{n} p_{ij}\, ln\, p_{ij} \left(\text{If } p_{ij} = 0, E_j = 0\right) \tag{3}$$

Determine the weight of each indicator

According to the calculation formula of information entropy, the information entropy of each index can be calculated as$E_1, E_2,\ldots,E_m$. The following formula can obtain the weight of each index W_2.

$$\omega_j = \frac{1 - E_j}{m - \sum E_j} \tag{4}$$

2.2.2 Combined weights

For the processing of combined weights, due to the strong subjectivity of the sub-target layer, we use the AHP method for determining the weights of the sub-target layer, and for the criterion layer, we adopt the EWM method for the weight measurement of the sub-target layer because the data is easy to query. The resulting combined weight formula is as follows:

$$W_3 = \alpha W_1 + (1 - \alpha) W_2$$
$$(\alpha\ is\ the\ preference\ coefficient, usually\ 0.5) \tag{5}$$

2.3. Evaluation methods

For the current evaluation of the degree of light pollution, people generally have a demarcation line, can clearly recognize where the degree of light pollution is the most serious area, where the degree of light pollution is not too serious. As described above, TOPSIS method is adopted to evaluate the degree of light pollution in different places.

TOPSIS is a method to obtain the relative merits and demerits of each evaluation scheme by comparing the two schemes with the information of the original data, and finally to get the "excellent" evaluation scheme by ranking it.

2.4. Cluster analysis

In order to accurately calculate the light pollution risk level of a region rather than the relative value, and make the evaluation results have practical significance, we carried out K-means clustering analysis on the comprehensive evaluation results of the light pollution risk analysis of a region obtained above, and calculated the critical value of the classification results, so as to obtain the light pollution risk classification set belonging to the risk level of each region.

The purpose of the k-means clustering algorithm is to divide n observed values into K categories and minimize the sum of distances between them by constantly selecting the points that are closest to the mean of the sample points

$$D(\mu_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{6}$$

Where x_i is an observation in the k-th cluster C_k

3. Results

3.1. The establishment of simulation model

As we all know, Shennongjia Forest area is one of the regions with the lowest degree of light pollution in the world. Under the condition of ensuring the consistency of geographical and climatic

factors as much as possible, we choose Hubei Province as an example to test the usefulness and accuracy of the established model.

3.2. Analysis of results

The collected data about relevant indicators in Hubei Province are substituted into the model, and the results are shown in the **Figure 2**.

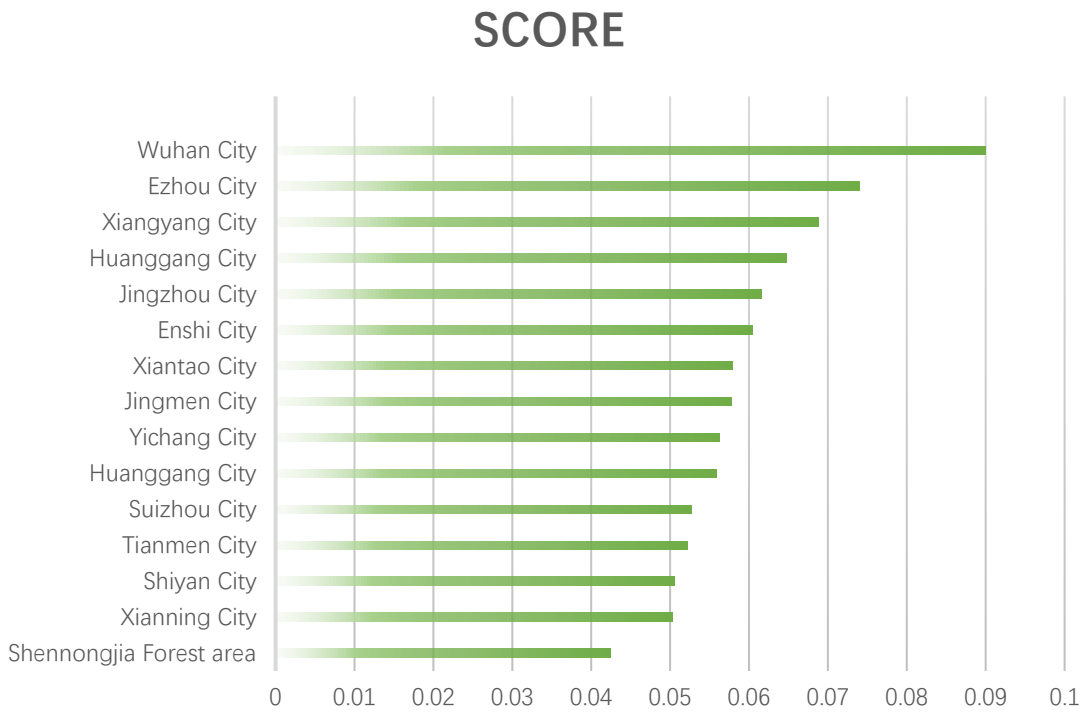


Figure 2. Map of light pollution risk level of cities in Hubei Province

3.3. Cluster analysis

Based on the K-means cluster analysis of the above results, the light pollution degree of each city can be divided into high pollution degree, medium pollution degree and low pollution degree. The resulting division criteria is shown in the **Table 2**.

Table 2. Risk classification table

hierarchy	Scoring range
High levels of light pollution risk	(0,0.053]
Moderate levels of light pollution risk	(0.053,0.07]
low levels of light pollution risk	(0.07,0.11]

The following are the results of the classification for Hubei Province, with red indicating a low risk of light pollution, blue indicating a moderate risk, and yellow indicating a high risk. The results are shown in the **Figure 3**.

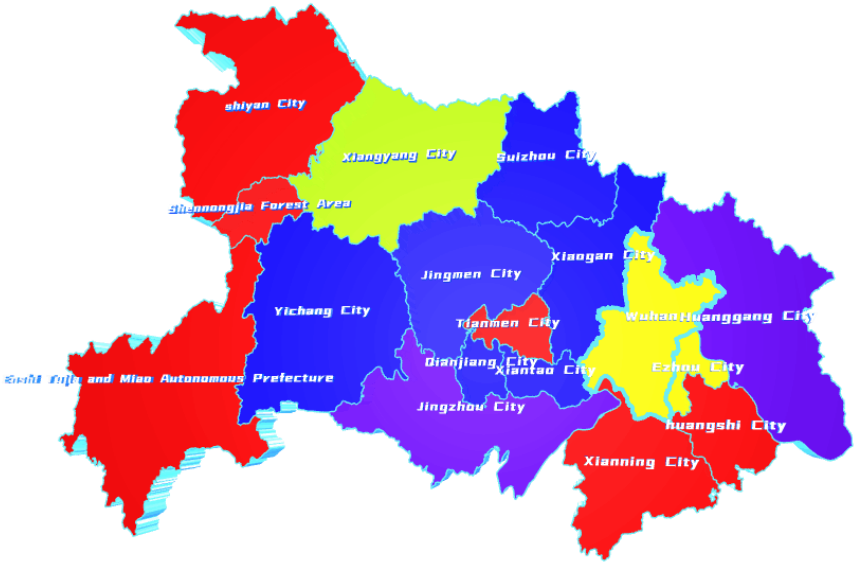


Figure 3. Classification map of light pollution risk level of cities in Hubei Province

To verify the accuracy of our model, we tested our model with night light data collected by DMSP/OLS. The results are shown in the **Figure 4**.

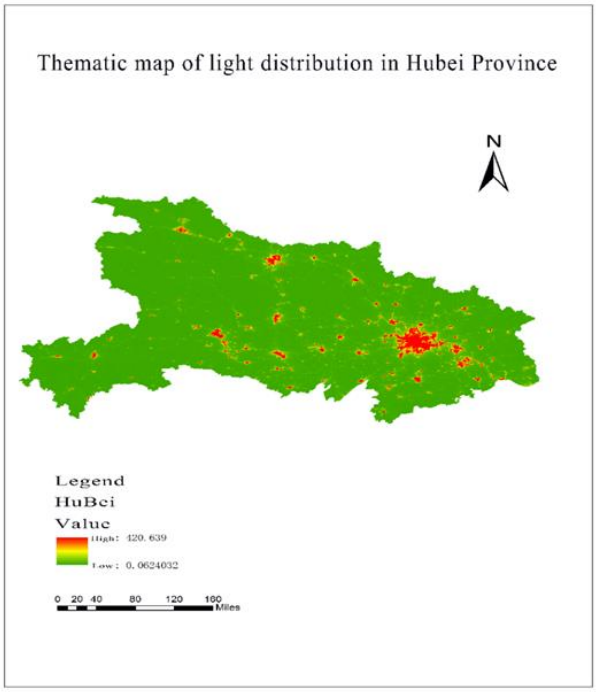


Figure 4. Thematic map of light distribution in Hubei Province

According to the night light data, the most heavily polluted areas in Hubei Province are Wuhan City and Xiangyang City, but some of the surrounding areas do not have serious light pollution, which is similar to the classification map obtained by our model, indicating the accuracy of our model

4. Conclusions

In order to accurately evaluate the degree of light pollution in a certain area and provide preventive measures, this paper divides 14 factors affecting the risk of light pollution into five dimensions: economic and energy consumption, residents' health status, natural environment, urban planning and optical index through principal component analysis (PCA). Then, analytical hierarchy analysis (AHP)

and entropy weight method were applied to assign values to each factor, and the ideal solution similar order preference technique (TOPSIS) was used to obtain the evaluation results. Finally, K-means clustering method was used to divide the assessment results of light pollution risk into high, average and low levels, which were used as the criteria for judging the degree of light pollution. Taking Hubei Province as an example, the collected data were fitted with the model, and the light pollution degree was similar to that obtained from the luminous data, and the model fitting effect was good.

References

[1] Svechkina, A., Portnov, B.A. & Trop, T. The impact of artificial light at night on human and ecosystem health: a systematic literature review. Landscape Ecol 35, 1725–1742 (2020).

[2] Hussein, A.A.A., Bloem, E., Fodor, I. et al. Slowly seeing the light: an integrative review on ecological light pollution as a potential threat for mollusks. Environ Sci Pollut Res 28, 5036–5048 (2021).

[3] Sanders D, Kehoe R, Cruse D, van Veen FJF, Gaston KJ. Low Levels of Artificial Light at Night Strengthen Top-Down Control in Insect Food Web. Curr Biol. 2018;28(15):2474-8 e3.

[4] Cinzano, P., Falchi, F., & Elvidge, C. D. (2001). The first world atlas of the artificial night sky brightness. Monthly Notices of the Royal Astronomical Society, 328(3), 689-707.

[5] Rich, C., & Longcore, T. (2006). Ecological consequences of artificial night lighting. Island Press.

[6] Falchi, F., Cinzano, P., Elvidge, C. D., Keith, D. M., & Haim, A. (2011). Limiting the impact of light pollution on human health, environment, and stellar visibility. Journal of Environmental Management, 92(10), 2714-2722.

[7] Gaston, K. J., Bennie, J., Davies, T. W., & Hopkins, J. (2013). The ecological impacts of nighttime light pollution: a mechanistic appraisal. Biological Reviews, 88(4), 912-927.

[8] Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., ... & Furgoni, R. (2016). The new world atlas of artificial night sky brightness. Science Advances, 2(6), e1600377.

[9] Aubé, M. (2015). Physical behaviour of anthropogenic light propagation into the nocturnal environment. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1667), 20140117.

[10] Luginbuhl, C. B., Walker, C. E., & Wainscoat, R. J. (2009). Light pollution reduction: A catalogue of successes. Journal of the British Astronomical Association, 119(2), 74-82.