Assessment and mitigation of light pollution

Botao Liu*, Huiru Yang, Jia Li

School of Economics and Management, Beijing Jiaotong University, 264401

*Corresponding author: 22711052@bjtu.edu.cn

Abstract. With the improvement of human development level, excessive optical radiation has caused adverse effects on human living and production environment, namely, light pollution. In this paper, a system for assessing the degree of light pollution is studied and methods for more effective intervention of light pollution are proposed. On the basis of collecting data related to light pollution, we first preprocessed and extracted information related to 14 indicators, including brightness, light range, population and distance from the sea, from five dimensions of light intensity, social development, human activities, biodiversity and geographical climate, so as to quantitatively analyze the degree of light pollution. Then combined with the analytic hierarchy process (AHP) and entropy weight method (EWM) to calculate the weight of each index, and at last, the sensitivity analysis was used to verify its robustness. Among them, the building distribution density, brightness and distance from the sea have the greatest weight. At the same time, K-means clustering algorithm was used to calculate the standard, and finally an evaluation system was calculated to measure the severity of light pollution.

Keywords: AHP, EWM, K-means, GM (1,1), TSF, CPM.

1. Introduction

With more than half of the world's population living in urbanized areas, light pollution is a growing problem, and the current increase in light pollution is rapid and widespread. Outdoor lighting is considered necessary for production in modern society. However, the widespread use of artificial light has greatly changed the natural pattern of darkness, affected people's circadian rhythm, and caused adverse effects on people's body, such as insomnia, depression and other problems; Artificial light can also interfere with migration and animal reproduction, and unnecessary light pollution can also contribute to air pollution.

In the view of those problems, we can reduce the impact of light pollution by adopting intervention strategies such as reducing the intensity and duration of light exposure, thus reducing the risk level of urban or rural areas. In addition, through these intervention strategies, a balance can be struck between the actual needs of social development and green and low-carbon development.

1.1. Restatement of the Problem

In this problem, it is very important to propose an effective method to evaluate the degree of light pollution in an area. First of all, we need an index system to evaluate the degree of light pollution. Four different types of regions are evaluated through the index system. Then we have to propose feasible intervention strategies to curb light pollution and use indicators to calculate the most effective intervention strategies. Finally, we publicize effective intervention strategies for light pollution in an area through posters.

2. Data Analysis

2.1. Indicator Selection

In order to establish an indicator system relatively completely, it is highly necessary to search for sufficient data. In this part, to measure the risk level of light pollution, we first defined the assessment index of light pollution risk level. Then, in the next step, we checked a large number of authoritative websites, and after screening, we initially selected 14 three-level indicators from five aspects, and

integrated them into the light pollution risk level indicators using the entropy weight method (EWM) and the analytic hierarchy process (AHP), thus establishing a three-level evaluation system.

Then, we use the K-mean clustering method to divide the risk level of light pollution into three categories: excellent good and normal, forming the evaluation criteria.

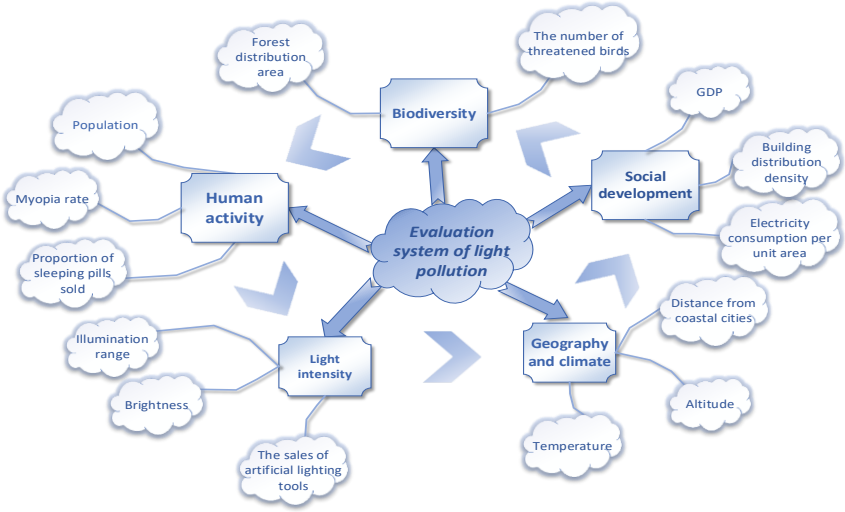


Figure 1: Three-level index chart

2.2. Data Description

2.2.1 Data Collection

The first step in establishing a good indicator system is to collect enough data.

In the process of finding data, we first used the website of the World Bank, and found the corresponding data of the three indicators of threatened birds, namely, the number of bird species, population, forest area, per capita power consumption, and per capita GDP, which are threatened by the impact of light pollution.

After continuing to search for other indicator data, we chose the Lancaster University Library for data screening, and successfully found out the lighting market revenue, that is, the sales of lamps and other lighting tools, the temperature anomaly, that is, the number of times that the temperature was considered to be higher or lower than the previous temperature, the proportion of anxiety or depression, and the global proportion of the sales of sleeping pills in various countries, The number of high-rise buildings in each country. Finally, we use the light pollution map tool to find out the brightness, light range, altitude and distance from the sea of the corresponding area.

2.2.2. Data normalization

We first completed the identification of missing values, and counted the missing values and their proportions.

Because the dimensions of the data are different, we cannot directly compare the data. In order to eliminate the adverse effects brought by unfamiliar sample data and normalize the data, we tried to convert all data into 0 to 1 for the convenience of the following calculation and comparison.

We can divide the index into four types: benefit index, cost index, intermediate index and interval index. We generally believe that the larger the benefit index is, the better it will be, while the opposite is true for the cost index. The value of the intermediate exponent cannot be too large or too small. It needs to have a specific value. Interval exponents need to fall within a certain range.

For benefit-type index:

$$z_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}, i = 1, \cdots, n; j = 1, \cdots, m$$

For cost-type index:

$$z_{ij} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}}, i = 1, \cdots, n; j = 1, \cdots, m$$

For intermediate-type index:

$$M_j = \max\{|x_{ij} - x_{j-\text{ best}}|\}, \quad z_{ij} = 1 - \frac{|x_{ij} - x_{j-best}|}{M_j}$$

For interval-type index:

$$M_j = \max\{a - \min\{x_{ij}\}, \max\{x_{ij}\} - b\}, \quad z_{ij} = \begin{cases} 1 - \frac{a - x_{ij}}{M_j}, & x_{ij} < a \\ 1, & a \le x_{ij} \le b \\ 1 - \frac{x_{ij} - b}{M_j}, & x_{ij} > b \end{cases}$$

Where x_{ij} denote the j-th index of the i-th object, z_{ij} is the normalized value of x_{ij}.

We divided the indicators into different types, as shown below:

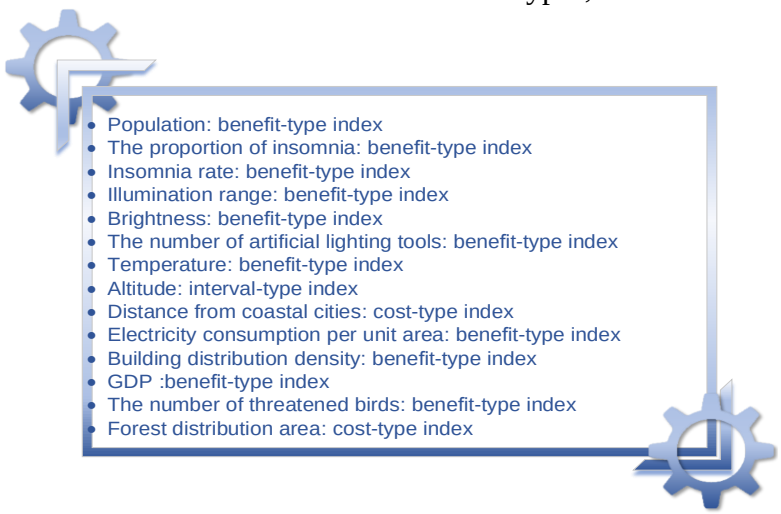


Figure 2: Index classification chart

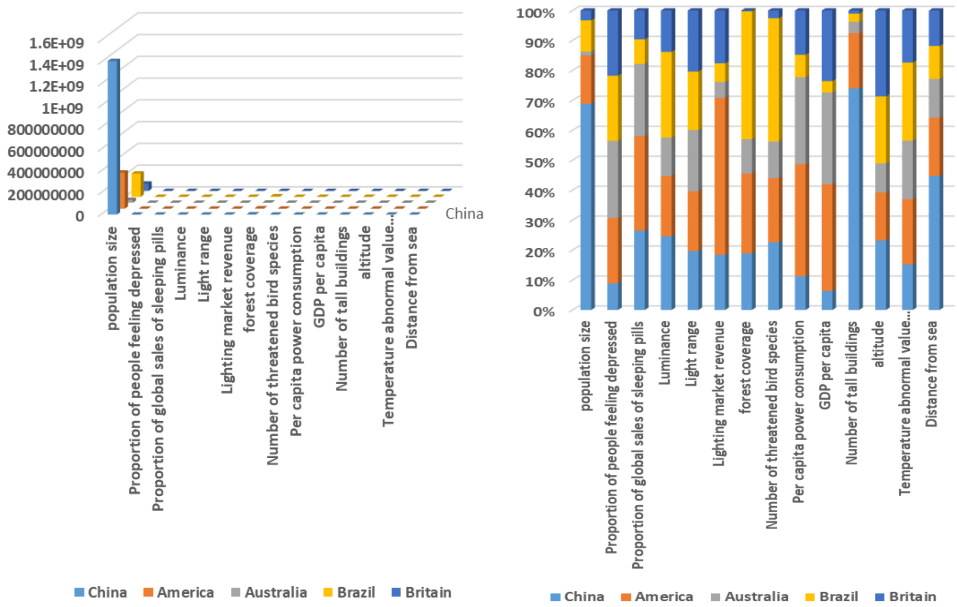


Figure 3: Data comparison graph before and after normalization

We have normalized the original data for the following reasons. First, normalization can eliminate the influence of dimension, that is, unit. In all the collected data, the meaning of each data is different, and the unit used is also different, which leads to the different size of data and directly affects the effect of visualization.

In order to eliminate this negative impact, we have normalized the original data of different units and mapped the data to the range of 0 to 1 for processing, which is more convenient and clear. Secondly, normalization can improve the accuracy and make each feature make the same contribution to the results. At the same time, normalization also eliminated the dimension effect and mechanism effect to a certain extent. Our results are as follows:

Table 1: Forward matrix

country	*population*	*Myopia rate*	*Sales of sleeping pills*
China	**0.9619**	**0.7354**	**0.5370**
America	**0.2260**	**0.2965**	**0.6376**
Australia	**0.0175**	**0.0119**	**0.4892**
Brazil	**0.1460**	**0.5976**	**0.1660**
Britain	**0.0459**	**0.1186**	**0.1953**
country	*Number of recipient birds*	*Per capita power consumption*	*GDP*
China	**0.4253**	**0.2188**	**0.1199**
America	**0.4032**	**0.7279**	**0.6708**
Australia	**0.2304**	**0.5642**	**0.5772**
Brazil	**0.7753**	**0.1463**	**0.0717**
Britain	**0.0487**	**0.2874**	**0.4442**
Light intensity	**Illumination range**	**Lighting market sales**	**Forest cover area**
0.5261	**0.4396**	**0.3127**	**0.3967**
0.4308	**0.4459**	**0.8906**	**0.2679**
0.2736	**0.4578**	**0.0899**	**0.5199**
0.6126	**0.4366**	**0.1060**	**0.0000**
0.2957	**0.4558**	**0.2997**	**0.7075**
Building distribution density	**altitude**	**temperature**	**Distance to sea**
0.9684	**0.4027**	**0.3345**	**0.0000**
0.2416	**0.5576**	**0.4778**	**0.4066**
0.0494	**0.5576**	**0.4300**	**0.5108**
0.0357	**0.4647**	**0.5734**	**0.5421**
0.0137	**0.0000**	**0.3823**	**0.5291**

2.3. Calculation of Index Weight

In order to obtain comprehensive evaluation indicators, we need to calculate the weight of X indicators. Next, we will use analytic hierarchy process (AHP) and entropy weight method (EWM) to judge the importance of indicators.

Weighted model based on analytic Hierarchy process

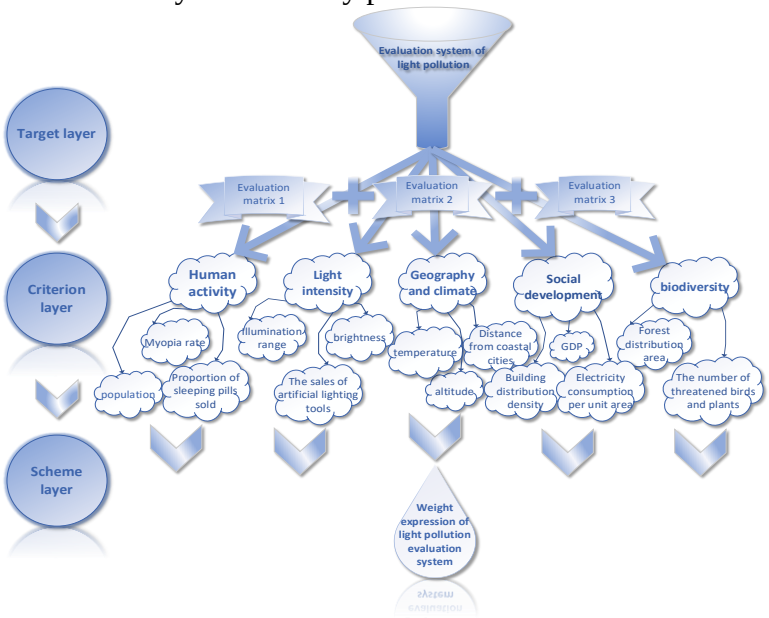


Figure 4: Data comparison graph before and after normalization

We get the importance correlation table between objects by asking the experts questions. Improve the accuracy through pairwise comparison form and digital form. Through the investigation, we give the following matrix: (from left to right, respectively:

Matrix 1: $\begin{bmatrix} 1 & 1/8 & 1/4 & 1/2 & 2 \\ 8 & 1 & 2 & 3 & 8 \\ 4 & 1/2 & 1 & 2 & 4 \\ 2 & 1/3 & 1/2 & 1 & 3 \\ 1/2 & 1/8 & 1/4 & 1/3 & 1 \end{bmatrix}$ **Matrix 2:** $\begin{bmatrix} 1 & 1/7 & 1/2 & 1/2 & 1 \\ 7 & 1 & 4 & 4 & 7 \\ 2 & 1/4 & 1 & 1 & 2 \\ 2 & 1/4 & 1 & 1 & 2 \\ 1 & 1/7 & 1/2 & 1/2 & 1 \end{bmatrix}$ **Matrix 3:** $\begin{bmatrix} 1 & 1/8 & 1/6 & 1 & 2 \\ 8 & 1 & 3 & 5 & 7 \\ 6 & 1/3 & 1 & 3 & 5 \\ 1 & 1/5 & 1/3 & 1 & 2 \\ 1/2 & 1/7 & 1/5 & 1/2 & 1 \end{bmatrix}$

Consistency test was carried out on the matrix:

Step1: Calculate consistency index:

$$CI = \frac{\lambda \max - n_l}{n-1}$$

Step2: Find the corresponding average random consistency index RI:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54	1.56	1.58	1.59

Step3: Calculate the consistency ratio CR:

$$CR = \frac{CI}{RI}$$

Conclusion:

If CR < 0.1, the consistency of the judgment matrix is acceptable; otherwise, the judgment matrix needs to be corrected. Adjust it to the consistent matrix as far as possible.

Our CR computer for the secondary index is shown below:

Matrix 1:CR=0.012 Matrix 2: CR= 0.0006 Matrix 3: CR= 0.035

All the three matrices have passed the consistency test, and the evaluation matrix of the three-level indicators should also undergo the following test. After passing the consistency test, we need to calculate the weight of each indicator.

Idea1: The weights of the three columns of data are calculated respectively, and the arithmetic mean roots are solved.

$$w_i = \frac{1}{n}\sum_{j=1}^{n}\frac{a_{ij}}{\sum_{k=1}^{n} a_{k_j}}$$

Idea2: Geometric mean method:

Step1: Multiply the elements of the judgment matrix by rows to get a new column vector

Step2: Raise the new vector to the power N

Step3: Normalize the column vector to get the weight vector

$$W_i = \frac{\left(\prod_{j=1}^{\eta} a_{ij}\right)^{\frac{1}{n}}}{\sum_{k=1}^{n}\left(\prod_{j=1}^{n} a_{kJ}\right)^{\frac{1}{n}}}$$

Idea3: Feature root:

Step1: Find the maximum eigenvalue of the judgment matrix and its eigenvector

Step2: The weight can be obtained by unitization of the obtained feature vector

In order to improve the robustness of the calculated weights, we choose to average the three to improve the robustness.

$$W_1 = (0.0735,0.4772,0.2530,0.1451,0.0513)^T$$
$$W_2 = (0.0753,0.5561,0.1466,0.1466,0.0753)^T$$
$$W_3 = (0.0730,0.5184,0.2681,0.0898,0.0507)^T$$

The result of average value calculation is as follows:

$$W_i = (0.0739,0.5172,0.2226,0.1272,0.0591)^T$$

The weights of other three-level indicators can also be calculated according to the above evaluation matrix that passes the consistency test.

We can get the weight results as follows:

Table 2: Second-class indicator and primary indicator and weight

Second-class indicator	Primary indicator and weight
LI(Light intensity): 51.7%	Bri. (57%)
	IR' (35%)
	TSOALT (9%)
GAC(Geography and climate): 22.3%	DFCC (57%)
	Alt. (29%)
	Tep. (14%)
SD(Social development): 12.7%	GDP (57%)
	BDD (29%)
	ECPUA (14%)
Bi(Biodiversity): 5.9%	FDA (67%)
	TNOTB (33%)
HA(Human activities): 7.4%	POP. (73%)
	MR (18%)
	IR (9%)

2.3.1. Weighted model based on Entropy Weight Method (EWM)

Here we all understand that AHP is associated with subjectivity because of the score matrix. In order to reduce the influence of this property, we use the entropy weight method (EWM) to help calculate the weight. EWM is a method to determine quantitative weights based on the variability of indicators. Generally speaking, the smaller the information entropy of the index, the greater the degree of change of the index value. The more weight it provides. Conversely, the greater the information entropy of the index, the smaller the change degree of the index value, the less information provided, and the smaller its weight. The specific algorithm flow is shown in the figure below.

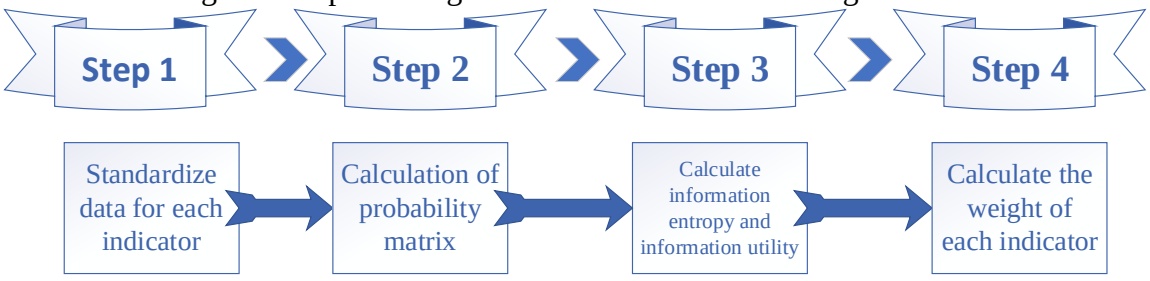


Figure 5: Data comparison graph before and after normalization

Since the index data has been forward and normalized in the data processing part, we will omit the first step and directly conduct subsequent operations on the processed index data. And we get an initialized matrix, as shown in the figure below:

$$X_{ij} = \left(\begin{array}{cc|cc} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{array}\right)$$

Step 1: Calculate the probability matrix P, ensure that the column sum is one, and all objects under each index correspond to the probability sum of 1. The calculation formula of each element P in P is as follows:

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} (j = 1, 2, ..., m)$$

Step2: The information entropy of each indicator is represented by E, and the information utility value is represented by D. The calculation formula is as follows:

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} P_{ij} \ln(P_{ij}) (j = 1, 2, ..., m)$$

$$d_j = 1 - e_j$$

Step3: Normalize the calculated information utility value and weights of each index to facilitate subsequent work. The calculation formula is as follows:

$$W_{ji} = \frac{d_j}{\sum_{jj=1}^{m} d_j}$$

The weight results of EWM obtained by calculation are as follows:

Table 3: EWM weight

17.7%	**9.8%**	**3.2%**
population	***Myopia rate***	***Sales of sleeping pills***
6.1%	**4.3%**	**6.5%**
Number of recipient birds	***Per capita power consumption***	***GDP***
1.3%	***0.0%***	***8.8%***
Light intensity	***Illumination range***	***Lighting market sales***
21.8%	***6.3%***	***0.5%***
Building distribution density	***altitude***	***temperature***
7.6%		
Forest cover area		
6.2%		
Distance to sea		

Finally, the final weight obtained is calculated as follows:

$$W_i = \theta_1 \cdot W_{ai} + \theta_2 \cdot W_{ei}$$

Here, Wai is used to represent the weight of the 1-th index calculated by AHP, and Wbi is the weight calculated by EWM. Let's assume θ1 = 0.3 and θ2 = 0.7. After calculating the weighted average of the above calculation results, the final weight is obtained. The results are as follows:

W_i= (0.1401,0.0725,0.0497,0.0974,0.0543,0.0755,0.0654,0.0485,0.0356,0.0668,0.1577,0.0634,0.0126,0.0815)^T

3. Metric of light pollution

After selecting indicators and calculating appropriate weights, we need to set appropriate criteria for light pollution levels. In this section, we use the k-means algorithm to establish a suitable standard. According to the K-means algorithm, we get the relevant criteria. According to the collected data, we first cluster each indicator and calculate 3 clustering centers. Then, taking the mean value of index center as the standard boundary, the light pollution degree was divided into normal, good and excellent. The pseudo-code of clustering is shown below:

Table 4: K-means method step diagram

Algorithm 1: K-means
Input: K: Number of clustering centers
Output: X: Cut-off points and different indicators
1: A set of data is given and k clustering centers are given. 2: Calculate the distance of each data to the cluster center, and the closest one is classified as one cluster. 3: According to the obtained cluster, the cluster center is recalculated. 4: Iterate until you reach the appropriate value.
end while
return X
end function

As shown in the figure below, the classification standards of the five combination indicators are different. Take HA as an example. When the score is less than 0.5912, it indicates that the HA is in a good level. Scores between 0 and 0.2333 indicate low levels of light pollution. When the score is greater than 0.5912, light pollution Chengdu is considered to be more serious. In order to make the result more vivid, we use the accumulation map and radar map to complete the visualization processing.

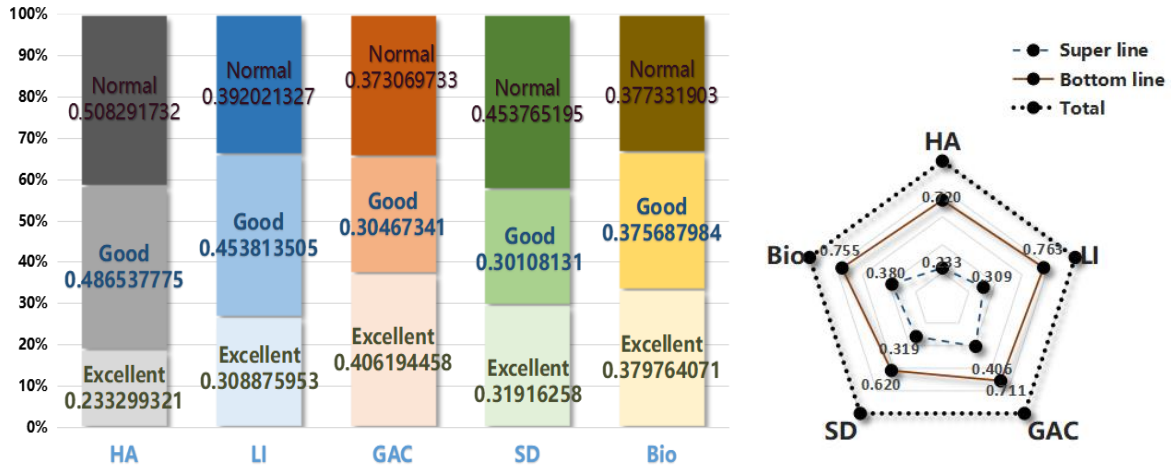


Figure 6: Evaluation system diagram

4. Conclusion

To sum up, we first established a light pollution risk assessment system to characterize the light pollution risk level of a region. After consulting the data, we should use the combination of AHP and entropy weight method to determine the weight of each index, and make a sensitivity analysis. After that, we use the k-means clustering method to determine the three light pollution levels of excellent, good and normal. Finally, we multiply the weight with the data to obtain the light pollution risk index, and determine the light pollution risk level of a region according to the size of the index. The higher the index, the more serious the light pollution. In addition, reducing light intensity can be a good way to reduce light pollution.

References

[1] https://data.stats.gov.cn/
[2] http://www.city-data.com/
[3] http://hao.199it.com/
[4] http://www.nmc.cn/publish/satellite/fy2.htm
[5] http://www.nsmc.org.cn/nsmc/cn/image/index.html
[6] https://www.lightpollutionmap.info/
[7] http://satellite.nsmc.org.cn/PortalSite/Default.aspx
[8] https://data.worldbank.org.cn/
[9] www.xihe-energy.com