Research of Light Pollution Levels Based on Principal Component Analysis and Topsis

Zhicheng Guo [1, *], Yao Yao [2], Shihao Zhang [2]

[1] School of Science, Chang'an University, Xi'an, Shaanxi, China 710018

[2] Department of Engineering Machinery, Chang'an University, Xi'an, Shaanxi, China 710018

* Corresponding Author Email: g1207xc@163.com

Abstract. With the development of society, a large amount of light has a bad impact on people's lives. The assessment and governance of light pollution levels also brought difficulties to government management in various regions. An evaluation model of light pollution level under the influence of main factors is constructed based on EWM-TOPSIS scoring method, combined with the advantages of principal component analysis in dealing with complex factors. This paper divides the cities into four areas based on hierarchical cluster analysis according to population density: protected land location, rural community, suburban community and urban community. Each region takes a city as an example group to calculate in our evaluation model, and the light pollution level score is as follows: Greater Khingan Mountains region: 3.7, Panzhihua City: 4.09, Suzhou City: 6.2, New York: 9.97. The results indicate that the areas with high population density have serious light pollution, which is consistent with the actual situation and verifies the rationality of the model.

Keywords: Light Pollution, PCA, EWM-TOPSIS, Hierarchical Cluster Analysis.

1. Introduction

The light pollution is so widely understood today to describe the degradation of the view of the night sky and the human experience of the night [1]. And the light pollution hazards can cause accelerated aging and death of normal cells in people who are active and work for long periods of time and may induce various diseases [2][3]. Ecological light pollution includes direct glare, chronically increased illumination, and temporary, unexpected fluctuations in lighting [4]. Therefore, a method to determine the level of light pollution is needed to give an appropriate light source use strategy.

Mu Haowei, et al. developed a framework to evaluate the light pollution status in global PAs, using the global NTL time series data [5]. First, they classified global PAs (30,624) into three pollution categories: non-polluted (5974), continuously polluted (8141), and discontinuously polluted (16,509), according to the time of occurrence of lit pixels in/around PAs from 1992 to 2018. Then, they explored the NTL intensity (e.g., digital numbers) and its trend in those polluted PAs and identified those hotspots of PAs at the global scale with consideration of global urbanization. To evaluate light pollution, O. Rabaza, et al. compared the sky background flux to that of many stars of known brightness [6]. Ramon Alamús, et al. demonstrated the possibility of using a hyperspectral imager (routinely used in airborne campaigns) to measure the ground-level spectral radiance of the urban nightscape and to retrieve several magnitudes of interest for light pollution studies [7]. Yali Katz, et al. estimated urban light pollution (mainly in open areas within the city) by combining data from high spatial resolution nighttime imagery acquired by EROS-B (Levin et al., 2014), and from ground measurements using the photosensitive "Sky Quality Meter" (SQM) produced by Unihedron (Cinzano, 2005) [8]. However, so far, there is no methodology nor a quantitative procedure to determine an optimal light pollution monitoring network [9].

This paper studied the main influencing factors of light pollution, including the radiation brightness value of the image response, the scale index of local urbanization, the relative geographical location evaluation index and the air particle pollution index. The relationship between these factors and light pollution level is analyzed. Secondly, based on hierarchical clustering, four regions are divided into protected land location, rural community, suburban community and urban community

by population density. The corresponding light pollution level is calculated by taking the cities as an example from each region to compare whether the level is consistent with the actual situation.

2. The basic fundamental of model

2.1. Index selection of influencing factors for light pollution

This paper quantifies different locations based on the background of the topic. The light pollution risk level of a region can be evaluated by the score. First of all, we need to select indicators. Through literature search, we find many indicators that directly cause light pollution or indirectly cause light pollution. This paper will analyze these indicators separately and discuss direct and indirect factors separately.

2.1.1 Selection of direct factor index

Among many factors of light pollution, the main direct influencing factor is light radiation intensity [10], Artificially lit surface of Earth increasing in radiance and extent [11], and the artificial light source not only brings light to people, but also brings light pollution at night, luminous remote sensing image can intuitively reflect the degree of light pollution in each city. Therefore, this paper chooses the radiation brightness value of the image response as the evaluation index. This paper use Arcmap to read the gray value *DN* [10] of the images of each city, and then convert it into the radiation brightness value *L* [12], which is the sum of the radiated energy of a certain area. The formula for calculating the radiation brightness value is as follows:

$$L = DN^{\frac{3}{2}} \times 10^2 \tag{1}$$

2.1.2 Indirect factor index selection based on PCA.

There are many indirect factors that cause light pollution. This paper select six main indicators that are closely related: population density, per capita disposable income, visibility [13], cloud cover[14], vegetation coverage[15], and pollution index as figure 1. Hothis paperver, too many variables greatly increased the difficulty of our analysis, and there is a great correlation betthis paperen some indicators. Therefore, PCA is used to reduce dimension and remove noise, and the six indicators with indirect influence are extracted into several most representative and comprehensive indicators, which have little correlation.

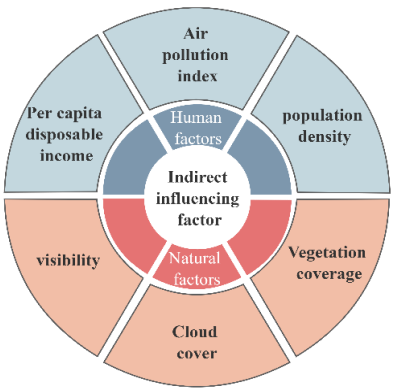


Figure 1. Indirect influencing factor of light pollution

This paper finds the index data of nearly 60 cities such as New York, Chicago and Shanghai as sample data, and set 6 indicators for PCA dimension reduction. The process of dimension reduction is as follows:

1. Compute standardized data X_{ij}:

$$\bar{x}_j = \frac{\sum_{i=1}^{60} x_{ij}}{60} \tag{2}$$

$$S_j = \sqrt{\frac{\sum_{i=1}^{6}\left(x_{ij} - \bar{x}_j^2\right)}{6}} \tag{3}$$

$$\bar{X}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \tag{4}$$

Where x_{ij} (i=1, 2…,60 j=1, 2…,6), i is the i*th* sample and j is the j*th* index data.

2. Calculate the correlation coefficient matrix *R*:

$$r_{ij} = \frac{1}{5}\sum_{k=1}^{6}(X_{ki} - \bar{X}_\iota)\left(X_{kj} - \bar{X}_j\right) = \frac{1}{5}\sum_{k=1}^{6} X_{ki}X_{kj} \tag{5}$$

$$R = \left(r_{ij}\right)_{60\times 6} \tag{6}$$

3. Calculate the eigenvalue μ and the eigenvector α for *R*:

$$tr(R) = \sum_{k=1}^{6}\mu_k = 6 \tag{7}$$

$$\alpha_j = \left(\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{6j},\right)' \ (j = 1,2,\ldots,6) \tag{8}$$

4. Calculate principal component contribution rate and cumulative contribution rate:

$$C = \frac{\mu_i}{\sum_{k=1}^{6}\mu_k} \tag{9}$$

$$AC = \frac{\sum_{k=1}^{i}\mu_k}{\sum_{k=1}^{6}\mu_k} \ (i = 1,2,\ldots,6) \tag{10}$$

Where *C* is the main component contribution rate, and *AC* is the cumulative contribution rate.

5. Take the i*th* principal component F_i corresponding to the eigenvalue whose cumulative contribution rate exceeds 85%:

$$F_i = \alpha_1 X_1 + \cdots + \alpha_6 X_6 \tag{11}$$

The digitals show that the cumulative contribution rate of corresponding feature vectors in F_3 is as high as 89.03%, that is, the first three principal components well contain most information of all indicators, and the three principal components have a linear relationship with the original data. Therefore, this paper chooses the first three principal components as new indicators to evaluate the level of light pollution risk in a region. The principal component feature vectors are as Table 1:

Table 1: Principal component eigenvector value

eigenvector	**F_1**	**F_2**	**F_3**
Population density	0.6571	-0.209	0.0869
Per capita disposable income	0.5201	-0.2108	0.1121
visibility	-0.426	0.5089	-0.7606
Cloud cover	0.1199	-0.2013	0.1018
Vegetation coverage	-2.899	0.7391	-0.2682
Pollution index	0.1211	-0.2575	0.5649

Therefore, we can obtain the calculation formulas of three new index data:

$$F_1 = -0.6571x_1 + 0.5201x_2 - 0.4267x_3 + 0.1199x_4 - 0.2899x_5 + 0.1211x_6 \tag{12}$$

$$F_2 = -0.209x_1 - 0.2108x_2 + 0.5089x_3 - 0.2013x_4 + 0.7391x_5 - 0.2575x_6 \tag{13}$$

$$F_3 = 0.0869x_1 + 0.1121x_2 - 0.7606x_3 + 0.1018x_4 - 0.2682x_5 + 0.5649x_6 \quad (14)$$

This paper has a reasonable explanation for the above results:

1. For F_1: Population density and per capita disposable income have a large positive load, while visibility has a large negative load. Therefore, the first principal component is the scale index of local urbanization. The more developed the city economy is, the denser the population is, the lower the visibility of the city people will be.
2. For F_2: There is a relatively large positive load in visibility and vegetation coverage, and a relatively large negative load on population density and per capita disposable income. Therefore, the second principal component is the relative geographical location evaluation index. The smaller the population density, the less per capita disposable income, the greater the visibility, and the higher the vegetation coverage rate, indicating that the region is relatively far away from the city.
3. For F_3: There is a large negative load in visibility, and there is a large positive load in the air pollution index, so the third principal component is called the air particle pollution index. The more serious the air pollution, the lower the visibility, the refraction and reflection of the pollution particles on the light will also cause the intensification of light pollution.

2.2. Calculation of light pollution level based on EWM-TOPSIS model

2.2.1 Index weight calculation based on EWM.

In PCA dimension reduction, this paper builds four indicators, which have different influences on the importance, so we need to assign different weights to the indicators. In order to avoid the influence of human subjectivity on the results, the entropy weight method is adopted in this paper, and the weights are assigned according to the changes of the four indicators, and the TOPSIS model is used to quantify the light pollution level.

Entropy is a measure of uncertainty. Entropy weight method uses the information entropy to calculate the weight of each indicator, and then corrects the weight of each indicator through the weight, so as to obtain a relatively objective index weight. The uncertainty increased with the growth of the entropy. The information contained is in direct proportion to the corresponding weight. Similarly, the decrease in uncertainty leads to the decrease in entropy, which contains less information.

1. Indicator forward:

For the smaller the better index $Z_{ij:}$

$$Z'_{ij} = \max Z_{ij} - Z_{ij} \quad (15)$$

For the index z_{ij} with the optimal value Z_{best}:

$$Z'_{ij} = 1 - \frac{\left|z_{ij} - Z_{best}\right|}{\max\left(\left|Z_{ij} - Z_{best}\right|\right)} \quad (16)$$

For the bigger is better index, we don't need to change it.

2. Index standardization:

Each index has different units. In order to eliminate the influence of different index dimensions, the paper first standardize the data.

$$Z''_{ij} = \frac{Z'_{ij}}{\sqrt{\sum_1^m Z'^2_{ij}}} \quad (17)$$

3. Calculate the proportion of each index under the corresponding sample:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^n z_{ij}} \quad (18)$$

4. Calculate information entropy e_j and information utility value d_j:

$$e_{ij} = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln(p_{ij})\ (j = 1,2,\dots,m) \tag{19}$$

$$d_j = 1 - e_j \tag{20}$$

5. Calculate weight ω_j:

$$\omega_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{21}$$

2.2.2 Evaluation of light pollution level based on TOPSIS model

TOPSIS method is a commonly used comprehensive evaluation method, which can make full use of the information of index data, and its results can accurately reflect the gap between light pollution levels.

Scores in four evaluation aspects:

1. Define A_j as the sample data under the j*th* index:

$$A_j = \left[z_{1j}, z_{2j}, z_{3j}, z_{4j}\right]' \tag{22}$$

$$A = [A_1, A_2, A_3, A_4] \tag{23}$$

2. Calculate the maximum and minimum values for each indicator:

$$Z^+ = \left({A_1}^+, {A_2}^+, {A_3}^+, {A_4}^+\right)$$

$$= (\max A_1, \max A_2, \max A_3, \max A_4) \tag{24}$$

$$D_i^+ = \sqrt{\sum_{j=1}^{4} \omega_j (Z_j^+ - z_{ij})^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^{4} \omega_j \left(Z_j^- - z_{ij}\right)^2} \tag{25}$$

3. Define the score S_i of the i*th* evaluation object:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{26}$$

The S_i value obtained is the score value of light pollution level.

3. Results

3.1. Regional classification criteria based on hierarchical cluster analysis

The division of regions is mainly determined by population density, so this paper take cities as the standard to divide four regions. However, nature reserves are characterized by large land area and small population, so this paper define areas with population density less than 10 as nature reserves, such as Qilian Mountain region, Greater Khingan Mountains region, Jiuzhaigou region, Jiuquan, Ordos, Hegang. However, we do not have a clear system for the division of urban, suburban and rural areas, so this paper use population density as the index of division, and use the systematic clustering model, 56 domestic and foreign regions are redefined as urban, suburban and rural. System clustering is to calculate multiple indexes according to the distance between classes and generate a tree of clustering family tree for classification. The algorithm flow chart is shown in Figure 2.

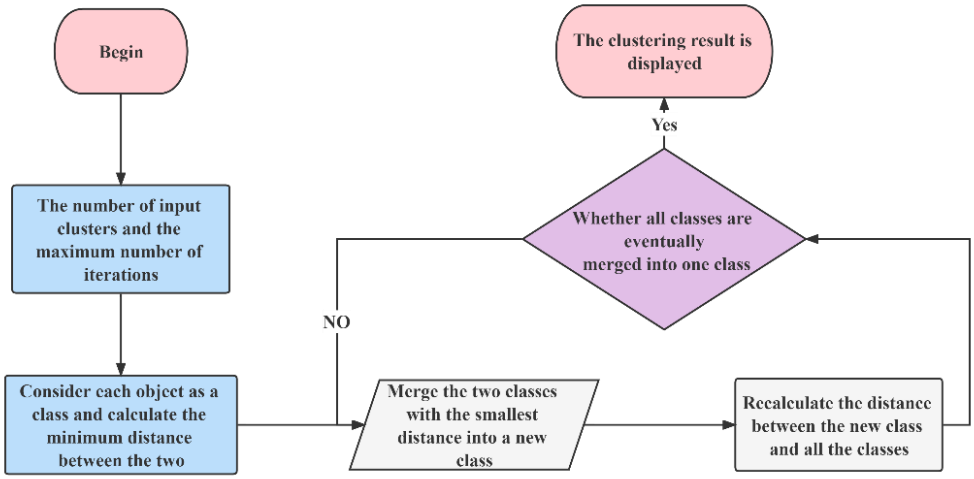


Figure 2: Cluster analysis flow chart

This paper gets the three types of results as Table 2 by substituting the population density data of each city. The three categories of clustering results show that the population density of the first category is all greater than 3000 people per square kilometer, which is divided into cities. The second category has a population density of between 1,000 and 3,000 people per square kilometer, dividing it into suburbs. The third category classifies areas as rural with a population density of less than 1,000 people per square kilometer. By combining the per capita disposable income level of the above locations, that is, reflecting the economic development level of the region, and comparing the differences of economic development level in urban, suburban and rural areas, it can be verified that the division of urban, suburban and rural locations is consistent with the actual situation, and the classification standard is reasonable.

Table 2: Regional classification

Categories	Cities
Urban community	Los Angeles, Chicago, Luoyang, Shanghai, Philadelphia, Shenzhen, Philadelphia, New York
Suburban community	Xi'an city, Beijing, Nanjing, Suzhou City, Houston, Yi Wu, Guangzhou city, Shantou, Foshan
Rural community	Lijing, Xishuangbanna, Yan'an,Yulin, Zhangjiajie, Jilin, Chaoyang, Mount Huangshan, Li Shui City, Panzhihua City, Guilin, Baoji, Shaoguan City, Hohhot, Leshan,Qu Zhou, Yun Fu City, Wuhai City, XiningLanzhou, Meizhou City, Kunming, Meishan City, The City of Wen, Dalian city, Ma On Shan, Nanchong, Shenyang, Shangqiu City, Hangzhou City, Nantong

3.2. Calculation of light pollution levels in different locations

According to the classification of the four regions obtained through cluster analysis, this paper takes six samples from each classification and set one index of each of the four regions as a group, which is then divided into six groups and substituted into the calculation model and to obtain its light pollution level *S* respectively. Taking the value of one group of light pollution level *S* as an example, the results are as Table 3:

Table 3: Light pollution levels in different locations

Locations	S
Greater Khingan Mountains region	3.7
Panzhihua City	4.09
Suzhou City	6.2
New York	9.97

From the comparison of light pollution levels, the digitals show that the level of light pollution from protected land locations to urban community continues to increase, indicating that light pollution is serious in locations with high population density, which accords with the actual situation and the model is reasonable. At the same time, in order to explore the influence of each influencing factor index on the light pollution level, correlation analysis is used to obtain the influence degree of each index on the light pollution level.

4. Conclusions

This paper uses PCA to simplify multiple indicators into three main components, and a few indicators can also reflect most information of all indicators. At the same time, TOPSIS is used to quantify the light pollution risk into scores, which is convenient to compare the light pollution risk levels in different regions and the entropy weight method (EWM) avoids the influence of people's subjectivity and makes the weight obtained objective. This paper concludes that the calculation of light pollution level of our model is consistent with the actual situation, has a certain rationality, and has a certain practical application value.

We will combine the model with the calculation of factor level and find out the minimum value of light pollution level and reasonable value of corresponding influencing factors under the premise of rational use of light source, so as to give the best strategy to reduce light pollution.

References

[1] Longcore T, Rich C. Ecological light pollution [J]. Frontiers in Ecology and the Environment, 2004, 2(4): 191-198.

[2] Falchi F, Cinzano P, Elvidge C D, et al. Limiting the impact of light pollution on human health, environment and stellar visibility [J]. Journal of environmental management, 2011, 92(10): 2714-2722.

[3] S.L. Chen. The impact of light pollution on the environment and health [J]. Chinese Tropical Medicine, 2007, 7(6): 5.

[4] Longcore, T. and Rich, C. (2004), Ecological light pollution. Frontiers in Ecology and the Environment, 2: 191-198. https://doi.org/10.1890/1540-9295(2004)002 [0191: ELP]2.0.CO;2

[5]] Mu Haowei, Li Xuecao, Du Xiaoping, Huang Jianxi, Su Wei, Hu Tengyun, Wen Yanan, Yin Peiyi, Han Yuan, Xue Fei. Evaluation of Light Pollution in Global Protected Areas from 1992 to 2018 [J]. Remote Sensing, 2021, 13(9).

[6] O. Rabaza, D. Galadí-Enríquez, A. Espín Estrella, F. Aznar Dols, All-Sky brightness monitoring of light pollution with astronomical methods, Journal of Environmental Management, Volume 91, Issue 6, 2010, Pages 1278-1287, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2010.02.002.

[7] Ramon Alamús, Salvador Bará, Jordi Corbera, Jaume Escofet, Vicenç Palà, Luca Pipia, Anna Tardà, Ground-based hyperspectral analysis of the urban nightscape, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 124, 2017, Pages 16-26, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2016.12.004.

[8] Yali Katz, Noam Levin, Quantifying urban light pollution — A comparison between field measurements and EROS-B imagery, Remote Sensing of Environment, Volume 177, 2016, Pages 65-77, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2016.02.017.

[9] Lopez-Farias Rodrigo, Valdez S. Ivvan, Paredes-Tavares Jorge, Lamphar Hector. Optimization of sensor locations for a light pollution monitoring network [J]. Journal of Quantitative Spectroscopy and Radiative Transfer,2023,304.

[10] NOAA/NGDC - Earth Observation Group. https://www.noaa.gov.

[11] Kyba CCM, Kuester T, Sánchez de Miguel A, Baugh K, Jechow A, Hölker F, Bennie J, Elvidge CD, Gaston KJ, Guanter L. Artificially lit surface of Earth at night increasing in radiance and extent. Sci Adv. 2017 Nov 22; 3(11): e1701528. doi: 10.1126/sciadv.1701528. PMID: 29181445; PMCID: PMC5699900.

[12] Fangfang. Research on power load forecasting based on Improved BP neural network [D]. Harbin Institute of Technology, 2011.

[13] Guo Xiaowei. Luminance observation and regularity study of urban night light polluted area [D]. Dalian University of Technology,2018. Ma Kunlong. Short term distributed load forecasting method based on big data [D]. Changsha: Hunan University, 2014.

[14] Kocifaj, M., Barentine, J.C. Air pollution mitigation can reduce the brightness of the night sky in and near cities. Sci Rep 11, 14622 (2021). https://doi.org/10.1038/s41598-021-94241-1

[15] Lian X, Jiao L, Zhong J, Jia Q, Liu J, Liu Z. Artificial light pollution inhibits plant phenology advance induced by climate warming. Environ Pollut. 2021 Dec 15; 291: 118110. doi: 10.1016/j.envpol.2021.118110. Epub 2021 Sep 6. PMID: 34525438.