Quantitative Assessment of Light Pollution: A Study on Index Construction Based on Entropy Weight and Topsis Methods

Junyi Yang [1, *, #], Song Zeng [1, #], Jiahao Yang [1, #, *], Yutong Wu [1], Zhengguo Yang [2]

[1] School of Software, South China Normal University, Foshan, China, 510631

[2] School of Future Technology, South China University of Technology, Guangzhou, China, 510641

* Corresponding Author Email: 15347538922@163.com

[#]These authors contributed equally.

Abstract. In recent years, the awareness of light pollution's hazards has grown. This article aims to develop metrics to assess light pollution, propose intervention strategies, and analyze their effectiveness. Key indicators such as population density, GDP, urban electricity consumption, high-rise buildings, vehicle ownership rate, AOD index, and remote sensing brightness were evaluated. Using entropy weight and Topsis methods, a Light Pollution Index (LPI) was derived. A multivariate regression model provided the regression equation for LPI. K-means++ clustering categorized LPI into five levels, delineating light pollution degrees in different regions. This framework aids urban planning assessments and establishes a comprehensive model for light pollution assessment. Through data analysis and evaluation, the feasibility of light pollution reduction measures was demonstrated.

Keywords: Light pollution index, Entropy weight method, Topsis, comprehensive evaluation model, K-means + +.

1. Introduction

With the rapid development of the global economy, light pollution has become a new kind of environmental pollution, which hurts the ecological environment, human normal life, disease infection, astronomical detection, and so on. According to the International Dark Skies Association [1], light pollution is defined as “The inappropriate or excessive use of artificial light that may have serious environmental consequences for humans, wildlife, and our climate". The effects of light pollution go beyond interference with astronomical observations to include the well-being of humans, animals, and plants [2]. It has been linked to circadian disturbances in humans, which may have flow effects, such as depression, obesity, and cancer [3]. Moreover, light pollution suppresses our observations of the night sky, disrupting modern astronomical discoveries [4] and ancient cultural practices [5]. The light pollution degree of the city can be evaluated well by establishing an appropriate evaluation index of light pollution. At the same time, it is also of great significance for formulating feasible intervention strategies.

For this study, we use the TOPSIS model, multiple linear regression, and KMEANS to study the degree of light pollution. These models play a role in assessing impact, with [6] employing the similarity prioritization technique with an idealized solution (Topsis) and using objective and subjective weighting techniques [7] proposing a method for comprehensive assessment of rural flood vulnerability in China based on Shannon entropy method and Topsis method. Using multiple linear regression, [8] used daily active cases or newly diagnosed COVID-19 cases [9], used the combination of multiple linear regression and structural equation modeling to study the characterization of groundwater quality [10], used Kmeans to study the voluntary loss of trust in large-scale social network decision-making. [11] summarized the application of K-means in air pollution analysis.

Data for this study are from the Chinese Bureau of Statistics and the U.S. Bureau of Statistics. Given this research, we first analyze the importance of indexes by entropy weight method and carry on weight distribution objectively. Then, a Light Pollution Index (LPI) is obtained by using the

TOPSIS method based on the weight. Furthermore, to study the relationship between primary index and LPI more intuitively, we set up a multiple regression model and get the regression equation about LPI. At the same time, we use the K-means + + Clustering Method to classify LPI into five grades of light pollution.

2. Assessment of Light Pollution Factors

Due to the multifaceted and complex nature of light pollution's causes, impacts, and existing measurements, a singular indicator cannot comprehensively and objectively describe the extent of light pollution. Therefore, a comprehensive consideration of various factors is crucial for the establishment of a model analyzing light pollution-related indicators.

To comprehensively consider these indicators, we propose a method for in-depth exploration of light pollution factors. Specifically, this method revolves around a thorough exploration of the causes, impacts, and existing measurements of light pollution. This involves considering the underlying factors behind the causes of light pollution, such as the use of glass walls in buildings and excessive nighttime lighting, which directly contribute to the formation of light pollution. Furthermore, we delve into the deeper layers behind these indicators. The formation of light pollution may also be associated with urbanization, with indicators related to urbanization including population density, GDP, building height, and urban electricity consumption, among others.

In considering the adverse effects of light pollution, such as reduced biodiversity and degraded air quality, important biodiversity indicators include species richness, vegetation coverage, and forest coverage. Air quality indicators include ozone concentration, carbon dioxide levels, PM2.5, and others. Additionally, in examining existing measurements of light pollution, parameters such as aerosol optical depth (AOD), SQM index, and the mean DN value of nighttime satellite remote sensing are taken into account.

Furthermore, it is evident that excessive nighttime lighting, as a cause of light pollution, has already been measured through the mean DN value of satellite remote sensing. Results stemming from different perspectives can be interconnected and mutually corroborated. Relevant schematic diagrams are illustrated in Figure 1.

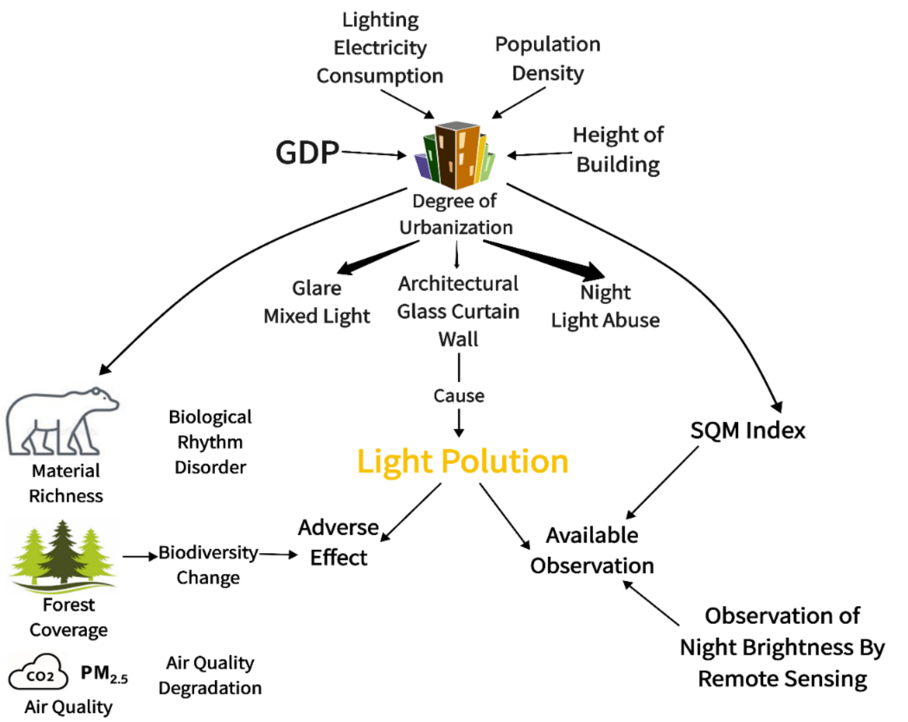

Figure. 1 In-depth exploration of light pollution factors thinking map

In the ensuing discourse, a comprehensive analysis of pertinent indicators concerning light pollution shall be undertaken, delineating its causative factors, ecological impact assessment, and extant methodologies for light pollution assessment. This trilateral investigation ensures a holistic and nuanced examination of indicators germane to the issue of light pollution.

2.1. Light pollution causative factors

As urbanization increases, the Earth's ecosystems are increasingly exposed to artificial nighttime lighting. Improper lighting at night and the reflection of sunlight by glass walls of city buildings during the day are the main causes of light pollution.

Improper urban lighting includes improper use of lamps and ineffective lighting. Improper use of lamps is the use of traditional streetlamps with high power consumption and lower luminous efficiency. Ineffective lighting specifically refers to the fact that the city is still very bright at late night. Some cities' lighting duration and intensity do not change with the change in meteorological conditions and solar illumination time.

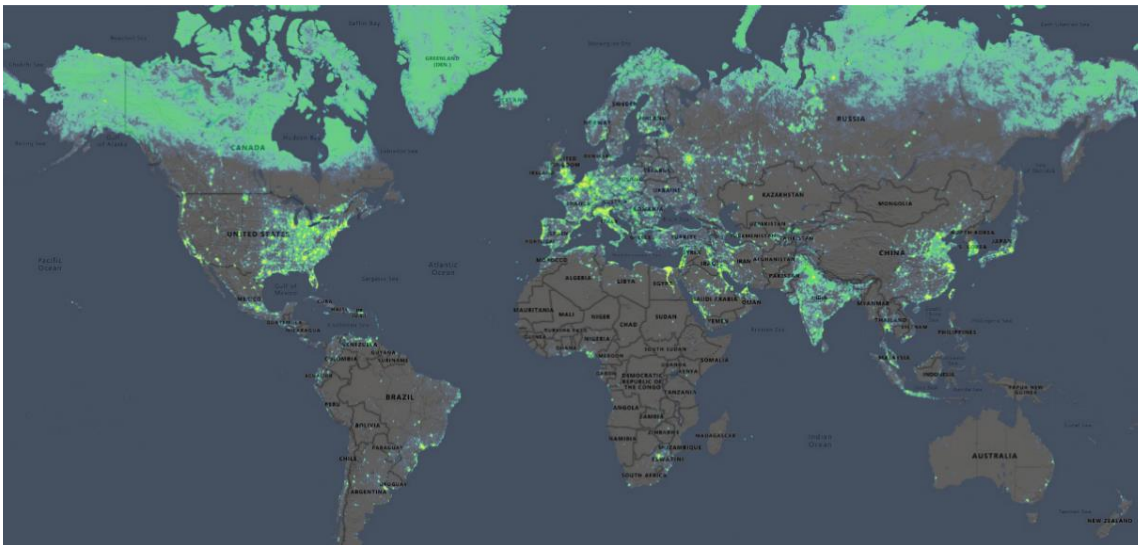

Figure. 2 2021 Global light pollution distribution map (The map is sourced from DMSP.)

As can be seen from Figure 2 2021 global light pollution distribution map, light pollution in coastal areas is more serious in economically developed areas, while light pollution in less developed areas, underdeveloped areas, and inland areas is less. Therefore, the light pollution degree can be reflected to a certain extent by combining it with the evaluation criteria of urbanization. The general criteria for urbanization are population density (in person per square kilometer), GDP (in millions of US dollars), electricity consumption (in GWh) the number of buildings over 150 meters (in buildings), and car occupancy (in cars per person). Qualitative analysis can be made. The higher the population density, the higher the GDP, the more electricity consumption, the more high-rise buildings above 150 meters, the higher the degree of urbanization of the region, and the more serious the degree of light pollution. At the same time, generally speaking, the lighting electricity of the city accounts for about 13% of the total electricity consumption, so the city lighting electricity consumption can be calculated. Also, the higher the share of cars in a region, the more serious the air pollution caused by the reflection of car lights and exhaust emissions. This leads to higher levels of light pollution.

2.2. Ecological impact assessment

Light pollution affects how wildlife forage, reproduce, migrate, and communicate in natural systems. Artificial light can prolong the foraging behavior of animals, which can confuse creatures used to navigating in the dark. Artificial light can affect the ability of hatchlings and birds in the nest to navigate, affecting reproduction and survival. Artificial light also affects interspecies communication. Female fireflies use bioluminescence to attract males from up to 45 meters away. Artificial light reduces the visibility of such communication and affects reproduction.

Therefore, biodiversity is an important indicator of the extent of light pollution. The more diverse the organisms in the area, the lower the light pollution range of 5 degrees, and vice versa. Furthermore, we selected forest coverage as an indicator to assess biodiversity. Another reason is that the vegetation is rough, which helps absorb glare and reduces the harm caused by daytime light reflection, such as increased traffic accidents. Generally, the higher the forest coverage, the better the regional green conditions, the better the biodiversity, and the lower the degree of light pollution.

Light pollution affects air quality. Nitric acid, catalyzed by artificial light, breaks down into harmful gases, increasing ozone and PM2.5 concentrations.PM2.5 and light pollution can work together to affect bird reproduction and migration. So, it is necessary to introduce indicators to determine air quality. The AQI index is a typical index for determining air quality. The lower the AQI index is, the better the air quality is and the lower the light pollution level is. The opposite is true.

2.3. Extant methodologies for light pollution assessment

At present, the measures of light pollution include light intrusion, light source intensity, overspill, AOD, etc. AOD is aerosol optical depth (unit: microgram/cubic meter), which is a physical quantity describing the degree of aerosol absorption and scattering of light in the atmosphere. First of all, increasing aerosol optical depth will enhance the ability of the atmosphere to scatter and absorb light, leading to increased brightness of nighttime lighting facilities and city lights, thus aggravating the degree of light pollution. Especially in the case of haze and other pollution, aerosol particles will increase the scattering and absorption in the atmosphere, making lighting light more easily reflected and diffused, increasing the consumption and energy consumption of nighttime lighting. The global night brightness (W/cm2.μm. rs) can be observed through satellite remote sensing. For example, the OLS sensor carried by the US DMSP satellite can detect faint light on the surface at night and can monitor the spatial distribution of large-scale night brightness. It is a good reflection of how much artificial light is used.

2.4. Overall research structure

Primarily, we categorize urban areas into four distinct types based on population density, ensuring the broad applicability of our study beyond specific regions. Table 1 shows the classification criteria.

Table. 1 Region selection criteria and results

Population Density	Regional type	Screened area
<1	Protected land location	Songshan Mountains, Greater Khingan Mountains, Changbai Mountains, Yellowstone National Park, the Great Rift Valley of East Africa
1~25	Rural community	Nyingchi, Turpan and Sansha
25~100	Suburban community	Qingyuan, Hohhot, Sanya, Columbus
>100	Urban community	New York, Hong Kong, Shenzhen, Guangzhou, Shanghai

Subsequently, we assess light pollution factors across three primary dimensions and synthesize these dimensions, employing the Topsis method based on entropy weighting to compute Topsis scores. Then, we establish a correlation between Topsis scores and light pollution factors using multivariate regression analysis to derive the Light Pollution Index (LPI).

The Topsis scores computed through the Topsis method based on entropy weighting effectively reflect the extent of light pollution in urban areas. To provide a more insightful evaluation of the impact of each indicator on LPI, we employ a multivariate regression model, establishing a regression equation concerning LPI for analysis. classified cities by population density. Figure 3 illustrates the overall research structure diagram.

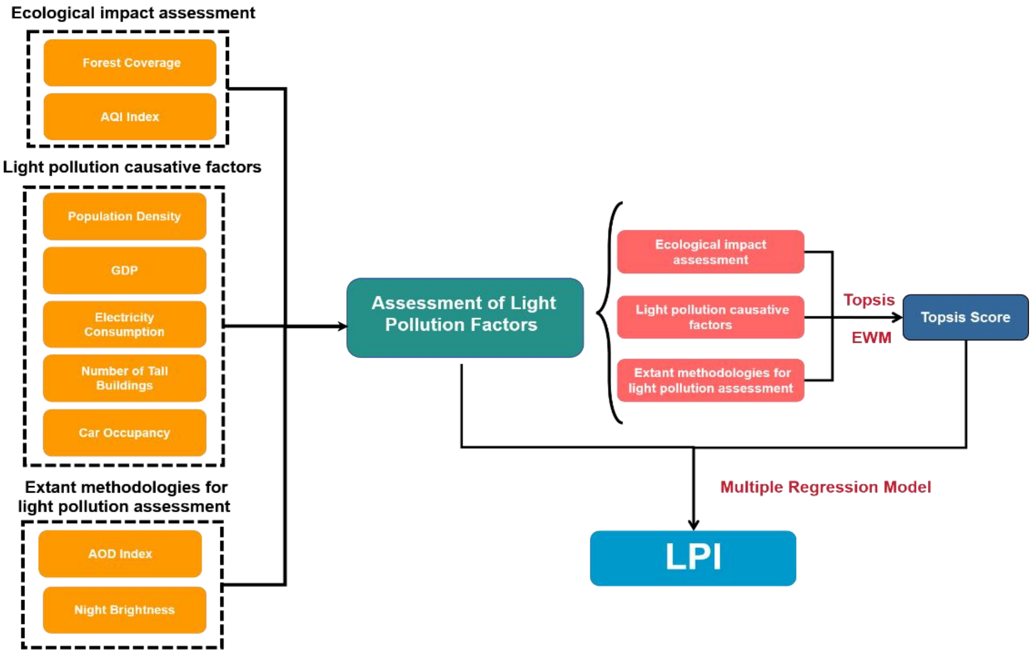

Figure. 3 Overall research structure diagrams

3. Comprehensive Evaluation and Regional Division of Light Pollution Factors

3.1. Topsis model based on entropy weight method

The entropy weight method is adopted to objectively determine the weight of indicators due to the different importance degrees among indicators. Specifically, according to the varying degree of each index, the entropy weight of each index is calculated by using information entropy, and then the weight of each index is corrected by entropy weight, to obtain a relatively objective. The higher the information entropy of an index is, the higher its information content is; on the contrary, the smaller the information entropy is, the lower the information content is. The main steps of the model are shown in Figure 4:

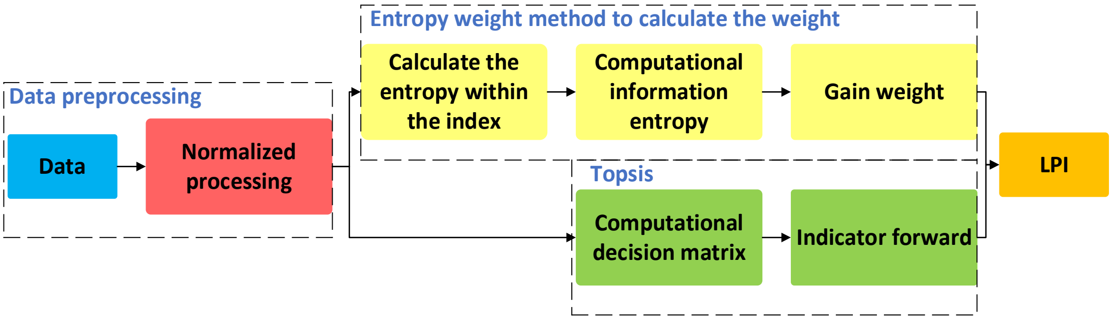

Figure. 4 Flow chart of entropy weight method

The maximum minimum standardized method to sta ndardize data. The formula is $r_{ij} = \frac{x'_{ij}-min(x'_j)}{\max(x'_j)-\min(x'_j)}$, including i for standardized data matrix elements and x_{ij} index after the positive data matrix elements. For the data matrix $R = (r_{ij})_{m\times n}$, for an index r_j, Information entropy for $E_j = -\frac{1}{\ln m}\sum_{i=1}^{m} p_{ij} lnp_{ij}$. The $p_{ij} = \frac{r_{ij}}{\sum_{j=1}^{n} r_{ij}}$. Further, we can determine the weight of each indicator:

$$w_j = \frac{(1-E_j)}{\sum_{j=1}^{n}(1-E_j)} \tag{1}$$

Furthermore, based on the index weights obtained by the entropy weight method, the Topsis method was adopted to make full use of the information of original data to calculate the gap between evaluation schemes. The main steps are shown in Figure 4. Based on the entropy weight method to get the weight, the evaluation objects and the maximum distance $D_i^+ = \sqrt{\sum_{j=1}^{m} w_j\left(Z_j^+ - z_{ij}\right)^2}$, With the minimum distance$D_i^- = \sqrt{\sum_{j=1}^{m} w_j\left(Z_j^- - Z_{ij}\right)^2}$. Then, the final normalized score $S_i = \frac{D_i^-}{D_i^+ + D_i^-}$of the ith evaluation object can be calculated to normalize the score and get the final score.

$$\tilde{S}_i = \frac{S_i}{\sum_{i=1}^{n} S_i} \tag{2}$$

After data substitution, the weights of indicators determined by the entropy weight method are shown in Table 2.

Table. 2 The weighted result of the entropy weight method

Variable	Weight
Electricity consumption	0.152669
GDP	0.168848
Population density	0.146184
AOD	0.096719
AQI	0.052033
Vehicle occupancy	0.179908
Number of tall buildings	0.079864
Radiance value	0.024078
Forest coverage	0.099698

Based on the weight obtained, the comprehensive score of Topsis can be calculated.Table 3 shows the Topsis comprehensive scores of some cities from 2016 to 2020.

Table. 3 Tosis comprehensive scores of some cities 2016-2020

Year / City	2016	2017	2018	2019	2020
Yellowstone National Park, USA	29.37	29.37	29.37	29.37	29.37
Sansha	195.03	194.15	191.22	199.51	195.18
Hohhot	415.90	407.97	401.06	402.74	425.12
New York	1401.55	1412.96	1437.07	1440.2	1421.01

The LPI of the selected regions in descending order are New York, Tokyo, Hong Kong, Chicago, Shanghai, Shenzhen, Beijing, Columbus, Guangzhou, Hohhot, Sanya, Qingyuan, Turpan, Nyingchi, Sansha, Lake District National 8 Park, England, Greater Khingan Mountains Nature Reserve, Yellowstone National Park, USA, Songshan Nature Reserve, Changbai Mountain Nature Reserve. It can be seen that generally speaking, the more heavily urbanized areas are, the more serious light pollution is, and the protected areas are almost unaffected by light pollution.

3.2. Multiple regression model

The scores obtained by the Topsis method can well reflect the degree of urban light pollution. To more intuitively evaluate the relationship between 9 indicators and LPI and the influence of each indicator on LPI, we adopted a multiple regression model and established a regression equation about LPI for analysis.

LPI is selected as the dependent variable, that is, the score after Topsis training, which is set as y. Population density, AQI, electricity consumption, night brightness, AOD, GDP, forest coverage, number of tall buildings, and automobile share are selected as independent variables, which are x1, x2, x3, x4, x5, x6, x7, x8, x9. Then there is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_9 x_9 + \varepsilon \tag{3}$$

Where, β_0, β_1, $\cdots$, β_9 are 10 unknown parameters, ε is random error.

Expressed in matrix form and estimated parameters by the least square method:

$$X'X\hat{\beta}_0 = X'y \tag{4}$$

Due to the possibility of multicollinearity between variables, we adopted stepwise regression to introduce indicators that have a significant impact on LPI one by one. After each independent variable is introduced, the F test is carried out on the selected variables one by one, and variables that no longer have significant influence are eliminated to eliminate the multicollinearity problem.

The results of stepwise regression are shown in Table 4.

Table. 4 Stepwise regression results table

Coefficient	B	Standard Error	P	VIF	R^2	$R^2 - Requested$
Population density	0.03	0.002	0***	9.12	0.999	0.998
AQI	1.381	0.069		3.08		
Electricity consumption	1.773	0.072		5.6		
Night brightness	0.625	0.024		2.14		
AOD	133.223	6.279		1.68		
GDP	0.023	0.001		6.55		
Forest coverage	146.928	6.535		1.9		
Number of tall buildings	0.879	0.039		8.78		
Automobile share	385.931	18.33		5.34		
Constant	150.993	5.021		-		

***, **, * respectively represent significance levels of 1%, 5%, and 10%. Among the calculated results, the R^2-requested was 0.998, indicating a high degree of model fitting. The rank correlation coefficient passed the significance test, and there was no heteroscedasticity. All coefficients passed the significance test. VIF values are all less than 10, and there is no obvious multicollinearity. Finally, the regression equation of LPI is:

$$\begin{aligned} LPI = 150.993 + \ 0.030 * x_1 + \ 1.381 * x_2 + \ 1.773 * x_3 + 0.625 * x_4 + 133.223 * x_5 \\ +0.023 * x_6 - 146.928 * x_7 + 0.879 * x_8 + 385.931 * x_9 \end{aligned} \tag{5}$$

This is a good indicator of model evaluation. It can be seen that the coefficient of other variables except x_7 is positive and has a positive correlation with LPI. The coefficient of x7 is negative and has a negative correlation with LPI, that is, the higher the forest coverage, the lower the LPI.

3.3. Clustering

To better improve LPI, we need to grade LPI after obtaining a comprehensive evaluation of regional light pollution. We adopted the clustering method to classify the scores trained by the Topsis model and obtained relatively objective grading standards. K-means ++ cluster analysis [5]is an unsupervised learning algorithm, which can be divided into k different clusters for similar features of light pollution data sets. The flow of its algorithm is as follows:

i. Randomly select k points as the center points of the cluster.

ii. For each sample, calculate its distance from k central points and assign it to the cluster where the nearest central point is located.

iii. For each cluster, its center point is recalculated.

iv. Repeat steps 2 and 3 until the cluster allocation does not change or the preset maximum number of iterations is reached.

As depicted in Figure 5, the clustering flow chart illustrates the process.

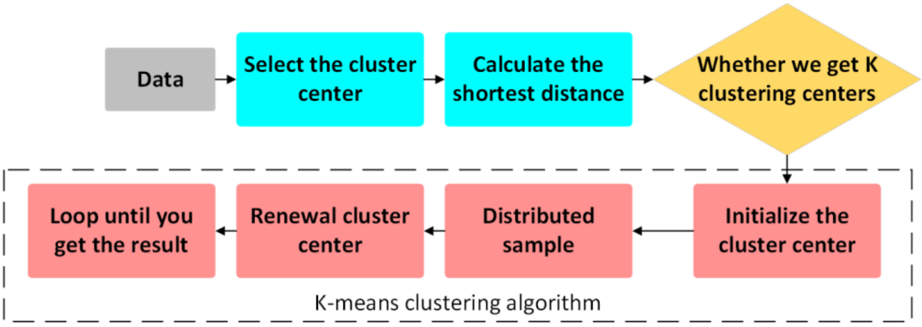

Figure. 5 Clustering flow chart

According to the clustering results, a comparison diagram is generated to illustrate the clustering numbers, as shown in Figure 6. According to the elbow rule, it can be found that the effect is better when divided into 5 categories.

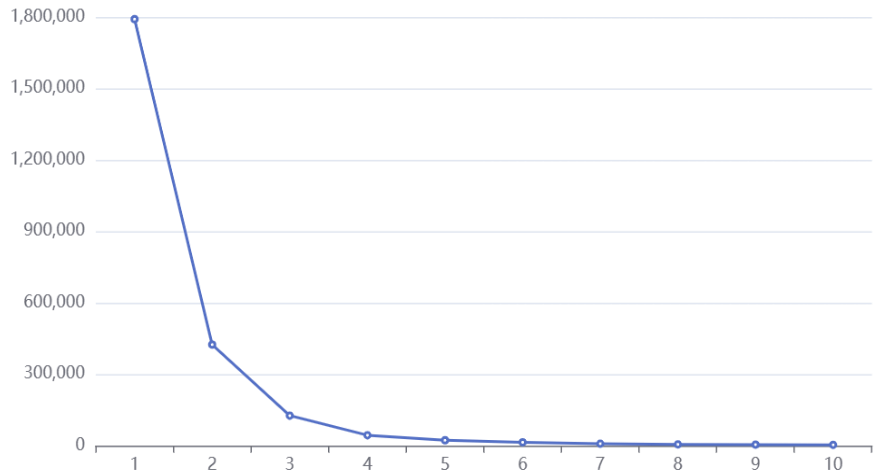

Figure. 6 Cluster number comparison

This indicates that after variable LPI is divided by cluster analysis, there are significant differences among categories, and the model has excellent performance.

Further, we can divide it into pollution-free areas, low-degree pollution areas, moderate pollution areas, heavy pollution areas, and extremely heavy pollution areas. Table 5 presents the LPI classification table.

Table. 5 LPI classification table

Degree of light pollution	LPI	Area
Pollution-free area	0-150	Songshan Nature Reserve. Greater Khingan Mountains Nature Reserve. Changbai Mountain Nature Reserve. Yellowstone National Park, USA. Lake District National Park, England
Low pollution area	150-500	Nyingchi. Turpan. Sansha. Qingyuan. Hohhot. Sanya
Moderately polluted area	500-800	Columbus. Beijing. Guangzhou. Shenzhen
Heavily polluted area	800-1100	Shanghai. Hong Kong. Chicago
Extremely heavily polluted area	1100-1500	New York. Tokyo

4. Conclusions

This text conducted a comprehensive evaluation of light pollution factors through an in-depth exploration, encompassing Light Pollution Causative Factors, Ecological Impact Assessment, and Extant Methodologies for Light Pollution Assessment. Ultimately, this paper selected population density, GDP, urban electricity consumption, quantity of high-rise buildings, vehicle ownership rate, forest coverage, AQI index, AOD index, and nocturnal remote sensing brightness as primary indicators.

Subsequently, employing the entropy weight method and relevant literature, the significance of these indicators was analyzed, and weights were objectively allocated. Then, based on these weights, the Topsis method was utilized to score and comprehensively assess the degree of light pollution, yielding a Light Pollution Index (LPI). Furthermore, to elucidate the relationship between primary indicators and LPI more intuitively, this text constructed a multivariate regression model to derive the regression equation for LPI. Concurrently, employing the k-means++ clustering method, LPI was categorized into five levels, thus delineating the degree of light pollution in different regions.

This paper presents a research framework applicable to urban planning evaluation and establishes a comprehensive model for assessing light pollution based on this framework. Through systematic data analysis and comprehensive evaluation, this text have demonstrated the feasibility of this method: not only does it provide insights into various factors and ecological impacts of light pollution, but it also objectively assesses the degree of light pollution in different regions, thereby furnishing a scientific basis for urban planning.

References

[1] Miao, C., Ting, X., & Daqiang, Y. (2023). Understanding light pollution: Recent advances on its health threats and regulations [J]. Journal of Environmental Sciences, 127, 589-602.

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

[3] Hatori, M., Gronfier, C., Van Gelder, R. N., Bernstein, P. S., Carreras, J., Panda, S., ... & Tsubota, K. (2017). Global rise of potential health hazards caused by blue light-induced circadian disruption in modern aging societies. *npj Aging and Mechanisms of Disease*, *3*(1), 9.

[4] Hamacher, D. W., De Napoli, K., & Mott, B. (2020). Whitening the Sky: light pollution as a form of cultural genocide. *arXiv preprint arXiv:2001.11527*.

[5] Khan, A., Gong, Z., Shah, A. A., & Haq, M. (2024). A comprehensive assessment model for livelihood vulnerability using fuzzy TOPSIS technique: A case of 2022 flood in Pakistan. *International Journal of Disaster Risk Reduction*, 104351.

[6] Khan, A., Gong, Z., Shah, A. A., & Haq, M. (2024). A comprehensive assessment model for livelihood vulnerability using fuzzy TOPSIS technique: A case of 2022 flood in Pakistan. *International Journal of Disaster Risk Reduction*, 104351.

[7] Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & metabolic syndrome: clinical research & reviews*, *14*(5), 1467-1474.

[8] Chenini, I., & Khemiri, S. (2009). Evaluation of groundwater quality using multiple linear regression and structural equation modeling. *International Journal of Environmental Science & Technology*, *6*, 509-519.

[9] Yu, S. M., Du, Z. J., Zhang, X. Y., Luo, H. Y., & Lin, X. D. (2021). Trust Cop-Kmeans clustering analysis and minimum-cost consensus model considering voluntary trust loss in social network large-scale decision-making. *IEEE Transactions on Fuzzy Systems*, *30*(7), 2634-2648.

[10] Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric pollution research*, *11*(1), 40-56.