# Research on the comprehensive evaluation of light pollution

**Yongchen Zhao**

SWUFE-UD INSTITUTE OF DATA SCIENCE AT SWUFE, Southwestern University of Finance and Economics, Chengdu, 611130, China

yczhao1666@163.com

**Abstract.** Light pollution refers to the inappropriate and excessive use of artificial light. The increasing extent and intensity of artificial light has impacted the biology and ecology of species significantly. Admittedly, the widespread use of light benefits people to a large extent and is positively associated with modernization, security, and wealth. But its catastrophic effects can never be ignored. To be specific, light pollution can arouse negative health impacts such as headaches, dizziness, increased anxiety, pressure, and fatigue. The paper wants to find applicable indicators regarding the risk levels of light pollution and establish criteria to judge the risk of light pollution in different areas. In the process, the research first measures the interconnection between the indicators, which are chosen to reflect the risk of light pollution, and then uses PCA to implement dimensionality reduction so as to simplify the model. After that, EWM and TOPSIS are applied to determine the weight of each indicator and the rating of the cities. Institutions or governments that are responsible for managing light pollution can then use the model to judge the risk level of different cities. The model can help to avoid overlooking or overemphasizing a city's light pollution risk level, providing a more accurate estimation. In this case, institutions and the government can take better and more effective measures to restrict light pollution.

**Keywords:** Light Pollution, PCA, EWM, TOPSIS.

## 1. Introduction

The problems caused by light pollution have become increasingly important in the past few decades. For a deeper understanding of light pollution, many researchers have focused on the development of light pollution monitoring technology, the effects of light pollution on ecosystems, and the effect on health status. These studies draw the conclusion that light pollution does have a negative impact on humans and the overall environment. Also, studies on how to control the impact of light pollution are implemented. However, few studies have been done to classify the risk level of light pollution. So the research theme is about the comprehensive evaluation of light pollution. The paper wanted to establish a model to help define the risk level of light pollution in different areas. In the model, the researchers use PCA, EWM, and TOPSIS to implement dimensionality reduction and rating. At last, criteria are set to justify the risk of light pollution in different cities. The standard could help researchers identify areas most affected by light pollution to better understand the impact of light pollution on ecosystems. Also, criteria can help governments and health authorities alert residents in high-risk areas and take appropriate precautions to reduce health problems associated with light pollution.

## 2. General Assumptions and Justification

In order to find widely applicable indicators which illustrate the risk level of light pollution, the paper set 7 first-level indicators and 16 second-level indicators at first. To simplify the model and find the most influential factors, Principle Component Analysis is applied to reduce dimension. Then based on processed indicators, we use Entropy Weight Method and TOPSIS and determined the weight and merit ranking of each city.

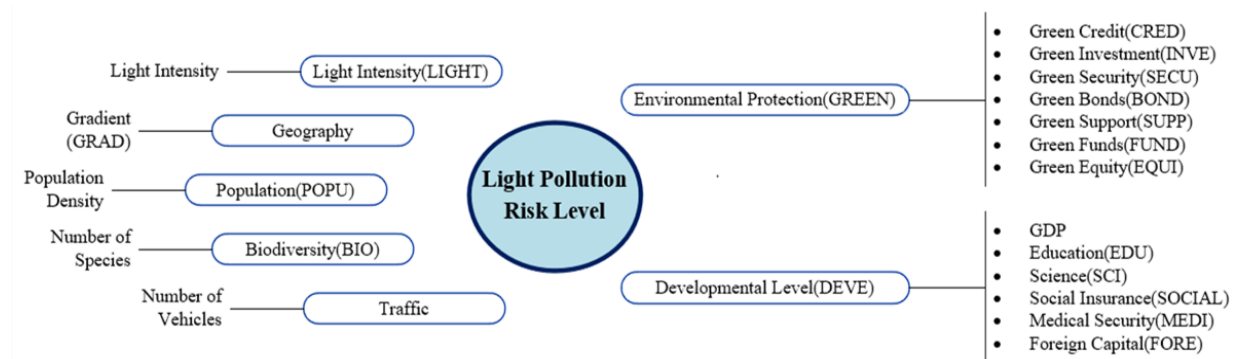We assume that data collected from different sources are accurate and reliable.



**Figure 1.** Light pollution and related indicators.

We collected data from valid websites like Statistical Bureau, Ministry of Science and Technology, People's Bank of China and other authoritative institutions' websites and various authoritative statistical yearbooks [1].

## 3. Model

In this section, at the beginning, the paper worked on two first-level indicators which have more than five second-level indicators. By using PCA, the research gets the most primary second-level elements affecting its first-level indicators, in this way first-level and second-level indicators are related to each other. Then the usage of Entropy Weight Method and TOPSIS enables me to rate cities with different grades regarding this two indicators. After that, the paper repeats the above process with all first-level indicators and get every city a grade so as to find the influence of light pollution to them [2]. Finally, a criteria is set to help judge the risk level of light pollution.

### 3.1. Data

Collecting data is an indispensable process solving the tasks. So it is crucial to obtain accurate and reliable database and integrate them carefully. The paper has access to data through different authoritative sources such as National Bureau of Statistics and China Statistical Yearbook. However, ascribing to the fact that not all data are on the websites, the research has to handle missing data by ameliorating them in the following ways:

• Some cities lack much data (more than 80 percent) about second-level indicators, however, since these cities only account for a very small proportion and the data are insignificant to the overall figures, the paper used deletion method to delete them.

• For the two secondary indicators with less data missing ( less than 2 percent), the research filled them according to the data distribution. These secondary indicators have skewed distribution of data, and the research used the median to fill in.

• During the integration process, the paper also populated the secondary indicators by interpolation, and the research used random interpolation to process the secondary indicators in the level of development of the primary indicator.

### 3.2. Primary parameters

In order to analyze the effect of light pollution, the paper set 7 first-level indicators and 16 secondary indicators as shown in table 1.Among first-level indicators, light intensity is the only direct factor associating with effect of light pollution, while others are all indirect measurements [3].

**Table 1.** Primary Indicators.

| Object | First-Level Indicators | Second-Level Indicators |
|---|---|---|
| Effect of light pollution | Light Intensity(LIGHT) | Light Intensity |
| | Environmental Protection(Green) | Green Credit(CRED) |
| | | Green Investment(INVE) |
| | | Green Security(SECU) |
| | | Green Bonds(BOND) |
| | | Green Support(SUPP) |
| | | Green Funds(FUND) |
| | | Green Equity(EQUI) |
| | Geography | Gradient(GRAD) |
| | Developmental Level(DEVE) | GDP |
| | | Education(EDU) |
| | | Science(SCI) |
| | | Social Insurance(SOCIAL) |
| | | Medical Security(MEDI) |
| | | Foreign Capital(FORE) |
| | Population(Popu) | Population Density |
| | Biodiversity(Bio) | Number of Species |
| | Traffic | Number of Vehicles |

### 3.3. Correlation Coefficient

The paper analyzed correlation Coefficient and covariance between different secondary indicators which reflect the same first-level indicators and analyzed the relevance between 7 first level indicators as well. As showed in the pictures below, the research asserted that many secondary factors about environmental protection and developmental level are closely related to each other, as a result, the research can further simplify the model by PCA. However, the covariance is low between 6 first-level indicators, so the research can still retain them.

Correlation coefficient and covariance are calculated as following:

$$Cov(X, Y) = E\big[(X - E(X))(Y - E(Y))\big] \qquad (1)$$

### 3.4. Primary Indicator System

16 indicators are selected initially to determine the effect of light pollution . Among all the indicators, secondary indicators regarding environmental protection and developmental level account for the most proportion. And by calculating their covariance, the paper found that they are closely related to each other so than the research can then use PCA according to figure 2 and 3. PCA is the most commonly used linear dimensionality reduction method. Its goal is to map high-dimensional data into a low-dimensional space through some linear projection, and expect the most
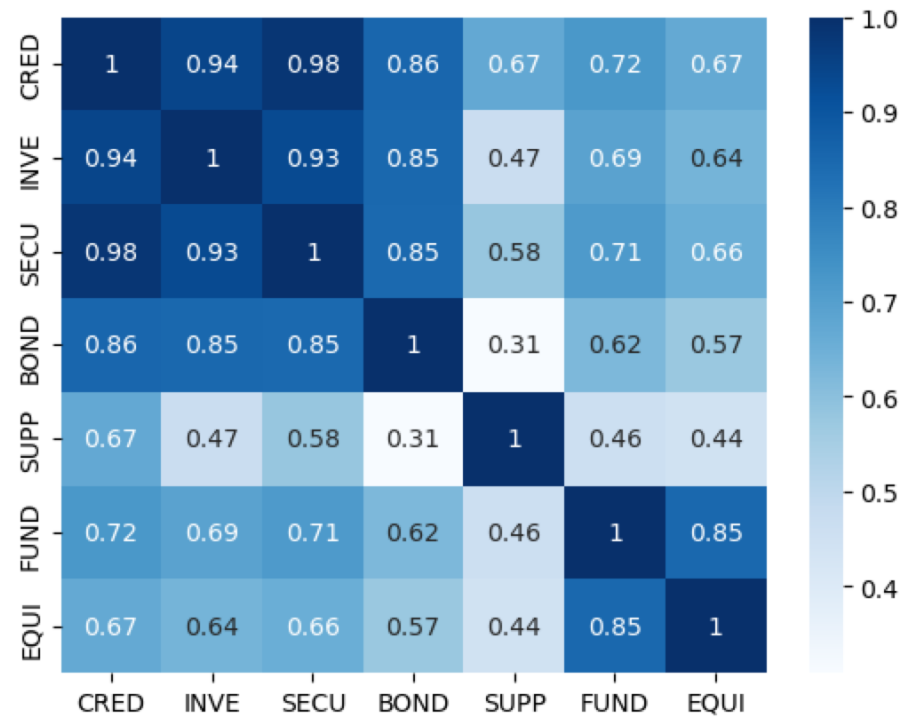
**Figure 2.** Environmental Protection.



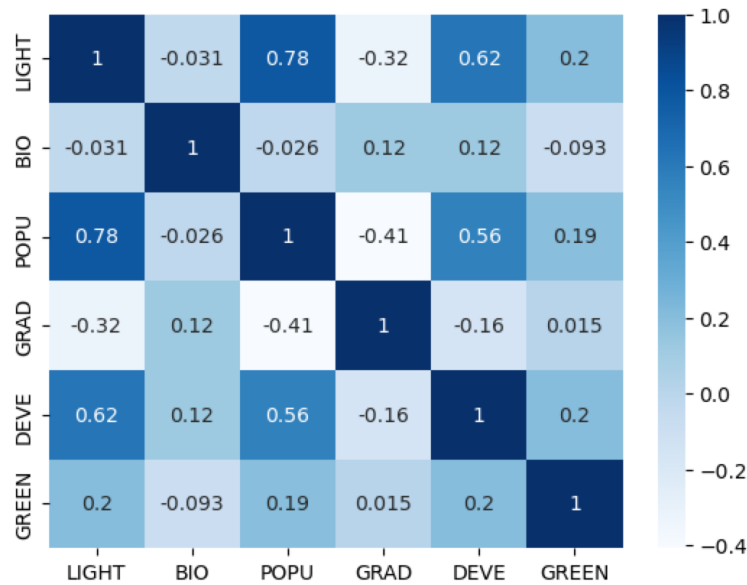**Figure 3.** Developmental Level.

**Figure 4.** First level Indicators.

Information (maximum variance) of the data in the projected dimension, which fit the research's request and can integrate different secondary indicators into one indicator, thus simplifying the whole model.

As what is shown in Figure 5 and Figure 6, cumulative explained variance far overweight individual explained variance. So the paper can conclude that by using PCA and mapping data into lower dimension, the research can have access to more information. After this process, all of the secondary indicators about one single first-level indicator are integrated or they represent the first-level indicator themselves ,the research can then only consider first-level indicators regardless of secondary factors in the further steps.

*3.5. Entropy Weight Method*

The entropy weight method (EWM) is an information weight model which is commonly used to measure value dispersion in decision-making. The greater the degree of dispersion, the greater the degree of differentiation, and more information can be derived. The biggest advantage considering EWM is its objective property compared with other weight- ing methods, for it avoids interference from human factors [4].
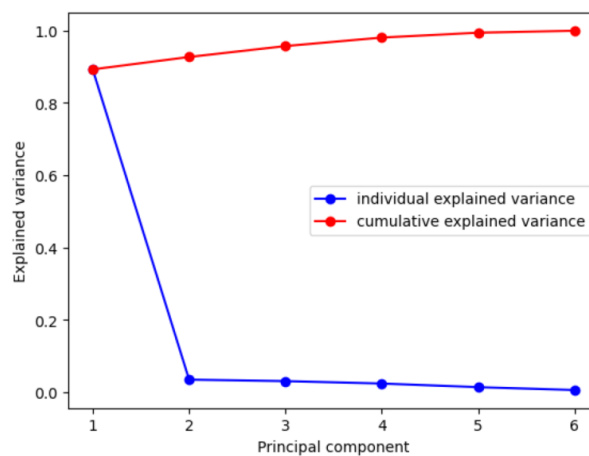


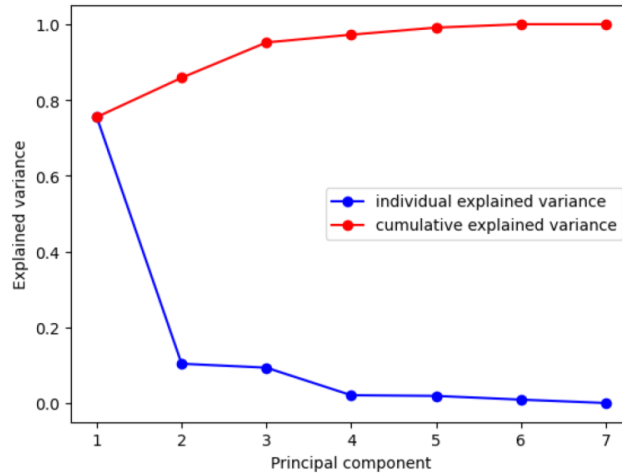**Figure 5.** PCA Environmental Protection.

**Figure 6.** PCA Developmental Level.

To be specific, in the task, the paper first focused on first-level indicators with more than five secondary factors: environmental protection and developmental level and gave them different weights through calculation. If the figures vary significantly in one indicator, then it is likely to get higher weight.

Step 1:

Determine the evaluation object, establish the evaluation index system, and construct the level matrix

$$\begin{bmatrix} IndexI_1 \\ IndexI_2 \\ \vdots \\ IndexI_m \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{n2} & \cdots & x_{mn} \end{bmatrix} \tag{2}$$

Step 2:

Data normalization: Data comes from different sources and have different dimension and

units. Obviously, indicators with greater units will occupy more weigh when using EWM, which is unjustified. Then it is imperative to unify the unit and convert all of them into numbers between 0 and 1 so that different indicators can directly compare with each other. This process is called data normalization. In the task, both positive and negative indicators are included and they have different forms of normalization:

Positive indicator:

$$r_{ij} = \frac{x_{ij} - \underset{i}{Min}(x_{ij})}{\underset{i}{Max}(x_{ij}) - \underset{i}{Min}(x_{ij})} \tag{3}$$

Negative indicator:

$$r_{ij} = \frac{\underset{i}{Max}(x_{ij}) - x_{ij}}{\underset{i}{Max}(x_{ij}) - \underset{i}{Min}(x_{ij})} \tag{4}$$

We know that the smaller the negative indicators are, the more influential the object is. By using different data normalization, negative indicators reverse their effect. All the indicators work in the same direction now, which is beneficial in establishing further model.

Step 3:

Calculate entropy values for each indicator: The greater the amount of information, the smaller the uncertainty and the smaller the entropy; the smaller the amount of information, the greater the uncertainty and the larger the entropy.

$$H_j = -k \sum_{i=1} f_{ij} \cdot \ln f_{ij} \tag{5}$$

$$f_{ij} = \frac{r_{ij}}{\sum_{i=1}^{m} r_{ij}}, k = \frac{1}{\ln m} \tag{6}$$

Step 4:
Calculate the entropy weight of the jth indicator:

$$w_j = \frac{1 - H_j}{\sum_{i=1}^{n}(1 - H_j)} = \frac{1 - H_j}{n - \sum_{i=1}^{n} H_j} \tag{7}$$

*3.6. TOPSIS*

TOPSIS is a multiple criteria method to identify solutions from a finite set of alternatives based upon simultaneous minimization of distance from an ideal point and maximization of distance from a nadir point [5].

In the task, the research combines TOPSIS with EWM. After determining weights of secondary factors of environmental protection and developmental level, the paper can than calculate grade of each city and find how light pollution affects them.

step1:
Multiply weight got from EWM to each column of indicator:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{n2} & \cdots & x_{mn} \end{bmatrix} \tag{8}$$

step 2:
Find optimal and inferior sets: After weighting every data, the research then needs to find optimal and inferior values considering each indicator and form the optimal as well as inferior set as shown in table 2.

In the task, since all the indicators are now positivizing factors, the research choose largest numbers as optimal values and smallest numbers as inferior values. Here is the optimal set and the inferior set is 0. The paper notes optimal set as $z_j^+$ and inferior set as $z_j^-$

**Table 2.** Optimal set.

| Light | 0.161221 |
|---|---|
| BIO | 0.456925 |
| POPU | 0.066535 |
| neg_GRAD | 0.013861 |
| TAXI | 0.170793 |
| pca_develop | 0.110252 |
| neg_pca_green | 0.020412 |

step3:
Rating and Ranking: The research first calculate Euclidean distance between each data and optimal as well as inferior set.

$$D_i^+ = \sqrt{\left(z_{j_{ij}^{\pm z}}\right)^2} \tag{9}$$

$$D_i^- = \sqrt{\left(z_j^- - z_{ij}\right)^2} \tag{10}$$

Then the research set an evaluation method to rate and rank every city.

$$s_{ij} = \frac{D_i^-}{D_i^{\mp D_i^+}} \tag{11}$$

Since each city has a rating now, the research can further develop a unified evaluation form judging light pollution risk levels of cities as shown in table 2. Degrees are divided based on quantile.

**Table 3.** Classical standard of light pollution risk.

| Grade | Degree | Risk |
|---|---|---|
| 0 | Very Low | <2.54 |
| 1 | Low | [2.54,3.46) |
| 2 | Middle | [3.46,5.26) |
| 3 | High | [5.26,9.64) |
| 4 | Very High | >9.64 |

*3.7. Sensitivity Analysis*
The paper makes sensitivity analysis to judge the stability of the model. The picture below shows the first order sensitivity of each indicator. Commonly speaking, the higher the first order sensitivity, the more influence an indicator can exert on the model. According to the graph, the sensitivity of biodiversity is high, however, this indicator seldom changes, so it has tiny impact on model. The first level sensitivity of other indicators is low. As a result, the overall sensitivity is low and the model is stable.
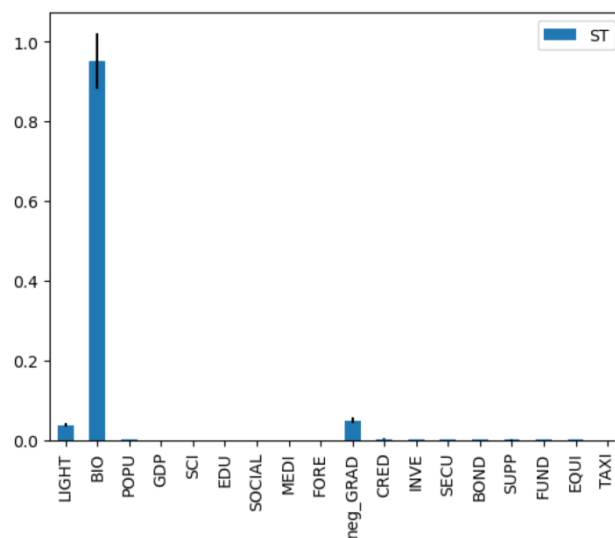


**Figure 7.** Sensitivity Analysis.

## 4. Conclusion

In conclusion, this research aimed to address the issue of light pollution by establishing a comprehensive evaluation model to determine the risk levels of light pollution in different areas. The study began by recognizing the growing importance of light pollution and its adverse effects on biology, ecology, human health, and the overall environment. While artificial light has undeniable benefits, its negative impacts cannot be ignored.

The primary goal of this research was to find applicable indicators that could reflect the risk levels of light pollution accurately. To achieve this, a multi-step methodology was employed, which involved the measurement of interconnections between selected indicators, dimensionality reduction through Principal Component Analysis (PCA), and the application of the Entropy Weight Method (EWM) and TOPSIS to determine the weight of each indicator and the rating of cities.

The establishment of this model has several significant implications. First and foremost, institutions and government bodies responsible for managing light pollution can utilize the model to assess the risk level of different cities. This allows for a more accurate estimation of light pollution risk, helping to prevent underestimation or overemphasis on the issue in specific areas. In turn, this enables governments to implement more effective strategies and policies to curb light pollution and its adverse effects.

Despite these achievements, it is essential to acknowledge certain limitations of this study. One limitation is that the model's accuracy is contingent on the reliability of the data sources used. Therefore, continuous efforts should be made to collect accurate and up-to-date data. What's more, as our understanding of light pollution deepens, it is crucial to refine and expand the set of indicators used in the model to make it even more comprehensive.

Future research in this field should focus on refining the model and expanding its applicability to different regions and contexts. Moreover, collaborative efforts among researchers, governments, and communities are vital to develop effective strategies for mitigating light pollution and preserving the environment and public health.

## References

[1] National Bureau of Statistics, 2019, Classification of Health Industry, http://www.stats.gov.cn/xxgk/tjbz/gjtjbz/201904/t20190411_1758929.html

[2] Zielińska-Dabkowska, K. M., Xavia, K., & Bobkowska, K. (2020). Assessment of citizens' actions against light pollution with guidelines for future initiatives. Sustainability, 12(12), 4997.

[3] Hölker F, Bolliger J, Davies TW,Giavi S, Jechow A, Kalinkat G,Longcore T, Spoelstra K, Tidau S,Visser ME and Knop E (2021) 11Pressing Research Ques- tions on How Light Pollution Affects Biodiversity.Front. Ecol. Evol. 9:767177.doi: 10.3389/fevo.2021.767177

[4] Zhu Y, Tian D, Yan F. Effectiveness of entropy weight method in decision-making [J]. Mathematical Problems in Engineering, 2020, 2020: 1-5.

[5] Olson D L. Comparison of weights in TOPSIS models [J]. Mathematical and Computer Modelling, 2004, 40(7-8): 721-727.