Research on Light Pollution Model Based on Clustering Algorithm

Xiaomeng Li *, Baoyue Zhang
School of Mathematics, Southeast University, Nanjing, China, 211189
* Corresponding author: 213200790@seu.edu.cn

Abstract. Over the past decade, artificial light sources have brightened the night sky so much that the number of stars visible to the naked eye has plummeted, the US journal Science reports. In addition, light pollution can also destroy bio-diversity and even affect the ecological balance. Therefore, it is very important to establish an evaluation index of light pollution and put forward effective treatment strategies.In this paper, we set up a light pollution evaluation model (Model I) through the integration of a large number of data and principal component analysis, which is used to evaluate the degree of light pollution in a region, and it is applied to four types of regions, and the clustering algorithm is used for analysis.Through statistical analysis of a large number of samples, light pollution was first divided into four risk levels: no pollution, light pollution, moderate pollution and heavy pollution, and the four risk levels were assigned a score range. Then, we selected seven parameters affecting light pollution, and made statistical analysis of parameters in four regions to obtain eight groups of data. Through principal component analysis, we reduced their dimensions, and finally got the relationship between the grade fraction of light pollution and the three principal components. So we used the score as a measure of universality to determine the level of light pollution risk in a place.

Keywords: Light Pollution, PCA, Intervention Strategies.

1. Introduction

In the past decade, artificial light sources have increased the brightness of the night sky so much that the number of stars visible to the naked eye has plummeted, the US journal Science reports. If this continues, there will be less than half as many stars visible to the naked eye in less than 20 years [1]. Some studies point out that light pollution can affect the natural life of plants and animals and destroy the balance of the ecosystem. In humans, it can disrupt the body clock and disrupt endocrine balance, leading to physical and psychological problems [2]. Excessive lighting also consumes unnecessary electricity, creating a huge waste of energy [3]. Moreover, with the development of science and technology in the future and the continuous increase of population, light pollution will face a more severe situation. If we do not pay attention to the control of light pollution, it is likely that in the coming decades, the global ecological environment will be seriously damaged, and more organisms will be extinct. Fortunately, more and more countries and more regions have paid attention to the problem of light pollution. The International Lighting Commission has formulated the light pollution evaluation index for outdoor lighting facilities, recommended limits and calculation methods. North America, Europe and Asian countries have also introduced relevant restriction policies to improve light pollution [4]. In this paper, PCA method will be used to analyze the light pollution situation in different regions according to multiple indicators, to develop a more targeted, more effective strategy. The figure 1 shows light pollution around the world. As can be seen from the picture, light pollution has gradually become one of the inevitable problems in our life. Therefore, the management and control of light pollution is very important.

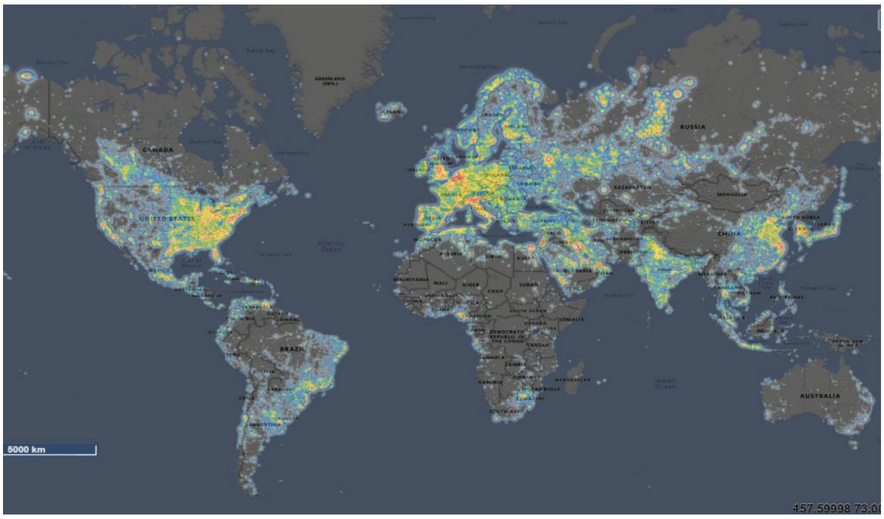

Figure 1: Lighting pollution worldwide
(https://www.lightpollutionmap.info/)

2. Light pollution risk level Measurement Model

Light pollution is determined by a variety of factors, namely, Illumination, zenith brightness, effective illumination, illumination time, per capita electricity consumption, per capita GDP [5], urbanization rate and other factors, all of which will affect the risk level of light pollution[6]. Therefore, a method of multivariate analysis should be used to solve such a complex multivariable system. While principal component analysis can eliminate the correlation influence among evaluation indicators and reduce the dimension of the data space studied, it can also know which factors take the heaviest proportion among the dependent variables we want, so as to facilitate the subsequent evaluation of the scheme and model. Therefore, the model based on principal component analysis is a good approximation for predicting the level of light pollution risk.

2.1. Light Pollution Risk Level Standard

After referring to the classification standards of other countries and consulting a large amountofliterature data, the paper divided the risk of light pollution into four levels: no pollution, light pollution,moderate pollution and heavy pollution[7]. The specific fractions are shown in the Table 1:

Table 1: Level standard

Levels	Score Interval
No Pollution	[0 , 20)
Light Pollution	[20 , 60)
Moderate Pollution	[60 , 90)
Heavy Pollution	[90 , 100]

According to the above scoring criteria, the paper has selected several typical samples worldwide respectively corresponding to the four regions. Through statistical analysis, the following conclusions are preliminarily drawn:

The four sections can approximately correspond to four kinds of sites, that is, protected lands correspond to no pollution, rural communities correspond to light pollution, suburban communities correspond to moderate pollution, and urban communities correspond to heavy pollution.

2.2. Principal Component Analysis

First of all, according to relevant data and obtained the parameters of eight locations as shown in the Table 2 after calculating the mean value through statistical analysis:

Table 2: Parameters for 8 different regions

	IL	*ZB*	*EI*	*IT*	*EC*	*GDP*	*UR*	*ARL*
UC_1	500	150	30	14	8500	65000	100	95
SC_1	200	50	70	10	6000	35000	90	85
RC_1	50	10	90	8	3500	20000	50	50
PL_1	1	1	100	2	80	800	7	5
UC_2	200	50	10	12	5000	20000	85	80
SC_2	50	20	40	7	3000	10000	60	45
RC_2	1	1	50	4	1000	2500	30	20
PL_2	0.01	0.01	60	1	20	40	1	2

Then, according to the data in the table, with the second to seventh columns as independent variables and RL as dependent variable, principal component analysis was carried out on these eight groups of data [8]. After data standardization and principal component extraction, the following results were obtained in the Table 3:

Table 3: Cumulative contribution rate

	IL	*ZB*	*EI*	*IT*	*EC*	*GDP*	*UR*	*ARL*
Rate	83.4677	94.2728	99.3696	99.7544	99.9381	98.9882	98.6785	99.7783

As can be seen from the figure above, the cumulative contribution rate of the first three principal components has reached 99.3969%, so we choose the first three principal components, and the following fig 2 can be obtained:

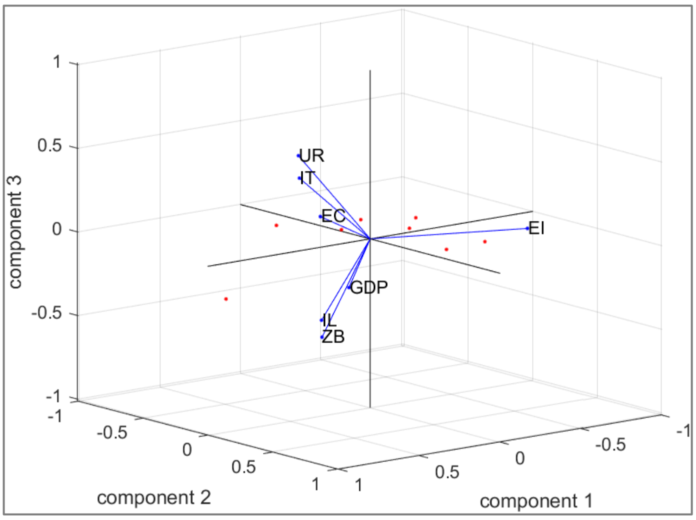

Figure 2: The relationship between the principal component and the original variable

This figure reflects the relationship between the extracted three principal components and the original variables and is drawn with the coefficients of each original variable and the three principal components Z_1, Z_2 and Z_3 as coordinates. The red dots are sample scores. According to the calculated coefficients, the paper can write the expressions of the extracted three principal components as follows:

$$\mathbf{Z_1 = 0.3987IL + 0.3918ZB - 0.2434EI + 0.3968IT + 0.4070EC + 0.3930GDP + 0.3883UR} \quad (1)$$

$$\mathbf{Z_2 = 0.1210IL + 0.1161ZB + 0.9176EI - 0.0530IT + 0.1223EC + 0.3272GDP + 0.0713UR} \quad (2)$$

$$Z_3 = -0.3970IL - 0.5005ZB + 0.2097EI + 0.4168IT + 0.2245EC - 0.1610GDP + 0.5458UR \quad (3)$$

Thus, for the principal componentZ_1, all variables except EI play a role. For the principal component Z_2, EI plays a major role, and for the principal component Z_3, IT and UR play a major role.

Finally, according to the calculated principal component and the cumulative contribution rate, the expression of fitting risk level FRL can be calculated as follows:

$$FRL = 83.4677Z_1 + 10.8105Z_2 + 5.1187Z_3 \quad (4)$$

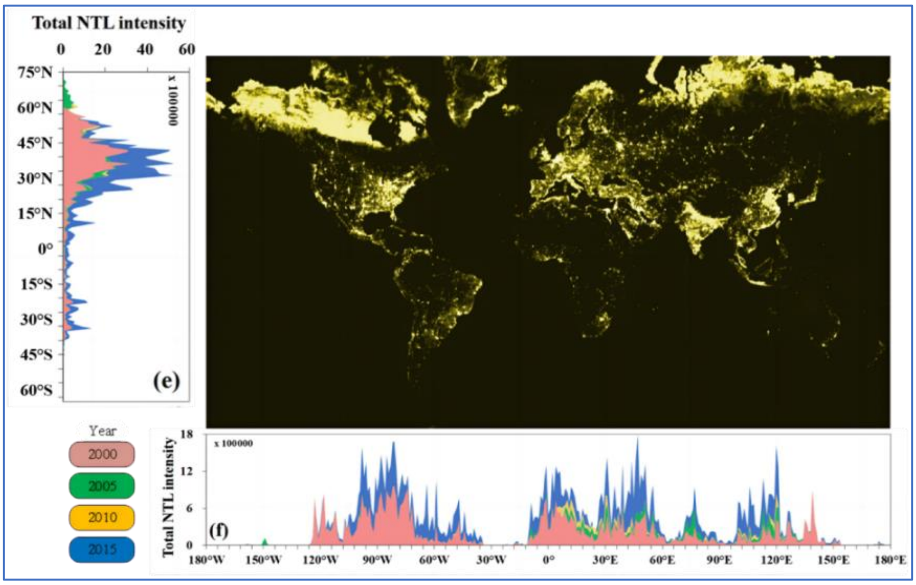

Figure 3: Global night light data in 2021 and dynamics of NTL intensity from 2000 to 2015

(Data Source: NCEI, the National Environmental Information Center under the National Oceanic and Atmospheric Administration NOAA)

3. Results

In thr paper, when getting the formula for the model, According to the global visualization of the night light data in Figure 3,nine different sites were selected to apply our light pollution risk level model[9]. Through cluster analysis, the three principal component scores of each sample were used to divide these nine sites into three categories. The clustering results are as follows Figure 4:

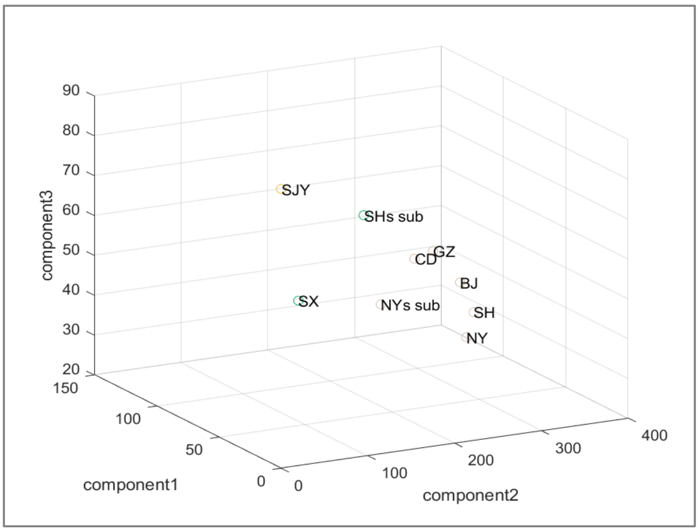

Figure 4: Cluster analysis results

Based on the above results, we can obtain the visual results obtained by the cluster analysis in China as shown in Figure 5:

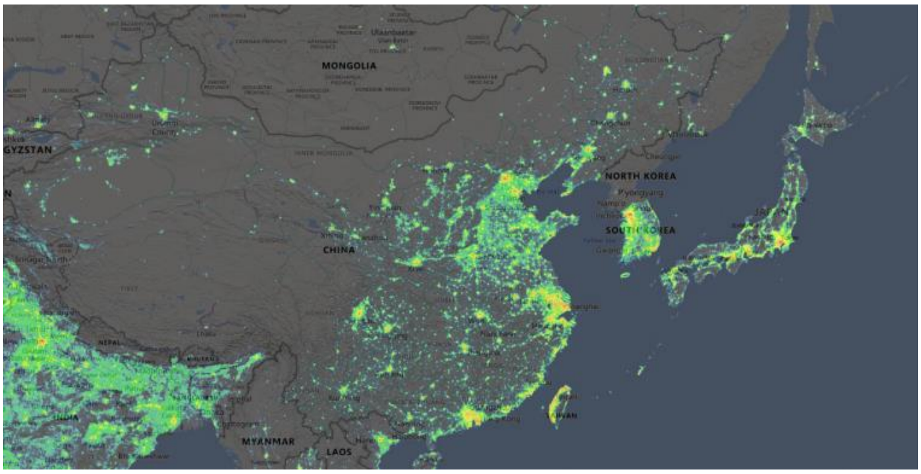

Figure 5: Visualization results of China region by Cluster analysis [10] (https://www.lightpollutionmap.info/)

The symbols in the figure4 represent the following regions in Table 4:

Table 4: The regions corresponding to the symbols.

Symbols1	Symbols2
NY -- New York	NYs sub -- Suburb of New York
BJ -- Beijing	SH -- Shanghai
SHs sub -- Suburb of Shanghai	GZ -- Guangzhou
CD -- Chengdu	SJY -- Sanjiangyuan
SX -- Shanxing County	

As can be seen from the figure above, the third principal component of SHs sub and SX is higher than that of other regions except SJY, and the first and second principal components also account for a certain proportion, so they are classified into the first category. NY, NYs sub, BJ, SH, GZ and CD have high first and second principal components, while the third principal component is relatively low, so they are divided into the second category. The third principal component of SJY is relatively high, and the first and second principal components almost have no effect on SJY, so SJY is classified into the third category[11].

Next, the actual parameters of 9 regions are substituted into the expression of FRL obtained by us, and the results are shown in the table 5 below:

Table 5: Parameters for 9 different regions

	IL	ZB	EI	IT	EC	GDP	UR	FRL
NY	400	130	20	13	8300	63000	100	93
NYs sub	240	87	40	10	8000	51000	96	83
BJ	350	100	40	12	5000	16000	85	70
SH	380	110	30	13	5400	22000	84	74
SHs sub	170	52	70	8	2900	11000	57	45
GZ	310	94	50	10	4500	18000	84	63
CD	280	88	50	10	3000	10000	76	55
SX	30	7	60	10	3500	21000	50	52
SJY	0.5	0.2	90	2	60	540	3	3

Compared with the FRL calculated from the above table and the existing AFL, it can be seen that the model well reflects the actual light pollution risk level, and the conclusion reached at last is consistent with the conclusion reached by the statistical analysis of the data: Protected land location and rural community are rated as light polluted, and suburban and urban community are rated as moderately polluted. Their scores increase in turn, and there are few areas with no pollution or heavy pollution. However, due to the special geographical conditions of the protected land, the paper do not rule out the problem of incomplete collection of relevant parameters or partial distortion of parameters.

4. Conclusions

The above results are mainly due to different development strategies in different places. For national central cities, such as New York, Shanghai and Beijing, economic development is the first priority, and the economic aggregate created by them is very important to the national economy. A large number of resources and talents gather together in such cities, creating a huge consumer market. The amount of light used by businesses and households is enormous.

Similarly, for the suburbs of Shanghai and Yamagata Prefecture in Japan, where economic development is not a priority, businesses and households will use less light.

Finally, for a nature reserve like Sanjiangyuan, the development goal is to ensure that the environment is not destroyed. The economic development here is relatively backward, and the consumption desire of residents is not strong. The life goal of residents here is to enjoy nature and ensure good health. As a result, light pollution indicators in such areas are very low.

Although the model may differ from the ideal model to lack of several domain statistics and errors, because there are some subjective factors that are difficult to quantify in light pollution, the model does not respond to the effects of these factors on light pollution, such as those on sleep quality. However, based on the above results, we can still conclude that formulating different policies for different regions can play a certain role in improving light pollution.

For Beijing, New York and other major cities, it is very important to improve people's awareness of light pollution. Try to save electricity without affecting their work. At the same time, different standards are set for enterprises of different sizes, and relevant reward and punishment measures are formulated. For areas where economic development is relatively less priority, the ecological environment is more complex, so relevant publicity and education should be carried out to ensure that the light pollution level in these areas is maintained at a low level.

References

[1] Kevin J. Gaston, Thomas W. Davies, Jonathan Bennie and John Hopkins. Reducing the ecological consequences of night-time light pollution: options and developments. Journal of Applied Ecology 2012, 49, 1256–1266.

[2] Xu Rui. Research on the status quo and prevention measures of urban light pollution [J]. Leather production and Environmental protection Technology, 2021,2 (16): 158-159.

[3] Zhang Yue, Xu Yongming, Xiong Wencheng, Li Xuwen, Jiang Sheng, Cui Jiayu, Wang Tiantian. Remote sensing monitoring and prevention measures of light pollution at night [J]. Environmental Monitoring and Early Warning, 2019,11 (05): 108-112

[4] Li Zhen, Huang Shenghua, Xu Haiwen, Wu Duxiong, Zhou Gang, Wu Chunhai. Research on the status quo of light pollution standards for urban nightscape [J]. China Lighting Appliance, 2020, (05): 7-11

[5] Wang Zhide, Yuan Jingyu, Yao Sheng, et al. Current status of light pollution studies by artificial illumination at night [J]. Journal of Lighting Engineering, 2021,32 (03): 94-99.

[6] Feng Kai, Hao Luoxi, Zeng Kun. Discussion on the evolution characteristics of light pollution in urban lighting Take VIIRS images as the analysis object [J]. Journal of Lighting Engineering, 2022,33 (06): 1-10.

[7] Luo Zhenqing, Su Hang, Wang Zhuoli, Yang Biao. Urban lighting planning and the hierarchical classification of dynamic light pollution in the night light environment [J]. Journal of Lighting Engineering, 2022, 33 (04): 193-202.

[8] Li Weishan. Observation method and spatial distribution characteristics of urban nighttime light pollution [D]. Dalian University of Technology, 2017.

[9] Li Jiayi, Xu Yongming, Cui Weiping, Wu Yuyang, Wang Jing, Su Boyang, Ji Meng. Nighttime light pollution monitoring in Nanjing based on Liaojia-1 night light remote sensing data [J]. Remote Sensing of Natural Resources, 2022, 34(02): 289-295.

[10] Hao Qingli. Digital observation and spatial model construction of multi-dimensional urban nighttime light environment [D]. Dalian University of Technology,2019.

[11] Yang Yu, Li Hongyan, Deng Zeyuan, et al. Based on principal component analysis method [J]. Food Journal of China, 2023, 23 (3): 196-205.