ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

Sustainable Management Strategies for Urban Light Pollution Based on Decision Tree and Time Series Modeling

Jianwei Yu[1,*], Ying Zhang[2], Minlian Fan[1], Guorui Zhao[1]

[1]*School of Computer Science and Engineering, Guangdong Ocean University Yangjiang Campus, Yangjiang, China*
[2]*School of Materials Science and Engineering, Guangdong Ocean University Yangjiang Campus, Yangjiang, China*
[*]*Corresponding author: 2889778284@qq.com*

Abstract: *This study focuses on sustainability issues related to urban light pollution. Utilizing a comprehensive evaluation model, created by integrating ExtraTrees and CatBoost decision tree models along with the ARIMA Time Series model, the study establishes management strategies under varying light pollution levels. Initially, 363 cities were sampled to identify factors affecting light pollution through the analysis of Night Sky Brightness (NSB). Subsequently, four representative regions within Guangzhou were empirically evaluated to affirm the model's robustness and accuracy. Lastly, interventions involving the modification of GDP growth, urbanization, and forest cover indices were simulated to assess their impact on light pollution levels. The findings indicate that targeted interventions can effectively mitigate light pollution.*

Keywords: *Light Pollution, ARIMA, Time Series, Intervention Strategies*

1. Introduction

In the era of technological advancements, artificial lighting has become ubiquitous in human life, contributing not only to convenience but also to a less-discussed issue—light pollution. While initial studies focused on the impact of light pollution on wildlife, emerging research underscores its far-reaching effects on human health and ecosystems. However, existing methods for measuring light pollution, such as satellite imagery or naked-eye observations, are limited in scope and accuracy. This study aims to address these gaps by simplifying influential factors to improve model accuracy, developing a comprehensive light pollution index and evaluation model with better generalizability, and applying this model to diverse settings—from nature reserves to metropolitan areas—to validate its efficacy. Additionally, the paper presents three practical strategies aiming to address the issue of light pollution while also considering economic benefits, environmental effects, and social requirements.

2. Comprehensive Light Pollution Index and Level Classification

This study employs data from 1419 cities in countries including the United States, China, Japan, and the United Kingdom to investigate the impact of both natural and non-natural factors on urban light pollution. Using 17 statistical indicators, such as GDP and elevation, culled from public reports and satellite imagery, the study narrows its focus to 363 cities in China. The choice of China as a case study allows for a more generalized light pollution evaluation metric, given its diverse economic conditions and levels of light pollution due to its large territorial extent and unequal resource distribution [1].

2.1 Indicator Correlation with Light Pollution Levels

The study employs Night Sky Brightness (NSB) as an indirect measure of artificial light intensity, serving as a proxy for assessing light pollution levels. An initial analysis was conducted to explore the relationships between NSB and 17 selected indicators. Preliminary results indicate that NSB shows a significant positive correlation with 12 indicators, such as GDP, Urbanization Rate, and Light Intensity Per Capita. Conversely, indicators like Forest Cover and Altitude exhibited minor negative correlations with NSB. Specifically, NSB tends to increase with a rise in GDP and decrease with higher altitudes.

ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

Among these, GDP exhibits the strongest influence on NSB variation in Figure 1.

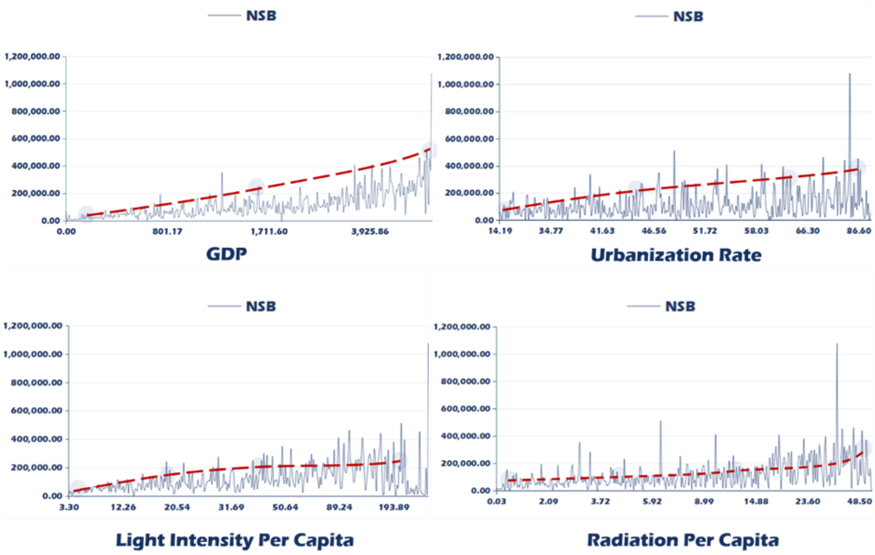

Figure 1: Relationship between NSB and Indicators

2.2 Weighted Analysis of Light Pollution Indicators

To comprehensively evaluate light pollution, this study applies the Lasso model for initial feature selection. The model effectively eliminates less influential variables, thereby mitigating the issue of multicollinearity among indicators. One notable finding is the exclusion of the 'Light Intensity Per Unit GDP' variable due to its high correlation with GDP, affirming the model's capability to home in on relevant features.

Subsequently, ten key indicators were identified based on established scientific literature for the study of light pollution. Advanced machine learning algorithms, including CatBoost and ExtraTrees, were employed to quantify the significance of each chosen indicator, further refining the evaluation model in Figure 2.

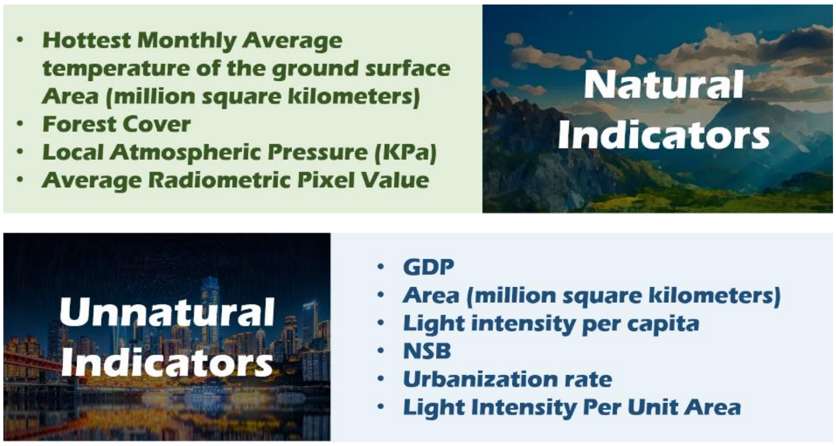

Figure 2: Two-level Indicators

2.3 Optimal CLPI Formation via TOPSIS Method

The Comprehensive Light Pollution Index (CLPI) is defined as a methodology to quantify the level of light pollution in each area through the following steps and formulas.

First, the standardized matrix is constructed. A matrix X_{ij} with n rows and m columns is constructed, and X in the matrix denotes the value of the j-th evaluation index of the i-th region.

In the second step, the gap between the 10 level II indicators and the optimal and inferior vectors is calculated.

$$D_i^- = \sqrt{\sum_{j=1}^{m} w_j (Z_j^- - z_{ij})^2} \ , D_i^+ = \sqrt{\sum_{j=1}^{m} w_j (Z_j^+ - z_{ij})^2} \tag{1}$$

Where, w_j is the feature importance of the jth indicator.

In the third step, the indicator data of the selected cities are substituted to derive the CLPI for the region.

$$CLPI = \frac{D^-{}_i}{(D^+{}_i + D^-{}_i)} \tag{2}$$

The CLPI was calculated for 363 cities in China, and the size of each local highlight was plotted on a map of China (partly) at the CLPI scale. Figure 3 shows light pollution and its CLPI score for only five regions. Among them, Taiwan, Beijing, Guangzhou, and Shanghai received the top four CLPI scores, while Shennongjia Forestry District received the lowest CLPI score.

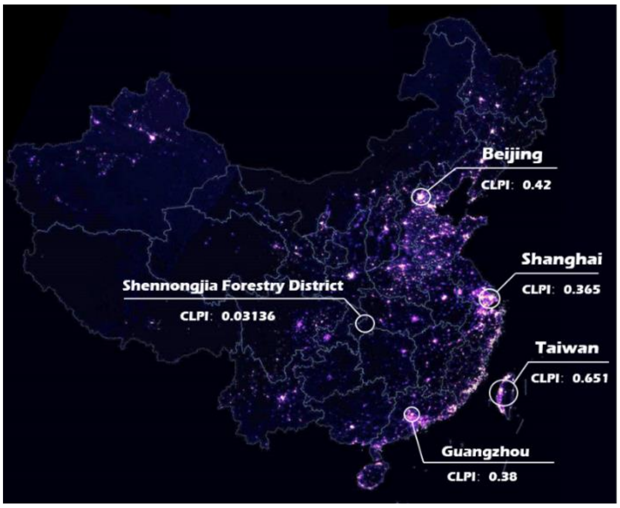

Figure 3: CLPI Scores for Selected Locations in China

2.4 K-Means-Based Light Pollution Level Categorization

To minimize the intra-classification gap and maximize the inter-class gap, the similarity between different influencing light pollution indicators will be explored to classify the sample set. This classification aims to minimize the intra-classification gap and maximize the inter-class gap. An analysis using unsupervised learning model clustering algorithms is considered for setting out different categories of cores based on the CLPI of different cities. Subsequently, the cities will be clustered into different categories based on the similar metric between CLPI and cores. This clustering will be used to classify the light pollution level [2].

By setting the criterion function for clustering, the k-value is dynamically changed in the process of clustering until the criterion function is the smallest, then the clustering is considered complete.

The sum of the squared distances from each sample point in the clustering set to the center of that cluster, and for the j-th cluster set, the criterion function is defined as:

$$J_j = \sum_{i=1}^{N_j} \left\| X_i - Z_j \right\|^2 , X_i \in S_j \tag{3}$$

Where S_j denotes the j-th cluster set. The center of the jth light pollution center is z_j; N_j is the number of cities contained in the jth light pollution level S_j.

For all k model classes, there are:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{N_j} \left\| X_i - Z_j \right\|^2 , X_i \in S_j \tag{4}$$

The jth level of light pollution center Z_j should be chosen so that the criterion function J is extremely

ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

small. It also means that the value of J_j is extremely small. This must be met and should $\frac{\partial_{J_j}}{\partial_{Z_j}} = 0$. Namely:

$$\frac{\partial}{\partial_{Z_j}} \sum_{i=1}^{N_j} \left\| X_i - Z_j \right\|^2 = \frac{\partial}{\partial_{Z_j}} \sum_{i=1}^{N_j} \left(X_i - Z_j \right)^T \left(X_i - Z_j \right) = 0 \tag{5}$$

The solution is $Z_j = \frac{1}{N} \sum_{i=1}^{N_j} X_i$. The cluster center of class S_j is the mean center of the light pollution level at level *j*.

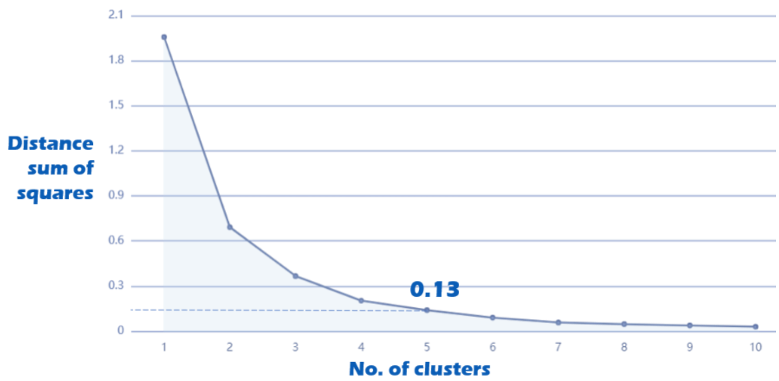

Figure 4: Elbow Method Diagram

As shown in Figure 4, when the number of clusters is 5, the slope tends to be smooth, and the intercept k=0.13, Therefore, the light pollution is divided into five categories, and the name of each category of light pollution level (LPI) and the determination interval of CLPI is shown in Table 1.

Table 1: Light Pollution Degree Division Interval

LPI	**Slight Pollution (SP)**	**Low Pollution (LP)**	**Moderate Pollution (MP)**	**Heavy Pollution (HP)**	**Severe Pollution (SeP)**
CLPI	0.056~0.104	0.104~0.175	0.175~0.345	0.345~0.641	0.641~∞
Representative Cities	Aba Tibetan and Qiang Autonomous Prefecture	Zhongshan	Nanjing	Beijing	Taiwan

3. Analysis and Categorization of Light Pollution in Diverse Urban Environments

3.1 Selection of Four Types of Areas and Basic Introduction

Among the several regions with high CLPI scores, considering that Guangzhou covers a large area, the LPI of different regions is highly differentiated. As shown in Figure 5, Tianhe District was selected as the urban representative, Conghua District as the suburban repre-sentative, Haizhu National Wetland Park as the protected land location representative, and Xiancun Town in Zengcheng District as the rural community representative [3].

ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

Figure 5: Four types of areas and initial perceived luminance levels

The study systematically analyzes light pollution levels in various regions of Guangzhou City, China, namely Tianhe District, Conghua District, Haizhu National Wetland Park, and Xiancun Town. Tianhe District is identified as a highly urbanized area with significant commercial activity, thus expected to exhibit elevated levels of light pollution. Conversely, Xiancun Town, due to its low population density and limited commercialization, is expected to maintain low light pollution levels. Conghua District and Haizhu National Wetland Park fall in between, with the former influenced by nighttime entertainment and tourism, and the latter being a protected area but still subject to potential light pollution due to tourism. The importance of various indicators is evaluated using CatBoost and ExtraTrees models, providing a scientific basis for future light pollution management.

3.2 CLPI and LPI and Rationalization Analysis for Four Types of Areas

Based on the comprehensive light pollution evaluation model developed in this paper, the empirical findings across distinct regions in Guangzhou align well with model predictions. Specifically, Tianhe District's elevated levels of light pollution can be attributed to its low forest cover and high performance on other key indicators. Conghua District, although less developed, shows significant levels of light pollution due to its closeness to Tianhe District and reliance on hot spring tourism. Despite being a protected natural park, Haizhu National Wetland Park manifests moderate levels of light pollution, a consequence of its commercial orientation. Lastly, the low light pollution levels in Xiancun Town are explained by its minimal urbanization rate and the dominance of residential and street lighting as the primary light sources, corroborating the model's assessments in Figure 6.

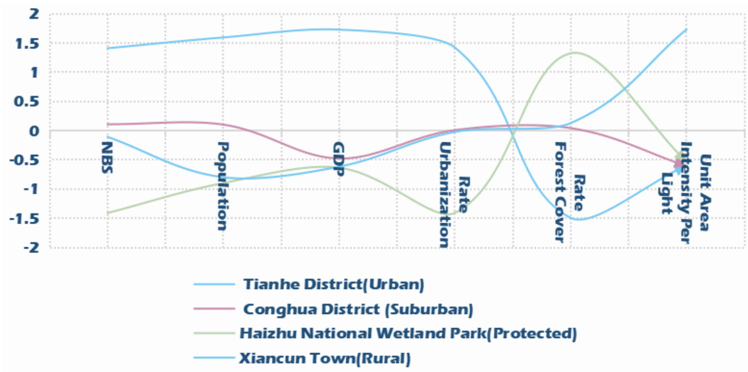

Figure 6: Comparison of Data for Six Indicators at Four Locations (Normalized)

4. Effectiveness of Intervention Strategies and Impact on Risk Levels

4.1 Building an ARIMA Time Series Model

In this study, the effectiveness of a proposed intervention strategy was validated by its application to two disparate locations, Guangzhou, and Harbin, chosen for their contrasting geographical and economic characteristics. Utilizing composite indicators of light pollution gathered over the past 20 years, future projections were generated for both cities over a seven-year period, both with and without intervention. The intervention strategy primarily targeted modifiable variables like GDP, Urbanization Rate, and

ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

Forest Cover, aligning with the study's earlier analytical framework to assess their impact on light pollution risk levels.

An ARIMA Time Series Model explores a particular time series. After a long-term change process, the statistical regularities that exist are studied and predicted. Following this, the model is built to extend the timeline and extrapolate the resulting statistical patterns to predict possible future developments and trends over a long timeline [4]. If you want to forecast the macroeconomic environment of China in the next five years, you need to build a time series model, and the steps to build it are as follows:

(1) Generate Sample Sets

Let the time series be $X(i), i = 1,2,3,\cdots,n$, X(t) constitutes the input vector and X(t+1) constitutes the output vector. The data were also subjected to standard normalization.

$$X = \frac{X-\mu}{\delta} \tag{6}$$

where μ is the meaning of the sample data and δ is the standard deviation of the sample data.

(2) Set Structure Parameters

The training set can determine the structural parameters. After the structural parameters are determined, the nodes of the hidden layer are set:

$$H = sqrt(m+n) + a \tag{7}$$

where H is the number of implicit layer nodes. m is the number of layer nodes input. n is the number of output layer nodes. a is the regulation constant.

(3) Model Evaluation Metrics

Let the predicted value be $\hat{Y}(i)$,The true value is Y(i), $i = 1,2,3,\ldots,n$,then:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid \hat{Y}_i - Y_i \mid \tag{8}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \mid \frac{\hat{Y}_i - Y_i}{Y_i} \mid \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2} \tag{10}$$

4.1 ARIMA Time Series Model Predicts Future Unintervened Indicators

Firstly, it is necessary to perform an ADF test on the time series with indicators of GDP, Forest Cover, and Urbanization Rate. Only when these two-time series pass the smoothness test can the model continue to be applied to the calculations.

The results show that the ADF test results of the time series for both days show significance (**P<0.05**), Then the time series with the three as indicators is proved to be a smooth time series.

Considering the intervention strategies implemented above, there will be an impact on the growth rates of the three-level II indicators. Therefore, the GDP Growth Rate and Urbanization Growth Rate are subtracted from Figure.7, and the growth rate of Forest Cover is added. The data were simulated for a total of seven years from 2023 to 2030 [5]. The data of the three level II indicators before and after the intervention were obtained as shown in Figure.7.

Region	Time Period	GDP_1	Forest cover$_1$ (%)	Urbanization Rate$_1$(%)	…	GDP_7	Forest cover$_7$ (%)	Urbanization Rate$_7$(%)
Guangzhou	Pre-intervention	26171.1	53.87%	82.40%	···	37973.3	55.98%	84.63%
	Post-intervention	25920.9	54.41%	81.60%	···	37616.6	56.54%	83.78%
Harbin	Pre-intervention	5,468.70	44.14%	52.15%	···	6,784.47	46.31%	56.22%
	Post-intervention	5416.2	44.58%	51.64%	···	6718.81	46.77%	55.66%

Figure 7: Comparison of Changes Indexes Before and After the Intervention

Where the value of α is determined based on the unit of growth rate and a reasonable selection within a certain range

ISSN 2616-5775 Vol. 6, Issue 11: 144-150, DOI: 10.25236/AJCIS.2023.061118

4.3 Comparison of Changes in Factors Before and After the Intervention

The effects of the above intervention strategies on CLPI were explored using the control variables method. This analysis aimed to identify the most suitable interventions for the region by examining the effects of each strategy individually.

Region	Time Period	Growth Rate$_1$	Growth Rate$_2$	Growth Rate$_3$	Growth Rate$_4$	Growth Rate$_5$	Growth Rate$_6$	Growth Rate$_7$	SUM
Guangzhou	GDP	-0.0007	-0.0044	-0.0047	-0.0052	-0.0057	-0.0056	-0.0012	-0.0275
	Urbanization Rate	-0.0029	-0.0071	-0.0062	-0.0044	-0.0034	-0.0033	-0.0032	-0.0306
	Forest Cover	-0.0269	-0.0163	-0.0088	-0.0064	-0.0088	-0.0177	-0.028	-0.1128
Harbin	GDP	-0.0085	-0.0092	-0.0089	-0.0109	-0.0126	-0.0121	-0.0107	-0.0104
	Urbanization Rate	-0.0016	-0.0016	-0.0017	-0.0023	-0.0024	-0.0021	-0.0017	-0.0019
	Forest Cover	-0.0015	-0.0031	-0.0045	-0.0048	-0.0038	-0.0025	-0.0016	-0.0031

Figure 8: Comparison of Impact Rates Before and After Intervention

As shown in Figure 8, the intervention of increasing Forest Cover was more effective for Guangzhou. As shown in Figure 9, with Forest Cover's interventions, the LPI level de-creases directly by one level in three of the next seven years; And for Harbin, interventions that use an increased share of green GDP. This helps the local area even more for the next four years within the next seven years and the LPI drops one level directly.

City	Before and after intervention	Grade$_1$	Grade$_2$	Grade$_3$	Grade$_4$	Grade$_5$	Grade$_6$	Grade$_7$
Guangzhou	Pre-intervention	LP	MP	HP	HP	SeP	SeP	SeP
	Post-intervention	LP	MP	MP	HP	HP	HP	SeP
Harbin	Pre-intervention	HP	HP	HP	HP	HP	HP	SeP
	Post-intervention	MP	MP	MP	HP	HP	HP	HP

Figure 9: Comparison of Light Pollution Levels Before and After Intervention

5. Conclusions

In conclusion, this study presents a robust and generalizable light pollution rating model by capitalizing on China's unique economic and geographical landscape. Methodological rigor is maintained using maximum likelihood estimation and the integration of multiple predictive models. Despite these strengths, the study acknowledges limitations in the lack of a unified night sky brightness evaluation framework and the selection of a limited set of indicators, potentially affecting the model's universality. For future research, the study aims to enrich the night sky brightness data set from 2020 to 2022 and investigate the potential correlation between light pollution levels and population mental health during pandemic lockdowns, thereby contributing to a more comprehensive understanding of light pollution's societal impacts.

References

[1] Xiao Qiaoli, Wang Yue. Correlation characteristics of nighttime lighting and GDP in regional economic research [J]. Remote Sensing Information, 2022, 37(03):42-50.
[2] Zhu Yong, Jiang Qiang, Guo Naiju. Research on structured light centerline extraction method based on K-means clustering [J]. Journal of Shenyang University of Technology, 2023, 42(05):40-46.
[3] Yang Leyi, Chen Jieying, Zhang Luziyu et al. A study on the sleep effect of nighttime light pollution based on NPP-VIIRS luminous data--Taking Guangzhou City as an example [J]. Tropical Geomorphology, 2021, 42(01):47-53.
[4] Tian Meixue. Application of ARIMA model in GDP forecasting in Shandong Province [J]. China Management Informatization, 2023, 26(01):139-141.
[5] Qian Y .Sustainable management of urban water resources in China[J].Sustainable and Safe Water Supplies, 2007, 7(2):23---30.DOI:10.2166/ws.2007.037.