Transactions on Environment, Energy and Earth Sciences
ISSN: 2960-1924, eISSN: 2960-2300 | Volume 3
GRET 2024

Light Pollution Evaluation Research Based on the Entropy Weight Method Combined with the TOPSIS Model

Haicheng Lin*

School of Management, Guangdong University of Technology, Guangzhou, China, 510520

* Corresponding Author Email: 355194976lhc@gmail.com

Abstract. Light pollution has emerged as a significant environmental issue, joining the ranks of other pollutants such as waste gas and wastewater in posing serious threats to environmental safety and public health. Unlike traditional pollutants, light pollution is often less visible in regulatory frameworks, yet its effects on ecosystems, human health, and overall environmental quality are profound and increasingly recognized. As urbanization expands and artificial lighting becomes more prevalent, the need to accurately measure and mitigate the impacts of light pollution becomes ever more critical. To tackle the complexities associated with assessing light pollution across diverse geographical locations, this paper introduces a comprehensive and versatile evaluation model. The proposed model integrates the TOPSIS method, enhanced by the entropy weight method, and is further supported by multiple regression analysis and random forest algorithms. Together, these methodologies form a robust framework for quantifying light pollution risk levels across various environments. By providing a systematic approach to light pollution assessment, this research offers valuable insights for the development of global light pollution evaluation systems. Furthermore, the findings contribute to the broader discourse on effective management strategies, offering a reference point for policymakers and researchers involved in environmental protection and sustainable urban planning.

Keywords: Multiple Regression; Evaluation Model; Machine Learning.

1. Introduction

Light pollution is a new source of environmental pollution following waste gas, wastewater, waste residues, and noise. It primarily includes white light pollution, artificial daylight pollution, and colored light pollution. The issue was first raised by the international astronomical community in the 1930s [1]. Nowadays, light pollution poses a significant threat to human life and health. It is crucial for humans to better understand these threats to conduct risk assessments and better regulate artificial lighting [2].

Current research on light pollution varies widely across countries. Traditional measurement methods focus on framework standards for specific areas, mainly using receivers to distinguish and measure skyglow, light trespass, and discomfort glare in light [3]. For light pollution in densely populated areas, which is often overlooked, researchers have used low-cost and efficient methods involving TinyML and wireless communication to measure light pollution. However, these approaches mainly focus on hardware solutions and lack a comprehensive policy framework [4]. In addition to traditional physical methods, some have used mathematical approaches (weighted processing of different factors) to analyze satellite remote sensing data, resulting in a more comprehensive light pollution risk assessment model and corresponding intervention strategies [5]. Since 2015, an increasing number of researchers have incorporated machine learning and deep learning methods into light pollution studies, such as using a multi-level light pollution assessment and mitigation framework based on TOPSIS and machine learning to quantify light pollution risk levels using multiple key indicators [6], or directly employing neural networks to construct a regional light pollution risk measurement model with corresponding evaluation indicators [7].

Currently, the measurement methods and standards in the scientific community remain inconsistent. However, progress has been made in strategies aimed at reducing specific indicators of pollution, such as using energy-efficient LED lighting and "dark infrastructure" to reduce light pollution, with considerations for lighting color, timing, location, and shading [8]. Other strategies have been developed from three simple dimensions: reducing lighting time, intensity, and effective light range, and discussing the impacts of these strategies on light pollution [9].

While these studies have proposed various evaluation and intervention strategies, they also have some limitations. For example, many models rely too heavily on specific regional and environmental data, lacking generalizability. Other studies have proposed detailed evaluation indicators, but in practice, they may face challenges with data acquisition and computational complexity [10]. Additionally, some strategies may be difficult to implement due to economic and technical constraints. Therefore, our research aims to strike a balance between comprehensiveness and practicality, proposing solutions that are both efficient and feasible.

Our research data primarily comes from the Luojia-1 satellite remote sensing images. After analyzing the general distribution of light pollution nationwide, this article selected several areas with high significance for concentrated data collection. The data visualization is shown in Figure 1.

Figure 1. General distribution of light pollution in China

2. The basic fundamental of Light pollution evaluation model

2.1. The fundamentals of AHP

The Analytic Hierarchy Process (AHP) is a multi-criteria decision-making method designed to simplify complex decision-making problems by breaking them down into multiple levels and elements in a hierarchical structure.

AHP is based on pairwise comparisons, where the relative importance of decision elements is compared to construct a judgment matrix. By calculating the eigenvalues and eigenvectors of these matrices, this article derives the relative weights of each element, helping decision-makers rank and select different options.

This article utilizes main data build the Judgment Matrix A.

$$A = \left(a_{ij}\right)_{n\times n} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

Then, each row vector of matrix A is geometrically averaged and then normalized, that is, each evaluation weight and feature vector W are obtained.

$$w_i = \bar{W}_i \Big/ \sum_{i=1}^{n} \bar{W}_i \, , W = \begin{Bmatrix} w_1 \\ \vdots \\ w_n \end{Bmatrix} \tag{2}$$

Then on, consistency checks are performed on the eigenvectors.

$$CI = \frac{\lambda max \ - \ n}{n \ - \ 1} \tag{3}$$

where n mean the dimension of the Matrix.

$$CR = \frac{CI}{RI} \tag{4}$$

When the consistency ratio CR<0.1, it can be considered that the degree of inconsistency of A is within the allowable range and has full consistency. After passing the consistency test, its normalized eigenvector can be used as a weight vector. This results in the calculated weight of the evaluation index.

2.2. The Determination of Using the TOPSIS Method Based on the Entropy Weight Method

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method used to rank alternatives. The entropy weight method is a technique for determining the weights of indicators by considering the correlation between indicators. Combining these two methods forms the entropy weight-based TOPSIS model.

The calculation process mainly involves:

(1) Data Standardization:

$$r_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{5}$$

where r_{ij} is the standardized decision matrix element, x_{ij} is the original decision matrix element, and mmm is the number of alternatives.

(2) Calculate the Entropy Value of Indicators:

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

where e_j is the entropy value of the j-th indicator.

(3) Calculate the Weights of Indicators:

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

where w_j is the weight of the j-th indicator, and n is the number of indicators.

(4) Construct the Weighted Decision Matrix:

$$v_j = w_j \times r_{ij} \tag{8}$$

where v_j is the weighted decision matrix element.

(5) Calculate the Distance Between Alternatives and the Ideal Solution:

$$d_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2} \tag{9}$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2} \tag{10}$$

where d_i^* and d_i^- are the distances between alternative i and the positive and negative ideal solutions, respectively, and v_j^* and v_j^- are the values of the j-th indicator in the positive and negative ideal solutions, respectively.

(6) Calculate the Relative Closeness:

$$C_i = \frac{d_i^-}{d_i^* + d_i^-} \tag{11}$$

where C_i is the relative closeness of alternative i.

(7) Comprehensive Scoring and Ranking:

The ranking of alternatives is based on their relative closeness C_i, with higher closeness indicating a higher ranking.

2.3. The fundamental of Random forest model used to evaluate feature importance

The Random Forest model, a widely used ensemble learning method, is particularly effective for evaluating feature importance in predictive modeling. It operates by constructing a multitude of decision trees during training and outputting the mode of the classes for classification tasks or the mean prediction for regression tasks.

Feature importance in a Random Forest is determined by calculating the decrease in Gini impurity (for classification) or the reduction in variance (for regression) as the tree nodes split based on a particular feature. The importance score for each feature is then averaged over all trees in the forest. Mathematically, if ΔG_irepresents the decrease in Gini impurity from splitting on feature i, then the importance of feature i is given by:

$$Importance(i) = \frac{1}{T} \sum_{t=1}^{T} \Delta\, G_i^t \tag{12}$$

where T is the total number of trees in the forest. This metric provides an estimate of each feature's contribution to the model's predictive accuracy, allowing for a comprehensive analysis of feature significance.

3. Results

3.1. The establishment of the evaluation model of AHP Method

The first step of our research was to reduce the dimensionality of potentially correlated factors using the Principal Component Analysis (PCA) algorithm. This process involves mapping n-dimensional features to k-dimensions, where the k-dimensional features are reconstructed from the original n-dimensional features. The new orthogonal features generated during this process become the primary components for the next stage of analysis.

This article used MATLAB to centralize the raw data, calculate the covariance matrix of the data matrix, and then obtain the eigenvalues and eigenvectors of the covariance matrix. By employing eigenvalue decomposition and using the ‘eig’ function, this article performed PCA and calculated the variance contribution rate to determine the final dimensionality reduction. The result of variance contribution of factor is shown in Figure 2.

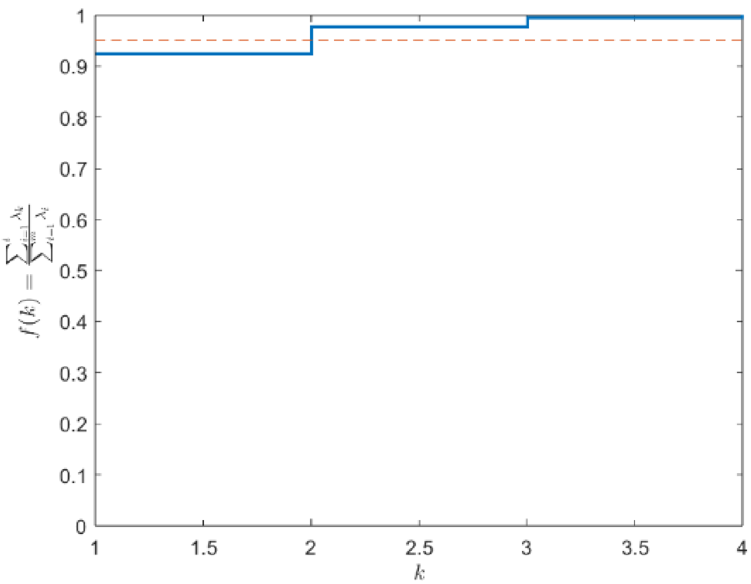

Figure 2. Variance Contribution of Factor Numbers

Generally, we select the smallest number of factors $f(r)$, where the variance contribution rate $\geq$ a certain threshold, so as to lose no more than 5% of the variance. Therefore, we retained three factors as secondary indicators and assigned appropriate weights to the other factors based on their contribution rates to the evaluation index.

The above graph shows the data point lattice of the top two factors with higher eigenvalue rankings in coordinate space.

We determined that the light pollution evaluation index ranges from 0 to 100, with higher values indicating higher light pollution risk. Initially, we used the Analytic Hierarchy Process (AHP) to determine the weights of various factors in the light pollution evaluation levels.

By calculating the consistency ratio of the equation, when $CR < 0.1$, it is considered that the inconsistency of A is within an acceptable range, with satisfactory consistency. After passing the consistency check, the normalized eigenvector can be used as the weight vector, thus deriving the calculation weights of the evaluation index. This completed the establishment of the first evaluation model.

3.2. The optimization of the evaluation model by TOPSIS method

After determining the weights of the first set of indicators, we used the light pollution levels from various regions (as cited) as a data validation set to verify the accuracy of the AHP-derived evaluation model. Here, we used the Kappa coefficient to perform a preliminary check on the results. The Kappa coefficient is mainly calculated based on the confusion matrix and is expressed as:

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \cdots a_C + b_C}{n \times n} \tag{13}$$

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{14}$$

where the sum of correctly classified samples p_0 is divided by the total number of samples, representing overall classification accuracy.

Assume that the actual number of samples in each category is a_1, a_2 , a_3 ... a_C,and the predicted number of samples in each category is b_1, b_2 , b_3 ... b_C, with a total sample number n.

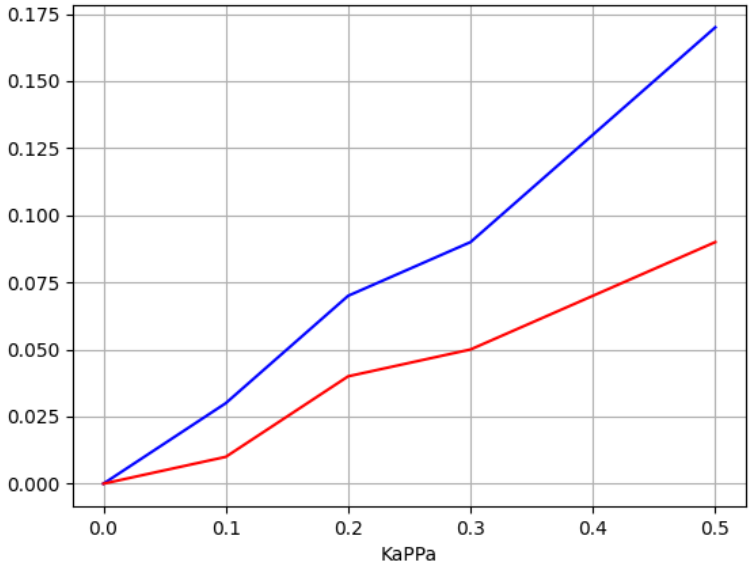

Figure 3. Accuracy Validation Results

Accuracy validation results are shown in Figure 3. As shown, the current fit is average, with an accuracy of 0.71, indicating room for improvement.

After re-evaluating the differences between the AHP model results and the data in the validation set, we found that the results for samples with higher original indicator data, i.e., more severe light pollution, tend to be weaker. This is because the weights assigned to the indicators in the AHP model are not sufficiently appropriate, leading to increased systematic error when the data is larger.

Based on this, we decided to correct the original model using the TOPSIS method based on the entropy weight method.

After recalculating the weights, we again used the Kappa coefficient for validation:

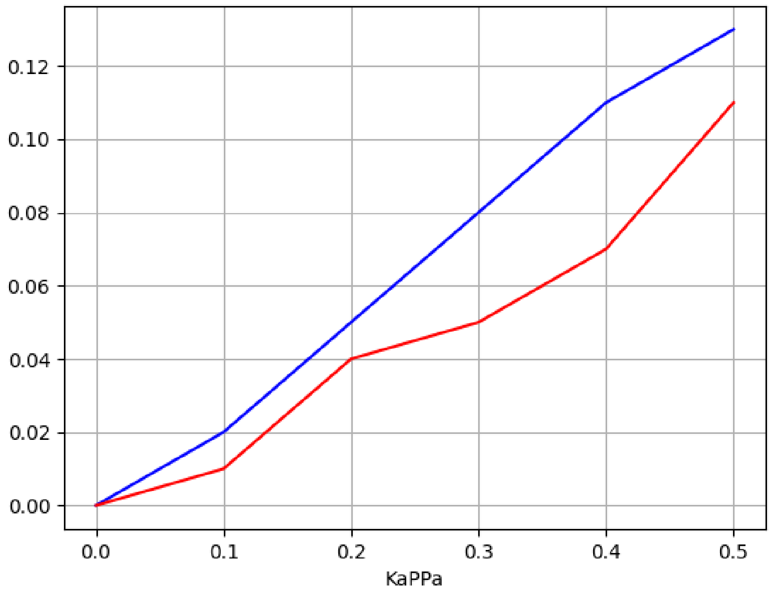

Figure 4. Accuracy Validation Results After Model Correction

Accuracy validation results after model correction are shown in Figure 4. As shown, the fit is better, with accuracy reaching 84%.

3.3. The Analysis of experimental results by random forest model

In our application of the light pollution evaluation model, we first collected data from a geopolitically significant area to serve as a focal point for analysis.

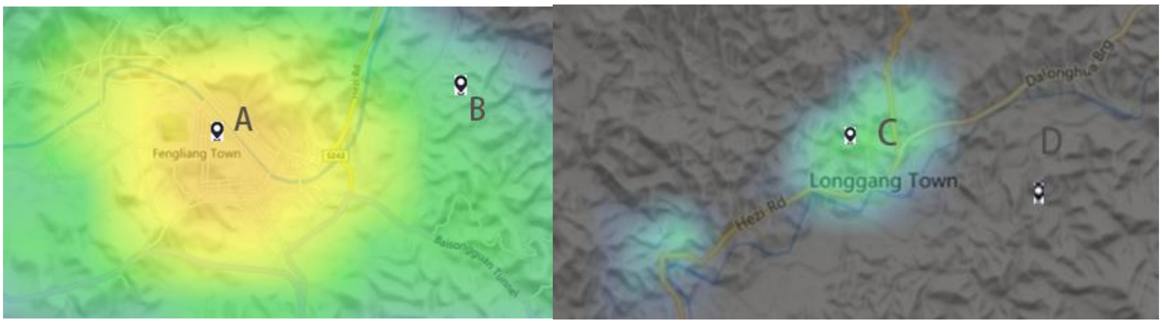

Figure 5. Marking points on the map

The data points labeled A, B, C, and D which shown in the Figure 5 correspond to an urban community, a suburban community, a rural community, and a protected area, respectively.

We applied the light pollution evaluation metrics from the previous section to assess the risk at couple locations. The light pollution evaluation index ranges from 0 to 100, with higher values indicating greater risk. We randomly selected ten samples from the locations in the previous section and used a multiple regression model to linearize and fit the secondary indicators and rating parameters.

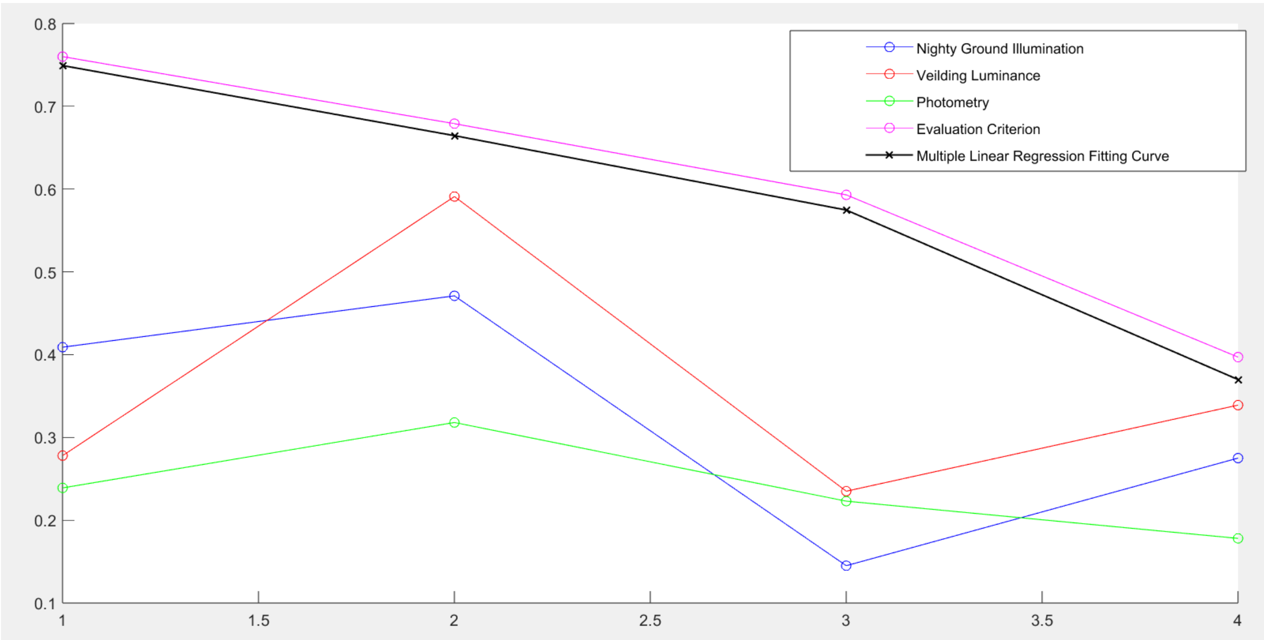

Figure 6. Fitting curve at four points

Fitting curve at four points is shown in Figure 6. The original data fit achieved an R-squared value of 91.094%, indicating that the fitted curve can roughly predict the effectiveness of the interventions.

We quantified the changes in the secondary indicators of the three strategies and predicted their outcomes, which are summarized in the Table 1:

Table 1. The impact of basic countermeasures on pollution

	R1	R2	R3	Forecast reduction percentage
Reduce lighting time and intensity	0.55	0.05	0.25	27.31%
Change lighting type	0.2	0.35	0.4	21.96%
Use shading device	0.3	0.05	0.3	16.29%

To assess the relative importance of factors, we used a random forest model. We trained the model with the original secondary indicator data and corresponding light pollution levels, then computed the feature importance based on the contribution of each decision tree to the model's predictive power.

Table 2. Feature Importance chart

Feature name	Feature importance
Nighty Ground Illumination	45.40%
Photometry	22.70%
Veiling Luminance	31.80%

Table 2 shows which factors were most influential in predicting light pollution levels. This analysis helps to refine our understanding of which interventions will be most effective in specific scenarios.

4. Conclusions

The vast amounts of light pollution data obtained from satellite remote sensing provide a foundation for establishing a trend evaluation model. However, classic evaluation methods are unable to handle the enormous time and computational resources required. The issue of overfitting in large sample sets can affect the accuracy of subsequent predictions.

In this paper, we initially reduced the dimensionality of the indicators using Principal Component Analysis (PCA) and Analytic Hierarchy Process (AHP) methods, based on practical considerations. When weighting the indicators, we chose the Topsis model based on the entropy weight method, which has lower dependency on sample size and uses only objective factors. By excluding subjective factors, this model is more stable during the evaluation process.

However, since a linear regression model was used to predict the potential impact of intervention strategies, the prediction progress is lower compared to predictions made by machine learning models. Additionally, the definition of light radiation varies by region, leading to data discrepancies and result deviations.

The evaluation of light pollution indicators is expected to evolve significantly as the field of environmental science continues to address this emerging concern. Future research will likely focus on refining existing methodologies and developing new models that are capable of capturing the dynamic and complex nature of light pollution. One potential direction involves the integration of advanced machine learning algorithms, such as deep learning models, to enhance the predictive accuracy of light pollution assessments. These models can process vast datasets from satellite imagery, ground sensors, and urban infrastructure to provide more precise and localized predictions of light pollution levels.

The feasibility of implementing advanced light pollution evaluation models will depend on several factors, including technological infrastructure, data availability, and stakeholder engagement. As digital infrastructure continues to expand globally, the feasibility of deploying sophisticated models in both developed and developing regions will improve. Ensuring that data collection methods, such as satellite monitoring and ground-based sensors, are standardized and widely accessible will be key to supporting these applications.

References

[1] Mander S, Alam F, Lovreglio R, et al. How to measure light pollution—A systematic review of methods and applications [J]. Sustainable Cities and Society, 2023, 92: 104465.

[2] Cao M, Xu T, Yin D. Understanding light pollution: Recent advances on its health threats and regulations [J]. journal of environmental sciences, 2023, 127: 589-602.

[3] Brons J A, Bullough J D, Rea M S. Outdoor site-lighting performance: A comprehensive and quantitative framework for assessing light pollution [J]. Lighting Research & Technology, 2008, 40 (3): 201-224.

[4] Botero-Valencia J S, Barrantes-Toro C, Marquez-Viloria D, et al. Low-cost air, noise, and light pollution measuring station with wireless communication and tinyML [J]. HardwareX, 2023, 16: e00477.

[5] Li X, Lu W, Ye W, et al. Enhancing Environmental Sustainability: Risk Assessment and Management Strategies for Urban Light Pollution [J]. Sustainability, 2024, 16 (14): 5997.

[6] Duan S, **ong Y, Yang R. Integrated Framework for Light Pollution Assessment and Mitigation: A Multifaceted Approach with TOPSIS and Machine Learning [J]. Highlights in Science, Engineering and Technology, 2024, 98: 164-174.

[7] Ma Y, Song Y, Zhang R. Research on regional light pollution risk level measurement based on neural network model [J]. Academic Journal of Environment & Earth Science, 2023, 5 (5): 41-45.

[8] Falchi F. Light pollution [J]. Urban pollution: Science and management, 2018: 147-159.

[9] Luo Y. Research on Light Pollution Risk Level Evaluation and Intervention Strategies [J]. Environment, Resource and Ecology Journal, 2023, 7 (5): 40-50.

[10] Mohamed A M O, Paleologos E K. Sustainable pollution assessment practices [M]//Pollution assessment for sustainable practices in applied sciences and engineering. Butterworth-Heinemann, 2021: 3-42.