Received: 18 March 2024 | Revised: 6 September 2024 | Accepted: 25 September 2024 | Published online: 30 September 2024

RESEARCH ARTICLE

Archives of Advanced Engineering Science
2024, Vol. 00(00) 1–9
DOI: 10.47852/bonviewAAES42022862

BON VIEW PUBLISHING

A Comprehensive Evaluation Method for Environmental Pollution in Tourist Attractions Based on Improved Principal Component Analysis

Shridhar Kedar[1,*], Nilesh Kolhalkar[1] and Prerna Mishra[2]

[1]*Department of Mechanical Engineering, MKSSS's Cummins College of Engineering for Women, India*

[2]*School of Mechanical and Manufacturing Sciences, JSPM University Pune, India*

Abstract: In order to address the problem of the reliability of current environmental pollution assessment methods, a comprehensive assessment method of environmental pollution in tourist scenic spots based on improved principal component analysis was proposed. First, air, water, and soil pollution for data tourist attractions in the top 5 metropolitan cities of India are collected. Then, the original pollution data are cleaned, the data clustering method is used to classify the pollution data, and the comprehensive evaluation index system of environmental pollution in tourist attractions is established. Finally, the comprehensive evaluation method is implemented to calculate the evaluation values of various indicators, and the comprehensive assessment results of pollution at tourist attractions are obtained. The experimental results show that the correlation of the evaluation index is always controlled above 0.90, the convergence rate of the method is fast, the optimal solution can be approached within 150 iterations, and the application effect is good.

Keywords: tourist attractions, comprehensive evaluation of pollution, principal component analysis, weighted assignment

1. Introduction

With the development of the economy, people's leisure lives have gradually shifted from traditional watching movies and gatherings to outdoor activities, tourism, and other activities. As a result of this, the development of the global tourism industry has been significantly improved. Furthermore, the findings indicate that the effect of air, water, and soil pollution on tourism income, tourist flow, tourist expenditure per capita, and duration of stay is non-linear. Pollution in the tourist business occurs in various forms, including increased emissions related to transportation and a higher requirement for electricity, and solid waste. The public was involved in the co-design of efficient solutions to air pollution problems through the use of a citizen science approach in which a real-time live air pollution data tool was developed [1]. Soil and food ecosystems are diversely affected by heavy metals and effects on human health and ecosystems through biologically cumulative [2]. Implementation of effects on the Internet of Things with the particulate matter network (PMN) of sensors which are economical with large numbers and with high spatiotemporal resolution and collection frequency with flexible distribution points was established for tracking and locating the air pollution sources [3]. At the same time, the development of the tourism industry has also exacerbated environmental pollution, such as noise pollution, water pollution, and solid waste pollution, causing a series of environmental problems in tourist attractions. These problems not only affect the local ecosystem but also constrain the further development of the tourism economy.

Almost the entire world population (99%) breathes air that violates WHO air quality regulations, which endangers their health. A record number of over 6000 cities in 117 countries are now monitoring air quality, yet residents continue to breathe harmful amounts of small particulate matter (PM) and nitrogen dioxide, with low- and middle-income nations suffering the most. A large number of death counts were reported (Figure 1) due to the emission of various air pollutants in millions of tons in the atmosphere (Figure 2) for the last 45 years. The reported death rate due to outdoor air pollution in 30 years where India has ranked second after China (Figure 3) therefore, more and more scholars believe that protecting and improving the environment of tourist attractions, improving environmental quality, and achieving sustainable development of tourist attractions are the only directions to achieve long-term development of the tourism industry.

Measurement data of realistic utilization of photochemical air are collected which ushers in a new age of regulatory applications at various geographical scales. It reduces inherent uncertainties and provides a realistic portrayal of atmospheric dynamics [4, 5]. The environmental issues in tourist attractions are a complex issue that involves multiple sources of pollution and environmental factors. The comprehensive evaluation method can comprehensively consider various factors, thereby more accurately reflecting the environmental conditions of the scenic area. Therefore, taking heavy metal pollution in scenic soil [6], after collecting soil surface samples, the content of 8 heavy metal ions and soil pH values was analyzed. Then, a combination of correlation analysis and the

*Corresponding author: Shridhar Kedar, Department of Mechanical Engineering, MKSSS's Cummins College of Engineering for Women, India. Email: shridhar.kedar@cumminscollege.in

Figure 1
Absolute number of deaths reported from particulate air pollution in 25 years

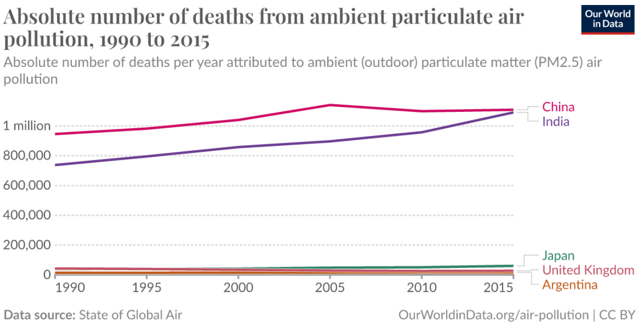
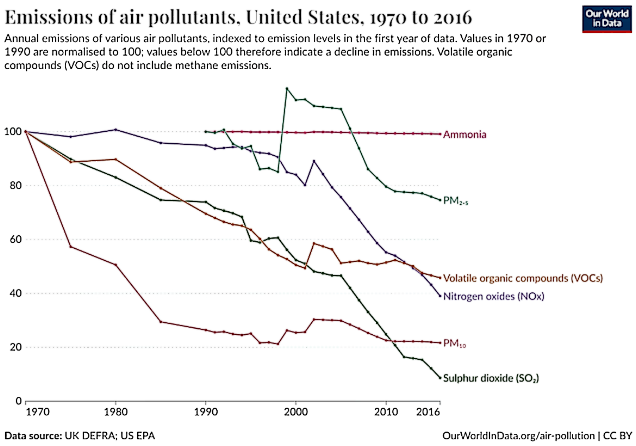
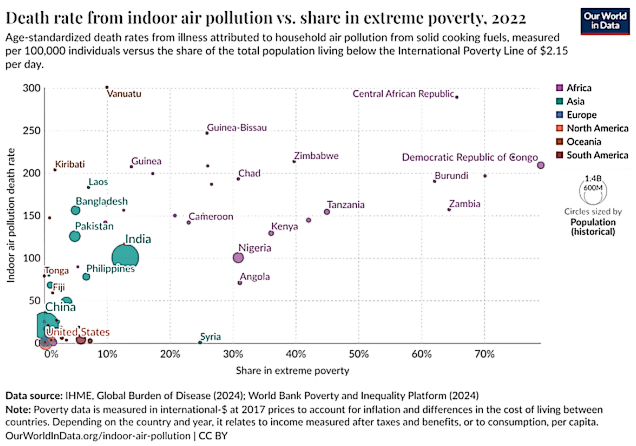


Figure 2
Emission of various air pollutants in million tons for 45 years



(Source: Our World in Data/outdoor-air-pollution)

Figure 3
Reported death rates due to outdoor air pollution in 30 years



positive definite matrix factorization model was used to share the pollution load index and potential ecological risk index, summarize the content and spatial distribution characteristics of heavy metals in surface soil, and carry out pollution evaluation. Taking groundwater pollution as an example, after selecting relevant evaluation indicators, the objective and subjective weights of the indicators were determined using the analytic hierarchy process and the entropy weight method, respectively. Then, a game theory aggregation model was used to calculate the comprehensive weights of the indicators. Finally, a risk assessment model was constructed to address the differences in conditions within the scenic area site. The emission of pollutants is the direct cause of environmental quality deterioration [7]. The entire method collected data on the discharge of solid waste from scenic area wastewater and used the factor analysis method to start from the direct reasons affecting environmental quality. Farmland, as a federally protected land resource, is essential in agriculture productivity and food safety, making soil quality and environmental health critical. Thus, researching the level of soil heavy metal contamination in farmland is very important for understanding the growing environment of food crops and conserving agricultural land resources [6]. The dual non-point source model was developed by the China in which it has been observed that the catchment area's farmlands and rural and urban residential areas pose the greatest danger of non-point source contamination to soil and natural water resources [7, 8]. The main factors affecting environmental quality were analyzed and summarized, and based on this, evaluation indicators were selected according to the main influencing factors to conduct a comprehensive evaluation of the pollution status. High-resolution numerically backward and forward tracer experiments were conducted concerning air pollution [9] and marine oil pollution in Indonesian waters [10, 11]. However, in practical applications, it has been found that the traditional methods mentioned above do not comprehensively and comprehensively analyze existing pollution data, and the selected evaluation indicators have poor correlation, resulting in low reliability of evaluation results. Figure 1 shows the absolute number of deaths reported due to particulate air pollution in 25 years. It mainly shows that as compared to Japan, the United Kingdom, Argentina and a more absolute number of death due to air pollution in India and China.

Due to the important significance of environmental quality and tourism economic development in tourist attractions, a thorough and scientific analysis of the current level of environmental quality in tourist attractions is conducted. The data are collected from the various pollution controlling and measurement authorities in India that measure the level of pollution across the most polluted cities of the country. The PM data are collected for the Capital City of India, i.e., Delhi city, and the frequency distribution of PM 1, PM 2.5, and PM 10 values is reported in Figure 4. The air quality data in the top 5 metropolitan cities of India are also collected, and maximum air pollution is reported in Delhi with an average value of 121 in comparison to other cities of the country (Figure 5).

Principal component analysis (PCA) is useful in a variety of domains, including population genetics, microbiome investigations, and atmospheric science. It is an analytical approach that includes identifying the linear combination of a set of variables with the greatest variation and reducing its influence, then repeating the process. When applying PCA, the first principal component of a collection of variables is the derived variable, which is generated by linearly combining the original variables and explains the most variation. The primary components are a linear arrangement of the initial variables in the dataset, arranged in decreasing order of importance. This study applies improved PCA technology and proposes a new comprehensive evaluation method for environmental pollution in tourist attractions. It is predicted that this method can offer additional theoretical references for the comprehensive evaluation of environmental pollution in tourist attractions. The specific content of this method is set as follows:

Figure 4
Frequency distribution of PM 2.5 value across Delhi and Pune

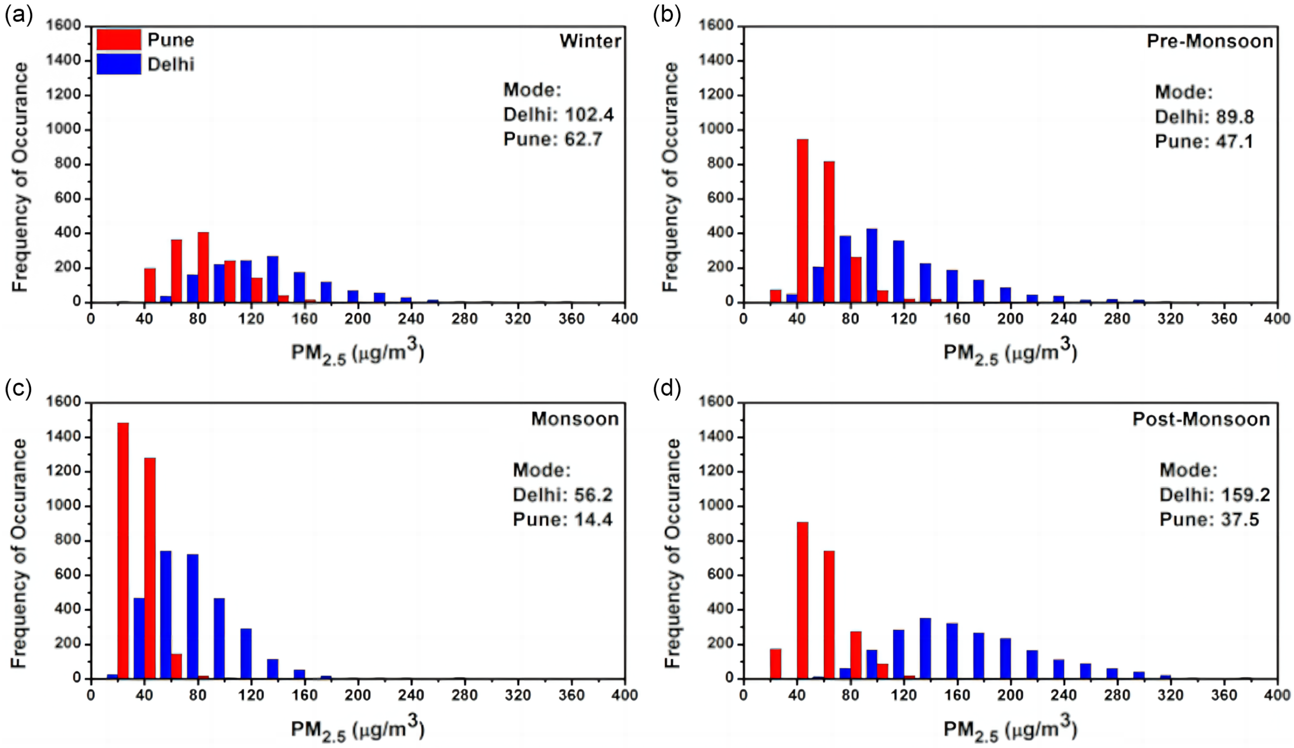
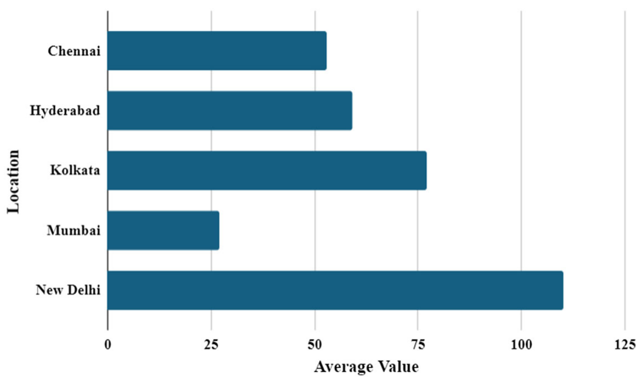


Figure 5
Air quality in the top 5 Metropolitan Cities of India with the most reported pollution in New Delhi



1) Collect air pollution data, water pollution data, and soil pollution data from tourist attractions separately, and store them uniformly on the data platform to provide a data foundation for subsequent processes.
2) Clean the raw pollution data collected above, remove duplicate values from the data, and then classify the sample similarity and similarity characteristics of the pollution data through clustering processing.

Considering that, there may be non-linear relationships between different indicators in the evaluation of environmental pollution in tourist attractions and conventional PCA methods are difficult. In these non-linear relationships, which may lead to distorted results, the PCA method is improved through weighted assignment. Established PCA determines a main vector that optimizes the summation of the second powers of principal components [12]. By introducing a weight matrix, the analysis of indicator correlation is supplemented, and the influence of the main indicator components is highlighted. The principal components selected from this are the available evaluation indicators, and then, a comprehensive evaluation index system for environmental pollution in tourist attractions is established.

After establishing an evaluation index system, the entropy method is used to calculate the evaluation values of various indicators in the index system. After layer-by-layer aggregation, the final comprehensive evaluation result of pollution in tourist attractions is obtained, and the evaluation results are classified into levels. Hyperspectral images were used to extract vegetation reflectance from polluted and control controlled. A sturdy correlation was reported between total petroleum hydrocarbons concentrations in soil and ground-assessed data, indicating that oil pollution changed pigment material in vegetation growing on impacted transects [13, 14].

Most of the earlier researchers have studied various analytical, experimental, and numerical investigations of environmental pollution. They mainly presented their findings worldwide. In the current study, we have mainly focused on environmental pollution data of five metropolitan cities of India and predicted that the comprehensive evaluation method should be implemented in this analysis.

2. Collect Environmental Pollution Data from Tourist Attractions

Due to the complex structure and extensive content of environmental pollution data in tourist attractions, in order to better complete the data collection work, the collection of environmental pollution data in tourist attractions is divided into three parts: air pollution data collection, water pollution data acquisition, and soil pollution data gathering.

2.1. Air pollution data collection

At present, the data obtained from the air pollution monitoring station in the location of the tourist attraction will be used as a sample of air pollution data for this study. Due to the large sample size of air pollution data, direct use will increase the difficulty of data processing in subsequent stages [15–19]. Therefore, first, monitor the air quality of tourist attractions and export all obtained air pollution data. The data are divided and processed by monitoring point equipment numbers. Then, divide the data of all monitoring points into days. In addition, set the sample data as daily atmospheric monitoring values. Arrange and combine them to obtain an air pollution dataset.

After the dataset construction is completed, the data are normalized to obtain the original air pollution data sample.

The normalization process is as follows:

$$\mathbf{X} = \frac{\mathbf{X_{AO}} - \mathbf{minX_{AO}}}{\mathbf{maxX_{AO}} - \mathbf{minX_{AO}}} \tag{1}$$

Among them, X_{A0} represents the original air pollution data of tourist attractions; minX_{A0} and maxX_{A0} represent the minimum and maximum values in the dataset, respectively [20]. By using Equation (1), the original air pollution data of tourist attractions is transformed into the range of (0, 1), thereby completing the normalization processing of the data.

2.2. Collection of water pollution data

In this study, the collection of water pollution data in the tourism environment was divided into two parts: sensor collection and data transmission. In the water area of tourist attractions, water quality monitoring sensors are used to collect data on pH value, dissolved oxygen, total nitrogen, total phosphorus, etc. For the convenience of data collection and organization, the collection cycle is set to collect water body data every 5 working days [21, 22]. After data calibration, the collected data are sent to the server or cloud platform through wireless communication or wired transmission, stored in a fixed data format for water body data and backup. The calibration results of water pollution data are as follows:

$$\mathrm{X} = \frac{\mathrm{X}_{WO} - \mathrm{Q}_W}{\gamma} \tag{2}$$

Among them, X_{W0} represents the sensor reading, Q_W represents the blank reading, and γ represents the sensor calibration coefficient.

2.3. Soil pollution data collection

Collect basic geological information about tourist attractions, including key data such as their geographical location, area, and main attractions, and determine the targets for soil pollution data collection, such as detecting the types, concentrations, and spatial distribution of pollutants [13, 23]. Select soil pollution data collection points based on terrain and topography. Based on the type of target pollutants, organize and analyze the detection data, calculate the pollutant concentration and exceedance multiples at each sampling point, and based on the calculation results, to visually display the soil pollution situation, establish a spatial distribution map of pollutants as follows

$$\mathrm{X_s} = \frac{\sum \mathrm{f(i,j)} * \omega(\mathrm{i,j})}{\sum \omega(\mathrm{i,j})} \tag{3}$$

Among them, XS represents the concentration value of soil pollutants at coordinates (i, j); f (i, j) represents the observed concentration of soil pollutants at coordinates i,j; represents the weight at i,j [12].

After organizing the data collected from the above three parts in the specified format, they will be uniformly stored in the comprehensive evaluation data platform for environmental pollution in tourist attractions, providing a data foundation for subsequent stages [24].

3. Preprocessing of Environmental Pollution Data in Tourist Attractions

Due to the presence of a substantial number of duplicate values in the original data, a data-cleaning step is required. The data-cleaning process in this study is set as follows:

- Firstly, detect and analyze the data from the original pollution data source and develop a data-cleaning plan.
- Then, execute a cleaning plan to clean various issues in the data source. Based on the current data characteristics, remove redundant data from the data and interpolate data gaps to ensure data integrity [25, 26].
- Finally, after executing the cleaning plan, the data that meet the requirements after cleaning will be returned to the data source.

Based on the diversity of polluted data, clustering algorithms were used to classify the sample similarity and similarity characteristics of the original polluted data based on data cleaning. Furthermore, the clustering processing of the data was further completed based on the degree of familiarity between the polluted data [27–29].

Given the original pollution data set X = {X_A, XW, XS}, C_N represents the basic features of the pollution data, and the number of features is set to N. Perform cluster analysis on data features, choose *k* initial cluster centers, and appoint each sample to the category for closely located cluster center. For each category, *I* calculate the average or other representative statistic of all samples in that category and use this statistic as the new cluster center. Then reassign each sample to the category where the nearest cluster center is detected. After each iteration, count the number of samples in class *i*. If P_i is set to represent the number of samples included in class *i*, then:

$$\mathrm{P} = \frac{\mathrm{c}_i * \mathrm{r}_i}{\mathrm{X}} \tag{4}$$

Among them, i ∈ N, ri represents the average value of the data sample.

Assuming k_i represents the clustering center of the *i*-th type of pollution, the clustering result of the *i*-th type of pollution data can be expressed as:

$$\mathrm{Q_i} = \mathrm{P}_i \frac{\sum \mathrm{r}_i \ \mathrm{k}_i}{m} \tag{5}$$

Among them, *m* represents the number of cluster centers.

4. Constructing a Comprehensive Evaluation Index System Using Improved PCA Method

After processing the environmental pollution data of tourist attractions, this study considers improving the PCA method to construct a comprehensive evaluation index system for environmental pollution in tourist attractions. The conventional PCA method is a commonly used multivariate statistical analysis method, which mainly focuses on the covariance and variance structure of the data, ignoring the correlation between indicators [28, 29]. However, in the evaluation of environmental pollution in tourist attractions, there may be non-linear relationships between different indicators, such as the concentration of certain pollutants and the intensity of tourism activities. Conventional PCA methods make it difficult to handle effectively these non-linear relationships, which may lead to information loss and distorted results. Therefore, this study improves the PCA method through weighted assignment, introduces a weight matrix to supplement the analysis of indicator correlation, and highlights the influence of main indicator components, making the evaluation results more accurate and reasonable.

If the preprocessed pollution data are set as $X_1, X_2, \ldots, X_N$ according to categories, and the pollution data of each category are independent of each other, the original pollution indicators can be expressed in the following form:

$$IX_1 = \mathrm{e}_1\mathrm{su}_1 * \frac{\mathrm{X}_1}{\mathrm{Q}_i}$$

$$IX_2 = \mathrm{e}_2\mathrm{su}_2 * \frac{\mathrm{X}_2}{\mathrm{Q}_i}$$

$$IX_N = \mathrm{e}_N\mathrm{su}_N * \frac{\mathrm{X}_N}{\mathrm{Q}_i} \tag{6}$$

Among them, $e_1, e_2, \ldots, e_N$ represents the eigenvectors corresponding to the multiple eigenvalues of the covariance matrix of the target indicator, and $su_1, su_2, \ldots, su_N$ represents the pollution indicator weights set according to the current pollution management requirements.

On this basis, the steps for improving PCA will be set as follows:

1) Based on the category of pollution data samples, preliminarily set the number of comprehensive evaluation indicators for pollution.
2) Eliminate dimensional differences between pollution indicators and standardize them [30].
3) Using standardized indicators, establish their correlation coefficient matrix $\mathrm{E} = [\mathrm{IX}_N]_{mn}$.
4) For the correlation coefficient matrix, based on the eigenvalues c_N, principal component contribution rate α, and cumulative variance contribution rate β of the original data, establish the weight matrix of the indicators as follows:

$$\frac{c_N * \alpha * \beta}{N} \quad (\mathrm{E}^{\mathrm{T}}\mathrm{E})^{-1} \tag{7}$$

5) By using the weight matrix to evaluate the indicators, the improved evaluation indicators are as follows:

$$\mathrm{IX}_{N'} = \mathrm{IX}_N * \omega_N \tag{8}$$

In the PCA method improved by the weighted assignment mentioned above, the principal component is the comprehensive indicator obtained through weight matrix and correlation coefficient matrix analysis. Based on this, a comprehensive evaluation index system for environmental pollution in tourist attractions is established as shown in Table 1 [31].

4.1. Realize comprehensive evaluation of environmental pollution in tourist attractions

Once the evaluation index system is reformed, the entropy method is used to calculate the evaluation values of various indicators in the index system. After layer-by-layer aggregation, the final comprehensive evaluation result of pollution in tourist attractions is obtained. The systematic process to be followed is given below:

Step 1: Calculation process of three-level indicator evaluation index:

1) When the indicator is a positive indicator, there are:

$$\mathrm{D}_i = \frac{\mathrm{F}_i}{\mathrm{E}_i} \tag{9}$$

2) When the indicator is negative, there are:

$$\mathrm{D}_i = \frac{\mathrm{E}_i}{\mathrm{F}_i} \tag{10}$$

Among them, D_i represents the evaluation value of the target indicator; E_i represents the actual measurement value of the target indicator; F_i represents the standard value of the target indicator [31, 32].

Step 2: The calculation process of the secondary indicator evaluation index is as follows:

$$\mathrm{R}_i = \sum_{i=1}^{m} \mathrm{e}_i\mathrm{D}_i \tag{11}$$

Among them, R_i represents the secondary indicator of the target, m represents the number of indicators in this indicator layer, and e_i represents the information entropy of each indicator in this indicator layer. The calculation method is as follows:

Table 1
Comprehensive evaluation index system for environmental pollution

Primary indicator	Secondary indicators	Three-level indicators
Pressure Indicator	The pressure of tourist attractions	Per capita consumption expenditure of tourists' per capita wastewater discharge
	Pressure from pollution	• Per capita solid waste generation • Per capita smoke and dust emissions per capita GDP
Status Indicator	The current development status of tourist attractions	The proportion of the tourism industry to GDP green plant coverage rate
	The current situation of pollution	Per capita water resources Airborne inhalable particulate matter content
Environmental pollution Indicator	Comprehensive evaluation index system for environmental pollution.	A comprehensive evaluation of pollution in tourist attractions using principal component analysis method.

Table 2
Classification criteria for comprehensive evaluation of pollution in tourist attractions

Grade scoring	The result of Zi	Valuation
I	≥ 0.85	The ecosystem of tourist attractions is relatively complete, with a high level of balanced development and low pollution levels
II	0.70–0.85	The ecological system of tourist attractions is relatively complete, the level of sustainable development is average, and the degree of pollution is relatively low
III	0.60–0.70	The level of sustainable development is average, the pollution level is relatively severe but meets the requirements, and the natural and economic development is not coordinated.
IV	0.30–0.60	The obstacles to sustainable development are relatively low, the degree of pollution is severe, and the natural and economic development is not coordinated.
V	≤0.30	The ecosystem of tourist attractions has been damaged, with serious obstacles to sustainable development, severe pollution, and uncoordinated natural and economic development

$$e_i = \frac{\omega_i}{\sum \omega_i} * \frac{\log 2\ \omega_i}{\sum \omega_i},\ (i\epsilon N) \tag{12}$$

Through the above formula, a comprehensive evaluation of environmental pollution in tourist attractions was conducted, and corresponding values were obtained. In order to more intuitively display the evaluation results and the environmental quality of the scenic area, it is necessary to define the evaluation results, as shown in Table 2.

Enter the calculation of Equation (12) into Table 2 and obtain the final comprehensive evaluation and results of environmental pollution in tourist attractions based on corresponding values [33].

5. Overview of the Research Area

The selected experimental area is located between 30 °–32 °N and 117 °–120 °E, and the scenic area contains two parts: cultural and natural landscapes. The temperature range of this scenic area in 2022 is between 17.5 °C and 18.5 °C, with extreme highest temperatures ranging from 36.5 °C to 41.0 °C and extreme lowest temperatures ranging from −5 °C to −1.5 °C. The annual precipitation is 1100.6 mm~2160.8 mm, mainly caused by typhoon and plum rain. The pH value of the precipitation is 5.0, and the acid rain rate is 80%. The acid rain type is sulfuric acid; the total annual sunshine hours are 1305.9–1720.5 h. In this study, corresponding water quality monitoring, air monitoring, soil monitoring, and fixed waste monitoring devices were used to collect pollution data from this tourist attraction and integrate it into the target database. Currently, this database contains 125451 pieces of water pollution data; 147185 air pollution data; 168542 pieces of soil pollution data; and 321547 fixed waste pollution data. A total of 762725 articles. Using the above data as the data foundation, complete a comprehensive evaluation of environmental pollution in tourist attractions.

5.1. Experimental indicators

Experimental methods of the presented work are compared with the methods [18, 19], to highlight whether the method of this paper has the value of promotion and application in practical work. In the experiment, the application performance of three methods was analyzed using the following indicators.

1) Data recall rate: Due to the significant dimensional differences and complex overall categories of the collected raw pollution data, there may be situations where some data cannot be identified and processed, which may affect the final evaluation results. To obtain more reliable evaluations and results, it is necessary to calculate and analyze the data recall rate.
2) Data convergence speed: This indicator refers to the speed at which the data sequence approaches the limit value. The fast convergence speed means that the method can achieve the predetermined accuracy requirements with less iteration, thereby improving computational efficiency. This is particularly important for processing large-scale environmental pollution datasets, as fast convergence speed can reduce computation time and resource consumption.
3) Indicator correlation: It evaluates the degree of correlation between various environmental pollution indicators. By analyzing the correlation of indicators, it can be determined to handle effectively whether the correlation between each indicator is close and their contribution to the overall environmental pollution comprehensive evaluation system.

5.2. Experimental results and analysis

5.2.1. Data recall analysis

The data recall of different methods was tested, and the results are shown in Table 3.

After analyzing the data in Table 3, it has been observed that there are significant distinctions in the data query rates of the method of this paper, method [18], and the method [19] for different types of pollution data. Among them, the method of this paper can control the data query rate to be above 98%, while the method [18] has a data query rate

Table 3
Experimental results of data recall rate (Unit: %)

Data type	Method of this paper	Method [18]	Method [19]
Water pollution data	98.5	95.5	94.0
Air pollution data	98.5	96.2	96.2
Soil pollution data	99.0	96.0	93.1
Fixed waste pollution data	99.1	95.4	94.0
Comprehensive Pollution data	98.0	96.6	93.6

Table 4
Experimental results of convergence speed of indicator data (unit: times)

Data volume/piece	The number of iterations of the method of this paper	The number of iterations of method [6]	The number of iterations of method [7]
100000	50	100	50
200000	50	100	50
300000	100	150	100
400000	100	150	150
500000	150	200	150
600000	150	300	200
700000	150	300	200

between 95% and 96%. Compared to these two methods, the data query rate of method [19] fluctuates slightly, with values ranging from 93.1% to 96.2%. Based on the above-performed experimental results, it can be determined that the method of current research findings has strong data processing capabilities.

5.2.2. Analysis of data convergence speed

During the experiment, 100000 pieces of data were used as the initial data, and an additional 100000 pieces of data were added for each experiment until all the data were converged. The convergence effect of different methods was determined by comparing the number of iterations. The particular experimental results are reported in Table 4.

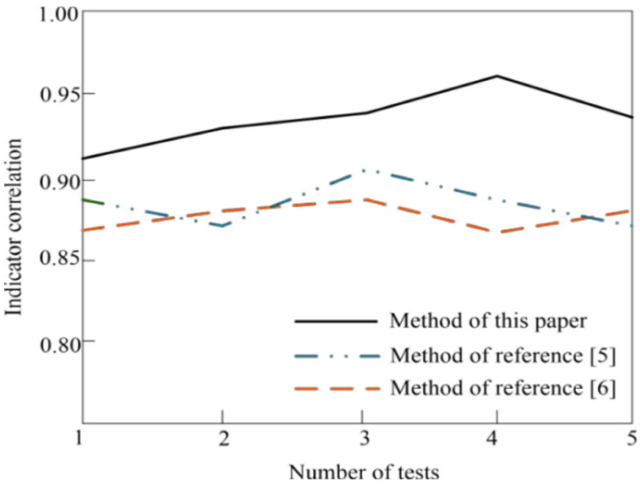
By comparing the data in Table 4, the following conclusion can be drawn: the convergence speed of the method [18] is relatively slow, and more iterations are required to achieve the target value under the same amount of data. Through specific analysis, it can be found that this method does not involve data preprocessing and directly classifies the data, resulting in slower data convergence speed. The convergence speed of method of this paper is fast, and it can approximate the optimal solution within 150 iterations, thanks to the use of data preprocessing in the method of this paper. The convergence speed of the method [19] is between the method of this paper and method [18], with moderate performance, which can improve the convergence speed to a certain extent. By comparing the convergence rates of different methods, it can be determined that the method of this paper performs the best in this experimental stage.

5.2.3. Index correlation analysis

In this experiment, the correlation of indicators was set as type parts: the correlation between indicators of the same type and the correlation between indicators of different types. The experimental results are reported in Figure 6.

Analyzing Figure 1, it has been observed that the method of this paper can consistently control the correlation of evaluation indicators above 0.90. The correlation between the indicators of method [18] and method [19] is relatively close, maintaining between 0.85 and 0.90. By contrast, after applying the method of this paper, the correlation between various evaluation indicators becomes closer. In this case of method [6] indicates to indicator correlation in the range of 0.85 to 0.88 (Decreasing) whereas method [5] indicates to indicator correlation in the range of 0.88 to 0.85 (Increasing). This can improve the contribution of the indicators to the overall environmental pollution comprehensive evaluation system.

Figure 6
Index correlation of different methods



6. Conclusion

This study designed a comprehensive evaluation method for environmental pollution in tourist attractions based on improved PCA. This method first collects air, water, and soil pollution data from tourist attractions and then cleans and classifies these pollution data. The data are collected over a few years in the top 5 metropolitan cities of India, wherein New Delhi city has reported the maximum air pollution compared to other cities. Then, by introducing a weight matrix to improve the PCA method, an evaluation index system is established. Finally, the comprehensive evaluation method is used to calculate the evaluation values of various indicators and determine the pollution evaluation level. The experimental results obtained show that this method has the characteristics of high data query rate, high correlation of evaluation indicators, and fast convergence and can effectively evaluate the environmental pollution situation of tourist attractions.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Shridhar Kedar: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization, Supervision. **Nilesh Kolhalkar:** Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing, Visualization. **Prerna Mishra:** Resources, Data curation, Writing – review & editing, Supervision.

References

[1] Abhijith, K. V., Kumar, P., Omidvarborna, H., Emygdio, A. P. M., McCallan, B., & Carpenter-Lomax, D. (2024). Improving air pollution awareness of the general public through citizen science approach. *Sustainable Horizons*, *10*, 100086. https://doi.org/10.1016/j.horiz.2023.100086

[2] Wu, F., Wang, X., Tan, K., & Liu, Z. (2020). Assessment of heavy metal pollution in agricultural soil around a gold mine area in Yitong County. In *International Geoscience and Remote Sensing Symposium*, 5034–5037. https://doi.org/10.1109/IGARSS39084.2020.9324309

[3] Li, X. (2021). Using sensor network for tracing and locating air pollution sources. *IEEE Sensors Journal*, *21*(10), 12162–12170. https://doi.org/10.1109/JSEN.2021.3063815

[4] Rickerby, D. G., & Skouloudis, A. N. (2016). Big data for innovative air-pollution assessments in the era of verifiable regulatory decisions. In *Proceedings of the 18th Mediterranean Electrotechnical Conference Intelligent & Efficient Technologies & Services Citizen,* 1–6. https://doi.org/10.1109/MELCON.2016.7495466

[5] Zhou, C. (2019). Spatio-temporal change of tropospheric no 2 column density of 'Cities in the past nine years based on Omi product's Republic of China, Beijing 100094, C. In *IEEE International Journal of Geoscience and Remote Sensing Symposium, no. 2*, 4419–4422.

[6] Chen, X., Zhang, C., Yan, K., Wei, Z., & Cheng, N. (2023). Risk assessment of agricultural soil heavy metal pollution under the hybrid intelligent evaluation model. *IEEE Access*, *11*, 106847–106858. https://doi.org/10.1109/ACCESS.2023.3319428

[7] Wang, S., Zhang, X., Zhu, H., Chen, X., Shuai, T., & Zhuang, Z. (2013). Non-point source P pollution risk assessment for basin area based on remote sensing image. In *International Journal of Geoscience and Remote Sensing Symposium*, 3706–3709. https://doi.org/10.1109/IGARSS.2013.6723635

[8] Zheng, Y., Shen, Q., Tao, S., Cao, Q., Feng, C., & Wang, M. (2020). Risk assessment of drinking water source based on high spatial resolution remote sensing. In *International Journal of Geoscience and Remote Sensing Symposium*, 2197–2200. https://doi.org/10.1109/IGARSS39084.2020.9324456

[9] Ceca, D., Unc-conae, I. M. G., & Ca, F. (2019). Monitoring air pollution from wildfires using ground data, satellite products and modeling: The Austral summer 2016–2017. In *Argentina Garc ´ Germ ´ nete, Argentina CETEMPS and Dipartimento di Fisica, Universit ' a degli Studi dell' Aquila, L' Aquil*, 7630–7633.

[10] Gade, M., Mayer, B., Meier, C., Pohlmann, T., Putri, M., & Setiawan, A. (2017). An assessment of marine oil pollution in Indonesia based on SAR imagery. In *International Journal of Geoscience and Remote Sensing Symposium*, *2017*, 1534–1537. https://doi.org/10.1109/IGARSS.2017.8127261

[11] Ambert, A., Germain, M., & Bourroubi, Y. (2023). Towards a method for taking account of uncertainties through the calculation of masses in the context of a fusion of heterogeneous multi-sensor data on air pollution. In *International Journal of Geoscience and Remote Sensing Symposium*, *2023*, 6826–6828. https://doi.org/10.1109/IGARSS52108.2023.10283283

[12] Battaglino, S., & Koyuncu, E. (2020). A generalization of principal component analysis. In *IEEE International Conference Acoustics, Speech Signal Processing – Proceedings*, *2020*, 3607–3611. https://doi.org/10.1109/ICASSP40776.2020.9054154

[13] Onyia, N. N., Balzter, H., & Berrio, J. C. (2018). Detecting vegetation response to oil pollution using hyperspectral indices. In *International Journal of Geoscience and Remote Sensing Symposium*, *2018*, 3963–3966. https://doi.org/10.1109/IGARSS.2018.8519398

[14] El Taher, A., & Massoud, E. E. (2023). Distribution of heavy metal and pollution assessment in environmental samples along the coastal area of Ras Tanura. *International Journal of Environmental Analytical Chemistry*, *103*(18), 6794–6803.

[15] Fang, Z., Hua, C., He, J., Zhu, M., Tian, Y., Yong, X., & Ren, L. (2023). Pollution assessment and source apportionment of heavy metal(loid)s in soil of Huangshui River basin, Qinghai Province, China. *Stochastic Environmental Research and Risk Assessment*, *37*(12), 4843–4855.

[16] Nourali, M. (2023). Improved treatment of model prediction uncertainty: Estimating rainfall using discrete wavelet transform and principal component analysis. *International Journal of Water Resources Management*, *37*(11), 4211–4231.

[17] Zhang, J., Peng, W., Lin, M., Liu, C., Chen, S., Wang, X., & Gui, H. (2023). Environmental geochemical baseline determination and pollution assessment of heavy metals in farmland soil of typical coal-based cities: A case study of Suzhou City in Anhui Province, China. *Heliyon*, *9*(4), 14841–14849.

[18] Jiang, Y., Yu, J., Wang, R., Wang, J., Li, Y., Yu, F., & Zhang, Y. (2023). Source analysis and pollution assessment of soil heavy metals in typical geological high background area in Southeastern Chongqing. *International Journal of Environmental Science*, *44*(07), 4017–4026.

[19] Yang, B. (2020). Influencing factors and comprehensive evaluation of urban environmental pollution. *International Journal of Environmental Impact Assessment*, *42*(02), 87–89+96.

[20] Liu, J., Wang, Y., Tang, J., Luo, J., Wang, P., & Luo, Z. (2023). Site groundwater environmental pollution risk assessment based on game theory comprehensive weight method. *International Journal of Safety and Environmental Engineering*, *30*(01), 221–230.

[21] Hareinda, N. L., Boukar, H., Robert, A., Dalatou, T., & Adamou, I. (2023). Pollution assessment of two urban soils in the city of Ngaoundere, Cameroon. *Journal of Geography, International Journal of Environment and Earth Science International*, *27*(9), 83–92.

[22] Zhao, Y., Zhou, M., Xu, X., & Zhang, N. (2023). Fault diagnosis of rolling bearings with noise signal based on modified kernel principal component analysis and DC-ResNet. *CAAI Transactions on Intelligence Technology*, *8*(3), 1014–1028.

[23] Li, G. Z., Wang, X. Y., Yan, M. X., Chen, W. K., Zhang, Q. H., Zhang, E. L., & Liu, E. F. (2023). Spatial-temporal variations and pollution assessment of heavy metals in sediments of Qionghai Lake in Xichang City. *Huan Jing ke Xue = Huanjing Kexue*, *44*(10), 5536–5545.

[24] Khan, J., Gupta, G., Singh, N. K., Bhave, V. N., Bhardwaj, V., Upreti, P., . . . , & Sinha, A. K. (2023). Geophysical and geostatistical assessment of groundwater and soil quality using GIS, VES, and PCA techniques in the Jaipur region of Western India. *Intentional Journal of Environmental Science and Pollution Research International*, *30*(31), 77713–77728.

[25] Atkins, P. J., & Cummins, M. (2023). Improved scalability and risk factor proxying with a two-step principal component analysis for multi-curve modelling. *European Journal of Operational Research*, *304*(3), 1331–1348.

[26] Mummidivarapu, S. K., Rehana, S., & Rao, Y. S. (2023). Mapping and assessment of river water quality under varying hydro-climatic and pollution scenarios by integrating QUAL2K, GEFC, and GIS. *International Journal Environmental Research*, *239*(P1), 117250–117250.

[27] Agwu, E. J., Odanwu, S. E., Ezewudo, B. I., Odo, G. E., Nzei, J. I., Iheanacho, S. C., & Islam, M. S. (2023). Assessment of water quality status using heavy metal pollution indices: A case from Eha-Amufu catchment area of Ebonyi River, Nigeria. *Acta Ecologica Sinica*, *43*(6), 989–1000.

[28] Sun, H., Jia, X., Wu, Z., Yu, P., Zhang, L., Wang, S., & Xia, T. (2023). Contamination and source-specific health risk assessment of polycyclic aromatic hydrocarbons in soil from a mega iron and steel site in China. *Journal of Environmental Pollution (Barking, Essex: 1987)*, *340*(P2), 122851–122856.

[29] Vinayagam, S., Sathishkumar, K., Ayyamperumal, R., Natarajan, P. M., Ahmad, I., Saeed, M., . . . , & Sundaram, T. (2023). Distribution and transport of contaminants in soil through mining processes and its environmental impact and health hazard assessment: A review of the prospective solutions. *Journal of Environmental Research*, *240*(P2), 117473–117483.

[30] Feng, J., Burke, I. T., Chen, X., & Stewart, D. I. (2023). Assessing metal contamination and speciation in sewage sludge: Implications for soil application and environmental risk. *International Journal of Reviews in Environmental Science and Bio/Technology*, *22*(4), 1037–1058.

[31] Mendili, M., Jrad, T. B., & Khadhri, A. (2023). Lichen diversity and bioaccumulation of heavy metals in northern Tunisia: A study to evaluate environmental pollution. *International Journal of Euro-Mediterranean Journal for Environmental Integration*, *8*(4), 847–862.

[32] Nicholas, D. O. (2023). Pollution assessment of private bore-well water supply using contamination index and contour in Ogbia local government area of Bayelsa State, Nigeria. *Asian Journal of Environment & Ecology*, *21*(2), 27–53.

[33] Ruiyao, L., Ru, Q., Qing, S., Li, X., Han, W., Longyu, C., & Zemeng, L. (2023). Assessment of soil heavy metal pollution in the dump of a Western Inner Mongolian coal mine. *Journal of Resources and Ecology*, *14*(4), 683–691.

How to Cite: Kedar, S., Kolhalkar, N., & Mishra, P. (2024). A Comprehensive Evaluation Method for Environmental Pollution in Tourist Attractions Based on Improved Principal Component Analysis. *Archives of Advanced Engineering Science*. https://doi.org/10.47852/bonviewAAES42022862