Uncovering Key Pollution Drivers in the Langat River Basin Using Factor Analysis on Water Quality Indicators

Siti Sarah Januri, Nur Najwa Ayuni Zaihan, Haslinda Ab Malek*

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia.

DOI: https://dx.doi.org/10.47772/IJRISS.2025.90800004

Received: 15 July 2025; Accepted: 22 July 2025; Published: 26 August 2025

ABSTRACT

Water quality has become a growing concern due to increasing human activities that contribute to river pollution, particularly in rapidly developing regions such as the Langat River Basin. Despite the regular monitoring of individual parameters, a deeper understanding of the interrelationships between water quality indicators remains limited. This study aims to examine the characteristics of water quality indicators and the Water Quality Index (WQI) across the Langat River Basin and to determine the underlying factors influencing the WQI indicators. A total of 360 water quality samples collected between 2020 and 2022 were obtained from the Department of Environment (DOE) Malaysia. Six key indicators which are dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, and ammonia nitrogen (NH3-N) were analysed in this study. Descriptive analysis was conducted to assess the variation of these parameters, while factor analysis was employed to identify pollutant underlying factors. The results revealed two distinct factors explaining over 60% of the total variance. The first factor is linked to organic pollution, while the second reflects physicochemical alterations such as sedimentation and pH disturbance. These findings highlight the potential sources of pollution and offer a simplified structure for interpreting water quality data. This study contributes to the understanding of major water quality drivers in the Langat River Basin and supports more effective environmental assessment, policymaking, and sustainable water resource management.

Keywords: Water Quality, Langat River Basin, Factor Analysis, Descriptive Analysis, Water Quality Index, Pollution Indicator

INTRODUCTION

Water is essential for sustaining life and maintaining environmental stability. It is necessary for body processes, climate regulation, and agricultural productivity. However, access to clean water remains a global issue. According to the [1], millions of people die each year from diarrhoea caused by inadequate water, sanitation, and hygiene. The problem is significant in Malaysia, where just 53% of river basins are classified as clean, 42% as contaminated, and 5% as extremely polluted [2].

Human activities including industrial discharge, agricultural runoff, and improper sewage disposal cause a mounting strain on the rivers, which are the primary supply of water. For instance, there have been numerous pollution incidents related to urbanization and industrialization in the Langat River Basin, which is vital for residential, commercial, and agricultural use. These contaminants pose serious risks to public health in addition to disrupting aquatic ecosystems.

Water quality assessment plays a vital role in ensuring environmental integrity and public health. The Water Quality Index (WQI) is a widely adopted tool that integrates multiple physicochemical and biological parameters into a single score, simplifying the interpretation of water status for various users [3]-[4]. Key

indicators commonly employed in WQI calculations include dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, and ammonia nitrogen (NH_3-N) [5].

Each parameter reflects specific environmental conditions: DO indicates oxygen availability for aquatic organisms; BOD and COD represent organic and chemical pollution levels; SS affects water turbidity and light penetration; pH regulates chemical solubility and aquatic health; and NH_3-N signifies nitrogenous waste often linked to human activities [6]-[7], Elevated levels of BOD, COD, or NH_3-N generally suggest anthropogenic pollution sources such as industrial discharge or agricultural runoff. Moreover, DO is a vital indicator of aquatic ecosystem health, with its concentration influenced by factors such as temperature, salinity, and organic pollution. High BOD levels signify increased organic matter in water, which leads to oxygen depletion and poses threats to aquatic life [6]. COD represents both organic and inorganic pollutant loads, reflecting the total oxidizable substances in water and serving as an indicator of industrial pollution. Meanwhile, SS affects water clarity and photosynthetic activity by reducing light penetration, while pH levels influence the solubility and toxicity of chemical compounds in water bodies. Elevated levels of NH_3-N can be toxic to aquatic organisms and contribute to eutrophication [7]

It offers a straightforward but thorough assessment of water quality, allowing monitoring and policymaking more efficient. Numerous models of WQI have been developed globally, with the DOE Malaysia adopting a model where a higher WQI score denotes better water quality. Table 1 shows the water quality index (WQI) classification by Department of Environment (DOE) along with the associated indicators.

TABLE 1 WATER QUALITY INDEX CLASSIFICATION

Parameter	Unit	CLASS				
		I	II	III	IV	V
NH3-N	mg/l	< 0.1	0.1 – 0.3	0.3 – 0.9	0.9 – 2.7	> 2.7
BOD	mg/l	< 1	1 – 3	3 – 6	6 – 12	> 12
COD	mg/l	< 10	10 – 25	25 – 50	50 – 100	> 100
DO	mg/l	> 7	5 – 7	3 – 5	1 – 3	< 1
pH	-	> 7.0	6.0 – 7.0	5.0 – 6.0	< 5.0	< 5.0
SS	mg/l	< 25	25 – 50	50 – 150	150 – 300	> 300
WQI	-	> 92.7	76.5 – 92.7	51.9 – 76.5	31.0 – 51.9	< 31.0

Therefore, this study applies Factor Analysis (FA) to analyse water quality data from the Langat River Basin, with the goal of identifying dominant pollutant groups and informing targeted environmental management strategies.is to implement WQI and its associated indicators to assess the water quality of the Langat River Basin and to apply factor analysis to determine the most important underlying factors. Comprehending these hidden factors is crucial for directing focused management tactics to reduce pollution and secure water security. In addition, uncover hidden patterns among these indicators, FA has proven effective in simplifying complex datasets and identifying latent pollution sources. As demonstrated by [3], FA can reduce data dimensionality while grouping correlated parameters under common factors, such as organic pollution or physicochemical changes. Moreover, [8] also emphasized FA's utility in enhancing the interpretability of multivariate water quality data.

METHODOLOGY

Study Area

This study was conducted in the Langat River Basin, located in Selangor, Malaysia. The Langat River Basin, a vital water source for the Klang Valley, has experienced degradation from rapid urbanization and industrialization. The Langat River, which spans approximately 78 kilometers and drains a basin area of around 2,350 square kilometers, plays a crucial role in supplying water to urban centers such as Putrajaya,

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

Kajang, and Cheras. The basin, being a critical freshwater resource for the Klang Valley region, has experienced increasing pressures from rapid urban and industrial development. The dataset used in this research was obtained from the Department of Environment (DOE) Malaysia and consists of secondary water quality data collected from 2020 to 2022. A total of 360 samples were used to ensure representative coverage of the basin.

Six water quality indicators were analyzed in this study: dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, and ammonia nitrogen (NH_3-N). These indicators were selected based on their significance in determining the Water Quality Index (WQI), which was calculated using the standard formula provided by the DOE.

Method of Analysis

Two statistical methods were applied to achieve the objectives of study. Descriptive statistics were used to summarize the distributions and central tendencies of each indicator. This provided insight into the general water quality conditions across the sampling period and locations. For each variable (DO, BOD, COD, SS, pH, NH_3-N, and WQI), the following statistical measures are computed:

Minimum Value(X) = lowest observed value of X

Maximum Value(X) = highest observed value of X

Mean: $\bar{X} = \frac{1}{n}\sum_i^n X_i$

where:

Xi = individual observation

n = total number of observations

Second, factor analysis was employed to uncover latent structures within the dataset. In water quality evaluation, factor analysis has been largely used to simplify the datasets with a few numbers of factors when an array of water quality parameters. This method is particularly effective for displaying patterns of variation of water quality parameters and to compress large datasets into understandable factors for interpretation. The factors generated by factor analysis are indicative of all potential sources contributing the water quality variation of interest. The factor analysis model is mathematically expressed as:

$X = \mu + LF + \varepsilon$

where:

X: vector of observed variables (e.g., DO, BOD, COD, SS, pH, NH_3-N)

μ: vector of means of the observed variables

L: factor loading matrix (shows the correlation between observed variables and common factors)

F: vector of common (latent) factors

ε: vector of unique (error/specific) factors

Principal Component Analysis (PCA) was used as the extraction method, followed by Varimax rotation to improve interpretability of the factors. Prior to conducting factor analysis, assumptions were tested to ensure data suitability, including the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

of sphericity. The correlation matrix was also examined to verify meaningful interrelationships among the indicators.

Through these methods, the study aimed to reduce the dimensionality of the water quality data and identify the most influential underlying factors contributing to pollution in the Langat River Basin.

RESULT AND DISCUSSION

Descriptive Analysis

The descriptive statistics revealed that the mean WQI for the basin was 81.22, which places most sampling points within the 'clean' category based on the Department of Environment (DOE) classification. Dissolved Oxygen (DO) exhibited a healthy average of 6.61 mg/L, surpassing the 5.0 mg/L Class II threshold, indicating generally supportive conditions for aquatic life.

Conversely, several pollution indicators exceeded acceptable ranges. Biochemical Oxygen Demand (BOD) had a mean value of 3.26 mg/L, slightly above the Class II threshold of 3.0 mg/L, suggesting moderate organic pollution. Chemical Oxygen Demand (COD) averaged 17.70 mg/L, below the cut-off value of 25.0 mg/L, but with a maximum value of 66.0 mg/L, indicating localized severe pollution. Suspended Solids (SS) and Ammonia Nitrogen (NH_3-N) also raised concerns; SS had a high mean of 87.99 mg/L, well above the 50.0 mg/L threshold, while NH_3-N recorded a mean of 0.72 mg/L—more than double the acceptable limit—indicating significant nutrient pollution likely from agricultural runoff or sewage. In addition, pH values, ranging from 5.89 to 8.44 (mean = 6.95), were generally within the acceptable range, indicating minimal acid–base imbalance in the river ecosystem. Table 2 represents the summary of descriptive statistics for water quality indicators and water quality index (WQI) and the cut-off values.

TABLE 2 DESCRIPTIVE STATISTICS OF WATER QUALITY INDICATORS AND WATER QUALITY INDEX (WQI)

Parameter	Minimum	Maximum	Mean	Cut-off Values
DO (mg/L)	3.04	9.83	6.61	5.00
BOD (mg/L)	0.90	16.00	3.26	3.00
COD (mg/L)	2.00	66.00	17.70	25.00
SS (mg/L)	0.90	838.00	87.99	50.00
pH	5.89	8.44	6.95	6.00 – 7.00
NH_3-N (mg/L)	0.009	8.19	0.72	0.30
WQI	57.00	98.00	81.22	76.5

Factor Analysis

Factor analysis was implemented to reveal a hidden pattern among the indicators. The Kaiser-Meyer-Olkin value (0.648) and Bartlett's test ($p < 0.001$) in Table 3 validated the adequacy of the data, indicating that the assumptions were evaluated and satisfied.

TABLE 3 KAISER-MEYER-OLKIN AND BARTLETT'S TEST

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.648	
Bartlett's Test of Sphericity	Approx. Chi-Square	567.227
	df	15
	Sig	<0.001

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume IX Issue VIII August 2025

Based on Table 4, two significant components were identified by Principal Component Analysis (PCA), which explained 62.77% of the cumulative variance. Factor 1 explained 42.05% of the variance, showing strong positive loadings for BOD (0.754), COD (0.603), and NH3-N (0.809), along with a strong negative loading for DO (−0.740). This pattern represents pollution from sewage, industrial effluents, and agricultural runoff and is typical of organic pollution carried on by microbial decomposition. Therefore, the Organic Pollution Factor was interpreted as Factor 1.

TABLE 4 COMPONENT MATRIX OF ROTATED FACTOR ANALYSIS

	Component	
	1	2
DO	-0.74	-0.345
BOD	0.754	0.171
COD	0.603	0.593
SS	-0.013	0.733
pH	-0.123	-0.704
NH3-N	0.809	-0.293

Moreover, Factor 2 which accounted for 20.72% of the variance, showed a strong negative loading for pH (-0.704) and a high positive loading for SS (0.733), suggesting both chemical and physical disruptions, most likely from chemical discharge, sedimentation, or erosion. Therefore, the factor was identified as the Physicochemical Factor. Varimax rotation was implemented to improve the interpretability of the factor structure. After the rotation, Factor 1 remained the dominant factor, contributing the largest share of variance, and highlighting organic pollution as the most critical concern for the Langat River Basin followed by the second factor, the Physicochemical Factor.

The inverse relationship between BOD and DO was consistent with literature (Tripathi & Singal, 2019; Nyantakyi et al., 2024), affirming that high BOD reduces DO availability in the water. Moreover, the findings in this study resonate with [9] and [10], who highlighted the role of SS and pH in affecting water clarity and aquatic health.

A comparative analysis with this study done by [11], which investigated the Langat River using chemometric techniques such as principal component analysis (PCA) and hierarchical cluster analysis (HCA). Both studies employed multivariate statistical approaches to evaluate key water quality indicators namely as dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solids (SS), pH, and ammonia nitrogen (NH_3-N) in assessing the state of the Langat River Basin. Consistent with their findings, this study also revealed that organic pollution (as reflected in BOD, COD, and NH_3-N levels) and physicochemical changes (associated with SS and pH) were the dominant contributors to water quality degradation. However, the use of FA in this research provided additional interpretative value by uncovering underlying pollutant structures, offering a more understanding of interrelated water quality variables. This advancement underscores the methodological complementarity of both studies while highlighting the potential of factor analysis to support more targeted environmental management strategies in the Langat River Basin.

CONCLUSION

In summary, this study revealed significant variations in several water quality indicators, with some exceeding safe limits due to urban, agricultural, and industrial influences. Factor analysis employed in the study successfully identified two key underlying factors. The Organic Pollution Factor includes DO, BOD, COD, and NH3-N, while the Physicochemical Factor consists of SS and pH. The Organic Pollution Factor was the most influential, indicating that organic waste and microbial activity are the primary contributors to water

quality variation in the Langat River Basin. Overall, the factor analysis successfully reduced the complexity of the dataset and offered insight into the dominant pollution sources affecting water quality in the Langat River Basin. These findings provide useful insights for future water management and support sustainable efforts to preserve the Langat River Basin.

Based on the outcomes of this study, several key recommendations are proposed to support improved water quality management in the Langat River Basin. Environmental monitoring and regulatory enforcement should be strengthened, particularly in high-risk zones such as industrial and agricultural areas, where elevated levels of pollutants were observed. For future research, it is recommended that additional water quality parameters such as heavy metals, nitrates, phosphates, and microbial indicators be included to provide a more comprehensive assessment of water quality conditions. Moreover, spatial and temporal analyses should be conducted to capture the variability of water quality across different sections of the river and across seasonal periods.

Moreover, the integration of machine learning techniques is strongly recommended for future water quality prediction and classification tasks. Supervised machine learning algorithms such as random forests, support vector machines, and gradient boosting could be employed to predict Water Quality Index classes based on complex interactions among water quality indicators.

ACKNOWLEDGEMENT

This research was supported by Universiti Teknologi MARA (UiTM) and funded under UiTM internal Grant No. 600-RMC/GPM LPHD 5/3 (077/2023). This support is gratefully acknowledged. The authors would like to thank other researchers, lecturers and friends for ideas and discussion.

REFERENCES

1. World Health Organization (WHO). (2023). Drinking-water. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/drinking-water
2. Global Environment Centre. (2023). River Care Programme. https://gec.org.my/programme/river-care-programme-rcp/
3. Tripathi, M., & Singal, S. K. (2019). Allocation of weights using factor analysis for development of a novel water quality index. Ecotoxicology and Environmental Safety, 183, 109510. https://doi.org/10.1016/J.ECOENV.2019.109510
4. Boyacıoğlu, H., & Boyacıoğlu, H. (2020). Ecological Water Quality Index associated with factor analysis to classify surface waters. *Water Supply*, *20*(5), 1884–1896. https://doi.org/10.2166/ws.2020.096
5. Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. Ecological Indicators, 122, 107218.
6. Nyantakyi, J. A., Sarpong, L., Mensah, R. B., & Wiafe, S. (2024). Surface water quality assessment and probable health threats of metal exposure in the Tano South Municipality, Ahafo, Ghana. Scientific African, 26, e02437. https://doi.org/ 10.1016/ J.SCIAF.2024.E02437.
7. Lu, L., Wang, Z., Wang, Z., Deng, L., Zou, S., Fan, L., & Yang, Y. (2024). Chemical characteristics and water quality assessment of groundwater in the Dongjiang-Hanjiang River Basin, China. Journal of Environmental Chemical Engineering, 12(6), 114721. https://doi.org/10.1016/J.JECE.2024.114721
8. Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11. https://doi.org/10.12691/ajams-9-1-2.
9. Valentini, M., dos Santos, G. B., & Muller Vieira, B. (2021). Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil. *SN Applied Sciences*, *3*(1), 1–11. https://doi.org/10.1007/S42452-020-04005-1/TABLES/11.

10. Kodli, M., Rajashekara, H. M., Subramoniam, S. R., & Jadi, R. (2023). Water Quality Studies in Urban Lakes Using Sentinel 2A Data. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2023-July*, 3671–3674. https:// doi.org/10.1109/ IGARSS52108.2023.10281762.
11. Kamarudin, M. K. A., Yusop, Z., & Yusof, M. J. M. (2015). *Analysis of Langat River water quality data using chemometric techniques*. Procedia Environmental Sciences, 30, 79–84. https://doi.org/10.1016/j.proenv.2015.10.014.