HEALTH CARE RESEARCH & CASE REPORTS JOURNAL (HCRCRJ)

Volume 02, Issue 01 Review Article Open Access

# Revolutionizing River Water Quality Assessment Through Novel PCA Integration and Standardized WQI Metrics in Malaysian River

**Khayell Balamurali, Zalina Mohd Ali***

*Department of Mathematical Sciences, Faculty of Science, Universiti Kebangsaan Malaysia, Malaysia*

## ABSTRACT

In today's globalized world, water quality faces escalating threats from various human activities and natural calamities, ranging from sewage and wastewater discharge to urbanization, deforestation, agriculture, industrialization, marine dumping, and radioactive waste exposure. The Water Quality Index (WQI), a universally acknowledged metric, evaluates the condition of surface and groundwater by translating complex water quality data into a single, comprehensible, and dimensionless figure. This simplification is pivotal in fostering public comprehension and aiding in the maintenance of healthy living practices. This study introduces innovative multivariate analysis models for assessing river water quality, with a special focus on Principal Component Analysis (PCA) and an advanced standardization technique. Our research utilized data from Malaysia's Department of Environment, concentrating on three heavily polluted rivers: Buloh, Langat, and Jawi. The data were standardized using a novel approach aligned with the National Water Quality Standards for Malaysia (NWQSM), ensuring a precise reflection of the water quality status. Spanning the last five years, our analysis incorporated six vital water quality parameters: Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), Suspended Solids (SS), Ammoniacal Nitrogen (AN), and pH. Following standardization, we explored PCA interrelations and computed the WQI. Our methodology also included crosstab analysis, biplot analysis, and correlation analysis, comparing PCA-based WQI with the Department of Environment's WQI. The results indicate that the PCA interrelations, when applied with our innovative standardization method, yield a more efficient and trustworthy model for water quality assessment. This proposed model holds significant promise for enhancing future river water quality evaluations and for advancing methods in water resource assessment and management.

**Keywords:** Water Quality Index (WQI), Principal Component Analysis (PCA), River water quality assessment, Multivariate analysis, Environmental standardization methodology

## INTRODUCTION

Water quality is defined by the physical, chemical, and biological characteristics of water. Assessing these parameters is crucial to determine water's suitability for various uses, ranging from cooking and daily activities to industrial and agricultural applications. Pollution from contaminants severely degrades the quality of water bodies. Factors such as unchecked urban development, extensive industrialization, rapid economic growth, and unregulated human activities pose significant environmental challenges, particularly disrupting the hydrological cycle. Consequently, there is an urgent need to address these issues by actively safeguarding water quality and ensuring its purity. Widyanti has discussed the impacts on the health and lifestyle through assessment of water quality and concluded that enhancing river water quality for a healthy living behaviour.

The Water Quality Index (WQI) is an essential tool for evaluating surface water quality, offering a simplified classification system. This system, as reviewed by Juliana et al., bases its assessments on various parameters, including river class or status.

**ARTICLE INFORMATION**

**Received date:** 18 January 2026 **Accepted date:** 20 January 2026
**Published date:** 15 March 2026 **Manuscript no.:** HCRCRJ-25-01258

***Corresponding author:**
Zalina Mohd Ali, Department of Mathematical Sciences, Faculty of Science, Universiti Kebangsaan Malaysia, Malaysia; **Email:** zalina99@ukm.edu.my

WQI aggregates these diverse parameters into a single indicative value, simplifying the process for stakeholders to assess water quality and make informed decisions about water resource usage. The assessment of water quality encompasses a range of physical, chemical, and biological parameters. These parameters are weighted according to expert opinions, introducing a subjective element to the evaluation. In the development of WQI, the selection and weighting of these parameters are critical to the accuracy of the calculation process. As a result, WQI values can vary significantly between rivers, influenced by factors such as geographical location, distribution, and surrounding land use patterns.

The escalating issue of non-point source pollution, characterized by the overflow of constituents from land surfaces into river systems, requires urgent attention. The Water Quality Index stands out as the most effective tool for assessing water quality. Additionally, Carlos et al. highlight that various nations employ distinct models for WQI calculation, considering their respective technologies, environmental conditions, and financial considerations.

The most recent annual report from Jabatan Alam Sekitar, also known as the Department of Environment Malaysia (DOE), relies on physiochemical parameters to assess water quality levels. Within this framework, six key parameters play a significant role, namely Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), Ammoniacal Nitrogen (AN), and pH. BOD, AN, and SS are acknowledged as primary indicators of river water quality due to their close correlation with pollutant discharges from either point or non-point sources. Point sources refer to determinable contaminants, such as discharges from industrial areas and wastewater treatment processes, often caused by heavy metals and chemicals. Non-point source pollution is more challenging to specify and includes pollutants like pesticides and fertilizers from agricultural runoff, as well as contaminants from urban runoff.

In Malaysia, the Department of Environment (DOE) adopts a Water Quality Index (WQI) that operates on a scale of 1 to 100, defining scores between 81 to 100 as clean, 60 to 80 as slightly polluted, and 0 to 59 as polluted. This assessment is not a simple task; it demands extensive data analysis, incorporating statistical tools and artificial intelligence. The DOE conducts regular monitoring programs, acknowledging the significant spatial and temporal variability in water conditions. This ongoing monitoring results in a comprehensive and complex data matrix, covering a wide range of parameters, which can be challenging to interpret. Xin Fang et al. discussed that there are several advanced models and modelling techniques are being discovered and explored to evaluate the water quality in detail in terms of its dynamics and spatial distribution. To address this, methodologies like those proposed by Shuquan An et al. are employed, wherein multivariate statistical techniques such as Principal Component Analysis (PCA) are instrumental in refining the interpretation of these datasets.

PCA, a renowned statistical method, is primarily used for reducing data dimensionality and simplifying complex datasets. It effectively transforms high-dimensional data into more manageable forms while striving to preserve the integrity of the original data as much as possible. The resulting uncorrelated variables, termed principal components (PCs), are orthogonal to each other. These PCs linearly and independently combine original values, facilitating a clearer understanding and interpretation of the data. Micheal Greencare et al. has strongly mentioned that PCA is among the versatile statistical method and importance of PC scores for further analysis. Essentially, PCA's goal is to reduce the complexity of a multivariate dataset, maintaining its core structure and minimizing any loss or distortion of information.

PCA efficiently minimizes the number of parameters used, thereby enabling the identification of the maximum variance within the dataset. This attribute is particularly beneficial in environmental studies, where large datasets are common. Aminu et al. recently demonstrated the effectiveness of PCA in water quality modelling, highlighting its ability to discern key components and analyze intricate relationships within extensive datasets. However, a fundamental step preceding the application of PCA is the normalization of the dataset. This process ensures a uniform and comprehensive analysis, irrespective of the dataset's original scale. Camacho et al. has reviewed about Principal Component Analysis deeply in terms of the computational scores, residuals and explained variation in the year 2020 and es expediated the review to its limitation and problems of deflation in the year 2021.

Normalization is not just a procedural step; it's a critical element in problem-solving, significantly impacting the outcomes of research studies. Its primary role is to minimize the effects of technical biases, thereby ensuring that the results accurately reflect the true nature of the data. This aspect is underscored by Sankpal, who stresses the indispensability of data normalization in extracting relevant information from complex datasets. By standardizing data, normalization enhances the reliability and validity of the findings, making it an indispensable tool in environmental data analysis.

The process of transforming input data into a numerical format is fundamental, typically undertaken through normalization, transformation, or standardization. These steps form an integral part of data pre-treatment, a phase crucially highlighted by Robert et al. prior to engaging in multivariate analysis. In database management and dataset handling, there are several key stages: Data pre-processing, pre-treatment, and analysis.

Data pre-processing involves refining raw data to produce a cleaner dataset. This refined dataset then serves as the foundation for more sophisticated analysis. Following pre-processing, the data is subjected to pre-treatment, a stage focused on transforming the cleaned data into a different scale. This transformation involves eliminating extraneous elements, such as measurement noise, to enhance data quality. Various techniques, including normalization, scaling, transformation, and centering, are employed during this stage. Each technique plays a pivotal role in refining and optimizing the dataset, ensuring the subsequent analyses are more robust and reliable. By rigorously preparing data through these stages, researchers can confidently engage in complex analyses, assured of the integrity and relevance of their data.

In this research, we scrutinized the interconnections among six chosen water quality parameters using Principal Component Analysis (PCA) across various datasets. The second section of the study is dedicated to a concise summary of the literature review, providing context and background for the research. The third section delves into the research methodology, detailing the approach and techniques employed. Subsequently, the fourth section presents an in-depth examination and discussion of the results obtained from the study. This final section culminates in a comprehensive analysis of the PCA interrelations, which are instrumental in shaping the Water Quality Index (WQI). This structure ensures a systematic and thorough exploration of the topic, from theoretical underpinnings to practical applications.

# LITERATURE REVIEW

## Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique designed to reduce the complexity of a dataset by generating new values known as principal components (PCs). These PCs are linear combinations of the original variables and serve as replacements for them. A distinctive aspect of PCA is its reliance on a correlation-based approach, which meticulously examines the interrelationships among variables. As a potent multivariate analysis tool, PCA excels in interpreting and illuminating the variability present in the original data, a feature underscored and validated by Zalina et al. This method adeptly condenses the original dimensions of a dataset, transforming a multifaceted array of variables into a more manageable and interpretable index. Importantly, PCA stands out for its ability to reduce dimensionality while retaining the essence and integrity of the original dataset, ensuring that critical information is preserved even in its simplified form.

## Standardization

The practice of data standardization or normalization is crucial, particularly in decision-making contexts. This process transforms an original dataset into a new format that is proportional, numerical, and refined. The aim of standardization is to address and eliminate outliers, reduce systematic biases inherent in research methodologies, and account for unavoidable variations. Over the past twenty years, data standardization has become a staple practice, leading to significant advancements in various research domains and technological fields. In today's digital era, the significance of big data analysis is unmistakably evident. Michal Gal et al., delivers progressive information of data standardization and emphasized that data standardization is crucial in this digital era as big data is the hot topic these days. Organizations leverage big data to augment employee skills, enhance databases, refine products or services, and drive growth across various industries. In such an environment, the value of data is unparalleled, and its omnipresence demands precise and careful handling. When dealing with large or seemingly infinite datasets, data standardization stands out as the preferred method. It aims at categorizing information as effectively as possible. Additionally, normalization plays a vital role in ensuring that technical variations and biases have minimal influence on the outcomes of the newly processed dataset. This meticulous approach to data handling is integral for yielding accurate, reliable, and unbiased results in the realm of big data analysis.

In addition, Aron et al. emphasized that data normalization, historically known as canonicalization, entails converting data into a standardized, canonical format, taking various parameters into account. In the context of experimental research, data undergoes several phases, including pre-processing, pre-treatment, and analysis. The process begins with pre-processing raw data to obtain a cleaner, more refined dataset. This dataset is then subjected to pre-treatment, which is tailored for specific analytical objectives. Pre-treatment is crucial because it adjusts the dataset to a relative measure and minimizes potential disturbances or distortions. Uddin outlined three primary methods of data pre-treatment: Centering, scaling, and transformation. Centering addresses fluctuations within the dataset by adjusting the differences between high and low values. Scaling involves adjusting the dataset based on either dispersion, highlighted by the standard deviation, or size, focused around the mean of the dataset. Transformation is used to reduce the effects of disproportionately large values, particularly those showing relative heteroscedasticity, thereby ensuring a more uniform distribution of data values for analysis.

Standardization of water quality datasets holds immense importance, playing a critical role in enabling meaningful water quality analysis, the development of comprehensive water quality models, and informed decision-making in water management. This process guarantees the reliability and interpretability of the data. The significance of data standardization in water quality datasets has been thoroughly researched and elucidated by Anastasios and Anastasia and Emily et al. Therefore, data standardization is an integral part of the data pre-treatment stage, preparing the data for more effective analysis. Additionally, normalization is a prerequisite to multivariate analysis, as it establishes the normality of the data and significantly influences the interpretation of subsequent data outputs.

## Water Quality Index (WQI)

The Water Quality Index (WQI) serves as a streamlined, yet effective metric for assessing water quality. As Kang Ide Soumalia et al. emphasize, it is a key tool for analysing weather trends, presenting current environmental conditions, and assisting authorities and the public in understanding water status. Besides, Ritabrata Roy has educated and illustrated the basic inductor to water quality analysis and the important measures to look in to in the water quality analysis process. By amalgamating diverse water quality parameters into a single numerical score, the WQI becomes instrumental in the ongoing monitoring of global water resources, encompassing both surface and groundwater. Developing WQI models typically involves four critical stages.

The initial stage focuses on selecting appropriate water quality parameters. This involves careful consideration of various physical, chemical, and biological parameters, with the goal of omitting less relevant ones and prioritizing those crucial for evaluating river water quality. The selection process is guided by factors such as data availability, expert input, and environmental significance, where techniques like Principal Component Analysis (PCA) are highly valued for their precision in environmental studies.

The subsequent step involves sub-indexing the chosen parameters. This involves processes like assessing parameter concentrations, applying linear interpolated functions, or using rating curves. This stage is crucial for transforming the values of water quality parameters, which may vary in units or dimensions, into uniform, dimensionless sub-indices. These sub-indices are essential components in the composition of the overall WQI model, ensuring that it accurately reflects the quality of water resources.

The third stage of developing the Water Quality Index model involves the weighting of parameters. Typically, this is done through unequal weighting, with the sum of weights for all parameters equating to 1. This approach of assigning different weights to each parameter is critical for enhancing the model's stability and integrity. It allows the WQI to accurately reflect the relative importance of each parameter in determining overall water quality.

The final stage involves the aggregation of these weighted parameters. Various methods, such as addition, multiplication, combined aggregation functions, or minimum operator functions, are employed for this purpose. The way these weights are aggregated plays a significant role in shaping the final value of the Water Quality Index, ultimately conveying the state of water quality.

In the case of the Malaysian Water Quality Index, which is endorsed by the Department of Environment, Malaysia, a localized approach is adopted for assessing surface water quality. The selection of parameters is carried out using the Delphi technique, which relies on the consensus of expert opinions. In Malaysia, the six principal water quality parameters are biological oxygen demand, dissolved oxygen, ammoniacal nitrogen, suspended solids, chemical oxygen demand, and pH. Sub-index values for these parameters are derived using a rating curve that fits the equation to the sub-index. The parameters are then weighted using techniques of unequal weighting, based on expert judgments, and aggregated by summing the products of the sub-index values. The Malaysian WQI categorizes surface water quality into three distinct classes: Clean, slightly polluted, and polluted, offering a clear and structured assessment of water conditions.

# METHODOLOGY

## Overview of Dataset

This study utilizes an initial dataset from the Department of Environment (DOE) Malaysia, focusing on water quality observations of three significantly polluted rivers in Malaysia over the past five years: The Buloh River, Langat River, and Jawi River. Understanding water quality is critical for various daily activities, including drinking, cooking, and personal hygiene. To compute the Malaysian Water Quality Index, six sub-indices were developed, reflecting key aspects of water quality. These crucial parameters include Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), Ammoniacal Nitrogen (AN), Suspended Solids (SS), and pH.

The comprehensive raw dataset from the DOE encompasses a range of information, including geographical data (states, basins, latitude, longitude), temporal data (sampling dates, time), and various water quality measurements (DO in both % and mg/l, BOD, COD, SS, pH, AN, along with corresponding indices). Additionally, it includes the overall Water Quality Index (WQI), as well as the classification and status of each river. According to the DOE's WQI Index, rivers are categorized into five classes (Class I to V) and are designated as either a clean river, slightly polluted river, or polluted river.

For the purpose of this study, specific parameters were selectively chosen for further data processing. These include DO (mg/l), BOD (mg/l), COD (mg/l), SS (mg/l), pH, AN (mg/l), the overall WQI, class of river, and the river's status. This selective approach ensures focused analysis on the most relevant data for assessing water quality in these rivers.

## Purpose of Dataset

The primary aim of this study is to evaluate standardized datasets through Principal Component Analysis (PCA) to derive principal components, diverging from the conventional use of sub-index values for Water Quality Index (WQI) calculations. This approach involves employing a select set of principal component scores to probe into the interrelationships among water quality parameters. PCA will facilitate the generation of a new set of artificial variables, known as Principal Components (PCs), representing linear combinations of the original dataset values.

The study focuses on three polluted rivers in Malaysia - Buloh River, Langat River, and Jawi River - chosen to illustrate the severity of water pollution affecting local communities, particularly in the Selangor and Penang basins. Prior research, including studies by Marina et al. and Juahir et al., highlights the Langat River basin's contamination from suspended solids (SS), largely due to poor planning and extensive land clearing. Similarly, Noh et al. observed that industrial activities heavily impact the Buloh River. Moreover, Sahabat Alam Malaysia has called for urgent action against the dumping of chemicals, food waste, and animal waste by factories, which are primary pollution sources in the Jawi River. Beyond the direct impact of water pollution, these issues create secondary challenges for residents, such as unpleasant odours and a degraded living environment, further disrupting their daily lives.

## Data Preparation

Initial steps in this research involve meticulous organization and structuring of the input data. Specifically, it is imperative to allocate each variable to a separate column and systematically arrange the corresponding observations row-wise. For this study, distinct water quality variables are treated as independent entities, with the collected data from selected rivers forming the observational dataset. Additionally, the analysis includes the elimination of missing values and the adjustment of any values below 1 to a fixed value of 0.5, to maintain consistency.

The research employs four different datasets for comprehensive analysis: The raw dataset, the log-transformed dataset, the normalized dataset (with a mean of 0 and a standard deviation of 1), and a newly normalized dataset. Principal Component Analysis (PCA) is applied to each of these datasets to extract PCA interrelations. These interrelations are then used to calculate the Water Quality Index (WQI), as well as to determine the class and status of the Langat River.

The raw dataset is obtained directly from the Department of Environment (DOE), Malaysia. A log transformation is then applied to this dataset to generate the log-transformed dataset. This transformation process utilizes the following formula.

$$y = \log(x+1) \text{ ........................... Equation (1)}$$

Following Equation 1, where y denotes the transformed value and x represents the original data value, we generate the log-transformed dataset by applying a logarithmic transformation to the raw data. This transformation is pivotal for normalizing data distribution and facilitating subsequent analyses.

In the next step, to create a dataset normalized with a mean of 0 and a standard deviation of 1, we standardize the raw dataset. This standardization process is crucial to ensure data comparability and to balance the influence of different scales of measurement.

Lastly, we apply a novel normalization method, introduced by Yong Cao et al. in 1999, for further data computations in this study. The computation of this normalization involves the formula given in Equation 2:

$$y = \frac{x_i - Stand_i}{c_i} \text{ .......................... Equation (2)}$$

In Equation 2, y is the transformed value, xi represents the raw value of each variable, Standi indicates the water quality standard for that variable, and ci is a constant reflecting the natural variability range of the variable. These elements are crucial in the transformation process, ensuring accurate normalization of the data. Following this step, PCA interrelations were analysed across four different datasets, aiming to support optimal decision-making in water quality assessment. According recent to DOE's Annual Environmental Quality reports, all three rivers selected in this study categorised into class III for class of river water quality.

**Table 1** illustrates the values of Standi and ci used to determine the novel standardization data set.

**Table 1:** Class III Standi and ci values according to NWQSM

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| $Stand_i$ | 4 | 6 | 50 | 150 | 7 | 0.9 |
| $c_i$ | 2 | 6 | 50 | 150 | 4 | 0.9 |

To add more context, the values for ci and Standi are derived from the National Water Quality Standards for Malaysia (NWQSM). These standards provide specific criteria for each water quality parameter and for different classes of river water quality. In this study, we have applied the standard ranges for Class III from the NWQSM to evaluate the six water quality parameters for the Buloh River, Langat River, and Jawi River. This decision was based on the categorization of these rivers in Class III over the past five years. The use of NWQSM standards as a benchmark ensures that our analysis is grounded in nationally recognized water quality metrics, providing a reliable basis for assessing and interpreting the water quality of these rivers.

## Data Analysis

In this study, Rstudio was indispensable as the main computational tool for executing a range of data transformations. This included log transformation, standardization to a mean of 0 and standard deviation of 1 and implementing a novel normalization method. Additionally, Rstudio facilitated the computation of the Principal Component Analysis (PCA) correlation matrix, eigen vectors, eigenvalues, and eigen loadings.

To assess the degree of linearity and compatibility between pairs of variables, correlation analysis was conducted. This preliminary analysis is crucial for understanding the strength and direction of the relationships between different water quality parameters. Following the correlation analysis, Principal Component Analysis (PCA) was applied to the resulting correlation matrices. This step is integral for distilling complex data into principal components, thereby simplifying the interpretation and analysis of the relationships among the variables.

Correlation analysis is a statistical method employed to assess the strength and direction of the relationship between two quantitative variables. It aids in comprehending how fluctuations in one variable correspond to changes in another. The correlation coefficient serves as a numerical indicator of the extent to which two variables are associated. In this study, PCA-WQI against DOE-WQI were analysed to illustrate the linear relationship of WQI in terms of PCA based results for three different transformations and raw data ser obtained from DOE. Then, the correlation analysis was conducted to evaluate the PCA-WQI and DOE-WQI. The correlation coefficient, often denoted by r, measures the strength and direction of the linear relationship between two variables. The correlation coefficient is a unitless measure, meaning it's not affected by the scale of the variables. Positive values indicate a positive correlation, while negative values indicate a negative correlation. It also provides a quantified measure of how strongly and in what direction two variables are related, making it a useful tool for analyzing relationships in data.

The eigenvalue, a key metric in PCA, indicates the quality of the projection and ensures that all factors are uncorrelated. Principal Component (PC) scores were derived by multiplying eigenvectors with the datasets. The selection of the most variant PC scores was guided by criteria set forth by Kambe et al., such as aiming for a cumulative proportion of 70-90%, excluding eigenvalues less than the mean, and discarding standard deviations below 1. This methodology is designed to capture a significant portion of the variance in the first two or three factors, reflecting the original data variability.

Loadings in PCA represent the linear transformation of standardized input variables, weighted to the principal components. These loadings are coordinated so that their squared sum equals one. To enhance the reliability of the Water Quality Index (WQI) assessment, normal distribution, as suggested by Wajid et al., was applied. This ensured an accurate distribution of values, which is particularly crucial in studies involving natural phenomena.

The final phase of the analysis involved a comparative assessment using crosstab analysis between the Department of Environment Water Quality Index (DOE-WQI) and PCA-derived WQI (PCA-WQI).

Building upon the correlation matrices, the study proceeds with the calculation of eigenvalues and eigenvectors. These are critical components in Principal Component Analysis (PCA), as they lay the groundwork for the derivation of principal component scores. To calculate these scores, the eigenvectors are multiplied by each of the four distinct datasets. This multiplication is an essential step in extracting meaningful patterns from the datasets. Once the principal component scores are obtained, they are aggregated to form the PCA-based Water Quality Index (PCA-WQI). This aggregation is done by summing the chosen principal components, employing the method suggested by Chow-Fraser. The PCA-WQI is then generated using a normal distribution approach. This PCA-WQI provides an alternative, nuanced view of water quality, differing from the traditional Department of Environment Water Quality Index (DOE-WQI).

To assess the relationship between the PCA-WQI and the DOE-WQI, a crosstab analysis is conducted. This analysis is instrumental in identifying the highest relation between these two indices. Additionally, biplot analysis and the correlation coefficients from correlation analysis were compared across PCA-WQI derived from four different standardization methods. The goal of this comparative analysis was to illuminate the accuracy, consistency, and reliability of these methods when correlated with DOE-WQI data, thereby validating the efficacy of our analytical approach. A correlation analysis was conducted to evaluate the predictive accuracy of the PCA-WQI in comparison to the DOE-WQI. **Figure 1** illustrates the flowchart of the analysis conducted using RStudio.

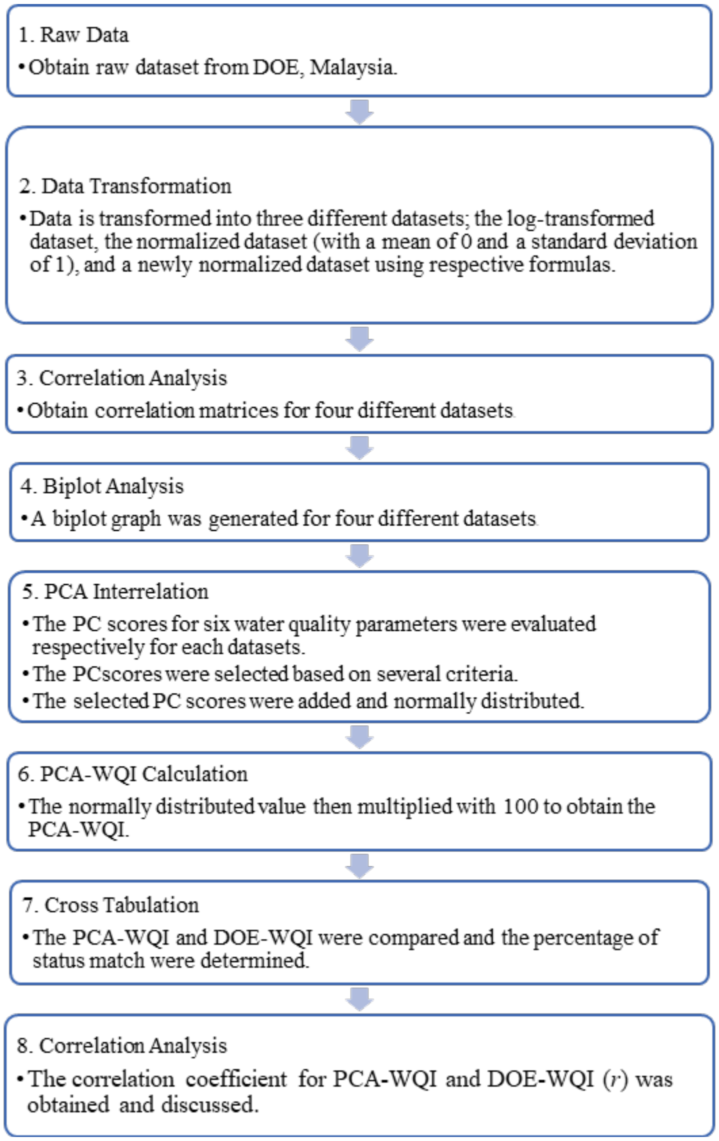

1. Raw Data
•Obtain raw dataset from DOE, Malaysia.
2. Data Transformation
•Data is transformed into three different datasets; the log-transformed dataset, the normalized dataset (with a mean of 0 and a standard deviation of 1), and a newly normalized dataset using respective formulas.
3. Correlation Analysis
•Obtain correlation matrices for four different datasets
4. Biplot Analysis
•A biplot graph was generated for four different datasets
5. PCA Interrelation
•The PC scores for six water quality parameters were evaluated respectively for each datasets.
•The PCscores were selected based on several criteria.
•The selected PC scores were added and normally distributed.
6. PCA-WQI Calculation
•The normally distributed value then multiplied with 100 to obtain the PCA-WQI.
7. Cross Tabulation
•The PCA-WQI and DOE-WQI were compared and the percentage of status match were determined.
8. Correlation Analysis
•The correlation coefficient for PCA-WQI and DOE-WQI ($r$) was obtained and discussed.

**Figure 1:** Flow chart of statistical analysis

# RESULTS AND DISCUSSION

## Correlation Analysis

The findings are systematically presented in Table 1, which is divided into three subsections: (a), (b), and (c). Each subsection details the correlation matrices obtained for the four distinct datasets, corresponding to the Buloh River, Langat River, and Jawi River, respectively. These matrices offer a comprehensive view of the inter-variable relationships within each dataset and provide a foundation for the subsequent PCA. By examining these correlation matrices, we gain valuable insights into how various water quality parameters interact and influence each other in each of the three rivers studied.

**Table 1(a):** Correlation matrix for (i) raw and normalized data and (ii) log transformed data for Buloh River

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| DO | 1 | -0.27 | -0.24 | -0.10 | 0.09 | -0.30 |
| BOD | | 1 | 0.92 | 0.42 | -0.12 | 0.05 |
| COD | | | 1 | 0.60 | -0.14 | 0.10 |
| SS | | | | 1 | -0.01 | 0.13 |
| PH | | | | | 1 | 0.18 |
| AN | | | | | | 1 |

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| DO | 1 | -0.41 | -0.39 | -0.35 | 0.05 | -0.31 |
| BOD | | 1 | 0.90 | 0.26 | -0.03 | 0.05 |
| COD | | | 1 | 0.30 | -0.10 | 0.07 |
| SS | | | | 1 | -0.18 | -0.07 |
| PH | | | | | 1 | 0.19 |
| AN | | | | | | 1 |

**Table 1(b):** Correlation matrix for (i) raw and normalized data and (ii) log transformed data for Langat River

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| DO | 1 | -0.43 | -0.43 | -0.16 | 0.33 | -0.49 |
| BOD | | 1 | 0.90 | 0.08 | 0.04 | 0.31 |
| COD | | | 1 | 0.07 | -0.07 | 0.31 |
| SS | | | | 1 | -0.01 | -0.03 |
| PH | | | | | 1 | 0.02 |
| AN | | | | | | 1 |

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| DO | 1 | -0.45 | -0.47 | -0.30 | 0.30 | -0.52 |
| BOD | | 1 | 0.84 | 0.22 | 0.03 | 0.36 |
| COD | | | 1 | 0.24 | -0.10 | 0.35 |
| SS | | | | 1 | -0.10 | 0.24 |
| PH | | | | | 1 | -0.02 |
| AN | | | | | | 1 |

**Table 1(c):** Correlation matrix for (i) raw and normalized data and (ii) log transformed data for Jawi River

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| DO | 1 | -0.21 | -0.06 | 0.06 | 0.33 | 0.12 |
| BOD | | 1 | 0.65 | -0.03 | 0.12 | 0.32 |
| COD | | | 1 | -0.24 | 0.20 | 0.22 |
| SS | | | | 1 | 0.09 | -0.12 |
| PH | | | | | 1 | 0.64 |
| AN | | | | | | 1 |

| | DO | BOD | COD | SS | PH | AN |
|---|---|---|---|---|---|---|
| DO | 1 | -0.19 | 0.11 | 0.02 | 0.16 | -0.23 |
| BOD | | 1 | 0.50 | -0.10 | 0.09 | 0.43 |
| COD | | | 1 | -0.31 | 0.08 | 0.08 |
| SS | | | | 1 | 0.11 | -0.17 |
| PH | | | | | 1 | 0.54 |
| AN | | | | | | 1 |

The analysis of the correlation matrices, as presented in Tables 1(a), (b), and (c), reveals interesting insights into the four datasets: The raw dataset, the dataset normalized to a mean of 0 and standard deviation of 1, the new normalization dataset, and the log-transformed dataset. It is observed that the correlation matrices for the raw dataset, the normalized dataset, and the new normalization dataset exhibit similar patterns. In contrast, the log-transformed dataset demonstrates notable variations.

Through the application of Principal Component Analysis (PCA), we sought to elucidate the interrelationships among the six key water quality parameters. The correlation analysis indicates a high degree of collinearity between the standardization and new standardization datasets with the raw dataset. This finding underscores the effectiveness of these standardization methods in maintaining the integrity of the original data.

Furthermore, the datasets were subjected to dimensional reduction, successfully transforming the data into uncorrelated variables while simultaneously normalizing them. This process resulted in minimal loss of information, preserving the essence of the original dataset. In summary, the correlation analysis demonstrates that the data was meticulously handled, particularly for the standardized and new standardized datasets. This careful data management ensures the validity of our analytical approach and the reliability of the resulting model, aligning well with the standards and requirements of the Department of Environment Malaysia. This alignment is critical in verifying the robustness of the study's findings and their applicability in real-world water quality assessments.

## Crosstab Analysis

The findings of this comparative evaluation are systematically presented in Table 2. By juxtaposing the PCA-WQI with the DOE-WQI, the study aims to verify the consistency and reliability of the PCA approach in water quality assessment. This comparative analysis is crucial in understanding the extent to which PCA-WQI aligns with or diverges from the established DOE-WQI, offering valuable insights into the effectiveness of PCA in environmental data analysis.

**Table 2:** Cross Tabulation Analysis Based on Standardization Method and Principal Component scores

| PC Scores | Standardization Method | | | |
|---|---|---|---|---|
| | Raw (%) | Log (%) | Normalization (%) | New Normalisation (%) |
| Buloh River | 67.27 | 40.91 | 60.00 | 71.82 |
| Langat River | 64.57 | 55.48 | 74.73 | 79.37 |
| Jawi River | 62.07 | 58.62 | 68.96 | 72.73 |

**Table 2** reveals a significant concordance between PCA-WQI and DOE-WQI across different datasets, with the new normalization datasets showing an impressive agreement of 71.82% for Buloh River, 79.37% for Langat River, and 72.73% for Jawi River. This high degree of consistency underscores the reliability of the PCA interrelation values derived from the new normalization method, especially when compared to the raw datasets, normalized datasets, and log-transformed datasets. Although there are some discrepancies in their respective calculation methodologies, both the normalization and new normalization methods follow a fundamental standardization framework. Consequently, the amalgamation of selected principal components using the new normalization approach resulted in the most significant alignment with DOE-WQI. In contrast, the log-transformed dataset exhibited the lowest degree of alignment. This finding highlights the efficacy of the new normalization method in aligning with established water quality indices, demonstrating its potential as a robust tool in water quality assessment.

## Biplot Analysis

A biplot serves as a visual representation presented in a scatterplot format, capturing the variance between two distinct variables. In the context of this study, the variability between PC1 and PC2 was portrayed via a biplot across four diverse datasets, as illustrated in Figure 2-4 for respective selected rivers. These biplots are graphical representations of Principal Component Analysis (PCA) results, showcasing relationships between the first two principal components (PC1 and PCS) for each dataset.

The progression of the biplots from (a) to (d) in Figures 2-4 reveals the effects of different data preprocessing techniques on the PCA of Buloh River, Jawi River and Langat River respectively. The raw data serves as a baseline for understanding the intrinsic variability of water quality parameters. As the data is transformed through log normalization and new normalization techniques, the distribution and orientation of the data points and vector shift, indicating changes in how the water quality parameters are represented and correlated.

The biplots provide a comprehensive view of how each preprocessing method impacts the PCA results. For instance, the log-transformed data tends to spread out the data points more along the principal components, which could help in identifying outliers or patterns not apparent in the raw data. The new normalized data, by balancing the contribution of the variables, might offer a refined perspective that potentially avoids the dominance of any single water quality parameter.

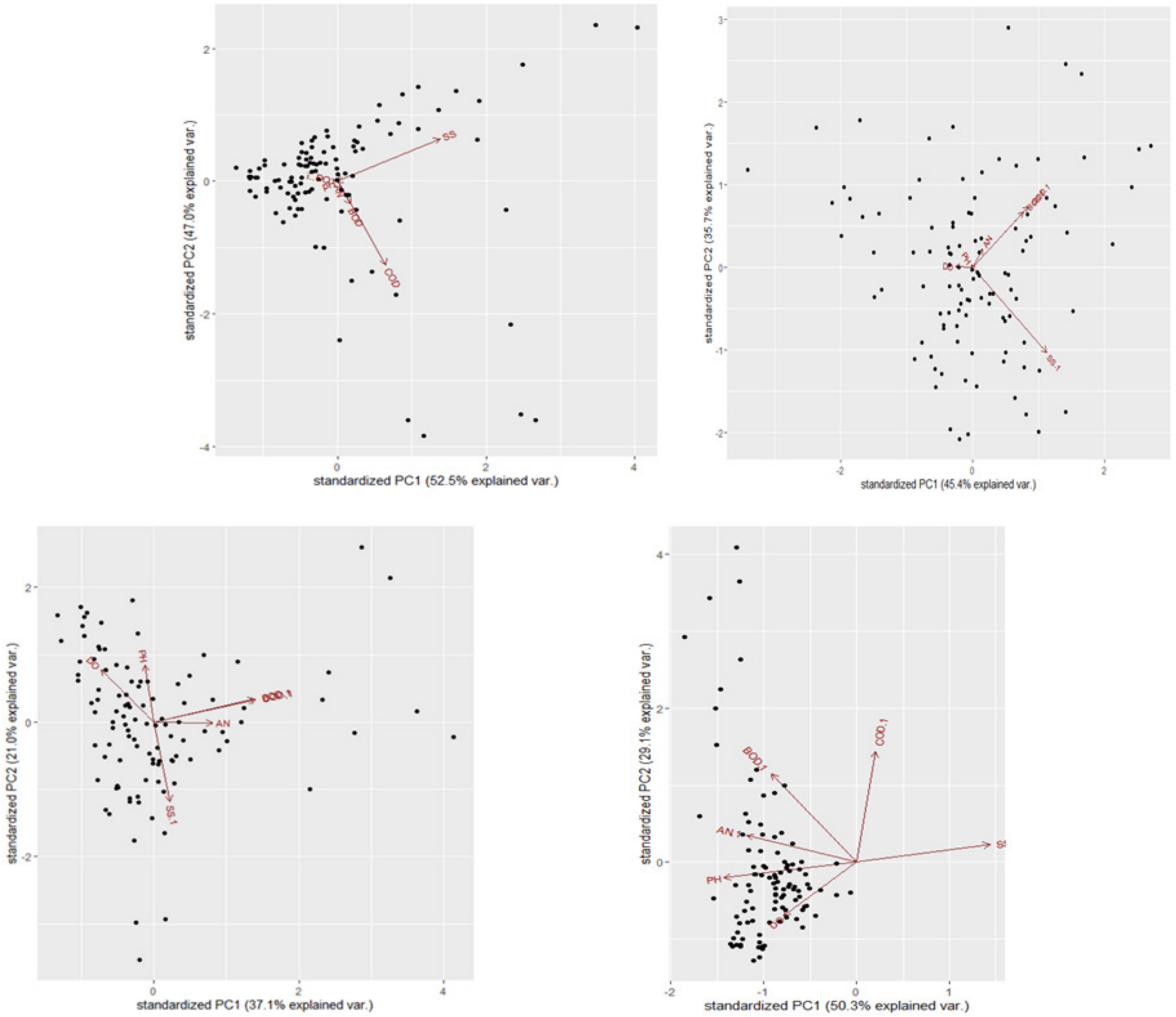
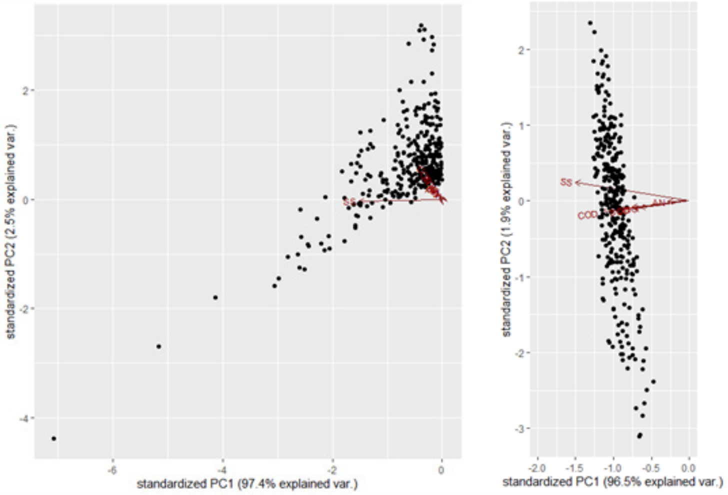


**Figure 2:** Biplot of PCA axis, PC1 and PC2 of (a) Raw data (b) Log Transformed Data (c) Normalised Data and (d) New Normalised Data for Buloh River



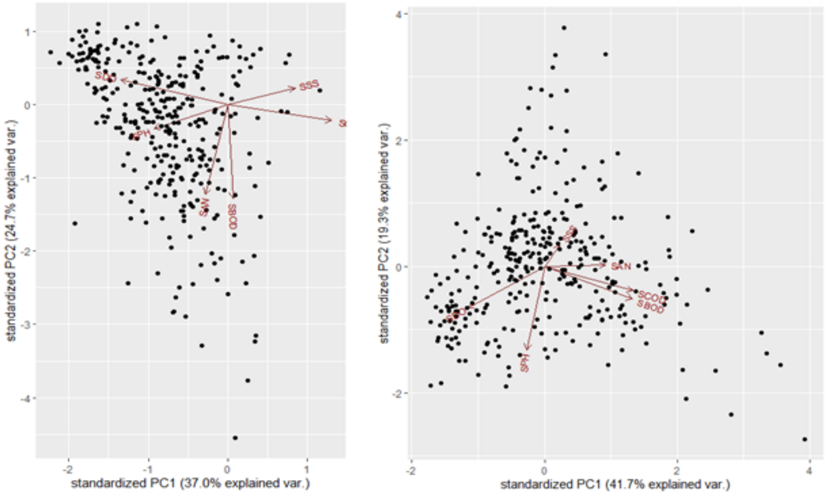


**Figure 3:** Biplot of PCA axis, PC1 and PC2 of (a) Raw data (b) Log Transformed Data (c) Normalised Data and (d) New Normalised Data for Langat River

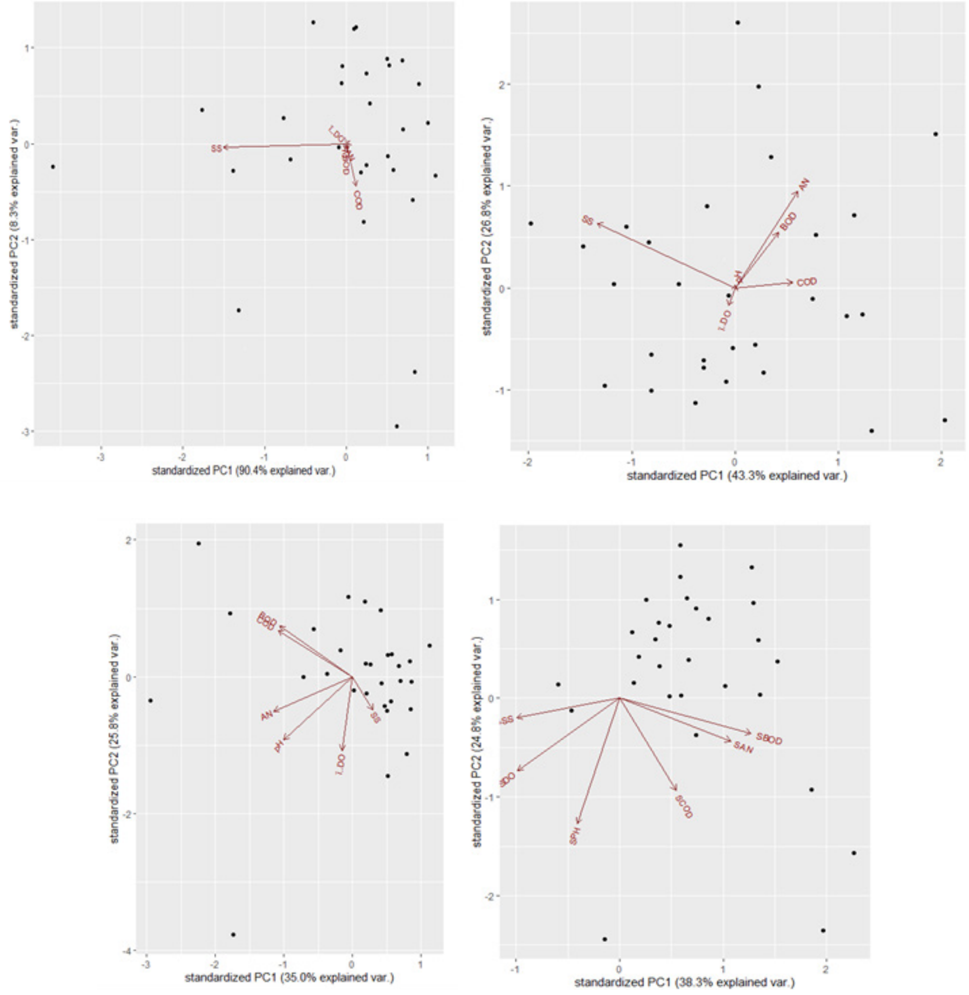


**Figure 4:** Biplot of PCA axis, PC1 and PC2 of (a) Raw data (b) Log Transformed Data (c) Normalised Data and (d) New Normalised Data for Jawi River

The initial Principal Component (PC) consistently captures the highest variance across the datasets, emphasizing a strong correlation between the variables. This robust correlation is further validated by a detailed examination of the absolute values and averages. The PCA biplots, which illustrate the sampling sites, reveal that the primary PCA axis predominantly governs the variance. Specifically, Suspended Solids and Chemical Oxygen Demand exert the most influence, with the highest absolute values on the first axis for the Buloh, Langat, and Jawi Rivers. However, the second PCA axis shows minimal correlation with the variables, contributing to only a minor portion of the overall variance for each river.

In Figure 1(b), the variance explained by the first and second PCA axes is 45.4% and 35.7%, respectively. When the log-transformed dataset is considered, there's a noticeable decrease in the eigenvalues, indicating a broader spread of samples along the two axes. This pattern persists in the biplots for the Langat and Jawi Rivers, where the variance explained by the PCA axes is substantially lower in the log-transformed data compared to the raw dataset.

Figures 1(c), 2(c), and 3(c) present the standardized data, showing the variance explained by the first and second PCA axes for each river. The normalized dataset, adjusted to a mean of 0 and a standard deviation of 1, exhibits the lowest values on the first axis and a wider spread on the second axis. Notably, the influence of smaller variables on the second PCA axis is diminished, resulting in an even distribution of samples across the first two dimensions. This standardization ensures that no single water quality parameter unduly influences the data.

Figure 1(d) and its counterparts for the other rivers demonstrate that the new normalization dataset creates a distinct PCA ordination compared to the raw, log-transformed, and standardized datasets. With most samples clustered around the origin and exhibiting low correlation, the data suggest that the water quality parameters are independent of each other. Moreover, the new normalization method reveals that no single parameter dominates the computation of the Water Quality Index (WQI); instead, each parameter contributes individually. The biplots from the new standardization method provide a refined visualization, where the water quality parameters are distinctly separated, signifying their individual and non-overlapping impact on water quality assessment.

These biplots provide crucial insights into the nature of water quality data for Buloh River. The varying degrees of variance explained by the principal components across different datasets highlight the impact of data transformation techniques on the PCA results. The findings suggest that normalization significantly influences the distribution and representation of water quality parameters, which could lead to different interpretations and conclusions regarding the status the pollution levels of Buloh River. In practical terms, these visualizations help in identifying which water quality parameters are most influential and how different preprocessing method can affect the outcome of PCA, ultimately guiding environmental scientist and policymakers in making more informed decisions about water quality management and intervention strategies.

In conclusion, these biplots for the selected rivers demonstrate the transformative effect of data preprocessing on PCA outcomes. They serve as a powerful diagnostic tool for environmental scientists, allowing for the assessment of water quality data's structure and the selection of appropriate preprocessing methods for further analysis. Such insights are essential for accurately interpreting environmental data and guiding effective water resource management.

## Correlation Analysis

**Table 3** furnishes the correlation coefficient for four different datasets.

**Table 3:** Correlation coefficient for different standardization methods for Buloh, Langat and Jawi River

| Correlation coefficient (r) | Standardization Method | | | |
|---|---|---|---|---|
| | Raw | Log | Normalisation | New Normalisation |
| Buloh River | 0.87 | 0.69 | 0.87 | 0.87 |
| Lanagt River | 0.86 | 0.77 | 0.86 | 0.86 |
| Jawi River | 0.82 | 0.74 | 0.82 | 0.82 |

However, the log transformation appears to reduce the explanatory power of the models, particularly for the Langat and Jawi Rivers. This suggests that log transformation may not effectively capture the variability in their water quality data.

The correlation coefficients for the normalized and new normalization datasets demonstrated high values, achieving 0.91 for Buloh River, 0.86 for Langat River, and 0.83 for Jawi River. These figures suggest that the new normalization dataset, in particular, offers a high degree of variable independence and an enhanced ability to represent water quality accurately. Furthermore, the PCA-WQI, when employing the new normalization method, aligns closely with the water quality status as determined by the Department of Environment, showing a consistent and reliable match.

Normalization and new normalization methods both preserve or even enhance the explanatory power of the models across all three rivers. They match or nearly mirror the high coefficients of correlation observed with the raw data models. This consistency is evident for the Buloh and Jawi Rivers, underscoring the robustness of the standardization methods applied. In contrast, the lower coefficient for the Langat River in the log-transformed data suggests the necessity for a more appropriate method to accurately capture its variability.

In summary, while models based on raw data are effective, normalization methods prove to be equally capable of elucidating the variance in water quality for the studied rivers. Log transformation, however, demonstrates less reliability, particularly for the Langat and Jawi Rivers. These findings could inform future water quality monitoring and management strategies for these rivers.

## Validation of the PCA-WQI

The effectiveness of the innovative standardization method has been rigorously assessed through four distinct analyses: Correlation analysis, crosstab analysis and biplot analysis. Each of these analyses consistently demonstrates the high accuracy and reliability of the new method implemented in the water quality model. Notably, the new standardization method was meticulously applied to examine the Buloh River, Langat River, and Jawi River, all classified as highly polluted by the Department of Environment (DOE).

Using the newly standardized values within Principal Component Analysis (PCA) interrelations significantly enhanced confidence, particularly for DOE-WQI. The proposed method incorporates the first two Principal Component (PC) scores into calculations, as they contribute to more than 70% of the weighted principal component scores.

Biplot analysis reveals that both the raw and log-transformed datasets are predominantly influenced by two water quality parameters, specifically Suspended Solids (SS) and Chemical Oxygen Demand (COD), highlighting centered PCA interrelations. In contrast, the normalization and the new normalization method, using non-centered PCA interrelations, exhibit no dominant water quality parameters. This suggests independence between the six DOE-specified parameters.

While centered PCA captures the maximum variance in the input data through its first principal component, non-centered PCA aligns this component primarily with the data sets' mean. The PCA interrelations derived from the new standardization method generate a PCA-WQI exhibiting the highest compatibility with PCA-WQI, as shown in Figure 5-7.

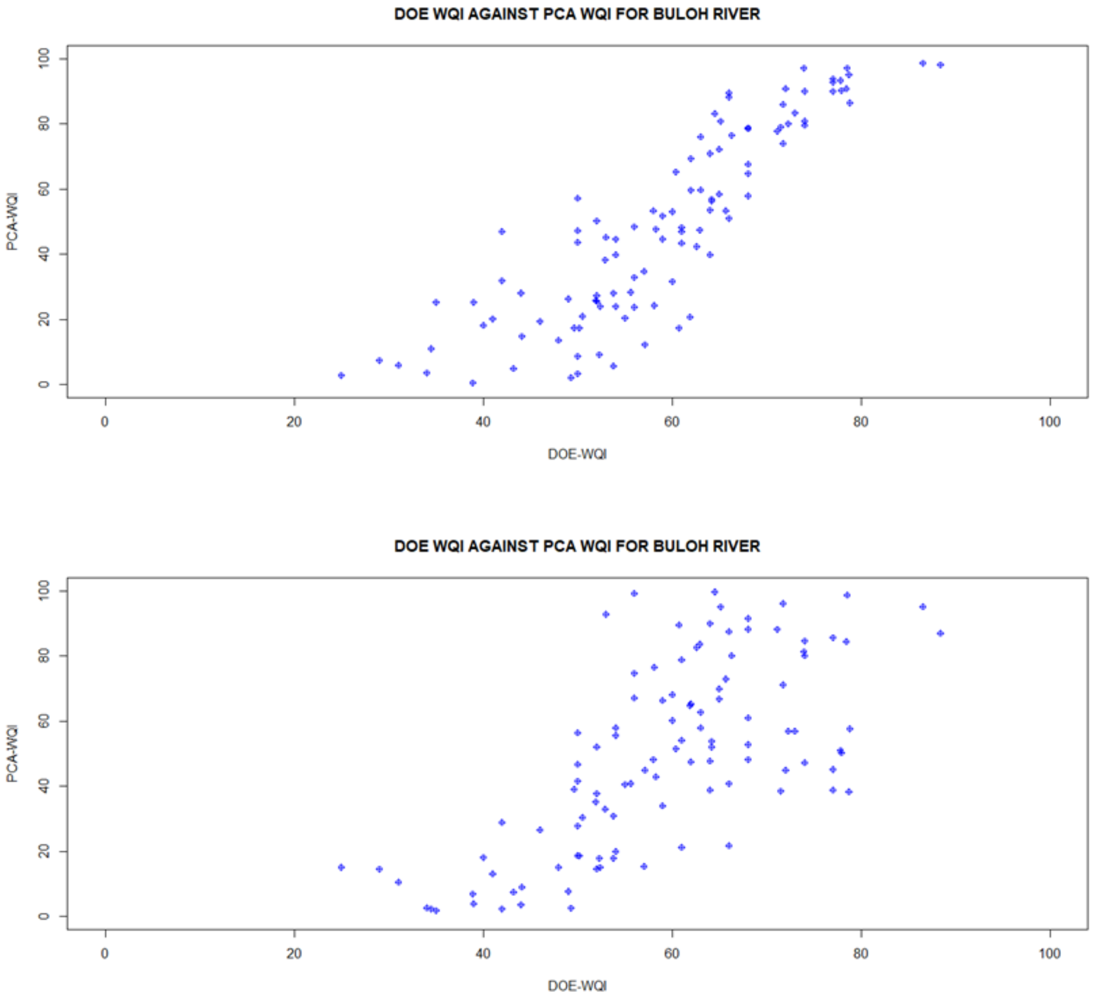
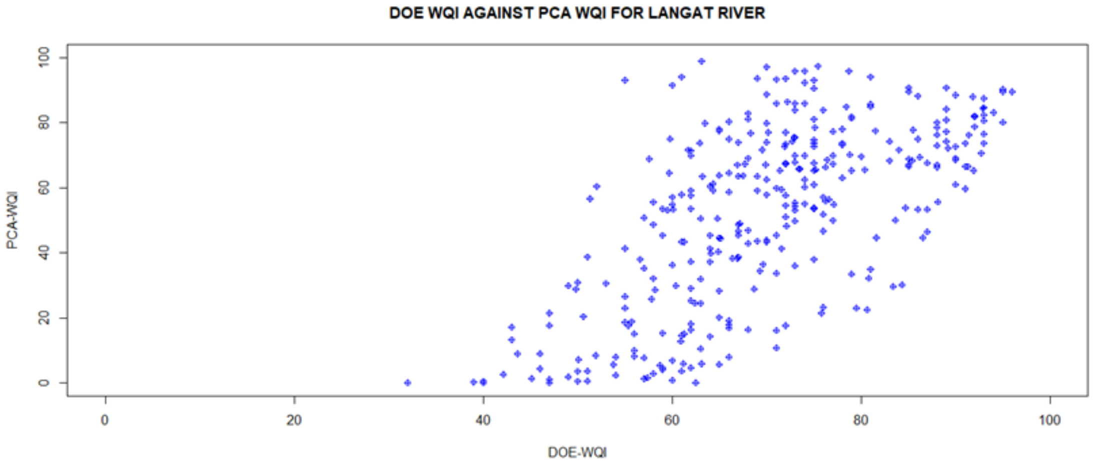


**Figure 5:** Scatter plot for (i) raw and normalized data and (ii) log transformed data for Buloh River



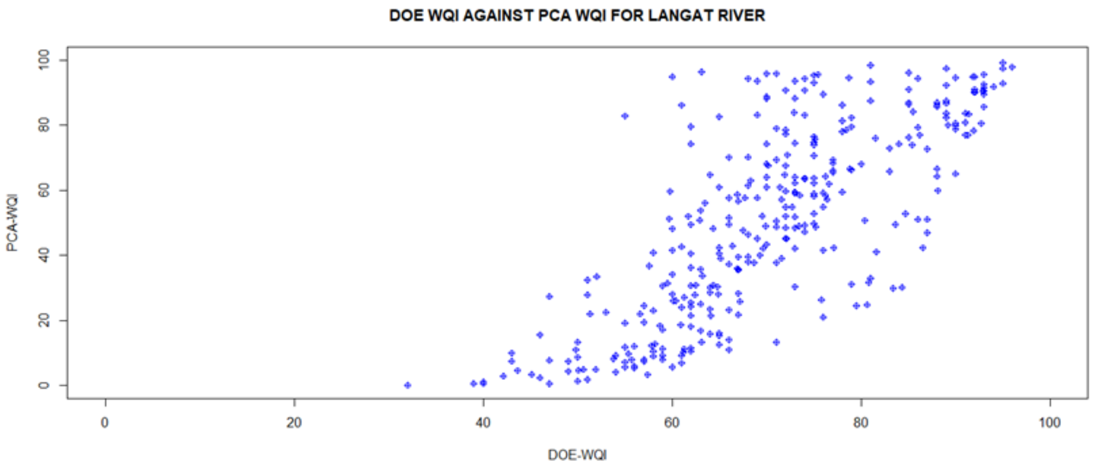


**Figure 6:** Scatter plot for (i) raw and normalized data and (ii) log transformed data for Langat River

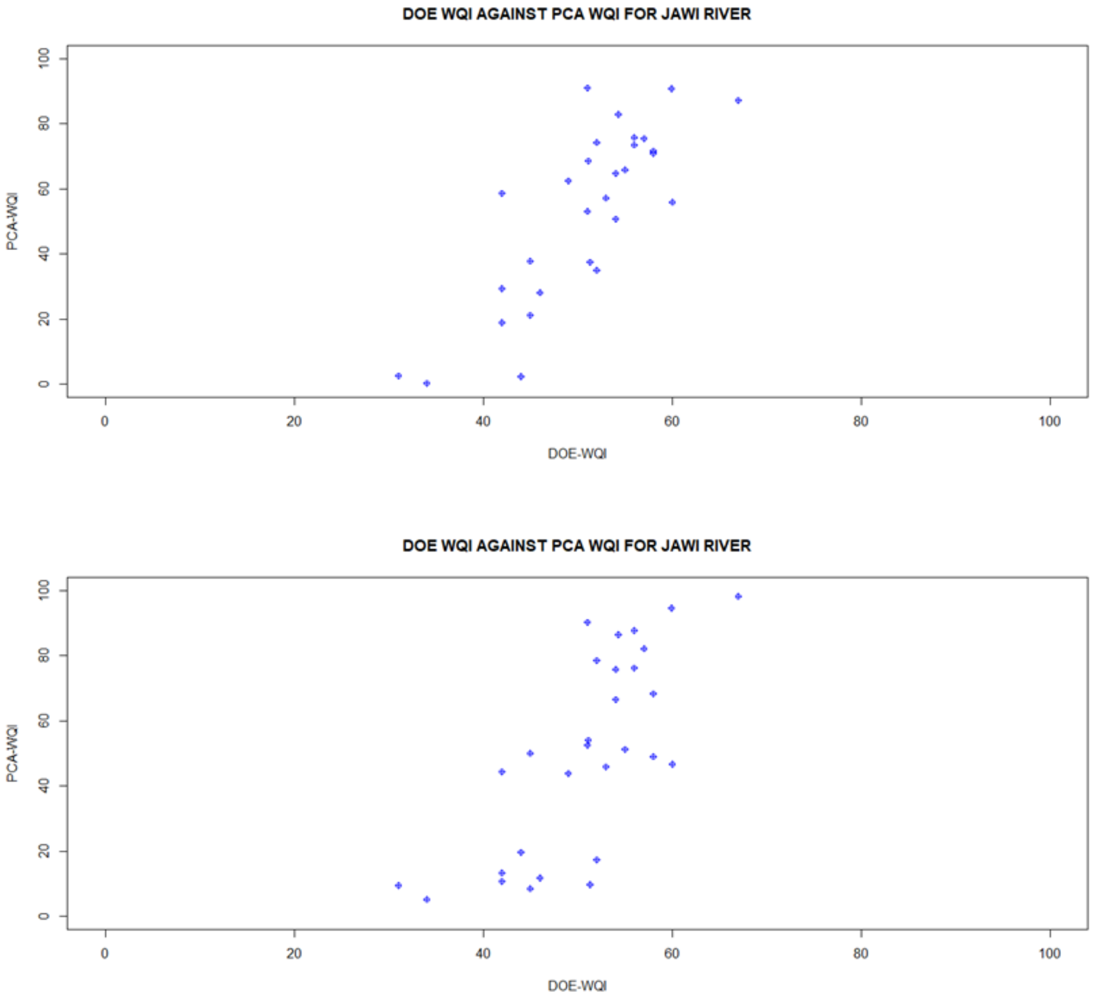


**Figure 7:** Scatter plot for (i) raw and normalized data and (ii) log transformed data for Jawi River

Figure 5-7 paints a compelling picture of the stability and conclusiveness of both DOE-WQI and PCA-WQI when empowered by the new standardization method. This decisive outcome resonates with a 2013 study by Zalina et al., which revealed a consistent linear relationship between PCA-WQI and DOE-WQI, solidifying the reliability and persistence of the findings. Likewise, Figure 2's scatter plot vividly illustrates a linear kinship between PCA-WQI and DOE-WQI for PCA interrelations facilitated by the new method, where higher PCA-WQI values champion better water quality.

Furthermore, the normalized dataset reveals PCA interrelations among water quality parameters that align closely with a normal distribution. By leveraging the 'area under the curve' concept, which reflects this statistical harmony, pollution levels can be effectively quantified into a Water Quality Index (WQI) on a scale from 0 to 100. Here, a score of 0 indicates 'poor' quality, while higher scores approach 'pristine' water conditions. This method provides a clear and accessible metric for assessing water quality.

This investigation revealed a significant interplay among the six water quality parameters and Principal Components (PCs) under the new standardization method. The leading PC showed substantial positive loadings for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH, and Ammoniacal Nitrogen (AN), whereas Suspended Solids (SS) exhibited a notable negative loading. These five variables collectively accounted for the majority of the variance in the initial phase of the PCA analysis. The second PC, characterized by COD's strong positive loading, explained a significant portion of the variance in the subsequent phase. These findings are in agreement with the results of the correlation analysis, supporting a unified interpretation.

The correlation coefficients, reaching impressive levels of 87%, 86%, and 82% for the Buloh, Langat, and Jawi Rivers respectively, when comparing PCA-WQI and DOE-WQI using the new standardization method, suggest that the six water quality parameters function as distinct variables, each contributing uniquely to the overall analysis. This distinctiveness highlights the method's effectiveness as a robust and reliable tool for water quality modelling. Beyond its high accuracy, the new method demonstrates increased reliability and consistency throughout the analytical process. Notably, its systematic outcomes, adeptly aligned with each river's specific characteristics, further enhance its practical utility.

Significantly, the new standardization method, specifically designed to align with the unique characteristics of the river, conclusively supports the Department of Environment (DOE)'s classification of the Langat River as polluted. The PCA interrelations, facilitated by this method, consistently yield results that are in complete agreement with the DOE-WQI across all four analyses, including correlation analysis, crosstabulation analysis, and biplot analysis. Thus, the PCA interrelations, enhanced by the new standardization method, provide a consistent and reliable water quality model. This model could be effectively utilized by the Department of Environment, Malaysia, to protect the country's vital water resources.

# CONCLUSION

This study introduces PCA-WQI as an innovative tool for evaluating DOE-WQI, utilizing a suite of analytical techniques including correlation, crosstab and biplot analyses to highlight the enhanced performance of new normalization datasets over raw, log-transformed, and normalized datasets. The PCA-WQI, when applied through the new normalization method, exhibited unparalleled accuracy and effectiveness. The primary objective was to develop a robust water quality model for the Langat River, informed by data spanning five years (2015-2020). A meticulous analysis led to the integration of a normally distributed water quality index, as used by the Department of Environment Malaysia, into the study, ensuring the exclusion of outliers.

The analysis revealed that key water quality parameters, such as BOD, COD, and AN, had significant positive loadings, pointing to pollution from animal waste, agriculture, and industrial discharges as major contributors to the degradation of water quality in the Langat River. This pollution has dire consequences for the ecosystem and the quality of life for local communities, as evidenced by the increased oxygen consumption by organic matter and the resulting decline in dissolved oxygen levels. In contrast, pH levels, a marker of the river's natural condition, remained relatively stable.

The advanced WQI model proposed herein aims to refine Malaysia's existing water quality framework by incorporating multivariate analysis with ecological relevance. Notably, BOD, AN, and SS emerged as critical indicators of water quality, reflecting the impact of pollutant discharges from both point and non-point sources. The PCA interrelations derived from the new standardization method showed superior accuracy, aligning with the river water quality classes designated by the National Water Quality Standards for Malaysia.

Ultimately, the PCA-WQI model, devised with the new standardization method, is poised for broader application across Malaysian rivers, offering deeper insights for strategic water management decisions. The model's precision and reliability have been affirmed, laying the groundwork for future enhancements. Potential improvements could include broadening the classification ranges to extend beyond the basic parameters set by the NWQSM and incorporating a wider array of biological and chemical parameters to establish an even more comprehensive water quality model.

# ACKNOWLEDGEMENT

This manuscript constitutes a segment of a research initiative financially supported by the Graduate Research Program (GGPM-2021-025) granted by Universiti Kebangsaan Malaysia. The authors extend their sincere appreciation to all individuals and entities, both directly and indirectly associated with this research endeavor. Furthermore, the Department of Environment (DOE) and Universiti Kebangsaan Malaysia (UKM) provided additional support for this study.

# ACKNOWLEDGEMENT

None.

# CONFLICTS OF INTEREST

None.

# REFERENCES

1. Aminu I, Ismail A, Juahir H, et al. (2023) Water quality modelling using principal component analysis and artificial neural network. Mar Pollut Bull. 187.
2. Karakostas A, Moumtzidou A (2021) Why data standardization is crucial for the supply of safe and secure water. Int Water Assoc.
3. Cullota A, Wick M, Hall R, et al. (2019) Canonicalization of database records using adaptive similarity measures (Modified). Proc 13th ACM SIGKDD Int Conf Knowl Discov Data Min. 201-209.
4. Camacho J, Smilde AK, Saccenti E, et al. (2021) All sparse PCA models are wrong, but some are useful. Part II: Limitations and problems of deflation. Chemometr Intell Lab Syst.
5. Camacho J, Smilde AK, Saccenti E, et al. (2020) All sparse PCA models are wrong, but some are useful. Part I: Computation of scores, residuals and explained variance. Chemometr Intell Lab Syst.
6. Garcia CAB, Silva IS, Mendonca MCS, et al. (2018) Evaluation of water quality indices: Use, evaluation and future perspectives. Adv Environ Monit Assess. 21-33.

7. Chow-Fraser P (2006) Development of the wetland water quality index for assessing the quality of Great Lakes coastal wetlands. Health Habitat Indicators. 137-166.
8. Read EK, Carr L, De Cicco L, et al. (2017) Water quality data for national-scale aquatic research: The Water Quality Portal. Water Resour Res. 53:1735-1745.
9. Juahir H, Zain SM, Yusoff MK, et al. (2011) Spatial water quality assessment of Langat River Basin. Environ Monit Assess. 625-641.
10. Paun J, Cruceru LV, Chiriac FL, et al. (2016) Water quality indices - methods for evaluating the quality of drinking water. Proc Environ Industry. 395-400.
11. Soumaila KI, Niandou AS, Naimi M, et al. (2019) A systematic review and meta-analysis of water quality indices. J Agric Sci Technol B. 9:1-14.
12. Malaysia SA (2020) Jawi River. FOE Malaysia, Penang.
13. Tripathi M, Signal SK (2019) Use of principal component analysis for parameter selection for development of a novel water quality index: S case study of River Ganga India. Ecol Indic. 96:430-436.
14. Abidin MZ, Kutty AA, Lihan T, et al. (2018) Hydrological change effects on Sungai Langat water quality. Sains Malaysiana. 1401-1411.
15. Uddin MG, Nash S, Olbert AI (2021) A review of water quality index models and their use for assessing surface water quality. Ecol Indic. 122.
16. Greenacre M, Groenen PJF, Hastie T, et al. (2023) Principal component analysis. Nat Rev Methods Primers.
17. Gal MS, Rubinfeld DL (2019) Data standardization. NYU Law Econ Res Pap.
18. Noh NSM, Sidek LM, Mohiyaden HAM, et al. (2020) Water pollution and water quality assessment based on water quality index method for Johor Rivers. J Energy Environ.
19. Van den Berg RA, Hoefsllot HCJ, Westerhuis JA, et al. (2006) Centering, scaling, and transformation: Improving the biological information content of metabolomics data. BMC Genomics. 7:142.
20. Roy R (2019) An introduction to water quality analysis. Int Res J Eng Technol. 201-205.
21. Sankpal KA (2020) A review on data normalization techniques. Int J Eng Res Technol. 115-118.
22. Sekitar JA (2022) Laporan Kualiti Alam Sekeliling 2022. Jabatan Alam Sekitar, Malaysia.
23. An S, Xie X, Ma Y (2015) Evaluation of water quality using principal component analysis. Nat Environ Pollut Technol. 855-858.
24. Aoyama T, Kambe J, Nagashima U (2007) Journal of Computer Chemistry Japan. 6:19-26.
25. Ali W, Nafees M, Turab SA, et al. (2019) Drinking water quality assessment using water quality index and geostatistical methods. J Himalayan Earth Sci. 52(1):65-89.
26. Widiyanti BL (2022) Assessment of water quality and its effect on health. Proc Int Conf Disaster Manage Clim Change. IOP Publ.
27. Fang X, Hu J, Sharma S (2023) Water quality modeling and monitoring. Water. 15(18):3216.
28. Cao Y, Williams DD, Williams NE (1999) Data transformation and standardization in the multivariate analysis of river water quality. Ecol Appl. 9(2):669-677.
29. Ali ZM, Ibrahim NA, Mengersen K, et al. (2013) The Langat River water quality index based on principal component analysis. AIP Conf Proc. 1522:1322-1336.
30. Ali ZM, Ibrahim NA, Mengersen K, et al. (2014) Robust principal component analysis in water quality index development. AIP Conf Proc. 1091-1097.

**How to Cite this article:** Khayell B, et al. (2026) Revolutionizing River Water Quality Assessment Through Novel PCA Integration and Standardized WQI Metrics in Malaysian River. Heal Care Res Case Rep J. 2(1): 1-14.