RT&A, No.3 (74)
Volume 18, September 2023

ASSESSMENT OF WATER QUALITY USING MULTIVARIATE TECHNIQUES

*[1]Olamiji T. Onafowokan, [2]Kazeem O. Obisesan & [3]Oladapo M. Oladoja

•

*[1,2,3]Department of Statistics, University of Ibadan, Ibadan, Nigeria
[2,3]Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria
*[1]olamijionafowokan17@gmail.com, [2]obidairo@gmail.com,
[3]oladapo.oladoja@tech-u.edu.ng

Abstract

When deciding if water is suitable for a particular usage, its quality"which includes its chemical, physical, and biological characteristics"is referred to. The quality of the water is influenced by many natural and human influences. Despite being in equilibrium, the natural ecosystem and water quality would certainly be disturbed by any large changes in the water quality. In order to assess the levels of water pollution in the Asejire and Eleyele reservoirs, this study conducted a Physico-chemical analysis of the two reservoirs. It also used multivariate techniques to identify the causes of water pollution in the two reservoirs under investigation, used a generalized linear model to analyze the variability in turbidity levels, and suggested regulatory solutions to address water pollution in the two reservoirs under study. In Ibadan, which has a population of about four million, the two main sources of pipe-borne water are the Eleyele and Asejire reservoirs. Between January 2003 and August 2019, water samples were taken from both locations and analyzed for 13 Physico-chemical parameters using the Principal Component Analysis and Cluster Analysis for feature extraction and finally a Generalized Linear model for prediction. Basic Tables and descriptive plots, Principal Component Analysis, Factor Analysis, and Generalized Linear Models were employed. Results: In the Asejire and Eleyele reservoirs, respectively, the PCA yields 5 significant main components explaining 76.56% and 60.97% of the variance, while the FA yields 5 significant major components explaining 94.90% and 79.97%. A generalized linear model (GLM) was used to study the variability in turbidity level, and the results indicate that two parameters"Iron and Silicon"in the Asejire reservoir are crucial for understanding turbidity variation and four"Colour, Alkaline, Silica, and Solids"contribute significantly to turbidity in the water level in the Eleyele Reservoir. With the exception of dissolved oxygen from either reservoir (Eleyele or Asejire) and iron from Eleyele Reservoir, many metrics in Asejire are within SON and WHO acceptable limits. This suggests that the water in the Eleyele reservoir is more contaminated than the Asejire reservoir.

Keywords: Turbidity, Reservoir, Pollution, Contaminated, Principal Component Analysis, Factor, Generalized Linear Model, Physico-chemical.

1. Introduction

Water is required for the development of water transportation, food production, industrial operations, and the development of renewable energy sources, and it is commonly associated with economic progress and human health. As a result, it is critical for human growth as a whole. The survival of living things, the health of ecosystems, the sustainability of human settlements, and economic progress all rely on having access to safe, clean, and sufficient water. However, as the human population has expanded and the demand for additional water has increased, industrial activity, agricultural activity, and climate change have all had a negative impact on water supply and quality (Oketola [1]). However, all human-caused activities have been shown to have a deleterious impact on water quality in both groundwater and surface water. Domestic use, agricultural operations, and industrial activities are the three main contributors to water contamination, according to the United Nations (2002) [2]. For example, excessive fertilizer use in agricultural activities has been shown to harm human health (Oketola [1]; UNEP [2]). While agriculture is the largest consumer of water, it also contributes significantly to water contamination. In essence, it pollutes water by releasing wastes (bacteria and viruses) from farms into waterways, posing a significant threat to water quality and harming people and wildlife (Nancy [3]).

An indicator of the quality of water is its chemical, physical, and biological characteristics, commonly as a measure of its suitability for a particular purpose. Water quality, according to World Health Organization in 2018 (WHO [4]), can be described by a variety of criteria that limit water use; it is a phrase used to demonstrate the suitability of water to validate certain purposes or processes. The researchers also stated that human and natural factors affect water quality. In regard to water availability and quality, geological, hydrological, and climatic factors are most important (Boyacioglu [5]). However, even when water is available in sufficient amounts, its poor quality restricts the uses that may be made. There is a greater need to maximize the use of limited water resources when available quantities are low. Even though the natural ecosystem and the water quality are in harmony, any large changes to the water quality are typically harmful to the environment.

2. Parameters of Water Quality

The qualities of water include turbidity, which is a fluid's cloudiness or haziness caused by countless small particles that are normally invisible to the human eye, similar to smoke in the air. The measuring of turbidity is an essential water quality test. For aesthetic reasons, water color is primarily an issue for water quality. The perception that colored water is unfit for consumption exists even when it is completely safe for ingestion by the general public. Although iron is a glossy, ductile, malleable, silver-gray metal (group VIII of the periodic table), color can also signal the presence of organic materials like algae or humic chemicals. It is known to exist in four different crystalline forms. To describe how acidic or alkaline a solution is, the pH scale is utilized. It is scored on a scale from 0 to 14. The term pH is made up of the letters "p," which in mathematics represent for negative logarithm, and "H," which chemically stands for hydrogen. The pH range that works best, which is normally between 6.5 and 9.5, is affected by the make-up of the water as well as the materials used to construct the distribution system. Similar to animals and people who live on land, aquatic animals need oxygen to survive. Oxygen from the atmosphere that has dissolved in river and lake water is absorbed by fish and other aquatic species.

Water flowing over rocks in creeks and rivers can introduce oxygen to the water. Since it enhances the taste of the water, a high dissolved oxygen water supply for the community is advantageous. However, high dissolved oxygen concentrations speed up the corrosion of water pipes. Healthy water typically has dissolved oxygen concentrations between 80 and 120 percent and over 6.5-8 mg/L. Total solids, or "TS," is the abbreviation for the sum of the dissolved, colloidal, and suspended solids in a sample of water. This consists of dissolved salts like sodium chloride, or NaCl, as well as solid particles like plankton and silt. An excessive amount of total solids in rivers and streams is a problem that arises rather frequently. The most common contaminant in the investigated streams and rivers is siltation, one of the major contributors to total solids, according to the Environmental Protection Agency's National Water Quality Inventory. The dry weight of non-dissolved suspended particles in a water sample that can be filtered out and measured using a filtering system is known as total suspended solids (TSS). It is a metric for measuring water quality that may be applied to any sort of water or body of water, including ocean water and wastewater that has undergone wastewater treatment. TSS was formerly known as non-filterable residue (NFR), but it was changed to TSS due to misunderstandings in other scientific fields. The National Drinking Water Quality Standard (NDWQS) has established a maximum recommended TSS limit of 25 mg/L.

A silicon and oxygen molecule, or silicon dioxide, is known as silica $(Si0_2)$. It is a tough, glassy mineral that can be found in sand, quartz, sandstone, and granite, among other forms. It is also present in the skeletons of both plants and animals. The majority of water supplies will contain some silica because it is the second most common element on Earth after oxygen. In every system of natural water, some is dissolved. A chemical indicator of water's capacity to neutralize acids is alkalinity. Alkalinity gauges a water's resistance to pH changes brought on by the addition of acids or bases. Strong bases (such OH-) may occasionally play a role in severe situations, but weak acid salts are primarily responsible for naturally alkaline water.

The two main causes of alkalinity in natural rivers are the partitioning of CO_2 from the atmosphere and the weathering of carbonate minerals in rocks and soil. Weak acid salts that may be present in trace amounts include borate, silicates, ammonia, phosphates, and organic bases produced from naturally existing organic compounds. The amount of calcium (Ca+2) ions in water expressed as calcium carbonate $(CaCO_3)$ is referred to as the calcium hardness. Calcium must be present in the pool water at a specified amount. Calcium and other minerals are dissolved from plaster pool surfaces and metal equipment parts when the calcium level in the water is too low (soft water). Calcium carbonate scale can develop on pool surfaces and recirculation equipment, especially heat-exchanging surfaces, when there is an excessive amount of calcium present (hard water, supersaturated). It is recommended to maintain calcium hardness levels between 150 and 1000 ppm. The optimal range for calcium hardness is 200 to 400 ppm.

Several research have used multivariate statistics to better understand natural and anthropogenic water contamination causes (Praus, 2007 [6]; Boyacioglu, 2008 [5]; Pejman *et al.*, 2009 [7]; Koklu *et al.*, 2010 [8]). However, because the majority of these research were conducted in Europe and Asia, there is a spatial-temporal variation that is heavily influenced by seasonality fluctuations.

Obisesan and Christopher (2018) [9] used statistical approaches such as principal component analysis and the General Linear Model to assess water pollution in the Asejire and Eleyele reservoirs, however the dataset was limited in scope (2003-2007). As a result, this analysis expanded on Obisesan and Christopher's (2018) work by extending the dataset from 2003 to 2019. This is to critically evaluate and statistically analyze changes in reservoir (Water) quality over time, as well as to statistically determine the influence of various statistical models on water quality metrics.

The two main reservoirs that provide water to the Ibadan Metropolis are Asejire and Eleyele, and they are the focus of this research. The measurements from these reservoirs were obtained from the Water Corporation of Oyo State in Ibadan, Nigeria, and are included in the data collection. The fact that these data are expansions of those from Obisesan and Christopher should be emphasized (2018). The 13 physicochemical parameters assessed monthly from January 2003 to August 2019 were turbidity, color, pH, dissolved oxygen, alkalinity, total hardness, calcium hardness, iron, silica, total solids, dissolved solids, and total suspended solids. These characteristics are necessary to determine the severity of the effects of water contamination. The aim of this study is to investigate water quality in Asejire and Eleyele Reservoirs using multivariate techniques.

3. Methodology

3.1. Principal Component Analysis

A large number of (potentially) linked variables are reduced to a smaller number of uncorrelated variables known as principal components through the mathematical process of principle component analysis (PCA). Each following component takes into account as much of the remaining variability as is practical whereas the first principal component takes into account as much of the data variability as is practicable.

The main goal of PCA is to transform a set of correlated qualities into a more manageable set of uncorrelated characteristics that account for the majority of the variation in the original attributes.

The sample data matrix of the n samples that were submitted to the n distinct characterization processes can be represented as matrix X.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{1}$$

The data in matrix X are mean-centered in order to generate the deviation matrix D. To accomplish this, the data mean is removed from each data point. Mean centering eliminates measurement bias.

$$D = \begin{bmatrix} x_{11} - \bar{X}_1 & \dots & x_{1n} - \bar{X}_n & \\ x_{21} - \bar{X}_2 & \dots & x_{2n} - \bar{X}_n & \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} - \bar{X}_m & \dots & x_{mn} - \bar{X}_n & \end{bmatrix} \tag{2}$$

The covariance matrix of the data set S, is constructed by,

$$S = \frac{D \cdot D^T}{n} \tag{3}$$

Resulting

$$S = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} \tag{4}$$

Where,

$$C_{ij} = 1/n\{(x_i - \bar{X}_i)(x_j - \bar{X}_j)\} \quad (i, j = 1, 2, ..., k) \tag{5}$$

The variance and covariance of the covariance matrix are real numbers. As a result, the covariance or variances cannot be compared when the variables in the covariance matrix are not measured in the same units. bigger values for the variables in the measurements will result in bigger variances, whereas lower values for the variables in the measurements will result in lower variances. By dividing each matrix member by its standard deviation, standardize the data to prevent the scale dependency of the covariance matrix.

Normalized matrix element C_{ij}

$$C_{ij} = \frac{c_{ij}}{\sqrt{var(i)var(j)}} \quad (i, j = 1, 2, ..., k) \tag{6}$$

Variance of *ith* element is given as (i). The maximum variation the *ith* and *jth* variable can have is (i) and $Var(j)$ respectively. Therefore, the correlation between *ith* and *jth* variable, C_{ij}, can never exceed $\sqrt{var(i)var(j)}$ resulting the maximum value a covariance matrix element to one. For two variables that are uncorrelated, the covariance is zero $(C_{ij} = C_{ji} = 0)$. Correlation matrix is symmetric due to the fact, $C_{ij} = C_{ji}$ and it is always real and positive definite.

PCA's main goal is to reduce the size of the data set while preserving as much variance as feasible from the original dataset. The covariance matrix describes the spread (variance) and orientation (covariance) of the data collection. As a result, a normally distributed K dimensional data set may be completely explained by a $K \times K$ covariance matrix along with the variable mean values.

The process of converting a square matrix into a diagonal matrix, which shares the same fundamental characteristics as the initial square matrix, is known as matrix diagonalization. A matrix is diagonalized when its original variables are replaced with a certain set of new variables, at which point the matrix assumes its canonical form. In other words, it's the same as figuring out a square matrix's eigenvalues. These eigenvalues will be the diagonal components of the resulting diagonal matrix. The new set of variables created by diagonalization, known as eigenvectors, correspond to the diagonal matrix.

We must diagonalize the correlation matrix, S, in order to describe it with directions and magnitudes (vectors).

$$S\overrightarrow{v_1} = \gamma_1 \overrightarrow{v_1} \quad (i = 1, 2, ..., k) \tag{7}$$

where γ_1 is the eigenvalue and $\overrightarrow{v_1}$ is the corresponding eigenvector of the correlation matrix, S.

$$S\overrightarrow{v_1} - \gamma_1 \overrightarrow{v_1} = 0 \tag{8}$$

$$(S - \gamma_1 I)\overrightarrow{v_1} = 0 \tag{9}$$

Where I is the identity matrix of the same dimensions as S. If v is not a null vector, then the equation above can only be define if $(S - \gamma_1 I)$ is not invertible. If a square matrix is not invertible then its determinant is zero.

$$Det(S - \gamma_1 I) = 0 \tag{10}$$

Solving the above equation gives a set of k eigenvalues and their corresponding orthogonal eigenvectors.

3.2. Factor Analysis

A statistical method known as factor analysis is used to translate variance among related, observable variables into a potentially more manageable set of unobservable variables known as factors. Alternatively, it is possible that changes in three or four of these observable variables, for instance, largely reflect changes in fewer of these unobservable variables. Such combined changes in response to unobservable hidden factors are what factor analysis searches for. The potential components are combined linearly to represent the observed variables using "error" terms. Later, the number of variables in a dataset can be reduced using the understanding of the relationships between observed variables that has been learned.

Suppose there exists a set of random variables, $x_1, ..., x_p$ and means $\mu_1, ..., \mu_p$. Suppose for some unknown constants I_{ij} and k unobserved random variables F_j where $i \in 1, ..., p$ and $j \in 1, ..., k$ where $k < p$, we have

$$x_i - \mu_i = L_{i1}F_1 + ... + L_{ik}F_k + \varepsilon_i$$

Here, the ε_i are independently distributed error terms with zero mean and finite variance, which may not be the same for all i. Let $Var(\varepsilon_i) = \psi$, so that we have,

$$Cov(\varepsilon) = Diag(\psi_1, ..., \psi_p) = \bar{\psi} \quad \text{and} \quad E(\varepsilon) = 0$$

In matrix terms, we have

$$x - \mu = LF + \varepsilon$$

Any solution of the above set of equations following the constraints for F is defined as the factors, and L as the loading matrix.

3.3. Generalized Linear Model

In a general linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + +\beta_q x_{qi} + \varepsilon \tag{11}$$

The response $y_i, i = 1, ..., n$ is modeled by a linear function of explanatory variables $x_j, j = 1, ..., q$ plus an error term.

Here general refers to the dependence on potentially more than one explanatory variable, versus the simple linear model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon \tag{12}$$

For the general linear model with $\varepsilon \sim N(0, \sigma^2)$, we have the linear predictor

$$\vartheta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_q x_{qi} \tag{13}$$

the link function

$$g(\mu_i) = \mu_i \tag{14}$$

and the variance function

$$var(\mu_i) = 1 \tag{15}$$

4. Results and Discussion

4.1. Exploratory Data Analysis

Table 1: *Descriptive Statistics for Asejire and Eleyele Reservoir*

	ASEJIRE RESERVOIR				**ELEYELE RESERVOIR**				**LIMITS**	
PARAMETERS	MIN	MAX	MEAN±SD	MED	MIN	MAX	MEAN±SD	MED	SON	WHO
Tur	0.0000	4.000	0.9278±1.2781	0.08	0.000	32.900	4.420±4.460	3.000	5	5
Col	4.00	7.00	5.01±0.2239	5.00	5.000	30.000	6.085±3.545	5.000	15	15
PH	6.400	8.800	7.476±0.3824	7.400	6.000	8.000	7.010±0.298	7.000	6.5-8.5	6.5-9.5
DO	3.500	15.700	7.638±1.4735	7.400	1.000	54.000	8.925±5.900	7.550	5	5
Alk	22.00	88.00	49.30±11.807	50.00	8.000	157.00	63.08±23.458	60.00	-	100
TH	36.00	104.00	60.83±11.169	58.00	64.00	148.00	94.65±12.947	94.00	150	100
CaH	9.00	80.00	41.89±10.899	41.00	30.00	108.00	62.88±13.769	63.50	-	75
Cl	10.40	67.00	23.51±8.4506	20.75	8.78	90.50	33.87±8.740	34.75	250	250
Fe	0.00	0.2800	0.08157±0.1125	0.00	0.00	3.20	2.27±0.471	2.30	0.3	0.3
Si	3.00	14.00	6.22±3.2534	4.00	4.00	17.00	7.51±4.199	5.00	-	-
Sol	40.0	1402.0	166.7±141.45	137.0	178.0	365.0	246.2±22.9526	245.0	-	-
DS	24.00	682.00	100.91±72.676	88.00	134.0	245.0	171.4±11.052	172.0	500	1000
SS	5.00	720.00	68.33±76.629	44.00	36.00	92.00	72.44±8.116	74.00	-	500

Asejire Reservoir: With high mean concentrations of 166.7 mg/L and 100.91 mg/L, respectively, Sol and DS are clearly the prominent parameters in Table 1. This demonstrates the shared origin of these variables. The typical PH value is 7:476 LU, which is little above neutral. Tur, Col, DO, Alk, TH, CaH, Cl, Fe, Si, and SS had average concentrations of 0.93, 5.01, 7.64, 49.30, 60.83, 41.89, 23.51, 0.08, 6.22, and 68.33 mg/L, respectively.

Eleyele Reservoir: According to Table 1, the dominating parameters in the Eleyele reservoir Sol are DS and TH, with mean concentrations of 246.2 mg/L, 171.4 mg/L, and 94.65 mg/L, respectively. This further demonstrates the human origin of these variables (Mustapha and Abdu 2012; Awoyemi et al. 2014). The average pH level is 7.010 LU, which is just above neutral. Tur, Col, DO, Alk, CaH, Cl, Fe, Si, and SS had average concentrations of 4.42, 6.09, 8.93, 63.08, 62.88, 33.87, 2.27, 77.51, and 72.44 mg/L, respectively.

Thus, in both the Asejire and Eleyele reservoirs, Total Solids and Dissolved Solids are the dominant parameters with high mean concentrations. Additionally, Asejire and Eleyele have dissolved oxygen concentrations that are 7.64 and 8.93 mg/L, respectively, above the allowable limit. Iron (Fe) levels in Eleyele Reservoir are also higher than allowed. This can be viewed in Figure 1 below.

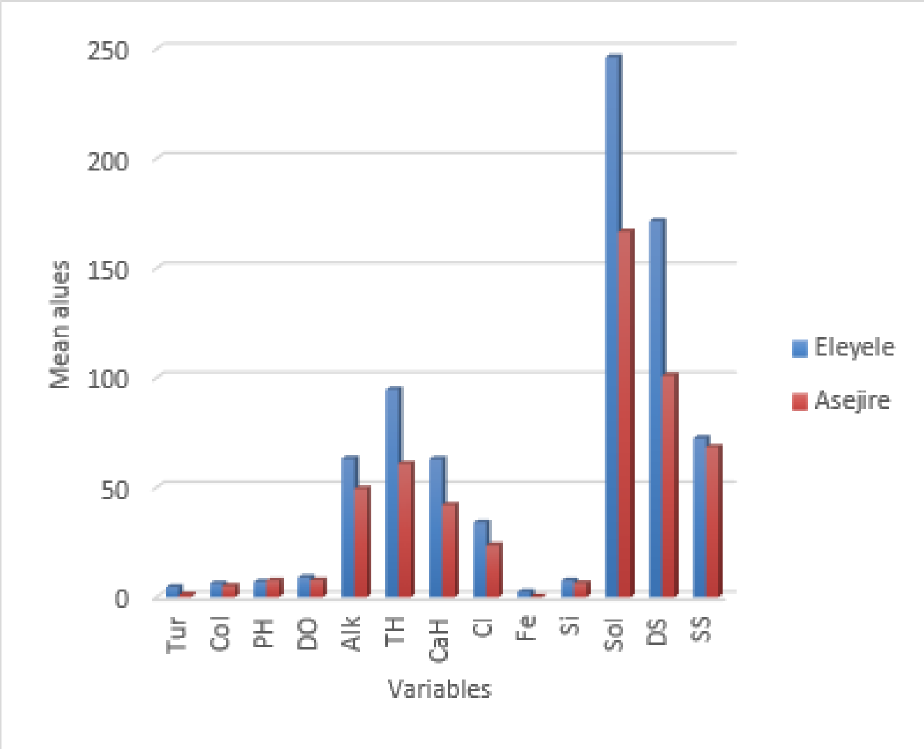

Figure 1: *Mean Plot Variable Values for Asejire and Eleyele Reservoirs*

4.2. Principal Component Analysis

Table 2: *Eigenvalues, Percentage Variance and Percentage Cumulative Variance for Both Reservoir PCs*

	ELEYELE RESERVOIR			ASEJIRE RESERVOIR		
PCs	Eigenvalue	%Variance	%Cum. variance	Eigenvalue	%Variance	%Cum. variance
PC1	2.2649278	17.422522	17.42252	3.47776690	26.7520530	26.75205
PC2	1.7912737	13.779028	31.20155	2.78797470	21.4459593	48.19801
PC3	1.4394865	11.072973	42.27452	1.52965004	11.7665388	59.96455
PC4	1.2299078	9.460829	51.73535	1.16678960	8.9753046	68.93986
PC5	1.2005036	9.234643	60.97000	0.99114982	7.6242294	76.56409
PC6	0.9619500	7.399615	68.36961	0.79671430	6.1285716	82.69266
PC7	0.9576967	7.366898	75.73651	0.77404385	5.9541834	88.64684
PC8	0.8029249	6.176345	81.91285	0.46483935	3.5756873	92.22253
PC9	0.6192486	4.763451	86.67630	0.31852158	2.4501660	94.67269
PC10	0.5870515	4.515781	91.19209	0.29688666	2.2837435	96.95644
PC11	0.5053452	3.887271	95.07936	0.19322174	1.4863211	98.44276
PC12	0.3389613	2.607395	97.68675	0.13872619	1.0671245	99.50988
PC13	0.3007223	2.313248	100.00000	0.06371526	0.4901174	100.00000

The water sample of dairy waste was subjected to a main component analysis, as indicated in Table 2. It contains loading for the component matrix that has been rotated, eigenvalues for each component, variance percentages, and cumulative variance percentages explained by each component. 13 physico-chemical parameters are taken into account during PCA, and the findings are summarized in a table. It shows that the first five principal components together account for 76.56% of the total variance in the dataset, where the first principal component accounts for 26.75%, the second for 21.44%, the third for 11.77%, the fourth for 8.98%, and the third for 7.62% of the total variance. To accurately assess the clustering behavior, PCA is used. Principal components (PC) are only extracted as components when Eigen values are greater than one. Factor loadings are used to represent PCs with Eigen values greater than the unit value. Factor loading is divided into three categories: strong, moderate, and mild. It ranges from 0.75 to 0.5. High factor loadings are present in the following principal components: PC1: Alk (0.7122), PC2: Sol (0.9026), PC3: DS (0.8397), PC4: SS (0.8313), and PC5: Col (0.7975). These factors have high factor loadings, indicating that they are the main polluters among other parameters and have high concentrations. PC3 and PC4 showed loadings ranging from strong to moderate.

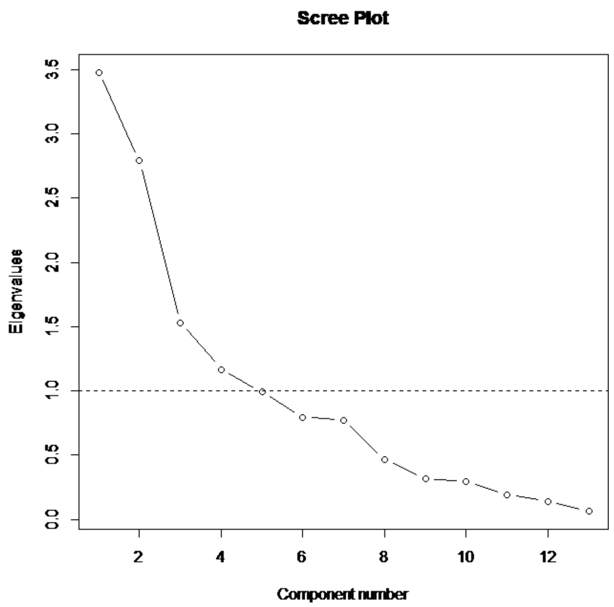

Figure 2: *A Scree Plot of the Principal Components (Eleyele Reservoir)*

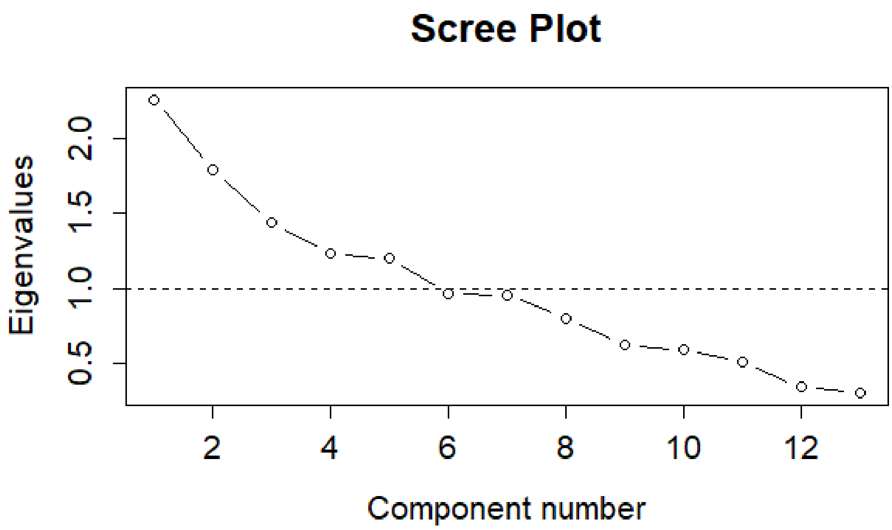

Figure 3: *A Scree Plot of the Principal Components (Asejire Reservoir)*

Table 3: *Factor Analysis for the Asejire Reservoir*

Parameter	PA1	PA2	PA3	PA4	PA5	h2	u2
Tur	-0.88	-0.04	-0.11	0.07	-0.05	0.790	0.21
Col	-0.06	0.00	-0.08	0.14	0.01	0.030	0.97
PH	-0.01	-0.08	0.09	0.65	-0.03	0.432	0.57
DO	0.05	-0.08	0.11	0.28	-0.04	0.099	0.90
Alk	0.55	0.01	0.31	0.29	0.23	0.528	0.47
TH	0.13	-0.05	0.84	-0.02	0.10	0.731	0.27
CaH	0.12	-0.09	0.78	0.17	0.13	0.681	0.32
Cl	0.21	0.09	0.19	-0.06	0.61	0.467	0.53
Fe	-0.93	-0.07	-0.08	0.09	-0.05	0.890	0.11
Si	-0.82	-0.03	-0.06	0.03	-0.15	0.700	0.30
Sol	0.10	1.01	-0.07	-0.11	-0.02	1.050	-0.05
DS	-0.13	0.85	-0.02	-0.03	0.02	0.739	0.26
SS	0.21	0.84	-0.08	-0.15	0.14	0.800	0.20
Eigenvalue	7.7661e-02	3.284e-02	8.9677e-03	5.7938e-03	4.265e-03		
% of variance	56.90	24.06	6.57	4.24	3.12		
Cum. % of variance	56.90	80.96	87.53	91.78	94.90		

4.3. Factor Analysis

The Factor Analysis (FA) helped to identify and extract the variables affecting water quality. Table 12 shows that FA identified latent components that accounted for 94.90% of the variation. Alkaline had a positive loading of 0.55 and Turbidity, Iron, and Silicon all exhibited very substantial negative loadings of -0.88, -0.93, and -0.82, respectively. PA1 explained 56.90% of the total variation. This could be considered anthropogenic input. PA2 demonstrated high positive loadings on suspended particles (0.84) and dissolved solids (0.85) and explained 24.06 percent of the overall variation. With substantial positive loadings on Total Hardness (0.84) and Calcium Hardness (0.78), PA3 explained 6.57 percent of the total variance. With mildly positive loadings on PH (0.65), PA4 explained 4.24 percent of the overall variance. With moderately positive loadings on chlorine (0.61), PA5 explained 3.12% of the overall variation.

Testing the idea that five criteria are adequate. The null model has 78 degrees of freedom and an objective function of 7.14 with a 1383.64 Chi Square. The objective function was 0.16, and the model has 23 degrees of freedom. Root mean square residuals (RMSR) are equal to 0.01. Root mean square of the residuals corrected for degree of freedom is 0.03.

The Factor Analysis (FA) helped to identify and extract the variables affecting water quality. FA identified latent components in table 13 above that accounted for 79.97% of the overall variance. With positive loadings on Turbidity (0.55) and Silicon (0.92), PA1 accounted for 33.60% of the total variance, while PA2 explained 18.89%, PA3 explained 12.95%, PA4 explained 8.28% of the total variance with positive loadings on Total Solids (0.59) and Dissolved Solids (0.53), and PA5 accounted for 6.25% of the total variance.

The five factors are adequate theory is put to the test. The null model has 78 degrees of freedom, an objective function of 2.12, and a Chi Square of 411.88. The model has 23 degrees of freedom, and its objective function was 0.26. The residuals' root mean square root (RMSR) is 0.04. The residuals' df-corrected root mean square is 0.07, their harmonic number is 200, and their empirical chi square is 39.9 with a probability of less than 0.016.

Table 4: *Factor Analysis for the Eleyele Reservoir*

Parameter	PA1	PA2	PA3	PA4	PA5	h2	u2
Tur	0.55	-0.23	-0.02	0.27	0.06	0.431	0.57
Col	0.34	-0.01	-0.04	0.10	-0.04	0.129	0.87
PH	0.43	0.15	0.04	-0.05	-0.09	0.223	0.78
DO	0.32	0.17	0.48	0.02	0.03	0.361	0.64
Alk	-0.07	1.01	0.00	-0.01	0.12	1.040	-0.04
TH	-0.05	0.41	-0.57	0.08	0.00	0.498	0.50
CaH	0.12	0.04	-0.65	0.06	0.05	0.448	0.55
Cl	0.15	-0.15	-0.28	0.23	-0.04	0.174	0.83
Fe	-0.03	-0.05	0.05	0.04	-0.28	0.088	0.91
Si	0.92	-0.17	-0.03	-0.05	-0.02	0.879	0.12
Sol	0.04	0.00	-0.09	0.59	0.13	0.378	0.62
DS	0.10	0.07	-0.02	0.53	-0.42	0.469	0.53
SS	-0.17	0.00	0.09	0.14	0.48	0.282	0.72
Eigenvalue	1.211e-02	6.809e-03	4.667e-03	2.986e-03	2.253e-03		
% of variance	33.60	18.89	12.95	8.28	6.25		
Cum. % of variance	33.60	52.49	65.44	73.72	79.97		

4.4. Generalized Linear Model

The Generalized Linear Model (GLM) estimates of the relationship between Turbidity and other water variables are presented in Table 5 below. It is clear that Silicon (Si) and Iron (Fe) contribute greatly to the water's level of turbidity in Asejire. We could also find that the water variables taken into account in this investigation could explain 72.73% of the turbidity.

Table 5: *GLM estimates and SD for Asejire Reservoir*

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-1.1636528	1.4200685	-0.819	0.41358
Col	0.2974706	0.2144660	1.387	0.16708
PH	-0.0559984	0.1340955	-0.418	0.67672
DO	0.0405164	0.0334850	1.210	0.22781
Alk	-0.0042912	0.0053421	-0.803	0.42284
TH	0.0029505	0.0060358	0.489	0.62553
CaH	-0.0075631	0.0064229	-1.178	0.24048
Cl	-0.0008144	0.0063186	-0.129	0.89759
Fe	7.8805683	0.6996219	11.264	< 2e-16 ***
Si	0.0694667	0.0243463	2.853	0.00481 **
Sol	-0.0004600	0.0010578	-0.435	0.66419
DS	0.0000971	0.0013802	0.070	0.94398
SS	0.0010290	0.0014926	0.689	0.49145
(Dispersion parameter for Gaussian family taken to be 0.4454064)				
Null deviance: 325.061 on 199 degrees of freedom				
Residual deviance: 83.291 on 187 degrees of freedom				
AIC: 420.38				
Multiple R-squared: 0.7438, Adjusted R-squared: 0.7273				

Only Iron and Silicon contributes greatly to turbidity in the water level. The fitted regression equation for turbidity level is given by

$$Tur = -1.1636528 + 7.8805683 * Fe + 0.0694667 * Si$$

Each one's impact on turbidity is explained by other factors. Turbidity will decrease at

1.1636528 with zero contribution from all independent factors. When all other factors are held equal, a rise in turbidity of 7.881 units is projected for every unit increase in iron (Fe). Similarly, assuming all variables remain constant, it is expected that for every unit increase in silicon (Si), turbidity level will increase by 0.069 units.

The Generalized Least Squares (GLS) estimates of the regression between Turbidity and other water variables are presented in Table 6 below. It is clear that the turbidity level of the Eleyele water level is substantially influenced by color, alkalinity, silicon (Si), and solids. We could also find that the water variables taken into account in this investigation might explain 37.56% of the turbidity.

Table 6: *GLM estimates and SD for Asejire Reservoir*

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-6.873238	8.399258	-0.818	0.4142
Col	0.176279	0.078312	2.251	0.0256 *
PH	-1.239459	0.999494	-1.240	0.2165
DO	0.093316	0.051428	1.814	0.0712 .
Alk	-0.034439	0.014186	-2.428	0.0161 *
TH	0.032033	0.025496	1.256	0.2105
CaH	-0.007554	0.021783	-0.347	0.7291
Cl	0.048267	0.032100	1.504	0.1344
Fe	0.449737	0.583198	0.771	0.4416
Si	0.480872	0.080525	5.972	1.16e-08 ***
Sol	0.026909	0.012235	2.199	0.0291 *
DS	0.013312	0.025536	0.521	0.6028
SS	0.034759	0.033860	1.027	0.3060
(Dispersion parameter for Gaussian family taken to be 13.21585)				
Null deviance: 3957.9 on 199 degrees of freedom				
Residual deviance: 2471.4 on 187 degrees of freedom				
AIC: 1098.4				
Multiple R-squared: 0.3756, Adjusted R-squared: 0.3355				

Colour, Alkaline, Silica and Solids contributes greatly to turbidity in the water level. The fitted regression equation for turbidity level is given by

$$Tur = -6.873238 + 0.176279 * Col - 0.034439 * Alk + 0.480872 * Si + 0.026909 * Sol$$

Each one's impact on turbidity is explained by other factors. Turbidity will be decreasing at 6.87323 if none of the independent variables contribute. Keeping all variables equal, it is expected that for every unit increase in Color (Col), there will be a 0.1763 unit rise in Turbidity level. Similar to this, it is predicted that for every unit increase in silicon (Si), there will be a 0.4809 unit increase in turbidity level, for every unit increase in alkaline (Alk), there will be a 0.4809 unit decrease in turbidity level, and for every unit increase in solids (Sol), there will be a 0.0269 unit increase in turbidity level, all other variables being held constant.

5. Conclusion

In this study, a variety of multivariate exploratory techniques were used to assess changes in the quality of the surface water in the reservoirs of Eleyele and Asejire. This study's primary goal is to determine the levels of turbidity and water pollution in the Asejire and Eleyele Reservoirs utilizing multivariate and generalized linear model techniques. According to the descriptive statistics, all of the parameters with the exception of dissolved oxygen from both reservoirs (Eleyele and Asejire) and iron from Eleyele Reservoir were within the acceptable ranges.

Since the goal of cluster analysis is to make heterogeneous groups homogeneous, the cluster means were divided into five groups. This revealed that all of the groups were largely homogenous, with the exception of Total Solids, Dissolved Solids, and Suspended Solids, whose cluster means are clearly heterogeneous. The groups were similarly homogeneous for Asejire Reservoir and for Eleyele Reservoir, but Alkaline and Turbidity's cluster means

The PC findings from the PCA were those with eigenvalues greater than 1, according to established measures. Results indicate that 5 of the 13 PCs chosen to effectively explain the variance in the data were taken into consideration. The PCs accounted for 76.56% of the total variance in the Asejire Reservoir and 60.97% of the total variance in the Eleyele Reservoir. According to the PCA results, Tur, Si, TH, CaH, and Alk were the main pollutants among other parameters in Asejire Reservoir, while Tur, Si, TH, and CaH were the major pollutants among other parameters in Eleyele Reservoir. This indicates that these parameters are high in concentration and are the major pollutants among other parameters in Asejire Reservoir.

To further simplify the data structure produced by the PCA, factor analysis was used to remove the contribution of less significant variables. It discovered latent components that, for Asejire and Eleyele Reservoirs, respectively, accounted for 94.90% and 79.97% of all variance. The findings make it abundantly evident that the main pollutants in Asejire Reservoir were Alk, DS, SS, TH, CaH, PH, and Cl, whereas the main pollutants in Eleyele Reservoir were Tur, Si, Sol, and DS.

According to the GLM results, only Iron (Fe) and Silicon (Si) have a significant impact on water turbidity; other physico-chemical factors have a less significant role in describing turbidity variance in the Asejire Reservoir. Additionally, turbidity in the water level is greatly influenced by color (Col), alkalinity (Alk), silica (Si), and total solids (Sol); other physico-chemical parameters are less significant in explaining turbidity variation in Eleyele Reservoir.

Intensive farming practices, home wastes, animal waste, organic wastes, inorganic wastes, and industrial areas close to the river channel all contribute to the general pollution of both reservoirs, which is caused by anthropogenic influence and industrial activity.

In a nutshell, this study shows the value of multivariate statistical techniques for complex data analysis and visualization in water quality assessment, pollution source/factor identification, and realization of temporal/spatial variations in water quality for effective river water quality management.

Through a careful examination of the environment, efforts should be made to reduce anthropogenic influence in the reservoir (Obisesan and Oladoja [10]). Human and industrial activity in the city must be carefully controlled, and human activity near the reservoir's route must be limited. Continual public education regarding the effects of water pollution and the implementation of environmental/water management laws is also necessary.

Acknowledgements

Acknowledgements goes to Oyo State Water Corporation for providing pollution data for this study.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Oketola A., Adekolurejo S., and Osibanjo, O. (2010). Water Quality Assessment of River Ogun Using Multivariate Statistical Techniques. *Journal of Environmental Protection*, 4 (5): 466-479.

[2] UNEP (United Nations Environment Programme) (2002). Convention on Biological Diversity. United Nations. *http://www.Biodiv.Org/Convention/Articles.Asp*

[3] Nancy, D. (2009). Water Quality: Frequently Asked Questions. Florida Brooks National Marine Sanctuary, Key West, FL.

[4] WHO, (2018). Global Nutrition Report. *https://globalnutritionreport.org/0b98fb*

[5] Boyacioglu, H. (2008). Water Pollution Sources Assessment by Multivariate Statistical Methods in the Tahtali Basin, Turkey. *Environ Geol.*, 54, 275-282.

[6] Praus, P. (2007). Urban Water Quality Evaluation Using Multivariate Analysis. *Acta Montanistica Slovaca Ro?n?k*, 12(2): 150-158.

[7] Pejman A. H., Bidhendi G. R. N., Karbassi A. R., Mehrdadi N. and Bidhendi M. E. (2009). Evaluation of Spatial and Seasonal Variations in Surface Water Quality Using Multivariate Statistical Techniques. *International Journal of Environmental Science and Technology*, 6(3): 467-476.

[8] Koklu R., Sengorur B. and Topal B. (2010). Water Quality Assessment Using Multivariate Statistical Methods — A Case Study: Melen River System (Turkey). *Water Resource Management*, 24: 959-978.

[9] Obisesan K. O. and Christopher P. (2018). Statistical Models for Evaluating Water Pollution: The Case of Asejire and Eleyele Reservoirs in Nigeria. *Journal of Environmental Statistics* Volume 8, Issue 5.

[10] Obisesan, K.O. and Oladoja, O.M. (2022). On Normal Process of Diffusion Equation in Monitoring Carbon Monoxide Concentrations in Nigeria. *International Journal of Statistical Distributions and Applications*. 8(2)