MULTIVARIATE ANALYSIS FOR ASSESSEMENT OF WATER QUALITY PARAMETERS: A CASE STUDY LAMINGO DAM JOS, PLATEAU STATE, NIGERIA

ABSTRACT

The water quality in Jos, Plateau State, Nigeria, has become a critical issue with significant public health, ecological, and economic implications. Despite ongoing efforts to manage water resources, substantial challenges remain in ensuring safe and clean water provision. This study aims to improve water quality by analyzing the various water quality parameters using Principal Component Analysis (PCA) and Factor Analysis (FA). Secondary data from the Plateau State Water Board's Laminga Treatment Plant for the year 2023 was used. The correlation analysis revealed significant interdependencies among parameters such as pH, Alkalinity, and Chlorine, highlighting the need for integrated management. PCA identified two principal components explaining 72.44% of the variance in water quality, with the primary component driven by pH, Alkalinity, and Chlorine, and the secondary component influenced by Total Hardness and Turbidity. FA uncovered two underlying factors: the "Water Chemistry Factor" driven by “Human Activities” and the "Water Mineral Factor" influenced by “Geological Processes”. The study concludes with recommendations for targeted monitoring, optimized chlorine dosing, advanced filtration systems, and holistic water quality improvement strategies. Additionally, infrastructure improvements, public awareness campaigns, and strict regulation of industrial effluents and agricultural runoff are essential. These findings provide a comprehensive understanding of the factors affecting water quality and offer actionable insights for enhancing water quality and management practices in Jos, Plateau State.

Keyboards: Water, Water Quality, Principal Component Analysis, Factor Analysis, Eigenvalue, Eigenvector

INTRODUCTION

Multivariate analysis techniques have gained widespread acceptance in various organizations enabling them to create knowledge and thereby improve their decision making. Multivariate analysis encompasses all statistical techniques that simultaneously analyzes multiple measurements on individual or object under investigation (Hair et al., 2019). It serves several purposes, including data reduction and simplification, categorization, and clustering, examining relationships, and interdependencies among variables, predictive modelling, and hypothesis testing (Johnson & Wichern, 2007).

Water is essential to all life. Without water, life as we know it could not exist, and civilizations would not be able to develop or thrive. Water is a major component of most living matter (Ayoade, 1988). Water is not only one of the most essential commodities of our day-to-day life, but the development of this natural resources also plays a crucial role in economic and social development processes. Therefore, the need to access a reliable, secure, safe, and sufficient source of fresh water is a fundamental requirement for the survival, well-being, and the socio-economic development of all humanity (Tebbutt, 1990). The term "water quality" is defined as “the physical, chemical, and biological characteristics of water” (Spellman, 2013; Alley, 2007). Assessment of water is not only for suitability for human consumption but also in relation to its agricultural, industrial, recreational, and commercial uses and its ability to sustain aquatic life. Water quality assessment is the overall process of evaluation of the physical, chemical, and biological nature of water in relation to natural quality, human effects, and intended uses (Chapman, 1992). Assessment of water quality is important for pollution control and the protection of surface and groundwaters (Gray, 2017).

In literature, Aliyu et al. (2022) investigated water quality parameters from samples taken from River Galma in Zaria, Northwestern Nigeria, during both wet and dry seasons. They analyzed forty-five samples per season using Microsoft Excel and SPSS Statistics 23. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO-MSA) was employed to assess data adequacy. Using PCA with varimax rotation, eighteen initial water quality parameters were reduced to nine, revealing that two principal components accounted for more than 79% of the total variance in both seasons. The study highlighted anthropogenic and natural sources as significant contributors to pollution and recommended measures to mitigate future surface water pollution risks.

Another study by Thomas (2023) analyzed water quality datasets from municipal deep wells in the Ibadan Metropolitan area, Nigeria. Eleven parameters with a sample size of sixty were normalized using log-normal distribution. Kaiser-Meyer-Olkin (KMO) and Bartlett's tests confirmed the suitability of data for factor analysis, which resulted in four factors explaining 98.35% of the total variance. Key findings showed significant correlations between anthropogenic factors and water quality parameters, with pollutants sourced from agricultural land use and natural processes.

Furthermore, the study of Tijjani et al. (2023) focused on groundwater quality around dumpsites in Bauchi Metropolis, Nigeria. Thirty samples from ten boreholes were analyzed using standard methods and Pearson's correlation coefficient for PCA analysis. Three components explained 100% of the cumulative total variance. The first component (46.984% variance) correlated with salinity, TSS (Total Suspended Solids), Sulfate (SO_4^{2-}), Nitrate (NO_3^-), and Nitrite (NO_2^-), likely due to leachates from dumpsites. The second component (27.181% variance) correlated with TDS (Total Dissolved Solids) and Bicarbonate (HCO_3^-), attributed to automobile emissions and machinery. The third component (25.835% variance) correlated with pH, temperature, and conductivity, indicating lithogenic origins. The study recommended consistent monitoring to detect early changes in water quality parameters.

The current water quality monitoring practices in Jos, Plateau State often focus on analyzing individual parameters without considering the interrelationships and combined effects of multiple contaminants. This approach may overlook underlying factors that significantly influence water quality and fail to identify the primary sources of pollution, thus hindering the development of effective strategies for water management and pollution control. The complexity of water quality issues necessitates a comprehensive assessment that goes beyond traditional methods. There is an urgent need for a robust analytical framework that can integrate and interpret complex water quality data to provide a holistic understanding of the factors affecting water quality. Thus, studying the correlations and interactions between different water quality parameters to understand their interdependencies, identifying the principal components that explain the variance in water quality parameters, determining the underlying factors or latent variables that influence water quality, and providing recommendations for improving water quality in Jos, Plateau State, Nigeria, which will serve as the objectives to achieve the aim of the study which is to improve water quality in the region.

MATERIALS AND METHODS

Principal Component Analysis (PCA) and Factor Analysis (FA) are the multivariate analysis techniques employed in this study. Principal Component Analysis (PCA) is a reduction dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset (Jolliffe, 2002), while Factor Analysis (FA) refers to a class of multivariate statistical analysis whose main purpose is to define the underlying structure in a data matrix. It analyzes the structure of correlations between a large dataset by defining a set of common latent dimensions called factors (Hair et al., 2019). In this study, the analysis was performed using the R software package.

Standardization

To facilitate statistical analysis and comparison of the water quality data, the water quality parameters were standardized by subtracting the mean and dividing by the standard deviation for each parameter. This transformation resulted in a mean of zero and a deviation of one for each parameter, removing the problem of scale dependence (Jolliffe, 2002). This will be done using the equation

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \tag{1}$$

where Z_{ij} is the standardized value, X_{ij} is the original value, μ_j is the mean, and σ_j is the standard deviation of the j-th variable.

Kaiser-Meyer-Olkin and Bartlett's Test of Sphericity

The adequacy of the data is evaluated on the basis of the results a Kaiser-Mayer-Olkin (KMO) sampling adequacy test and Bartlett's test of Sphericity. KMO is used to evaluate whether the relationship between variables is truly reflexive of an underlying process. Bartlett's test of Sphericity was employed to test that variables in the matrix are uncorrelated, an undesirable result. Overall, KMO value uses above 0.5 and $p < 0.05$ for Bartlett's sphericity test are considered acceptable (Hair et al., 2019).

Correlation Matrix

A correlation matrix was constructed to examine the relationships between the water quality parameters. The matrix revealed the strength and direction of linear associations between each pair of the parameters, with correlation coefficients (r_{ij}) ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). Coefficients between 0.7 and 0.9 (or -0.7 to -0.9) indicate a strong positive (or negative) linear relationship, while values from 0.4 to 0.6 (or -0.4 to -0.6) suggest a moderate correlation. Coefficients between 0.1 and 0.3 (or -0.1 to -0.3) denote a weak correlation, and a coefficient of 0 indicates no linear relationship.

The sample correlation which is a measure of the strength and direction of the linear relationship between the i^{th} and j^{th} variables is given by:

$$r_{i,j} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{j=1}^{n}(x_j - \bar{x})^2}} \tag{2}$$

The sample correlation matrix which shows the interrelationships between each pair of variables in a dataset is given by:

$$R = r_{i,j} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \ldots & 1 \end{bmatrix} \tag{3}$$

Eigenvalue Decomposition (EVD)

The derivation and properties of principal components (PCs) and factors are based on the concept of eigenvectors and eigenvalues of the covariance matrix or correlation matrix. PCs can be found using purely mathematical arguments, they are given by an orthogonal linear transformation of a set of variables optimizing a certain algebraic criterion that is PCs are defined by transformation (Jolliffe 2002).

The eigenvalues measure the amount of variation in the total sample accounted for by each of the component or factor. The eigen values of the correlation matrix R is obtain using the expression

$$|R - \lambda I| = 0 \tag{4}$$

where λ_i are the eigen values

The eigenvector is a set weights that define the relative importance of each variable in a principal component or factor. The eigenvector for each eigenvalue is obtain using the expression

$$(R - \lambda I)X = 0 \tag{5}$$

where the X_s' are the eigenvectors.

Extraction of Principal Components

Principal Components are particular (algebraically) i.e. linear combinations of the p random variables $X_1, X_2, \dots, X_p$. It depends solely on the covariance matrix or correlation matrix of $X_1, X_2, \dots, X_p$.

Consider the linear combinations

$$PC_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = a_1'X$$

$$PC_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = a_2'X$$

$$\vdots$$

$$PC_p = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p = a_p'X$$

$$PC_i = \sum_{j=1}^{p} a_{ij}x_j \tag{6}$$

Where $a_1', a_2', \dots, a_p'$ are row vector and X is column vector, such that

$$a_1' = (a_{11}, a_{12}, \dots, a_{1p}),\ a_2' = (a_{21}, a_{22}, \dots, a_{2p}), \dots, a_p' = (a_{p1}, a_{p2}, \dots a_{pp}).$$

Principal components are those uncorrelated linear combinations P (i.e. P is column vector) whose variances are as large as possible. Each component is a linear combination of the p variables. The first components account or makes up for the largest possible amount of variance while the second largest amount of variance is accounted or made up for by the second components, formed from the variance remaining after that related with the first components has been removed, etc. The restriction (assumption upon which) principal components are isolated or taken out is that they are orthogonal. They may be geometrically viewed as being dimensions in p-dimensional space where each dimension is perpendicular to the other dimension.

Variance Explained

Proportion of variance explained by the i^{th} principal component is

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \tag{7}$$

where, λ_i is the i^{th} eigenvalue and p is the total number of components.

Cumulative Proportion

Cumulative proportion of variance explained by the first k^{th} principal components is

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_i} \tag{8}$$

Component Loadings

The component loadings in Principal Component Analysis (PCA) quantify the correlation between the original variables and the principal components is given as:

$$a_{ij} = \sqrt{\lambda_i} \, . \, v_{ij} \tag{9}$$

Where:

a_{ij} is the loadings of the jth original variable on the ith principal component.

λ_i is the eigenvalue associated with the ith principal component.

v_{ij} is the element of the eigenvector corresponding to ith principal component

and the jth variable.

Stopping Rule

To determine the optimal number of components and factors to extract, this study employed the cumulative variance explained, Kaiser's stopping rule, and scree test as stopping rules criteria. Jolliffe (2002) states that a sensible cutoff point is when 70-90% of the variance is accounted for by the components. The Kaiser's stopping rule (Kaiser, 1960) simply means that all eigenvectors with an eigenvalue of 1.0 or greater are extracted from the data and retained as part of the solution. The value of 1.0 is not arbitrary in this context; 1.0 is the variance of a standardized variable. The scree test (Cattel. 1966) is dependent on a scree plot, or graphical display of the eigenvalues for successive eigenvectors.

Factor Analysis

The factor analysis model expresses each variable as a linear combination of few underlying common factors $f_1, f_2, \ldots, f_m$ with an accompanying error term account for that part of the variable that is unique (Rencher & Christensen, 2012). Thus, for $y_1, y_2, \ldots, y_p$ in any observation vector y, the model in matrix notation is given as:

$$y = \mu + Lf + \varepsilon\,, \tag{10}$$

Where:

$y = (y_1, y_2, \ldots, y_p)' =$ vector of observations

$\mu = (\mu_1, \mu_2, \ldots, \mu_p)' =$ population mean vector

$f = (f_1, f_2, \ldots, f_m)' =$ vector of common factors

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)' =$ vector of specific factors

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{pmatrix} = \text{matrix of factor loadings}$$

The assumptions can be expressed concisely using vector and matrix notation:

1. $E(f_i) = 0,\ i = 1,2, \ldots, m$, becomes

$$E(f) = 0, \tag{11}$$

2. $\mathrm{Var}(f_i) = 1,\ j = 1,2,\dots,m$ and $\mathrm{cov}(f_i, f_j) = 0,\ j \neq k$, become

$$\mathrm{cov}(f) = I, \tag{12}$$

3. $E(\varepsilon_i) = 0,\ i = 1,2,\dots,p$, becomes

$$E(\varepsilon) = 0, \tag{13}$$

4. $\mathrm{Var}(\varepsilon_i) = \psi_i, i = 1,2,\dots,p$ and $\mathrm{cov}(\varepsilon_i, \varepsilon_k) = 0, i \neq k$, therefore (14)

$$\mathrm{cov}(\varepsilon) = \Psi = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix}. \tag{15}$$

5. $\mathrm{cov}(\varepsilon_i, f_j) = 0$ for all i and j becomes

$$\mathrm{cov}(f, \varepsilon) = O. \tag{16}$$

The notation $\mathrm{cov}(f, \varepsilon) = O$ indicates a rectangular matrix containing the covariances of the f's with the ε's:

$$\mathrm{cov}(f, \varepsilon) = \begin{pmatrix} \sigma f_1 \varepsilon_2 & \sigma f_1 \varepsilon_2 & \dots & \sigma f_1 \varepsilon_p \\ \sigma f_2 \varepsilon_1 & \sigma f_2 \varepsilon_2 & \dots & \sigma f_2 \varepsilon_p \\ \vdots & \vdots & \vdots & \vdots \\ \sigma f_m \varepsilon_1 & \sigma f_m \varepsilon_2 & \dots & \sigma f_m \varepsilon_p \end{pmatrix} \tag{17}$$

since μ does not affect variances and covariances of y, we have, from equation (10),

$$\Sigma = \mathrm{cov}(y) = \mathrm{cov}(Lf + \varepsilon). \tag{18}$$

By equation (18), Lf and ε are uncorrelated; therefore, the covariance matric of their sum is the sum of their covariance matrices:

$$\begin{aligned} \Sigma &= \mathrm{cov}(Lf) + \mathrm{cov}(\varepsilon) \\ &= L\ \mathrm{cov}(f)\ L' + \Psi \\ &= L\ I\ L' + \Psi \\ &= L\,L' + \Psi. \end{aligned} \tag{19}$$

In general, loadings themselves represent covariances of the variables with the factors,

$$\mathrm{cov}(y_i, f_j) = l_{ij}\ ; \quad i = 1,2,\dots,p \ \ j = 1,2,\dots,m \tag{20}$$

Since l_{ij} is the $(ij)^{th}$ element of L, (3.7.11) can be written in the form

$$\mathrm{cov}(y, f) = L. \tag{21}$$

If standardized variables are used, equation (18) is replaced by $P_\rho = L\,L' + \Psi$, and the loadings become correlations:

$$\mathrm{corr}(y_i, f_j) = l_{ij} \tag{22}$$

In partitioning variance of y_i into a component due to the common factor, called the communality, and a component unique to y_i, called the specific variance:

$$\begin{aligned}\sigma_i^2 = \mathrm{Var}(y_i) &= (\, l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2) + \psi_i \\ &= h_i^2 + \psi_i \\ &= \text{communality} + \text{specific variance}\end{aligned} \tag{23}$$

The Communality h_i^2 is also called common variance and the specific variance ψ_i has been called specificity, unique variance, or residual variance.

Factor Extraction

The extraction method used in this study was the principal axis factoring method (PAF). PAF is generally used when the research purpose is to identify latent variables, which contribute to the common variance of the set of measured variables, excluding variable-specific variance. The principal axis factoring method (also called the principal axis or principal factor method) uses an initial estimate $\widehat{\Psi}$ and factors $R - \widehat{\Psi}$ to obtain

$$R - \widehat{\Psi} \cong \hat{L}\,\hat{L}' \tag{24}$$

Where $\hat{L}$ is $p \times m$ which is obtained using eigen values and eigenvectors of $R - \widehat{\Psi}$.

The i^{th} diagonal element of $R - \widehat{\Psi}$ is given by $1 - \widehat{\psi}_i$, which is the i^{th} communality,

$\hat{h}_i^2 = 1 - \widehat{\psi}_i$. With these diagonal values, $R - \widehat{\Psi}$ have the form,

$$R - \widehat{\Psi} = \begin{pmatrix} \hat{h}_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & \hat{h}_p^2 \end{pmatrix} \tag{25}$$

After obtaining communality estimates, the eigenvalues and eigenvectors of $R - \widehat{\Psi}$ is use to obtain estimates of factor loadings, $\hat{L}$. Then the columns and rows of $\hat{L}$ can be used to obtain new eigenvalues (variance explained) and communalities, respectively. The sum of squares of the j^{th} column of $\hat{L}$ is the eigenvalue of $R - \widehat{\Psi}$, and the sum of squares of the i^{th} row of $\hat{L}$ is the communality of y_i. The proportion of variance explained by the j^{th} factor is

$$\frac{\theta_j}{\mathrm{tr}(R - \widehat{\Psi})} = \frac{\theta_j}{\sum_{i=1}^{p} \theta_i} \tag{26}$$

Where θ_j is the j^{th} eigenvalue of $R - \widehat{\Psi}$. The matrices $R - \widehat{\Psi}$ are not necessarily positive semidefinite and will often have some small negative eigenvalues. In such case, the cumulative proportion of variance will exceed 1 and then decline to 1 as the negative eigenvalues are added.

Factor Rotation

Rotation is a mathematical manipulation meant to minimize the factor loadings close to 0 and maximize the loadings that are close to 1.0 for the purpose of simplifying interpretability of factors without changing the solution (Brown, 2006). This study used the orthogonal varimax rotation which was applied to the subset of factors extracted, aiming to estimate uncorrelated factors with a simpler loading which are considered easier to interpret. Thus, the varimax method attempts to make the loadings either larger or small to facilitate interpretation. The output typically includes the rotated loading matrix.

Interpreting the Components and Factors

In order to ensure that the interpretation of components and factors loadings is meaningful, this study took into account practical significance as the criteria to assess the loadings. Factors loadings less than $\pm.10$ are considered to be zero for the purposes of assessing simple structure, $\pm.30$ is considered to meet the minimal level for interpretation of structure, $\pm.50$ or greater are considered practically significant and loadings exceeding $\pm.70$ are considered indicative of well-defined structure Hair et al. (2019). A simple structure is achieved when each variable(parameter) has a significant loading on one factor only and each factor has significant loadings for only a subset of items. The variables (water quality parameters) considered are the ones with the highest loadings on any factor. In addition, labelling of factors is based on a subjective, theoretical and conceptual intent, (Hair et al., 2019) states that researcher should assign some meaning to the pattern of the acceptable loading, variable with higher loadings are considered more significant and have greater influence on the name or label selected to represent a factor. Thus, this study attempts to assign some meaning to the patterns of factor loadings based on its appropriateness for representing the underlying dimensions of a particular factor affecting water quality parameters.

Data Description

The data used for this study were secondary data (16th January, 2023-31th December, 2023) extracted from the records of experiments conducted on raw water by the Plateau State Water Board, Laminga Treatment Plant. The data exclusively focuses on water quality parameters of Lamingo Dam Jos, Plateau State, Nigeria. The research aims to analyze and interpret existing data to assess water quality in Plateau State, particularly in relation to the parameters (pH, Total Hardness, Alkalinity, Turbidity, and Chlorine) measured in the experiments conducted by the water board.

RESULTS AND DISCUSION

Table 1: Summary Statistics of Parameters

Vars	Parameter	Unit	N	mean	standard deviation	Median	minimum	maximum
1	pH	-	188	6.84	0.3	6.9	6.2	7.4
2	Total Hardness	mg/L	188	25.13	7.83	24	8	57
3	Alkalinity	mg/L	188	30.48	7.84	30	19	46
4	Turbidity	NTU	188	4.98	3.02	4.04	1.32	20.5
5	Chlorine	mg/L	188	0.69	0.21	0.74	0.32	1.31

Exploratory Data Analysis (EDA)

The Kaiser-Meyer-Olkin sampling adequacy test and the Bartlett's test of sphericity are performed before conducting Principal Component Analysis and Factor Analysis to ensure that the data is appropriate for the analysis. The water quality parameters data was standardized (scaled) before the analysis was carried out since Principal Component Analysis and Factor Analysis are sensitive to units of measurement. Standardizing the data ensures that all parameters are treated equally, preventing any single variable's units from influencing the analysis, thereby providing a more accurate and meaningful interpretation of the results.

Table 2: Kaiser-Meyer-Olkin Sampling Adequacy Test

Kaiser-Meyer-Olkin factor adequacy **Call: KMO (r = water_quality_scaled)**

Overall MSA = 0.67				
MSA for each item =				
pH	**Total Hardness**	**Alkalinity**	**Turbidity**	**Chlorine**
0.63	0.54	0.75	0.59	0.69

Table 2 shows that the overall KMO value is 0.67 (> 0.5) which means that the data is appropriate for the analysis. The KMO for the individual parameters signifies a suitable level of sampling adequacy and indicating that it is acceptable for inclusion in the analysis.

Table 3: Bartlett's Test of Sphericity

Chi-square	285.4294
p.value	1.860868e-55
Df	10

Table 3 shows that the Bartlett's test of sphericity is statistically significant (p.value < 0.05), indicating sufficient correlation between the parameters to proceed with the analysis.

Correlation Analysis

Table 4: Correlation Matrix

	pH	**Total Hardness**	**Alkalinity**	**Turbidity**	**Chlorine**
pH	**1**	0.0894298	0.6665304	0.1255853	-0.749308
Total Hardness	0.0894298	**1**	0.0130593	0.3075231	-0.09188
Alkalinity	0.6665304	0.0130593	**1**	0.0561912	-0.533024
Turbidity	0.1255853	0.3075231	0.0561912	**1**	-0.150567
Chlorine	-0.749308	-0.09188	-0.533024	-0.150567	**1**

Table 4 shows the relationship between the water quality parameters. The correlation analysis reveals that pH has a weak positive relationship with total hardness ($\mathrm{r}_{12} = 0.0894298$) and turbidity ($\mathrm{r}_{13} = 0.1255853$), indicating that these factors do not significantly impact pH levels in water. However, there is a moderate to strong positive correlation between pH and Alkalinity ($\mathrm{r}_{14} = 0.6665304$), suggesting that maintaining proper alkalinity is crucial for stabilizing pH. Most notably, there is a strong negative correlation between pH and Chlorine levels ($\mathrm{r}_{15} = -0.749308$), underscoring the importance of pH management to ensure Chlorine remains effective as a disinfectant. Overall, these findings emphasize the need for integrated water quality management practices to maintain stable pH levels, which is vital for effective disinfection and overall water quality. Total Hardness has an almost negligible relationship with alkalinity ($\mathrm{r}_{23} = 0.0130593$), indicating these two parameters can be managed independently in water treatment. There is a weak to moderate positive correlation between Total Hardness and Turbidity ($\mathrm{r}_{24} = 0.3075231$), suggesting that higher mineral content might be associated with increased particulate matter, affecting water clarity. Additionally, the very weak negative correlation between Total Hardness and Chlorine ($\mathrm{r}_{25} = -0.09188$) implies that hardness minimally impacts Chlorine levels and its disinfecting effectiveness. There is very weak positive correlation between Alkalinity and Turbidity ($\mathrm{r}_{34} = 0.0561912$), indicating that these parameters can be managed independently as changes in one do not significantly affect the other. In contrast, the moderate negative correlation between Alkalinity and Chlorine levels ($\mathrm{r}_{35} = -0.533024$) can decrease Chlorine effectiveness by neutralizing it, which can impact water disinfection. The very weak negative correlation between Turbidity and Chlorine ($r_{45} = -0.150569$) indicates that as turbidity slightly increases, Chlorine levels tend to decrease, but the relationship is not strong. This suggests that while higher Turbidity can somewhat reduce Chlorine's effectiveness in disinfection by shielding microorganisms and consuming Chlorine, this effect is minor. Therefore, while managing turbidity is important for optimal Chlorine disinfection and the overall water quality.

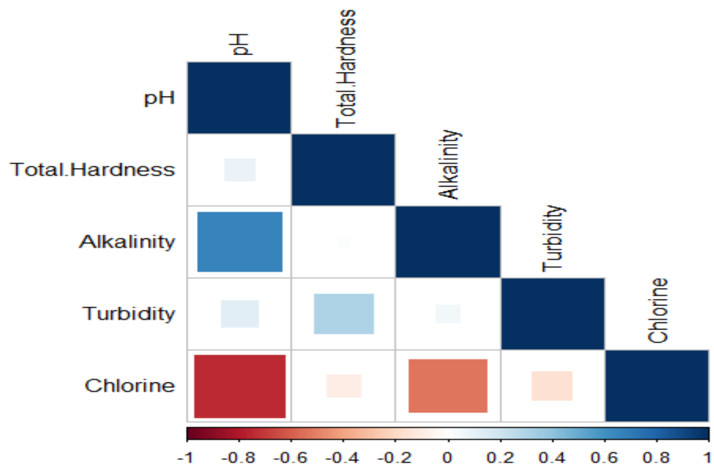


Figure 1: Correlogram

Figure 1 gives a visualization of the correlation matrix. Positive correlations are displayed in a blue scale while negative correlations are displayed in a red scale. The strength of the relationships is indicated by the size and darkness of the squares: larger and darker squares represent stronger correlations, while smaller and lighter squares represent weaker correlations.

Principal Component Analysis
Table 5: Total Variance Explained

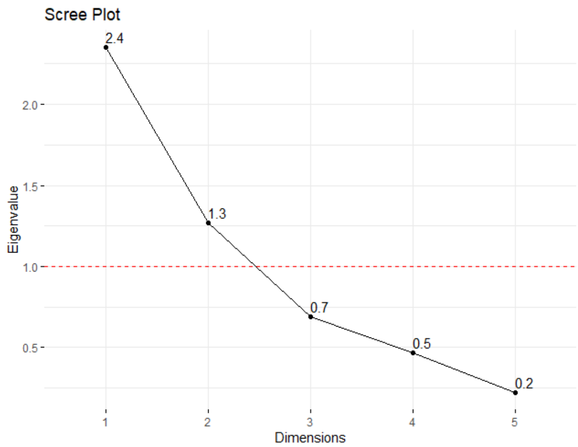
Importance of components:					
	PC1	**PC2**	**PC3**	**PC4**	**PC5**
Standard deviation	1.5332	1.1274	0.8312	0.68262	0.47044
Proportion of Variance	0.4702	0.2542	0.1382	0.09319	0.04426
Cumulative Proportion	0.4702	0.7244	0.8626	0.95574	1

Table 5 shows the standard deviation, proportion of variance, and the cumulative proportion of the principal components. The Principal Component Analysis (PCA) results reveal a clear hierarch of variance explanation, with the first two components (PC1 and PC2) accounting for a combined 72.44% of the total variance, indicating that they capture the most significant patterns and relationship in the data. PC1, explaining nearly half of the variance (47.02%), represents the primary structure or correlation among variables, likely related to the overall water quality or a dominant pollution source. PC2, explaining approximately one-quarter of the variance (25.42%), represents a secondary pattern or correlation, possibly related to a specific pollutant. The remaining components (PC3, PC4, and PC5) explain smaller proportions of variance (13.82%, 9.319%, and 4.426%, respectively), representing a minor aspect of water quality.

Table 6: Component Matrix

Rotation (n x k) = (5 x 5):					
	PC1	**PC2**	**PC3**	**PC4**	**PC5**
pH	0.5981823	0.10316947	-0.038095592	-0.14093976	-0.78116497
Total Hardness	0.1215913	-0.70160956	-0.699343331	0.05738367	0.02419893
Alkalinity	0.5270153	0.19657081	-0.03205922	0.77300942	0.29162201
Turbidity	0.1684001	-0.67449923	0.713002518	0.09051773	-0.0112314
Chlorine	-0.5668243	-0.05925187	-0.008200445	0.60918476	-0.55138567

Table 6 presents the component loadings, which represent the correlation coefficients between the original water quality parameters and the Principal Components (PCs). These component loadings quantify the strength and direction of the relationships between each parameter and the patterns captured by the PCs.



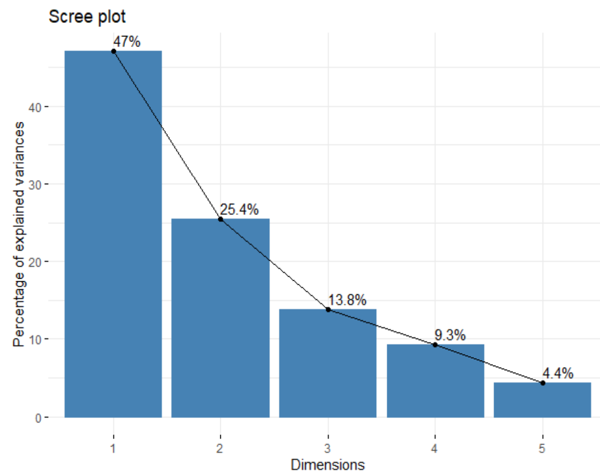


Figure 2: Eigenvalue Scree Plot **Figure 3: Percentage of Explained Variances Scree Plot**

Figure 2 reveals that the first two principal components (PC1 and PC2) have eigenvalues greater than 1, indicating that they explain a significant amount of variance in the data. The cutoff at eigenvalue greater than 1 suggests that these two components are the most important and should be retained. The first principal component (PC1) has an eigenvalue of 2.4, explaining the most variance, followed by PC2 (1.3). The third, fourth, and fifth components (PC3, PC4, and PC5) have eigenvalues less than 1, indicating that they explain relatively little variance and can be ignored.

Figure 3 visually displays the percentage of variance explained by each principal component, providing a graphical representation of the relative contribution of each component to the overall variability in the data.

Table 7: Loading Matrix of the First Two Principal Components

	PC1	PC2
pH	**0.598182**	0.1031695
Total Hardness	0.121591	**-0.70161**
Alkalinity	**0.527015**	0.1965708
Turbidity	0.1684	**-0.674499**
Chlorine	**-0.566824**	-0.059252

Table 7 shows the loading matrix of the first two components which capture the majority of the variability in the data. The selection was determined by examining the cumulative variance explained and the scree plot, the first two principal components together represent 72.44% of the variability with eigenvalues greater than 1 making them a suitable stopping point for component extraction.

First Principal Component (PC1): The component loadings on PC1 reveal the correlations between the water quality parameters and the first principal component, which explains 47.02% of the total variance. The primary pattern of water quality variation is driven by pH, alkalinity, and chlorine, which are strongly interconnected, as evidenced by their high component loadings (above 0.5 and below -0.5). As the primary pattern of water quality variation increases, the positive correlation with pH and Alkalinity indicates a shift towards more alkaline water conditions, potentially leading to scaling and pipe corrosion, reduced disinfection effectiveness, and increased pathogen growth, while also providing a higher buffering capacity to stabilize water chemistry. The negative correlation between Chlorine and PC1 indicates that as water quality variation increases, Chlorine levels decrease, leading to inadequate disinfection and increased risk of waterborne diseases such as cholera, typhoid, and dysentery. Total hardness and Turbidity have weaker correlations, contributing less to the primary pattern. The result highlights the importance of monitoring and managing pH, alkalinity, and chlorine to maintain acceptable water quality standards.

Second Principal Component (PC2): The second principal component (PC2) accounts for 25.42% of the variation in water quality parameters. Total Hardness and Turbidity have the strongest associations with PC2 (below < -0.5), with increasing levels of these parameters corresponding to decreasing PC2 values, suggesting a potential decline in water quality. The strong negative correlation with Total Hardness indicates that higher levels of mineral content in the water are associated with decreased water quality, leading to increased risk of scaling and corrosion, as well as aesthetic issues like staining and spotting, and potentially harming aquatic life by disrupting the balance of essential nutrients. The negative correlation of Turbidity on PC2 (-0.674499) indicates that higher levels of suspended solids and particulate matter are associated with decreased water quality, leading to cloudy water that may harbor harmful pollutants or pathogens, interfere with water treatment processes, and impact aquatic life by reducing light penetration and harming habitats, ultimately compromising water quality and potentially affecting human health and water treatment efficacy. The weaker correlations with pH, Alkalinity, and Chlorine indicates minor contributions to PC2.

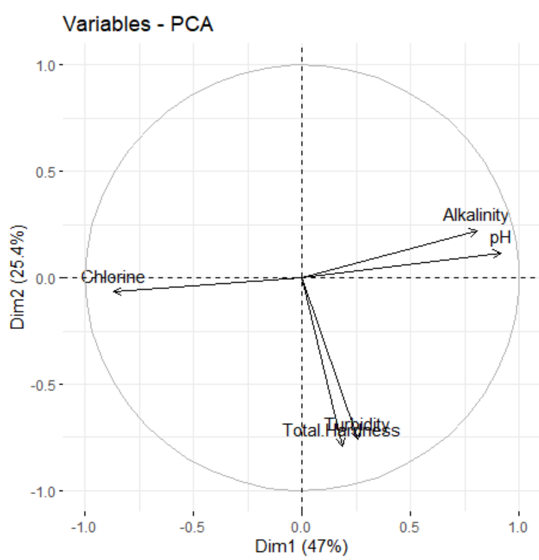


Figure 4:Biplot of the Parameters with Respect to the Principal Components

Figure 4.4 shows a visualization of the impact of each parameter on the first two principal components. The parameter that are grouped together are positively correlated and variables that are negatively correlated are displayed to the opposite sides of the biplot's origin.

Factor Analysis
Table 8: Principal Axis Factoring

Factor Analysis using method = pa
Call: fa (r = water_quality_scaled, nfactors = 2, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix

	PA1	PA2	h2	u2	Com
pH	**0.95**	0.10	0.92	0.083	1.0
Total Hardness	0.03	**0.55**	0.30	0.702	1.0
Alkalinity	**0.70**	-0.01	0.49	0.513	1.0
Turbidity	0.08	**0.56**	0.32	0.680	10
Chlorine	**-0.77**	-0.14	0.61	0.386	1.1

	PA1	PA2
SS loadings	2.00	0.64
Proportion Variance	0.40	0.13
Cumulative Variance	0.40	0.53
Proportion Explained	0.76	0.24
Cumulative Proportion	0.76	1.00

Table 8 shows the Principal Axis Factoring (PAF) analysis:

Factor 1 (PA1): This factor can be aptly named “Water Chemistry Factor”, it is driven by water chemistry parameters, specifically pH and Alkalinity, which have a strong positive correlation with factor loadings exceeding 0.5, and Chlorine, which has a strong negative correlation with a factor loading less than -0.5. pH has a high communality of 0.92, indicating its major role in this factor, while chlorine also has a strong influence with a communality of 0.61, but alkalinity contributes to a lesser extent with a communality of 0.49. This factor captures the combined effects of these water chemistry parameters. The underlying factor driving the Water Chemistry Factor is “Human Activities”, particularly industrial effluents that release acidic and basic substance, altering water pH, and agricultural runoff from fertilizers and waste, which enters water bodies and increases Alkalinity. Additionally, industrial processes, such as paper bleaching, textile manufacturing, and chemical production, can release chlorine-containing compounds into water bodies, further impacting water chemistry. These human activities collectively influence the Water Chemistry Factor, highlighting the need for sustainable practices to mitigate their effects on water quality.

Factor 2 (PA2): Based on the results, PA2 can be interpreted as a factor related to “Water Mineral Factor”, driven by presence of minerals, rocks, and sediments influencing Total Hardness (0.55) and Turbidity (0.56) levels. The high factor loadings (>0.5) suggest “Geological Processes” like weathering, dissolution, erosion, and sedimentation from surrounding rocks and soil, affecting water, leading to high levels of Total Hardness and Turbidity. This may impact water quality, potentially requiring treatment to remove excess minerals and improve clarity, and could also affect health, causing gastrointestinal issues or other problems if consumed without proper treatments. Furthermore, the high levels of minerals and turbidity may harm the aquatic ecosystem of the dam, alter the food chain, and lead to scaling and damage to infrastructure like pipes and water treatment plants.

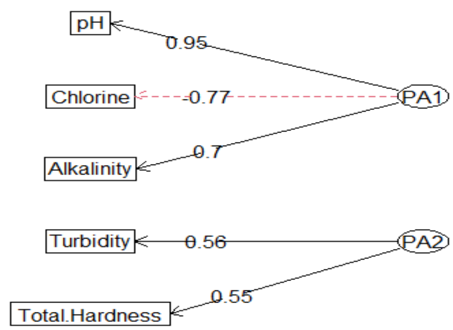


Figure 5: Loading Plot of Selected Parameters on the Factor

Figure 5 displays analysis diagram, which highlights the selected parameters that most strongly represent each factor, based on the criteria that the factor loading is greater than 0.5 or less than -0.5. PA1 (Water Chemistry Factor) is primary represented by pH, Chlorine, and Alkalinity, while PA2 (Water Mineral Factor) is primary composed by Turbidity and Total Hardness.

Conclusion

This study utilizes Principal Component Analysis (PCA) and Factor Analysis (FA) to analyze water quality parameters at the Laminga Treatment Plant in Jos Plateau State, Nigeria. The analysis reveals significant interdependencies among pH, alkalinity, and chlorine, highlighting the need for integrated management. PCA identified two principal components explaining 72.44% of the variance in water quality, with pH, alkalinity, and chlorine being crucial for maintaining proper water chemistry. FA identified two underlying factors: the "Water Chemistry Factor," driven by human activities, and the "Water Mineral Factor," influenced by geological processes. The high communalities of pH, alkalinity, and chlorine in the FA imply that these parameters are strongly influenced by underlying factors and are critical in assessing overall water quality.

To improve water quality, the Plateau State Water Board (PSWB) and the Laminga Treatment Plant should implement targeted monitoring and management of pH, alkalinity, and chlorine, with regular adjustments using appropriate chemicals and optimizing chlorine dosing with stabilizers. Addressing total hardness and turbidity through water softening techniques and advanced filtration systems will further enhance water quality. Additionally, the Plateau State Government should invest in infrastructure improvements, support new and upgraded water treatment facilities, and launch public awareness campaigns on water conservation and pollution prevention. Encouraging research and innovation through funding and collaboration with academic institutions is crucial. Specific measures for the Lamingo Dam include regulating industrial effluents and agricultural runoff, promoting sustainable agricultural practices, and implementing sediment control measures. By following these recommendations, significant improvements in water quality in Jos Plateau State can be achieved, ensuring safe and clean water for the population and preserving the aquatic ecosystem.

REFERENCES

Aliyu, G.A., Zailani, S. & Akuha, K.D. (2022). Spatial and Seasonal Variation of Water Quality in Savanna River System. *Journal of Science Technology and Education, 10*(2), ISSN: 2277-0011.

Alley, E.R. (2007). *Water Quality Control Handbook*. (2nd ed.). WEF Press: McGraw-Hill, New York.

Ayoade J. O. (1988). *Tropical Hydrology and Water Resources*. Macmillan, Ibadan, Nigeria.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* Guilford Press, New York.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Chapman, D. (1992). Water Quality Assessments. A Guide to Use of Biota, Sediments and Water in Environmental Monitoring. (2nd ed.). 10.4324/9780203476710.

Gray, N.F. (2017). *Water Technology*. (3rd ed.). CRC Press. London.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis.* (8th ed.). Cengage Learning EMEA, Hamoshire, United Kingdom.

Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. (6th ed.). Pearson Education, Inc. USA.

Jolliffe I. T. (2002). *Principal Component Analysis*. (2nd ed.). Springer-Verlag, New York. USA.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Rencher, A.C. & Christensen, W.F. (2012). *Methods of Multivariate Analysis.* (3rd ed.). John Wiley & Sons, Inc., Hoboken, New Jersey.

Spellman, F.R. (2013). *Handbook of Water and Wastewater Treatment Plant Operations*. (3rd ed.). CRC Press, Boca Raton.

Tebbutt, T. H. Y. (1990). *Basic Water and Wastewater Treatment*. Elsevier Butterworths Heinemann, Oxford. UK.

Thomas, E.O. (2023). Spatial evaluation of groundwater quality using factor analysis and geostatistical kriging algorithm: a case study of Ibadan Metropolis, Nigeria. *Water Practice & Technology, 18*(3), 592.

Tijjani, A.F., Alhassan, B., Salisu, U. & Hassan, A.M. (2023). Multivariate Statistical Analysis for Assessment of Groundwater Quality around Dumpsites in Bauchi metropolis, Bauchi State, Nigeria. *J. Appl. Sci. Environ. Manage, 27*(11), 2509-2514.