ADEJUWON, E.O., OGWUELEKA, T.C., OGUNGBEMI, E.O., PRABHU, R., RENDON-NAVA, A. and YATES, K. 2025. Assessment of surface water quality using chemometric tools: a case study of Jabi Lake, Abuja, Nigeria. *Iranian journal of science and technology, transactions of civil engineering* [online], 49(1), pages 829-852. Available from: https://doi.org/10.1007/s40996-024-01712-2

Assessment of surface water quality using chemometric tools: a case study of Jabi Lake, Abuja, Nigeria.

ADEJUWON, E.O., OGWUELEKA, T.C., OGUNGBEMI, E.O., PRABHU, R., RENDON-NAVA, A. and YATES, K.

2025

Iranian Journal of Science and Technology, Transactions of Civil Engineering (2025) 49:829–852
https://doi.org/10.1007/s40996-024-01712-2

RESEARCH PAPER

Assessment of Surface Water Quality Using Chemometric Tools: A Case Study of Jabi Lake, Abuja, Nigeria

E. O. Adejuwon[1] · T. C. Ogwueleka[2] · E. O. Ogungbemi[3] · R. Prabhu[1] · A. Rendon-Nava[1] · K. Yates[1]

Received: 31 March 2022 / Accepted: 15 December 2024 / Published online: 2 January 2025

Abstract
Water pollution has become a growing threat to human society and natural ecosystems in recent decades. It increases the need to understand surface water quality assessment better using chemometric tools within aquatic systems. This study sampled the water quality of 21 parameters at multiple sampling points in Jabi Lake during wet and dry seasons (August–December 2021) using various statistical methods including cluster analysis, principal component analysis/factorial analysis, discriminant analysis, and box plot analysis. These samples were examined for physicochemical parameters employing standard techniques. The study revealed significant seasonal variations in water quality. During the wet season, key measurements included total dissolved solids (100.40 mg/l), dissolved oxygen (13.72 mg/l), and electrical conductivity (97.14 µs/cm). The dry season showed higher levels of most parameters, with total dissolved solids at 137.91 mg/l and electrical conductivity at 230.93 µs/cm. Statistical analysis identified strong correlations between various parameters, notably between phosphate and total hardness in the wet season ($r = 0.978$, $\alpha = 0.05$) and between pH and temperature in the dry season ($r = 0.995$, $\alpha = 0.05$). The study identified four principal components explaining 98.5–100% of the variance, representing various pollution sources including organic waste, domestic sewage, and natural factors. The findings indicated that dry season water samples were more polluted, with some parameters exceeding World Health Organisation standards, suggesting potential health risks. The research demonstrated the effectiveness of multivariate statistical techniques in analysing complex water quality data and provided valuable insights for water resource management, particularly regarding seasonal variations' impact on water quality.

Keywords Multivariate tools · Principal component analysis · Factorial analysis · Cluster analysis · Box plot analysis · Discriminant analysis · Water pollution

Abbreviations

CA	Cluster analysis
PCA	Principal Component Analysis
FA	Factorial Analysis
DA	Discriminant Analysis
BPA	Box Plot Analysis
EC	Electrical Conductivity
Cl^{-1}	Chloride ion
EC	Electrical conductivity
DO	Dissolved Oxygen
BOD_5	Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
Zn	Zinc
TA	Total Alkalinity
TDS	Total Dissolved Solids
Na^{+1}	Sodium
KMO	Kaiser–Meyer–Olkin
HCA	Hierarchical Cluster Analysis
WHO	World Health Organization

E. O. Adejuwon
e.adejuwon@rgu.ac.uk

T. C. Ogwueleka
toochukwu.ogwueleka@uniabuja.edu.ng

E. O. Ogungbemi
emmanuel.ogungbemi@ofgem.gov.uk

R. Prabhu
r.prabhu@rgu.ac.uk

A. Rendon-Nava
a.rendon-nava@rgu.ac.uk

K. Yates
k.yates@rgu.ac.uk

[1] Robert Gordon University, Aberdeen, UK

[2] Department of Civil Engineering, The University of Abuja, Abuja, Nigeria

[3] The Office of Gas and Electricity Markets, London, UK

1 Introduction

Rivers, Lakes, and streams play a significant societal role (Shakhman and Bystriantseva 2021a). Water is a vital and fateful natural resource for survival and sustenance because it provides water for agriculture, human needs, industry, and transportation (Chadli and Boufala 2021; Markad et al. 2021). Most of man's activities depend on quality water resources (Fadel et al. 2021). Many natural and anthropogenic factors can affect water quality (Dash and Kalamdhad 2021). Moreover, urbanisation and rapid population growth are highly associated with the deterioration of surface water quality (Egbueri and Mgbenu 2020). Water pollution is a significant problem in developing countries like Nigeria (Ikpeze and Aririguzoh 2023; Imam et al. 2023; Matouke and Abdullahi 2020). The surface water quality of Jabi Lake is getting highly polluted daily with the rapid population growth, urbanisation, and haphazard agricultural and industrial production, all giving rise to increased emissions of organic and inorganic pollutants into the aquatic environment. The deterioration of water resources, both in quality and quantity, is a consequence of pollution, inadequate management, diminished drinking water quality, and heightened public health risks (Balcerowska-Czerniak and Gorczyca 2024a). Bhatt et al. (2024) in their comprehensive study of the Rispana River's aquatic environment highlighted the degradation of the river due to pollution and climate change. In general, microorganisms, organic compounds, and toxic heavy metals are the sources of contamination of water (Sharma et al. 2023). Many are products of the natural mobilisation of these elements, such as weathering reactions, biological activity, volcanic emissions, and rock solubilisation which can mix into the surface water (Sager and Wiche 2024; Keerthanan et al. 2023; Apestegui et al. 2023; Mishra 2023). The negative impact could be rendered by indirect factors, such as atmospheric precipitation, land management and climate change (Ahmed et al. 2022). For instance, Kumar et al. (2022, 2023) reported that only 48% of the urban and semi-urban areas and 39% of the rural regions have access to a portable water supply. Tazoe (2023) reported that on the average 40% of the lakes and rivers of the planet have been polluted by heavy metals and proposed techniques for water quality monitoring which should be cost-effective, environmentally friendly, selective, and sensitive enough to detect traces with good precision. Thus, continuous water quality monitoring is required to sustain water resources and registered changes (Kumar et al. 2022; Luo et al. 2020). It is one of the crucial activities of environmentalists, so to this end, governments need to construct sampling stations along the rivers and lakes for regular checks. Although several water studies have been conducted in Nigeria in recent years There is an urgent need for the constant monitoring of surface water as a significant component of water resource management. Lakes and spring waters, the primary water sources in Nigeria and other developing countries have historically been neglected in quality assessment and management (Nnaji et al. 2023; Egun and Oboh 2023; Isukuru et al. 2024; Ubuoh et al. 2023; Chen 2024). Some of the rivers include Eha-Amufu Ebonyi River-Enugu State, Iwofe River-Port-Harcourt-Rivers State, River Otamiri-Imo State, Ona River-Ibadan, Iju River-Ogun State, Ogbor Hill River Water-Southern Nigeria, and many others (Okey-Wokeh et al. 1359; Famuyiwa et al. 2023; Akintola et al. 2024; Ndukwe et al. 2023; Eze et al. 2023).

Moreover, there has been increasing awareness of and concern about surface water pollution worldwide in recent years. New approaches toward the sources of pollutants and achieving sustainable exploitation of water resources have been developed (Hue and Thanh 2020; Elkorashey 2022). The best approach to avoid misinterpretation of environmental monitoring data is the application of chemometric methods for environmental data classification and modelling (Hue and Thanh 2020; Elkorashey 2022; Banda and Kumarasamy 1584). Also, Omeka et al. (2024) in their review identified the commonly used water quality assessment techniques in Nigeria, he mentioned the drawbacks in the application of these techniques as well as the gaps in water quality assessment and monitoring using an evidence-based method approach. However, the combined use of environmental tools such as multivariate statistical techniques such as cluster analysis (CA), factor analysis (FA), principal component analysis (PCA), discriminant analysis (DA) and box plot (BP) enables the classification of water samples into distinct groups, source apportionments, relationships, and differences in the parameters used based on hydrochemical characteristics (Zavareh et al. 2021; Liu and You 2023). Compared to the conventional method, these analyses can detect long-range correlations that are artificial non-stationeries. Conventionally, the usual interpretation technique of surface water quality is only a univariate procedure that is inadequate to characterise similarities and differences between samples or variables in a complex environment. For example, Chitrakar (2020) applied cluster analysis (CA) to delineate surface water quality monitoring sites. At the same time, Rangeti et al. (2021) and Shafii (2019) used it in addition to discriminant analysis (DA) to identify significant parameters and optimise monitoring networks of groundwater quality data. Cluster analysis helps group objects (cases) into classes (clusters) based on similarities within a category and dissimilarities between different classes. The results of cluster analysis (CA) help interpret the data and indicate patterns (Chitrakar et al. 2020).

Multivariate data analysis has been revealed to reduce data without losing the original information (Markad et al. 2021; Zavareh et al. 2021). Liu et al. (2023) explained that these methods were used to identify water pollution, water pollution parameters, and the classification of stations by principal components analysis. Hammoumi et al. (2024) made use of the water quality index (WQI) principal component analysis (PCA) to determine the basic parameters of pollution. The evaluation of the qualification of water resources using these methods has been successful and has been used by many researchers (Markad et al. 2021; Soares et al. 2020; Jahin et al. 2020; Abdel-Fattah et al. 2020; Edoreh et al. 2021; Bhatt et al. 2024). Researchers have found that principal component analysis, a method within multivariate data analysis, effectively reduces data dimensions, highlights key parameters affecting water quality variations, evaluates variable relationships, and uncovers patterns in the data distribution, ultimately aiding in the assessment of water quality (Rautela et al. 2023; Talukdar et al. 2023; Zahoor and Mushtaq 2023). Principal Component Analysis (PCA) has allowed the identification of a reduced number of latent factors with a hydro chemical meaning: mineral contents, man-made pollution and water temperature (Vega et al. 1998). Pollution sources, such as spatial (pollution from anthropogenic origin) and temporal (seasonal and climatic) sources of variation affecting the quality and hydrochemistry of river water have been differentiated and assigned to polluting sources (Markad et al. 2021; Ali et al. 2024; Wieczorek et al. 2024; Dimri et al. 2023; Shulembayeva et al. 2023; Muniz and Oliveira-Filho 2001; Zhou et al. 2023; Olalekan et al. 2023). At the same time, PCA has allowed the explanation of related parameters by only one factor (Markad et al. 2021; Elkorashey 2022; Zhou et al. 2023) and exposed the vital factors responsible for seasonal changes in river water quality. This data-mining technique will further help reduce the number of pollution parameters to be tested and the subsequent analysis cost (Ibrahim et al. 2023).

The current pollution status of Jabi Lake is alarming. Several scientific reports about the condition of water quality of surface waters worldwide have been published with several pollutants such as Organic Pollutants, Inorganic Pollutants, Radioactive Pollutants, Suspended Solid, Pathogens, Nutrients and Agricultural Pollutants, Thermal Pollution (Shakhman and Bystriantseva 2021a, 2021b; Egbueri and Mgbenu 2020; Bhatt et al. 2024; Yassin et al. 2024; Matta et al. 2023). Egbueri et al. (2020) measured the degree of heavy metal contamination, identified potential sources of pollution, and evaluated the health risks posed to humans by consuming water from Ojoto Province, Nigeria. Yassin et al. (2024) performed a study to assess the spatial as well as indexical water quality, identifying contamination sources, hotspots, and evaluated associated health risks pertaining to nitrate and fluoride in the Al-Hassa region, KSA. Shakhman et al. (2021b) reported modern anthropogenic load on the surface water of the Southern Bug River Basin in a changing climate. His study shows that the use of surface water of the Southern Bug River Basin for drinking, fishery, cultural and recreational needs is related to certain environmental risks. Balcerowska-Czerniak et al. (2024b) employed a multivariate statistical quality control chart based on principal component analysis (PCA) to present a universal methodology for monitoring many parameters simultaneously and early detection of out-of-control samples in a real-time mode. Since the lake is used for beneficial purposes, it becomes necessary to assess its pollution using some physicochemical parameters and the potential risks to residents and tourists. To date, considering the lake's physicochemical characteristics using chemometric tools, requisite data have not been collated and analysed to document the current pollution status of Jabi Lake and its environs. This uncollated document on the current pollution status of Jabi Lake and many other factors informed this research by contributing to local literature. This study can lead to more targeted water quality management strategies. However, chemometrics can be used to create predictive models for changes in water quality based on both historical and present data. These models might be useful for management and monitoring in the future, allowing stakeholders to foresee and proactively solve possible quality problems.

This research aims to evaluate the water quality in terms of physicochemical characteristics of the man-made Jabi Lake in the Federal Capital Territory, Abuja, Nigeria, using standard methods. It aims to assess the water quality of Jabi Lake using chemometric tools by determining the level of physicochemical parameters and comparing values with the World Health Organization (WHO) permissible limits. In this research, 21 water quality parameters were selected and collected between August to September 2021 for the wet season and November to December 2021 for the dry season at three and five sampling points, respectively, in Jabi Lake. The multivariate statistical methods (i.e., cluster analysis (CA), principal component analysis (PCA), factorial analysis (FA), discriminant analysis (DA), and box plot analysis (BPA)) were applied to analyse the water quality data to interpret better, understand and define the water quality parameters and specific sources of water quality deterioration and contamination in the area. Firstly, similarities and dissimilarities among the parameters and sampling points were classified utilising cluster analysis (CA) and box plot analysis (BPA). The complex water quality data sets were analysed to extract latent water quality factors using factorial analysis (FA)/ principal component analysis (PCA) and discriminant analysis (DA). Finally, the effects of possible pollution sources on water quality were identified.

2 Materials and Methods

2.1 Study Area

The sample used in this study was from Jabi Lake, Abuja, Nigeria. Jabi Lake is a natural water body in the Jabi district of Abuja, Nigeria. Jabi Lake, a human-made reservoir of water from the foot of Katampe rocks, is located within the Kado and Jabi districts of the federal capital territory (FCT) (Daniel et al. 2023), about 1.78 km long and 0.55 km wide, as shown in Fig. 1. It's located between geographical coordinates of 9° 3′ 45″ North, and 7° 25′ 27″ East respectively (Ogoko and Sylvester 2020). The peculiarity of Jabi Lake and its environs as a savannah zone vegetation of the west African sub-region makes it unique for change pattern analysis due to urbanisation and population growth. Jabi has a tropical climate. The climate of the lake basin is characterised by temperate weather, influenced by the lake's surrounding conditions. The temperature of the lake ranges between 26.3 and 31.5 °C annually. The warmest month is March, with an average temperature of 30.1 °C | 86.3 °F. In august, the average temperature is 23.3 °C | 74.0 °F. It is the lowest average temperature of the whole year. The average temperatures vary during the year by 6.8 °C | 12.3 °F. Winter is considered December, January, and February; spring is March through May; summer is June through August; and fall or autumn is September through November (Matta et al. 2023).

2.2 Sample Collection and Analytical Techniques

The water sample was collected during August–December 2021 at 3 and 5 locations for different seasons at the same time intermittently. Sample collection and analytical techniques were conducted according to edition (Edition 2011; Rice 2012) from predetermined points. The study used a purposive sampling method to select the three and five sampling points. Criteria for selecting sampling points were based on patterns of land use in the area and the types and pathways of contamination and anthropogenic activities. The first sampling point is the vegetation area at Jabi Park. The second sampling point is under the Jabi Lake bridge. The third sampling point is the recreation section. The fourth sampling point is the boat ride section, while the fifth sampling point is the discharge point of municipal waste—the Jabi mall section. The samples were collected between the hours of 12:00–2:00 pm.

The water sample was collected between August and September for the wet season and between November and December for the dry season. Samples were taken 0.1–0.3 m below the water surface. A volume of 1 L of water was collected in sterile bottles. Samples were stored at 4 °C and transferred to the laboratory within two hours after collection. All water measurements were carried out within 24 h after sampling (Edition 2011) for physicochemical, biological and metal analysis, except for pH, total dissolved solids (TDS), electrical conductivity (EC) and temperature, which was measured on the site using a multi-meter (Hanna 9813–6) because they change with storage time. The instruments in situ were calibrated using a specific calibration solution before each measurement (Rice 2012). The general framework for Jabi Lake water quality assessment involved the comprehensive evaluation of river integrity from the perspective of physicochemical, water quality and biological aspects. The parameters analysed were colour, temperature, electrical conductivity, total dissolved solid, total alkalinity, pH, turbidity, dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD5), sulphate, nitrate, iron, chloride ion, zinc, total hardness (TH), salinity, sodium, manganese, phosphate and copper. Appropriate quality assurance and control procedures ensured the data collection was reliable, repeatable, and unbiased (Rice 2012). Collection, handling, and transportation of the water sample were done according to WHO and Rice et al. (2012) ways of collecting, handling, and transporting water samples (Edition 2011; Rice 2012; World Health Organization 2020).

Salinity was measured using the Eutech salinity pocket tester SaltTestr. Dissolved oxygen was examined using dissolved oxygen (DO) meter and treatment with manganous sulphate solution and alkaline iodide-azide solution (Winkler reagents). The water sample was poured into a bottle and incubated for five days to measure BOD5 using Winkler's azide methods (Edition 2011; Rice 2012). Alkalinity was determined using American Public Health Association, Standard (APHA) methods (Rice 2012), after which samples for metal determination were collected in a one-litre container and preserved with 4 ml of concentrated nitric acid. Water samples were taken to analyse iron, manganese, nitrate and sulphate, and measurements were taken spectrophotometrically after reduction with the appropriate solution (Rice 2012; World Health Organization 2021) COD was determined in the laboratory using a dichromate reflex technique (Edition 2011; Rice 2012). Turbidity was directly measured with a turbidity meter (Hach 2100 AN). All the water quality parameters are expressed in mg/l, except temperature (°C), pH, turbidity (NTU), EC (µs/cm), and colour (PtCo). The statistical summary of the water quality parameters sampled at three and five sampling points is shown in Tables 1 and 2, respectively.

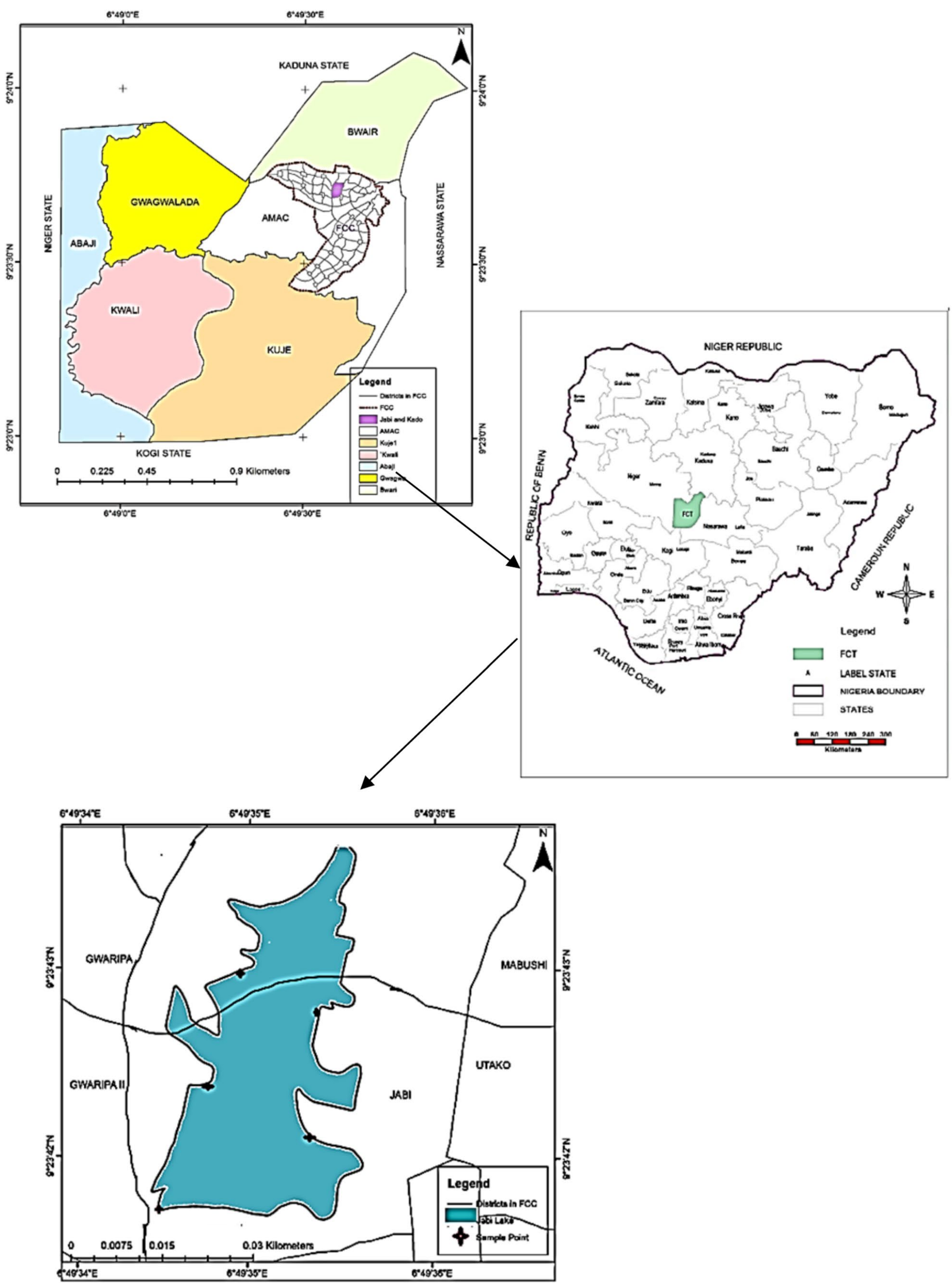


Fig. 1 Map showing Jabi Lake, Nigeria (the study area) (Kaanayochukwu et al. 2019)

Table 1 Water quality data and their descriptive statistic parameters at the sampling sites Across Jabi Lake during the wet season

S/N	Parameter	Units	S1	S2	S3	Mean	SD	Min	Max
1	Colour	PtCo	113.33	66.33	54.33	77.99	31.18	54.33	113.33
2	Temperature	°C	27.33	27.37	27.43	27.38	0.05	27.33	27.43
3	Electrical Conductivity	µs/cm	106.07	95.30	90.07	97.14	8.15	90.07	106.07
4	Total Dissolved Solid	mg/l	111.37	94.03	95.80	100.40	9.54	94.03	111.37
5	Total Alkalinity	mg/l	78.33	80.00	80.00	79.44	0.96	78.33	80.00
6	pH	pH	7.37	7.47	7.43	7.42	0.05	7.37	7.47
7	Turbidity	NTU	8.26	5.70	4.53	6.16	1.91	4.53	8.26
8	Dissolved Oxygen	mg/l	14.10	12.03	15.03	13.72	1.54	12.03	15.03
9	Chemical Oxygen Demand	mg/l	22.80	21.20	22.20	22.07	0.81	21.20	22.80
10	Biochemical Oxygen Demand (BOD_5)	mg/l	5.80	5.43	5.63	5.62	0.19	5.43	5.80
11	Sulphate	mg/l	9.00	6.67	6.67	7.45	1.34	6.67	9.00
12	Nitrate	mg/l	4.67	1.53	4.03	3.41	1.66	1.53	4.60
13	Iron	mg/l	0.42	0.33	0.34	0.36	0.05	0.32	0.42
14	Chlorine ion	mg/l	24.81	24.17	22.91	23.96	0.97	22.91	24.81
15	Zinc	mg/l	0.97	1.01	0.82	0.933	0.10	0.82	1.01
16	Total Hardness	mg/l	102.00	81.333	83.33	88.89	11.40	81.33	102.00
17	Salinity	PSU	22.33	20.00	22.00	21.44	1.26	20.00	22.33
18	Sodium	mg/l	15.13	13.90	15.53	14.86	0.85	13.90	15.53
19	Manganese	mg/l	0.63	0.70	0.27	0.53	0.23	0.27	0.70
20	Phosphate	mg/l	1.24	0.25	0.49	0.66	0.52	0.25	1.24
21	Copper	mg/l	0.55	0.44	0.417	0.469	0.07	0.42	0.55

where *S1* Sampling points 1, *S2* Sampling points 2, *S3* Sampling points 3, *SD* Standard Deviation

Table 2 Water quality data and their descriptive statistic parameters at the sampling sites across Jabi Lake during the dry season

S/N	Parameter	Units	S1	S2	S3	S4	S5	Mean	SD	Min	Max
1	Colour	PtCo	49.67	26.33	54	91.33	30	50.27	25.91	20	194
2	Temperature	°C	30.8	30.63	30.63	30.9	30.03	30.60	0.34	29.8	31.2
3	Electrical Conductivity	µs/cm	237.33	219	227.33	235.67	235.33	230.93	7.71	214	247
4	Total Dissolved Solid	mg/l	141.23	132.57	135.27	140.5	140	137.91	3.79	129.7	144.5
5	Total Alkalinity	mg/l	87.33	82	82	86	83.33	84.13	2.42	74	94
6	pH	pH	7.4	7.47	7.47	7.37	7.63	7.47	0.10	7.3	8.1
7	Turbidity	NTU	5.39	3.82	4.15	7.45	4.04	4.97	1.52	2.72	6.94
8	Dissolved Oxygen	mg/l	7.64	7.17	7.47	7.8	7.97	7.61	0.31	6.9	8.4
9	Chemical Oxygen Demand	mg/l	12.43	13.8	12.53	12.17	11.67	12.52	0.79	10	14.2
10	Biochemical Oxygen Demand (BOD_5)	mg/l	4.27	4.70	4.23	4.07	3.93	4.24	0.29	3.4	4.8
11	Sulphate	mg/l	6.70	5.33	4.53	5.4	6.67	5.73	9.39	2.2	9
12	Nitrate	mg/l	5.77	4.07	1.77	2.43	1.73	3.15	1.74	0.5	10.4
13	Iron	mg/l	0.36	0.41	0.55	0.79	0.35	0.49	0.18	0.18	1.04
14	Chloride ion	mg/l	26.51	24.61	25.09	26.98	29.13	26.46	1.78	21.3	31.24
15	Zinc	mg/l	1.00	0.89	0.95	0.91	0.9	0.93	0.05	0.70	1.3
16	Total Hardness	mg/l	109.33	98.00	102.00	106.67	110.00	105.2	5.11	68	180
17	Salinity	PSU	43.74	40.61	41.39	44.52	38.27	41.71	2.50	35.15	51.55
18	Sodium	mg/l	10.03	12.47	13.9	15.23	11.37	12.6	2.05	9.9	15.5
19	Manganese	mg/l	1.44	0.58	1.53	0.98	1.16	1.14	0.38	0.1	3.4
20	Phosphate	mg/l	0.71	0.73	0.53	0.61	0.63	0.64	0.08	0.37	0.89
21	Copper	mg/l	0.59	0.61	0.62	0.59	0.52	0.59	0.09	0.5	0.63

where *S1* Sampling points 1, *S2* Sampling points 2, *S3* Sampling points 3, *SD* Standard Deviation

2.3 Multivariate Statistical Techniques

Water quality monitoring procedures produce a complex matrix of various parameters (physical, chemical, microbiological, and biological). These parameter patterns are often complex to be interpreted to elicit meaningful conclusions. Statistical analysis is used to understand the patterns in water quality datasets. Applying different multivariate statistical tools allows the recognition of the possible pollution sources that impact water resources and suggests the most potential solutions. It extracts the most illustrative information from the overall water quality data to investigate the spatial and temporal variations resulting from anthropogenic factors (Banda and Kumarasamy 1584).

2.3.1 Principal Component Analysis, Factorial Analysis, and Descriptive Analysis

Principal component analysis (PCA) decreases the dimensions of the data matrix by rotating the variables (Banda and Kumarasamy 1584). Principal component analysis (PCA) determines the normalised data for component analysis (Markad et al. 2021; Balcerowska-Czerniak and Gorczyca 2024a). Principal component analysis (PCA) has allowed the identification of a reduced number of latent factors with pollution sources, such as spatial (pollution from anthropogenic origin) and temporal (seasonal and climatic) sources of variation affecting the quality and hydrochemistry of river water have been differentiated and assigned to polluting sources (Liu and You 2023; World Health Organization 2021; Yu et al. 2551; Pratama et al. 2020). According to Ogwueleka (2015) Principal component analysis (PCA)/ factorial analysis (FA) for each group formed by cluster analysis (CA) helped to identify spatiotemporal dynamics of water quality in the Kaduna River. At the same time, Principal component analysis (PCA) has allowed the explanation of related parameters by only one factor and exposed the vital factor responsible for seasonal changes in river water quality (Markad et al. 2021; Elkorashey 2022). The general equation expresses the j-th principal component as shown in Eqs. 1–7:

$$PC_J = \sum b_{ij} x_j + e_j \tag{1}$$

where PCj is the j-th principal component.

bij is the loading of x_i on PCj; xi is an independent variable, and ej is the error.

Principal components, PC1, PC2, PC3, PC4, and PC5, are expressed as:

$$PC_1 = \sum b_{i1} x_i + e_1 \tag{2}$$

$$PC_2 = \sum b_{i2} x_i + e_2 \tag{3}$$

$$PC_3 = \sum b_{i3} x_i + e_3 \tag{4}$$

$$PC_4 = \sum b_{i4} x_i + e_4 \tag{5}$$

$$PC_5 = \sum b_{i5} x_i + e_5 \tag{6}$$

The relationship between the observed variable and the factors is given as:

$$x_i = \sum w_{ij} f_i + e_i \tag{7}$$

x_i is the measured variable, fi is the factor score, wij is the factor loading of the i-th factor on the j-th variable, and ei is the error.

2.3.2 Cluster Analysis

The primary goal of cluster analysis is to categorise objects (cases) into classes (clusters) where objects placed within a class are similar but different from those in other classes (Markad et al. 2021; Dash and Kalamdhad 2021). The Euclidean distance is used as a similarity measure to construct the dendrogram diagram, and Ward's method procedure is used as a linkage algorithm to calculate the distances between points in clusters as follows in Eq. 8:

$$d(1,2) = \min\{d(x_i + x_n)\} \text{ for } x_i \text{ in 1 and } x_n \text{ in 2} \tag{8}$$

where d $(x_i = x_n)$ = Euclidean distance.

The distance is measured at each step between every pair of clusters, and the two clusters with the smallest distance are united in this method.

2.3.3 Discriminant Analysis

Discriminant analysis (DA) is a supervised pattern recognition that can classify objects or cases into exhaustive and mutually exclusive groups based on independent variables. Discriminant analysis (DA) is used in this study to predict the variables which discriminate between two natural groupings in river-quality water. The objective of discriminant analysis (DA) is to maximise the similarities of the between-group relative to the within-group variance (Markad et al. 2021; Rangeti and Dzwairo 2021). The model parameters were Wilks' Lambda, an index of the discriminating power ranging between 0 and 1 (the lower the value, the higher its discriminating power); eigenvalue, a measure of variance in the dependent variable for each function; and canonical correlation (discriminant functions), a measure of association between the groups formed by the dependent and the given discriminant function (Markad et al. 2021; Banda and Kumarasamy 2020). Discriminant analysis (DA) is achieved by calculating the variate weight for each independent variable. The variate for the discriminant analysis (DA) is known as the discriminant function and is derived from Eq. 9:

$$z_{jk} = a + w_1 x_{1k} + w_2 x_{2k} + \cdots + w_n x_{nk} \tag{9}$$

where Zjk is the z score of the discriminant function j for object k, a is an intercept, w_1 is the discriminant weight for independent variable 1, and X1k is the independent variable 1 for object k.

2.3.4 Box Plot Analysis

The box plot is used to plot the distribution of a data set. Box plots are also known as box-and- whiskers plots. Box plot analysis is mostly used to investigate water quality and the effect of pollution on stream/river water.

3 Result and Discussion

The observed mean values for the 21 parameters tested in the water quality carried out for each of the five sampling points are shown in Tables 1 and 2. Out of all the samples analysed, colour, turbidity, DO, iron and Mn are out of range, while other parameters are within the recommended limits of WHO guidelines (Edition 2011; World Health Organization 2021). Descriptive statistics were calculated as range, mean, and standard deviations on all data sets. Cluster analysis, discriminant analysis/descriptive analysis, principal component analysis/factor analysis and box plot were applied to determine water quality. SPSS 26 version statistical programme was used for descriptive and multivariate data analysis.

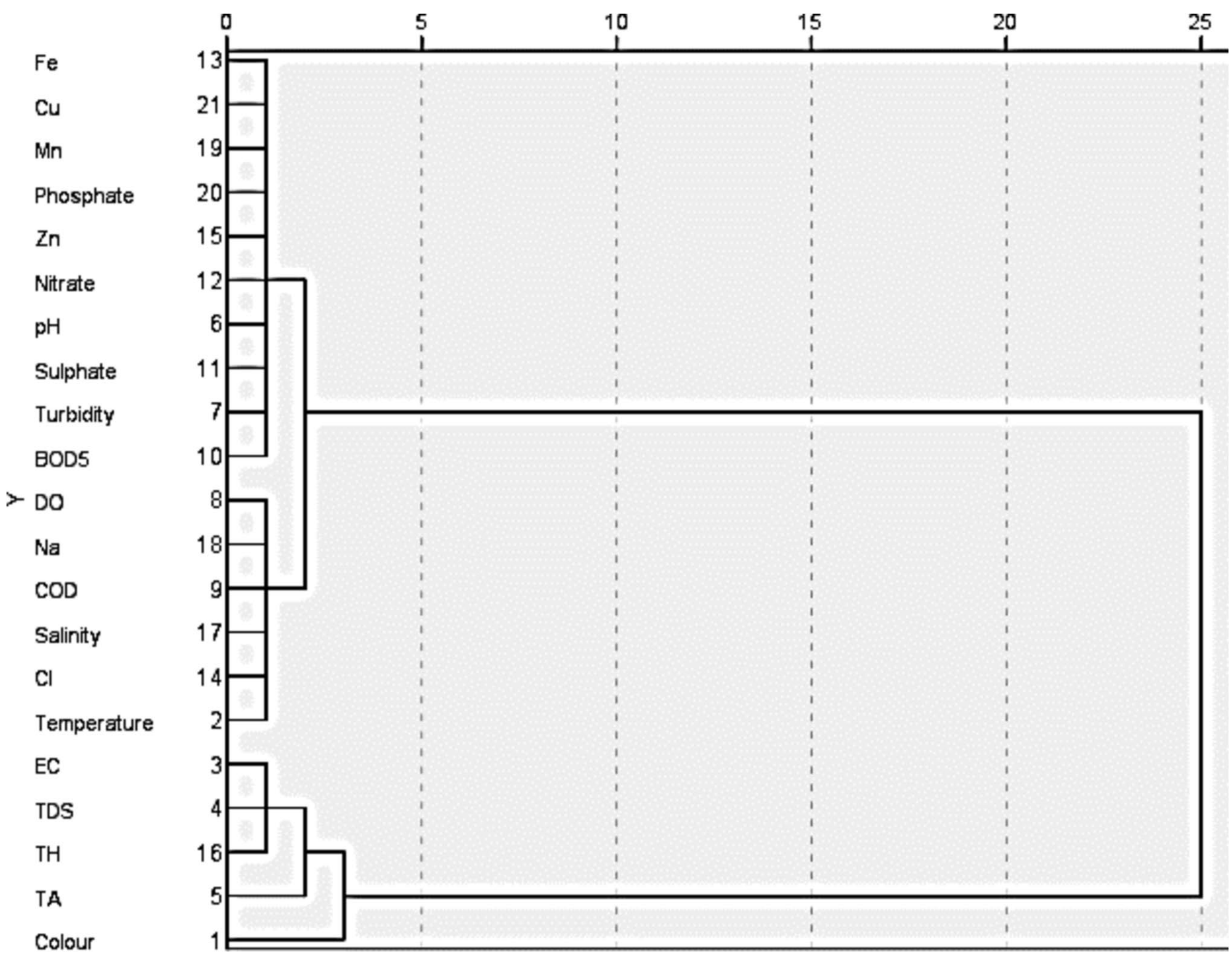


Fig. 2 Dendrogram based on agglomerative hierarchical clustering of the parameters during the wet season

3.1 Cluster Analysis

Cluster analysis (CA) was applied to detect similar groups among the parameters for both seasons. Figures 2, 3, 4 and 5 shows the representation of the cluster analysis for the parameters and the sampling points for both seasons. From the result of the study done during the wet season, cluster 1 was formed by five parameters: electrical conductivity (EC), total dissolved solids (TDS), Total hardness (TH), total alkalinity (TA) and colour, while cluster 2 was ascribed the remaining parameters with two subgroups containing six and ten parameters, respectively. The first subgroup has dissolved oxygen (DO), sodium (Na), chemical oxygen demand (COD), salinity, chlorine (Cl^-) and temperature. The second subgroup contains iron (Fe), copper (Cu), manganese (Mn), phosphate, zinc (Zn), nitrate, pH, sulphate, turbidity and biochemical oxygen demand (BOD5). The classifications were statistically significant because parameters within the same group had similar natural and anthropogenic backgrounds.

The parameters in cluster one contained pollutants believed to have come from water-ions, suspended particles, and crustal materials. In contrast, most of the parameters possessed pollutants with anthropogenic or natural origin, depending on the subgroup. Cluster 1 shows the concentration of EC, TDS, TH, TA, and colour within the WHO recommended value for drinking water except for colour. The interrelated association among colour—EC shows similar positive loadings in PC1, while EC–TDS and EC–TA show similar and negative loadings in PC2 and PC3, respectively. There is the presence of tannins chlorophyll in the leaves present in the water- ions that can impact colour; also, anthropogenic activities are carried out, landfills and pipelines, reduction in water clarity, which could contribute to reduced photosynthetic activities and possibly increase water temperature (Matouke and Abdullahi 2020). Figure 2 shows the dendrogram with two statistically different clusters at (Dlink/Dmax) × 100 < 25, consisting of different subgroups. PCA analysis grouped

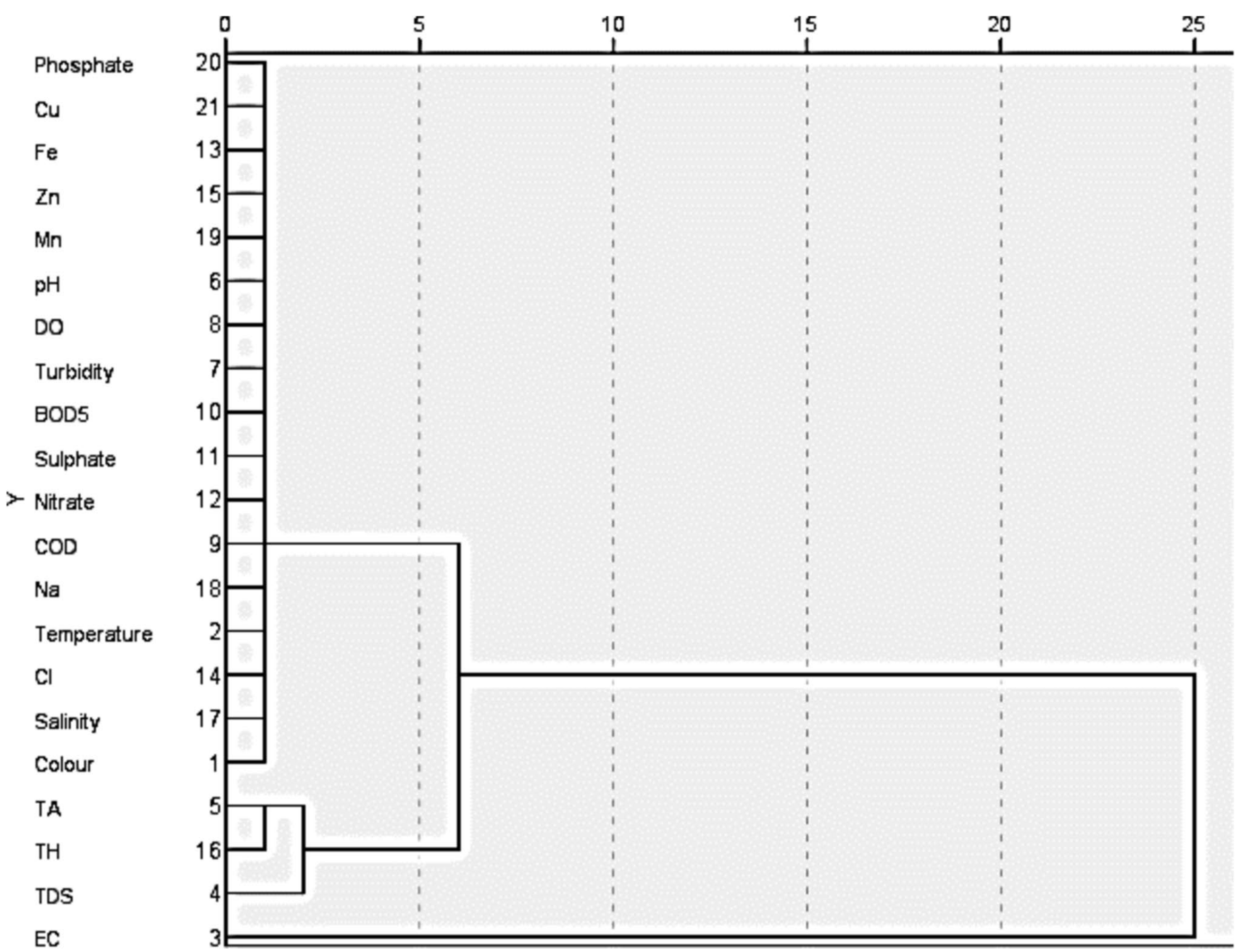
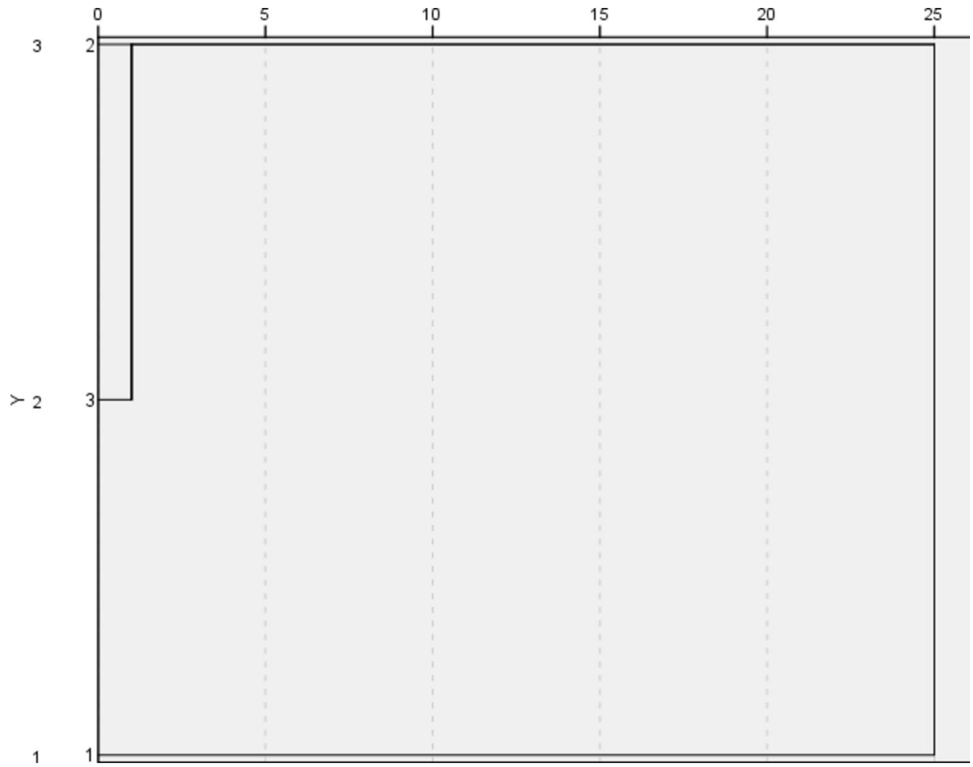


Fig. 3 Dendrogram based on agglomerative hierarchical clustering of the parameters during the dry season

Fig. 4 Dendrogram based on agglomerative hierarchical clustering of the three sampling points



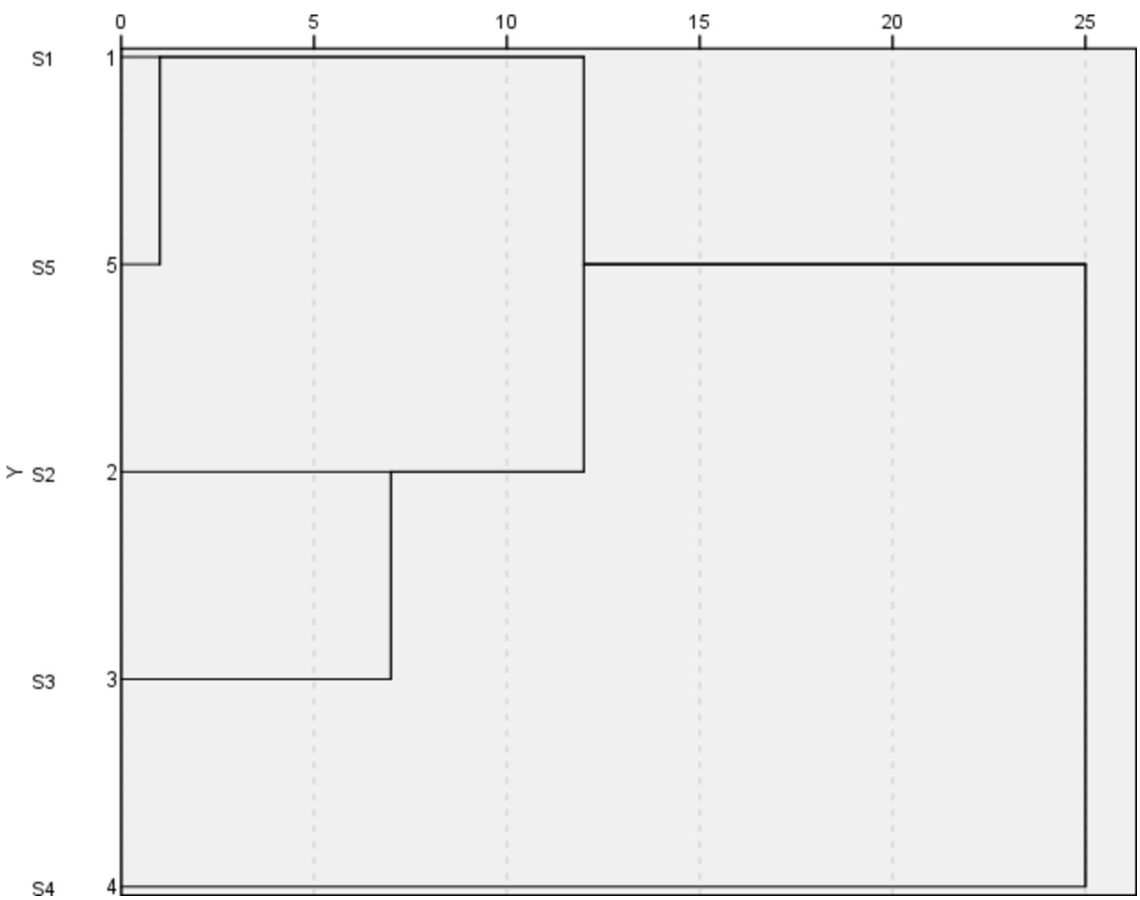


Fig. 5 Dendrogram based on agglomerative hierarchical clustering of the five sampling points

the chemical constituents into three (3) main cluster components.

Parameters in the first subgroup include DO, Na, COD, salinity, Cl^{-}, and temperature. The interrelated association between DO-COD and DO-Na shows similar positive loadings in PC1. The concentrations of these parameters indicate the dissolution of rocks surrounding the aquiver, which is probably due to sewage leakage (poor sewage system) and anthropogenic pollution. The occurrence can be harmful to aquatic life, troublesome for irrigation, corrode concrete used for construction purposes and make water unfit for drinking or livestock watering.

The parameters in the second subgroup contain Fe, Cu, Mn, Phosphate, Zn, nitrate, pH, sulphate, turbidity and BOD5. The interrelated association among Fe-Cu, Fe-phosphate, phosphate-nitrate, and turbidity-pH shows similar positive loadings in PC2 and PC4 with a negative loading in PC1. The presence of these parameters signifies the leaching of domestic effluent, weathering of rocks, agricultural runoffs, refuse dumps or contamination with human or animal wastes. Clusters that are closed together are similar, and far ones have more variability; for example, using Fig. 2, iron is far from colour, which means iron has a wide variation with colour but closer and less variability from Cu and Mn, moreover DO, and COD are more connected with lower variability. Likewise, colour, EC, TDS, TA, and TH are more connected and correlated with low variability.

Moreover, the analysis during the dry season stated that cluster 1 was formed by one parameter, EC, while cluster 2 was ascribed to the remaining parameters, with two subgroups containing three and 17 parameters, respectively. The first subgroup has TA, TH and TDS. The second subgroup contains phosphate, Cu, Fe, Zn, Mn, pH, DO, turbidity, BOD5, sulphate, nitrate, COD, Na, temperature, Cl^{-}, salinity and colour. The parameter in cluster 1 contained pollutants believed to have come from water- ions, suspended particles, and crustal materials, whereas most of the parameter possessed pollutants with anthropogenic or natural origin, depending on the subgroup.

Parameters in the first subgroup include TA, TH and TDS. The interrelated association among TA-TDS shows similar positive loadings in PC1. The concentrations of these parameters indicate the dissolution of rocks surrounding the aquiver, which is probably due to sewage leakage (poor sewage system) and anthropogenic pollution. This concentrated pollution can be harmful to aquatic life, troublesome for irrigation, corrode concrete used for construction purposes and make water unfit for drinking or livestock watering.

The parameters in the second subgroup contain phosphate, Cu, Fe, Zn, Mn, pH, DO, turbidity, BOD5, sulphate, nitrate, COD, Na, temperature, Cl^{-}, salinity and colour. The interrelated association among Fe-temperature, Fe-turbidity, nitrate-COD, nitrate- BOD_5, BOD_5- DO, BOD_5—COD, Cl- DO, Cl- COD, Cl- BOD_5 and Cl- sulphate shows similar positive loadings in PC2, PC3 and PC4 with a negative loading in PC1. The presence of these parameters signifies the leaching of domestic effluent, weathering of rocks, agricultural runoffs, refuse dumps or contamination with human or animal wastes.

3.2 Pearson's Product-Moment Correlation

However, the correlation matrix (Tables 3, 4) also further shows a distribution pattern of linear and non- linear relationships among the studied variables during the dry and wet season investigation period. The correlation results are as follows: it is fascinating that COD has a linear relationship with DO and BOD_5 in the wet season and a non-linear relationship with BOD_5 in the dry season. This strong correlation indicated that the lake was polluted with oxidisable organic and inorganic pollutants, which identified an increase in biological activities that deplete DO in water. During the wet season, the correlation results revealed that although sewage water is polluting the river, precipitation dilutes the effect of sewage water on biological activities. Turbidity has a relationship with colour for both seasons. EC has a linear relationship with TA, DO, COD, BOD_5 and phosphate for the dry season. At the same time, EC has a linear relationship with Cl^{-} for the wet season.

Geological deposits may explain the strong correlation between nitrate and phosphate with Total hardness (TH), natural organic matter decomposition and deep percolation of nitrates from fertiliser applications. There is a high correlation between phosphate and nitrate (0.805) for the wet season; this may also be explained by the runoff of chemicals used in agricultural fields, which contributes significantly to the number of nutrients on the water body's surface. Correlation analyses clearly show that colour positively correlates to turbidity and Cl^{-} with a positive loading of 0.937 and 0.810, respectively. These correlations show that the primary mechanism for releasing heavy metals into the lake is the cause of chemical oxygen demand (COD). Cu has a non-linear relationship with colour, TDS, TA and Fe with the following numbers for the dry season: − 0.889, − 0.890, − 0.808, and − 0.824. The correlations of these parameters only show anthropogenic activities near the lake (Chitrakar et al. 2020). Gani et al. (2023) suggested that industrial and domestic waste along the lake mainly contributes to river metal pollution.

Several factors affect total dissolved solids (TDS) concentrations in surface water. The presence of TDS includes water discharge and weather, industrial activities, domestic and agricultural runoff (Abdel-Fattah et al. 2020). Fe shows a linear relationship with turbidity, colour, and temperature for the dry season. Nitrate shows a linear relationship with COD and BOD_5. Phosphate has a linear relationship with

Table 3 Correlation matrix showing the relationship between the parameters during the wet season

Correlation Matrix[a]

	Colour	Temperatur e	EC Conductivity	TDS	TA	pH	Turbidity	DO	COD	BOD5	Sulphate	Nitrate	Iron	Chlorine ion	Zinc	TH	Salinity	Sodium	Manganese	Phosphate	Copper
Correlation																					
Colour	1.000																				
Temperature	− 0.275	1.000																			
Electrical Conductivity	0.711	− 0.784	1.000																		
TDS	− 0.046	0.824	− 0.699	1.000																	
Total Alkalinity	− 0.689	0.363	− 0.712	0.378	1.000																
Ph	− 0.780	0.584	− 0.770	0.297	0.489	1.000															
Turbidity	**0.937**	− 0.397	0.735	− 0.098	− 0.639	− 0.703	1.000														
DO	0.266	− 0.514	0.538	− 0.378	− 0.035	− 0.721	0.238	1.000													
COD	0.547	− 0.696	0.737	− 0.508	− 0.353	**− 0.918**	0.449	**0.808**	1.000												
BOD5	0.521	− 0.695	0.714	− 0.507	− 0.287	**− 0.894**	0.434	**0.816**	**0.995**	1.000											
Sulphate	0.609	− 0.226	0.387	0.170	− 0.183	− 0.446	0.783	0.137	0.145	0.142	1.000										
Nitrate	0.032	0.703	− 0.327	0.609	− 0.089	0.121	− 0.116	− 0.061	− 0.286	− 0.338	− 0.124	1.000									
Iron	0.492	0.532	− 0.081	0.634	− 0.052	− 0.289	0.347	0.207	0.073	0.070	0.237	0.666	1.000								
Chlorine ion	**0.810**	− 0.650	**0.885**	− 0.502	− 0.638	− 0.649	**0.875**	0.247	0.536	0.533	0.576	− 0.453	− 0.012	1.000							
Zinc	0.376	0.123	0.054	0.066	− 0.419	0.053	0.302	− 0.527	− 0.088	− 0.075	− 0.072	− 0.156	0.099	0.328	1.000						
Total Hardness	0.560	0.490	0.013	0.492	− 0.405	− 0.292	0.325	− 0.018	0.096	0.055	0.002	0.727	**0.840**	0.038	0.358	1.000					
Salinity	0.383	0.020	0.364	− 0.156	− 0.470	− 0.528	0.104	0.464	0.539	0.485	− 0.271	0.496	0.418	0.072	− 0.053	0.639	1.000				
Sodium	0.370	− 0.587	0.658	− 0.413	− 0.308	− 0.765	0.342	**0.825**	0.748	0.705	0.240	0.052	0.071	0.342	− 0.600	0.046	0.535	1.000			
Manganese	0.550	− 0.475	0.624	− 0.420	− 0.574	− 0.214	0.684	− 0.137	0.153	0.167	0.368	− 0.461	− 0.161	**0.821**	0.609	− 0.057	− 0.229	− 0.109	1.000		
Phosphate	0.497	0.503	− 0.022	0.524	− 0.352	− 0.296	0.269	0.041	0.105	0.058	0.007	**0.805**	0.835	− 0.052	0.234	**0.978**	0.670	0.115	− 0.153	1.000	
Copper	0.085	**0.831**	− 0.482	0.763	0.220	0.247	− 0.036	− 0.188	− 0.381	− 0.363	− 0.068	0.680	0.840	− 0.333	0.228	0.696	0.140	− 0.394	− 0.205	0.683	1.000

The bold values explain strong correlations between parameters and correlations that are statistically significant

[a]This matrix is not positive definite

Table 4 Correlation matrix showing the relationship between the parameters during the dry season

Correlation Matrix[a]

	Colour	Temperature	EC	TDS	TA	pH	Turbidity	DO	COD	BOD5	Sulphate	Nitrate	Iron	Chloride ion	Zinc	TH	Salinity	Sodium	Manganese	Phosphate	Copper
Correlation																					
Colour	1.000																				
Temperature	0.704	1.000																			
EC	0.531	− 0.207	1.000																		
TDS	0.788	0.354	0.747	1.000																	
TA	0.528	− 0.186	**0.988**	**0.811**	1.000																
Ph	−0.760	− **0.995**	0.119	− 0.443	0.092	1.000															
Turbidity	**0.936**	0.626	0.578	**0.930**	0.627	− 0.700	1.000														
DO	0.330	− 0.425	**0.969**	0.624	**0.962**	0.339	0.409	1.000													
COD	− 0.340	0.419	− **0.951**	− 0.506	− **0.904**	− 0.347	− 0.328	− **0.966**	1.000												
BOD5	− 0.391	0.372	− **0.961**	− 0.540	− **0.915**	− 0.298	− 0.378	− **0.967**	**0.988**	1.000											
Sulphate	− 0.342	− 0.733	0.489	0.272	0.559	0.675	− 0.072	0.644	− 0.488	− 0.460	1.000										
Nitrate	− 0.269	0.371	− 0.697	− 0.134	− 0.592	− 0.335	− 0.103	− 0.717	**0.854**	**0.855**	−0.106	1.000									
Iron	**0.981**	**0.806**	0.369	0.708	0.373	− **0.851**	**0.912**	0.164	− 0.169	− 0.224	− 0.461	− 0.149	1.000								
Chloride ion	0.035	− 0.643	0.841	0.476	**0.859**	0.564	0.192	**0.942**	− **0.861**	− **0.848**	**0.855**	− 0.565	− 0.125	1.000							
Zinc	0.332	0.246	0.134	− 0.096	0.000	− 0.281	0.014	− 0.007	− 0.227	− 0.226	− 0.583	− 0.458	0.278	− 0.258	1.000						
TH	− 0.442	− 0.474	− 0.005	0.070	0.115	0.441	− 0.120	0.171	0.052	0.063	0.760	0.317	− 0.433	0.448	− **0.965**	1.000					
Salinity	− 0.583	− 0.462	− 0.329	− **0.828**	− 0.452	0.533	− **0.889**	− 0.248	0.043	0.83	− 0.229	− 0.308	− 0.599	− 0.212	0.529	0.409	1.000				
Sodium	− 0.392	0.062	− 0.716	− **0.866**	− **0.815**	0.032	−0.660	− 0.709	0.518	0.532	− 0.694	0.049	− 0.296	− 0.728	0.480	− 0.523	**0.803**	1.000			
Manganese	− 0.591	0.220	− 0.431	− 0.140	− 0.307	0.231	− 0.302	− 0.317	0.515	0.538	0.489	0.790	− 0.541	− 0.023	− 0.786	0.775	− 268	− 0.234	1.000		
Phosphate	0.443	− 0.228	**0.927**	0.593	**0.877**	0.162	0.409	**0.873**	− **0.923**	− **0.920**	0.373	− 0.710	0.260	0.722	0.425	− 0.261	− 0.059	− 0.504	− 515	1.000	
Copper	− **0.889**	− 0.352	− 0.790	− **0.890**	− **0.808**	0.437	− 0.922	− 0.672	0.625	0.669	− 0.072	0.454	− **0.824**	− 0.454	− 0.057	0.094	0.671	0.676	0.484	− 0.602	1.000

The bold values explain strong correlations between parameters and correlations that are statistically significant

[a]This matrix is not positive definite

EC, TA, and DO, and a non-linear relationship with COD and BOD_5, while salinity has a linear relationship with TDS, Na and turbidity. During the wet season, phosphate has a relationship with nitrate and TH. Cu has a relationship with temperature. Generally, Table 5 shows high positive correlations ($p < 0.01$) with correlation coefficients varying from 0.959 to 1.000 observed for S1, S2 and S3 for the wet season. This shows that the parameters in S1 correlate well with the parameters in S2 and S3, with correlation coefficients varying from 0.959 to 1.000. However, Table 6 shows high positive correlations ($p < 0.01$) with correlation coefficients varying from 0.974 to 1.000 were observed for S1, S2, S3, S4 and S5 for the dry season. Luo et al. (2020) suggested that elements with high correlation coefficients in the water body could have similar hydrological characteristics.

3.3 Principal Component Analysis and Factorial Analysis

The correlation matrix in Tables 3 and 4 was used to identify the inter-relationship between the parameters, while the correlation matrix in Tables 5 and 6 was used for the sampling points. The classification method adopted by Markad et al. (2021), Liu et al. (2023) was used as follows: $r < 0.3$ was considered of no relevance; $0.3 < r < 0.5$ as less relevant/weak; $0.5 < r < 0.75$ as median relevance/moderate; and $r >$ 0.75 as of high relevance /strong (Matouke and Abdullahi 2020).

Before the principal component analysis (PCA), the Kaiser–Meyer–Olkin test (KMO) and Bartlett's test were performed on the data set with a scree plot to examine the validity of the principal component analysis (PCA). KMO is a measure of sampling adequacy that indicates the proportion variance (Fadel et al. 2021). A score near 1.00 shows that the data are suitable for principal component analysis (PCA), while a value below 0.05 shows that the data may not be suitable for PCA (Hammoumi, et al. 2024). KMO values below 0.5 indicate that the factorial analysis (FA)/principal component analysis (PCA) will not be applicable, whereas values ranging from 0.5 to 0.7 are considered sufficient, and higher values (above 0.7) are excellent. The result of principal component analysis (PCA) in Tables 7 and 8 shows the Kaiser–Meyer–Olkin (KMO) result for the wet season as 0.583 and 0.623 for the dry season for sampling adequacy test, which suggests a substantial correlation in the

Table 5 Pearson's correlation matrix of the sampling points during the wet season

Correlation Matrix[a]		S1	S2	S3
Correlation	S1	1.000		
	S2	**0.976**	1.000	
	S3	**0.959**	**0.996**	1.000
Sig. (1-tailed)	S1		0.000	0.000
	S2	0.0000		0.000
	S3	0.0000	0.000	

[a]The *P*-value indicates a statistically significant result ($p < 0.01$) thereby explaining the dataset suitability for factor analysis in bold

Table 7 Result of the KMO and Bartlett's test during the wet season

KMO and Bartlett's Test		
Kaiser–Meyer–Olkin Measure of Sampling Adequacy		**0.583**
Bartlett's Test of Sphericity	Approx. Chi-Square	152.848
	Df	3
	Sig	0.000

KMO value shows that the dataset is suitable for factor analysis in bold

Table 6 Pearson's correlation matrix of the sampling points during the dry season

		S1	S2	S3	S4	S5
Correlation Matrix[a,b]						
Correlation	S1	1.000				
	S2	0.989	1.000			
	S3	0.985	0.995	1.000		
	S4	0.989	0.974	0.986	1.000	
	S5	0.997	0.996	0.988	0.978	1.000
Sig. (1-tailed)	S1					
	S2	0.000				
	S3	0.000	0.000			
	S4	0.000	0.000	0.000		
	S5	0.000	0.000	0.000	0.000	

[a]Determinant = 1.286E−10 shows high multicollinearity meaning that the variables are highly correlated with each other

[b]($p < 0.01$) *p*-value indicates a statistically significant result

Table 8 Result of the KMO and Bartlett's test during the dry season

KMO and Bartlett's Test		
Kaiser–Meyer–Olkin Measure of Sampling Adequacy		**0.623**
Bartlett's Test of Sphericity	Approx. Chi-Square	421.322
	Df	10
	Sig	0.000

KMO value shows that the dataset is suitable for factor analysis in bold

data (statistically significant). The eigenvalue, percentage of variance, cumulative percentage of variance, the communalities and the factor loadings are presented in Tables 9 and 10 below. In doing this, only independent factors with eigenvalues greater than one were extracted. Figures 10 and 11 show the use of varimax rotation for the derivation of factors.

According to the results of the principal component analysis/factorial analysis (PCA/FA) shown in Tables 7 and 8, the principal component analysis (PCA) used varimax rotation (Abdel-Fattah et al. 2020). Eigenvalue selection criteria were selected to explain the sources of variance one and greater than one. The screen plots of eigenvalues are as shown in Figs. 6, 7, 8 and 9. Figures 8 and 9 show that four of the eigenvalues have over 1. Other eigenvalues are below 1. Thus, a new set of data was obtained. These eigenvalues may explain the variation of the data set with fewer variables. The total variance is significant in principal component analysis (PCA) analysis. The principal components (PCs) are arranged according to their size. Eigenvalues belonging to the total variance explained before and after rotation are given in Table 11. Contributions to the total variance of groups with converted and unconverted Eigenvalues over one occurred at a rate of 98.469% and 91.446%, respectively, for the wet season and 99.014% and 100% for the dry season. Eigenvalues indicate the degree of importance of one factor. Consequently, the eigenvalue with the highest eigenvalue number contributes the most to the variance (Gani, et al. 2023). Considering the rotated component matrix in Figs. 10 and 11, Table 9 shows that three principal components were obtained with a cumulative variance that ranges between 98.5 and 100%. The factor loadings were also sorted according to the classification method above adopted (Hammoumi et al. 2024; Abdel-Fattah et al. 2020). The descriptive analysis goes alongside the principal component analysis. The main aim of principal component analysis (PCA) is to describe how the data correlate with each other (Chadli and Boufala 2021; Passos et al. 2021) Tables 5 and 6 show the correlation matrix of S1 to S3 and S1 to S5 to be above the standard value, 5% of the probability factor using Pearson's r correlation coefficient. Table 9 shows that component 1 has 98.5% of cumulative variance (CV%), component 2 has 99.9% of cumulative variance (CV%), and component 3 has 100% of cumulative variance (CV%), respectively. Table 9 buttressed that component 1 correlated well with S1, S2, and S3 and was characterised by a robust favourable loading of its parameters. For components 2 and 3, the S1 and

Table 9 Summary of the total variance explained for the sampling points during the wet season

Total Variance Explained						
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of variance	Cumulative%	Total	% of variance	Cumulative%
1	2.954	98.469	98.469	2.954	98.469	98.469
2	0.044	1.475	99.943	0.044	1.475	99.943
3	0.002	0.57	100.00	0.002	0.057	100.000

Extraction Method: Principal Component Analysis

Table 10 Summary of the total variance explained for the sampling points during the dry season

Total Variance Explained						
	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of variance	Cumulative%	Total	% of variance	Cumulative%
1	4.951	99.014	99.014	4.951	99.014	99.014
2	0.030	0.607	99.621	0.030	0.607	99.621
3	0.018	0.366	99.988	0.018	0.366	99.988
4	0.001	0.011	99.998	0.001	0.011	99.998
5	8.696E−5	0.002	100.000	8.696E−5	0.002	100.000

Extraction Method: Principal Component Analysis

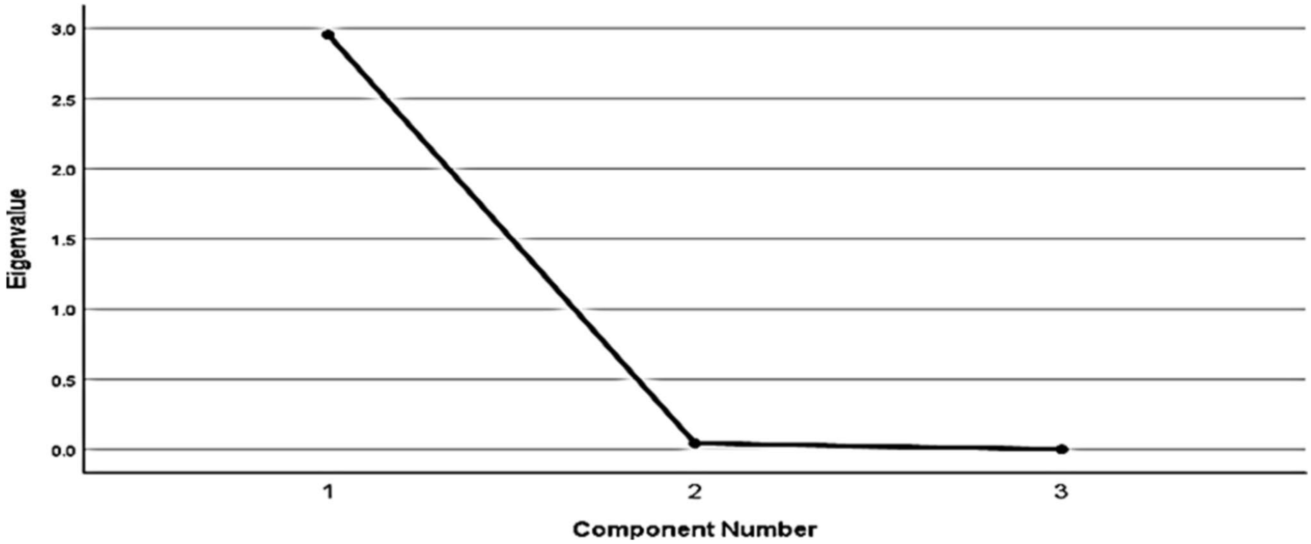


Fig. 6 Scree plot—relationship between factors (component number) and eigenvalues of the sampling points during the wet season

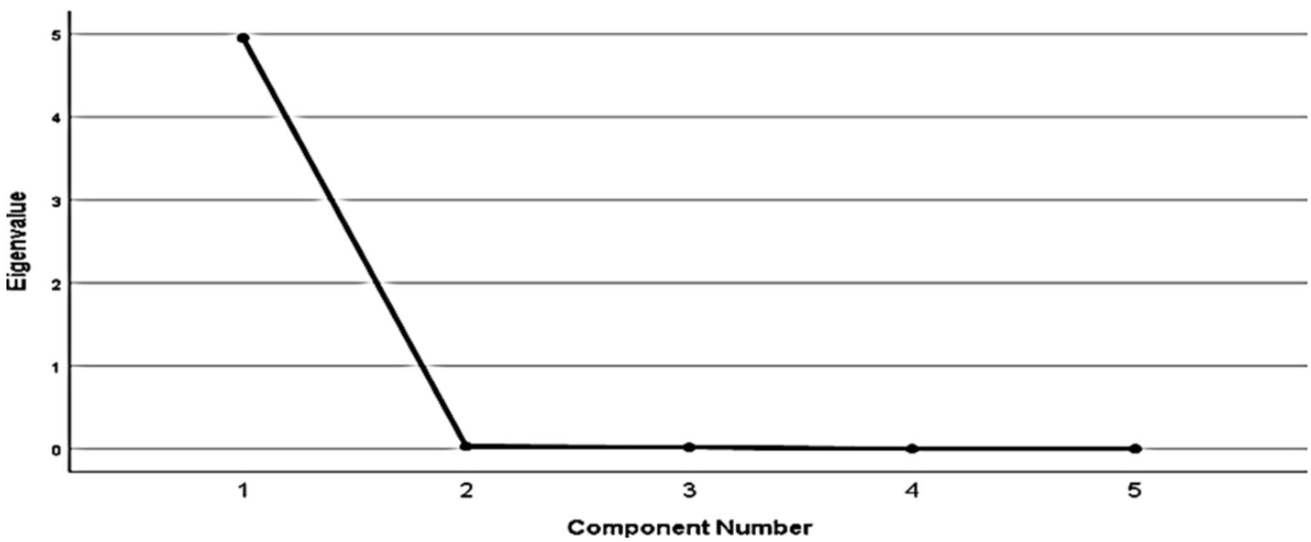


Fig. 7 Scree plot—relationship between factors (component number) and eigenvalues of the sampling points during the dry season

S3 have no negative relevance loading of their parameters. Table 10 shows component 1 has 99.014% of cumulative variance (CV%). Component 2 has 99.621% of the cumulative variance (CV)%, components 3 and 4 have 99.998% of the cumulative variance (CV)% while component 5 has 100% of the cumulative variance (CV)%.

3.4 Pollution Source Identification

Four components obtained by the principal component analysis for the parameter are presented in Table 11 for eigenvalues greater than 1, summing the totals of 100% for both seasons for the total variance in data sets. Factors are generally classified by Liu et al. (2021) above. The first principal component (PC1) in the wet season samples signifies an increase in loading of some parameters such as EC, pH, DO, BOD5, COD and Na. PC1 explains 39.824% variance and heavy positive loading on EC, DO, COD, BOD5 and Na, with a negative loading on pH. These PC1 results indicate that the lake is polluted. The pollution is a result of the human factor in the account of the application of fertilisers and wastewater discharge into the lake from agriculture, which contributes to the pollution and the erosion effect during cultivation of the soil and associated organic matter; higher DO value is as a result of increased water volume in the lake (Benateau et al. 2019).

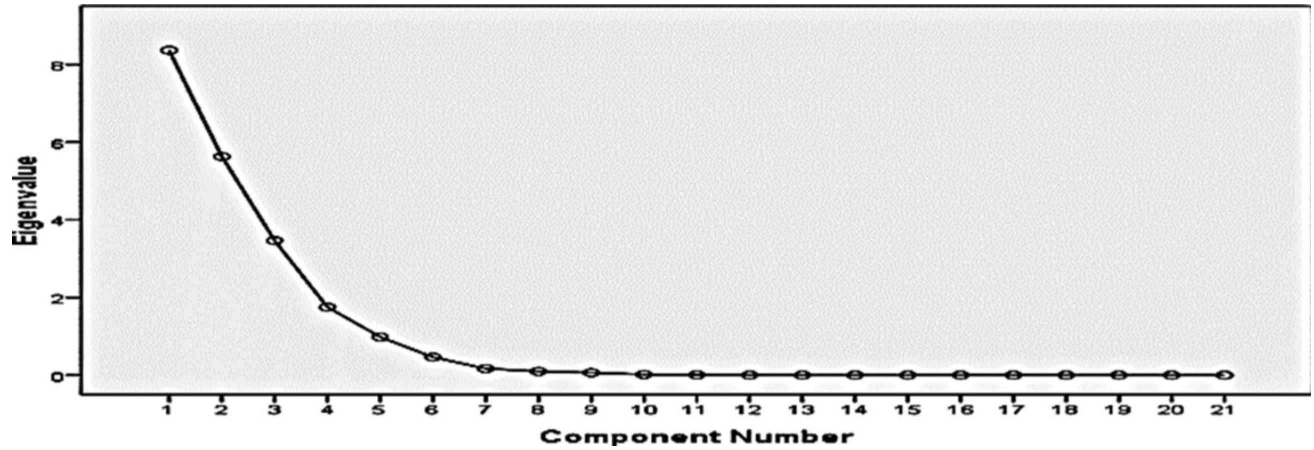


Fig. 8 Scree plot—relationship between factors (component number) and eigenvalues of the parameters during the wet season

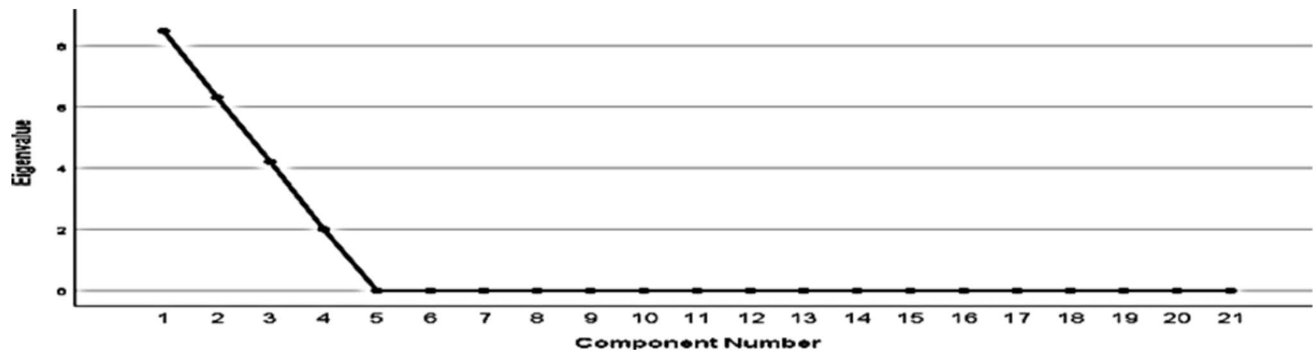


Fig. 9 Scree plot—relationship between factors (component number) and eigenvalues of the parameters during the dry season

The strong loading of BOD5 and COD explained that since they measure oxygen demand by biodegradable and non-biodegradable pollutants, the high value obtained suggests that a large amount of the product was lost to the lake, indicating that a large amount of the product lost might cause an increase in biological activities in the lake. The presence of water hyacinth shows the presence of faeces in the water body reduction in COD. PC1 for dry season samples explained a 40.376% variance and had a strong factor loading of EC, TDS, DO, Cl^{-}, TH and Cu, with a negative loading of COD and BOD5. These parameters are reactive components of partial anthropogenic activities. Conductivity reflects the status of inorganic pollution and a measure of TDS in water.

For the second principal component, (PC2), high concentrations of Temperature, TDS, Nitrate ($NO^{-}3$), Fe, TH, phosphate and Cu in the Lake may originate from some sources: geological deposits, natural organic matter decomposition and deep percolation of nitrate resulting from fertiliser applications (Dash and Kalamdhad 2021). The strong loading of those parameters explained a 26.803% variance, which implies titration from rocks and soil infected through weathering. Though Fe does not have many health implications when minimal, it only stains laundry (brownish).

However, when it is on the high side, it might have health implications and block the kidney, causing kidney problems. Nitrate concentrations in surface water like Jabi Lake may originate from geological deposits, natural organic matter decomposition and deep percolation of nitrate. A heavily positively loaded TDS indicates the field dust settled in the water body. PC2 for dry season samples explained 30.036% of the total variance and had a loading of colour, temperature, turbidity, Fe, salinity, and pH. This factor represents the erosion effect during soil cultivation and associated organic matter. Principal component 3 (PC3) for wet season samples had a high positive loading of colour, Cl^{-}, Zn, and Mn with a negative loading of TA. PC3 explained 16.48% of the total variance. This factor is due to local anthropogenic activities such as agricultural and domestic waste. PC3 for the dry season sample had loaded on nitrate and phosphate and explained 20.03% of the total variance. This factor's higher value of nutrients could have been due to surface runoff from the surrounding farmlands, which might have brought ionic substances such as NO_3, Cl and PO^{3-} from fertiliser.

Table 11 Communalities and principal component analysis vector of coefficients for first four principal components (PCs) with eigenvalues greater than one (> 1.0) for Jabi Lake water quality data for both dry and wet season. *Source*: PCA results from IBM SPSS Statistics

Parameter	Component for dry				Component for wet				Communalities		
	1	2	3	4	1	2	3	4	Initial	Extraction for dry	Extraction for wet
Colour	0.231	**0.933**	– 0.264	0.140	0.531	0.324	**0.717**	0.597	1.000	1.000	0.984
Temperature	– 0.408	**0.873**	0.207	0.293	– 0.663	**0.719**	– 0.345	– 0.250	1.000	1.000	0.967
EC	**0.903**	0.292	0.126	0.453	**0.782**	– 0.267	0.657	0.362	1.000	1.000	0.944
TDS	**0.898**	0.331	0.234	0.388	– 0.517	**0.720**	– 0.373	0.162	1.000	1.000	0.922
TA	0.527	0.586	0.562	0.478	– 0.412	– 0.100	**– 0.848**	– 0.094	1.000	1.000	0.797
Ph	0.315	**– 0.901**	– 0.245	– 0.311	**– 0.907**	– 0.088	– 0.389	– 0.399	1.000	1.000	0.946
Turbidity	0.365	**0.910**	0.074	0.059	0.452	0.120	0.673	**0.788**	1.000	1.000	0.992
DO	**0.986**	0.049	– 0.136	0.093	**0.878**	– 0.037	– 0.138	0.121	1.000	1.000	0.887
COD	**– 0.935**	– 0.030	0.288	– 0.258	**0.936**	– 0.115	0.309	0.137	1.000	1.000	0.894
BOD5	**– 0.911**	– 0.067	0.380	– 0.162	**0.910**	– 0.145	0.286	0.157	1.000	1.000	0.848
Sulphate	0.689	– 0.306	0.674	0.175	0.198	0.008	0.164	**0.956**	1.000	1.000	0.924
Nitrate	– 0.231	0.140	**0.913**	0.415	– 0.097	**0.867**	– 0.202	– 0.255	1.000	1.000	0.823
Iron	– 0.003	**0.838**	– 0.482	– 0.202	0.117	**0.909**	0.021	0.246	1.000	1.000	0.928
Chloride ion	**0.962**	– 0.218	0.008	– 0.082	0.502	– 0.247	**0.770**	0.621	1.000	1.000	0.955
Zinc	0.095	0.232	0.304	**0.984**	– 0.322	0.141	**0.759**	0.039	1.000	1.000	0.840
TH	**0.955**	0.042	0.196	0.381	0.122	**0.923**	0.338	– 0.037	1.000	1.000	0.985
Salinity	– 0.057	0.941	0.239	0.397	0.632	0.512	0.237	– 0.402	1.000	1.000	0.948
Sodium	– 0.211	0.582	– 0.697	– 0.412	**0.918**	– 0.046	– 0.020	0.133	1.000	1.000	0.894
Manganese	0.361	0.033	– 0.268	**0.873**	0.059	– 0.314	**0.829**	0.480	1.000	1.000	0.904
Phosphate	– 0.219	0.012	**0.933**	– 0.149	0.156	**0.943**	0.228	– 0.066	1.000	1.000	0.975
Copper	**– 0.837**	0.479	– 0.079	0.224	– 0.377	**0.851**	– 0.123	0.000	1.000	1.000	0.847
Eigenvalue (˃1.0)	8.48	6.31	4.21	2.01	8.36	5.63	3.46	1.75			
% of total variance	40.38	30.04	20.02	9.56	39.82	26.80	16.49	8.33			
% of cumulative	40.38	70.41	90.44	100.00	39.82	66.63	83.12	91.45			

Values in bold correspond to the absolute loading value > 0.70

Four PCs were extracted using PCA as the extraction method. Rotation method: Oblimin with Kaiser Normalization and rotation converged in several iterations

Principal component factor 4 (PC4) in the wet season samples explained a strong loading on turbidity and sulphate with an 8.331% variance. These concentrations represent the contribution of nonpoint pollution from agriculture and soil erosion processes, even though the natural input of particulates to the river through erosion and sediment transport. In this area, especially in the north, farmers use ammonium sulphate fertilisers, and the surface runoff receives sulphate via surface runoff and irrigation waters (Abdel-Fattah et al. 2020). PC4 in the dry season samples explained 9.564% of the total variance and loading on Zn and Mn. This factor is due to local anthropogenic activities such as agricultural and domestic waste. The primary pollution sources of Jabi Lake were urban, agricultural, industrial, and domestic wastewater. The results indicated that pollution sources differed significantly among the sampling sites. From the above discussion, PCA/FA proved to be a reliable tool for distinguishing sources of pollution among the parameters. This technique could be used to inform policies of pollution control. It could strengthen government initiatives to improve the quality of drinking water sources.

3.5 Discriminant Analysis

The equality of group means tests using Wilks' Lambda method of analysis with factorial shown in Table 12 that the eigenvalue of 0.333, Wilks' Lambda result of 0.75, % variance and cumulative variance% of 100% at a significant of 0.17, df of 3, chi-square of 5.031, with 81% classification check respectively shows that the factors are excellently correlated. The canonical result of 0.005 indicates a median relevance with the group function. Table 13 shows that 81% of original grouped cases are correctly classified; cross-validation is done only for those in the analysis for both seasons.

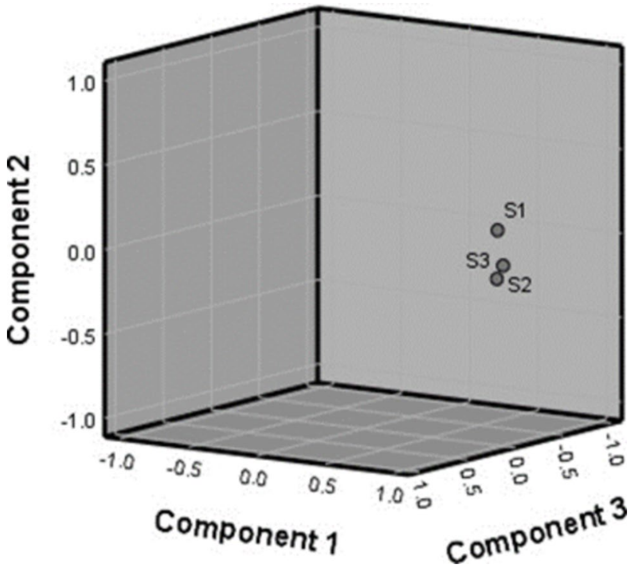


Fig. 10 3 Component plot extracted from the Sampling points during the wet season

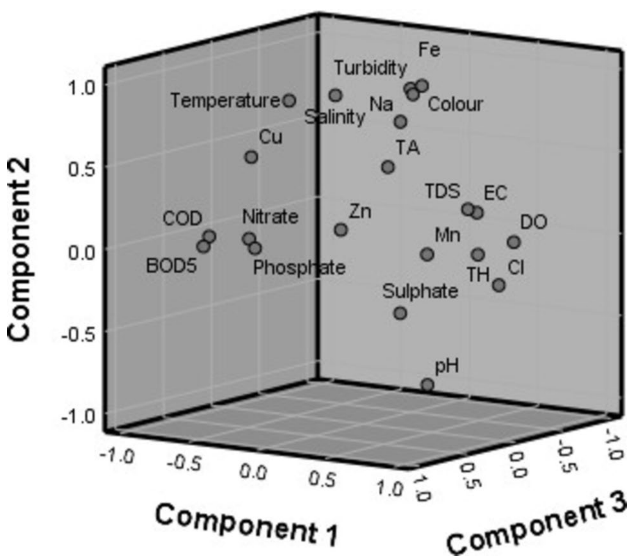


Fig. 11 3 Component plot in rotated space extracted from the Parameters during the dry season

In cross-validation, each case is classified by the functions derived from all cases other than that case. Moreover, 81.0% of cross-validated grouped cases were correctly classified.

3.6 Box Plot Analysis

Box plot analysis was also used to investigate water quality and the effect of pollution on river water in the Abuja metropolis using Jabi Lake as a case study. The box plot analysis of the three and five sampling points are shown in Figs. 12 and 13, respectively.

Figures 14 and 15 below classified the parameters into three factors for both seasons. For the wet season samples, the first factor increases the loading of some parameters, which entails total alkalinity (TA), colour, EC, TDS, and TH. The second factor entails DO, COD, Cl-, Na, temperature and salinity, while the third factor consists of pH, turbidity, BOD_5, sulphate, nitrate, Zn, Cu, Mn, Fe and phosphate.

F1 contains a substantial load of EC, TDS, TA and TH. A heavily positively loaded TDS indicates the field dust settled in the water body. TA and TH are natural processes of dissolution of soil components. High TDS influences the other qualities of water, such as taste, hardness, corrosion properties, and osmoregulation of freshwater organisms. Conventional methods do not generally remove them, and finally, they reduce the utility of water for drinking and irrigation purposes.

F2 reveals a substantial load of DO, COD, Cl-, Na, temperature and salinity. Chloride concentration due to weathering and dissolution of salt deposits, seawater intrusion and irrigation runoff—higher DO value results from increased water volume in the lake (Benateau et al. 2019). The strong loading of biochemical oxygen demand (BOD_5) and chemical oxygen demand (COD) explained that since they measure oxygen demand by biodegradable and non-biodegradable pollutants, the high value obtained suggests that a large amount of the product was lost to the lake, indicating that a large amount of the product lost might cause an increase in biological activities in the lake.

F3 contains a substantial load of pH, turbidity, BOD_5, sulphate, nitrate, Zn, Cu, Mn, Fe and phosphate. These concentrations represent the contribution of nonpoint pollution from agriculture and soil erosion processes. In this area, especially in the north, farmers use ammonium sulphate fertilisers, and the surface runoff receives sulphate via surface

Table 12 Summary of standardised canonical discriminant functions, eigenvalues, Wilks' Lambda for both seasons

Test of Function	Eigenvalue	%of variance	Cumulative %	Canonical Correlation	Wilks' Lambda	Chi-square	df	Sig
1	0.333[a]	100.0	100.0	0.500	0.750	5.031	3	0.170

This value of % Correct shows that there are no issues with the data, meaning that the analysis is on the right side. Assuming it shows a value less than 50%, that means there is a problem with the data

Table 13 Classification results

	Predicted Membership		Group	
	Group	0	1	Total
Original				
Count	0	1	4	5
	1	0	16	16
%	0	20.0	80.0	100.0
	1	0.0	100.0	100.0
Cross-validated				
Count	0	1	4	5
	1	0	16	16
%	0	20.0	80.0	100.0
	1	0.0	100.0	100.0

81% of original grouped cases are correctly classified

Cross-validation is done only for those cases in the analysis. In cross-validation, each case is classified by the functions derived from all cases other than that case

81% of cross-validated grouped cases are correctly classified

runoff and irrigation waters (Abdel-Fattah et al. 2020). The application of fertilisers and wastewater discharge into the lake from agriculture contributes to the pollution and the erosion effect during soil cultivation and associated organic matter. Intensive agricultural activities have been reported around Jabi Lake, such as (Matta et al. 2023) BOD_5, which establishes self-purification of surface waters and indicates water load with dissolved organic matter (Chadli and Boufala 2021). The maximum concentration of BOD5 in the lake water maybe because of the excess of organic matter. Their presence is due to local anthropogenic activities such as agricultural and domestic waste. Nitrate concentrations in surface water like Jabi Lake may originate from geological deposits, natural organic matter decomposition and deep percolation of nitrate. Fe contamination in water is due to weathering of rocks and industrial waste. For the dry season samples, the first factor increases the loading of one parameter: EC. The second factor entails TA, TDS and TH. In contrast, the third factor consists of colour, temperature, pH, turbidity, DO, COD, BOD_5, sulphate, nitrate, Fe, Cl-, Zn, salinity, Na, Mn, phosphate and Cu like that of the cluster analysis above.

These correlations show that PCA/FA, cluster, descriptive and box plots are good chemometric techniques for assessing water quality and the effect of pollution on Jabi Lake.

4 Conclusion and Recommendation

In conclusion, Jabi Lake water quality was investigated, employing chemometric tools to identify the significant sources of pollution and the variation of water pollution in the three and five sampling points, joining with

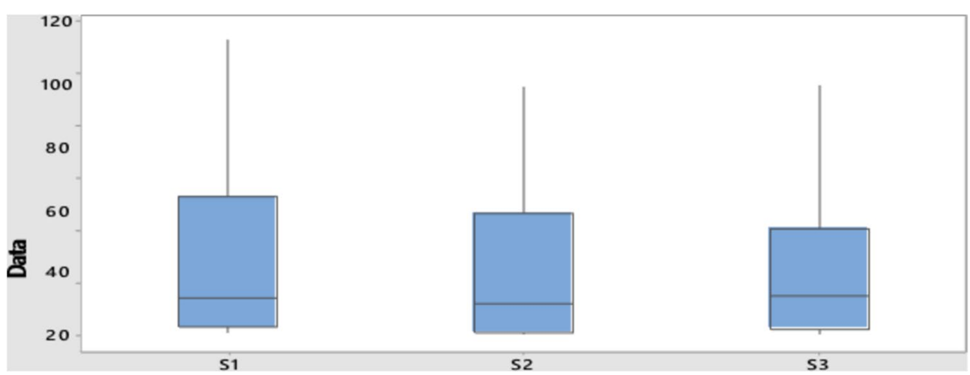


Fig. 12 Box plot analysis of the three sampling points during the wet season (S1, S2, S3)

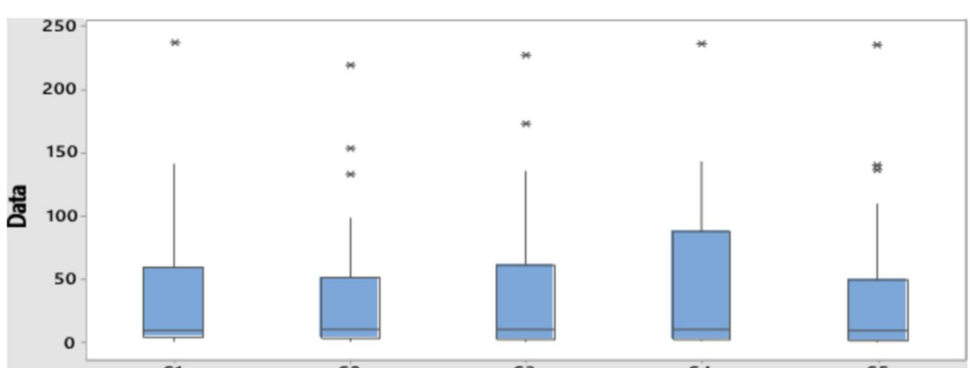


Fig. 13 Box plot analysis of the five sampling points during the dry season

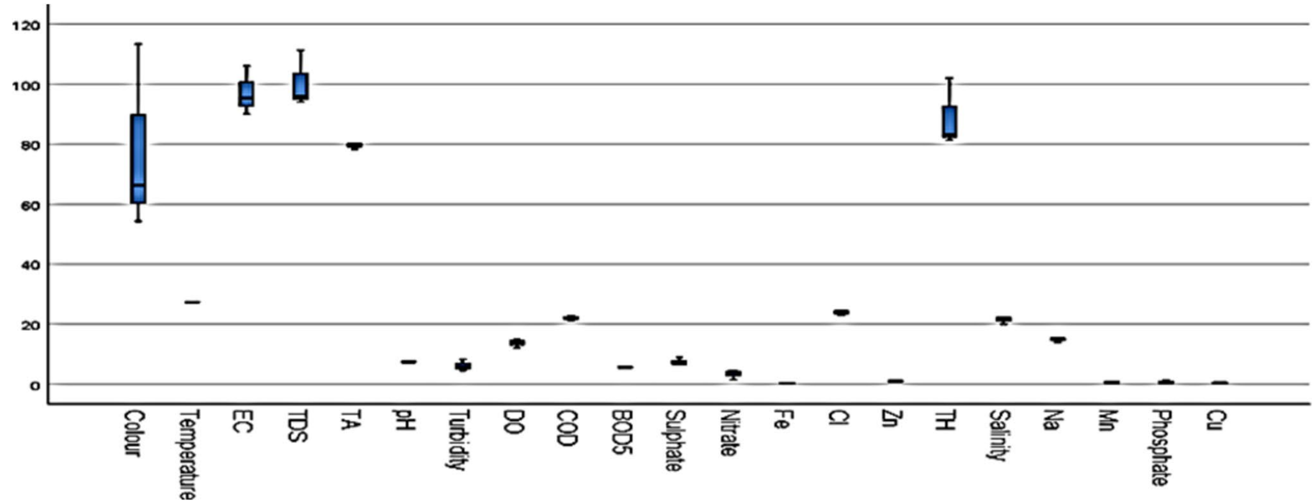


Fig. 14 Box plot analysis of the mean of the parameters during the wet season

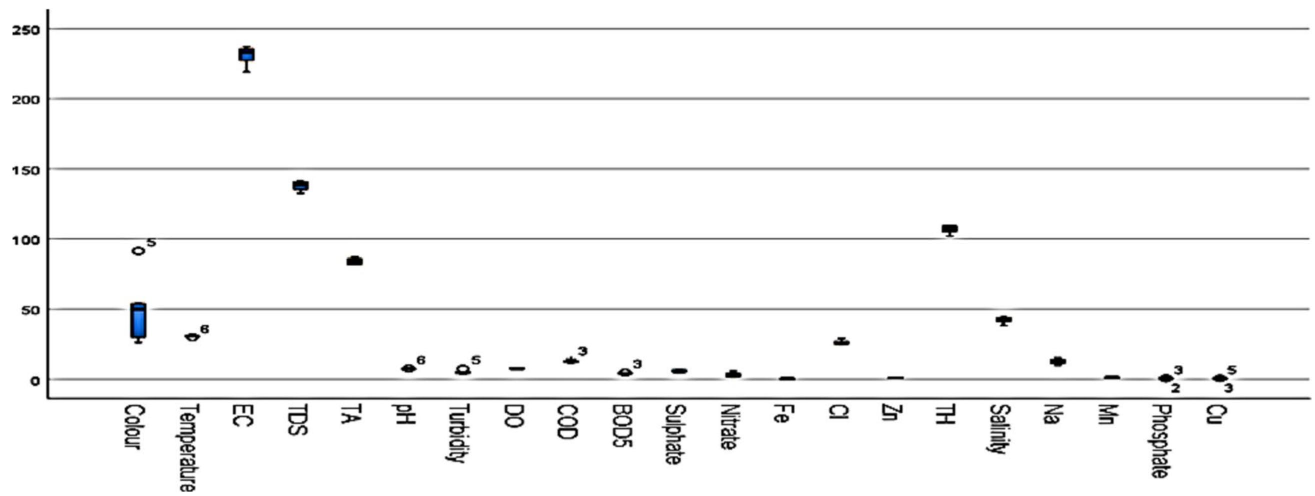


Fig. 15 Box plot analysis of the mean of the parameters during the dry season

a correlation coefficient of 21 water quality parameters. Four main factors with an eigenvalue more significant than one were retained using PCA coupled with factorial analysis (FA) on the available data, which indicated that the Jabi Lake water quality variations are mainly due to municipal/domestic waste disposal, metals, organic pollution and natural processes (runoff of chemicals used in the agricultural field, erosion). Discriminant analysis (DA) rendered a vital data reduction using sixteen parameters, affording 81% correct assignation. Furthermore, box plot analysis (BPA), cluster analysis (CA) and principal component analysis (PCA) help recognise constituents that impact water quality. These chemometric tools provided a more objective interpretation of surface water physicochemical parameters and identification of water pollution source apportionment as part of managing a sustainable river like Jabi Lake. Suitable water pollution control measures should treat industrial and domestic sewage before mixing with Jabi Lake. These measures would protect the threatened biodiversity in the water bodies. Therefore, it is required that researchers and government agencies organise educative programmes to enlighten the people on the proper use of water, be it surface water or groundwater, and the potential dangers associated with human consumption of contaminated water.

Moreover, appropriate authorities should take proactive measures to stop the discharge of pollutants into Jabi Lake. These pollutants could cause serious havoc on the domestic water supply, increase the water treatment cost, and finally destroy the potential for the aquatic organisms in the system.

The study recommends further studies on the sources and causes of pollutants in Jabi Lake.

References

Abdel-Fattah MK et al (2020) Multivariate analysis for assessing irrigation water quality: a case study of the Bahr Mouise Canal, Eastern Nile Delta. Water 12(9):2537

Ahmed M et al (2022) Impact of climate change on dryland agricultural systems: a review of current status, potentials, and further work need. Int J Plant Prod 16(3):341–363

Akintola OA et al (2024) Simple fuzzy classification of metal qualities of Ona River, Ibadan, Nigeria, and its implication for fish production and other uses. Discover Water 4(1):79

Ali MS et al. (2024) Multivariate analysis of water quality in the Dhaleshwari River, Bangladesh: identifying pollution sources and environmental implications

Apestegui A et al (2023) Trace element distribution and pollution status of surface sediments in lakes impacted by volcanic activity. J Soils Sedim 23(3):1552–1567

Balcerowska-Czerniak G, Gorczyca B (2024a) Rapid assessment of surface water quality using statistical multivariate analysis approach: Oder River system case study. Sci Total Environ 912:168754

Balcerowska-Czerniak G, Gorczyca B (2024b) Rapid assessment of surface water quality using statistical multivariate analysis approach: Oder River system case study. Sci Total Environ 912:168754. https://doi.org/10.1016/j.scitotenv.2023.168754

Banda TD, Kumarasamy M (2020) Application of multivariate statistical analysis in the development of a surrogate water quality index (WQI) for South African watersheds. Water 12(6):1584

Benateau S et al. (2019) Climate change and freshwater ecosystems: Impacts on water quality and ecological status

Bhatt S et al (2024) Characterizing seasonal, environmental and human-induced factors influencing the dynamics of Rispana River's water quality: implications for sustainable river management. Results Eng 22:102007. https://doi.org/10.1016/j.rineng.2024.102007

Bhatt S et al (2024) Characterizing seasonal, environmental and human-induced factors influencing the dynamics of Rispana River's water quality: Implications for sustainable river management. Results Eng 22:102007. https://doi.org/10.1016/j.rineng.2024.102007

Chadli K, Boufala M (2021) Assessment of water quality using Moroccan WQI and multivariate statistics in the Sebou watershed (Morocco). Arab J Geosci 14(1):27

Chen P (2024) Unlocking policy effects: Water resources management plans and urban water pollution. J Environ Manage 365:121642

Chitrakar P et al (2020) Multivariate statistical technique in the assessment of coastal water quality of Oman. J Environ Eng Sci 15(3):141–153

Daniel M et al (2023) Water quality assessment of the man-made jabi lake, federal capital territory, Abuja, Nigeria. J Appl Sci Environ Manag 27(6):1141–1146

Dash S, Kalamdhad AS (2021) Hydrochemical dynamics of water quality for irrigation use and introducing a new water quality index incorporating multivariate statistics. Environ Earth Sci 80(3):73

Dimri D et al (2023) Spatio-temporal variation of trace elements distributed over surface water of Upper Ganga River Basin in Western Himalayan Region. J Mt Sci 20(1):145–162

Edition F (2011) Guidelines for drinking-water quality. WHO Chron 38(4):104–108

Edoreh JA et al (2021) Assessment of Water Pollution Indices of Two Anthropogenic Impacted Rivers in Southern Nigeria. J Aquat Sci 35(2):125–137. https://doi.org/10.4314/jas.v35i2.15

Egbueri JC, Mgbenu CN (2020) Chemometric analysis for pollution source identification and human health risk assessment of water resources in Ojoto Province, southeast Nigeria. Appl Water Sci 10(4):98

Egun NK, Oboh IP (2023) Assessment of water quality for suitability and human health risk: a study of the Owan River, Edo State, Nigeria. Afr J Aquat Sci 48(1):19–27. https://doi.org/10.2989/16085914.2022.2156468

Elkorashey RM (2022) Utilizing chemometric techniques to evaluate water quality spatial and temporal variation. A case study: Bahr El-Baqar drain—Egypt. Environ Technol Innov 26:102332

Eze VC et al (2023) Source apportionment of polychlorinated biphenyls in surface water and sediments from River Otamiri, Imo State. Sci Afr 22:e01957

Fadel A, Kanj M, Slim K (2021) Water Quality Index variations in a Mediterranean reservoir: a multivariate statistical analysis relating it to different variables over 8 years. Environ Earth Sci 80:1–13

Famuyiwa AO et al (2023) Physicochemical quality, potentially toxic elements characterization and toxicological risk assessment of industrial effluents in Iju River, Ogun State, Nigeria. J Res for Wildlife Environ 15(3):126–135

Gani A et al. (2023) Water quality index assessment of river ganga at haridwar stretch using multivariate statistical technique. Mol Biotechnol, pp 1–24

Hammoumi D et al (2024) Seasonal variations and assessment of surface water quality using water quality index (WQI) and principal component analysis (PCA): a case study. Sustainability 16(13):5644

Hue NH, Thanh NH (2020) Assessment of surface water quality by using multivariate statistical analysis techniques: a case study of Nhue River, Vietnam. Int J Environ Sci Dev 11:488–492

Ibrahim A et al (2023) Water quality modelling using principal component analysis and artificial neural network. Mar Pollut Bull 187:114493

Ikpeze OV, Aririguzoh GC (2023) Control methods and management of water pollution in Nigeria. Unizik Law J 19(3)

Imam N et al (2023) Progress on drinking water quality monitoring in the northern part of nigeria: a catalyst to achieving sustainable development goals. Fudma J Sci 7(2):152–158

Isukuru EJ et al (2024) Nigeria's water crisis: abundant water, polluted reality. Cleaner Water 2:100026. https://doi.org/10.1016/j.clwat.2024.100026

Jahin HS, Abuzaid AS, Abdellatif AD (2020) Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate, Egypt. Environ Technol Innov 17:100532. https://doi.org/10.1016/j.eti.2019.100532

Kaanayochukwu EC et al (2019) Assessing the pollution status of Jabi Lake in the Federal Capital Territory, Abuja, Nigeria. Indonesian J Geogr 51(3):324–331
Keerthanan S, Bhattacharya P, Vithanage M (2023) Geochemical provenance of metalloids and their release: implications on medical geology. Med Geol En Route to One Health, pp 217–234
Kumar L et al (2023) Water quality assessment and monitoring in Pakistan: a comprehensive review. Sustainability 15(7):6246
Kumar M, Mohapatra S, Acharya K (2022) Contaminants of Emerging Concerns and Reigning Removal Technologies. (1st ed.). https://www.taylorfrancis.com/books/9781000551082. https://doi.org/10.1201/9781003247869
Liu L, You X (2023) Water quality assessment and contribution rates of main pollution sources in Baiyangdian Lake, northern China. Environ Impact Assess Rev 98:106965
Liu J et al (2021) Water quality assessment and source identification of the Shuangji River (China) using multivariate statistical methods. PLoS ONE 16(1):e0245525
Liu W et al (2023) A novel deep learning ensemble model based on two-stage feature selection and intelligent optimization for water quality prediction. Environ Res 224:115560. https://doi.org/10.1016/j.envres.2023.115560
Luo P et al (2020) Historical assessment and future sustainability challenges of Egyptian water resources management. J Clean Prod 263:121154
Markad AT et al (2021) A multivariate statistical approach for the evaluation of spatial and temporal dynamics of surface water quality from the small reservoir located in the drought-prone area of South-West India: a case study of Tiru reservoir (India). Environ Sci Pollut Res 28:31013–31031
Matouke MM, Abdullahi KL (2020) Assessment of heavy metals contamination and human health risk in Clarias gariepinus [Burchell, 1822] collected from Jabi Lake, Abuja, Nigeria. Sci Afr 7:e00292
Matta E et al (2023) Data integration for investigating drivers of water quality variability in the Banja reservoir watershed. Water 15(3):607
Mishra S et al. (2023) Heavy metal/metalloid contamination: their sources in environment and accumulation in food chain. In: Heavy Metal Toxicity: Environmental Concerns, Remediation and Opportunities. Anonymous
Muniz DH, Oliveira-Filho EC (2023) Multivariate statistical analysis for water quality assessment: a review of research published between 2001 and 2020. Hydrology 10(10):196
Ndukwe MK et al (2023) Effect of abattoir waste on surface water quality parameters of Iwofe River, Port-Harcourt, Rivers State, Nigeria. J Geogr Environ Earth Sci Int 27(9):93–101
Nieder R, Benbi DK (2023) Potentially toxic elements in the environment–a review of sources, sinks, pathways and mitigation measures. Rev Environ Health
Nnaji CC et al (2023) Impact of anthropogenic and environmental conditions on surface run-off quality: a case study of Nsukka, Eastern Nigeria. Int J Environ Sci Technol 20(11):12351–12362
Ogoko EC, Sylvester AO (2020) Investigation on the quality of water from Jabi Lake in Abuja, Nigeria. J Chem Soc Nigeria 45(5)
Ogwueleka TC (2015) Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. Environ Monit Assess 187:1–17
Okey-Wokeh CG et al (2023) Anthropogenic impacts on physicochemical and heavy metal concentrations of Ogbor Hill River Water, Southern Nigeria. Water 15(7):1359
Olalekan AS et al (2023) Comparative assessment of seasonal variations in the quality of surface water and its associated health hazards in gold mining areas of Osun State, south-west Nigeria. Adv Environ Eng Res 4(1):011
Omeka ME et al (2024) A review of the status, challenges, trends, and prospects of groundwater quality assessment in Nigeria: an evidence-based meta-analysis approach. Environ Sci Pollut Res 31(15):22284–22307
Passos JBDMC et al (2021) Multivariate statistics for spatial and seasonal quality assessment of water in the Doce River basin, Southeastern Brazil. Environ Monit Assess 193(3):125
Pratama MA, Immanuel YD, Marthanty DR (2020) A multivariate and spatiotemporal analysis of water quality in code river, Indonesia. TheScientificWorld 2020:8897029–8897111. https://doi.org/10.1155/2020/8897029
Rangeti I, Dzwairo B (2021) Interpretation of water quality data in uMngeni basin (south africa) using multivariate techniques. In: River Basin Management-Sustainability Issues and Planning StrategiesAnonymous
Rautela KS et al (2023) Multivariate statistical analysis to assess the surface water quality of a snow and glacier-fed river: a case from Alaknanda River basin. Water Sci Technol 88(11):2873–2888
Rice EW (2012) Bridgewater and American Public Health Association, Standard Methods for the Examination of Water and Wastewater
Sager M, Wiche O (2024) Rare Earth Elements (REE): origins, dispersion, and environmental implications—a comprehensive review. Environments 11(2):24
Shafii NZ et al. (2019) Application of chemometrics techniques to solve environmental issues in Malaysia. Heliyon 5(10)
Shakhman I, Bystriantseva A (2021a) Water quality assessment of the surface water of the southern bug river basin by complex indices. J Ecol Eng 22(1):195–205
Shakhman I, Bystriantseva A (2021b) Spatio-temporal analysis of the ecological state of the dniester river transboundary water. J Ecol Eng 22(9):119–128
Sharma A et al. (2023) Heavy metal contamination in water: consequences on human health and environment. Metals in Water
Shulembayeva K et al (2023) Assessment of the hydrophysical and hydrochemical characteristics of Lake Burabay (Akmola Region, North Kazakhstan). Sustainability 15(15):11788
Soares ALC, Pinto CC, Oliveira SC (2020) Impacts of anthropogenic activities and calculation of the relative risk of violating surface water quality standards established by environmental legislation: a case study from the Piracicaba and Paraopeba river basins, Brazil. Environ Sci Pollut Res 27(12):14085–14099. https://doi.org/10.1007/s11356-020-07647-1
Talukdar P, Kumar B, Kulkarni VV (2023) A review of water quality models and monitoring methods for capabilities of pollutant source identification, classification, and transport simulation. Rev Environ Sci Bio/technol 22(3):653–677
Tazoe H (2023) Water quality monitoring. Anal Sci 39(1):1–3
Ubuoh EA et al (2023) Environmental pollution loads on surface water chemistry and potentially ecological risks of inland aquatic ecosystem in South-Eastern State, Nigeria. Environ Syst Res 12(1):22
Vega M et al (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res 32(12):3581–3592. https://doi.org/10.1016/S0043-1354(98)00138-9
Wieczorek K et al (2024) A holistic approach to the spatio-temporal variability investigation of the main river water quality–The importance of tributaries. Sci Total Environ 906:167588
World Health Organization (2020) WHO global water, sanitation and hygiene: annual report 2019
World Health Organization (2021) WHO guidelines on recreational water quality: volume 1: coastal and fresh waters. http://www.who.int/iris/handle/10665/342625

Yassin MA et al (2024) Toward decontamination in coastal regions: groundwater quality, fluoride, nitrate, and human health risk assessments within multi-aquifer Al-Hassa, Saudi Arabia. Water 16(10):1401

Yu Y et al (2020) Assessment of water quality using chemometrics and multivariate statistics: a case study in Chaobai River Replenished by Reclaimed Water, North China. Water (Basel) 12(9):2551. https://doi.org/10.3390/w12092551

Zahoor I, Mushtaq A (2023) Water pollution from agricultural activities: a critical global review. Int J Chem Biochem Sci 23(1):164–176

Zavareh M, Maggioni V, Sokolov V (2021) Investigating water quality data using principal component analysis and granger causality. Water 13(3):343

Zhou Y et al (2023) Water quality evaluation and pollution source apportionment of surface water in a major city in Southeast China using multi-statistical analyses and machine learning models. Int J Environ Res Public Health 20(1):881