International Journal of Zoology and Applied Biosciences **ISSN: 2455-9571**
Volume 10, Issue 5, pp: 176-188, 2025
https://doi.org/10.55126/ijzab.2025.v10.i05.020 **http://www.ijzab.com**

Research Article

ASSESSMENT OF WATER QUALITY USING MULTIVARIATE STATISTICAL TECHNIQUES: CASE STUDY OF THE MAN SAGAR LAKE, JAIPUR, RAJASTHAN, INDIA

Anamika, *V. Kumari, and S. Meena

Department of Zoology, University of Rajasthan, Jaipur (Rajasthan), India 302004.

Article History: Received 19th August 2025; Accepted 17th September 2025; Published 30th September 2025

ABSTRACT

The present study aimed to assess the water quality of Man Sagar Lake using multivariate statistical techniques. Water samples were collected over two years (2021–2023) from four sampling sites, and ten critical physicochemical parameters, namely temperature, pH, electrical conductivity (EC), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), alkalinity, hardness, nitrate, and phosphate, were analyzed following APHA standard methods. To interpret the complex dataset, Principal Component Analysis (PCA), Factor Analysis (FA), and Cluster Analysis (CA) were applied to identify major pollution sources, underlying factors, and site groupings. The results indicated relatively stable surface water temperatures (~26 °C), consistently alkaline pH (~8.95), and high EC (~2.38 mS/cm), while critically low DO (3.39–3.42 mg/L) reflected severe oxygen stress. Elevated BOD and COD, particularly at Site 1, signified heavy organic pollution, while nitrate (~4.2 mg/L) and phosphate (~1.19 mg/L) suggested eutrophication risk. PCA revealed a dominant pollution gradient accounting for 92.6% of total variance, mainly driven by organic and nutrient enrichment. FA extracted four latent factors linked to organic decomposition, chemical buffering, ion–nutrient imbalance, and seasonal variability. CA classified the sites into two distinct clusters, separating highly polluted zones (Sites 1 and 3) from relatively cleaner areas (Sites 2 and 4). Correlation analysis showed DO positively correlated with alkalinity ($r = 0.738$), COD ($r = 0.562$), BOD ($r = 0.514$), and hardness ($r = 0.555$). COD was strongly related to BOD ($r = 0.505$) and alkalinity ($r = 0.657$). pH correlated positively with alkalinity ($r = 0.504$) and COD ($r = 0.398$), but negatively with phosphate ($r = -0.156$). Both DO and BOD were weakly but significantly negatively correlated with phosphate. This study demonstrates the effectiveness of multivariate statistical techniques in diagnosing pollution drivers and provides a comprehensive framework for water quality assessment and restoration of stressed urban lakes.

Keywords: Cluster Analysis, Correlation analysis, Factor Analysis, Multivariate Statistical Methods, Urban lakes.

INTRODUCTION

The quality of surface water is a critical determinant of both ecological stability and human health. With rapid urbanization, water bodies are increasingly exposed to untreated wastewater, surface runoff, and altered land use. Urban lakes are particularly sensitive because they accumulate pollutants from multiple sources, making them hotspots of ecological stress. In developing regions, this challenge is more pronounced due to inadequate infrastructure and weak regulatory enforcement (Rahman *et al*., 2021). Lake water quality is shaped by interactions between natural factors such as rainfall variability, sediment dynamics, and geology, together with human-induced inputs including sewage discharge, solid waste dumping, and agricultural inflows. Parameters such as temperature, pH, dissolved oxygen (DO), electrical conductivity (EC), biochemical oxygen demand (BOD), chemical oxygen demand (COD), alkalinity, hardness, nitrate, and phosphate are interlinked in complex ways that cannot be fully explained by single-variable approaches.

To address these challenges, multivariate statistical methods such as Principal Component Analysis (PCA), Factor Analysis (FA), and Cluster Analysis (CA) are widely applied. These tools simplify large datasets, identify

*Corresponding Author: Dr. Vinod Kumari, Associate Professor, Lab No 16, Department of Zoology, University of Rajasthan, Jaipur (Rajasthan), India 302004 Email: vins.khangarot@yahoo.com.

key drivers of pollution, and differentiate natural processes from anthropogenic sources (Niu *et al*., 2021; Ramos-Pacheco *et al*., 2023;). In India, such approaches have been used successfully to assess pollution patterns in rivers (Singh *et al*., 2020; Mishra *et al*., 2024). The present study applies these techniques to Man Sagar Lake, Jaipur, over two years (2021–2023), with the objective of identifying critical water quality parameters, pollution gradients, and seasonal-spatial variability for sustainable management.

MATERIALS AND METHODS

Study Area: Man Sagar Lake, Jaipur

Man Sagar Lake, a 300-acre (121-hectare) artificial freshwater body in Jaipur, Rajasthan, India (26°57′22″N, 75°50′46″E), was constructed in 1610 by damming the Dravyavati River to alleviate severe water shortages (NIUA, 2020; Pradhan and Chauhan, 2016). The lake is bordered by the Aravalli Hills on three sides and densely populated urban settlements to the south. It experiences seasonal variations in water volume, ranging from 3.13 million cubic meters during the peak season to 360,000 cubic meters in the dry season (Indian Institute of Science, 2020) and in depth, from a maximum of 4.5 meters (15 feet) to a minimum of 1.5 meters (4.9 feet) (NIUA, 2020). With a catchment area of 23.5 km², the lake receives inflow from 325 seasonal streams and two perennial municipal streams, Brahmpuri and Nagtalai which carry untreated sewage and sediment into the lake (Sharma and Choudhary, 2021; Pradhan and Chauhan, 2016). A two-year study (2021-2023) was carried out on four fixed sampling sites (Sites 1–4) as illustrated in Figure 1.

Sample collection and analytical procedures

The assessment of physicochemical parameters of water was conducted over a continuous two-year period (2021–2023). Water samples were collected fortnightly from four fixed sampling sites established across Man Sagar Lake. All sampling was carried out during the morning hours between 7:00 AM and 12:00 PM. The analytical procedures for all physicochemical parameters followed the standard methods outlined in the APHA 23rd Edition (APHA, 2017) as mentioned in table 1.

Multivariate statistical methods

To examine relationships among physicochemical parameters of Man Sagar Lake, multivariate statistical analyses—Principal Component Analysis (PCA), Factor Analysis (FA), and Cluster Analysis (CA)—were employed. PCA transformed correlated variables into uncorrelated components, identifying dominant pollution sources (Ebrahimi *et al*., 2017). FA with Varimax rotation extracted latent factors linked to nutrient and organic contamination (Xiao *et al*., 2023). CA, using Ward's linkage and Euclidean distance, grouped sites into pollution clusters (Khan & Wen, 2021). Data suitability was confirmed by KMO (>0.5) and Bartlett's test (Shrestha, 2021). Analyses were performed using Origin 2024, IBM SPSS 26, and PAST 4.02.

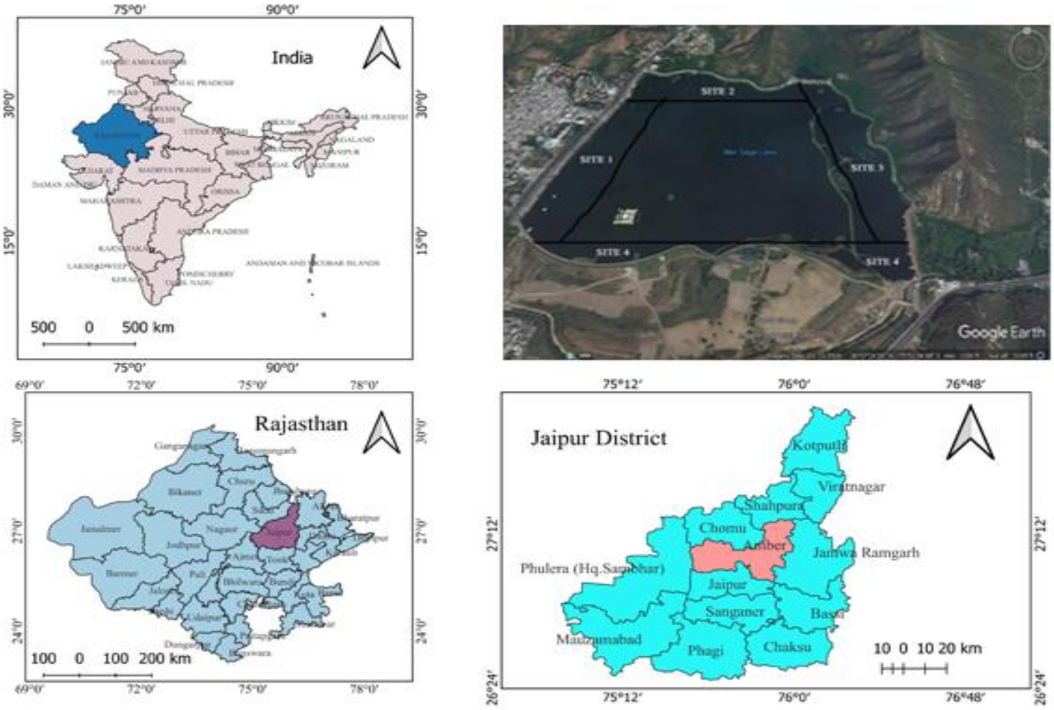


Figure 1. Location of Man Sagar Lake including India, Rajasthan, Jaipur District, and satellite image (QGIS version 3.40.2).

RESULT AND DISCUSSION

The basic statistics of lake water quality were based on 576 total water samples (4 sampling sites×3 replications×2 sampling frequency×24 months) as summarized in Table 2, which gives the range, mean and the standard error of the results for each of the 10 parameters. Water temperature showed minimal spatial variation across sites, with Site 1 recording the highest mean and Site 2 the lowest. This thermal stability aligns with earlier observations (Singh & Jain, 2021; Pradhan & Chauhan, 2016; Sharma & Choudhary, 2021), unlike wider fluctuations seen in Ana Sagar Lake (21–29°C) due to size and hydrological variability (Koli & Ranga, 2011). Fateh Sagar and Govardhan Sagar reached higher seasonal peaks of 26.2°C and 30.75°C, respectively (Mangal & Pathania, 2014; Mishra *et al*., 2012). pH remained consistently alkaline across all sites, slightly exceeding CPCB (2008) limits but within biologically tolerable ranges. These levels were driven by carbonate buffering, photosynthesis, and evaporation, similar to Ana Sagar (6.7–10.2) and Kishore Sagar (approx. 9.0) lakes (Koli & Ranga, 2011; Jain & Singh, 2013). Electrical conductivity (EC) exceeded the 1.5 mS/cm guideline at all sites, highest at Site 1, indicating ionic enrichment from chlorides, sulphates, nitrates, and phosphates due to sewage and runoff inputs (Singh & Jain, 2021; Pradhan & Chauhan, 2016). Comparable EC levels (2.0–2.6 mS/cm) were reported in Ana Sagar, Pichola, Fateh Sagar, and Govardhan Sagar lakes (Koli & Ranga, 2011; Shivani *et al*., 2019; Mangal & Pathania, 2014; Mishra *et al*., 2012).

Dissolved Oxygen (DO) levels were below the CPCB threshold of 4.0 mg/L, with Site 2 showing the highest and Site 1 the lowest. Low DO was attributed to organic matter decomposition, poor circulation, and stagnation (Pradhan & Chauhan, 2016; Singh & Jain, 2021). In contrast, Ana Sagar and Pichola Lakes showed generally higher DO levels (6.7–10.7 mg/L and 4.9–9.4 mg/L), despite hypoxic zones near drains (Koli & Ranga, 2011; Shivani *et al*., 2019). Both BOD and COD were elevated at all sites, highest at Site 1, indicating significant organic and chemical pollution from domestic sources and stagnant conditions (Singh & Jain, 2021; Pradhan & Chauhan, 2016). Similar levels were observed in Ana Sagar (BOD: 4.2–8.6 mg/L; COD: up to 78 mg/L), Govardhan Sagar, and Kishore Sagar (Koli & Ranga, 2011; Mishra *et al*., 2012; Jain & Singh, 2013). Total hardness and alkalinity were high throughout, peaking at Site 1, likely influenced by the region's mineral-rich geology and anthropogenic inputs (Singh & Jain, 2021; Pradhan & Chauhan, 2016). Comparable values were recorded in Ana Sagar (alkalinity: 215–240 mg/L; hardness: 280–310 mg/L), Pichola, Kishore Sagar, and Fateh Sagar Lakes (Koli & Ranga, 2011; Shivani *et al*., 2019; Jain & Singh, 2013; Mangal & Pathania, 2014). Nitrate and phosphate concentrations were moderately high and consistent across sites, peaking at Site 1. Nitrate ranged from 4.21–4.24 mg/L, and phosphate from 1.175–1.193 mg/L, pointing to persistent inputs from sewage and agriculture (Pradhan & Chauhan, 2016). These levels exceeded ecological safety thresholds, raising concerns over eutrophication and oxygen depletion (EPA, 2024; NOAA, 2024). By contrast, Pichola Lake reported lower nutrient concentrations (Shivani *et al*., 2019), while Ana Sagar showed periodic exceedances post-monsoon (Koli & Ranga, 2011). KMO and Bartlett's Test were used to evaluate the dataset's suitability for factor analysis. The KMO value was 0.570, indicating mediocre sampling adequacy and modest shared variance among variables. Bartlett's Test of Sphericity was significant ($\chi^2 = 764.980$, df = 45, $p < 0.001$), confirming that the inter-variable correlations were sufficient to proceed with factor analysis.

Table 1. Water quality parameters analyzed in the study along with their units, and standard analytical methods based on APHA (2017) guidelines.

Variables	Units	Analytical methods
Water Temperature	°C	Water analyzer kit (ME 972, Max Electronics, Mumbai, India)
pH	pH unit	water analyzer kit (ME 972, Max Electronics, Mumbai, India)
Electrical Conductivity	µS/cm	water analyzer kit (ME 972, Max Electronics, Mumbai, India)
Dissolved Oxygen	mg/L	water analyzer kit (ME 972, Max Electronics, Mumbai, India)
Biochemical Oxygen Demand	mg/L	5-day BOD test (APHA 5210 B)
Chemical Oxygen Demand	mg/L	Dichromate Method (APHA 5220 B)
Total Alkalinity	mg/L	Titrimetric Method (APHA 2320 B)
Total Hardness	mg/L	Titrimetric Method (APHA 2340 C)
Nitrate	mg/L	UV Spectrophotometric (APHA 4500-NO3)
Phosphate	mg/L	Stannous Chloride Method (APHA 4500-P)

*Corresponding Author: Dr. Vinod Kumari, Associate Professor, Lab No 16, Department of Zoology,
University of Rajasthan, Jaipur (Rajasthan), India 302004 Email: vins.khangarot@yahoo.com.

Table 2. Range, mean and standard error of water quality parameters at different sites of Man Sagar Lake during 2021-23.

Parameter	CPCB Limits		Site 1	Site 2	Site 3	Site 4
Water Temperature (°C)	-	**Range**	19-32	18.9-30.6	19.4-31.6	18.92-30.2
		Mean ± Std. Error	26.13± 0.464	25.97± 0.461	26.08± 0.461	26.05± 0.4626
pH	6.5-8.5	**Range**	8.45-9.5	8.44-9.2	8.44-9.4	8.43-9.3
		Mean ± Std. Error	8.951± 0.0386	8.943± 0.0374	8.947± 0.0364	8.946± 0.0374
Electrical Conductivity (mS/cm)	1.5	**Range**	1.620–3.268	1.600–3.260	1.610–3.265	1.608–3.263
		Mean ± Std. Error	2.3786± 0.6601	2.375± 0.6652	2.3785± 0.6617	2.3781± 0.6613
Dissolved Oxygen(mg/l)	4	**Range**	5.1-2.2	5.5-2.5	5.2-2.3	5.3-2.4
		Mean ± Std. Error	3.391± 0.1011	3.422± 0.1067	3.404± 0.101	3.411± 0.1008
Biological Oxygen Demand (mg/l)	3	**Range**	3.3-13.5	3.0-12.9	3.2-13.3	3.1-13.0
		Mean ± Std. Error	5.702± 0.7975	4.937± 0.347	4.965± 0.355	4.946± 0.349
Chemical Oxygen Demand (mg/l)	-	**Range**	49.9-92	48.5-87	49.5-89	48.8-86.5
		Mean ± Std. Error	62.5426± 1.769	62.29±1.7258	62.41± 1.722	62.36± 1.7476
Total Alkalinity (mg/l)	-	**Range**	132-436	128-433	130-435	129-434
		Mean ± Std. Error	222.48± 11.73	221.56± 11.704	222.2± 11.789	221.98± 11.725
Total Hardness (mg/l)	-	**Range**	192-396	188-390	190-393	189-391
		Mean ± Std. Error	290.16± 7.197	288.31± 6.99	289.42± 7.0944	288.62± 6.9796
Nitrate (mg/l)	-	**Range**	2.48-10.6	2.45-10.3	2.47-10.4	2.46-10.5
		Mean ± Std. Error	4.242± 0.2201	4.21± 0.221	4.23± 0.2206	4.22± 0.2218
Phosphate(mg/l)	-	**Range**	0.25-3.38	0.22-3.25	0.24-3.34	0.23-326
		Mean ± Std. Error	1.193± 0.0937	1.175± 0.092	1.191±0.0934	1.182± 0.0923

Table 3. Eigenvector and eigenvalues on the correlation matrixes of concentration of physico-chemical parameters in Man Sagar Lake.

Parameter	PC 1	PC 2	PC 3	PC 4
Water Temperature	1.24925	0.12601	-0.1895	-0.20027
pH	1.23609	-0.1863	-0.5563	0.88292
Electrical Conductivity	1.01317	1.0025	-0.5778	-5.93E-04
Dissolved Oxygen	-1.2353	0.28554	0.21208	0.14356
Biological Oxygen Demand	1.18813	0.22408	1.26405	0.0311
Chemical Oxygen Demand	1.19252	-0.2856	-1.1319	-0.31882
Total Alkalinity	1.25227	0.0256	-0.198	-0.20467
Total Hardness	1.20692	-0.3943	0.64627	-0.10049
Nitrate	1.24191	-0.2311	0.15336	-0.02134
Phosphate	1.22933	0.175	0.72994	0.06842
Eigenvalue	9.25898	0.49412	0.2469	0
Percentage of Variance	92.59%	4.94%	2.47%	0.00%
Cumulative (%)	92.59%	97.53%	100.00%	100.00%

Table 4. Total variance explained by the extracted components using Principal Component Analysis. The table presents the initial eigenvalues, the extraction sums of squared loadings, and the rotation sums of squared loadings.

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.505	35.051	35.051	3.505	35.051	35.051	3.228	32.277	32.277
2	1.476	14.761	49.812	1.476	14.761	49.812	1.444	14.442	46.719
3	1.357	13.569	63.381	1.357	13.569	63.381	1.363	13.631	60.350
4	1.009	10.095	73.476	1.009	10.095	73.476	1.313	13.126	73.476
5	0.753	7.526	81.002						
6	0.587	5.871	86.873						
7	0.538	5.383	92.256						
8	0.426	4.264	96.520						
9	0.258	2.582	99.102						
10	0.090	0.898	100.000						
Extraction Method: Principal Component Analysis.									

Table 5. Component and Rotated Component Matrix of PCA (Varimax rotation with Kaiser Normalization).

	Component Matrix[a] Component				Rotated Component Matrix[s] Component			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Temperature	0.009	-0.037	-0.722	0.44	-0.031	0.208	-0.056	-0.818
pH	0.524	0.056	-0.393	-0.64	0.236	-0.03	0.882	-0.082
DO	0.849	-0.054	0.218	-0.005	0.845	-0.097	0.189	0.114
EC	-0.353	0.758	0.035	0.268	-0.27	0.781	-0.259	0.151
COD	0.816	0.177	0.058	0.088	0.798	0.177	0.199	0.004
BOD	0.708	-0.05	0.162	0.049	0.714	-0.068	0.122	0.054
Total Alkalinity	0.859	0.231	-0.153	-0.061	0.753	0.231	0.434	-0.1
Hardness	0.622	-0.299	0.252	0.546	0.79	-0.201	-0.39	-0.146
Nitrate	-0.098	0.451	0.658	-0.115	0.01	0.277	-0.163	0.746
Phosphate	-0.29	-0.717	0.288	-0.09	-0.218	-0.758	-0.238	0.105

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
s. Rotation converged in 5 iterations.

Table 6. Component transformation matrix displaying the correlation coefficients among the rotated principal components after Varimax rotation with Kaiser normalization.

Component Transformation Matrix				
Component	1	2	3	4
1	0.939	-0.003	0.336	-0.070
2	-0.022	0.951	0.127	0.280
3	0.217	-0.189	-0.430	0.856
4	0.265	0.243	-0.828	-0.429

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

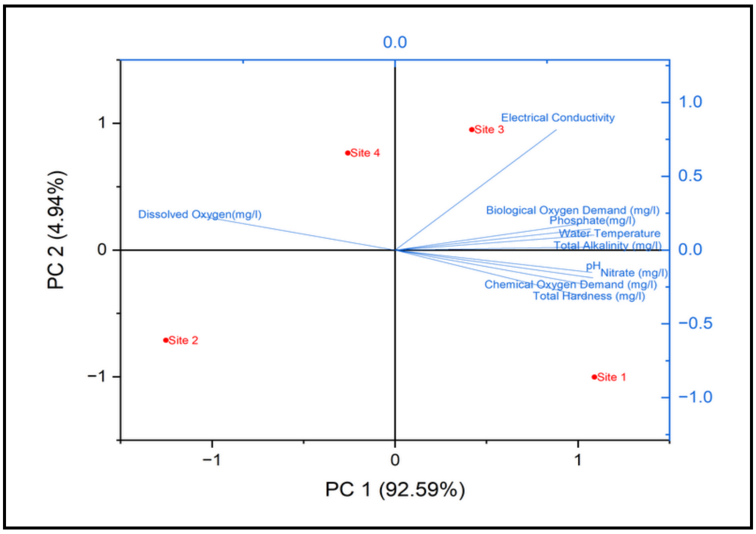
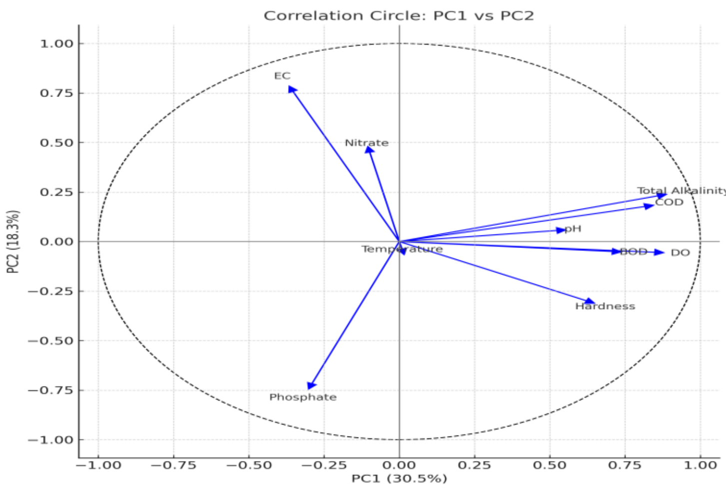
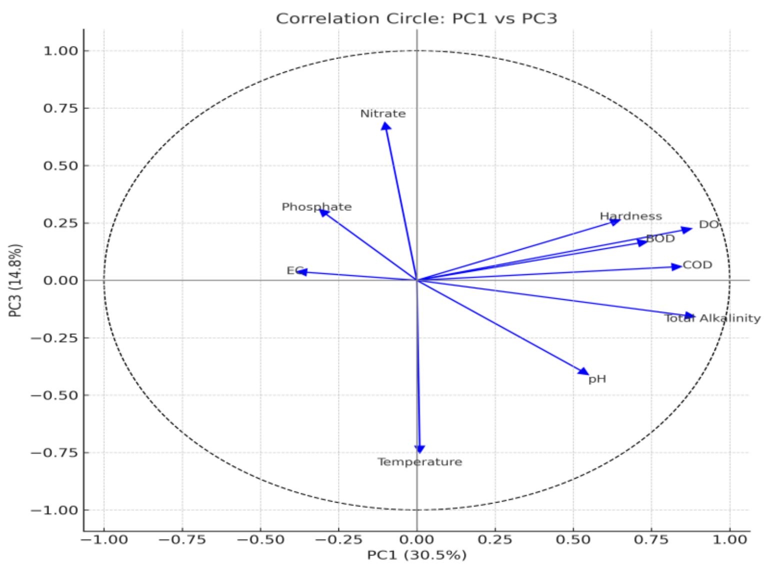


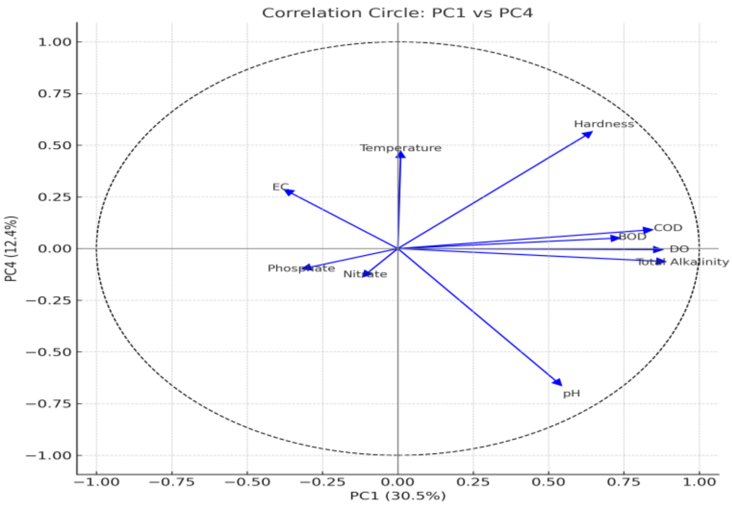
Figure 2. Scores of the first two principal components (PC1 and PC2) PC1 and PC2 explained 97.53% of the total variance.



(A)



(B)



(C)

Figure 3. Correlation circle plots showing variable loadings on PC1 vs PC2 (A), PC1 vs PC3 (B), and PC1 vs PC4 (C), highlighting distinct variance patterns linked to organic load, nutrients composition.

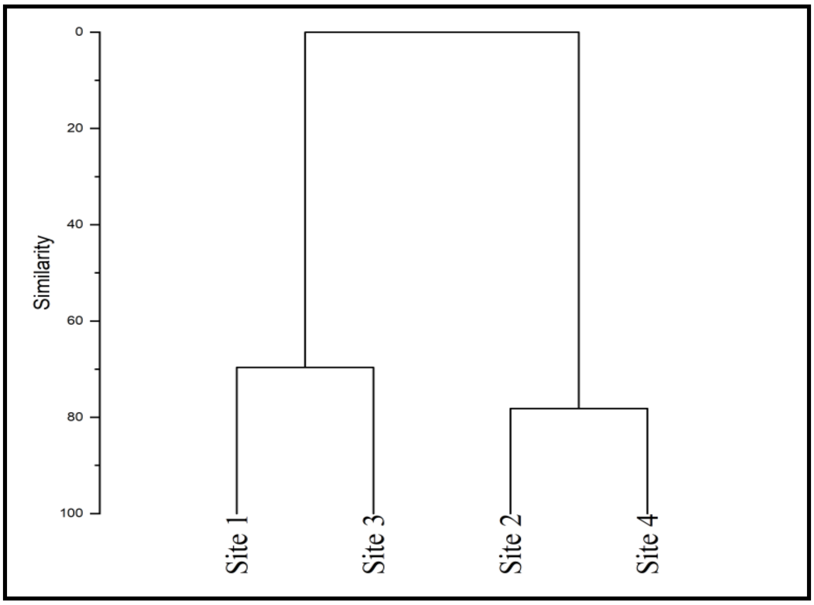


Figure 5. Dendrogram representing hierarchical cluster analysis (HCA) of the four sampling sites in Man Sagar Lake based on ten physico-chemical water quality parameters.

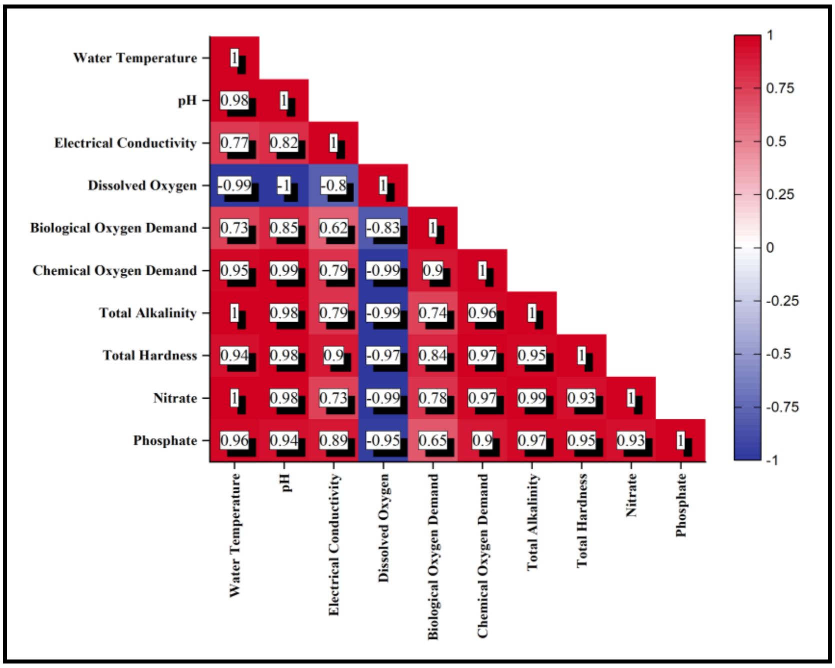


Figure 6. Correlation matrix showing Pearson correlation coefficients among physicochemical water quality parameter.

The result of the PCA, based on the correlation matrix of chemical components, is presented in Table 3 and site-wise in Figure 2. PCA extracted four components explaining 100% of total variance. PC1 accounted for 92.59% of the variance, with strong positive loadings from temperature, pH, EC, BOD, COD, alkalinity, hardness, nitrate, and phosphate, reflecting the dominant pollution gradient. PC2 explained 4.94% and was associated mainly with EC and DO, while PC3 contributed 2.47%, characterized by high BOD, phosphate, hardness, and a strong negative loading from COD. The PCA biplot revealed that PC1 and PC2 together accounted for 97.53% of the total variance. PC1 primarily reflected pollution from organic matter and nutrients, likely originating from urban runoff, sewage, and domestic effluents. Site 1, strongly aligned with PC1, indicated high pollution levels, whereas Site 2, positioned oppositely and associated with DO, showed relatively better water quality. Similar patterns were observed in the Kali River (Singh *et al*., 2020), Kankaria Lake (Mishra *et al*., 2024), the Yamuna River (Lokhande & Tare, 2021), and a drought-prone reservoir in southwest India (Markad *et al*., 2021). Internationally, PCA has been widely applied to identify pollution sources. Ghaemi and Noshadi (2022) linked urban and agricultural runoff to water quality issues in Iran's Fars Province. In the Gharaso River, domestic wastewater and nutrients were the main drivers (Khaledian *et al*., 2018), while Garizi *et al*. (2011) attributed seasonal variation in the Chehelchay watershed to monsoonal flow and natural geochemical inputs. These findings affirm PCA's role in tracing pollution origins and guiding water resource management.

To explore latent relationships among the water quality parameters, FA was conducted using Principal Component extraction with Varimax rotation. The Scree tesr indicated a distinct inflection point after the fourth component, supporting the retention of four components in accordance with Kaiser's criterion (eigenvalues > 1). These four components cumulatively accounted for a substantial proportion of the total variance, underscoring the multidimensional structure of the underlying dataset. The initial eigenvalue distribution indicated that the first four components possessed eigenvalues exceeding 1, satisfying the Kaiser criterion and justifying their retention for interpretation. Together, these four components explained 73.476% of the total variance in the dataset. Specifically, Component 1 accounted for the largest share of variance at 35.05%, followed by Component 2 with 14.76%, Component 3 with 13.57%, and Component 4 with 10.10%. After Varimax rotation, the rotation sums of squared loadings redistributed the variance more evenly: Component 1 explained 32.28%, Component 2 accounted for 14.44%, Component 3 for 13.63%, and Component 4 for 13.13% as shown in table 4. The initial component matrix (Table 5) revealed that Component 1 exhibited high positive loadings for DO (0.849), COD (0.816), BOD (0.708), alkalinity (0.859), and hardness (0.622). Component 2 was characterized by a strong loading for EC (0.758) and moderate loading for nitrate (0.451). Component 3 was heavily influenced by temperature (−0.722) and nitrate (0.658), while Component 4 showed significant associations with temperature (0.440), pH (−0.640), and hardness (0.546).

Following Varimax rotation, the rotated component matrix (Table 5) yielded a more interpretable factor structure. Factor 1 retained high positive loadings for DO (0.845), COD (0.798), BOD (0.714), alkalinity (0.753), and hardness (0.790). Factor 2 demonstrated a strong positive loading for EC (0.781) and a strong negative loading for phosphate (−0.758). Factor 3 was almost exclusively defined by pH (0.882), implicating it as a chemical equilibrium factor governing acid–base and solubility dynamics. Factor 4 exhibited a negative loading for temperature (−0.818) and a positive loading for nitrate (0.746). The component transformation matrix (Table 6) illustrated the mathematical rotation of the axes, confirming orthogonality of the components post-rotation. This transformation improved the interpretability the final factor structure while preserving the statistical independence of the factors. Collectively, the four extracted factors encapsulated the key latent dimensions governing water quality in the study area: organic–mineral buffering, ionic nutrient contrast, pH regulation, and temperature-linked nutrient enrichment.

The correlation circle plots derived from PCA illustrated the relationships among physicochemical parameters across different principal component combinations as shown in figure 3. In the PC1 vs PC2 plot, which explained 48.8% of the total variance, DO, BOD, COD, total alkalinity, and pH showing strong positive associations along PC1, while EC and Nitrate aligning more with PC2. In the PC1 vs PC3 projection (45.3% cumulative variance), Temperature, Nitrate, and Phosphate dominated PC3, indicating their roles in capturing variations related to thermal and nutrient factors, orthogonal to the organic pollution indicators of PC1. The PC1 vs PC4 circle (42.9% variance) highlighted hardness, temperature, and nitrate as contributors to PC4, separated from the DO, BOD, COD, and alkalinity cluster in PC1, reflecting differing pollution sources. These projections together revealed clear groupings of water quality parameters, with PC1 capturing the most organic load variability, while PC2–PC4 distinguished nutrient, ionic, and temperature-related influences. FA extracted four latent factors that collectively explained 73.47% of the total variance, highlighting the multidimensionality of water quality in Man Sagar Lake. Factor 1, accounting for 32.28% of the variance, included DO, BOD, COD, alkalinity, and hardness, indicating an organic–mineral buffering system driven by untreated sewage, microbial decomposition, and carbonate equilibrium. Similar clusters were reported in urban lake studies by Rahman *et al.* (2021) and Sheela *et al.* (2012), and in riverine contexts by Sharma *et al.* (2021;2022).

Factor 2, with positive EC and negative phosphate, reflected ionic contamination and nutrient uptake or dispersion, as observed in Udaipur lakes (Bharadwaj *et al.* 2025) and Lake Tianmuhu, China (Zhang *et al.* 2011). Factor 3, dominated by pH, likely represented acid–base buffering, echoing findings from Yuqiao Reservoir (Chen *et al.* 2018). Factor 4, with negative temperature and positive nitrate loadings, suggested seasonal nutrient cycling, consistent with studies in glacial lakes (Zhang *et al.* 2021), Mettur Reservoir (Saha *et al.* 2021), and national-scale satellite analyses (Deoli *et al.* 2021). These patterns affirmed the utility of FA in diagnosing ecological and chemical dynamics in freshwater systems. Cluster analysis was performed using Ward's linkage method and Euclidean distance to classify the water quality of the four sampling sites in Man Sagar Lake, as illustrated in Figure 5. The resulting dendrogram distinctly grouped the sites into two primary clusters. Cluster 1, comprising Sites 1 and 3, reflected similar pollution profiles characterized by elevated concentrations of BOD, COD, nitrate, and phosphate. These locations were indicative of high pollution loads due to domestic wastewater discharge and nutrient-rich surface runoff. Cluster 2, including Sites 2 and 4, showed comparatively better water quality, marked by lower organic and nutrient levels and higher dissolved oxygen content.

This spatial pattern demonstrated the utility of CA in identifying pollution hotspots and guiding lake management strategies. A similar application was reported from Kurichi Lakes, Big Lake Tamil Nadu, where Yogeshwaran and Priya (2025) used CA to differentiate nutrient-loaded zones from less impacted areas. Dhanush *et al.* (2024) also applied CA in semi-arid lake systems, successfully distinguishing high- and low-pollution zones, highlighting effectiveness of CA in regional water quality assessment. In the Ramsar-listed Rudrasagar wetland, Debnath *et al.* (2023) utilized K-means clustering to identify sensitive versus disturbed areas, showcasing relevance of CA for protected ecosystems. Ramya *et al.* (2022) used BOD and EC as key variables in CA to delineate pollution clusters in Veeranam Lake, closely aligning with the EC- and nitrate-driven clusters seen in Man Sagar. Andrabi *et al.* (2024) reported similar biotic–abiotic cluster patterns in Manasbal Lake. Globally, CA has proven effective in Kenya (Githaiga *et al.* 2021), Egypt (Eid *et al.* 2024), China (Tian *et al.* 2024), and the Philippines (Estorosos *et al.* 2023), reinforcing its value in global lake and wetland monitoring efforts.

Several significant correlations among physicochemical parameters in Man Sagar Lake were identified (Figure 6). DO showed strong positive correlations with alkalinity ($r = 0.738$, $p < 0.01$), COD ($r = 0.562$), BOD ($r = 0.514$), and hardness ($r = 0.555$), reflecting linked oxygen dynamics, buffering, and organic degradation. COD correlated strongly with alkalinity ($r = 0.657$) and BOD ($r = 0.505$), indicating common sewage and organic sources. pH was positively correlated with alkalinity ($r = 0.504$) and COD ($r = 0.398$), but negatively with phosphate ($r = –0.156$, $p < 0.05$). Similarly, DO and BOD were inversely related to phosphate ($r = –0.158$ and $–0.145$, respectively; $p < 0.05$), suggesting nutrient uptake or sediment binding in alkaline conditions. Moon *et al*. (2024) observed a strong COD–BOD correlation in Amalnala Lake, linked to sewage and agricultural runoff, echoing DO–alkalinity–COD patterns in Haryana wetlands (Singh & Khalid, 2022), Deepor Beel (Bormudoi *et al*., 2022), and the Eastern Himalayas (Ganie *et al*., 2023). The negative phosphate–pH relationship aligned with studies in Lake Dianchi (Wang *et al*., 2021), Chaohu (Bao *et al*., 2023), and Chenhu (Ma *et al*., 2022). Temperature, turbidity, and TDS co-variation was also reported via remote sensing in Manchar Lake (Imran *et al*., 2022). Nutrient buffering correlations appeared in the Western Ghats (Saranya *et al*., 2021), Central Asian rivers (Khan *et al*., 2022), and Naltar Lakes (Muhammad, 2023).

Correlation frameworks have further supported hazard assessments in Nepal's glacial lakes (Begam, 2024) and microplastic–nutrient studies in Jilin, China (Yin *et al*., 2021). These findings affirm correlation analysis as a vital tool for understanding pollutant interactions and ecological risks in freshwater bodies like Man Sagar Lake.

CONCLUSION

The study applied a suite of multivariate statistical tools to analyze the spatial and temporal dynamics of ten key physicochemical parameters in Man Sagar Lake, Jaipur. The temperature remained relatively stable (~26°C) across sites, reflecting minimal thermal stratification. The lake exhibited widespread pollution, as evidenced by persistently high pH levels (alkaline range ~8.9), electrical conductivity above 2.3 mS/cm, and DO concentrations below CPCB standards across all sampling sites. High BOD (5.7 mg/L) and COD (62.5 mg/L) values, particularly at Site 1, reflected organic pollution from untreated sewage inflows, while elevated nitrate and phosphate levels suggested ongoing nutrient loading and eutrophication risk. PCA identified a dominant pollution vector governed by these parameters, explaining over 92% of the variance. FA revealed four influential factors tied to organic-microbial activity, ionic-nutrient profiles, buffering mechanisms, and seasonal nutrient cycling. Cluster analysis spatially classified the lake into two zones: Cluster 1 (Sites 1 and 3), characterized by poor water quality, and Cluster 2 (Sites 2 and 4), with relatively better conditions. Correlation analysis further validated these findings, revealing significant positive correlations between DO, COD, and alkalinity, and negative associations with phosphate. These insights highlighted the importance of DO and buffering capacity as integrative indicators of water quality health. The integrated use of PCA, FA, and CA offers a comprehensive diagnostic framework to guide targeted interventions and sustainable lake management practices. The findings are especially relevant for similarly stressed urban water bodies in semi-arid regions undergoing rapid development.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Council of Scientific and Industrial Research–Human Resource Development Group (CSIR-HRDG), New Delhi, for the financial assistance provided under Reference File No: 09/149(0763)/2019-EMR-1. We are also thankful to the Head of the Department of Zoology, University of Rajasthan, Jaipur, for providing the necessary laboratory and infrastructural facilities to carry out this research.

CONFLICT OF INTERESTS

The authors declare no conflict of interest

ETHICS APPROVAL

Not applicable

FUNDING

Council of Scientific and Industrial Research–Human Resource Development Group (CSIR-HRDG), New Delhi, for the financial assistance provided under Reference File No: 09/149(0763)/2019-EMR-1.

AI TOOL DECLARATION

The authors declares that no AI and related tools are used to write the scientific content of this manuscript.

DATA AVAILABILITY

Data will be available on request

REFERENCES

Andrabi, S., Bakhtiyar, Y., Yousuf, T., Akhtar, M., & Nissar, S. (2024). Water quality assessment in relation to fish assemblage using multivariate analysis in Manasbal Lake, Kashmir. *Water Science, 38,* 92–108.

APHA. (2017). *Standard methods for the examination of water and wastewater* (23rd ed.). Washington, DC: American Public Health Association.

Bao, Y., Huang, T., Ning, C., Sun, T., Tao, P., Wang, J., *et al*. (2023). Changes of DOM and its correlation with internal nutrient release during cyanobacterial growth and decline in Lake Chaohu, China. *Journal of Environmental Sciences, 124,* 769–781.

Begam, S. (2024). Potentially dangerous moraine-dammed glacial lakes for outburst in the Nepal Himalaya. In *The nature of geomorphological hazards in the Nepal Himalaya* (pp. 259–282). Cham: Springer International Publishing.

Bharadwaj, D., Kumar, P., & Choudhary, M. P. (2025). Geospatial and multi-index assessment of urban lake water quality: Insights into pollution sources and irrigation suitability in Udaipur, the lake city of India. *Water Environment Research, 97,* e70116.

Bormudoi, A., Hinge, G., Nagai, M., Kashyap, M. P., & Talukdar, R. (2022). Retrieval of turbidity and TDS of Deepor Beel Lake from Landsat 8 OLI data by regression and artificial neural network. *Water Conservation Science and Engineering, 7,* 505–513.

Chen, R., Ju, M., Chu, C., Jing, W., & Wang, Y. (2018). Identification and quantification of physicochemical parameters influencing chlorophyll-a concentrations through combined principal component analysis and factor analysis: A case study of the Yuqiao Reservoir in China. *Sustainability, 10,* 936.

Debnath, P., Roy, S., Bharadwaj, S., Hore, S., Nath, H., Mitra, S., *et al*. (2023). Application of multivariable statistical and geo-spatial techniques for evaluation of water quality of Rudrasagar wetland, the Ramsar site of India. *Water, 15,* 4109.

Deoli, V., Kumar, D., Kumar, M., Kuriqi, A., & Elbeltagi, A. (2021). Water spread mapping of multiple lakes using remote sensing and satellite data. *Arabian Journal of Geosciences, 14,* 2213.

Dhanush, S. K., Murthy, M. M., Sathish, A., & Verma, R. (2024). Evaluating semi-arid lake water quality: A synergy of water quality indices, multivariate statistics and geospatial technology. *Water Environment Sustainability, 4,* 52–70.

Ebrahimi, M., Gerber, E. L., & Rockaway, T. D. (2017). Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis. *Journal of Environmental Management, 193,* 234–246.

Eid, M. H., Eissa, M., Mohamed, E. A., Ramadan, H. S., Czuppon, G., Kovács, A., *et al*. (2024). Application of stable isotopes, mixing models, and K-means cluster analysis to detect recharge and salinity origins in Siwa Oasis, Egypt. *Groundwater for Sustainable Development, 25,* 101124.

EPA. (2024). The effects: Dead zones and harmful algal blooms. U.S. Environmental Protection Agency. Retrieved from https://www.epa.gov/nutrientpollution/effects-dead-zones-and-harmful-algal-blooms

Estorosos, J. S., Genterolizo, J. M. B., Rios, W. V., Saavedra, K. C., Andres, M. I. C., Villeta, R. B., *et al*. (2023). Assessment of spatial variations in surface water quality of Laguna Lake stations using cluster analysis approach. *Journal of Biodiversity and Environmental Sciences, 22,* 1–8.

Ganie, P. A., Posti, R., Baruah, D., Kunal, K., Kunal, G., Sarma, D., *et al*. (2023). Land suitability modelling for rainbow trout farming in the Eastern Himalayan Region, India, using GIS–MCE approach. *Modeling Earth Systems and Environment, 9,* 2437–2462.

Garizi, A. Z., Sheikh, V., & Sadoddin, A. (2011). Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods. *International Journal of Environmental Science and Technology, 8,* 581–592.

Ghaemi, Z., & Noshadi, M. (2022). Surface water quality analysis using multivariate statistical techniques: A case study of Fars Province rivers, Iran. *Environmental Monitoring and Assessment, 194,* 178.

Githaiga, K. B., Njuguna, S. M., Gituru, R. W., & Yan, X. (2021). Water quality assessment, multivariate analysis and human health risks of heavy metals in eight major lakes in Kenya. *Journal of Environmental Management, 297,* 113410.

Imran, U., Zaidi, A., Mahar, R. B., & Khokhar, W. A. (2022). Assessment of the lake water quality using Landsat 8 OLI imagery: A case study of Manchar Lake, Pakistan. *Arabian Journal of Geosciences, 15,* 1094.

Jain, A., & Singh, S. K. (2013). Limnological studies related to physico-chemical characteristics of water in Anasagar Lake, Ajmer, Rajasthan. *International Journal of Advanced Research in Innovation, 1,* 65–72.

Khaledian, Y., Ebrahimi, S., Natesan, U., Basatnia, N., Nejad, B. B., Bagmohammadi, H., *et al*. (2018). Assessment of water quality using multivariate statistical analysis in the Gharaso River, Northern Iran. In *Urban ecology, water quality and climate change* (pp. 227–253). Cham: Springer International Publishing.

Khan, M. A., & Wen, J. (2021). Evaluation of physicochemical and heavy metals characteristics in surface water under anthropogenic activities using multivariate statistical methods, Garra River, Ganges Basin, India. *Environmental Engineering Research, 26,* 6.

Khan, M., Ellahi, A., Niaz, R., Ghoneim, M. E., Tag-eldin, E., & Rashid, A. (2022). Water quality assessment of alpine glacial blue water lakes and glacial-fed rivers. *Geomatics, Natural Hazards and Risk, 13,* 2597–2617.

Koli, V. K., & Ranga, M. M. (2011). Physicochemical status and primary productivity of Ana Sagar Lake, Ajmer (Rajasthan), India. *Universal Journal of Environmental Research and Technology, 1,* 3.

Lokhande, S., & Tare, V. (2021). Spatio-temporal trends in the flow and water quality: Response of river Yamuna to urbanization. *Environmental Monitoring and Assessment, 193,* 117.

Ma, N., Tong, L., Li, Y., Yang, C., Tan, Q., & He, J. (2022). Distribution of antibiotics in lake water–groundwater–sediment system in Chenhu Lake area. *Environmental Research, 204,* 112343.

Mangal, H., & Pathania, S. (2014). Aquatic resources: A case study of Udaipur ‘City of Lakes’, Rajasthan. In *Aquatic ecosystem: Biodiversity, ecology and conservation* (pp. 13–20). New Delhi: Springer India.

Markad, A. T., Landge, A. T., Nayak, B. B., Inamdar, A. B., & Mishra, A. K. (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). *Environmental Science and Pollution Research, 28,* 31013–31031.

Mishra, M., Singhal, A., Rallapalli, S., & Sharma, R. (2024). Innovative lake pollution profiling: Unveiling pollutant sources through advanced multivariate clustering techniques. *Environmental Management, 74,* 818–834.

Mishra, V., Sharma, S. K., Sharma, B. K., Upadhyay, B., & Choubey, S. (2012). Phytoplankton, primary productivity and certain physico-chemical parameters of Goverdhan Sagar lake of Udaipur, Rajasthan. *Universal Journal of Environmental Research and Technology, 2,* 569–574.

Moon, U. R., Durge, A. A., & Wadhai, V. S. (2024). Impacts of agricultural practices (pesticides and fertilizers) and Amalnala Stream on Amalnala Lake in Gadchandur, India. *Journal of Pure and Applied Microbiology, 18,* 1.

Muhammad, S. (2023). Evaluation of heavy metals in water and sediments, pollution, and risk indices of Naltar Lakes, Pakistan. *Environmental Science and Pollution Research, 30,* 28217–28226.

Niu, S., Song, X., Yu, J., & Wu, J. (2021). Evaluation of spatial and seasonal water quality variation of urban lake by multivariate statistical approaches for water quality improvement. *Desalination and Water Treatment, 239,* 137–149.

NIUA (National Institute of Urban Affairs). (2020). *Revival of Man Sagar Lake, Jaipur: A case study.* Urban Climate Change Resilience Strategy Series. New Delhi: NIUA.

NOAA (National Oceanic and Atmospheric Administration). (2024). What causes nutrient pollution? NOAA's National Ocean Service. Retrieved from https://oceanservice.noaa.gov/education/tutorial-coastal/harmful-algal-blooms/habs01-sub-02.html

Pradhan, C., & Chauhan, S. (2016). Physicochemical analysis of water quality of Man Sagar Lake, Jaipur. *International Research Journal of Environmental Science, 5,* 21–27.

Rahman, K., Barua, S., & Imran, H. M. (2021). Assessment of water quality and apportionment of pollution sources of an urban lake using multivariate statistical analysis. *Clean Engineering and Technology, 5,* 100309.

Ramos-Pacheco, B. S., Choque-Quispe, D., Ligarda-Samanez, C. A., Solano-Reynoso, A. M., Choque-Quispe, Y., Aguirre Landa, J. P., *et al*. (2023). Water pollution indexes proposal for a high Andean river using multivariate statistics: Case of Chumbao River, Andahuaylas, Apurímac. *Water, 15,* 2662.

Ramya, M., Elumalai, S., & Umamaheswari, A. (2022). Study of physicochemical and statistical analysis of water quality parameters in Veeranam Lake, Cuddalore district, Tamil Nadu, India. *Materials Today: Proceedings, 49,* 2934–2942.

Saha, A., Ramya, V. L., Jesna, P. K., Mol, S. S., Panikkar, P., Vijaykumar, M. E., *et al*. (2021). Evaluation of spatio-temporal changes in surface water quality and their suitability for designated uses, Mettur Reservoir, India. *Natural Resources Research, 30,* 1367–1394.

Saranya, P., Krishnakumar, A., Sinha, N., Kumar, S., & Krishnan, K. A. (2021). Isotopic signatures of moisture recycling and evaporation processes along the Western Ghats orography. *Atmospheric Research, 264,* 105863.

Sharma, G., & Choudhary, M. (2021). Assessment of water quality parameters of Man Sagar Lake Jaipur, Rajasthan, India. *Pollution Research, 40,* 90–96.

Sharma, G., Lata, R., Thakur, N., Bajala, V., Kuniyal, J. C., & Kumar, K. (2021). Application of multivariate statistical analysis and water quality index for quality characterization of Parbati River, Northwestern Himalaya, India. *Discover Water, 1,* 5.

Sharma, S., Mittal, A., Joshi, B. P., & Pandey, P. (2023). Predictive model for dissolved oxygen in water body (Nainital Lake) of Nainital district by multivariate regression analysis. In *Proceedings of the 3rd International Conference on Advances in Computing, Innovation and Technology in Engineering (ICACITE)* (pp. 376–379).

Sheela, A. M., Letha, J., Joseph, S., Chacko, M., Kumar, S. P., & Thomas, J. (2012). Water quality assessment of a tropical coastal lake system using multivariate cluster, principal component and factor analysis. *Lakes & Reservoirs: Research and Management, 17,* 143–159.

Shivani, A. P., Sharma, S. K., Sharma, B. K., Upadhyay, B., & Jat, G. (2019). A study on physicochemical parameters of Lake Pichola, Udaipur (Rajasthan). *Journal of Entomology and Zoology Studies, 7,* 752–755.

Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics, 9,* 4–11.

Singh, G., & Khalid, M. A. (2022). To study seasonal variation and co-relation in water quality parameters of different wetlands in Haryana. *Journal of Experimental Zoology, India, 25,* 2.

Singh, G., Patel, N., Jindal, T., Srivastava, P., & Bhowmik, A. (2020). Assessment of spatial and temporal variations in water quality by the application of multivariate statistical methods in the Kali River, Uttar Pradesh, India. *Environmental Monitoring and Assessment, 192,* 394.

Singh, Y., & Jain, M. K. (2021). Assessment of physicochemical parameters and metal contents of Mansagar Lake of Jaipur. *Environmental Earth Sciences, 80,* 284.

Tian, D., Zhao, X., Gao, L., Liang, Z., Yang, Z., Zhang, P., *et al*. (2024). Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, South China. *Environmental Pollution, 342,* 123104.

Wang, J. H., Li, C., Xu, Y. P., Li, S. Y., Du, J. S., Han, Y. P., *et al*. (2021). Identifying major contributors to algal blooms in Lake Dianchi by analyzing river-lake water quality correlations in the watershed. *Journal of Cleaner Production, 315,* 128144.

Xiao, J., Gao, D., Zhang, H., Shi, H., Chen, Q., Li, H., *et al*. (2023). Water quality assessment and pollution source apportionment using multivariate statistical techniques: A case study of the Laixi River Basin, China. *Environmental Monitoring and Assessment, 195,* 287.

Yin, K., Wang, D., Zhao, H., Wang, Y., Guo, M., Liu, Y., *et al*. (2021). Microplastics pollution and risk assessment in water bodies of two nature reserves in Jilin Province: Correlation analysis with the degree of human activity. *Science of the Total Environment, 799,* 149390.

Yogeshwaran, V., & Priya, A. (2025). Assessment of surface water quality parameters using multivariate analysis—A case study of Kurichi and Big Lakes in Coimbatore. *Water Environment Research, 97,* e70055.

Zhang, M., Chen, F., Zhao, H., Wang, J., & Wang, N. (2021). Recent changes of glacial lakes in the High Mountain Asia and its potential controlling factors analysis. *Remote Sensing, 13,* 3757.

Zhang, Y., Yin, Y., Feng, L., Zhu, G., Shi, Z., & Liu, X. (2011). Characterizing chromophoric dissolved organic matter in Lake Tianmuhu and its catchment basin using excitation-emission matrix fluorescence and parallel factor analysis. *Water Research, 45,* 5110–5122.