



# Assessment and interpretation of surface water quality in Jhelum River and its tributaries using multivariate statistical methods

Sarvat Gull · Shagoofa Rasool Shah ·  
Ayaz Mohmood Dar

Received: 12 December 2022 / Accepted: 3 May 2023 / Published online: 27 May 2023  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

**Abstract** Water is an essential part of human life, and its pollution is a hotly debated topic on both national and international scales. Surface waterbodies in the beautiful Kashmir Himalayas are already deteriorating. In this study, fourteen physio-chemical parameters were tested in water samples taken during the spring, summer, autumn, and winter seasons from twenty-six different sampling points. The findings showed a consistent decline in the water quality of river Jhelum and its adjoining tributaries. The upstream section of the river Jhelum had the least pollution, whereas the Nallah Sindh had the poorest water quality. The water quality of Jhelum and Wular Lake was strongly impacted by the water quality of all the adjoining tributaries. To examine the link between the selected water quality indicators, descriptive statistics and a correlation matrix were used. Analysis of variance (ANOVA) and principal component analysis/factor analysis (PCA/FA) were used to identify the key variables that influenced seasonal and sectional water quality fluctuations. The ANOVA analysis revealed that there were significant

differences in water quality characteristics among the twenty-six sampling locations throughout all four seasons. The PCA findings identified four principal components that accounted for 75.18% of the total variance and could be utilized to evaluate all data. The study revealed that chemical, conventional, organic, and organic pollutants were significant latent factors influencing the water quality of rivers in the region. The findings of this study could contribute to the vital management of surface water resources in Kashmir's ecology and environment.

**Keywords** Water quality · PCA · ANOVA · Kashmir Himalayas · Variance

## Introduction

The quality of surface water is critical to human survival and is a worldwide concern owing to its fragility. Riverine systems have been essential to the growth of human civilizations throughout history because they provide an adequate and essential source of fresh water for household, agricultural, and industrial uses. However, uncontrolled urbanization and industrialization, which have facilitated modern cultures' economic advancement, have seriously endangered life on the planet by polluting water. Both human and natural causes of pollution have an influence on the water quality of rivers (Njuguna et al., 2020). Water quality is affected by a

---

S. Gull (✉) · S. R. Shah · A. M. Dar  
Department of Civil Engineering, National Institute  
of Technology Srinagar, Srinagar 190006, India  
e-mail: r.sarvat@gmail.com

S. R. Shah  
e-mail: shagoofa2002@yahoo.com

A. M. Dar  
e-mail: ayazmohmood@hotmail.com

number of natural processes, including variations in precipitation, surface runoff, erosion, and weathering (Akhtar et al., 2021). However, human activities including irrigational operations, sewage, municipal waste, and effluents also have a significant influence (Hanjra et al., 2012).

Every human possesses a fundamental right to access clean water which is also essential to their well-being. Unfortunately, a number of global issues, including industrial activity, population increase, urbanization, and climate change, make it difficult for many parts of the world to acquire clean water. Anthropogenic activity's adverse effect on water quality has a big impact on the environment and people's health. The geographical structure of regions and seasons affects surface water quality even when pollution is not present (Mutlu, 2021; Mutlu & Kurnaz, 2017). Thus, it is necessary to assess surface water quality in order to assure its safe usage for a variety of purposes, including drinking, industrial, and agricultural operations. For the betterment of both people and the environment, effort needs to be taken to stop water contamination and preserve the quality of surface water.

The natural resources of the Kashmir valley have given the fairly large number of people residing there a source of sustenance and livelihood for as long as humans have lived. However, despite the variety of benefits that the area's rivers and lakes provide, they are frequently used as dumping sites for wastewater, which has led to a significant degradation in water quality (Qayoom et al., 2022). Additionally, a number of factors, including changes in land use, uncontrolled application of pesticides and fertilizers, deforestation, a great deal of tourism pressure, and unplanned urbanization, have exerted a significant impact on the biological and physicochemical qualities of the valley's water resources (Mir & Gani, 2019). For instance, Showqi et al. (2014) observed that anthropogenic activities, changes in hydrometeorological climate, and changes in land use/land cover could potentially be responsible for the deterioration of the river Jhelum. Similar to this, Ganaie et al. (2021) claim that changes in land use land cover and subsequent alterations to hydrological patterns, such as increased erosion, reduced runoff, and sedimentation, are to blame for the decline of Wular Lake. However, it is noteworthy that the hydrogeochemistry of Wular Lake and the water quality of the river Jhelum are both significantly influenced by the water quality of

several adjoining tributaries, viz., Sindh, Sukhnag, Aripal, Rambiar, and Lidder (Khanday et al., 2021).

To assess the quality of surface freshwater, researchers have utilized a variety of methodologies, such as the Trophic Status Index (TSI) and the Water Quality Index (WQI), to acquire comprehensive information on a wide range of complex water quality indicators (Lee et al., 2022). Further, Factor analysis (FA) and Analysis of Variance (ANOVA) are two multivariate statistical approaches that have been frequently employed to get helpful information from the analysis of water quality data (Aydin et al., 2021; Najar & Khan, 2012). These techniques have proven useful for characterizing and assessing water quality and analyzing spatial and temporal changes caused by both naturally occurring and human-induced factors across different river sections (Machiwal et al., 2018). Multivariate statistical techniques make it feasible to uncover information regarding potential environmental effects on water quality that could not have been evident from the raw data alone (Schreiber et al., 2022). Factor analysis is a valuable tool for identifying fundamental elements that explain relationships between observed factors, including water quality features (Tung & Yaseen, 2020). The approach involves calculating the parameter correlation matrix and subsequently deriving eigenvalues and factor loadings from the matrix. Factor loadings represent the degree of correlation between variables and factors, whereas eigenvalues relate to eigenvectors that highlight categories of variables with a high degree of relation (Bowman & Goodbody, 2020). Most of the variability in the data is often explained by the first few components. On the other hand, ANOVA is a statistical method used to examine whether the mean values of two or more groups differ when they are based on two distinct factors (Keysers et al., 2020).

This study's objective was to collect preliminary data on water pollution by examining water samples taken from different places along the river Jhelum and its tributaries. This research possessed three objectives: (1) to evaluate the water quality of the various streams of the Kashmir region and their impact on the river Jhelum, (2) to detect regional and temporal changes in water quality as well as potential sources of pollution using descriptive statistics (DS) and principal component analysis (PCA), and (3) to use analysis of variance (ANOVA) for evaluating the variation of water quality measures at various spots.

The findings of this study could provide substantial insight to decision-makers in managing water quality, preventing pollution sources, and safeguarding the Kashmir Valley's water resources.

## Materials and methods

### Study area and sampling stations

The river Jhelum flows through the Kashmir Division's 140-km stretch, with most villages and towns located along its banks. The river originates from an elegant spring named "Verinag." The Lidder nallah, the greatest of all effluents and the source of the river Jhelum's headwaters, meets the river Jhelum at a distance of two kilometers on the right side of the mountain. Furthermore, the Sindh nallah is the second-largest tributary, entering it on the right bank at Shadipora. The river finally drains itself into Wular Lake at Banyari, where it is joined by the Arin and Madhumati streams.

Twenty-six sample stations on the riverbank of Jhelum and its adjoining tributaries in 2022 were selected based on the anticipated pollution strength and riverbed geology, as can be shown in Fig. 1, to achieve the investigation's aims and objectives. The samples were taken in the springtime (March–April), summertime (June–July), fall (September–October), and wintertime (December–January).

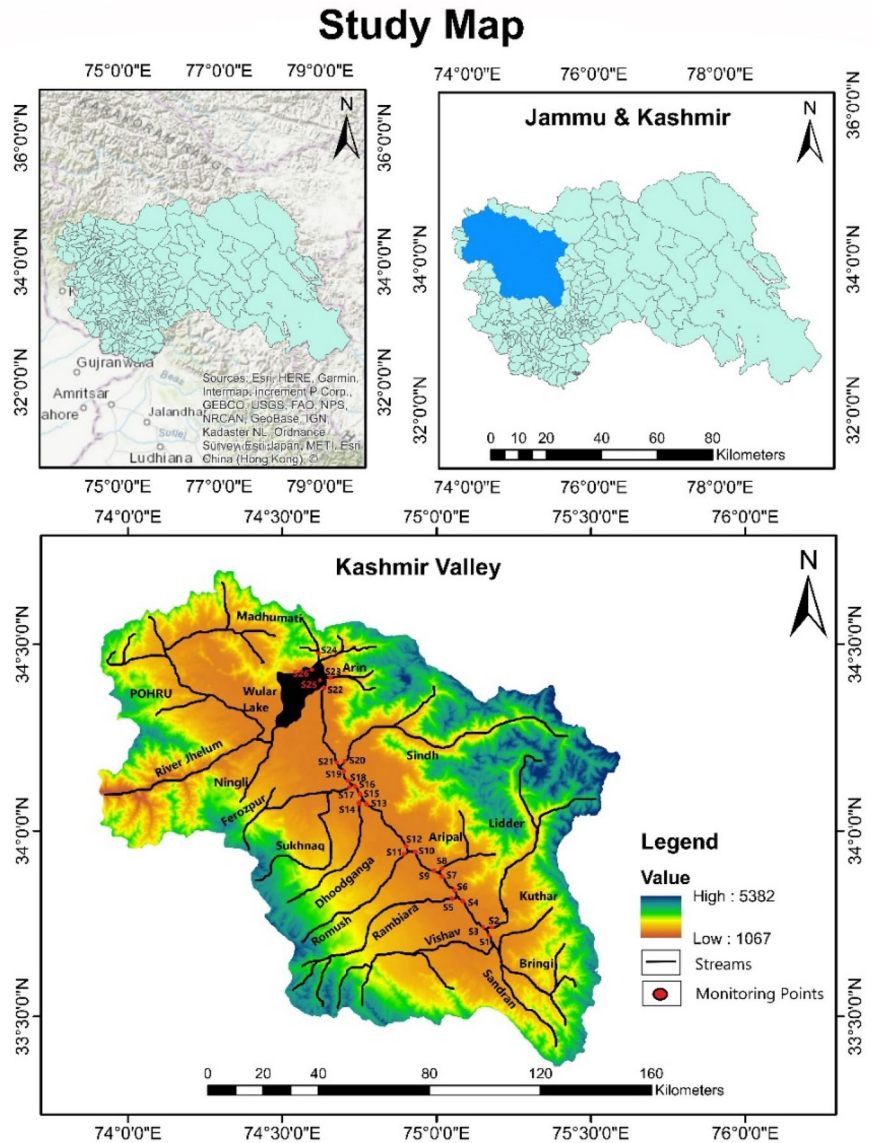
### Testing procedure

Surface water samples taken from 26 different locations during the year 2022 were analyzed during the four seasons of spring, summer, autumn, and winter. The collected samples were placed in 3-L airtight vials which were considered to be adequate for the sampling procedure. The samples were tested at the State Pollution Control Board's (SPCB) water laboratory in Srinagar and Tehkeek International's environmental laboratory in accordance with APHA guidelines (2017). Transparency, pH, and conductivity levels were measured instantly at the sample location while other parameters were evaluated in the laboratory. Table 1 lists the various analytical techniques used for estimating the values of water quality parameters.

### Statistical analysis

A variety of graphs have been utilized by the researchers to present visual overviews of data that highlight the data's significant information and provide an understanding of the data in a timely and effective manner (Mubarak et al., 2021; Whitlock et al., 2019). Graphs aid in determining whether or not more intricate modelling is required (Hohman et al., 2019). In this study, line diagrams and box plots were employed to summarize the dataset. For all parameters related to water quality, a two-way ANOVA (analysis of variance) was used at the 0.05% level of significance to identify significant differences between locations and seasons. ANOVA allows you to discover whether differences in mean values between two or more groups are by chance or if they are indeed significantly different (Oleson et al., 2019). A multivariate approach known as principal component analysis (PCA) was used to reduce the dataset's dimension and find trends in the backdrop of disorganized or perplexing data. The PCA approach defines the number of factors required to comprehend the observed variation in the data (Alavi et al., 2020). Eigenvalues and eigenvectors are calculated mathematically while computing PCs, using covariance or other cross-product matrices that represent the dispersion of the observed parameters and initial variables (Mikis et al., 2022). The relationship between datasets is explained by PCA in terms of underlying factors that are not immediately clear. For standard statistical analysis, the SPSS software (v.26) was employed. Using principal component analysis (PCA), we can identify the most significant underlying variables and patterns in a large dataset by reducing its degree of dimensionality (Nobre & Neves, 2019). The relationships between the various water quality parameters measured at several sites and across various seasons were examined in this study using PCA. The resulting principal components (PCs) can be seen as completely new variables that accurately represent the most significant patterns of data variance (Maity et al., 2022). The researchers are able to find greater detail regarding the underlying factors of the observed patterns of water quality by examining the loadings of the different variables on each PC. When analyzing huge datasets with several variables, PCA is very helpful since it enables us to simplify the data without losing crucial information (Goldrick et al., 2020).

**Fig. 1** Study area showing drainage network and sampling stations



## Results and discussions

The trends in water quality parameters observed along the Jhelum River's stretch from Vishav to Wular Lake (S1–S26) are noteworthy. The levels of pH, DO, hardness, calcium, chloride, alkalinity, transparency, ammoniacal nitrogen, and nitrate nitrogen have all been decreasing. This indicates that the water quality has deteriorated over time in terms of these parameters. Phosphorus, carbon dioxide, TDS, TSS, and conductivity, on the other side, have increased. Overall, the above trend indicates that water quality

in the study area has declined over time. Figure 2a–n depicts the regional and temporal variations in water quality, as well as a visual representation of the changes observed in each of the parameters over the study period.

The water quality of river Jhelum varies significantly along its course due to instream pollution and abrupt changes at input sampling stations viz., S2, S5, S8, S11, S14, S17, and S20, which reflect the water quality of adjacent tributaries. Furthermore, the water quality of the river Jhelum changes as it pauses in Wular Lake owing to its direct input sources such as

**Table 1** Water quality parameters and their analytical procedures

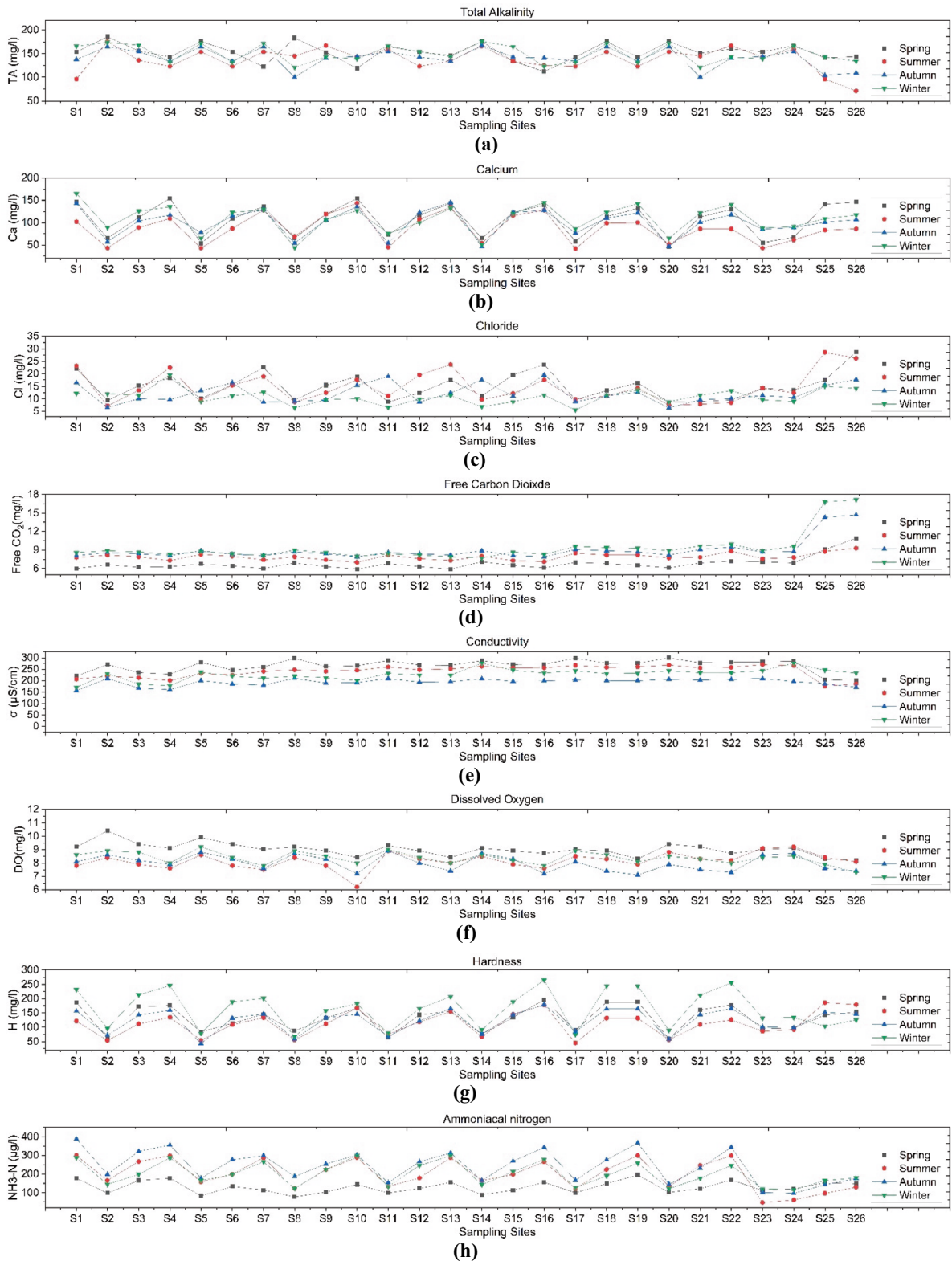
Parameters	Abbreviation	Method	Unit
Hardness	H	EDTA method	mg/l
Calcium	Ca	EDTA method	mg/l
Chloride	Cl	Argentometric titration	mg/l
Free carbon dioxide	CO <sub>2</sub>	Titration method	mg/l
Dissolved oxygen	DO	Winkler’s method	mg/l
Phosphorus	PO <sub>4</sub>	Perchloric acid method	µg/l
Ammoniacal nitrogen	NH <sub>3</sub> -N	Phenate-spectrophotometric method	µg/l
Nitrate nitrogen	NO <sub>3</sub> -N	Flotation-spectrophotometric method	µg/l
Total dissolved solids	TDS	Gravimetric analysis	mg/l
Total suspended solids	TSS	Gravimetric analysis	mg/l
H + concentration	pH	Digital pH meter	–
Transparency	T	Secchi disk	cm
Conductivity	EC	Conductivity meter	µS/cm
Total alkalinity	TA	Titrimetric method	mg/l

Arin and Madhumati, whose quality parameters are shown at S24 and S25. The water quality indicators of Wular Lake are shown at S25 and S26, and they differ significantly from the values at the river Jhelum’s upstream stretch.

Based on the line diagrams, it is evident that the river Jhelum’s tributaries cause its water quality to change dramatically over its entire course until it reaches Wular Lake. Due to the greater concentrations of dissolved organic matter and sediment yield generated by its input streams of water, TDS, and TSS levels in Wular Lake are at their peak values. Ammoniacal nitrogen and nitrate nitrogen levels are higher in all the sampling stations of river Jhelum, which may be due to the presence of residential and industrial areas near its banks, leading to a greater amount of animal and human waste being discharged into the river. Higher conductivity levels of the river Jhelum’s adjoining tributaries indicate that the water has greater levels of dissolved ions and mineralization. This is probably because more household and municipal waste effluents are being discharged into the river, especially when these streams pass through an urbanized area. Furthermore, the pH levels in the adjoining tributaries are higher as well. This could be due to interactions between different contaminants, such as chemicals, minerals, and pollution in the flowing water, as well as variations in the composition of the soil or bedrock. Despite these challenges, all of the nearby tributaries have far higher dissolved oxygen levels

than the river Jhelum. This is probably because of things like lower water temperature and lower levels of dissolved salt content. Additionally, side streams have low calcium and hardness levels, which may be caused by the lack of calcium-containing rocks and minerals like gypsum, dolomite, and limestone in the riverbeds of these tributaries. The total concentration of chloride ions does not vary significantly over the length of the river Jhelum; however, it is generally higher in Wular Lake sampling stations. The chloride ions are often present in the environment as NaCl, CaCl, KCl, and MgCl, and because they are mobile, they can be easily released into river systems near the source (Huang et al., 2021). Phosphorus levels are higher in Lidder (S2), Aripal (S11), Dhoodganga (S11), and Sindh (S20) compared to other sampling spots possibly due to severe bank erosion in these rivers and insufficient sewage disposal infrastructure in their watersheds. Increased CO<sub>2</sub> concentrations in Wular Lake indicate a significant level of decomposition of dead matter, which could be created by various types of water contaminants or natural processes. The line diagrams show the notable seasonal fluctuations at each of the 26 sampling points. DO, TSS, TDS, and conductivity readings are higher in the spring, while hardness, calcium, alkalinity, and carbon dioxide are higher in the winter. However, only a small number of the elements at some sites that were studied showed either a high concentration (positive peak) or a low concentration (negative peak) during





**Fig. 2 a–n** Spatial–temporal variation of water quality parameters during all four seasons at different sites

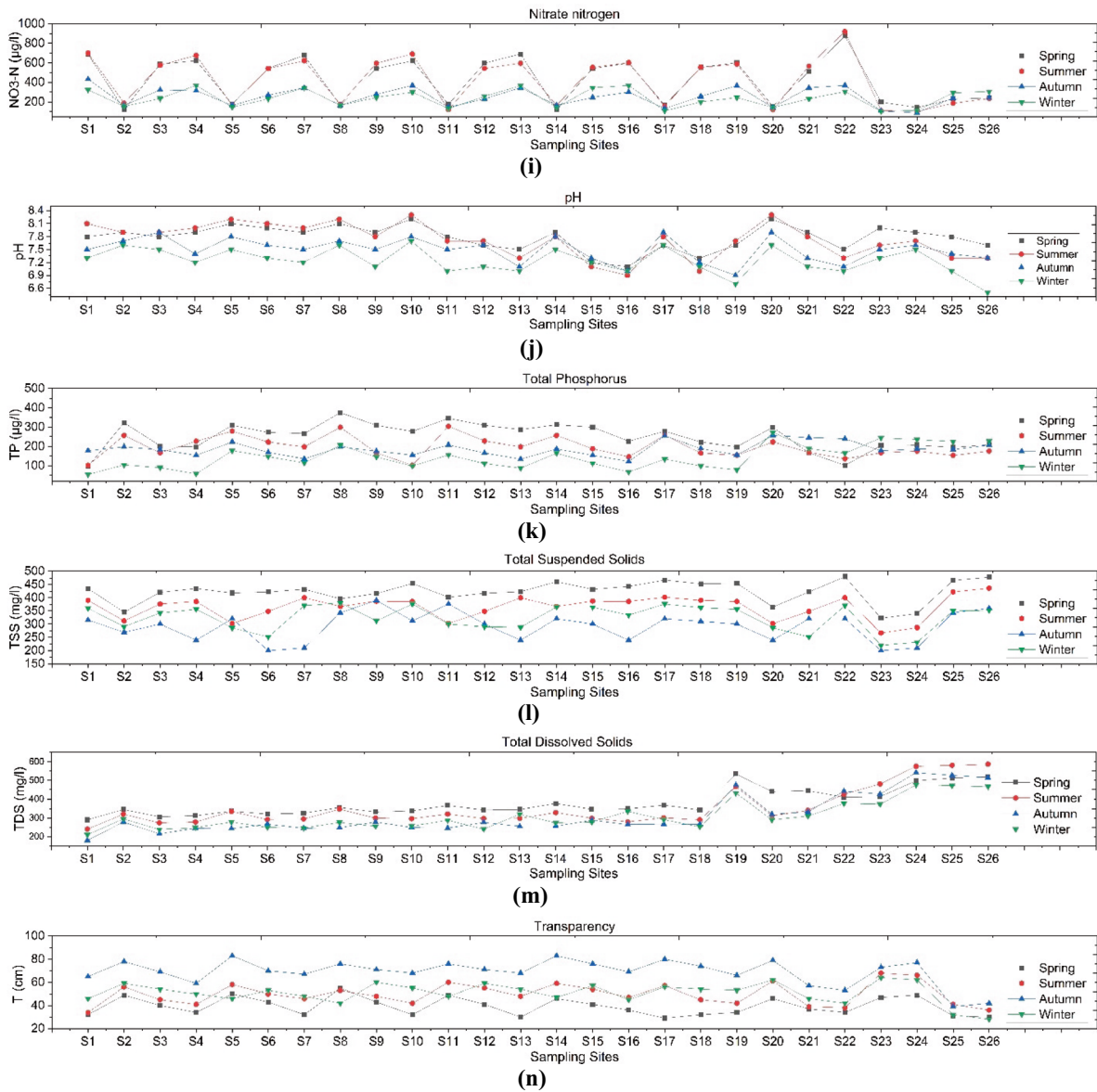


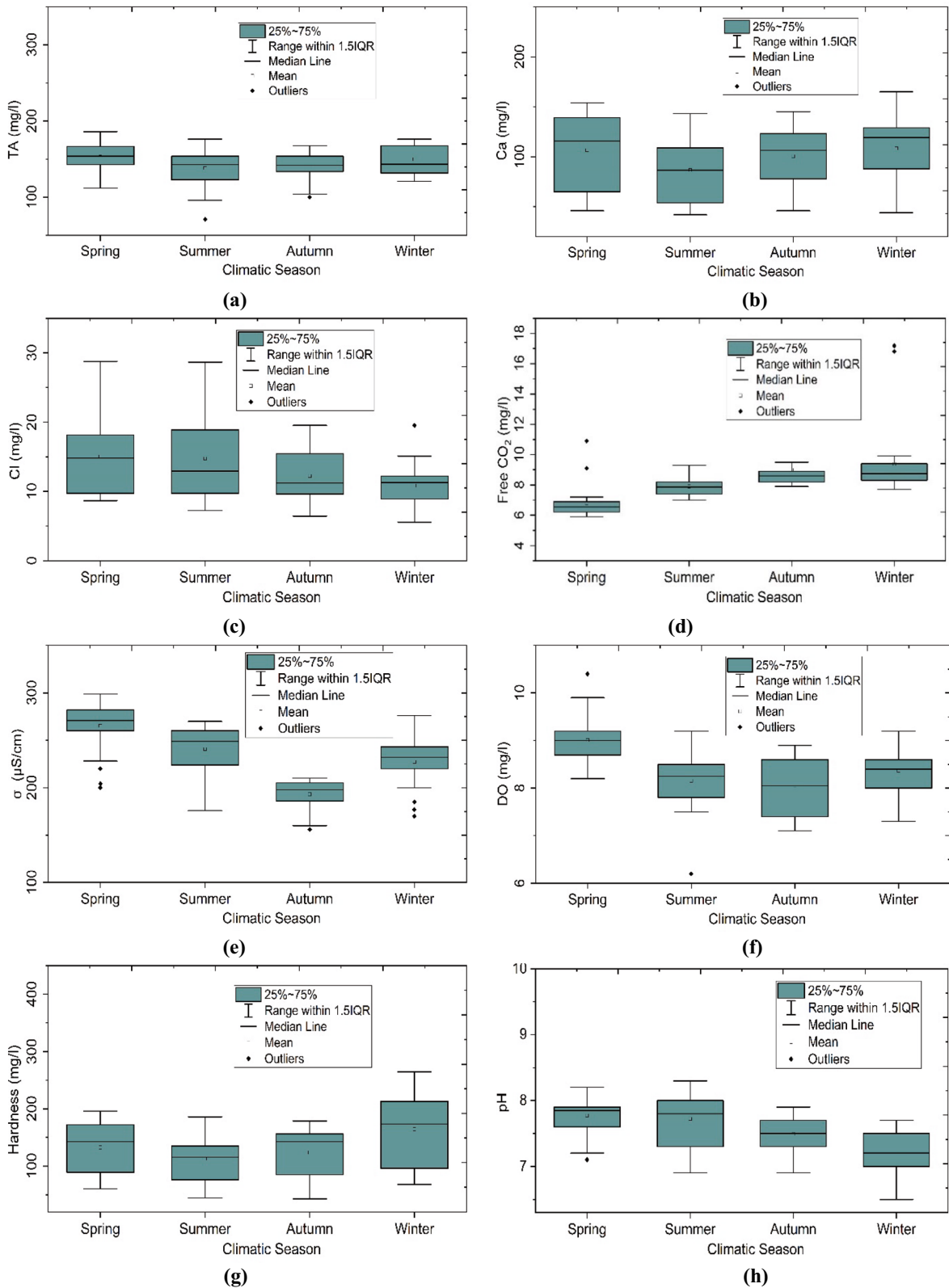
Fig. 2 (continued)

all of the four seasons. The box plots in Fig. 3a–n show the dataset’s extreme values, median, dispersion, skewness, and outliers.

ANOVA analysis

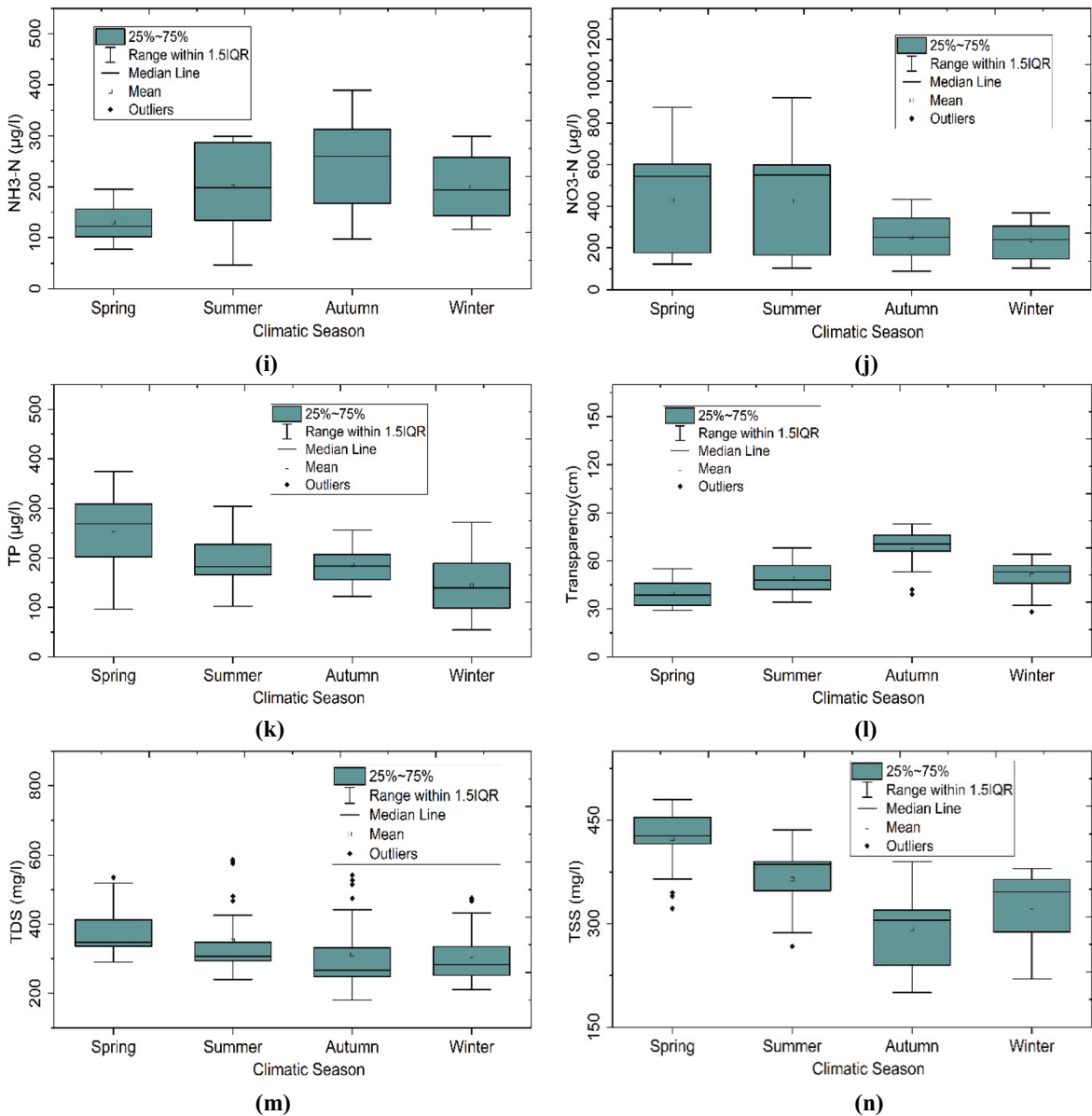
The two-way analysis of variance (ANOVA) was used in this study to evaluate the variability of the parameters impacting water quality. At a probability of 5%, the value of parameters with significant *F* was

compared between the stations as well as seasons. The results indicated a substantial difference between *F* and *F*-critical values for all samples and that the *P* value is nil in comparison to the alpha value (0.05) except for the phosphorus. The null hypothesis was rejected for all parameters, indicating significant variation in parameter values across all sampling stations and seasons. Table 2 displays the results of the two-way ANOVA analysis. Phosphorus is the only parameter for which the null hypothesis is accepted across



**Fig. 3 a–h** Box plots showing seasonal variation of various water quality parameters





**Fig. 3** (continued)

sample stations, showing that there is little variation in values across the twenty-six sampling sites.

**Factor analysis**

Factor analysis is a statistical technique that is used to analyze data with multiple dependent interactions between variables. The purpose of factor analysis is to identify a small number of underlying factors that can

explain the interdependence between the variables (Tokatli et al., 2021). These factors, or abstract indicators, can capture a significant amount of information that is reflected by many of the original variables (Uncumusoglu & Mutlu, 2022). This approach can provide a scientific basis for decision-making by analyzing and evaluating data in a rational and systematic way. In essence, factor analysis seeks to extract information from a large number of variables, reduce

**Table 2** Results of two-way ANOVA analysis

Parameter	Source of variation	Sum of squares	Degree of freedom	Mean squares	F-value	F-critical	P-value
pH	Seasons	7.9578	25	0.3183	10.29	1.65	0
	Stations	4.678	3	1.5593	50.41	2.72	0
NO3-N	Seasons	1,283,238.90	25	102,043.13	9.90	1.65	0
	Stations	872,659.33	3	290,886.44	28.23	2.72	0
TDS	Seasons	736,240.77	25	29,449.63	35.9117	1.65	0
	Stations	88,360.41	3	29,453.47	35.9163	2.72	0
DO	Seasons	19.26	25	0.77	6.34	1.65	0
	Stations	15.61	3	5.20	42.87	2.72	0
Conductivity	Seasons	42,283.11	25	1691.32	8.42	1.65	0
	Stations	70,317.30	3	23,439.10	116.78	2.72	0
TSS	Seasons	153,184.00	25	6127.36	6.40	1.65	0
	Stations	248,799.76	3	82,933.25	86.66	2.72	0
CO <sub>2</sub>	Seasons	182.89	25	7.31	9.66	1.65	0
	Stations	105.80	3	35.26	46.61	2.72	0
Hardness	Seasons	201,935.46	25	8077.41	15.35	1.65	0
	Stations	38,272.34	3	12,757.44	24.24	2.72	0
Calcium	Seasons	92,114.58	25	3684.58	22.40	1.65	0
	Stations	7566.49	3	2522.16	15.33	2.72	0
Alkalinity	Seasons	26,635.46	25	1065.41	4.71	1.65	0
	Stations	3985.61	3	1328.53	5.87	2.72	0
Phosphorus	Seasons	155,994.03	25	6239.76	2.72	1.65	0.06
	Stations	145,515.42	3	48,505.14	21.20	2.72	0
Chloride	Seasons	1355.26	25	54.21	4.58	1.65	0
	Stations	322.89	3	107.63	9.09	2.72	0
NH <sub>3</sub> -N	Seasons	375,205.65	25	15,008.22	12.31	1.65	0
	Stations	170,065.15	3	56,688.38	46.50	2.72	0
Transparency	Seasons	6169.11	25	246.76	7.32	1.65	0
	Stations	11,718.80	3	3906.26	115.88	2.72	0

it into a few factors, and minimize the loss of information in the process (Sellbom & Tellegen, 2019). Principal component analysis (PCA) was employed in this study to analyze 14 water quality parameters at twenty-six monitoring sites in the study area. The Kaiser–Meyer–Olkin (KMO) and Barlett tests were done before the analysis to check the suitability of PCA. The KMO value was 0.699, which was higher than the recommended minimum of 0.5 (Elsaman et al., 2022), indicating that the data was appropriate for PCA. The Barlett test value was 0.00, which is less than the statistical significance level of 0.05, showing that the variables were independent and appropriate for PCA.

To analyze the data, this study used SPSS (v. 26.0) software from IBM located in Armonk, NY, USA. The original monitoring data was standardized before generating a correlation coefficient matrix. The

descriptive statistics of the experimental results are presented in Table 3, while the correlation matrix is shown in Table 4.

The typical factor analysis levels are shown in Table 5, which includes the initial communalities that have a value of 1 for all parameters. After factor extraction, the communalities of the variables are displayed in the third column. The table shows that the communalities of TDS, TSS, hardness, calcium, DO, alkalinity, calcium, and transparency are all high (>0.80), indicating that they all provide comprehensive information. The communalities of pH, conductivity, CO<sub>2</sub>, phosphorus, chloride, NH<sub>3</sub>-N, and NO<sub>3</sub>-N, on the other hand, are low (<0.80), indicating that they provide insufficient information.

Table 6 presents the results of a factor analysis that shows each of the four common factors has eigenvalues greater than one, and together, they account

**Table 3** Descriptive statistics

Parameter	Range	Minimum	Maximum	Mean		Std. deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Std. error		
pH	1.80	6.50	8.30	7.5567	0.03736	0.38105	0.145
TDS	407.00	180.00	587.00	338.4327	9.09510	92.75218	8602.966
EC	143.00	156.00	299.00	231.6154	3.45207	35.20434	1239.346
TSS	280.00	200.00	480.00	351.0000	6.65033	67.82029	4599.592
DO	4.20	6.20	10.40	8.3817	0.06407	0.65341	0.427
CO <sub>2</sub>	11.28	5.90	17.18	8.2635	0.17958	1.83134	3.354
H	222.00	43.00	265.00	133.4808	5.10955	52.10743	2715.184
Ca	123.00	42.00	165.00	100.6058	3.23369	32.97731	1087.503
TA	115.00	71.00	186.00	145.7308	2.10769	21.49429	462.004
PO <sub>4</sub>	321.00	54.00	375.00	194.0962	6.64536	67.76967	4592.729
Cl	23.20	5.56	28.76	13.2229	0.48938	4.99074	24.907
NH <sub>3</sub> -N	342.00	47.00	389.00	193.3077	7.70953	78.62212	6181.438
NO <sub>3</sub> -N	832.00	89.00	921.00	335.6635	19.79251	201.84482	40,741.332
T	55.00	28.00	83.00	52.1346	1.38055	14.07887	198.215

for 75.182% of the total variance. Only the first four components are extracted and rotated, and the variance of the original variables of the multiple factors is redistributed by the factor, which brings the variance of the factors closer together. This implies that these four factors represent the fundamental elements of the original data.

Additionally, the scree plot shown in the Fig. 4 can be used for understanding the underlying data structure. It displays information about the eigenvalues of all factors, which helps in identifying the ideal number of primary components based on the selection principles of principal component analysis (Isiyaka et al., 2019). In this study, it was observed that the slope of the scree plot significantly flattened after the fourth component. Therefore, the first four main components, which had eigenvalues greater than one and accounted for maximum of the dataset’s variance, were retained. The variation in eigenvalues is greater between factors 1 and 2, 3 and 4, and 4 and 5. However, the difference between factors 5 and 6 and beyond is minimal. This implies that the top four factors include more reliable general information, and they could be considered the primary composition factors to represent all the 14 variables.

Table 7 displays the factor loading matrix, which shows the load each variable has on principal components before and after rotation; it is clear that there is a significant polarization in the loading factors after rotation.

It is evident from the results that the first rotated principal component (PC1) and second rotated principal component (PC2) have a strong link with soil erosion and natural pollution, respectively. PC1 has a substantial positive loading on TDS (0.877) and TSS (0.852), and it accounts for 28.863% of the overall data variance. The second principal component (PC2), which accounted for 22.715% of the overall variation, had a significant positive loading on EC (0.849) and hardness (0.843). These findings imply that the most severe pollution problem in the stream is caused by significant sediment production as a result of numerous factors that include bank erosion, agricultural runoff, and watershed disintegration. Seasonal fluctuations, which have a direct impact on water quality, could also be responsible for the presence of these pollutants and signs of soil erosion in the water. For instance, excessive rain can cause topsoil to erode and become contaminated as a result of animal and human contact. As a result, this polluted soil can change the physicochemical and microbiological characteristics of the surface water resources.

In addition to natural factors, human activities also contribute to water pollution. For instance, land use activities such as agriculture, industrial activities, and improper waste disposal can lead to the emission of pollutants that interact with water during runoff. This interaction can result in the transport of pollutants into water bodies, further compromising water quality. Overall, the information provided suggests

**Table 4** Correlation matrix

Parameter	pH	TDS	EC	TSS	DO	CO <sub>2</sub>	H	Ca	TA	PO <sub>4</sub>	Cl	NH <sub>3</sub> -N	NO <sub>3</sub> -N	T
pH	1.000													
TDS	-0.065	1.000												
EC	0.237	0.254	1.000											
TSS	0.183	0.184	0.395	1.000										
DO	0.338	0.062	0.490	0.202	1.000									
CO <sub>2</sub>	-0.445	0.281	-0.374	-0.243	-0.383	1.000								
H	-0.555	-0.015	-0.278	0.137	-0.375	0.021	1.000							
Ca	-0.389	-0.105	-0.299	0.203	-0.440	-0.011	0.831	1.000						
TA	0.199	-0.189	0.334	-0.108	0.422	-0.226	-0.271	-0.258	1.000					
PO <sub>4</sub>	0.441	0.189	0.479	0.199	0.399	-0.154	-0.598	-0.456	0.160	1.000				
Cl	-0.034	0.220	-0.171	0.410	-0.192	-0.034	0.358	0.437	-0.487	-0.082	1.000			
NH <sub>3</sub> -N	-0.327	-0.421	-0.570	-0.211	-0.712	0.068	0.483	0.547	-0.196	-0.571	0.081	1.000		
NO <sub>3</sub> -N	0.018	-0.097	0.127	0.505	-0.262	-0.323	0.404	0.583	-0.211	-0.159	0.417	0.376	1.000	
T	0.015	-0.335	-0.390	-0.722	-0.079	0.021	-0.313	-0.334	0.193	-0.066	-0.389	0.155	-0.539	1.000

**Table 5** Communalities

Parameter	Initial	Extraction
pH	1.000	0.775
Total dissolved solids	1.000	<b>0.811</b>
Conductivity	1.000	0.611
Total suspended solids	1.000	<b>0.849</b>
Dissolved oxygen	1.000	<b>0.876</b>
Carbon dioxide	1.000	0.740
Hardness	1.000	<b>0.842</b>
Calcium	1.000	<b>0.817</b>
Alkalinity	1.000	<b>0.814</b>
Total phosphorus	1.000	0.637
Chloride	1.000	0.685
Ammoniacal nitrogen	1.000	0.729
Nitrate nitrogen	1.000	0.785
Transparency	1.000	<b>0.806</b>

that natural and human activities can both contribute to water pollution and erosion. Therefore, there is a need for proper land use management and waste disposal practices to protect water resources from contamination.

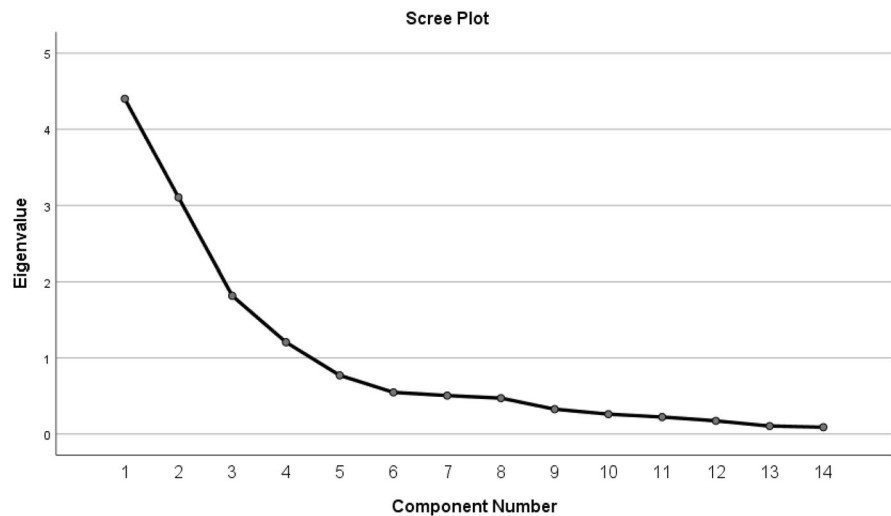
The third rotated principal component (PC3) accounts for 15.516% of the total data variance and is associated with both anthropogenic and geogenic sources of pollution. This PC has a strong positive loading on NO<sub>3</sub>-N (0.888), NH<sub>3</sub>-N (0.848), and Cl (0.871) and has a strong negative loading on pH (-0.899). The positive loadings of NO<sub>3</sub>-N and NH<sub>3</sub>-N may be attributed to anthropogenic influences such as industrial and domestic waste. Similarly, the high concentration of chloride in natural water sources could be a result of human activities, such as domestic waste disposal, agriculture, and industry-based activities. On the other hand, the negative loading of pH may be associated with organic matter oxidation resulting from anthropogenic activities. Overall, this principal component is associated with both anthropogenic pollution and geogenic sources. Therefore, it is crucial to implement effective strategies for managing human activities and protecting water resources to prevent further contamination.

The fourth rotated principal component (PC4) accounts for 8.089% of the total data variance and has a positive loading on pH (0.853) and strong negative loading on DO (-0.915). The relationship between a river water pH and other water quality indices is

**Table 6** Total variance explained

Initial eigenvalues				Extraction sums of squared loadings			Rotation sums of squared loadings		
C. No	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.618	28.863	28.863	4.618	28.863	28.863	3.648	22.802	22.802
2	3.634	22.715	51.577	3.634	22.715	51.577	3.638	22.738	45.541
3	2.483	15.516	67.094	2.483	15.516	67.094	2.478	15.487	61.028
4	1.294	8.089	75.182	1.294	8.089	75.182	2.265	14.154	75.182
5	0.815	5.096	80.278						
6	0.715	4.467	84.746						
7	0.516	3.222	87.968						
8	0.481	3.008	90.976						
9	0.331	2.069	93.045						
10	0.284	1.777	94.823						
11	0.258	1.613	96.436						
12	0.196	1.225	97.660						
13	0.106	0.659	98.320						
14	0.104	0.651	98.971						
15	0.094	0.585	99.556						
16	0.071	0.444	100.000						

**Fig. 4** Scree plot



complex, but it is claimed that toxic pollution from industrial manufacturing may be a contributing factor. The negative loading and DO suggest that pollution sources are of anthropogenic origin. Anthropogenic pollution sources can include wastewater discharges, agricultural runoff, and industrial activities, all of which can contribute decreased DO in water bodies. Low levels of DO can also have negative impacts on aquatic life, leading to reduced biodiversity and decreased water quality. Therefore, effective

management of anthropogenic pollution sources is necessary to prevent further degradation of water quality and protect aquatic ecosystems.

*Factor score*

The factor scores for various sampling stations were calculated using the SPSS statistical package (v 26). To do this, open the data editing box in SPSS, copy, and paste the four columns of component matrix data from



**Table 7** Component matrix and rotated component matrix

Parameter	Component matrix				Rotated component matrix			
	PC (1)	PC (2)	PC (3)	PC (4)	PC (1)	PC (2)	PC (3)	PC (4)
PH	-0.566	0.214	0.331	-0.547	0.757	0.122	<b>-0.899</b>	<b>0.853</b>
TDS	-0.131	0.334	-0.756	0.105	<b>0.877</b>	0.301	-0.022	-0.767
EC	-0.582	0.549	0.066	0.342	0.455	<b>0.849</b>	0.659	-0.083
TSS	0.003	0.865	0.020	-0.019	<b>0.852</b>	0.122	0.069	0.007
DO	-0.730	0.248	0.117	0.261	0.469	0.125	0.664	<b>-0.915</b>
CO <sub>2</sub>	0.241	-0.369	-0.734	0.087	-0.213	-0.331	-0.302	-0.703
H	0.815	0.206	0.053	0.364	0.337	<b>0.843</b>	-0.136	-0.006
Ca	0.825	0.286	0.172	0.159	-0.739	0.419	-0.260	0.166
TA	-0.497	-0.156	0.433	0.504	0.055	-0.233	0.774	0.240
PO <sub>4</sub>	-0.697	0.261	-0.123	-0.262	0.761	0.142	0.173	-0.092
Cl	0.442	0.529	-0.210	-0.407	-0.038	0.594	<b>0.871</b>	-0.065
NH <sub>3</sub> -N	0.776	-0.295	0.357	-0.112	-0.647	-0.159	<b>0.848</b>	0.429
NO <sub>3</sub> -N	0.481	0.636	0.372	-0.105	-0.288	<b>0.710</b>	<b>0.888</b>	0.404
T	-0.167	-0.844	0.181	-0.182	0.103	-0.858	-0.078	0.231

Values in the bold indicate substantial loading

Table 7 and label the variables as  $\alpha$ ,  $\beta$ ,  $\mu$ , and  $l$ . Then, use “transform → compute variable” to input the formula “ $A = \alpha / \text{SQR}(Xi)$ ,  $B = \beta / \text{SQR}(Xi)$ ,  $C = \mu / \text{SQR}(Xi)$ ,  $D = l / \text{SQR}(Xi)$ ,” where  $Xi$  represents the eigenvalue for each principal component, as shown in Table 6, and  $A$ ,  $B$ ,  $C$ , and  $D$  represent the corresponding eigenvectors. Multiplying the standardized data by these eigenvectors produces the four extracted principal component expressions.

$$F1 = \sum_{i=1}^{26} A_i Z y_i \tag{1}$$

$$F2 = \sum_{i=1}^{26} B_i Z y_i \tag{2}$$

$$F3 = \sum_{i=1}^{26} C_i Z y_i \tag{3}$$

$$F4 = \sum_{i=1}^{26} D_i Z y_i \tag{4}$$

where  $F$  is the principal component score;  $A$ ,  $B$ ,  $C$ , and  $D$  is the corresponding eigenvectors;  $Z$  is the standardized  $Z$  score; and  $y$  is the corresponding principal component value.

The function  $F$  for comprehensive evaluation can be calculated using formula (5), taking into account the varying weights of variance for the four primary components ( $e1$ ,  $e2$ ,  $e3$ , and  $e4$ ).

$$\begin{aligned}
 F &= F1 \times \frac{e1}{e1 + e2 + e3 + e4} + F2 \times \frac{e2}{e1 + e2 + e3 + e4} \\
 &+ F3 \times \frac{e3}{e1 + e2 + e3 + e4} + F4 \times \frac{e4}{e1 + e2 + e3 + e4} \\
 &= 0.38390F1 + 0.30210F2 + 0.20641F3 + 0.10757F4 \tag{5}
 \end{aligned}$$

The results obtained using the above formulas are shown in Table 8. The results in the table reveal that of the 26 sampling stations, S5 has the poorest water quality, and S1 is the least polluted stretch. Comprehensive data indicate that the pollution status of 26 sections analyzed using evaluation function  $F$  is in the following order: site 5 > site 14 > site 25 > site 20 > site 17 > site 19 > site 26 > site 22 > site 11 > site 8 > site 15 > site 23 > site 24 > site 21 > site 2 > site 18 > site 7 > site 10 > site 16 > site 4 > site 6 > site 12 > site 13 > site 9 > site 3 > site 1.

### Conclusions

The goal of this study was to analyze the overall quality of surface water in the Kashmir using standard testing procedures and multivariate statistical tools. The study is also aimed at assessing the impact of adjoining tributaries on the water quality of the river Jhelum and Wular Lake. Water quality monitoring was carried out at 26 sampling locations

**Table 8** Water quality principal component score and order of 11 sampling stations

Site	F1	F2	F3	F4	F	Order
Site 1	-1.2135	0.2439	-0.2765	-0.0043	-0.44972	26
Site 2	0.2659	-0.6544	0.8765	0.8976	0.181858	15
Site 3	-0.8123	0.0376	0.1012	0.4740	-0.22861	25
Site 4	-0.0657	0.3164	-0.2475	-0.5644	-0.04144	20
Site 5	1.4875	-0.6654	0.4188	0.6654	0.528055	<b>1</b>
Site 6	-0.3211	0.0189	0.3122	-0.1223	-0.06628	21
Site 7	-0.4564	0.7998	0.4889	-0.8766	0.073026	17
Site 8	0.6743	-0.0234	-0.0654	0.0987	0.248913	<b>10</b>
Site 9	-0.4377	-0.0699	-0.0094	0.0321	-0.18764	24
Site 10	-0.6592	0.5432	0.5466	0.0065	0.024557	18
Site 11	0.6588	0.0094	-0.0076	-0.0876	0.244761	<b>9</b>
Site 12	0.0034	-0.2389	-0.0999	-0.0090	-0.09245	22
Site 13	-0.0784	-0.6895	0.0322	0.8766	-0.13745	23
Site 14	0.3426	1.1099	0.0233	0.3987	0.51452	<b>2</b>
Site 15	0.0098	1.0022	0.0321	-0.8777	0.218738	11
Site 16	0.0065	0.0098	0.0755	-0.0965	0.010659	19
Site 17	1.0986	-0.0087	0.0087	0.1009	0.431774	<b>5</b>
Site 18	-0.0654	0.0789	0.5465	0.0076	0.112349	16
Site 19	-0.0056	0.8765	0.4534	0.0054	0.356808	<b>6</b>
Site 20	0.1022	1.3022	-0.0019	-0.0876	0.45993	<b>4</b>
Site 21	0.4321	-0.7687	0.6777	1.1088	0.192817	14
Site 22	-0.0433	1.0098	-0.0086	-0.0654	0.279628	<b>8</b>
Site 23	0.9236	-0.7174	0.2443	0.2332	0.213355	12
Site 24	0.8112	-0.7011	0.4162	0.0876	0.194948	13
Site 25	1.1209	0.8098	-1.112	0.2213	0.469231	<b>3</b>
Site 26	1.1133	0.9716	-1.1933	-1.1102	0.355183	<b>7</b>

across all four seasons in 2022 to investigate spatial and temporal variations. Line diagrams and box plots revealed that some parameters exceeded allowable limits at certain points during particular sampling seasons, rendering the water unsafe for drinking, farming, fishing, or other domestic purposes (Cotruvo, 2017). The increasing amount of side-stream pollution was an indicator of increased human influence in the core watersheds of Kashmir.

A two-way ANOVA analysis was performed on the sampled parameters to determine any seasonal or sectional variations. The results revealed significant spatiotemporal variability. The principal component analysis (PCA) method was employed to identify the most important indicator parameters affecting water quality and potential sources of pollution. PCA simplifies the high-dimensional variable system by integrating and optimizing the retention of the original data information. Four significant

principal components were extracted from the 14 water quality measures, accounting for 75.18% of the variation in the initial dataset. PC1 (28.86%) and PC2 (22.71%) represented chemical and conventional pollutants which indicates the influence of bank erosion and other deposited sediments on water quality. PC3 (15.51%) showed a positive correlation with DO and total alkalinity, representing pollution due to bed decomposition and industrial and residential wastes. PC4 (8.08%) had positive loadings of pH, representing toxic pollution from municipal areas surrounding surface water resources.

This study utilized various techniques to assess and interpret sources of pollutants and fluctuations in water quality in the river Jhelum and its tributaries. The aim was to enhance decision-making and efficient management of water resources. The study's outcomes could stimulate thoughtful considerations and lead to improved management practices for

surface water resources in the Kashmir valley, benefiting the ecology and environment.

**Author contribution** Sarvat Gull wrote the literature review, methodology, and results. This author carried out original manuscript preparation. Shagoofa Rasool Shah is behind conceptualization, investigation, manuscript editing, and supervision. Ayaz Mohmood Dar carried out statistical analysis and graphical explanation. Further, all authors reviewed the manuscript.

**Data availability** The authors certify that the data supporting this study's conclusions are included in the publication.

#### Declarations

**Ethics approval** All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the instructions for authors and are aware that with minor exceptions, no changes can be made to authorship once the paper is submitted.

**Conflict of interest** The authors declare no competing interests.

#### References

- Akhtar, N., SyakirIshak, M. I., Bhawani, S. A., & Umar, K. (2021). Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water*, *13*(19), 2660.
- Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use. *Journal of Advanced Nursing*, *76*(8), 1886–1889.
- American Public Health Association (APHA). (2017). *Standard methods for examination of water and wastewater* (23rd ed.). APHA.
- Aydin, H., Ustaoglu, F., Tepe, Y., & Soylu, E. N. (2021). Assessment of water quality of streams in northeast Turkey by water quality index and multiple statistical methods. *Environmental Forensics*, *22*(1–2), 270–287.
- Bowman, N. D., & Goodboy, A. K. (2020). Evolving considerations and empirical approaches to construct validity in communication science. *Annals of the International Communication Association*, *44*(3), 219–234.
- Cotruvo, J. A. (2017). WHO guidelines for drinking water quality: first addendum to the fourth edition. *Journal-American Water Works Association*, *109*(7), 44–51.
- Elsaman, H. A., El-Bayaa, N., & Kousihan, S. (2022). Measuring and validating the factors influenced the SME business growth in Germany—descriptive analysis and construct validation. *Data*, *7*(11), 158.
- Ganaie, T. A., Jamal, S., & Ahmad, W. S. (2021). Changing land use/land cover patterns and growing human population in Wular catchment of Kashmir Valley, India. *Geojournal*, *86*(4), 1589–1606.
- Goldrick, S., Sandner, V., Cheeks, M., Turner, R., Farid, S. S., McCreath, G., & Glassey, J. (2020). Multivariate data analysis methodology to solve data challenges related to scale-up model validation and missing data on a micro-bioreactor system. *Biotechnology Journal*, *15*(3), 1800684.
- Hanjra, M. A., Blackwell, J., Carr, G., Zhang, F., & Jackson, T. M. (2012). Wastewater irrigation and environmental health: Implications for water governance and public policy. *International Journal of Hygiene and Environmental Health*, *215*(3), 255–269.
- Hohman, F., Park, H., Robinson, C., & Chau, D. H. P. (2019). S ummit: Scaling deep learning interpretability by visualizing activation and attribution summarizations. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 1096–1106.
- Huang, X., Ding, C., Su, Q., Wang, Y., Cui, Z., Yin, Q., & Wang, X. (2021). A simplified method for determination of short-, medium-, and long-chain chlorinated paraffins using tetramethyl ammonium chloride as mobile phase modifier. *Journal of Chromatography A*, *1642*(1), 462002.
- Isiyaka, H. A., Mustapha, A., Juahir, H., & Phil-Eze, P. (2019). Water quality modelling using artificial neural network and multivariate statistical techniques. *Modeling Earth Systems and Environment*, *5*(2), 583–593.
- Keyzers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799.
- Khanday, S. A., Bhat, S. U., Islam, S. T., & Sabha, I. (2021). Identifying lithogenic and anthropogenic factors responsible for spatio-seasonal patterns and quality evaluation of snow melt waters of the River Jhelum Basin in Kashmir Himalaya. *CATENA*, *196*(1), 104853.
- Lee, S., Park, J. R., Joo, J. C., & Ahn, C. H. (2022). Application of WQI<sub>EUT</sub> and TSI<sub>KO</sub> for comprehensive water quality assessment immediately after the construction of the Yeongju Multipurpose Dam in the Naeseong Stream Basin, Republic of Korea. *Science of the Total Environment*, *819*(1), 152997.
- Machiwal, D., Cloutier, V., Güler, C., & Kazakis, N. (2018). A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environmental Earth Sciences*, *77*(19), 1–30.
- Maity, S., Maiti, R., & Senapati, T. (2022). Evaluation of spatio-temporal variation of water quality and source identification of conducive parameters in Damodar River. *India. Environmental Monitoring and Assessment*, *194*(4), 308.
- Mikis, D. S., Robert, A. R., Nikolaos, G., & Fernanda, D. B. (2022). Principal component regression in GAMLSS applied to Greek-German government bond yield spreads. *Statistical Modelling*, *22*(1–2), 127–145.
- Mir, R. A., & Gani, K. M. (2019). Water quality evaluation of the upper stretch of the river Jhelum using multivariate statistical techniques. *Arabian Journal of Geosciences*, *12*(14), 1–19.
- Mubarak, A. A., Cao, H., Zhang, W., & Zhang, W. (2021). Visual analytics of video-clickstream data and prediction of learners' performance using deep learning models in MOOCs' courses. *Computer Applications in Engineering Education*, *29*(4), 710–732.

- Mutlu, E. (2021). Determination of seasonal variations of heavy metals and physicochemical parameters in Kildir pond yildizeli - sivas. *Fresenius Environmental Bulletin*, 30(6), 5773–5780.
- Mutlu, E., & Kurnaz, A. (2017). Determination of seasonal variations of heavy metals and physicochemical parameters in Sakiz Pond (Kastamonu-Turkey). *Fresenius Environmental Bulletin*, 26(4), 2806–2815.
- Najar, I. A., & Khan, A. B. (2012). Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis. *Environmental Earth Sciences*, 66(8), 2367–2378.
- Njuguna, S. M., Onyango, J. A., Githaiga, K. B., Gituru, R. W., & Yan, X. (2020). Application of multivariate statistical analysis and water quality index in health risk assessment by domestic use of river water. Case study of Tana River in Kenya. *Process Safety and Environmental Protection*, 133(1), 149–158.
- Nobre, J., & Neves, R. F. (2019). Combining principal component analysis, discrete wavelet transforms and XGBoost to trade in the financial markets. *Expert Systems with Applications*, 125(1), 181–194.
- Oleson, J. J., Brown, G. D., & McCreery, R. (2019). The evolution of statistical methods in speech, language, and hearing sciences. *Journal of Speech, Language, and Hearing Research*, 62(3), 498–506.
- Qayoom, U., Bhat, S. U., Ahmad, I., & Kumar, A. (2022). Assessment of potential risks of heavy metals from wastewater treatment plants of Srinagar city, Kashmir. *International Journal of Environmental Science and Technology*, 19(9), 9027–9046.
- Schreiber, S. G., Schreiber, S., Tanna, R. N., Roberts, D. R., & Arciszewski, T. J. (2022). Statistical tools for water quality assessment and monitoring in river ecosystems—a scoping review and recommendations for data analysis. *Water Quality Research Journal*, 57(1), 40–57.
- Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428.
- Showqi, I., Rashid, I., & Romshoo, S. A. (2014). Land use land cover dynamics as a function of changing demography and hydrology. *GeoJournal*, 79(3), 297–307.
- Tokatli, C., Mutlu, E., & Arslan, N. (2021). Assessment of the potentially toxic element contamination in water of Şehriban Stream (Black Sea Region, Turkey) by using statistical and ecological indicators. *Water Environment Research*, 93(10), 2060–2071.
- Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670.
- Uncumusaoglu, A. A., & Mutlu, E. (2022). Water quality index and multivariate statistical approach in assessing the quality of irrigation water of The Caykoy Pond. *Fresenius Environmental Bulletin*, 31(3 A), 3447–3459.
- Whitlock, M., Wu, K., & Szafir, D. A. (2019). Designing for mobile and immersive visual analytics in the field. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 503–513.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.