Journal of Human Environment and Health Promotion

Print ISSN: 2476-5481 Online ISSN: 2476-549X

Classification of Water Quality in the Shafarood River Based on Physicochemical Parameters and Heavy Metal Concentration Using Multivariate Analysis Methods

Elham Gholami Deljomanesh [a], Mahsa Hakimi Abed [a *], Ebrahim Fataei [b **], Fatemeh Shariati Feyzabadi [a], Ali Akbar Imani [c]

a. Department of Environment, Lahijan Branch, Islamic Azad University, Lahijan, Iran.
b. Department of Environment, Ardabil Branch, Islamic Azad University, Ardabil, Iran.
c. Department of Agriculture, Ardabil Branch, Islamic Azad University, Ardabil, Iran.

***Corresponding author:** * Mahsa Hakimi Abed: Department of Environment, Lahijan Branch, Islamic Azad University, Lahijan, Iran. Postal Code: 39515-44169. E-mail: hakimimah@gmail.com; ** Ebrahim Fataei: Department of Environment, Ardabil Branch, Islamic Azad University, Ardabil, Iran. Postal Code: 5615731567. E-mail: eb.fataei@iau.ac.ir

A R T I C L E I N F O

Article type:
Original article

Article history:
Received: 21 January 2025
Revised: 1 March 2025
Accepted: 3 April 2025

https://doi.org/10.61186/jhehp.11.2.129

Keywords:
Water quality
Statistical analysis
Cluster analysis
Discriminant analysis
Principal component analysis
Factor analysis

A B S T R A C T

Background: The present study aimed to classify the water quality of the Shafarood River in Gilan Province, Iran, using multivariate analysis methods.
Methods: This research employed Cluster Analysis (CA), Discriminant Analysis (DA), Principal Component Analysis (PCA), and Factor Analysis (FA) as effective methodologies for decision-making in river water quality management. The analysis was based on selected data from physicochemical parameters and heavy metal concentrations measured at five water sampling stations over a six-year period (2017-2022).
Results: The CA results classified the sampling stations into two clusters: medium pollution (Stations 1 and 2) and high pollution (Stations 3, 4, and 5). PCA results confirmed the quality clustering of CA data. The findings of CA and FA methods facilitated the reduction of parameters identified in the first cluster including arsenic (As), lead (Pb), cadmium (Ca), chromium (Cr), nitrate (NO^{3-}), and phosphate (PO_4^{3-}). In contrast, the second cluster was characterized by total suspended solids, turbidity, hardness, ammonia, fecal coliform, electrical conductivity, biological oxygen demand, and chemical oxygen demand. PCA analysis revealed that the first principal component, accounting for 69.5% of the total variance, identified Pb, As, Cd, and Cr as the most important factors influencing changes in water quality. The second principal component, explaining 15.8% of the total variance, identified ammonia, nitrate, turbidity, and total suspended solids as the main parameters affecting the water quality of the Shafarood River.
Conclusion: The findings suggest that multivariate statistical techniques are valuable for interpreting large data sets, assessing water quality, and elucidating relationships between parameters and pollutant sources. These methodologies provide essential information regarding water quality and represent an effective approach to decision-making in the management of the Shafarood River′ s water quality.

1. Introduction

Surface water, as one of the most basic water sources, plays an essential role in supplying water needed for various agricultural, industrial, drinking, and health activities (Ebadati, 2017). Accordingly, being aware of the quality of surface water resources is considered one of the fundamental requirements (Jalili, 2020). Water quality is

How to cite: Gholami Deljomanesh, E., Hakimi Abed, M., Fataei, E., Shariati Feyzabadi, F., & Imani, A. A. (2025). Classification of Water Quality in the Shafarood River Based on Physicochemical Parameters and Heavy Metal Concentration Using Multivariate Analysis Methods. *Journal of Human Environment and Health Promotion*, *11*(2), 129-137.

affected by natural and anthropogenic pollution sources, including erosion, weathering, atmospheric chemistry, geology, organic matter dissolution, and activities associated with urbanization, industry, and agriculture (Mohammadi et al., 2023). Due to the surface water sources being in direct contact with these anthropogenic pollution sources, they are more exposed to pollution and vulnerability (Safizadeh et al., 2021). Therefore, the decrease in the quality of surface water resources is inevitable because of the population explosion, the increase in the diversity of pollutants, the destruction of forests and vegetation, and natural events such as droughts and floods, thus highlighting the necessity of monitoring and controlling the quality of these valuable resources (Sajjadi et al., 2019). The health of human communities is closely related to water quality, which is threatened by any pollutants that alter water quality (Akhoni pourhassani & Ghorbani, 2016). To address these challenges, numerous researchers have employed various statistical and data mining methods to classify the quality of water resources (Salarian et al., 2022) and have conducted many studies on surface water quality monitoring around the world (Mostafavi & Teimori, 2018). One of these statistical methods is multivariate analysis techniques. For instance, Ouyang (2005) applied PCA and FA to identify the most suitable parameters for assessing water quality at monitoring stations along the main stem of the lower St. Johns River in Florida, USA. Their results showed that nine parameters were critical in explaining the variations in water quality. Similarly, Alkarkhi et al. (2008) applied FA and DA methods to evaluate spatial changes in water quality data from two rivers, Juru and Jejawi in Malaysia, based on ten parameters measured at ten sampling stations. The FA method accounted for about 82% of the total variance in the water quality data through the first two factors. Nosrati et al. (2011) also applied multivariate analysis methods to assess the water quality of the Haraz-Ghara Soo Watershed and reported that two, three, and two parameters as the most significant explanatory factors at homogeneous stations 1, 2, and 3 respectively. In another research, Mirzaei et al. (2014) recruited the PCA and cluster CA methods to monitor the water quality of rivers in Mazandaran, analyzing monthly data from 16 parameters. The results indicated that total dissolved solids (TDS), pH, sulphate (SO_4^{2-}), NO^{3-}, and PO_4^{3-} were the most important parameters, with the first two principal components explaining 96.94% of the variances. Arain et al. (2014) evaluated the water quality in the Bannu watershed, Pakistan using multivariate analysis methods and showed that electrical conductivity (EC) and TDS were the most effective parameters in water quality. Kiyani et al. (2016) studied temporal and spatial variations of water quality in the Karoon basin (Khuzestan province, Iran) through multivariate analysis methods using 13 quality variables measured at 26 stations over 12 years. The CA method classified the stations into four pollution levels: low, medium, high, and very high. PCA and FA results highlighted Na%, SAR, Na^+, Cl^-, EC, and TDS as effective variables. Kazemzadeh and Malekian (2017) investigated temporal and spatial variations of water quality parameters in the Aji-Chai Watershed (Iran) using CA, DA, and PCA methods. The DA results identified three significant functions, explaining 73.50%, 20.30%, and 3.40% of the total variance, respectively. Zhao and Cui (2009) investigated the properties of surface water in the Luan River, northern China, by CA and FA methods, which demonstrated that the study months were divided into three quality groups based on seasonal characteristics. Rahnama and Sayari (2019) investigated the trends in chemical parameters of Tajan River water in Iran using PCA and Aqua Chem software, indicating that the first two components explained 77.57% and 81.54% of the variance during the first and second halves of the year, respectively, with the highest impacts related to pH, TDS, EC, SO_4^{2-} and flow rate. Soltani et al. (2019) analyzed the temporal and spatial variations of Aras Watershed water quality parameters at gauge stations of Khodaafarin, Khazangah, and Jolfa (Iran) using FA and PCA techniques. Their findings revealed that the first component accounted for the highest percentage of explained variance, primarily associated with Magnesium (mg^{2+}), CA^{2+}, HCO^{3-}, EC, and TDS at Khodaafarin Station, with EC (0.98) identified as the principal parameter of this component. At Khazangah Station, the first three components had the highest eigenvalues. These components explained 53.6%, 17.5%, and 12.9% of the variance, respectively. Collectively, these components explained a total of 84% of the variance. Babolhakami and Gholami Sefidkouhi (2019) analyzed the water quality of the Talar River (Mazandaran Province, Iran) using CA, PCA, and FA techniques to evaluate spatial variations based on 14 chemical parameters at six stations, identifying three clusters through the CA method. Their PCA and FA analyses showed that 80% of water quality variances were caused by the effective parameters in the first three components. These components explained 50.7%, 15.8%, 13.2%, and 5.8% of the variances, respectively. To reveal spatial and temporal variations in the Jajrood River (Tehran, Iran), Razmkhah et al. (2010) examined monthly samples of water quality parameters over three years at 18 stations via statistical analysis of PCA and CA. The CA classified similar water quality stations and identified Out-Meygoon as the most polluted site. The PCA facilitated the identification of a reduced number of mean five varifactors, pointing out 85% of both temporal and spatial changes. The CA and PCA led to similar results, though Out-Meygoon was identified as the most polluted station in both methods. Given that the Shafarood River, the second high-water river in western Gilan, Iran, discharges into the Caspian Sea and is exposed to various natural and human pollutants, it is of great importance to assess its water quality and determine the sources of pollutants. Accordingly, the present study aimed to classify the water quality of the Shafarood River and to determine the sources of pollution through multivariate analysis methods.

2. Materials and Methods

2.1 Study area

The Shafarood River is the second-largest river in terms of water flow in the western part of Gilan province in northern Iran, which flows into the Caspian Sea. The length of this river is 55 km, with a watershed of about 350 km^2. The width of the river is up to 25 m, with an average depth of 1.5 m, varying between 0.5 m and 3 m. The river with an area of 349.9 km^2, is located in the Talesh area, positioned between the cities of Rezvanshahr and Pareh Sar, at geographical coordinates of 49-06-30 East to 48-41 West latitude and 37-25 South to 37-34-30 North longitude. The maximum and minimum heights of this watershed are 2903 m and 60 m, respectively (Figure 1).

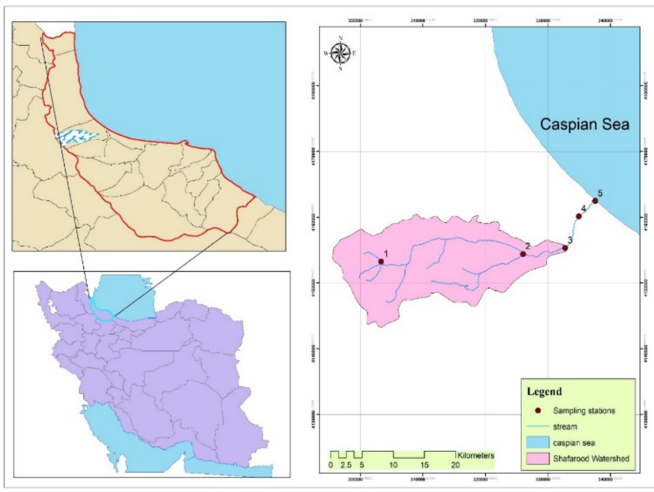

Figure 1. Location of Shafarood River Watershed (Gilan Province, northern Iran) and selected sampling stations from 2017 to 2022

2.2 Multivariate analysis techniques and data

This descriptive-analytical study analyzed the data from 16 water quality parameters of the Shafarood River, including total suspended solids (TSS), EC, pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), NO_3^-, PO_4^-, turbidity (Turb.), total hardness, ammonium (NH_4^+), fecal coliform (F. Coliform) and As, Ca, Cr, and Pb. These parameters were seasonally measured at five stations (Siyahrud or Station 1, Sarak or Station 2, Rezvanshahr or Station 3, Sandian or Station 4, and Shafarood Estuary or Station 5) over six years from 2017 to 2022. Data collection adhered to the Standard Methods for the Examination of Water and Wastewater (Rodger et al., 2017). The obtained data were used to cluster the studied stations and to identify the characteristics and similarities among these stations. Clustering was performed using CA, PCA, and FA statistical methods. DA was used to confirm the resulting clusters. The normality of the data was checked by the Kolmogorov-Smirnov (KS) test. Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were used to check data fit for PCA and FA. The KMO test was applied to determine the sampling adequacy, which represented the ratio of variance. Bartlett′s test was used to determine the homogeneity of variances. The data were analyzed by SPSS 24 and MINITAB 15 software (Fataei & Shiralipoor, 2011).

2.2.1 Principal component analysis (PCA)/factor analysis (FA)

The PCA is used to reduce the dimensions of data sets by describing the variability of many variables through linear combinations between source variables and principal factors (Bierman et al., 2011). Therefore, the principal component provides information about the most significant parameters that describe the whole data set obtained from the data reduction process with minimal loss of primary data (Equation 1) (Salarian et al., 2022):

Zij=ai1x1j+ai2x2j+ai3x3j+…+aimxmj Eq. (1)

Where *Z* is the component score, *a* is the component of loading, *x* is the measured value of the variable, *i* is the component number, *j* is the sample number, and *m* is the total number of variables. FA follows PCA, aiming to reduce the contribution of variables with a low significance level to simplify the data structure obtained from PCA as much as possible, which can be implemented with Varimax rotation through PCA and create new variables called variance factors. Therefore, the main component is a linear combination of the observable variables of water quality, while the varimax rotation can include latent, hypothetical, and hidden variables as well (Noori et al., 2012).

2.2.2 Cluster analysis (CA)

The CA method categorizes a set of sampling stations into two or more clusters based on their similarities, taking into account a set of characteristics resulting from the influence of the measured parameters (Fataei et al., 2012). Therefore, stations with similar water quality are placed in the same clusters (Hajigholizadeh & Melesse, 2017). A tree diagram provides a summary of the clustering process, a representation of the clusters and their adjacency, along with a significant reduction in the dimensions of the original data. The similarity between clusters and the separation of homogeneous clusters are determined based on the Euclidean distance. This research used hierarchical CA for normalized datasets through Ward′s minimum-variance method. Ward's method uses analysis of variance to evaluate the distances between clusters. In this research, the changes in water quality in the Shafarood River Watershed were evaluated using CA and linkage distance ($D_{link}/D_{max} \times 100$).

2.2.3 Discriminant analysis (DA)

This analytical statistical method extracts variables that can be discriminated between two or more groups that are naturally grouped (Koklu et al., 2010). Therefore, the discriminant function is used to classify the variables according to clustering values. If the DA is effective for the data set, the clustering table of accuracy and inaccuracy of the evaluations will give a high percentage of accuracy. Unlike CA, DA provides statistical clustering of samples that are implemented with prior knowledge of the membership of each parameter in a specific group or cluster. Moreover, it helps to group stations assuming common characteristics of parameters. Therefore, the DA technique consists of a discriminant function for each group, which operates on the raw data and creates a discriminant function for each group.

3. Results and Discussion

The mean and standard deviation of the measured data during the years 2019 to 2022 at the five study stations in the Shafa Rood River are presented in Table 1.

Table 1. Mean and Standard Deviation of Meseared Parameters in the study stations on the Shafarood River

Stations / Parameters	1		2		3		4		5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
BOD	1.04	1.49	1.33	1.20	1.19	2.11	2.29	1.78	3.33	2.79
COD	4.46	8.04	4.62	4.41	6.30	8.21	5.95	4.10	8.62	6.56
DO	9.00	1.53	9.07	1.50	8.73	2.31	8.76	1.95	7.65	2.70
EC	326.75	46.93	321.79	46.23	314.38	48.22	366.13	90.42	755.25	1206.98
F.Coli	209.43	327.69	545.76	987.25	489.49	957.70	326.69	389.21	828.91	969.72
NH4	0.15	0.11	0.20	0.23	0.25	0.48	0.41	0.81	0.33	0.55
NO3	1.70	0.44	1.40	0.32	1.46	0.64	1.29	0.78	1.22	0.63
PO4	5.27	20.11	0.91	3.09	1.52	4.88	0.66	2.07	1.19	4.35
Total Hardness	200.74	37.70	193.73	39.44	191.53	30.04	195.48	67.92	280.61	239.39
Turbidity	58.49	101.88	67.44	94.10	136.89	251.39	150.12	355.63	161.11	418.61
pH	7.95	0.35	8.06	0.31	805	0.29	7.92	0.42	7.92	0.33
TSS	44.79	89.29	50.59	98.21	47.90	132.67	57.33	155.58	116.02	128.66
As	7.82	0.41	5.76	0.67	2.56	0.18	1.70	0.00	1.50	0.00
Cd	5.50	0.69	3.11	0.45	0.98	0.12	0.40	0.00	0.10	0.00
Cr	0.10	0.00	0.10	0.00	0.06	0.04	0.00	0.00	0.00	0.00
Pb	0.17	0.15	0.22	0.10	0.10	0.00	0.10	0.00	0.10	0.00

SD = Standard Deviation

The results of the multivariate statistical analysis methods used based on the data obtained from the stations for the 16 measured parameters are presented as follows.

3.1 The CA results

Figure 2 shows the dendrogram of the results of the quality classification of the studied stations based on 16 parameters in the Shafarood River Watershed through the CA method. The results of the CA method put the stations studied in two homogeneous clusters. The similarity between clusters and the separation of homogeneous clusters were determined based on Euclidean distance. The quality distance between each sampling station in the clusters obtained was the result of correlation and correlation itself between the quality parameters measured in the river.

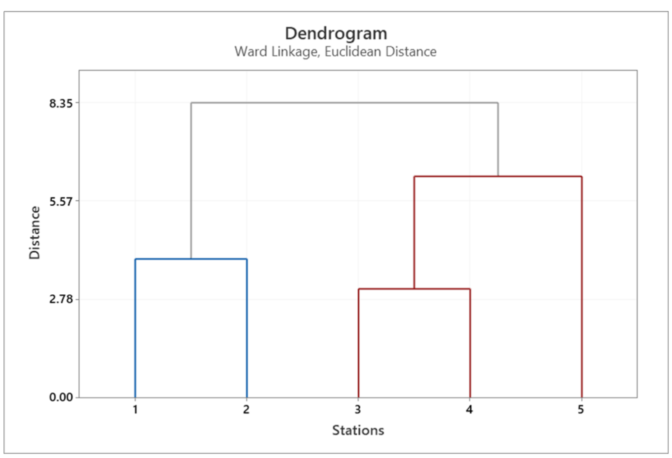

Figure 2. Dendrogram obtained from cluster analysis of sampling stations based on parameters measured in Shafarood River Watershed (Gilan Province, northern Iran) from 2017 to 2022

Based on the clustering obtained from CA (Figure 2), Stations 1 and 2 in the upstream area of the river with medium pollution level were placed in the first cluster, and Stations 3, 4, and 5 in the downstream area of the river with high pollution level were placed in the second cluster. In the evaluation of Dez River water quality (Khuzestan province, Iran) using multivariate analysis methods, the CA results for the classification of sampling stations demonstrated that the two stations in the downstream area of the river were placed in the same category in terms of similarity of water quality variables and that one station in the upstream area of the river had no similarity to other stations in terms of water quality variables; therefore, the urban area and the anthropogenic pollution sources played a significant role in the quality grouping of the studied stations (Neissi & Tishehzan, 2019). However, the results of water quality parameters analysis in Aji-Chai Watershed (East Azerbaijan, Iran) classified the studied stations into three clusters since there was a large number of sampling stations, as well as there was a great difference in quality changes between the stations under study (Kazemzadeh & Malekian, 2017). Based on the results of the mean value of parameters measured in each cluster (Table 1), the concentrations of Pb, Cr, Cd, NO_3^-, and PO_4^{3-} were higher than the total mean value. The reason for the high concentration of heavy metals in the first cluster (Figure 2 and Table 1), given that there was no industrial or mineral activity in the upstream area of the stations of this cluster, could be attributed to the structure of the river bed and dissolution of soluble minerals and their penetration into river water, which might be due to the presence of coal veins in the geological structure and river bed in the first and second upstream stations (it is necessary to explain that there are small quantities of different metals such as nickel, copper, aluminum, lead, arsenic and mercury in coal) or

JHEHP. 2025; 11(2): 129-137

could be caused by the sewage from the use of fertilizers and pesticides in the gardens, as well as the rural sewage to the river water in these stations. At the stations of the second cluster, the mean values of TSS, Turb., TH NH_4^+, F. Coliform, EC, BOD, and COD were higher than the total mean value, and the DO level was less than the total mean value, which could be attributed to various reasons, including forest clearcutting within the limits of these stations and as a result of introducing eroded sediments into the river, taking sand from the riverbed, the presence of villa gardens and tourist restaurants and scattered residential areas in the upstream area of second cluster stations, and the discharge of sewage into the river. Accordingly, the stations of the first cluster were considered as sources of natural and anthropogenic pollution, while the stations of the second cluster were introduced as sources of anthropogenic pollution. A research team investigated the surface water quality of Shallow Valley Lake, Kashmir, India, using PCA and CA methods based on 13 parameters and reported that the main cause of changes in lake water quality could be due to discharge of domestic sewage and agricultural runoff (Ishtiyaq et al., 2017). Bhardwaj et al. (2010) assessed the water quality of the Chhoti Gandak River, Ganga Plain, India, using PCA, and found that the main factors of water quality changes in the river were reasons such as poor drainage and the introduction of pollution from domestic sewage. In the classification of the surface water quality parameters in the Aji-Chai Watershed, Kazemzadeh and Malekian (2017) announced the reason for the change in the quality of the stations as different residential, agricultural, and industrial uses. However, in the classification of Water Quality Parameters in the Tadjan River using CA, Faryadi et al. (2013) stated that the quality changes were caused by the composition of geological formations in the basin of this river. Fan et al. (2010) used the cluster statistical analysis method to examine the quality of the stations in the north, east, and west of the Pearl River in China and grouped the rivers into several clusters based on the severity of pollution. Bu et al. (2010) also used the multivariate statistical technique to examine the spatial and temporal changes in the water quality of the Jinshui River in China and to identify the main pollution factors and their sources and stated that the pollution of the river in the study was caused by the inflow of domestic wastewater and agricultural runoff.

Table 2. Mean value of water quality parameters tested in two clusters resulting from the cluster analysis method in Shafarood River Watershed (Gilan Province, northern Iran)

Clustering	BOD	COD	DO	EC	F. Coliform	NH_4^+	NO_3^-	PO_4^{3-}	TH	Turb.	pH	TSS	As	Cd	Cr	Pb
1.00	1.19	4.55	9.04	324.27	377.60	0.17	1.55	3.09	197.24	62.96	8.00	47.69	6.79	4.31	0.10	0.19
2.00	2.51	6.96	8.38	478.58	548.36	0.33	1.32	1.12	222.54	149.37	7.96	73.75	1.92	0.49	0.02	0.10
Total	1.98	5.99	8.64	416.86	480.06	0.27	1.41	1.91	212.42	114.81	7.98	63.33	3.87	2.01	0.05	0.14

3.2 The DA results

In this research, the DA method was used to further investigate the impact of the measured water parameters of the Shafarood River on changes in the quality of the sampling stations categorized through the CA method in two clusters (Table 3).

Table 3. Results of discriminant analysis for water quality parameters tested in Shafarood River Watershed (Gilan Province, northern Iran)

A comparison of equality of means		
Parameters	Wilks' Lambada	*p-value*
BOD	0.348	0.098
COD	0.377	0.112
DO	0.607	0.258
EC	0.803	0.453
F. Coliform	0.843	0.0509
NH_4^+	0.341	0.095
NO_3^-	0.542	0.209
PO_4^{3-}	0.680	0.321
TH	0.869	0.549
Turb.	0.036	0.003**
pH	0.900	0.604
TSS	0.771	0.415
As	0.088	0.011**
Cd	0.158	0.028
Cr	0.247	0.057
Pb	0.110	0.016**

** = indicating a significance level of 1%

Table 2 provides the results of the DA method related to the Equity of Means Test for water quality parameters in studied stations. In the DA method, the Wilks' Lambada index was used to determine the best function and assess the Equity of Means test for parameter values at different stations. The Wilks' Lambda index ranges between zero and one, with lower values indicating a more appropriate discriminant function (Alili & Krstev, 2019). The results showed that the lowest value of the Wilks' Lambada index was related to Turb. (0.36), As (0.088), and Pb (0.11). F-statistics also confirmed that the mean values of these parameters were statistically significant among the studied stations ($P < 0.01$). Consequently, it was found that these three parameters played the most role in discriminating and grouping the quality of the stations, which also confirmed the CA findings. In evaluating the temporal and spatial variations of water quality parameters in the Zohreh River, Iran, the DA results demonstrated that the flow rate, EC, temperature, HCO_3^-, CL^-, Na%, and TH were responsible for quality changes among the four studied stations and confirmed the clustering of the CA method (Ravanbakhsh et al., 2019). Similarly, Koklu et al. (2010) used 22 parameters along with PCA, DA, and multivariate regression analysis methods to monitor the water quality of the Melen River in Turkey, reporting that the use of multivariate statistical methods yielded satisfactory results in evaluating the water quality for this river.

3.3 The PCA/FA results

The FA technique is one of the multivariate analysis methods to determine the most important influencing parameters in river water quality. To this end, the current research employed the PCA method based on 16 water quality parameters in the Shafarood River Watershed. As previously explained, the data must have the required conditions for factor analysis, which was determined through the results of the KMO and Bartlett's tests. The KMO and Bartlett's index value should be more than 0.5 (Fataei & Shiralipoor, 2011). Based on the results of this test (Table 4), the calculated index value was 0.61, indicating that the data were suitable for application in the PCA test.

Table 4. Results of Kaiser-Meyer-Olkin (KMO) and Bartlett's tests for data measured in the Shafarood River Watershed (Gilan Province, northern Iran)

sampling adequacy		
Kaiser-Meyer-Olkin (KMO) test		0.611
Bartlett's test	χ^2	1314.74
	Degrees of freedom	120
	Significance level	0.000

In the PCA method, the criterion for factor selection was based on eigenvalues greater than 1, which helped identify the most effective parameters influencing river water quality (Kazemzadeh & Malekian, 2017). Therefore, the Scree Plot (Figure 3) indicated that out of the 16 components, each having eigenvalues above 1 and explained a total of 69.14% of the variance, were selected for further analysis.

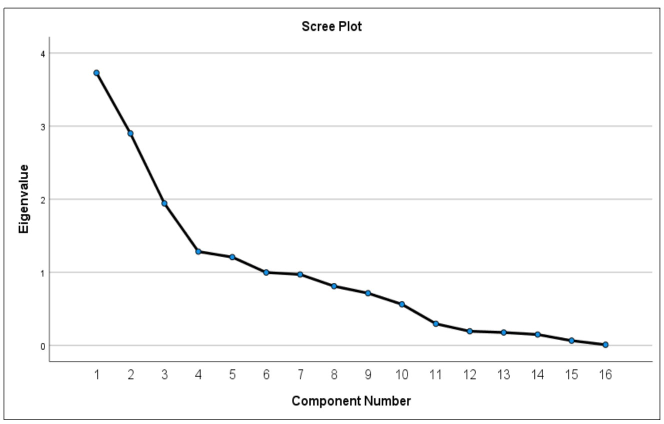

Figure 3. Scree plot for principal component analysis of water quality parameters in Shafarood River Watershed (Gilan Province, northern Iran)

According to the obtained eigenvalues (Figure 3), the first two components accounted for the largest amount, with eigenvalues of 3.73 and 2.90, respectively, explaining 69.5% and 15.8% of the total variance. Together, these components accounted for 85.3% of the variance for the measured parameters (Table 4), thereby serving as the best components in explaining the variations in the water quality of the studied stations. As previously explained, factor rotation was used through the Varimax rotation method to extract the main water quality parameters from the principal components. This method identifies parameters with the highest factor loading (positive or negative) in each component as the most representative variables for that component (Soltani et al., 2019). Table 5 shows the factor loadings of each measured parameter with respect to their association with the two principal components.

Table 5. Results of principal component analysis of water quality parameters in Shafarood River Watershed (Gilan Province, northern Iran)

Parameters	Cluster 1	Cluster 2
BOD	-0.309	0.345
COD	-0.134	0.285
DO	0.099	-0.025
EC	-0.073	-0.030
F. Coliform	-0.011	-0.033
NH_4^+	-0.133	0.903
NO_3^-	0.347	0.756
PO_4^{3-}	0.152	-0.117
TH	-0.068	-0.101
Turb.	-0.120	0.568
pH	0.159	0.126
TSS	-0.088	0.893
As	0.953	-0.063
Cd	0.933	-0.054
Cr	0.900	-0.014
Pb	0.542	0.020
Initial Eigenvalues	11.12	2.53
Percent variance	69.5	15.8
Cumulative percent variance	69.5	85.3

Based on the dendrogram obtained from PCA (Figure 3 and Table 4), Stations 1 and 2 were grouped, with their water quality influenced by natural and anthropogenic sources of pollution. These influences include the geological structure and type of the riverbed in the upstream area, the discharge of sewage from garden lands containing residues of used fertilizers and pesticides, as well as the discharge of sewage from regional rural areas into the river. The parameters NH_4^+, NO_3^-, Turb., and TSS were identified as the main factors affecting water quality in the second component, which accounted for 15.8% of the total variance. Stations 3, 4, and 5 were included in this group (Figure 3), where quality changes have been affected by anthropogenic activities, including illegal logging, unauthorized sand extraction from the riverbed, and sewage discharge from rural areas and tourist areas in the upstream areas of these stations. Soltani et al. (2019) evaluated the temporal and spatial variations of Aras Watershed water quality parameters in the period 1999-2011 at the gauge stations of Khodaafarin, Khazangah, and Jolfa gauge stations in Iran using FA and PCA techniques. They found that the highest percentage of explained variance in the first component was related to mg^{2+}, CA^{2+}, HCO_3^-, EC, and TDS at Khodaafarin Station. Among the parameters in the first component, EC had the highest factor loading (0.98) and was recognized as the main parameter of this component. At Khazangah Station, the first three components had the highest eigenvalues. These components explained 53.6%, 17.5%, and 12.9% of the variance, respectively, and collectively accounting for 84% of the total variance. In the first component, the parameters Mg^{2+}, Ca^{2+}, SO_4^{2-}, Cl^-, HCO_3^-, EC, and TDS had the highest correlations. At Jolfa Station, the first four components had the highest eigenvalues. These components explained 50.7%, 15.8%, 13.2 and 5.8% of the variances, respectively. Yidana et al. (2008) used PCA to extract the main factors influencing hydrochemical changes of surface water in the Ankwaso, Dominase, and Prestea regions of Ghana, identifying four principal components that explained 86% of the total variance. The parameters of EC, pH, TDS, HCO_3^-, Na, SiO_2^-, PO_4^{3-}, K^+, and total alkalinity (TA) were introduced as the most important effective parameters in the hydrochemical changes of surface water in these areas, which were caused by agricultural activities, domestic sewage discharge and weathering. Nosrati et al. (2011) used FA/PCA to evaluate the water quality of the Haraz-Ghara Soo Watershed, revealing that pollution from agricultural and garden activities, domestic sewage, and weathering were the key influencing factors based on the identified important parameters. Kazemzadeh and Malekian (2017) determined the effective parameters in surface water quality in the Aji-Chai Watershed (Iran) and determined that the first two components explained 78.75% and 14.71% of the community variance, respectively, with sodium adsorption ratio (SAR) and pH having the highest factor loadings were identified as the main parameters affecting the river quality. To identify the principal factors influencing hydrochemical changes in surface water in Mazandaran Province, Mirzaei et al. (2014) recognized five key components describing the region's water quality, including TDS, pH, SO_4^{2-}, NO_3^-, and PO_4^{3-}. Mohammadi Ghaleni and Kardan Moghaddam (2022) determined the parameters affecting the water quality of the Sefidroud River (Iran) using PCA, concluding that TDS and K^+ exhibited the highest and lowest weights, respectively.

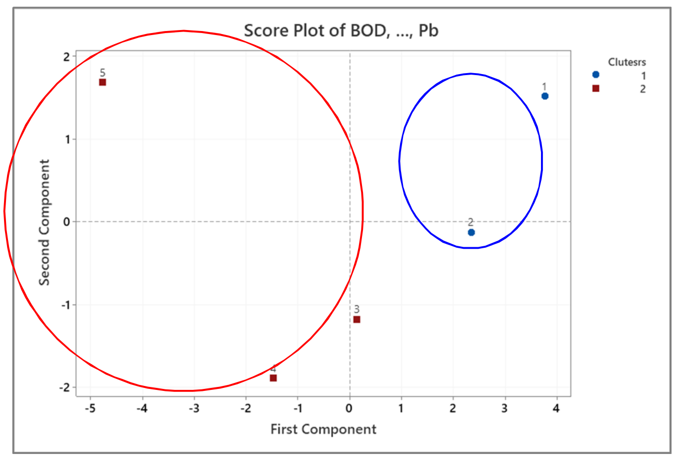

Figure 4. Score plot of sampling stations according to the first and second components of principal component analysis based on water quality parameters measured in Shafarood River Watershed (Gilan Province, northern Iran)

As seen in Figure 4, the classification of PCA confirmed the results of CA clustering. According to the score plot of PCA (Figure 4), the studied stations in the Shafarood River Watershed were classified into two quality groups based on the effect of the measured parameters on the quality of the river water, similar to CA (Figure 2), so that both methods of statistical analysis put Stations 1 and 2 in one group and Stations 3, 4 and 5 in the next group. In the classification of groundwater quality using multivariate analysis methods (a case study of aqueducts in the east of Tehran, Iran), the CA and PCA results classified the studied stations into three groups with high pollution (HP), medium pollution (MP) and low pollution (LP) levels (Salarian et al., 2022). The results showed that the CA method provided better quality classification needs of the stations for PCA. Therefore, these reliable analytical techniques can be suitable for monitoring the quality of river water for quality classification and can be used to detect the parameters influencing the quality classification of sampling stations.

4. Conclusion

Rivers are considered to be the main sources of water supply for various domestic, industrial, and agricultural sectors, fundamentally supporting sustainable development. Therefore, it is vital to protect the quality of these valuable resources. The present research was conducted to evaluate the water quality of the Shafarood River Watershed in Gilan Province, northern Iran, using multivariate analysis methods based on 16 water quality parameters during a six-year period with seasonal sampling. The results of CA, DA, and FA indicated that the most effective parameters in the quality of the five studied stations could be classified into two quality groups with medium and high pollution levels. Both anthropogenic and natural pollution factors were involved in the first cluster, while the main source of water quality pollution in the stations of the second cluster was caused by anthropogenic activity. To conclude, our results indicated that CA and PCA statistical methods can properly determine the effective parameters influencing spatial variations in the water quality of surface water sampling stations. The two-dimensional display of the stations using PCA confirmed the CA clustering and was able to discriminate the investigated stations from one another.

Authors' Contributions

Elham Gholami Deljomanesh: Conceptualization; Funding acquisition; Investigation; Visualization; Writing. **Mahsa Hakimi Abed:** Supervision. **Ebrahim Fataei:** Methodology; Project administration; Supervision; Validation. **Fatemeh Shariati Feyzabadi:** Resources. **Ali Akbar Imani:** Formal analysis; Software.

Funding

No funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

The authors would like to express their gratitude for the support provided by the university in implementing the current project.

Ethical considerations

This article has been adapted from the Ph.D. dissertation (No. 162821187) in Environmental Science and Engineering by Ms. Elham Gholami Deljomanesh at Islamic Azad University, Lahijan Branch, Lahijan, Iran.

Using artificial intelligence

Artificial intelligence was not used in this research.

References

1. Akhoni Pourhosseini, F., & Ghorbani, M. A. (2016). Application of Shannon entropy in determining the most effective chemical parameter in surface water quality (case study: Sofi Chay watershed). *Environment and Water Engineering, 2*(4), 322-332.

2. Alili, A., & Krstev, D. (2019). Using spss for research and data analysis. *Knowledge-International Journal, 32*(3), 363-368.

3. Alkarkhi, A. F., Ahmad, A., Ismail, N., Easa, A., & Omar, K. (2008). Assessment of surface water through multivariate analysis. *Journal of Sustainable Development, 1*(3), 27-33.

4. Arain, M. B., Ullah, I., Niaz, A., Shah, N., Shah, A., Hussain, Z., . . . & Kazi, T. G. (2014). Evaluation of water quality parameters in drinking water of district Bannu, Pakistan: multivariate study. *Sustainability of Water Quality and Ecology, 3*, 114-123.

5. Babolhakami, A., & Gholami Sefidkouhi, M. A. (2019). Analyze of Talar River water quality using multivariate techniques. *Journal of Watershed Management Research, 9*(18), 250-259.

6. Bhardwaj, V., Singh, D. S., & Singh, A. K. (2010). Water quality of the Chhoti Gandak River using principal component analysis, Ganga Plain, India. *Journal of Earth System Science, 119*(1), 117-127.

7. Bierman, P., Lewis, M., Ostendorf, B., & Tanner, J. (2011). A review of methods for analysing spatial and temporal patterns in coastal water quality. *Ecological Indicators, 11*(1), 103-114.

8. Bu, H., Tan, X., Li, S., & Zhang, Q. (2010). Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. *Ecotoxicology and Environmental Safety, 73*(5), 907-913.

9. Ebadati, N. (2017). Statistical analysis of Dez River water quality, southwest of Iran. *Anthropogenic Pollution, 1*(1), 46-60.

10. Fan, X., Cui, B., Zhao, H., Zhang, Z., & Zhang, H. (2010). Assessment of River water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Sciences, 2*, 1220-1234.

11. Faryadi, S., Shahedi, K., & Nabatpoor, M. (2013). Investigation of water quality parameters in Tadjan River using multivariate statistical techniques. *Journal of Watershed Management Research, 3*(6), 75-92.

12. Fataei, E., & Shiralipoor, S. (2011). Evaluation of surface water quality using cluster analysis: a case study. *World Journal of Fish and Marine Sciences, 3*(5), 366-370.

13. Fataei, E., Mosavi, S., & Imani, A. A. (2012). Identification of anthropogenic influences on water quality of Aras River by multivariate statistical techniques. *2nd International Conference on Biotechnology and Environment Management IPCBEE, 42*(1), 1-8.

14. Hajigholizadeh, M., & Melesse, A. M. (2017). Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses. *Catena, 151*, 247-258.

15. Ishtiyaq, A. N., Anisa, B. K., & Abdul, H. (2017). Evaluation of seasonal variability in surface water quality of Shallow Valley Lake, Kashmir, India, using multivariate statistical techniques. *Pollution, 3*(3), 349-362.

16. Jalili, S. (2020). Water quality assessment based on HFB I& BMWP index in Karoon River, Khouzestan Provience, (Northwest of Persian Gulf). *Anthropogenic Pollution Journal, 4*(1), 36-49.

17. Kazemzadeh, M., & Malekian A. (2017). Analysis of the surface water quality parameters in the Aji-Chai watershed based on the multivariate statistical techniques. *Journal of Range and Watershed Management, 70*(2), 465-478.

18. Kiyani, M., Kiyani, S., & Sadeghi, H. (2016). Analysis of spatial and temporal variation in water quality in Karoon basin using multivariate statistical techniques. *Journal of Irrigation and Water Engineering, 6*(3), 34-48.

19. Koklu, R., Sengorur, B., & Topal, B. (2010). Water quality assessment using multivariate statistical methods-a case study: Melen River system (Turkey). *Water Resources Management, 24*, 959-978.

20. Mirzaei, M., Riahi-Bakhtiari, A., Salman-Mahini, A., & Gholamalifard, M. (2014). Analysis of the physical and chemical quality of Mazandaran province (Iran) rivers using multivariate statistical methods. *Journal of Mazandaran University of Medical Sciences, 23*(108), 41-52.

21. Mohammadi, J., Fataei, E., Ojaghi Aghchekandi, A., & Taghavi, L. (2023). Investigation and determination of land use effects on surface water quality in semi-arid areas: case study on Qarasu River in Iran. *Anthropogenic Pollution, 7*(2), 1-7.

22. Mohammadi Ghaleni, M., & Kardan Moghaddam, H. (2022). Introducing a new drinking water quality index for surface water resources using multivariate analysis (case study: Sefidroud River). *Journal of Water and Soil, 36*(4), 439-458.

23. Mostafavi, H., & Teimori, A. (2018). Investigating multiple human pressure types in the southern Caspian Sea Basin Rivers at different spatial scales toward integrating water resource management (IWRM) in Iran. *Anthropogenic Pollution Journal, 2*(1), 38-47.

24. Neissi, L., & Tishehzan, P. (2019). Dez River water quality assessment by using multivariate statistical methods. *Journal of Irrigation and Water Engineering, 9*(1), 139-150.

25. Noori, R., Karbassi, A., Khakpour, A., Shahbazbegian, M., Mohammadi Khalf Badam, H., & Vesali-Naseh, M. (2012). Chemometric analysis of surface water quality data: case study of the Gorganrud River basin, Iran. *Environmental Modeling & Assessment, 17*, 411-420.

26. Nosrati, K., Derafshi, Kh., Ghrachahi, S., & Ghrachahi, Kh. (2011). Assessment of surface water quality in Haraz-Ghara Soo watershed using multivariate statistical techniques. *Researches in Earth Sciences, 2*(1), 41-55.

27. Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research, 39*(12), 2621-2635.

28. Rahnama, S., & Sayari, N. (2019). Survey and trends of chemical water quality parameters of Tajan River water quality using principal component analysis and Aqua Chem software. *Human & Environment, 17*(1), 13-25.

29. Ravanbakhsh, M., Tahmasebi Birgani, Y., Dastoorpoor, M., & Ahmadi Angali, K. (2019). Evaluation of temporal and spatial variations of water quality parameters in Zohreh River, Iran. *Avicenna Journal of Environmental Health Engineering, 6*(2), 75-82.

30. Razmkhah, H., Abrishamchi, A., & Torkian, A. (2010). Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management, 91*(4), 852-860.

31. Rodger, B. B., Andrew, D. E., & Eugene, W. R. (2017). *Standard methods for the examination of water and wastewater, 23rd edition.* American Public Health Association, American Water Works Association, Water Environment Federation, 1545.

32. Safizadeh, E., Karimi, D., Gahfarzadeh, H. R., & Pourhashemi, S. A. (2021). Investigation of physicochemical properties of water in downstream areas of selected dams in Aras catchment and water quality assessment. *Anthropogenic Pollution Journal, 5*(1), 41-48.

33. Sajjadi, N., Davoodi, M., & Jozi, S. A. (2019). The quality assessment of Kan River's resources in terms of agricultural and drinking purposes. *Anthropogenic Pollution Journal, 3*(1), 46-53.

34. Salarian, M. B., Mackialeagha, M., & Behbahaninia, A. (2022). The use of multivariate statistical methods for the classification of groundwater quality: a case study of aqueducts in the east of Tehran, Iran. *Anthropogenic Pollution Journal, 6*(2), 1-9.

35. Soltani, S. H., Ghohroudi, M., & Sadoogh, S. H. (2019). Evaluation of the surface water quality using statistical multi-variate techniques, case study: Aras watershed. *Iran-Water Resources Research, 15*(2), 319-328.

36. Yidana, S. M., Ophori, D., & Banoeng-Yakubo, B. (2008). A multivariate statistical analysis of surface water chemistry data-the Ankobra Basin, Ghana. *Journal of Environmental Management, 86*(1), 80-87.

37. Zhao, Z. W., & Cui, F. Y. (2009). Multivariate statistical analysis for the surface water quality of the Luan River, China. *Journal of Zhejiang University-Science A, 10*(1), 142-148.