Agriculture Association of Textile Chemical and Critical Reviews Journal (2024) 405-412

AATCC Review

27 August 2024: Received
29 September 2024: Revised
06 November 2024: Accepted
27 November 2024: Available Online

https://aatcc.peerjournals.net/

Original Research Article

Open Access

Surface Water Quality Management Using Multivariate Analysis Methods (A Case Study: Shafarood River, Gilan Province, Iran)

Elham Gholami Deljomanesh[1], Mahsa Hakimi Abed*[1], Ebrahim Fataei*[2], Fatemeh Shariati Feyzabadi[1], Ali Akbar Imani[3]

[1]Department of Environment, Lahijan Branch, Islamic Azad University, Lahijan, Iran
[2]Department of Environment, Ardabil Branch, Islamic Azad University, Ardabil, Iran
[3]Department of Agriculture, Ardabil Branch, Islamic Azad University, Ardabil, Iran

ABSTRACT

Surface water management is directly related to determining the correlation between physical, chemical and biological variables as well as identifying their natural and anthropogenic origin. The Shafarood River (Gilan Province, Northern Iran) is subject to the discharge of residential and agricultural pollutants, so it is very important to control its quality and determine the sources of pollutants and its quality status. The present study was conducted to scrutinize the correlation between the major parameters affecting the water quality of the Shafarood River and to monitor the water quality in different areas of the river using canonical correlation analysis and cluster analysis models, respectively. Five physical and four chemical parameters were measured at five stations based on Standard Methods for the Examination of Water and Wastewater 2015 over six years. The results revealed a significant correlation between two categories of response (physical parameters) and predictor (chemical parameters) variables, which mainly originated from anthropogenic pollution (effluents from residential and garden areas). According to the results of cluster analysis, the stations were grouped into two clusters based on the degree of pollution, and the cluster grouping confirmed the canonical correlation matrix. The research findings revealed the effectiveness of the obtained linear combinations for the physical parameters, including total suspended solids and turbidity, as well as the chemical parameters, including biochemical oxygen demand and nitrate. To conclude, the efficiency of canonical correlation analysis and cluster analysis methods was confirmed in identifying the determinant variables of water quality and in classifying the water quality monitoring stations in the optimal management of rivers. Considering the importance of different parameters in changing the water quality of the Shafarood River, the multivariate statistical methods have been used in a proper way in classifying the quality and determining the pollution sources.

Keywords: *Surface water management, Quality control, Hydrology and water resource, Statistical Analysis, Multivariate analysis methods, Shafarood river*

1. Introduction

One of the fundamental issues in the water quality management is to determine the connection between different physical, chemical, and biological parameters effectively influencing the water bodies [20]. The physical pollutants are generally caused by the natural factors and climate conditions of the watershed, while the chemical pollutants are often generated by man-made activities [17]. Hence, a high correlation between these two categories of parameters can naturally indicate the same origin of the pollution source. Finding the relationship between pollution sources using laboratory experiments requires huge costs. To this end, multivariate analysis methods have been widely reported to be useful for finding such correlations [15]. For example, Canonical correlation analysis (CCA) and Cluster Analysis (CA) methods have been exploited to interpret complex and large-volume data and to better understand the water quality and environmental status of the studied system. They have been applied as a valuable tool for the reliable management of water resources by identifying possible temporal and spatial factors caused by natural and anthropogenic pollution sources [7], because these techniques help better interpret water quality data [12].

Khoda Bakhsh et al. [11] identified the origin of a dissolved load of Khar-River Water (Qazvin Province, Iran) by statistical methods (correlation coefficient, cluster analysis and box plot) using 35 samples in two high-water and low-water seasons. Their results showed that the parameters of calcium sulfate and sodium chloride originated mainly from Miocene evaporates and played the most important role in increasing the salt content of the river water. Abdolabadi et al. [1] studied the correlation between physical and chemical parameters, as well as the origin of their transfer in the water of the Atrak River (Iran) watershed using CCA method. Their findings revealed a strong correlation between the two sets of physical and chemical parameters, which were also confirmed by the CA data. Of the five canonical categories, only the first two categories had statistical validity based on P-value statistic.

Khanduzi et al. [10] used the CCA method to find a correlation between quality parameters and heavy metal concentrations in samples collected from the Ramian area (Golestan province, northern Iran) and found that three canonical categories with correlation coefficients of 0.973, 0.795 and 0.624, respectively, showed high dependence between the variables under investigation. In addition, the results of the principle component analysis (PCA) revealed that 79.67 % of the overall variance was explained by the first two components, reflecting the effective

*Corresponding Author: **Mahsa Hakimi Abed***
Ebrahim Fataei

DOI: https://doi.org/10.21276/AATCCReview.2024.12.04.405

water quality parameters. According to the CCA findings, the parameters of electrical conductivity (EC), bicarbonate, nitrite, pH, nitrate, and phosphate belonging to the "predictor" variables, as well as the concentration of heavy metals including zinc, cobalt, cadmium, and nickel belonging to the "response" variables were the most important factors among other studied variables to evaluate the drinking water quality in the study area.

Eskandare et al. [5] examined the CCA method to determine the correlation between the parameters of physical and chemical pollution in two periods of High-Water and Low-Water in the Sefidrood River watershed in Iran. Their results suggested a strong correlation between physical and chemical parameters in both periods. Therefore, they deduced that the origin of both groups of physical and chemical parameters in both periods could be the same. Boyacıoğlu et al. [3] investigated 22 surface water quality parameters using multivariate analysis methods (including CA, CCA, Water Quality Index, and Sensitivity Analysis) in Gediz River Basin, Turkey, and reported that the organic and physical-inorganic chemical parameters were affected by changes in physical conditions (discharge and temperature), unlike inorganic variables. Their findings clarified that information generated based on changes and correlations in water quality datasets could be useful for prioritizing water management measures. Nessi and Tishehzan (2018) evaluated the water quality of Dez River (Khuzestan province, Iran) using multivariate analysis methods and realized that there was a significant correlation between two categories of physical and chemical variables based on the canonical correlation matrix, which were mainly caused by the same anthropogenic pollution sources. Moreover, the CA results for the classification of sampling stations demonstrated that two stations in the downstream area of the river were placed in the same category in terms of the similarity of water quality variables, and one station in the upstream area of the river was not similar 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.

Rasuli et al. (2011) used the CCA method to model water quality management. To this end, the Sefidrood River was chosen as the study area, and the results indicated the anthropogenic origin of the physical and chemical parameters of the pollution. In a similar study, Chan et al. [4] analyzed measurements taken from a Storage Reservoir over ten years and concluded that EC (as a physical parameter) and chloride concentration (as a chemical parameter) had the highest canonical weight in the category of physical and chemical variables. In determining the correlation between the parameters affecting the water quality of the Bursa River in Turkey, Gumus [8] achieved successful results in the application of the CCA method to find the correlation between physical and chemical variables and to understand the difference in the measurement periods concerning these groups.

Regarding the construction of Shafarood Dam on the Shafarood River (Gilan Province, North of Iran), since determining the correlation between water quality variables is one of the basic pillars of water quality management [18], we used the CCA method to find the correlation between the effective variables and the CA method to identify the main effective variables in the water quality of this river and also to categorize the studied water quality monitoring stations.

2. Methods

2.1 Study area and site selection

Shafarood River is the second most watered river in the western part of Gilan province in northern Iran, which flows into the Caspian Sea. Shafarood Dam was built on this river with the aim of storing, regulating, and distributing the river water to meet the water needs of the area under cultivation of the region. The early branches of this river originate from the heights of Shirakli Dagh, located 17 km southeast of Khalkhal city (Ardebil province, Iran), which are connected in the village of Sang Deh and forms this river. Then, this river flows in the east direction and irrigates many villages on its way and also receives a branch that flows from the south of the main branch. Along this route, other big and small rivers join each other, thereby increasing the water level of this river. This river is considered the habitat of native freshwater fish, and for this reason, many ponds have been created around the river for fish breeding.

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 and its average depth is 1.5 m (with a range between 0.5 m and 3 m). With an area of 349.9 km^2, this river is located in Talesh region between the cities of Rezvanshahr and Pareh Sar, at the 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 height of this watershed is 2903 m and 60 m, respectively. A large part of this watershed is mountainous and about 80% of its surface is covered by forest trees. The geological structure of the second era and the expansion of layers exposed to erosion can be seen in most places. The limestone formations in the watershed have caused the emergence of many springs. Deep valleys and steep slopes are the two main reasons for the high speed of water flow in the region. This watershed has a humid climate with an annual rainfall of 1431.76 mm and a watery regime in autumn. Shafarood River has two high-water periods every year, one in spring and the other in autumn. The regime of this river is rainy and snowy. The maximum flow of the river is in April (9.09 m^3s^{-1}) and the minimum is in August (2.85 m^3s^{-1}). In addition, the average annual flow of this river is 5.56 m^3s^{-1}. The perimeter of the watershed is 100.7 km, with a slope of 6.06%. Figure 1 depicts the location map of the studied stations on the Shafarood River.

Considering that the Shafarood River is subject to the discharge of residential and agricultural pollutants, it is of vital importance to check the water quality of this river and determine the sources of its pollution.

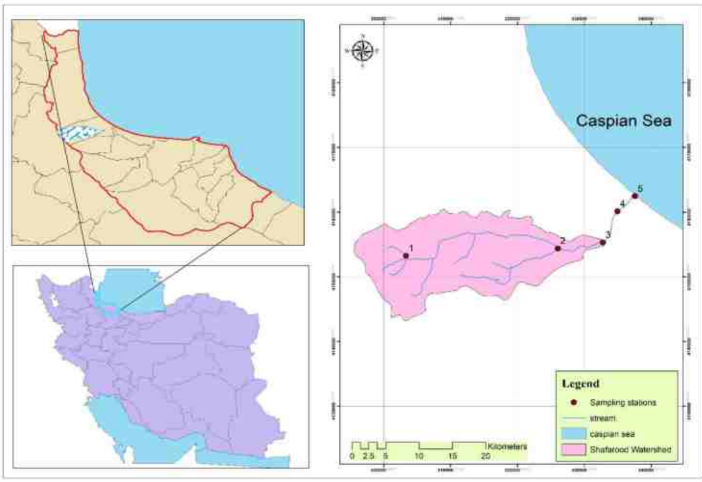

Figure 1. Location of Shafarood River Watershed (Gilan Province, northern Iran) and selected sampling stations during 2017 to 2022

2.2 Sampling and Measurements

Sampling stations were selected based on the natural conditions of the region and the possibility of access to the river and taking into account the natural and anthropogenic effects, including river sub-branches and pollution sources, including agricultural lands or residential and tourism centers. Finally, five sampling stations were selected, whose coordinates were determined on the map by the Global Positioning System (GPS) (Table 1).

Table 1. Coordinates and locations of water sampling stations in Shafarood River (Gilan Province, northern Iran) during 2017 to 2022

Stations	Locations	Height (m)	Coordinates	
			N	E
1	The intersection of Siyahrud-Shafarood-Rine village	990	4153237	303345
2	The intersection of Shafarood-Sarak river	184	4154381	326043
3	Under the Shafarood Bridge (Rezvanshahr-Talesh Road)	66	4155317	332788
4	Sandian Village (above Luleh Gaz Bridge)	0.3	4160137	334970
5	Shafarood estuary	019	4162464	337597

Samples were collected seasonally over a six-year period from March 20, 2017 to March 19, 2022. The samples were analyzed for nine parameters according to the Standard Methods for the Examination of Water and Wastewater 2015, as described in Table 2.

Table 2. Water quality parameters, units and methods of analysis used in Shafarood River (Gilan Province, northern Iran) during 2017 to 2022

Parameters	Abbreviations	Units	Analytical instrument and methods
Electrical conductivity	EC	$S\ cm^{-1}\mu$	Electrometer
Dissolved oxygen	DO	mgL^{-1}	Azide-Winkler titration method
Total suspended solids	TSS	mgL^{-1}	HACH LXV
Turbidity	Turb.	NTU	Turbidity Meter
pH	pH	pH unit	pH-meter
Nitrate nitrogen	No_3^-	mgL^{-1}	Spectrophotometry
Chemical oxygen demand	COD	mgL^{-1}	Dichromate reflex method
Biochemical oxygen demand	BOD	mgL^{-1}	Azide-Winkler titration method
Phosphate	PO_4^-	mgL^{-1}	Spectrophotometry

SPSS 24 and MINITAB 15 software were used to determine the correlation between parameters through the canonical correlation analysis (CCA) method, as well as to identify parameters influencing water quality and to group quality categories through the cluster analysis (CA) method. The normality of the data was checked by Kolmogorov-Smirnov (KS) test. Bartlett's test was used to determine the homogeneity of variances.

2.3 Canonical Correlation Analysis (CCA)

The CCA is one of the most common multivariate analysis methods to determine the linear correlation between multidimensional variables. Many statistical studies classify variables into two groups: predictor (independent) and response (dependent) variables. The CCA modeling examines the relationship between two sets of variables including *X* and *Y* matrices. For example, the present research used the canonical correlation matrix to interpret the relationship between two groups of response (physical parameters) and predictor (chemical parameters) variables. This method is considered as a tool to reduce the volume of data needed in calculations [1]. The CCA creates a set of canonical variables a for the category of variables *X* and a set of canonical variables b for the category of variables *Y*, similar to factor analysis. The goal of CCA is to construct two new categories of components as *U=Ax* and *v=By*, which are linear combinations of primary variables. In this method, the values of *a* and *b* are selected in such a way that *U* and *V* have the highest correlation with each other. This method continues until obtaining *m* sets of canonical variables, where *m* is the minimum number of variables of the examined groups. To find *a* and *b*, a correlation matrix must be formed between *X* and *Y* variables. For example, the matrix *(p+q)*(p+q)* is established between the variables *X1* to *Xp* as shown in Figure 2, which is obtained from sampling, that is, *p* stands for a physical parameter and *q* stands for a chemical parameter.

$$XY = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1mx} & y_{11} & y_{12} & \cdots & y_{1my} \\ x_{21} & x_{22} & \cdots & x_{2mx} & y_{21} & y_{22} & \cdots & y_{2my} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nmx} & y_{n1} & y_{n2} & \cdots & y_{nmy} \end{pmatrix}$$

Figure 2. Canonical correlation matrix

The correlation matrix was as follows:

$$R(xy) = \begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix}$$

In fact, this matrix was calculated based on the sample in which the variables were measured, and its eigenvalue could be calculated according to Equation (1).

$$(R_{yx} R_{yy}^{-1} R_{xy} R_{xx}^{-1} - \lambda I) f_i = 0 \quad \text{Eq. (1)}$$

In this case, the eigenvalues of *λ1〉 λ2〉 λ3〉 ...〉 λr* were the square of the correlation coefficients between the canonical variables. The eigenvectors of *fi* gave the coefficients of variables *Y* for the category of canonical variables. The coefficient *Ui* of the *i*th canonical variable for variables *X* was obtained by Equation (2).

$$g_i = R_{xx}^{-1} R_{xy} f_i \quad \text{Eq. (2)}$$

The amounts of U and V could be obtained based on *fi* and *gi* [14].

2.4 Cluster analysis (CA)

The CA method categorizes the data based on the distance between the members. Thus, the parameters with less distance from each other are placed in the same cluster. The purpose of CA is to create categories and groups where intra-group differences and variance are less than inter-group differences and dispersion. The distance method is often used to group two or more criteria, which takes help from Euclidean geometry to determine the distance of parameters from each other.

2.2 Sampling and Measurements

Sampling stations were selected based on the natural conditions of the region and the possibility of access to the river and taking into account the natural and anthropogenic effects, including river sub-branches and pollution sources, including agricultural lands or residential and tourism centers. Finally, five sampling stations were selected, whose coordinates were determined on the map by the Global Positioning System (GPS) (Table 1).

Table 1. Coordinates and locations of water sampling stations in Shafarood River (Gilan Province, northern Iran) during 2017 to 2022

The distance matrix is obtained according to the Euclidean distance between spatial or temporal points, and spatial and temporal groups are determined based on the distances of this matrix [22].

The steps of grouping the sampling stations via the CA method consisted of preparing the raw data matrix, determining the factor loading of each station using the factor analysis method, merging the groups using Ward's minimum-variance method, and finally grouping and drawing a dendrogram. At the beginning of the clustering process, there are clusters equal to the number of observations; and in the last step, the data are separated into a smaller number of clusters [13]. In this research, Ward's method was used for normalized datasets. Thus, first, the final mean value of the variables within each cluster was calculated, and then the squared Euclidean distance of the mean clusters was calculated for each observation [9].

3. Results and Discussion

In this research, CCA and CA methods were used to determine the correlation between physical and chemical parameters and to identify the main parameters affecting the Shafarood River water quality. Table 3 shows the results of the CCA method. The results indicated that the number of chemical variables (n=4) determined the number of canonical categories created. As can be seen, the canonical correlation coefficient for the category of canonical variables 1 and 2 was equal to 0.701 and 0.428 respectively, which assigned relatively high values. According to Tabachnick and Fidell [19], the correlation between the main canonical variables can be interpreted only with a value greater than 0.3. It is also noteworthy that the higher the correlation coefficients of the categories, the more decisively it is possible to judge the existence of a correlation between the categories of variables under investigation. Considering that the correlation coefficients of the third and fourth categories of canonical variables were lower than this value, they were not used for drawing conclusions. According to the P-value for each canonical category, it could be seen that only the canonical variables of the first and second categories were valid, and the first category showed a high degree of correlation compared to the second category; the first category was statistically significant at a probability level of 1 % and the second category was statistically significant at a probability level of 5 %. Therefore, with a confidence interval of 99%, it could be stated that the two categories of physical (dependent) and chemical (independent) variables in the first category, whose correlation coefficient was 0.701, had a high correlation, and then the correlation of the second category was justified with a confidence interval of 95%. Two other categories of canonical variables did not provide valid results regarding the existence of a correlation between parameters.

Table 3. The results of canonical correlation analysis for physical (dependent) and chemical (independent) parameters in Shafarood River watershed (Gilan Province, northern Iran) during 2017 to 2022

Canonical variables	Category 1		Category 2		Category 3		Category 4	
	V_1,U_1		V_2,U_2		V_3,U_3		V_4,U_4	
Canonical correlation	0.701		0.428		0.259		0.071	
Eigenvalues	9.66		2.24		0.72		0.05	
Degrees of freedom	20		12		6		-	
Significance level	6.086**		2.704*		1.416		-	
P-value								
Physical parameters	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading
DO	0.171	0.321	-0.864	-0.781	-0.108	-0.197	0.124	-0.113
EC	-0.202	-0.164	0.066	0.239	0.752	0.767	0.256	0.172
Turb.	-0.227	-0.546	0.001	-0.003	-0.419	-0.344	0.983	0.760
pH	0.145	-0.004	-0.502	-0.439	0.556	0.470	-0.294	0.346
TSS	-0.852	-0.925	-0.310	-0.288	0.074	-0.049	-0.584	-0.207
Chemical parameters	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading	Standard canonical coefficients	Canonical loading
BOD	-0.738	-0.698	1.236	0.576	-0.638	0.400	0.192	0.144
COD	0.128	-0.510	-0.813	0.059	1.406	0.798	-0.029	0.315
NO_3^-	-0.721	-0.739	-0.583	-0.603	-0.417	-0.279	-0.002	0.115
PO_4^-	0.151	0.118	0.147	-0.110	-0.307	0.054	0.996	0.985

** and ** = respectively indicating between-cluster significance at the levels of 5% and 1%*

Considering that an effective correlation was established between physical and chemical parameters with high canonical correlation for Categories 1 and 2, these two categories could indicate the fact that chemical and physical parameters probably had the same pollution sources. According to Table 3, the principle variables in the first canonical variables (Category 1) included physical variables related to U_1 and chemical variables related to V_1. The CCA results showed that the TSS variable in column U_1 of physical variables had canonical coefficients of 0.852, and thus they were more important, indicating the greater influence of the linear combination of U_1 (physical parameters) from this water quality variable. Among the chemical variables V_1, the parameters of BOD and NO_3^- had the canonical coefficients of 0.738 and 0.721, respectively, indicating that the increase in the value of BOD parameter had a relatively high correlation with V_1 compared to other parameters, so that increasing the value of this parameter enhanced the value of V_1. In the canonical category of physical variables U_1, the TSS parameter had a relatively high correlation with U_1; hence, as the value of TSS parameter increased, the value of U_1 also elevated. Based on the CCA results (Table 3), the physical variables of TSS and Turb. with canonical loading of 0.925 and 0.546, respectively, played an important role in predicting the canonical variable of chemical parameters. According to the canonical loading squared values (0.856 and 0.298, respectively), in predicting the variance of chemical variables, the share of TSS variable was 86% and the share of variable Turb. was 30%. On the other hand, the chemical variables of BOD and NO_3^- with canonical loading of 0.698 and 0.739, respectively, had a significant role in predicting the canonical variable of physical parameters, with the squared canonical load of 0.487 and 0.546, respectively. Therefore, in predicting the variance of chemical variables, the contribution of BOD and NO_3^- was 49%and 55%, respectively.

Rasi Nezami et al. [15] studied the water quality of Karkheh River (Iran) using CCA technique, the results of which demonstrated that the parameters of Turb., DO and TSS (as physical attributes) and the parameters of BOD, COD, total phosphorus and total nitrogen (as chemical attributes) had the highest scores among all canonical variables.

According to the results obtained for the correlation of each pair of *Ui* and *Vi*, it could be stated that the physical variables of TSS and Turb. as well as the chemical variables of NO_3^- and BOD were the dominant variables in terms of P-value among the parameters investigated in this research. Therefore, the results of this research could reveal a strong correlation between two categories of chemical and physical parameters. It could be claimed that both chemical and physical parameters originated from the same source. Since chemical factors in watersheds are often caused by anthropogenic activities, and physical factors can also be of anthropogenic or natural origin [4], based on the CCA results, it could be concluded that both chemical and physical parameters were of anthropogenic origin in the Shafarood River watershed in Iran. Nessi and Tishehzan (2018) as well as Rasuli et al. [16] used CCA and CA methods, and found that river water quality had been affected by anthropogenic pollution sources. Yaganoglu et al. [21] in the study of spatial and temporal changes in the water quality of the Filyos River (Turkey) using multivariate analysis methods concluded that the quality changes of the studied river were due to the introduction of agricultural pollutants.

Table 4 and Figure 3 show the results of CA and the dendrogram of parameters measured at five stations.

In this research, the CA method was used to check the accuracy of the correlation coefficient matrix obtained from the CCA method.

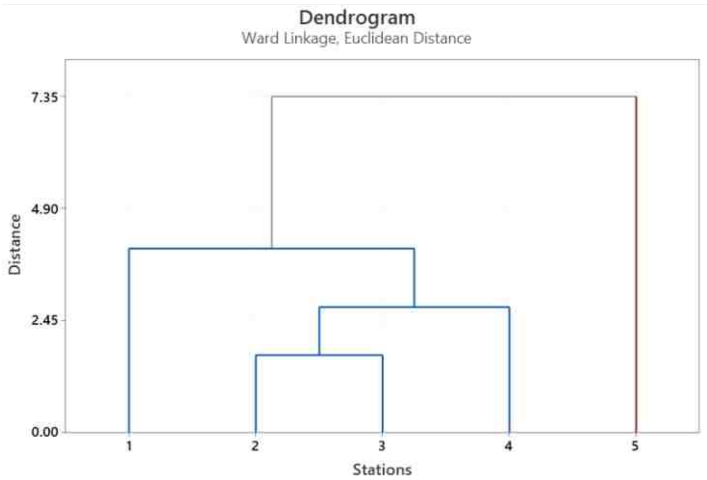

Figure 3. Dendrogram obtained from cluster analysis of sampling stations for parameters measured in Shafarood River Watershed (Gilan Province, northern Iran) during 2017 to 2022

Table 4. Total Mean and the averages in the two clusters of cluster analysis for the parameters evaluated in Shafarood River (Gilan Province, northern Iran) during 2017 to 2022

Clusters	Statistical parameters	TSS	pH	Turb.	PO_4^-	NO_3^-	EC	DO	COD	BOD
1	$\bar{x}$	50.15	7.99	103.2	2.09	1.46	332.3	8.89	5.34	1.64
2	$\bar{x}$	116.02	7.91	161.10	1.20	1.22	755.25	7.65	8.62	3.33
Mean	Total	83.09	7.95	132.15	1.65	1.34	543.77	8.27	6.98	2.48

Fan et al. [6] implemented the CA method to check the quality of stations in the north, east, and west of the Pearl River Delta in China and grouped the rivers into several clusters based on pollution severity. In the current work, as can be seen, the studied stations were categorized into two clusters based on the impact factor of the type and the values of the measured parameters using Ward's method. The first cluster covered Stations 1, 2, 3 and 4. These stations showed values higher than the average of all measured values in terms of NO_3^-, PO_4^- and pH parameters (Table 4), highlighting changes in water quality due to the introduction of anthropogenic pollution sources following agricultural (horticultural) and fish farming activities, as well as inflow of sewage from rural residential areas. The high concentration of NO_3^- and PO_4^- at Stations 1 and 2 could be attributed to the discharge of sewage as well as the infiltration of runoff from Rine and Sarak villages in these areas. The second cluster only included Station 5, which had higher average values for measured parameters including BOD, COD, EC, Turb. and TSS due to the introduction of scattered villas and residential areas as well as tourist restaurants in the upstream areas of the station, as well as erosion in the upstream areas due to clearcutting and smuggling of wooden trees and taking sand from the river bed. Bu et al. [2] investigated the spatial and temporal variations of the Jinshui River water quality in China using multivariate analysis methods and also identified the main pollution factors and their resources, and found that the pollution of the studied river was caused by the entry of domestic wastewater and agricultural runoff.

In general, it could be said that Station 5 (Second cluster) was the most polluted site, which had the greatest distance from the rest of the stations. The results of one-way analysis of variance (ANOVA) confirmed the significant difference between the resulting clusters in terms of most of the parameters studied at the probability level of 1 % and 5 %. The difference between the clusters indicated that the stations of each cluster had no significant difference in the measured parameters, while there was a significant difference between the clusters at the probability levels of 1% and 5% for most of the evaluated traits (Table 4).

To further investigate the accuracy of the canonical correlation matrix through the CA method, the water quality parameters of the Shafarood River watershed were clustered (Figure 4).

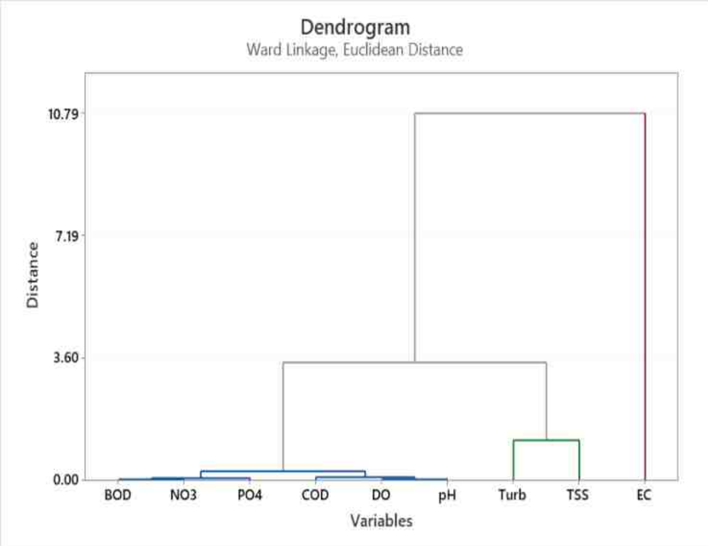

Figure 4. Dendrogram derived from cluster analysis of measured parameters for studied stations in the Shafarood River Watershed (Gilan Province, northern Iran) during 2017 to 2022

As can be seen in Figure 4, the EC parameter alone was placed in a cluster with a long distance from other parameters, and the TSS and Turb. parameters were in the same cluster at a close distance, and the rest of the parameters, which had a close connection in the correlation matrix, were in the third cluster. In the third cluster, it was found that the physical parameters of NO_3^- and BOD as well as the chemical parameters of DO and COD were closely spaced in the cluster diagram, so that there was a high correlation matrix coefficient between these parameters. In general, the CA dendrogram showed that the findings of the CA method for water quality parameters confirmed the results of the canonical correlation matrix.

Table 5 presents the Pearson correlation coefficient matrix for physical and chemical parameters. As it is clear, there was statistically significant between TSS and Turb., and also between TSS and BOD, COD and NO_3^-, as well as between Turb. parameter and BOD and NO_3^- (P<0.01). A significant correlation was found between BOD and COD as well as BOD and DO, Turb. and TSS and on the other hand between COD and TSS and also between NO_3^- parameter and Turb. and TSS (P<0.01). In addition, there was a significant correlation between BOD and EC, between COD parameter and EC and PO_4^-, as well as between DO and Turb. (P<0.05). The results of the canonical correlation matrix (Table 3) confirmed the correlation findings between the variables in the Pearson correlation coefficient (PCC), see Table 4. Similarly, Neissi and Tishehzan [13] investigated the physicochemical parameters affecting the water quality of Dez River in Iran, and declared that physical variables of EC and temperature and chemical variables of SAR, COD, SO_4^- and NO_3^- confirmed each other in affecting water quality based on canonical and Pearson correlations.

Table 5. Correlation matrix of physicochemical variables studied in Shafarood River (Gilan Province, northern Iran) during 2017 to 2022

Parameters	BOD	COD	DO	EC	NO_3^-	PO_4^-	Turb.	pH	TSS
BOD	1								
COD	0.754**	1							
DO	-0.371**	0.176	1						
EC	0.220*	0.229*	0.144	1					
NO_3^-	0.073	0.157	0.050	0.029	1				
PO_4^-	-0.026	0.201*	0.029	0.322	0.108	1			
Turb.	0.238**	0.141	-0.226*	0.085	0.314**	0.013	1		
pH	-0.054	0.098	0.120	0.055	0.086	0.039	0.156	1	
TSS	0.374**	0.310**	0.100	0.059	0.555**	0.076	0.375**	0.125	1

**and ** = respectively indicating between-cluster significance at the levels of 5% and 1%*

In total, the research results indicated a strong correlation between the physical and chemical parameters using the CCA method, as confirmed through the data obtained from the CA and PCC, in line with the research findings of Abdolabadi et al. [1] in examining the correlation between physical and chemical parameters and also determining the origin of their transmission in the water of the Atrak River basin.

4. Conclusions

The present research used multivariate statistical methods of CCA and CA to determine the correlation of the main pollution parameters and to classify the water quality monitoring stations in Shafarood River located in Gilan province in the north of Iran. According to the CCA results, the physical variables of TSS and Turb. and the chemical variables of BOD and NO_3^- were dominant among the parameters investigated in this research. Therefore, there was a strong correlation between two categories of chemical and physical parameters, indicating the same origin of pollution (anthropogenic pollution sources) in the studied river. Using the CA method, five sampling stations were grouped into two clusters with similar quality attributes in each cluster. The CA results indicated the anthropogenic pollution sources of the Shafarood River, and the grouping obtained from clustering based on parameters confirmed the CCA results. The decrease in the water quality of the Shafarood River in the studied stations was caused by the sewage of residential centers and agricultural (garden) drains and fish breeding ponds, as well as forest clearcutting and illegal sand harvesting from the bed (which causes an increase in mud and sediments). It is recommended to use multivariate statistical techniques in the interpretation of large-volume data, water quality assessment, detection and determination of the correlation of pollution parameters by providing appropriate information about water quality and as an effective and efficient method for decision making in river water quality management. The data obtained from the application of CCA and CA methods in the present study could provide appropriate information for the protection of the Shafarood River water resources, indicating the ability of these techniques to monitor water quality in the form of river water quality management.

5. Acknowledgements
This research has been extracted from the dissertation of the Ph.D student (Mr. Elham Gholami Deljomanesh) in Environmental Engineering, Islamic Azad University of Ardabil Branch, with the dissertation code of 162821187. Therefore, the authors appreciate the esteemed president, and educational and research deputies of Ardabil Islamic Azad University for their cooperation in facilitating the implementation of this project.

6. Conflict of Interest Statement
The authors have no conflicts of interest to declare.

7. Data Availability Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request

8. Authors' Contributions
Elham Gholami Deljomanesh: Conceptualization.
Dr. Ali Akbar Imani: Formal analysis
Elham Gholami Deljomanesh: Funding acquisition
Elham Gholami Deljomanesh: Investigation
Dr. Ebrahim Fataei: Methodology
Dr. Ebrahim Fataei: Project administration
Dr. Fatemeh Shariati Feyzabadi: Resources
Dr. Ali Akbar Imani: Software
Dr. Mahsa Hakimi Abed and Dr. Ebrahim Fataei: Supervision
Dr. Ebrahim Fataei: Validation
Elham Gholami Deljomanesh: Visualization
Elham Gholami Deljomanesh: Writing

9. Funding: No funding.
10. Scope of the study: The scope of this research is within the scope of water resource quality management in watersheds.

11. Conflict of interest: The authors declare no conflicts of interest.

12. Acknowledgment: This article has been adapted from the in Environmental Science)No 162821187(Ph.D. dissertation And Engineering by Ms. Elham Gholami Deljomanesh at Islamic Azad University, Lahijan Branch, Lahijan, Iran. The authors would like to express their gratitude for the support of this implementing the current project.university in

13. References

1. Abdolabadi, H., Ardestani, M., & Hasanlou, H. (2014). Evaluation of Water Quality Parameters Using Multivariate Statistical Analysis (Case Study: Atrak River). Journal of Water and Wastewater, Ab va Fazilab (In Persian), 25(3), 110-117.

2. 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, 907-913.

3. Boyacıoğlu, H., Gündogdu,V., &Boyacıoğlu, H. (2013). Investigation of priorities in water quality management based on correlations and variation. Marine Pollution Bulletin, 69(1,2), 48-54.

4. Chan, M. C., Lou, I. C., Ung, W. K., & Mok, K. M. (2013). Integrating Principle Component Analysis and Canonical Correlation Analysis for Monitoring Water Quality in Storage Reservoir. In Applied Mechanics and Materials, 284-287, 1458-1462. Trans Tech Publications, Ltd. https://doi.org/10.4028/www.scientific.net/amm.284-287.1458

5. Eskandare, A., Noure, R., Rasoule, A., & Vesale, M. (2015). Offering a Suitable Method for Water Quality Management of the Sefidrood River Based on Canonical Correlation Analysis. Environmental Researches, 5(9), 79-86.

6. 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.

7. 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

8. Gumus, E. (2022). A Case Study on the Relationship between Water Quality Parameters: Bursa. Sakarya University Journal of Science, 26(5), 867-878. https://doi.org/10.16984/saufenbilder.1083427

9. 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 Managment, 70(2), 465-478. doi: 10.22059/jrwm.2017.109430.779

10. Khanduzi, F., Parizanganeh, A.H., & Zamani, A.A. (2016). Canonical correlation analysis: the relationship between water qualityparameters and heavy metals in water samples (Ramian-Golestan province, Iran). Journal of Water and Soil Conservation (In Persian), 23(4),267-280.

11. Khoda bakhsh, S., Rafiei, B., Afshar Nia, M., Kabiri, S., & Ekhlas mand, R. (2017). Provenance determination of dissolved load of Khar-Rud River water (S Qazvin Province) using statistical parameters. Quaternary Journal of Iran, 2(4), 367-378. doi: 10.22034/irqua.2017.701966

12. Mackialeagha, M., Salarian, M. B., & 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, 6(2), 1-9. doi: 10.22034/ap.2022.1965587.1134

13. Neissi, L., & Tishehzan, P. (2018). Dez River Water Quality assessment by using multivariate statistical methods. Irrigation and Water Engineering, 9(1), 139-150.

14. Noori, R., Sahabi, M.S., Karbasi, A.R., Baghvand, A., & Taati-Zadeh, H. (2010). Multivariate statistical analysis of surface water quality based on correlation & variations in the data set. Desalination, 260, 129-136.

15. Rasi Nezami, S., Nazariha, M., Baghvand, A., &Moridi, A. (2013). Karkheh River Water Quality Using Multivariate Statistical Analysis and Qualitative Data Variations. Health System Research, 8(7), 1280-1292.

16. Rasuli, A., Noori, R., Vesali-Naseh, M. R., Nazariha, M., & Kiaghadi, A. (2011). Determination of the Relationship Between Physical and Chemical Pollutant Parameters in the Sefidrood River Basin Based on Canonical Correlation Analysis. In Proceedings of 4th Iran Water Resources Management Conference (pp. 23-24).

17. 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 (Case study: Aras catchment in the border area of Iran and Armenia). Anthropogenic Pollution, 5(1), 41-48. doi: 10.22034/ap.2021.1912491.1082

18. Sajjadi, N., Davoodi, M., & Jozi, S. (2019). The quality assessment of Kan River's resources in terms of agricultural and drinking purposes. Anthropogenic Pollution, 3(1), 46-53. doi: 10.22034/ap.2019.582368.1036

19. Tabachnick, B. L. & Fidell, S. (2000). Using multivariate statistics. A Pearson education company, Needham Heights, 966 pp., USA.

20. Valiallahi, J. (2022). Groundwater quality Zoning Based on Wilcox Index Using Geographic Information System in Jajarm district, north Khorasan, Iran. Anthropogenic Pollution, 6(2), 16-24. doi: 10.22034/ap.2023.1964980.1136

21. Yaganoglu, E., Yaganoglu, A. M., Arslan, G., &Sonmez, A. Y. (2020). Determination of Spatial and Temporal Changes in Surface Water Quality of Filyos River (Turkey) Using Principal Component Analysis and Cluster Analysis. Marine Science and Technology Bulletin, 9(2), 207-214. https://doi.org/10.33714/masteb.784959

22. Zamani, A., Pari Zanganeh, A.H., &Khandoozi, F. (2016). Canonical correlation analysis to determine the relationship between water quality parameters and heavy metals in water samples (Ramian City-Golestan Province). Journal of Water and Soil Conservation, 23(4), 267-280. doi: 10.22069/jwfst.2016.9936.2431