Journal of Applied Research in Water and Wastewater

E-ISSN: 2476-6283

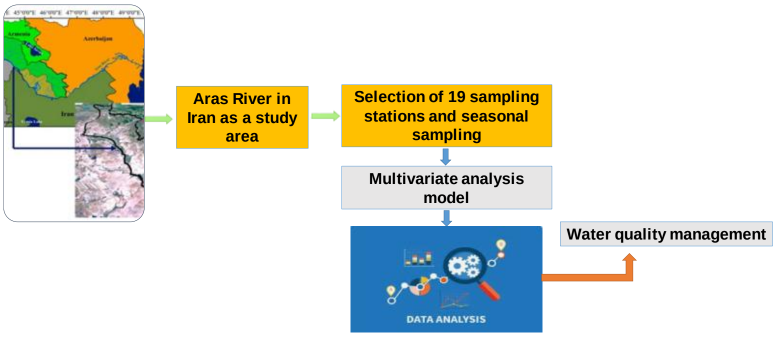
Journal homepage: https://arww.razi.ac.ir

Evaluation of spatial qualitative changes in surface water using cluster analysis and factor analysis: A case study on Aras River within the boundaries of Iran

Mohammad Mostafa Shabrang[1], Ebrahim Fataei[*,1], Ali Akbar Imani[*,2], Hooman Bahmanpour[3], Mohammad Shabani[5]

[1]*Department of Environment, Ardabil Branch, Islamic Azad University, Ardabil, Iran.*
[2]*Department of Agriculture, Ardabil Branch, Islamic Azad University, Ardabil, Iran.*
[3]*Department of Environment, Shahrood Branch, Islamic Azad University, Shahrood, Iran.*
[4]*Department of Water Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran.*

GRAPHICAL ABSTRACT



ARTICLE INFO

Article type:
Research Article

Article history:
Received 25 July 2024
Received in revised form 27 October 2024
Accepted 29 October 2024
Available online 4 November 2024

Keywords:
Water quality
Qualitative classification
Aras River
Multivariate analysis
Cluster analysis
Principal component analysis

ABSTRACT

Surface water quality classification for use in a variety of applications such as drinking, agriculture, and industry is very significant. Different methods and indicators are employed to determine water quality. In this study, to assess the quality of the international Aras River, 16 parameters were analyzed at 19 stations with seasonal data collected over two years (2020 and 2021), and multivariate statistical analysis methods, including cluster analysis and factor analysis, were utilized. The cluster analysis results categorized the studied stations into four clusters based on quality. The primary parameters influencing the grouping of water quality at the stations were as follows: BOD, COD, and T. Coli in the first cluster; T. Coli and NO_3^- in the second cluster; TDS, EC, and Turbidity in the third cluster; and BOD, COD, TDS, EC, and Turb. in the fourth cluster. The principal component analysis and factor analysis results indicated that the first two components explained 86% of total variance. In the first component, with an eigenvalue of 5.94, the most influential parameters in the qualitative classification of the stations included pH, DO, EC, T. Coli, NO_3^-, and Hg. In the second component, with an eigenvalue of 2.72, parameters of BOD, COD, turbidity. and AS played the most significant role in creating quality differences among the stations. Therefore, based on the obtained results, it was revealed that the reason for qualitative changes at different stations is due to the entry of human pollutants from various urban, industrial, mining, and agricultural sources as well as erosion in the river basin. Therefore, given the high precision of analytical methods used in evaluation of qualitative aspects of the studied river's water, it can be acknowledged that multivariable methods including cluster analysis and factor analysis can confidently determine water quality of rivers and significant parameters, affecting their quality, as well as identify pollutants in the management of river water quality.

1. Introduction

The surface water is influenced by a wide range of natural and artificial pollutants (Kazemi Noredinvand,Takdastan and Jalilzadeh Yengejeh, 2015; Ebadati, 2017). These factors affect the physical, chemical, and biological characteristics of freshwater (Jalilzadeh Yengejeh, Morshedi and Yazdizadeh, 2014; Mostafavi and Teimori, 2018), in such a way that the extensive length of rivers and their passage through various

*Corresponding author Email: eb.fataei@iau.ac.ir

Cite this article: M.M. Shabrang, E. Fataei, A.A.Imani,H. Bahmanpour, M. Shabani, Evaluation of spatial qualitative changes in surface water using cluster analysis and factor analysis: A case study on Aras River within the boundaries of Iran, *Journal of Applied Research in Water and Wastewater*, 11 (2), 2024, 143-150. doi: https://doi.org/10.22126/arww.2025.10788.1337

urban, industrial and agricultural areas have resulted in rivers being exposed to different types of physical, chemical, and biological pollution (Sadjadi, Davoodi and Jozi, 2019; Jalili, 2020). Therefore, assessing and managing river water quality for various uses are important (Muangthong and Shrestha, 2015). An important issue in water quality analysis is identifying and differentiating important pollutant parameters (Mohammadi *et al.,* 2023). Due to the size of the Aras River, quality of its water can be affected by various natural and human-made pollutants. Therefore, by assessing water quality of this river and categorizing the quality of its monitoring stations, pollutant sources and key parameters influencing the quality of the stations through the measurement of physicochemical and biological parameters can be identified.

Water quality monitoring is a critical priority for the preservation of water resources. Long-term water quality monitoring is the most common and effective method for assessing environmental issues. This approach delineates the spatial and temporal changes in quality parameters, and consequently provides researchers with greater assistance in evaluating pollution status (Kumar *et al.,* 2021). Various approaches are used to analyze and interpret variables determining observed changes in water quality in different sources. Nonetheless, interpretation of a complex and extensive data matrix consisting of many physicochemical parameters from long-term monitoring programs is difficult. Statistical methods, especially multivariate statistical techniques constitute the significant part of these approaches. Multivariable techniques include Cluster Analysis (CA), Principal Component Analysis (PCA), and Factor Analysis (Mackialeagha, Salarian and Behbahaninia, 2022). Cluster analysis is a suitable method for analyzing water sample data and is essential for evaluating water quality data and possibility for the qualitative grouping of samples (Zhang *et al.,* 2012). Therefore, given the fact that these methods are considered valuable tools for reliable assessment and management of water resources for the purpose of development of quick solutions in connection to pollution management (Sayemuzzaman *et al.,* 2019; Aydin *et al.,* 2021), the present study also utilized the aforementioned methods to better interpret the overall data for a clearer understanding of water quality in the Aras River.

Haji Gholizadeh *et al.* (2017) in a study conducted on 3500 water quality data related to the South Florida River, classified the river into some parts with low-to-high pollution levels using multivariable statistical techniques and determined that parameters of DO, TKN, TP, and T had the most significant impacts in terms of temporal changes among seasons (Haji Gholizadeh *et al.*, 2017). Belkhiri *et al.* (2015) in a study conducted on evaluation of groundwater resources quality in the Al-Ain Azel Plain located in Algeria using cluster analysis and factor analysis methods, classified the studied wells into two groups based on quality (Belkhiri *et al.*, 2015). They reported that, based on the factor analysis results, 85% of total variance of quality changes in this plain was created by the initial two factors. Dissolution of geological formations in the riverbed and pollution from human resources are among the main effective factors on the studied water resources quality. In the first component, values of parameters of EC, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, and SO_4^{2-} were high, which were due to geological formations. In the second component, values of parameters of NH_4^{+}, NO_3^{-}, and COD were higher, which were attributed to human activities.

Ildorom, Hassanzadeh and Hedayatzadeh (2022) in a study on evaluation of the spatial-temporal changes of the quality parameters of the Karkheh River in Iran using multivariate statistical analysis, showed that PCA analysis identified three principal components as responsible for the data structure, which accounted for 77.65% of total variance of water quality (Ildorom, Hassanzadeh and Hedayatzadeh, 2022). Based on the results obtained from this analysis, parameters of TDS, EC, HCO_3^{-} and pH were the most important parameters of these components and had a significant contribution to the spatial changes of surface water quality of the Karkheh River. The results of cluster analysis also showed three similar groups among the 7 monitoring sites. Rezaei *et al.* (2017), using analysis of principal components on the qualitative water data at two stations of the Qarasu River declared that the first two principal components justified a total of 36.90% of variations, and changes were due to chemical parameters and entry of industrial wastewater, human sewage and agricultural runoff (Rezaei *et al.*, 2017). Noshadi and Ghafourian (2016) utilized multivariate statistical techniques to assess the groundwater quality in Fars province (Noshadi and Ghafourian, 2016). Through factor analysis, it was shown that more than 80% of total variance of changes in the groundwater quality of the studied area was explained by the first three factors wherein the significant parameters were SAR, Ca^{2+}, SO_4^{2-}, mg^{2+}, Cl^{-}, TDS, EC, TH, and Na^{+} in the first factor, HCO_3^{-}in the second factor and NO_2^{-} and NO_3^{-} in the third factor.

Mohammadi Ghaleni and Kardan Moghaddam (2022) in a study on evaluation of water quality of the Sefidroud River in Iran, using multivariate analysis reported that the results of CA indicated that water quality parameters are located in separate clusters, so that only the parameters TDS, EC, Cl^{-} and Na^{+} in both Rudbar and Astaneh stations were located in the same cluster (Mohammadi Ghaleni and Kardan Moghaddam, 2022). The weights of the parameters showed that TDS and K^{+} are assigned, respectively, with the highest and lowest weights equal to 0.163 and 0.031 based on the PCA method. Also, PCA results showed that the first and second principal components covered 59.3% and 67.6 % of total variance of measured water quality parameters in Rudbar and Astaneh stations, respectively. Ghassemi Dehnavi *et al.* (2016) assessed the qualitative and quantitative aspects of surface waters using statistical analysis in the Azna River, located in Lorestan province (Ghassemi Dehnavi *et al.*, 2016). They announced that the first three components by over 70 % variability justified the statistical population. Based on cluster analysis, the stations were classified into four clusters in terms of quality in the first cluster, the main reason for changes was due to SO_4^{2-} and pH parameters, in the second cluster K^{+}, in the third cluster Ca^{2+}, and in the fourth cluster TDS, EC, Cl^{-}, Mg^{2+}, HCO_3^{-}, Na^{+}, and TH parameters. De Andrade Costa *et al.* (2020) in an assessment of water quality based on multivariable statistics and water quality index on the strategic rivers of the Pantanal in Atlantic Ocean forests in Brazil, announced that cluster analysis and principal component analysis identified the most effective parameters in water quality and their correlation with the primary pollution sources, highlighting wastewater discharge as the main source of river pollution (De Andrade Costa *et al.*, 2020). Yağanoğlu *et al.* (2020) used ten water quality parameters measured at seven stations over twelve months through principal component analysis and cluster analysis to investigate the spatial and temporal changes in water quality of the Filyos River in Turkey (Yağanoğlu *et al.*, 2020). PCA results indicated that 4 principal components explained 69.49 % of total variance in water quality changes of the Filyos River. The predominant parameters of river water quality included temperature, EC, DO, and pH. The reason for the qualitative differences between the studied stations was due to entry of wastewater from agricultural activities. Giao (2022), in a study on assessment of surface water quality in the Ca Mau peninsula in Vietnam, using multivariate statistical analysis revealed that in terms of principal component analysis, the most effective parameters in seasonal water variations were DO, TSS, BOD_5, and pH, wherein the first two components explained 76.91 % of the variations (Giao, 2022). Also, based on cluster analysis, water quality was classified into seven clusters from poor to severely polluted, wherein the most effective parameters in clustering were DO, TSS, BOD_5, COD, and total coliforms (T. Coli). Considering that the international Aras River passes through several countries including Turkey, Armenia, Iran, and Azerbaijan, receives various types of pollution, and its water is utilized for various purposes such as drinking, industrial, and agricultural use, therefore, this research aims to identify the parameters affecting river water quality to determine pollution resources at various stations and classify the quality of stations using multivariable statistical methods to establish an appropriate management approach for controlling water quality of the Aras river.

2. Methodology
2.1. Studied area

River is one of the important border rivers of Iran. The watershed basin of this river covers parts of the territories of Turkey, Armenia, Azerbaijan, and Iran. The length of this river is 1072 km, and the area of the watershed basin is 97,000 km^2. It is susceptible to various types of pollution including adjacent urban-rural sewage as well as agricultural and industrial effluents. This river is the main source for providing water for various agricultural, domestic, industrial, recreational, and drinking sections and electricity production in the region. River basin in Iran covers an area of 39,478 square kilometers, approximately 81.5 % of which is located in mountainous and foothill regions, while the remainder consists of plains scattered throughout various parts of the basin (Salahi, Saber and Mofidi, 2022). From a macroclimatic perspective, this basin exhibits characteristics typical of temperate regions at mid-latitudes. However, the presence of mountainous and elevated areas such as the slopes of Sabalan and Ararat, along with plains situated below 500 meters, have led to significant climatic diversity within this basin (Aghayari Samian *et al.,* 2022). A total of 45 rivers flow through this basin, with the Aras River being the main artery and the primary source of freshwater for the Republic of Azerbaijan. It also provides water for agricultural, industrial, and drinking purposes for other countries within the basin. This river flows through the borders of Iran and West Azerbaijan, East Azerbaijan, and Ardebil provinces and after receiving multiple branches such as

Zangmar River, Sari Su, Aland, Qatur, Arpachay, Darreh Rud, etc. is separated from the Iranian border near Tazehkand village of Parsabad Moghan in Ardebil province and enters the territory of the Republic of Azerbaijan and meets Kura River in this country and discharges about 0.21 km^3 of water into the Caspian Sea annually on average (Khoshnoodmotlagh *et al.,* 2020).

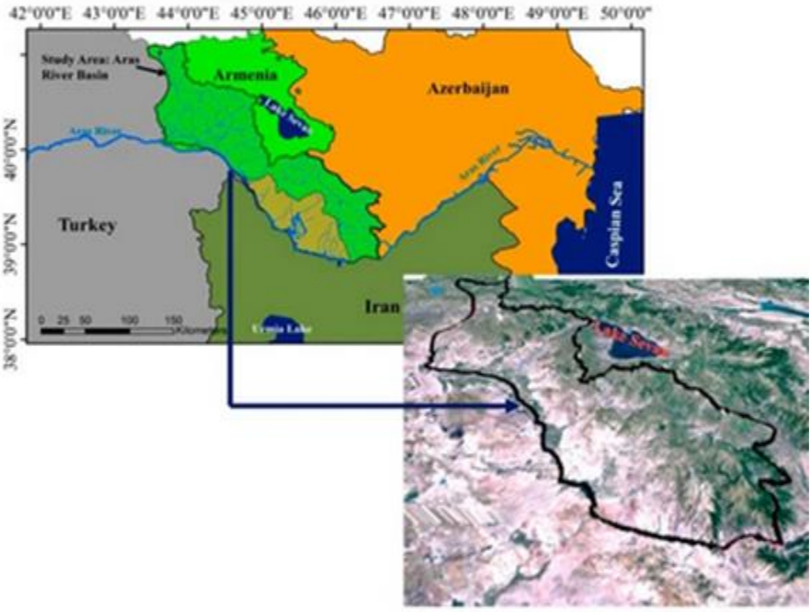


Fig. 1. Location of the Aras River and its studied area in Iran (Khoshnoodmotlagh *et al.*, 2020).

2.2. Data gathering and statistical methods

The present study aimed to assess water quality of the Aras River in Iran based on the qualitative data collected from 19 sampling stations from entry point of the river into Iran, located at the border of Turkey to its exit point into Turkey. Sampling was conducted seasonally over two years from spring 2020 to winter 2021 (Table 1). Parameters included 16 factors such as EC, pH, NO_3^-, COD, BOD, total Coli. (T. Coli), Fecal Coli. (F. Coli), TDS, SO_4^{2-}, Hg, Pb, Cd, Fe, Cu, and Ni, which were measured through following the guidelines outlined in the standard method (Rodger B.B., Andrew D.E. and Eugene W.R., 2017).

Table 1. Geographic location and coordinates of sampling stations in the Aras River within the boundaries of Iran.

No.	Location	Coordinates	
		X	Y
1	Entry to Iran	4381422	490451
2	Qarah Khvojalu	4363202	497322
3	Sanam Balaghi	4357243	504796
4	Zangmar entrance	4356332	506821
5	Downstream of Poldasht	4353802	506871
6	Dam tank	4328339	532306
7	Dam outlet	4325377	534376
8	Before Agh Chai station	4316485	538794
9	Agh Chai	4315187	542922
10	After a mix of Agh Chai	4313561	547311
11	Jolfa entrance	552469	4312934
12	Above of Mil Moghan dam	704238	4367735
13	Before the confluence of the sugar factory's sewage with the Aras river	736405	4387667
14	After the confluence of the sewage of the sugar factory with the Aras river	736440	4387718
15	Before the confluence of the urban sewage with the Aras River	754200	4397811
16	After the confluence of urban sewage with the Aras River	757491	4397770
17	After the confluence of Darreh Rud with the Aras River	706855	4371067
18	Before the confluence of Darreh Rud with the Aras River	705556	4371067
19	In front of Mil Moghan dam	704238	4367735

Sampling was conducted using glass bottles. The sampling containers were first rinsed with detergent and regular water, then with nitric acid, and finally with distilled water. For samples of COD and BOD parameters, dark glass containers were used. One-liter polymer containers were employed for sampling of heavy metal measurements, which ensures that they were thoroughly rinsed with detergent and distilled water before sampling, and rinsed once again with sampling water at the time of collection. Samples were taken from a distance of two meters from the water's edge and at a depth of 20 centimeters in the river. Before transporting the samples to laboratory, stabilization of the samples with the required chemicals, determination of water and air temperature, and preparation of labels (including station specifications, sampling time, and weather conditions) were taken place. After the necessary preparations, the samples were transported to the laboratory under appropriate temperature conditions and within the shortest possible time, and they were stored in a refrigerator until analysis. The analysis of the samples was conducted according to the protocol outlined in standard method (Rodger B.B., Andrew D.E. and Eugene W.R., 2017).

2.3. Statistical analysis

For analyzing the obtained data, multivariate statistical methods such as Cluster Analysis (CA), Discriminant Analysis (DA), Principal Component Analysis (PCA), and Factor Analysis (FA) were used as follows. The analyses were performed using SPSS24 and MINITAB15 software. The Kolmogorov-Smirnov (K-S) statistic was used for normal distribution and fit of the data. The Bartlett test was used to evaluate the normality of the data in assessing the principal components.

2.4. Cluster analysis

In the cluster analysis method, data grouping is done based on the distance between them, and Euclidean geometry was used to calculate the distances between clusters, in such a way that, by calculating the Euclidean distance between spatial or temporal points, a distance matrix is obtained to define spatial and temporal groups based on this matrix distances. At the beginning of the cluster analysis process, there are as many clusters as there are observations, and at the last stage, the data is separated into a smaller number of clusters. In this way, similar sets, being less distant from each other are placed in a cluster (Fataei and Shiralipoor, 2011).

Two distinct methods in cluster analysis include "Hierarchical Cluster Analysis" (HCA) and "Non-Hierarchical Cluster Analysis" (NCA). In the present study, HCA was used, which can be compressible, in such a way that each parameter is initially placed within its cluster, next two groups with higher similarities are combined or divided, which in the process of division, initially, all parameters are placed within one cluster, and then at each stage, the more distinct parameters are separated and create a smaller cluster. Hierarchical relationships between groups are visually represented in a tree diagram. In non-hierarchical cluster analysis, often referred to as "K-mean clustering", there is no diagram structure and it requires a predefined number of classes and parameters within each cluster are based on a pre-defined threshold distance (Muangthong and Shrestha, 2015).

2.5. Principal component /factor analysis

The principal component analysis is a multivariable statistical method, which can be used as an appropriate solution for reducing the volume of data and the complexity of initial variables as well as better interpretation of information in cases where the data volume is substantial, in such a way that these parameters may describe real variations in population (Saavedra *et al.,* 2013). The Kaiser-Meyer-Olkin (KMO) test is used to ensure appropriateness of the principal component analysis. Furthermore, Varimax rotation is employed to enhance the relationships between inputs and initial factors, and to better distinguish them for inclusion in the components. In fact, in PCA, a correlation matrix is constructed, and subsequently components, which cover a small portion of population variations are disregarded (Zamani-Ahmadmahmoodi *et al.,* 2019).

Choosing a correlation matrix and standardization mean that regardless of the actual data magnitude, each variable should have equal importance. To achieve this goal, standardized values using the method defined in Eq. 1 should be used instead of using the values of either variable (Fataei, *et al.,* 2017).

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

where, z is the standardized variable, x is the value of variable, and μ and σ are the mean and standard deviation of variables respectively. In the next step, some principal components capable of providing a comprehensive description of variables were determined, and they are grouped under these components based on their correlation. Finally, the primary variables in each component are determined based on the maximum factor loadings (Lausch and Herzog, 2002).

3. Results and discussion

3.1. Cluster analysis

The cluster analysis method was used for evaluation of water quality degrees, similarity at sampling stations and classifying them. The results obtained from the cluster analysis of 19 studied stations in the Aras River in Iran, in terms of the evaluated parameters, are shown in Fig. 2. In this diagram, X-axis represents the distances between homogeneous groups, and Y-axis represents the names of the sampling stations. Based on the specified Euclidean distance for clustering, indicated in gray within the diagram, the stations were divided into four clusters based on their qualitative similarity. Stations 1, 3, and 7 are in the first cluster, stations 2, 10, 4, 8, 5, 11, 6, and 9 are in the second cluster, stations 12, 13, 15, 17, 18, and 19 in the third cluster, and stations 14 and 16 in the fourth cluster.

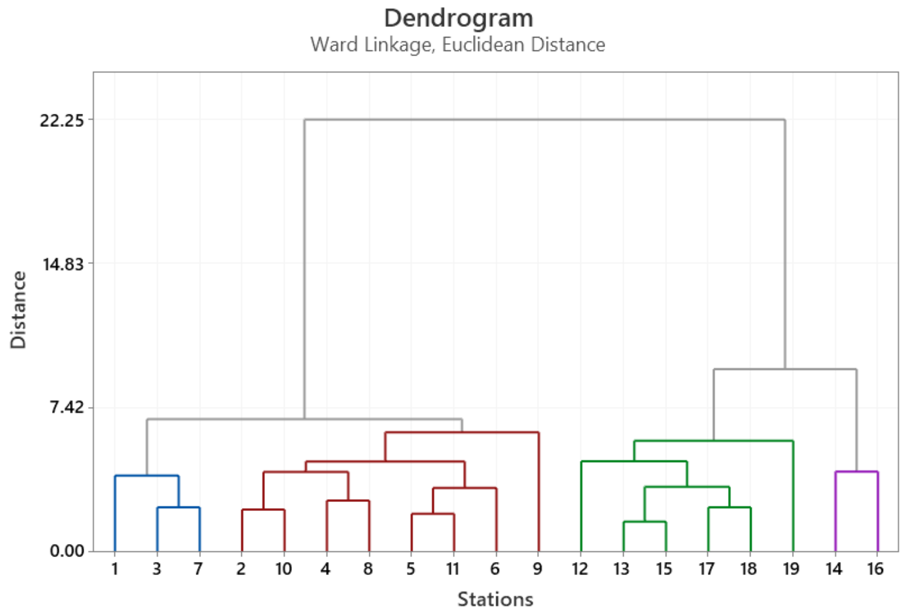


Fig. 2. Cluster diagram of qualitative grouping of the studied stations based on parameters measured in the Aras River, Iran.

Cluster analysis places a set of variables into homogeneous clusters so that the clusters obtained from this method have more internal homogeneity and external heterogeneity (Muniz *et al.,* 2023; Govender *et al.,* 2020). The Aras River is affected by various pollutants such as untreated sewage from residential communities, heavy metals, hydrocarbons, PCBs from industries and mines, nutrients, and organochlorine pesticides from agricultural wastewater in its watershed basin. Due to lack of development and proper efficiency of sewage collection and treatment systems in Armenia and Azerbaijan, the pollution of internal branches of the rivers in these countries is very high in terms of organic matter (TACIS 2003). Additionally, organic sediment load resulting from deforestation and agricultural activities further contributes to the above factors. Based on the results of cluster analysis and dendrogram chart shown in Fig. 2, the qualitative classification of the studied stations in this river within the territory of Iran is as follows:

Cluster 1 includes the entrance stations to Iran, Sanam Balaghi, and exit of Ares Dam. Based on the results obtained from average values of the measured parameters in the clusters and overall average (Table 1), values of BOD, COD, T. Coli., and NO_3^- parameters at these stations are higher than overall average of all stations. The high levels of these parameters at Sanam Balaghi station result from entry of agricultural runoff, fish farming wastewater, as well as wastewater from villages and towns located on fringes of the river within the vicinity of Nakhchivan (an autonomous republic of Azerbaijan), primarily consisting of untreated household sewage from the Sharur district and Shah Takhti City of this republic. Furthermore, in this area (an effective area on the Aras River), most agricultural activities are related to the Sharur district of this republic. In such a way the arable land area in this region is approximately 15,000 hectares and is currently being utilized. Agricultural drainage waters from this area are discharged into the Aras River through four main drainage channels. Furthermore, at the entrance station to Iran, the pollutants of the route traveled in Turkey due to the return effluents from the lands of Igdir Plain and change in water quality caused by the igneous nature of Ararat, entry of effluents from industrial areas in Armenia and a 200-hectare fish farm where different types of fish are grown, including sturgeon, which discharge all the polluted effluents directly into the river, which mostly include nutrients such as nitrates and phosphates as well as the sewage of the Poldasht city and Zangbar River in Mako, Iran, and the sewage of these cities all affect the outlet water quality of Aras Dam.

Cluster 2 includes the stations of Qarah Khvojalu (affected by the discharge of wastewater from upstream villages and wastewater from fish farming in Armenia), Zangmar (affected by entry of untreated sewage from Maku city), downstream of Poldasht (affected by the discharge of untreated sewage from Poldasht City), Aras Dam reservoir (affected by the discharge of mine wastewater and fish farming in

Armenia, as well as agricultural land runoff), before the Agh Chai station, Agh Chai, after the confluence of Agh Chai (quality of this station is affected by the connection of Qatur River subsidiary branch, which is severely affected by the erosion of watershed basin and has a high sediment load, and also receives treated wastewater from Khoy City and agricultural runoff) and entrance of Jolfa (affected by entry of septic tank effluents from the Aras residential town and also leakage from the absorptive wells of this town. In these stations, levels of T. Coli and NO_3^- parameters were higher than overall average.

Cluster 3 includes the areas, being higher than Mil Moghan Dam, before confluence of the sewage from the sugar cane factory of Moghan Agro-Industry and Livestock Co. to Aras River, before confluence of the urban sewage of Parsabad to Aras River, after confluence of the subsidiary branch of Darreh Rud with Aras River, and in front of the Mil Moghan Dam. In these areas, only the values of physical parameters such as TDS, EC, and turbidity were higher than overall average, which could be due to entry of drainage of extensive agricultural lands in the region, as well as sedimentation loads and high suspended solids in the Darreh Rud River resulting from increased destruction of vegetation coverage in the watershed basin of this river. However, stations in this cluster had lower levels of the chemical and biological pollutants compared to overall average, indicating the favorable quality status of these stations and absence of human pollutants within the range of these stations. This also signifies that the river has a suitable self-purification capacity along its path and has purified the received pollutants.

Cluster 4 includes the stations after confluence of effluents of the sugar factory of Moghan Agro-Industry and Livestock Co. into the Aras River, as well as the station after confluence of Parsabad urban sewage into the Aras River. In these stations, levels of BOD, COD, TDS, EC, and turbidity parameters were measured. Water quality changes in these stations are mainly due to discharge of industrial and urban pollutants. The high pollutant concentrations at station No. 14 are attributed to the discharge of effluents from the sugar factory of Moghan Agro-Industry and Livestock Co., while at station 16, it is due to the discharge of treated sewage from the Parsabad urban sewage treatment plant.

In contrast with similar studies conducted on other rivers, Giao (2022) assessed water quality of the rivers in the Ca Mau Peninsula, Vietnam, and classified it into seven clusters ranging from poor to heavily polluted based on a cluster analysis. The most significant parameters influencing this clustering included DO, TSS, BOD_5, COD, and total coliforms. Meanwhile, Ildorom, Hassanzadeh and Hedayatzadeh (2022) reported that water quality classification at the studied stations on the Karkheh River fell into three clusters due to variations in values of parameters such as TDS, EC, K^+, Mg^{2+}, and Ca^{2+}, particularly noting high values in the third cluster compared to other two ones (Ildorom, Hassanzadeh and Hedayatzadeh, 2022). Additionally, Bab al-Hakami and Ghollami (2018) in their research on water quality of the Talār River in Mazandaran, divided the studied stations into three groups based on quality, attributing the differences among the stations and the resulting clustering to the influx of industrial, domestic, hospital, and agricultural wastewater into the downstream stations (Hakami and Ghollami, 2018). In further research, Yağanoğlu *et al.* (2020) indicated that quality variations observed among the studied stations of the Filyos River in Turkey resulted from wastewater discharge from agricultural activities (Yağanoğlu *et al.*, 2020).

On the other hand, according to the obtained results (Table 2), heavy metals such as mercury and arsenic have been observed in all studied stations. Based on the analysis of pollutants in Aras River watershed basin, it was identified that this river is affected not only by the runoffs from accumulated tailings of past mining activities in the country of Armenia but also from active mines in this country, which is mainly located on the southern side of Aras River watershed basin on the border of Iran in East Azerbaijan Province, where the use of various organic and inorganic chemicals for separating metal from ore leads to entry of different chemical substances and heavy metals into water sources. Moreover, in industry, mercury pollution generally originates from lamp manufacturing industries. In this regard, there is a large lamp production factory called “Electrolamp” in Yerevan, the capital of Armenia, whose wastes were discharged without treatment to the water sources of the Ares River basin in this country in the past years. Another source of pollution in East Azerbaijan, Iran, may arise from mining and mineral processing activities, specifically the extraction of copper and gold. On the other hand, considering that there are vast agricultural lands in the watershed of the Aras River, wherein the effluents containing the remains of poisons and fertilizers enter this river, it can be considered as another source of the presence of these heavy metals in the different stations of the Aras River. A study conducted by Abbasi *et al.* (2017) on the impact of mining and exploitation of gold mines on the pollution of water in the Zarshooran region of Takab showed that the high levels of arsenic and mercury in the surface waters of the region are due to the gold deposit in this area (Abbasi *et al.*, 2017). The study conducted by Rafiei *et al.* (2010) also provided similar results regarding the pollution of heavy metals such as mercury and arsenic around the Dashkasan gold mine (Rafiei *et al.*, 2010). According to the findings of a study conducted by Behbahani Nia and Farahani (2016) on investigating the source of water pollution in the Hashtroud region in East Azerbaijan; the release of arsenic from the geological units of the region, absorption of arsenic by oxide minerals under aerobic and acidic conditions to nearly neutral, and absorption from oxide surfaces due to an increase in pH to alkaline conditions and an increase in arsenic concentration in the solution were identified as contributors to contamination (Behbahani Nia and Farahani, 2016).

Based on qualitative grouping obtained for the sampling stations of Aras River in Iran, pollution source was identified from various human activities, and cluster analysis grouping effectively classified the studied stations based on the susceptibility against parameter values and received pollution loads. Ling *et al.* (2017) also using cluster analysis multivariate statistical methods demonstrated that water quality in Kanto City, Malaysia, was influenced by various human pollutants including industrial effluents, domestic sewage, agricultural runoff, and erosion (Ling *et al.*, 2017). Ghassemi Dehnavi *et al.* (2016), in their study on qualitative and quantitative assessment of the Azna River in Lorestan, also classified the stations into four clusters based on the results from the cluster diagram (Ghassemi Dehnavi *et al.*, 2016). The first cluster was affected by SO_4^{2-} and pH parameters, the second cluster by the potassium ion parameter, the third cluster by high calcium ion levels, and the fourth cluster by TDS, EC, Cl^-, Mg^{2+}, HCO_3^-, Na^+, and TH parameters. Cheraghi *et al.* (2019) classified the studied stations into two groups in grouping groundwater quality in Khorramabad region using cluster analysis and stated that the main difference between the two clusters depended on factors such as type of formation and lithology (Cheraghi *et al.*, 2019). As explained, in the current research, water quality of Aras River in the qualitative grouping of the stations was mainly influenced by human factors

Table 2. Average of the measured parameters in the clustering resulted from cluster analysis.

Cluster	1	2	3	4	Total
BOD	6.43	3.81	3.73	5.10	4.33
COD	23.25	17.41	16.88	25.34	19.00
pH	8.23	8.34	7.90	8.12	8.16
TDS	458.6	527.4	882.4	1038.9	682.5
DO	8.56	8.48	7.79	7.70	8.20
EC	897.8	1049.3	1334.7	1537.6	1166.9
T.C.F	968.5	951.5	110.31	152.6	604.5
SO_4^{2-}	66.42	66.10	46.94	68.20	60.32
NO_3^-	5.23	6.44	3.13	3.35	4.88
Turb.	23	23	22.64	25.87	23.19
As	13	13	13.04	12.89	13
Fe	0.87	1.13	0.98	0.88	1.01
Hg	5.28	5.16	4.90	4.41	5.02
Ni	0.04	0.09	0.00	0.00	0.05

3.2. Principal component analysis and factor analysis

In the factor analysis method, the first component explains the most variation in the data, while subsequent components are ranked in descending order (Yağanoğlu *et al.,* 2020). Therefore, in cases where a number of stations and parameters is high, this method can justify the existing difference between sources of variation (e.g., stations herein) through a small number of principal components with significant variations. Therefore, important principal components can be used as a criterion to represent different aspects of the data (Zamani-Ahmadmahmoodi *et al.,* 2019). Therefore, when a limited number of primary components justify over 70% of total variations, it is possible to ascertain the identity of each component based on the eigenvectors' coefficients of each PC (for each parameter), so that it can be determined which parameters had the greatest impact on the process of changes between sources of change in the stations. Of course, a number, equal to 70 %, is variable for different designs with different sensitivities (Fataei and Shiralipoor, 2011).

The results of the analysis into principal components, including values of eigenvectors, eigenvalues, and relative and cumulative variance of the principal components of the examined parameters are presented in Table 3. The results clearly indicate that in PC_2 and PC_1, coefficients of pH, DO, EC, T. Coli, NO_3^-, Hg, BOD, COD, turbidity, and are almost high indicating the greater impact of these parameters in creating differences between stations compared to other parameters, mainly originating from pollution resulting from human activities. However, the results of the principal component analysis in a study conducted by Ravanbakhsh *et al.* (2019) on assessing changes in river water quality showed that the main responsible factors for changes in water quality at the studied stations are total dissolved solids (TDS), electrical conductivity (EC), chlorides (Cl^-), sodium (Na), sodium adsorption ratio (SAR), and Na parameters, which are mainly related to natural sources (Ravanbakhsh *et al.*, 2019). Therefore, it becomes clear that formation of analysis factors into principal components is carried out through the influence of several important parameters that can have either human or natural origins and can be used as a precise method in determining the most effective parameters on surface water quality and identifying the sources of contamination.

Table 3. Values of eigenvectors for principal component analysis and rotation matrix of factor analysis along with their eigenvalues and percentage of their variance explanation.

Variable	PC_1	PC_2
BOD	0.072	-0.459
COD	0.007	-0.527
pH	0.344	-0.117
TDS	-0.371	-0.085
DO	0.356	0.033
EC	-0.297	-0.076
T.C.F	0.393	0.007
SO_4^{2-}	0.263	-0.362
NO_3^-	0.365	0.045
Turbidity	-0.052	-0.382
As	-0.026	0.357
Fe	0.086	0.222
Hg	0.292	0.078
Ni	0.268	0.149
Eigenvalues	5.94	2.72
Variance percentage	0.42	0.19
Cumulative percentage	0.42	0.62

Table 3 shows that most parameters appeared in PC_1 with approximately equal coefficients, indicating that the first component has played the most significant role in creating variance between stations. Analysis of the principal components of the water parameters based on a 16×16 correlation matrix derived from the average data of two-year sampling periods for 19 stations revealed that the first two components justify 86% of changes between the stations, in such a way that the first and the second principal components justified 5.94 % and 2.72 % of qualitative variations between stations respectively. Other components partly contributed to justifying the water quality changes among the stations. Comparison of coefficients for either parameter (eigenvectors) for components 1 and 2 indicates that the first component has the most significant contribution to model variations, implying susceptibility and exist of a substantial difference between the stations due to pH, DO, EC, T. Coli., NO_3^-, and Hg parameters and in terms of the component 2, BOD, COD, turbidity and As played the most significant role in creating qualitative differences among the studied stations. The concept of high susceptibility refers to significant differences between values at different stations, indicating variations in pollutant sources at each station. Results show that based on the nature of parameters, water quality of the studied stations has been influenced by various human activities. As mentioned, Principal Component Analysis (PCA) for determination of the main parameters of river water quality has been used in many studies. In a study conducted by Ildorom, Hassanzadeh and Hedayatzadeh (2021), it was reported that the first three factors accounted for the most significant variations in the river water quality, with values of 53.36, 16.15, and 8.14, respectively (Ildorom, Hassanzadeh and Hedayatzadeh, 2021). The most important water quality parameters identified through these three factors including TDS, EC, HCO_3^-, and pH. Bab al-Hammami and Gholami Sefidkouhi (2018) also demonstrated that 80 % of changes in water quality were explained by the first three factors, with the first, second, and third factors accounting for 6.1%, 12.59%, and 8.3% of the variations, respectively (Bab al-Hammami and Gholami Sefidkouhi, 2018). They indicated that changes in water quality of the Talar River basin were primarily due to parameters such as TDS, EC, Cl, TH, SAR, K, Na, Mg, Ca, and SO_4^{2-}, which were influenced by agricultural runoff entering the river. The results obtained from a study conducted by Yağanoğlu *et al.* (2020) in assessment of variations in surface water quality of the Filyos River in Turkey by using the multivariate statistical method of Principal Component Analysis revealed that the first four components explained 69.49% of total variations (Yağanoğlu *et al.*, 2020). They demonstrated that the main parameters and showed that the main parameters of qualitative changes in water among the stations include temperature, EC, DO, and pH, all of which are affected by human activities.

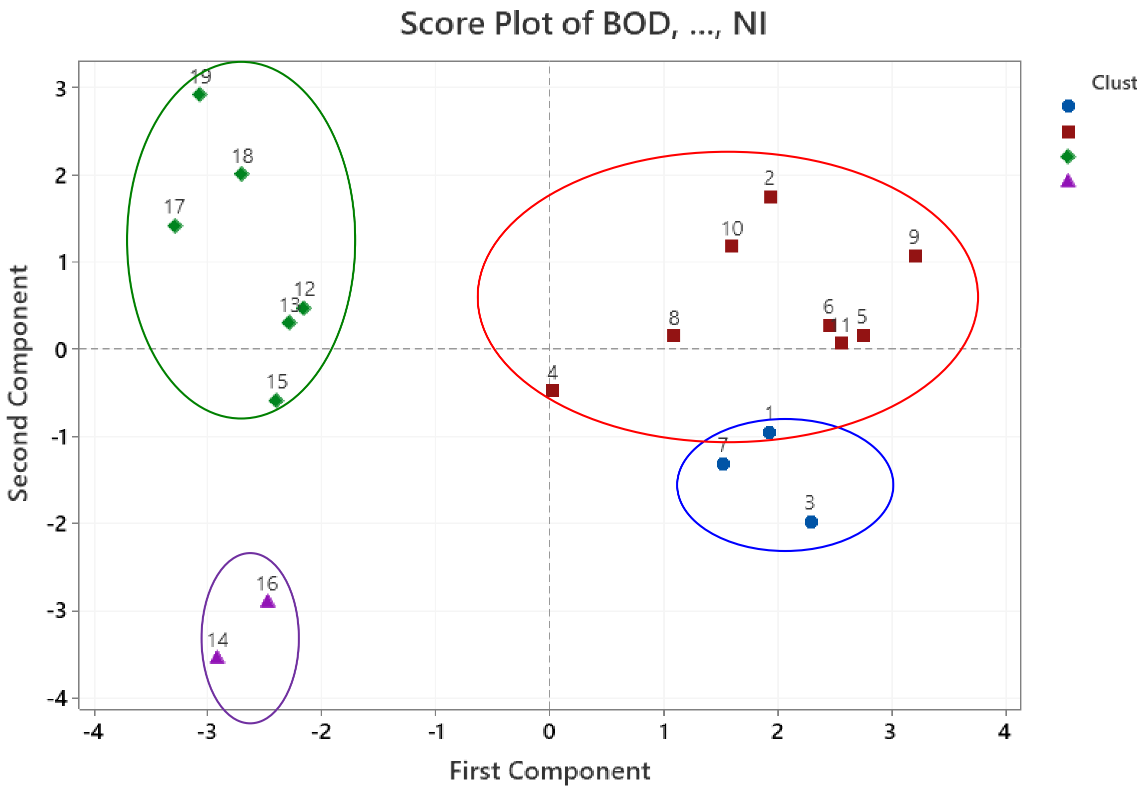


Fig. 3. Distribution of investigated stations (Biplot) based on the first and second components of the analysis of the principal components in the Aras River watershed basin.

Fig. 3 illustrates the grouping of the sampling stations using principal component analysis based on PC_1 and PC_2 components in connection with the studied parameters. As observed, this grouping, similar to the cluster analysis results discussed earlier, has classified the stations into four pollution categories based on two components PC_1 and PC_2. Therefore, the results obtained from the principal component analysis are compatible with the cluster analysis results and confirm one other, in such a way that these results are in line with the findings of Poyraz and Taspinar (2014) on evaluation of the drinking water quality in the industrial region of Marmara, Turkey, as well as the study by Giao (2022) on assessment of the surface water quality in the Ca Mau Peninsula, Vietnam (Poyraz and Taspinar, 2014; Giao, 2022).

The results of this research provide valuable information for interpreting a large set of data measured over two years on a seasonal basis regarding the assessment of water quality, identification of pollution sources, and understanding the spatial changes and trends in water quality of the Aras River. This information can be directly used in management plans aimed at improving water quality of aforementioned river. The findings indicate that pollutant sources are primarily of human origin, stemming from the discharge of industrial and mining wastewater, inadequately treated sewage effluents, and runoff from extensive agricultural land within the watershed. Given that the Aras River basin spans several countries, there is a critical need for coordinated and integrated watershed management planning to control the inflow of pollutants into this river, which serves as an important source of water for agricultural, industrial, and drinking purposes in these countries.

4. Conclusions

In the present study, multivariable cluster analysis and principal component analysis methods were used to determine the quality status and parameters affecting water quality at sampling stations on the Aras River. The results showed that water quality at different stations varied mainly due to the input of human pollutant sources. The origin of pollution was determined through the parameters affecting the quality of each station, in such a way that 19 studied stations were classified into four quality groups based on the parameters' effect on the river's water quality, and the changes in quality were explained by 86% of the variations through the first two components of factor analysis. Therefore, the grouping resulting from cluster analysis was confirmed through the results of principal component analysis. The parameter weighting, using PCA indicates that the highest and the lowest weights were assigned to parameters of BOD and Fe, respectively, with values of 0.527 and 0.222. This reflects that water quality of the stations is most affected by entry of organic matter through the sewage entry. Therefore, considering the issues related to the operation of very old and malfunctioning treatment plants in Azerbaijan and Armenia, along with influx of untreated raw sewage from some cities along the Aras River in Iran, there is a pressing need to renovate the existing treatment plants and establish sewage treatment facilities for cities lacking such an infrastructure in the Aras River watershed. According to the obtained results, it was determined that the qualitative changes in different stations are due to entry of different pollutant sources at these stations, including agricultural runoff and effluents from large fish farming ponds, erosion resulting in excessive sedimentation in some tributaries of the watershed basin, as well as the entry of effluents and wastewater from industrial and mining activities in the watershed basin of the Aras river through various countries including Turkey, Iran, Armenia, and Azerbaijan. Therefore, based on the results, there is a need for comprehensive management under the supervision of international environmental organizations and with the cooperation of all countries located in the watershed to maintain water quality of this important border and international river.

Author Contributions

Mohammad Mostafa Shabrang: Conceptualization, funding acquisition, investigation, visualization and writing.
Ali Akbar Imani: Formal analysis and software.
Ebrahim Fataei: Methodology, project administration and supervision.
Mohammad Shabani: Resources.
Hooman Bahmanpour: Validation.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgement

The authors would like to express their gratitude to Islamic Azad University, Ardabil Branch, Ardabil, Iran, for the financial support of this project.

Data Availability Statement

The analysis dataset of this study will be available on request.

References

Babolhakami, A. and Gholami Sefidkouhi, M.A. (2019) 'Analyze of Talar River water quality using multivariate techniques', Journal of Watershed Management Research, 9(18), pp. 250-259. doi: https://doi.org/10.29252/jwmr.9.18.250

Aghayari Samian, R. *et al.* (2023) 'The effect of climate change on surface runoff fluctuations in the Aras River basin', *Hydrogeomorphology*, 9(33), pp. 85-61. doi: https://doi.org/10.22034/hyd.2022.51870.1643

Abbasi, B., Maleki, R. Pirkharati H. (2017) 'Study effects of mining and gold extraction on amount of water contamination to As and Hg in Zarshoran area of Takab', *Journal of Environmental Geology* (in Persian) (2017), pp. 49:39-48. Available at: https://sanad.iau.ir/fa/Article/931362?FullText=FullText (Accessed date: 17 January 2024).

Aydin, H. *et al.* (2020) 'Assessment of water quality of streams in northeast Turkey by water quality index and multiple statistical methods', *Environmental Forensics*, *22* (1–2), pp. 270–287. doi: https://doi.org/10.1080/15275922.2020.1836074

Belkhiri, L. and Narany, T.S. (2015) 'Using multivariate statistical analysis, geostatistical techniques and structural equation modeling to identify spatial variability of groundwater quality', *Water Resource Management,* 29, pp. 2073–2089. doi: https://doi.org/10.1007/s11269-015-0929-7

Behbahaninia, A. and Farahani, M. (2016) 'Investigation of natural sources contamination with Arsenic in the Suburbs of Hashtrood city', East Azerbayjan Provincem, *Journal of Environmental Science and Technology*, 2, pp. 469-475. Available at: https://sanad.iau.ir/en/Article/839954 (Accessed date: 15 February 2024).

Cheraghi, Z. *et al.* (2019) 'Evaluation and statistical analysis of groundwater quality parameters and index for drinking in khorramabad region', *Water Engineering*, 11(39), pp. 1-14. Available at: https://sid.ir/paper/169398/en (Accessed date: 18 January 2024).

De Andrade Costa, D. *et al.* (2020) 'Water quality assessment based on multivariate statistics and water quality index of a strategic river in the Brazilian Atlantic Forest', *Scientific Reports*, 10, p. 22038. doi: https://doi.org/10.1038/s41598-020-78563-0

Ebadati, N. (2017) 'Statistical analysis of Dez River water quality, Southwest of Iran', *Anthropogenic Pollution*', 1 (1), pp. 46-60. doi: https://doi.org/10.22034/apj.2017.1.1.4660

Fataei, E. and Shiralipoor, S. (2011) 'Evaluation of surface water quality using cluster analysis: a case study', *World Journal of Fish and Marine Sciences*, 3(5), pp. 366-370. doi: https://doi.org/10.5829/idosi.wjfms.2012.04.01.6193

Ghassemi Dehnavi, A. *et al.* (2016) 'Qualitative and quantitative evaluation of surface water using statistical analysis in Azna River, Lorestan', *Environment and Water Engineering*, 2(4), pp. 306-321. Available at: https://www.jewe.ir/article_40977.html?lang=en (Accessed date: 20 January 2024).

Giao, N. T. (2022) 'Analysis of surface water quality using multivariate statistical approaches: A case study in Ca Mau Peninsula, Vietnam', *Pollution*, 8(2), pp. 463-477. doi: https://doi.org/10.22059/poll.2021.329252.1165

Govender, P. and Sivakumar, V. (2020) 'Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)', *Atmospheric Pollution Research,* 11, pp. 40–56. doi: https://doi.org/10.1016/j.apr.2019.09.009

Hajigholizadeh, M. and Melesse, A.M. (2017) 'Assortment and spatiotemporal analysis of surface water quality using cluster and discriminant analyses', *CATENA*, 151, pp. 247-258. doi: https://doi.org/10.1016/j.catena.2016.12.018

Ildorom, A., Hassanzadeh, N. and Hedayatzadeh, F. (2022) 'Evaluation of spatial and temporal variation of water quality parameters of Karkheh River and the suitability of its quality for irrigation purposes', Journal of Wetland Ecobiology, 13(50), pp. 33-54. Available at: https://sid.ir/paper/1059494/en (Accessed date: 24 January 2024).

Jalili, S. (2020) 'Water quality assessment based on HFBI & BMWP index in Karoon River, Khuzestan Province, (Northwest of Persian

Gulf)', *Anthropogenic Pollution*, 4(1), pp. 36-49. doi: https://doi.org/10.22034/AP.2020.1877482.1047

Jalilzadeh Yengejeh, R., Morshedi, J. and Yazdizadeh, R. (2014) 'The study and zoning of dissolved oxygen (DO) and biochemical oxygen demand (BOD) of Dez River by GIS software', *Journal of Applied Research in Water and Wastewater*, 1(1), pp. 23-27. Available at: https://arww.razi.ac.ir/article_47.html (Accessed date: 26 January 2024).

Kazemi Noredinvand, B., Takdastan, A. and Jalilzadeh Yengejeh, R. (2015) 'Removal of organic matter from drinking water by single and dual media filtration: a comparative pilot study', *Desalination and Water Treatment*, 57(44), pp. 20792–20799. doi: https://doi.org/10.1080/19443994.2015.1110718

Khoshnoodmotlagh, R. *et al.* (2020) 'Transboundary basins need more attention: anthropogenic impacts on land cover changes in Aras River Basin, monitoring and prediction', *Remote Sensing* 12(20), pp. 3329-3335. doi: https://doi.org/10.3390/rs12203329

Ling, T. *et al.* (2017) 'Application of multivariate statistical analysis in evaluation of surface river water quality of a Tropical River', *Journal of Chemistry*', 2017, p. 5737452. doi: https://doi.org/10.1155/2017/5737452

Lausch, A. and Herzog, F. (2002) 'Applicability of landscape metrics for the monitoring of landscape change: issues of scale, resolution, and interpretability', Ecological Indicators, 2(1-2), pp. 3-15. https://doi.org/10.1016/S1470-160X(02)00053-5

Mackialeagha, M., Salarian, M.B. and 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), pp. 1-9. doi: https://doi.org/10.22034/ap.2022.1965587.1134

Mohammadi, J. *et al.* (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*, https://doi.org/10.57647/j.jap.2023.0702.19

Mohammadi Ghaleni, M. and Kardan Moghaddam, H. (2022) 'Introducing a new drinking water quality index for surface water resources using multivariate analysis (Case study: Sefidroud River)', *Water and Soil*, 36(4), pp. 439-458. doi: https://doi.org/10.22067/jsw.2022.77413.1177

Mostafavi, A. 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*, 4(1), pp. 17-28. doi: https://oiccpress.com/ap/article/view/3402

Muangthong, S. and Shrestha S. (2015) 'Assessment of surface water quality using multivariate statistical techniques: Aase study of the Nampong River and Songkhram River, Thailand', *Environmental Monitoring Assessment*, 187(9), p. 548. doi: https://doi.org/10.1007/s10661-015-4774-1

Muniz, D. H. F. and Oliveira-Filho, E.C. (2023) 'Multivariate statistical analysis for water quality assessment: A review of research published between 2001 and 2020', *Hydrology*, *10*(10), p. 196. https://doi.org/10.3390/hydrology10100196

Neissi, L. and Tishehzan, P. (2018) 'Dez River water quality assessment by using multivariate statistical methods', *Irrigation and Water Engineering*, 9(1), pp. 139-150. Available at: https://www.waterjournal.ir/article_81860.html?lang=en (Accessed date: 5 January 2024).

Nirmalakhandan N. (2002) *Modeling tools for environmental engineers and scientists*', New York: CRC Press LLC.

Noshadi, M. and Ghafourian A. (2016) 'Groundwater quality analysis using multivariate statistical techniques (case study: Fars province, Iran)', *Environmental Monitoring Assessment*, 188(7), p. 419. doi: https://doi.org/10.1007/s10661-016-5412-2

Poyraz, B. and Taspinar, F. (2014) 'Analysis, assessment and principal component analysis of heavy metals in drinking waters of Industrialized Region of Turkey', *International Journal of Environmental Research*, 8(4), pp. 1261-1270. doi: https://doi.org/10.22059/ijer.2014.818

Rafiei, B. *et al.* (2010) 'Distribution of heavy metals around the Dashkasan Au Mine', *International Journal of Environmental Research*, 4(4), pp. 647-654. doi: https://doi.org/10.22059/ijer.2010.250

Ravanbakhsh, M. *et al.* (2019) 'Evaluation of temporal and spatial variations of water quality parameters in Zohreh River, Iran', *Avicenna Journal Environmental Health Engineering*, 6(2), pp. 75-82. doi: https://doi.org/10.34172/ajehe.2019.10

Rezaei, A. and Sayadi, M.H (2015) 'Long-term evolution of the composition of surface water from the River Gharasoo, Iran: A case study using multivariate statistical techniques', *Environmental Geochemistry and Health*, 37(2), pp. 251-61. doi: https://doi.org/10.1007/s10653-014-9643-2

Safizadeh, E. *et al.* (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), pp. 41-48. doi: https://doi.org/10.22034/ap.2021.1912491.1082

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

Saavedra, J. *et al.* (2013) 'Principal component analysis as an exploration tool for kinetic modeling of food quality: a case study of a dried apple cluster snack', *Journal of Food Engineering*, 119, pp. 229-235. doi: https://doi.org/10.1016/j.jfoodeng.2013.05.036

Sayemuzzaman M. *et al.* (2018) 'Multivariate statistical and trend analyses of surface water quality in the central Indian River Lagoon area', Florida', *Environmental Earth Science*, 77(4), pp. 1-13. doi: https://doi.org/10.1007/s12665-018-7266-0

Salahi, B., Saber, M. and Mofidi, A. (2022) 'Evaluation of evapotranspiration changes in the Aras basin using spatial high resolution MOD16 product', *Iranian Journal of Geophysics*, 16(2), pp. 119-137. doi: https://doi.org/10.30499/ijg.2022.317879.1382

Yağanoğlu E. et al. (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), pp. 207-214. https://doi.org/https://doi.org/10.33714/masteb.784959

Zamani-Ahmadmahmoodi, R. *et al.* (2019) 'Water quality assessment of the Beheshtabad River using Liou pollution index and principal component analysis', *Water and Soil*, 33(3), pp. 405-417. doi: https://doi.org/10.22067/jsw.v0i0.79102

Zhang B. *et al.* (2012) 'Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songnen plain, Northeast China', *Water Research*, 46(8), pp. 2737-2748. doi: https://doi.org/10.1016/j.watres.2012.02.033