World Journal of Fish and Marine Sciences 3 (2): 159-166, 2011
ISSN 2078-4589

Assessment of Surface Water Quality Using Principle Component Analysis and Factor Analysis

Ebrahim Fataei

Department of Environmental Engineering, Islamic Azad University, Ardabil Branch, Ardabil, Iran

Abstract: This research was carried out using multivariate statistical techniques for analyzing the quality of water and monitoring the variables affecting its quality in Gharasou river, in Ardabil province in northwest of Iran. During a year, 28 physical and chemical parameters were sampled in 11 stations. Results of these measurements were analyzed by multivariate procedures such as Principal Component Analysis (PCA), Factor Analysis (FA) and Discriminant Analysis (DA). The amount of pollutants resulting from the PCA showed that the three first components accounted for %78 percent of differences altogether, the first component accounting for 51.6, the second for 14.2 and the third for 12.2 percent of contribution, respectively. The values of coefficients calculated for the first and second components show the degree to which the places are susceptible to pollution. These values also showed that the main reason for the differences among stations' pollution level, in different periods, is the amount of parameters like NO_3^-, NH_3, Na, TDS, EC, Mg^{2+}, Turb., PO_4^{3-}, WT, COD, BOD, SO_4^{2-}, Fe, Ca^{2+} and diazinon pesticide compared to other parameters. The results of this analysis distinguished the stations as did CA. DA showed that NH_3, Ca^{2+}, Mg^{2+}, EC parameters and diazinon had the utmost importance in grouping the stations. The first two functions of DA accounted for 100% of changes. Therefore, DA allowed a reduction in the dimensionality of large data set, delineating a few indicator parameters responsible for large variation in water quality. This, in addition, confirmed that the model resulting from multi-linear regression analysis of the main component is a good indicator of each source or factor's loading in the distribution of pollution in the river. Thus, this study illustrated the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets and in water quality assessment, identification of pollution sources/factors and understanding spatial variations in water quality for effective river water quality management. This study also showed the effectiveness of these techniques for getting better information about the water quality and design of monitoring network for effective management of water resources.

Key words: *Multivariate statistical analysis* % *Water quality* % *Pollutant sources* % *Gharasou river* % *Iran*

INTRODUCTION

Surface waters are most vulnerable one of the methods in water resources assessment, environmental analyses and qualitative variables control to design qualitative monitoring management programs for rivers, is the use of multivariate statistical techniques which has been prevalent in recent years [1, 2]. Recently, the use of PCA and FA has become common in the analysis of water quality for reduction of the number of variables and better interpretation of the findings [3]. Some studies, too, have been carried out using principal components analysis (PCA), factor analysis (FA), cluster analysis (CA) and discriminate analysis (DA) methods to control qualitative variables and monitoring of sampling stations. As examples we can refer to determining the quality of surface waters in Turkey and also to the assessment of temporal and spatial fluctuations in the quality of water in Gomti river in India, Daliao river in China and Fuji river in Japan [4-6]. The quality of water in Daliao river in China was assessed by Zhang *et al*. [2], using multivariate techniques. They divided the stations under surveillance into three groups based the sources of pollution. The studies by Singh *et al*. [4] in India's Gomti river and Shrestha and Kazama's [6] studies in Japan's Fuji river, to assess the quality of water, confirmed the efficacy of multivariate statistical methods in controlling qualitative variables and monitoring of sampling stations.

Corresponding Author: Ebrahim Fataei, Department of Environmental Engineering, Islamic Azad University, Ardabil Branch, Ardabil, Iran, E-mail: ebfataei@Gmail.com.

The results of studies by Boyacioglu [5] in assessing the sources of pollution in the Tahtali river in Turkey showed that multivariate statistical techniques can reduce the number of parameters and monitoring stations and can also be used in determining the quality of surface waters.

This study was carried out to determine the quality of water, sources of pollution and effective parameters in the Gharasou river through using multivariate statistical techniques.

MATERIALS AND METHODS

Study Area: The Gharasou river is one of the main branches of the Aras river in the west side of Caspian sea, in Ardabil Province, northwest of Iran. This river originates from altitudes of Sabalan and Baghro mountains and after joining other streams in Ardabil plain, exits the plain in Samian Hydrometric station. This river has three hydrologic units and is a prennial river with a length of 255 kilometers and average slope of 5.7% and is considered as one of the Aras river's sub-rivers which itself is one the rivers in Caspian sea basin [7]. In this research the hydrologic unit of Ardabil plain as big as 4003 square kilometer was studied. The average water yield of this river, calculated in long term, in Samian station is about 228 cubic meters per year [8]. The Gharasou river, which is located in the area of study, because of the construction of Yamchi dam in upstream and Sabalan dam in downstream to provide fresh water and agricultural water for the capital city of Ardabil and Meshginshar city, respectively, is of high economical and social importance. Since this river passes through three urban (Ardabil, Nir and Sarein) and some rural areas, vast farmlands and some already established or under construction manufacturing units it is quite naturally exposed to pollution. For the decrease in the river's water production on the one hand and because of the ever-increasing amount of water consumption and urban, industrial and agricultural sewage discharges on the other hand, the quality of water in the river is endangered. Since, Ardabil is an agricultural center and is in the process of development, thus constant monitoring of water quality in the river is necessary. This research was done to control qualitative variables in Gharasou river for the same reason.

Table1: Water quality parameters, units and methods of analysis

Parameters	Abbreviations	Units	Analytical methods
EC	Electrical conductivity	μ S cm^{-1}	Electrometric
DO	Dissolved oxygen	mgl^{-1}	Winkler azide method
Turb.	Turbidity	NTU	Turb. -meter
pH	pH	pH unit	pH-meter
WT	Water Temperature	°C	Mercury thermometer
NO_3^-	Nitrate nitrogen	Meg/l	Spectrophotometric
NH_3	Ammonical nitrogen	Meg/l	Spectrophotometric
PO_4^{3-}	Phospate	Meg/l	Spectrophotometric
COD	Chemical oxygen demand	mgl^{-1}	Dichromate reflex method
BOD	Biochemical oxygen demand	mgl^{-1}	Winkler azide method
TColi.	Total coliform	MPN/100ml	Multiple tube method
FColi.	Feacal coliform	MPN/100ml	Multiple tube method
TDS	Total dissolved solids	mgl^{-1}	Gravimetric
HCO_3	bicarbonate	Meg/l	Titrimetric
Cl^-	Chloride	Meg/l	Spectrophotometric
SO_4^{2-}	Sulphate	Meg/l	Spectrophotometric
Ca^{2+}	Calcium	Meg/l	Flame AAS
Mg^{2+}	Magnesium	Meg/l	Flame AAS
Na	Sodium	Meg/l	Flame photometer
TH	Total hardness	CaCo3mgl^{-1}	Titrimetric
Mn	Manganese	μgl^{-1}	ICP-OES
Fe	Iron	μgl^{-1}	ICP-OES
Al	Aluminum	μgl^{-1}	ICP-OES
Cd	Cadmium	μgl^{-1}	ICP-OES
Cu	Copper	μgl^{-1}	ICP-OES
2,4-D	2,4-dichlorophenoxy acetic acid	μgl^{-1}	Gas chromatography
Zolon	Zolon	μgl^{-1}	Gas chromatography
Diazinon	Diazinon	μgl^{-1}	Gas chromatography

Methods of Sampling, Measuring and Analyzing Parameters: In determining the locations of study, pollution sources like: agricultural areas, residential and industrial areas, geological structure of land, main and subsidiary branches of the river and the ease of access were considered. After deciding on the location of sampling stations, their latitude and altitude were determined on a map using Geographical Position System (GPS).

Eleven sampling sites were chosen. The sampling process went on during a year from September 2007 until September 2008. The analyses were carried out based on the instructions introduced by Standard Method [9]). The methods and tools used are given in Table 1.

Sampling and analysis was done upon 28 physical and chemical parameters. These parameters were monitored monthly.

To analyze the data statistical methods like: Correlation, PCA, FA and DA. All the mathematical and statistical calculations were done by Excell_{2007}, SPSS_{16} and MINITAB_{15}.

The Q-Q plot (Kolmogorov-Smirnov) statistics were used to test the goodness-of-fit of the data to log-normal distribution. The same procedure was used to find the goodness of-fit for the assessment of principal components and factorial analysis using Bartlet test.

Multivariate Statistical Methods

Principal Component Analysis/Factor Analysis: PCA is designed to transform the original variables into new, uncorrelated variables (axes), called the principal components, which are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type [10] PC provides information on the most meaningful parameters, which describes a whole data set affording data reduction with minimum loss of original information [11, 12]. The principal component (PC) can be expressed as:

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \ldots + a_{im}x_{mj} \qquad (1)$$

Where z is the component score, a is the component loading, x the measured value of variable, i is the component number, j the sample number and m the total number of variables.

FA follows PCA. The main purpose of FA is to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA. This purpose can be achieved by rotating the axis defined by PCA, according to well established rules and constructing new variables, also called varifactors (VF). PC is a linear combination of observable water quality variables, whereas VF can include unobservable, hypothetical, latent variables [11, 13]. PCA of the normalized variables was performed to extract significant PCs and to further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation (raw) generating VFs [11, 12]. As a result, a small number of factors will usually account for approximately the same amount of informations as do the much larger set of original observations. The FA can be expressed as:

$$z_{ij} = a_{f1} f_{1i} + a_{f2} f_{2i} + a_{f3} f_{3i} + \ldots + a_{fm} f_{mi} + e_{fi} \qquad (2)$$

Where z is the measured variable, a is the factor loading, f is the factor score, e the residual term accounting for errors or other source of variation, i the sample number and m the total number of factors.

Discriminant Analysis: DA is used to classify cases into categorical-dependent values, usually a dichotomy. If DA is effective for a set of data, the classification table of correct and incorrect estimates will yield a high correct percentage. In DA, multiple quantitative attributes are used to discriminate between two or more naturally occurring groups. In contrast to CA, DA provides statistical classification of samples and it is performed with prior knowledge of membership of objects to a particular group formed with prior knowledge of membership of objects to a particular group or cluster. Furthermore, DA helps in grouping samples sharing common properties. The DA technique builds up a discriminant function for each group, which operates on raw data and this technique constructs a discriminant function for each group [4, 15, 18, 19], as in the equation below:

$$f\left(G_i\right) = K_i + \sum_{j=1}^{n} w_{ij} P_{ij} \qquad (3)$$

Where i is the number of groups (G), k is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, w_j is the weight coefficient, assigned by DA to a given selected parameters (p_j). The weight coefficient maximizes the distance between the means of the criterion (dependent) variable. The classification table, also called a confusion, assignment or prediction matrix or table, is used to assess the performance of DA. This is simply a table in which the rows are the observed categories of the dependent and the columns are the predicted categories of the dependents. When prediction is perfect, all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications.

In DA, the results of classification and their accuracy in error matrixes are determined by Kappa values. The kappa coefficient (k) was calculated from the error matrix Eq. (4). Kappa indicates to what extent classification accuracy is due to true agreement of the field data and the classified data and to what extent it could have been achieved by chance [20]. Its value varies between 0 and 1 and a value of 0.5 means that there is 50% better agreement than expected by chance alone.

$$k = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r}(x_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \times X_{+i})} \qquad (4)$$

Whereas:

r = Number of the rows/columns in the error matrix
X_{ii} = Number of observations in the cell *ii* (row *i* and column *i*)
x_{i+} = Marginal totals of row *i*
x_{+I} = Marginal totals of column *i*
N = Total number of observations

Also in DA a table called Wilk's Lambada is acquired. This is a table which assesses any meaningful differences among groups all over the predicting variables after eliminating the effects of each previous discriminating function [21]. On the other hand, Eigenvalues table, resulting from DA findings, show the value of the first discriminating function and canonical correlation [22].

In this study, three groups, in three sampling regions, which were resulted from cluster analysis, were selected for spatial assessment and the numbers of analytical parameters were used to assign a measure from a monitoring site into a group (monitoring area). DA was performed on each raw data matrix using step by step mode in constructing discriminant functions to evaluate the spatial variations in river water quality in Gharasou river basin. The sites (spatial) were the grouping (dependent) variables, whereas all the measured parameters constituted the independent variables.

RESULTS AND DISCUSSION

The description of the results of sampling and analyses for 28 physical and chemical parameters are given in Table 2.

According to the K-S test, all the variables are log-normally distributed with 95% or higher confidences.

Results of Principal Components Analysis and Factor Analysis: The results of PCA including: the values of particular vectors, egnivalues, the values of relative and cumulative variances of principal components parameters are given in Table 3. The results showed that the three first components accounted for the %78 of among stations. The two first components with Eigenvalue values of 14.44 and 3.98 together accounted for %65.8 of the total qualitative differences among the stations. The first component accounted for 51.6 and the second for 14.2 percent, respectively. The other components played a less important role in the qualitative changes of water among stations.

Table 2: Descriptive statistics of water quality variables

Parameters	Mean	Standard Deviation	Range Minimum	Maximum
EC	931	656	125	2069
DO	8.68	0.86	7.22	10.23
Turb.	15.01	6.17	6.13	26.57
pH	7.85	0.16	7.66	8.09
Temp.	10.71	1.44	8.22	12.67
NO_3^-	5.86	4.57	2.69	18.07
NH_3	0.37	0.50	0.09	1.50
PO_4^{3-}	0.91	0.74	0.46	2.82
COD	12.41	2.85	8.30	18.07
BOD	2.15	1.08	0.87	4.55
TColi.	475.3	229	127.5	764.1
FColi.	192.9	132.2	34.7	372.0
TDS	652	459	87	1448
HCO_3	3.289	1.37	0.72	5.89
Cl^-	2.48	2.24	0.35	8.05
SO_4^{2-}	3.36	3.13	0.17	8.10
Ca^{2+}	2.72	1.40	0.66	5.05
Mg^{2+}	1.56	0.82	0.35	2.75
Na	4.91	4.32	0.30	13.83
TH	214.5	111	51.3	390
Mn	34.32	24.38	15	97.50
Fe	333.3	150.5	96	557.5
Al	121.9	41.5	72	196.0
Cd	6.18	5.73	0.00	16.00
Cu	10.23	11.33	0.00	28
2,4-D	1.63	0.51	0.65	2.60
Zolon	0.55	0.34	0.15	1.35
Diazinon	0.35	0.26	0.05	1.05

Table 3: Loadings of experimental variables (28) on the first two rotated principal components and Factor analysis[1]

	PCA		FA	
Variables	PC_1	PC_2	F_1	F_2
EC	***0.244***	0.109	***0.952***	0.000
DO	-0.180	0.022	-0.642	-.240
Turb.	***0.225***	0.034	***0.840***	0.175
pH	-0.005	***0.388***	0.179	***-0.752***
TW	0.193	0.164	0.573	-.104
NO_3^-	***0.200***	***-0.279***	***0.790***	***0.745***
NH_3	***0.210***	***-0.261***	***0.729***	***0.716***
PO_4^{3-}	***0.230***	***-0.215***	***0.724***	0.649
COD	***0.230***	-0.103	***0.779***	0.445
BOD	***0.236***	-0.149	***0.778***	0.533
TColi.	0.080	***0.353***	0.480	-0.576
FColi.	-0.005	***0.309***	0.146	-0.580
TDS	***0.238***	0.135	***0.944***	0.000
HCO_3	-0.123	-0.167	-0.536	0.172
Cl^-	0.049	0.019	0.203	0.000
SO_4^{2-}	***0.246***	0.058	***0.934***	0.144
Ca^{2+}	***0.254***	0.100	***0.982***	0.000
Mg^{2+}	***0.247***	0.127	***0.982***	0.000
Na	***0.217***	0.158	***0.880***	0.000
TH	***0.252***	0.117	***0.984***	0.000
Mn	-0.029	***0.335***	0.000	-0.683
Fe	***0.201***	0.043	***0.745***	0.120
Al	0.134	0.197	0.601	-0.245
Cd	0.133	***-0.248***	0.347	0.622
Cu	0.156	-0.125	0.497	0.414
2,4-D	0.142	-0.041	0.502	0.208
Zolon	0.172	0.028	0.649	0.113
Diazinon	***0.221***	0.128	***0.739***	0.467
Eigenvalue	14.440	3.980	14.440	3.950
% Total variance	51.600	14.200	51.600	14.100
Cumulative % variance	51.600	65.800	51.600	65.700

[1*]Bold and Italic values indicate strong and moderate loadings, respectively

The comparison of the parameters' coefficients (especial vectors) for the first and second components shows that the first component has a major loading on the changes of model. This shows that the substantial differences among stations is mainly due to the following parameters: WT, TDS, Turb., BOD, COD, EC, NO_3^-, PO_4^{3-}, NH_3, Ca^{2+}, Mg^{2+}, Na, SO_4^{2-}, Fe and Diazinon. In this component the parameters pH, FColi, Cl^- and Mn had a minor role in the differences among stations. Stations 10 and 11 had the highest amount of PC_1, which shows that the quality of water in these stations is affected by the waste from agricultural drainage and urban and industrial units' sewage treatment plants. Since, these stations are located in a plain with big farmlands, industrial towns, sewage treatment plants and industrial slaughter house, the grouping of these stations is affected by these sources of pollution thereby having the highest difference in the amount of first component compared to other stations. Despite the fact that these stations are located next to the industrial zone, but the amount of affecting parameters which differentiate these stations shows that pollution in these stations is not due to the waste from industrial units, especially Ardabil's second industrial town, having heavy metals. This is indication of the efficient treatment of waste water by these units. Therefore, the quality of water in these stations is mainly affected by the drainage from agricultural activities and waste from Ardabil sewage treatment plant and industrial slaughter house. So that, in dry season (summer) a lot of fish are destroyed in these stations because of the entrance of these waste waters into the river. Consequently, PC_1 can be introduced as the polluting component from agricultural and industrial activities. We can conclude, therefore, that stations having high amounts of PC_1 are mainly affected by waste from agricultural and industrial activities.

In the second component Mn, FColi, TColi and pH with high positive coefficients and NH_3, NO_3^-, Cd, PO_4^{3-} with negative coefficients showed different reactions compared to other parameters. The pollution of water in stations 1, 3, 4, 5, 6, 7, 8 and 9, which are in the other group, is mainly because of the residential polluting sources. Among these stations the best quality of water belongs to stations 1, 3 and 5, which are qualitatively similar having negative coefficients for both the first and second components. The reason for this is that these stations are located close to the source of river. As the stations' distance increases from the source of river, with the entrance of pollutants from different sources the quality of water decreases. PC_2 can be introduced as the polluting component of residential areas. So that, stations 6, 8 and 9 which are located next to the small town of Nir, Yamchi and Hakim Gheshlagi villages and Ardabil city respectively have the highest level of pollution.

Pollution in the second station, which by itself is a distinctive station, is due to the waste from warm mineral springs of Borjloo and Gaynarjeh above the station. The quality of water in this station is affected by the values of both first and second components. It is worth noting that, station 8 which is qualitatively similar to station 2, as it is represented in Figure 2, in addition to residential waste waters, is affected by the waste from Saghezchi warm mineral spring.

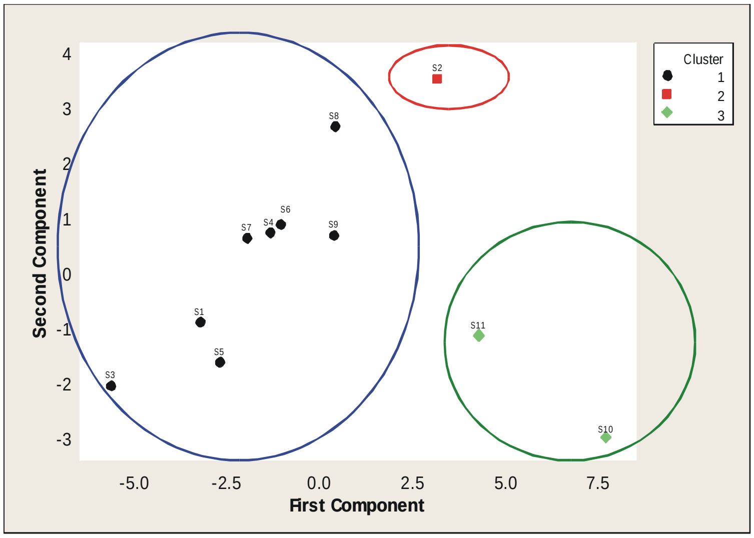

Fig. 2: Dispersion of investigated stations based on the first two principal components in the Gharasou river basin

Table 4: Eigenvalues and %Variance of Discriminant Functions

		Eigenvalue	% of Variance	Cumulative %
Function	1	683.35	92.3	92.3
	2	57.35	7.7	100.0

Table 5: Results of Wilk's Lambada test

Test of Function (s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.000	37.94	16	.002
2	.026	16.44	7	.021

Figure 3 shows the values for PC_1 and PC_2 along with the grouping of stations based on the investigated parameters. As we did in cluster analysis, these stations, based on the investigated parameters, can be classified into three groups by calculating values for PC_1 and PC_2. These pollution sources include warm mineral waters (station 2), agricultural and industrial activities (stations 10 and 11) and urban and rural wastes (other stations).

The coefficients for rotating varimax matrix (loading) in Table 3 show that the results from FA confirm the findings from PCA. In FA, too, the two first factors account for 65.7% of data variance. Therefore, it can be concluded that these two factors have been responsible for the great part of the differences between stations in this research.

Results of Discriminant Analysis: The total variance percentage and eigenivalues of the functions are given in Table 4. According to the data in Table 4, the first two discrninant functions account for the overall variance. The first factors accounts for 92.3% of variance by itself. Therefore, using the two first functions, we can determine the parameters which have the highest effect on the grouping of stations.

The results of Wilk's Lambada in DA, which is given in Table 5, indicated that the first and second rows in this test are meaningful with a probability level of 99%. This result shows the existence of a meaningful difference between groups under investigation, with regard to all parameters assessed in the area, after eliminating the effect of the first discrininant function.

It is possible to christen very discriminant function with regard to the variable with which it has the strongest relationship. A discriminate function is christened by assessing the standardized correlation for predicting variables of the two functions and correlation coefficients between predicting variables of the function within a group. With regard to the fact that Kappa coefficient value, shows the corrected accuracy of prediction, in this study the value calculated for Kappa is equal to +1, which shows the precision of prediction.

CONCLUSION

In this study for the assessment of the quality of water and to determine the sources of pollution in the Gharasou river basin in Ardabil plain, multivariate statistical procedures including PCA, FA and DA were used. The results show the effect of different pollutant factors in the environment on the quality of water. According to the findings of this research, these methods can be used, with high confidence, in the management of environmental monitoring of surface water resources.

The findings are in accordance with the findings of Boyacioglu in the Tahtali river in Turkey and Zhang *et al*., in the Daliao river in China.

Using CA method, 11 sampling stations were divided into three clusters with similar qualitative features. The results obtained from groupings, like the findings of Shrestha, et al. in the Fuji river in Japan, Singh *et al*., in the Gomti river in India and Zhang et al., in the Daliao river in China showed that the number of sampling stations and associated monitoring costs can be reduced without missing much information.

PCA and FA in determining the parameters and sources effective in the change of water quality helped dividing the river into three different areas. The first two principal components of PCA and the first two factors of FA showed that the main parameters responsible for the changes in the quality of water are heavy metals (existent in the region's soil) including Fe and Mn, natural soluble salts like Mg^{2+}, Na and Ca^{2+}, microbial pollutants such as FColi and TColi (from residential centers) and agricultural pollutants including PO_4^{3-}, BOD, COD, NO_3^-, NH_3, SO_4^{2-} and diazinon pesticide. DA indicated that the five parameters of NH_4^+, EC, Mg^{2+}, Ca^{2+} and diazinon pesticide had the highest importance in the grouping of stations. The first two function s of DA accounted for 100% of the changes. Thus, DA allowed reduction in dimensionality of large data set, delineating a few indicator parameters responsible for large qualitative changes in water. Furthermore, it confirmed that the model resulting from multi-linear regression analysis of the main component and effective parameters is a good indicator of each source or factor's loading in the distribution of pollution in the river. Bartlet test too, shows a correlation coefficient of 99% and confirms the multivariate statistical techniques used in this study. Therefore we can conclude that:

The erosion resulting from weathering, floods, warm mineral waters wastes, residential wastes, agricultural drainage and waste from industrial activities are the main factors responsible for deteriorating the quality of water in the Gharasou river.

ACKNOWLEDGMENT

I would like to thanks Dr. S.M. Monavari, Dr. A.H. Hasani, as supervisors and Dr. S.A. Mirbagheriand Dr. A.K. as Consultants in my Ph.D thesis. This study was carried out as part of Ph.D thesis in Islamic Azad University- Science and Research Branch.

REFERENCES

1. Astel, A., M. Biziuk, A. Przyjazny and J. Namiesnik, 2006. Chemometrics in monitoring spatial and temporal variations in drinking water quality. Water Res., 8: 1706-1716.
2. Zhang, Y., F. Guo and W. Meng, 2008. Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. Environmental Monitoring and Assessment. DOI 10.1007/S10661-008-0300-z.
3. Ouyang, Y., 2005. Application of principal component and factor analysis to evaluate surface water quality monitoring network. Water Res., 39: 2621-2635.
4. Singh, K.P., A. Malik, D. Mohan and S. Sinha, 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) Using multivariate statistical techniques: A case study. Analytica Chimica Acta, 538: 355-374.
5. Boyacioglu, H.U. and H. Boyacioglu, 2007. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. Evniron. Geol., 54: 275-282.
6. Shrestha, S. and F. Kazama, 2007. Assessment of surface water quality using multivariate statistical techniques: A Case study of the Fuji river Basin, Japan. Environmental Modeling and Software, 22: 464-475.
7. Ardabil Regional Water, 2004. Ardabil dam irrigation and drainage network design. Consulting engineers Band Ab., pp: 135.
8. Ardabil Regional Water, 2005. Environmental impact assessment of Sabalan dam and network. Nepta Consultative Engineering Corporation, pp: 239.
9. Standard Methods for the Examination of Water and Wastewater, 2005. 21th edn, American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA.
10. Sarbu, C. and H.F. Pop, 2005. Principal component analysis versus fuzzy principal component analysis. Acase study: the quality of Danube water, 1985-1996. Talanta, 65: 1215-1220.
11. Helena, B., R. Pardo, M. Vega, E.Barrado, J.M. Fernandez and L. Fernandez, 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by Principal component analysis. Water Res., 34: 807-816.

12. Fataei, E., S.M. Monavari, A.H. Hasani, S.A. Mirbagheri and A.R. Karbasi, 2010. Heavy Metal and Agricultural Toxics Monitoring in Garasou River in Iran for Water Quality Assessment. Asian J. Chemistry, 22(4): 2991-3000.
13. Vega, M., R. Pardo, E. Barrado and L. Deban, 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res., 32: 3581-3592.
14. Brumelis, G., L. Lampina, O. Nikodemus and G. Tabors, 2000. Use of an artificial model of monitoring data to aid interpretation of principal component analysis. Environmental Modelling and Software, 15(8): 755-763.
15. Singh, K.P., A. Malik, D. Mohan and S. Sinha, 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India) - a case study. Water Res., 38: 3980-3992.
16. Love, D., D. Hallbauer, A. Amos and R. Hranova, 2004. Factor analysis as a tool in groundwater quality management: two southern African case studies. Physics and chemistry of the Earth, 29: 1135-1143.
17. Abdul-Wahab, S.A., C.S. Bakheit and S.M. Al-Alawi, 2005. Principal component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentrations. Environmental Modelling and Software, 20(10): 1263-1271.
18. Johnson, R.A. and D.W. Wichern, 1992. Applied Multivariate Statistical Analysis. Prentice- Hall, Englewood Cliffs, NJ.
19. Wunderlin, D.A., M. Diaz, M.M.V. AME, S.F. Pesce, A.C. Hued and M. Bistoni, 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, A case study: Suquia river basin (Cordoba-Argentina). Water Res., 35: 2881-2894.
20. Lillesand, T.M., R.W. Kiefer and J.W. Chimpman, 2004. Remote sensing and image interpretation (5^{th} ed.). USA: John Wiley and Sons Inc.
21. Laurie Kelly, F.J. Bryan Manly, 2005. Multivariate Statistical Methods: A Primer. Chapman and Hall/Crc pp: 214.
22. Glasbery, C.A., 1998. Normal distribution assumptions in discrimination. Proc. of IGARSS'88 symposium, Ebinburgh, Scotland, pp: 1789-1791.