2012 2nd International Conference on Biotechnology and Environment Management
IPCBEE vol. 42 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCBEE. 2012. V42. 8

Identification of Anthropogenic Influences on Water Quality of Aras River by Multivariate Statistical Techniques

Ebrahim Fataei [+], Saeed Mosavi and Ali Akbar Imani

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

Abstract. Aras River is one of the major rivers of the Caspian Sea watershed. This river provides water needed for industry, agriculture and drinking uses in the study area. This study was done using multivariate statistical techniques to assess the effects of human activities upon the borderline Aras River which is located in northwest of Iran. 24 physical and chemical parameters were sampled in 9 stations during a year. Based on the findings from cluster analysis (CA) the stations were divided into two groups based on the amount of pollution. This 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 illustrates the usefulness of using multivariate statistical techniques in assessing water quality, determining the amount of pollutants, determining pollution sources and making relevant data available concerning the water quality, designing water quality monitoring network, and overall water quality management.

Keywords: Cluster Analysis, Water quality, Pollutant sources, Water management.

1. Introduction

Management of river quality need to study of morphological, hydraulic and biological characteristics of them. Quality of surface waters due to its importance to human health, crops and aquatic environment is very important [1].

There are several methods for monitoring surface water quality pollutants. One of the methods in water resources quality assessment is the use of multivariate statistical techniques which has been prevalent in recent years [1, 2, 3, 4, 5 and 6]. Multivariate statistical methods which are used for river quality assessment include cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA).

Some studies have been carried out using multivariate statistical techniques to control qualitative variables and monitoring of sampling stations. As examples we can refer to determining the quality of surface waters in Iran, China and Poland. Fataei et al. using CA, PCA and FA and discriminant analysis studied water quality of Gharasou river in Iran; in this study based on results of cluster analysis the river was divided into three groups of stations with high pollution, medium and low pollution [7]. Azarmdel (2010) and Nouri et al.(2007) in order to assessment of qualitative monitoring network of Gorganrood and Karoon river in Iran used PCA and FA [8,9]. Yang Li et al. used multivariate statistical techniques including CA, PCA and FA to evaluate spatial and temporal variations in Songhua river water quality in China[10]. Sojka et al. by comparing the physicochemical characteristics of water quality measurement stations introduced the important parameters in determining the temporal and spatial water quality changes at Mała Wełna River in Poland [11].

[+] Corresponding author. Tel.: + 989143549400; fax: +984517716199.
E-mail address: ebfataei@gmail.com & ebfataei@iauardabil.ac.ir

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

2. Study Area

The Aras river is one of the main branches of the Kora river and one of the main river of Caspian sea watershed in the west side of Caspian sea. Aras River which has over 1072 km in length; 451 km of it is located on the border of Iran, Azerbaijan and Armenia. Aras River originates in Hezar Berkeh in southern Turkey from Arzeroum mountains with an altitude of 3650 meters. In this research, study area was consisted entry of Aras river to Iran from turkey country in West Azarbaijan province until East Azarbaijan province in the city of Julfa.

The Aras river in study area provides uses of fresh water and agricultural water. Aras river in study area receive three sub basin river include Zangbar, Gotour Chai and Agh Chai which pass from municipal and rural residential, vast farmlands, and some already established or under construction manufacturing units. Thus constant monitoring of water quality in the Aras river is necessary. This research was done to control qualitative variables in Aras river for the same reason.

3. Materials and Methods

3.1. Methods of Sampling, Measuring and Analyzing Parameters

In determining of sampling stations was considered 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. After deciding on the location of sampling stations, their latitude and altitude were determined on a map using Geographical Position System (GPS). Nine sampling sites were chosen. The sampling process went on during a year from September 2010 until September 2011. The analyses were carried out based on the instructions introduced by EPA Standard Method. The methods and tools used are given in table 1.

Table1: Water quality parameters, units and methods of analysis

Parameters	Units	Abbreviations	Analytical methods
pH	pH unit	PH	4500- H^+-B
DO	mgl^{-1}	Dissolved oxygen	4500-O G
BOD	mgl^{-1}	Biochemical oxygen demand	2510 D
COD	mgl^{-1}	Chemical oxygen demand	5220 B
EC	µS cm-1	Electrical conductivity	2510-B
TSS	mgl^{-1}	Total suspended solids	2540-D
Turb.	NTU	Turbidity	2130-B
SO_4^{2-}	mgl^{-1}	Sulphate	4500-$SO4^{-2}$C
TP	mgl^{-1}	Total Phospate	4500-P D
Temp.	c °	Temperature	2550-B
TH	$CaCo3mgl^{-1}$	Total hardness	2340 C
Sal	%	Salinity	2520B
TColi.	MPN/100ml	Total coliform	9221-B
No_3^-	mgl^{-1}	Nitrate nitrogen	4500- No_3^- D
No_2^-	mgl^{-1}	Nitrite nitrogen	4500- No_2^- B
NH3	mgl^{-1}	Ammonical nitrogen	4500-NH3-C
Cl^-	mgl^{-1}	Chloride	4500-Cl^- D
Fe	gl^{-1}µ	Iron	ICP-OES
Ar	gl^{-1}µ	Arsenic	ICP-OES
Al	gl^{-1}µ	Aluminum	ICP-OES
Cr	gl^{-1}µ	Chromium	ICP-OES
Pb	gl^{-1}µ	Lead	ICP-OES
C6H5OH	gl^{-1}µ	phenol	Spectrophotometric
Br	gl^{-1}µ	Barium	ICP-OES

Sampling and analysis was done upon 24 physical and chemical parameters. To analyze the data statistical methods like ANOVA, correlation, cluster analysis (CA) was used. All the mathematical and statistical calculations were done by $Excell_{2007}$, $SPSS_{18}$, and $MINITAB_{15}$. The Kolmogorov-Smirnov (K-S) statistics were used to test the goodness of-fit of the data to log-normal distribution.

3.2. Multivariate Statistical Methods- Cluster Analysis

CA is a group of multivariate techniques whose primary purpose is to assemble objects based on the characteristics they possess. Cluster analysis classifies objects, so that each object is similar to the others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between clusters) heterogeneity. Hierarchical agglomerative clustering is the most common approach, which provides intuitive similarity relationships between any one sample and the entire data set, and is typically illustrated by a dendrogram (tree diagram)[12]. There are two types of cluster analysis: based on distance [13] and based on models [14].

Presently, methods based on distance are more frequently used. These methods themselves are divided into two groups: ordinal and chance models. Ordinal models are used more frequently compared to chance models. In this method, in the first stage of grouping, the number of parameters is equal to the number groups and each group includes only one parameter. In later stages the more similar groups are put together. Then these groups themselves join other similar groups. Finally, all the parameters are put in only one group [15]. There different grouping methods like: unweighted Paired Group Method Using Arithmetic Averages (UPGMA), Ward's Minimum Variance (WMV), single linkage (SL), and Complete Linkage (CL). Among these methods, UPGMA is the most commonly used one [16]. In this method the similarities and differences between parameters and related groups are equal to the similarities or differences between parameters in the group. The distance in different groups is calculated between pairs of parameters. This is while in Ward's method, grouping is done based on intra-group minimum and inter-group maximum variance [17]. In this study UPGMA method was used.

4. Results and Discussion

The description of the results of sampling and analyses for measured parameters is given in table 2. According to the K-S test, all the variables are log-normally distributed with 95% or higher confidences.

Table 2: Descriptive statistics of water quality variables

ean	Parameters	Standard Deviation	Range	
			Maximum	Maximum
PH	8.078	0.32	7.96	8.14
DO	8.221	1.36	7.94	8.39
BOD	5.33	1.55	3.17	6.5
COD	26.94	6.08	24	30.83
EC	1028	136	1010.5	1059.8
TSS	251.3	460.31	196.3	344.9
Turb	421	182.15	271	629
SO4	192.1	20.99	176.67	208
TP	1.645	1.33	1.55	1.7
Temp	10.19	1.61	10.025	10.5
Hard	449.4	77.9	406.8	490.8
Saline	0.047	0.01	0.047	0.048
Tcoliform	773	414.7	0	1100
NO3	10.67	8.08	0.3	24.5
NO2	7.335	2.54	6.88	8.086
NH3	0.14	0.07	0.073	0.226
CL	0.095	0.07	0.05	0.135
Fe	0.635	0.15	0.506	0.705
Ar	4.56	27.67	2.5	8.34
AL	83.28	26.72	77.17	93.17
Cr	44.33	12.61	28.5	55.5
Pb	1886.7	479.66	1880	1893.3
C6H3OH	518	153.03	328	740
Br	117.5	23.22	100	137.5

To classify the water quality in sampling stations and to determine the sources of pollution, CA with Ward method, using Euclidean distance based on the standardized mean of mean of the 24 measured parameters, was used. The stations were divided into two groups by dendogram cross-section, based on the farthest Euclidean distance inspired [18]. Figure 2 represents cluster analysis dendrogram based on the measured parameters.

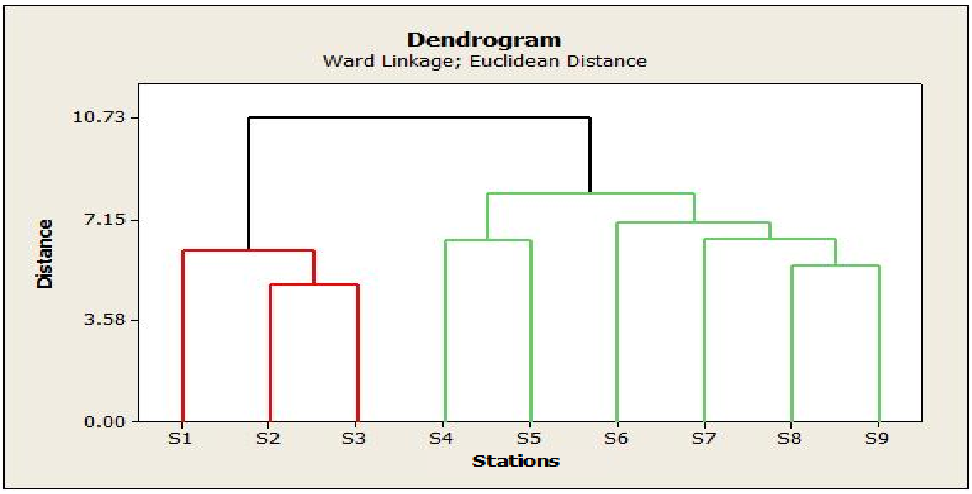

Fig. 2: Cluster analysis dendrogram of the sampling sites for surface water quality in Aras river basin in study area

The dendrogram of figure 2 shows that stations S_4,S_5,S_6,S_7,S_8 and S_9 have the highest pollution level (HP). These stations are distinguished from other stations concerning the level of pollution and have the biggest distance from other stations. The change in water quality in them was mainly due to agricultural pollutants, sewage from Maku and Poldasht city and rural residential, sewage treatment Khoy city, industrial twons of Maku and Khoy cities, and slaughter house. After that the second group with moderate pollution (MP) which is related to stations S_1,S_2 and S_3. Pollution in the second group was largely due to the waste water from rural areas, agricultural drainage and fish farm and Poldasht city landfill. Between stations of the first group(HP), station No. 7 and 6 have proper quality. This is due to locate Aras dam before these stations. Hereby, water pollution can be reduced by the reservoir. Among this group, the lowest quality is related to the Station No. 8 and 9.

The results of one-way ANOVA confirms the existence of meaningful differences among resulting groups concerning most parameters understudy with probability level of 5% and 1%. Within-group assessments showed that these parameters were not meaningfully different within groups. This is while there were meaningful differences among clusters concerning most studied parameters. Therefore, the difference between the groups shows the difference between the pollutants.

5. Conclusion

In this study to assess the quality of water and to determine the sources of pollution in the Aras river basin in West and East Azarbaijan provinces of Iran, multivariate statistical techniques including CA was used. This way, the study stations were differentiated like CA. The results show the effects of different pollutant resources 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 Fataei in the Gharasou river in Iran and Sojaka et al., in the Mala Welna River in Poland.

Using CA method, 9 sampling stations were divided into two clusters with similar qualitative features. The results obtained from groupings, like the findings of Yang Li et al. in Songhua river in China and Azardel in the Gorganrood river in Iran. Also results showed that the number of sampling stations and associated monitoring costs can be reduced without missing much information. Therefore we can conclude that:

1. The erosion resulting from weathering, floods, residential wastewater and waste from industrial activities are the main factors responsible for deteriorating the quality of water in the Aras river.

2. Using multivariate statistical techniques is useful in assessing water quality, determining the amount of pollutants, determining pollution sources and making relevant data available concerning the water quality, designing water quality monitoring network, and overall management of water quality

6. References

[1] USGS. Water quality in the Albemarle-Pamlico drainage basin, North Carolina and Virginia, 1992-95, Circular 1157. 2005.

[2] W. Z. Lu, W. j. Wang, X. K.Wang, Z. B. Xu And , A. Y. T. Leung. Using improved neural network to analyze RSP, NOx and NO2 levels in urban air in Mong Kok, Hong Kong. *Environmental monitoring and assessment.* 2003, 87: 235-254.

[3] S. A.Mingoti, & J. O. Lima. Comparing SOM neural network with fuzzy c-means, K-means and Traditional hierachical clustering algorithms. *European Journal of Operational Research*. 2006,174: 1742-1759.

[4] M. Kallionen, T. Huuhilo, R. Zbytniewski , J. Szpejna, & M. Manttari. Examination of membrane performance with multivariate methods: A case study within a pulp and paper mill filtration application. *Chemometrics and Intelligent Laboratory Systems*, 2006, 84: 744-752.

[5] R. D. Holbrook, J. H. Yen, & T. J. Grizzard. Characterizing natural organic material from the Occoquan watershed (Northern Virginia, US) using fluorescence spectroscopy and PARAFAC. *Science of the Total Environment*. 2006, 210/211: 69-109.

[6] Q., shi, X. Zhang, B. Huang, DS. Yu, I. Oborn, A. Blomback, H. J. Wang, T. F. Pagella, F. Sinclair. Surface water quality of factory-based and vegetable-based Peri-urban areas in the Uangtze River delta region, China, Catena. 2007, 69: 57-64.

[7] E. Fataei, S. M. Monavari, A.H. H asani, S.A. Mirbagheri and A.R. Karbasi. Heavy metal and agricultural toxic monitoring in Garasou River in Iran for water quality assessment using multivariate statistical methods. *Asian Journal of Chemistry*. 2010, 4: 2991-3000.

[8] H. Azarmdel, R. Mostafazadeh, and A. Ghasemi. Water resource quality stations network assessment of Gorganrood River, Iran. *Watershed Management J*. 2010, 10:57-61.

[9] R. Nouri, R. Karachiyan, A. Khodadadi Darban, and A. Shakibaeinya. Assessment of river quality monitoring stations using PCA and factor analysis (Case study: Karoon river, Iran). *Water and Wastewater Journal*.2007, 63: 60-69.

[10] Li Yang, Xu Linyu, Li Shun “Water Quality Analysis of the Songhua River Basin Using Multivariate Techniques”, *Journal of Water Resource and Protection*,2009.

[11] M. Sojka, M. Siepak, A. Zioła, M. Frankowski, S. Murat-BłaŜejewska, and J. Siepak. Application of multivariate statistical techniques to evaluation of water quality in the Mała Wełna River (Western Poland). *Journal of Environmental Monitoring Assessment.* 2007, 147:159–170.

[12] J.E. McKenna. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling & Software*. 2003, 18 (3): 205-220.

[13] I. T. Jolliffe. Principal component analysis. Springer-Verlag, 1986.

[14] S. A. Mohammadi, and B. M. Prasanna. Analysis of genetic diversity in crop plant salient statistical tools and considerations. *Crop Sci.* 2003, 43: 1235-1248.

[15] M.Vega, R. Pardo, E. Barrado, & L. Deban. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 1998, 32: 3581-3592.

[16] A. L. Panchen. Classification, evolution and the nature of biology. Cambridge Univ. press, Cambridge, England. 1992

[17] M. Otto. Multivariate methods. In: Kellner, R., Mermet, J.M., Otto, M., Widmer, H.M.(Eds.), Analytical Chemistry. Wiley-VCH, Weinheim. 1998

[18] F. Laurie Kelly and J. Bryan Manly. Multivariate Statistical Methods: A Primer . Chapman & Hall/Crc. 2005.