RESEARCH 3(11), July - September, 2017

ISSN 2394–8558
EISSN 2394–8566

Climate Change

Data analysis of metal contamination in the coastline sediments of the Persian Gulf using principal component factor analysis

Elaheh Sadatipour[1], Masoumeh Tahmasebi[1], Robabeh Vajdi[1], Mahsa Sanaei[1☼], Roohollah Noori[1]

Department of Environmental Engineering, Graduate Faculty of Environment, University of Tehran, Tehran, Iran

☼**Correspondence:**
Mahsa Sanaei, Department of Environmental Engineering, Graduate Faculty of Environment, University of Tehran, Enqelab Avenue, Qods Street, Azin Alley, P.O. Box: 14155-6135, Tehran, Iran, Tel: +98-9113956981, Fax: +982166407719; E-mail: mahsa.sanaei@ut.ac.ir

Publication History
Received: 08 May 2017
Accepted: 24 June 2017
Published: July-September 2017

Citation
Elaheh Sadatipour, Masoumeh Tahmasebi, Robabeh Vajdi, Mahsa Sanaei, Roohollah Noori. Data analysis of metal contamination in the coastline sediments of the Persian Gulf using principal component factor analysis. *Climate Change*, 2017, 3(11), 842-851

ABSTRACT

Study on heavy metal contamination of sediments and determining the monitoring stations and substations is an important step in improving the efficiency and updating the monitoring networks. In this study sampling stations located in the Persian Gulf were evaluated with principal component analysis and factor analysis techniques. For this purpose nine stations were selected and using qualitative data measured of (Pb, Cd, Cu, Ni) in the stations from 1386 to 1387, stations and substations were determined. Principal component factor analysis (PCFA) was performed using SPSS software. In this process, the correlation matrix with a VARIMAX rotation that perpendicularly of its axes is maintained in it, is used. After formation of the coefficient components matrix, it was attempted to determine the principal stations using PCFA. In this way, principal stations are the parameters which at least one of their coefficients used to form the relevant factors, have relatively high amount. In this study, because of the extent of the study area and a few parameters, this criterion was equal to 0.8. According to the criteria considered it is determined that only in the station G the value of this coefficient between each of the factors considered for this station is less than 0.8 and as a result, this station is a Substation and the other stations are principal stations.

Keywords: Principal component factor analysis, Persian Gulf, Sediment, Heavy metal pollution.

1. INTRODUCTION

Salty water pollution is a worldwide problem. Human activities on the Earth cause entering the various contaminants into marine environments (Dadolahi and Dehkordi 2013). Previously it was thought that seawater is so broad that its pollutants can be overlooked, but now the effects of water pollution on humans and other creatures directly and indirectly has been proved. Industrial wastewater, municipal solid waste, discharge of ballast water and water from washing the ship into the sea, leakage of petroleum material resulting from accidents of petroleum tankers, nuclear waste and natural events such as storms, dust, volcanoes and waste from marine life impact on marine ecosystems (Tahmasebi 2011).

Closed and semi-closed waters due to their characteristics have little power to self-decontaminate (Rezaie 2003). Persian Gulf is a semi-enclosed sea with an area of 251,300 square kilometers which on the one hand by the Strait of Hormuz has connection to the Indian Ocean and is located in the tropical and arid area and after Hudson Bay and the Gulf of Mexico is considered the third-largest Gulf in the world. Psychological and environmental situation of this region has caused that the scope of the tolerance of aquaculture in this water body due to the high pressures, is low relative to environmental changes and with entering the contaminants, be highly vulnerable (Jafari 2005; Lotfi et al. 2010).

The impacts of pollutants on the marine environment of the Persian Gulf may be significant due to its shallow depth, limited circulation, high salinity, and temperature. Discharge of industrial and urban effluents, shipping, oil spills, and recent wars in the Persian Gulf are of the major causes of its pollution (Delshab 2016)

Heavy metals in the sea are present in two form of soluble metal cations and suspended metal compounds and usually are found in the second form (suspended metal compounds) in sediments (Förstner 1979). Basically deposits as the final status of contaminants in the aquatic environment pose a significant role in the accumulation of metals in bent hoses and aquatic plants and their transfer to higher trophic levels. Heavy metals such as cadmium, lead, copper and nickel due to toxicity, high persistence in the environment, accumulating in living organisms and portability are important in the food chain, which this doubles the need to investigate these elements.

The sediments can be used as a tool for monitoring the concentration of pollutants in marine environments that with accurately determining the sediment sampling stations, data will be more reliable (Dadolahi and Dehkordi 2013). So determining the importance of the existing stations and specifying the principal stations, the stations that can express the most changes in the studied system can be effective in future decisions to optimize existing monitoring network, delete or add new stations and upgrading of the sampling frequency. For this purpose, the method of principal component factor analysis (PCFA) can be used. In recent years multivariate statistical methods is widely used in important environmental analysis (Moraleset al. 1999; Helena et al. 2000; Lu et al. 2003; Kikuchi et al. 2009; Zare Garizi et al. 2011). In the analysis of large volumes of water quality data Principal component factor analysis could be used for reducing the number of variables, data classification, and obtain the better results (Noori et al. 2007).

PCFA method for the study of the main components is one of the most valuable results of linear algebra because it is a simple and non-parametric method to extract relevant data sets from Confusing collections. PCA method transfer high-dimensional data space to down space but PFA is one of the statistical methods that can display several observed random variables by a smaller number of random variables (which are hidden in the data). Hidden random variables, called factors. This method tries to model observed random variables by linear combination of factors in addition to the amount of the error (Mohiuddin et al. 2011; Hemmatpur and Hashemi 2011).

2. THE STUDY AREA

Persian Gulf have different widths, the minimum width of it is 185 km from the cape of Nayband in Iran to Rakan vertex (Raken) in north of the Qatarand its widest width is 355 km from Jaze portin Iran to the United Arabic Emirates Selye coast. Its length is 830 kilometers from the mouth of the Arvand River to Abu Dhabi coast and its depth in the eastern part has the average of 50 to 80 meters and in western part about 10 to 30 meters. The deepest point of Persian Gulf is a hole to a depth of 93 meters, which is located at the 15 kilometers south of the Greate Tunb Island. Length of the Iranian coast in the Persian Gulf from Bandar Abbas to the mouth of the Arvand River is 1259 kilometers. Salinity of the Persian Gulf water is rather higher than the world oceans (around 37 to 50 g/l), (Mohseni 2002). Samplingarea was located in the north of the Hormuz Island.

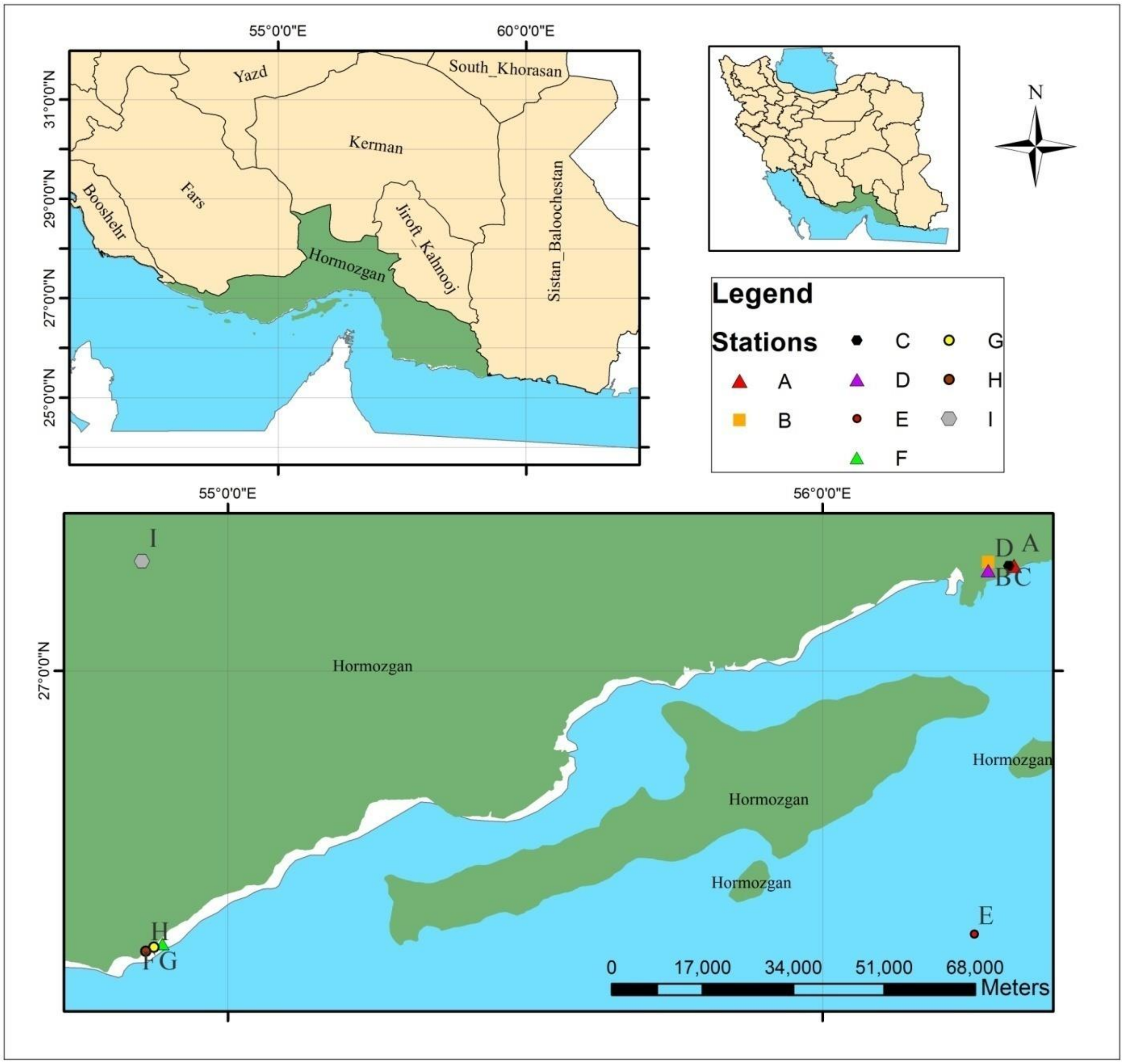

Figure 1 Location of the stations on Iran map and their situation in the Persian Gulf

Marine ecosystems, according to their size are the largest ecosystems on Earth that placed the range of creatures in it and the highest initial production is made in them. Aquatic fauna are directly affected of the marine environment and are sensitive to pollutants. In fact the Marine environments are the final destination of their surface waters region; it can also show their regional characteristics. All organisms on Earth are linked in some way to these ecosystems that in terms of species diversity, genetic diversity and diversity branch are the vast and pristine ecosystems. Part of the food chain in these ecosystems is directly used in human consumption Therefore; control of environmental factors on them has a particular importance for humans (Rezaie 2003).

According to recent reports presented in the world, unfortunately, the Persian Gulf and the Oman Sea due to the emission of pollutants from ship-oil trade and 10 million tons of pollutants and waste of war is one of the most polluted marine areas in the world. Now the infection rate is 3%, i.e., 47 times more than the international average (Aghili 2009). To investigate on the pollution of heavy metals with using modeling in this study, data was used from stations that have been selected randomly for taking the sediment and are located in the Persian Gulf coastline of the Hormozgan province and in the northern region of the Hormuz Island. Geographic Coordination of stations is provided in Table 1.

Table 1 Geographical coordination of stations

Stations	Geographical coordinate system							
A	27 °	10 ´	35 ˝	N	56°	19 ´	20 ˝	E
B	27 °	10 ´	57 ˝	N	56°	16 ´	43 ˝	E
C	27 °	10 ´	35 ˝	N	56°	18 ´	48 ˝	E
D	27 °	10 ´	7 ˝	N	56°	16 ´	43 ˝	E
E	26 °	33 ´	27 ˝	N	56°	15 ´	19 ˝	E
F	26 °	32 ´	28 ˝	N	54°	53 ´	25 ˝	E
G	26 °	32 ´	8 ˝	N	54°	52 ´	32 ˝	E
H	26 °	31 ´	44 ˝	N	54°	51 ´	42 ˝	E
I	27 °	11 ´	2 ˝	N	54°	51 ´	20 ˝	E

Location of the study area on map of Iran and the station position in the Persian Gulf area with using the ArcGIS 10.2 software is shown in Figure 1.

3. METHODOLOGY

Principal component factor analysis is a multivariate statistical method that it can be used to reduce the complexity of the analysis of the primary variable (quality monitoring stations)and In cases we are facing a large volume of information and as well as to better interpret the information (Camdevyren et al. 2005; Mohiuddin et al. 2011). With these methods, the primary components are converted to new variables independent of each other (with a correlation coefficient of zero for both components). New components are linear combinations of the primary variable (Lu et al. 2003). The primary variables used in this study, are the stations that four quality parameters with the highest importance are selected in them. With using this technique, the combinations of P primary variables, X_1, X_2, ... X_P to create an P independent component (equal to the number of primary studied variables), Z_1, Z_2, ... Z_P is created. The lack of correlation between these components is a useful feature because the lack of correlation shows that the components of the primary variables revealed the different aspects (Manly 1986). In the method (Principal component factor analysis) instead of direct use of basic variables, first we convert them into components, and then the component is used instead of the original variables. Because in the formation of components all of the variables are used, so the information of primary variables with minimal losses is presented by caused components (Mohiuddin et al. 2011).

The main components are determined with the following Equation (1):

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{ip}X_P \quad (1)$$

Z_i represents the desired component, a_{i1}represents the primary variables coefficients and X_1 is primary variables. The coefficient of the primary variables obtained from the Equation (2) (Legates and McCabe 1999).

$$|R - \lambda I| = 0 \quad (2)$$

In Equation (2) I is unit matrix, R is correlation matrix between primary variables and λ is the eigenvalues. Eigenvectors is obtained from eigenvalues. To perform this method the following steps is conducted:

1. The standardizing of the input variables: At this stage, the input data are being standardized in way, which have a mean of zero and a standard deviation of one.
2. The calculation of the correlation matrix for primary variables: The correlation matrix shows the correlation between each of primary variables which is used and is a symmetric matrix. Each element of this matrix, a_{ij} represents the correlation between variables i and j.

3.Calculate the eigenvalues λ and eigenvectors: Eigenvalues and eigenvectors for each eigenvalues are obtained by solving equation (2). Each eigenvalues and eigenvectors gives us the features of a component. In addition, each component of primary variables expressed as a percentage of the information contained and Equivalent to a part of the problem. When the first component gives us more than 95 percent of primary variables information, this means that the first component can be used instead of primary variables and in addition can retain 95% of the information and the analysis process easier. The correlation matrix is a symmetric matrix of order n and the sum of eigenvalues for this matrix is equal to n. Whatever the eigenvalues are larger, indicates that the created components from it also encompasses a greater percentage of primary variable data. Eigenvectors obtained are as the coefficients of primary variables in the formation of respective component (Noori et al. 2010).

4.Applying the proper rotation on the coefficient matrix components: Since in the formation of any component all of primary variables are used, it will be difficult to interpret components. Therefore ways can be used that cause to interpret components simpler. Some of these ways are rotating components, which are divided into two types' vertical rotation and tilting miles. Vertical rotation is used more than the rotating miles because in vertical approach, independence is maintained between the components. One of the vertical rotations used in scientific studies, called VARIMAX rotation. The results of this method are better than other methods and are recommended as a standard rotation. VARIMAX rotation method is known as PFA (Noori et al. 2012; Vega et al. 1998; Simeonov et al. 2003).

4. RESULTS AND DISCUSSION

Concentrations of cadmium, nickel, copper and lead from 9 stations is shown in the following charts during different seasons of a year in Figures 2, 3, 4 and 5.

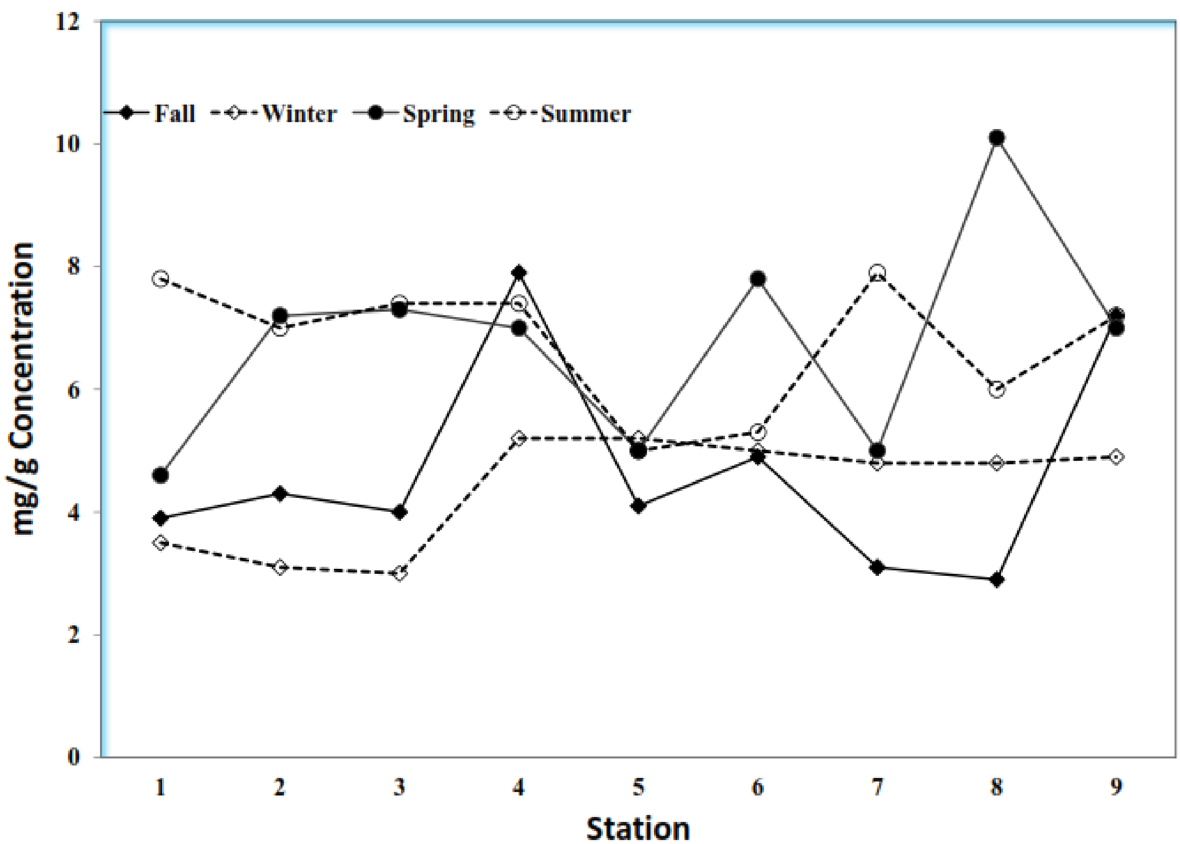

Figure 2 Concentration of cadmium from nine stations in different seasons during a year

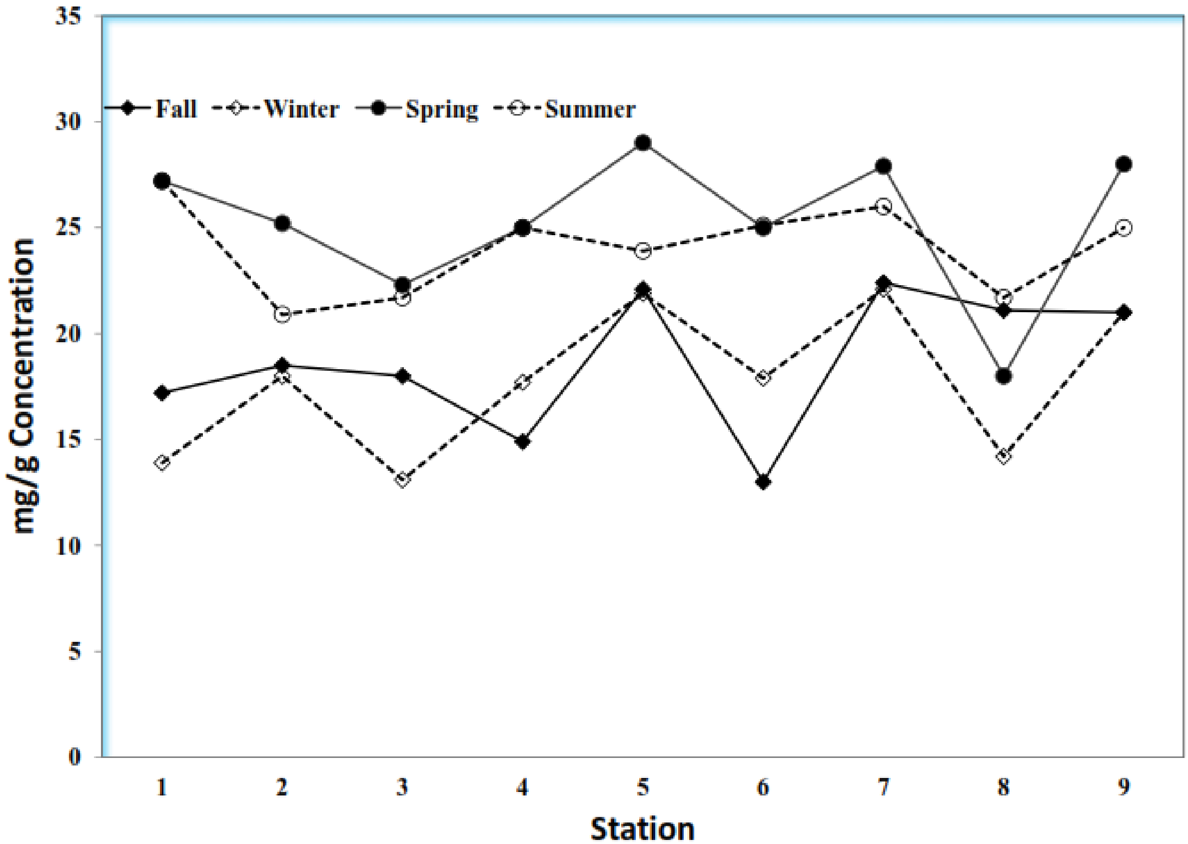

Figure 3 Concentration of lead from nine stations in different seasons during a year

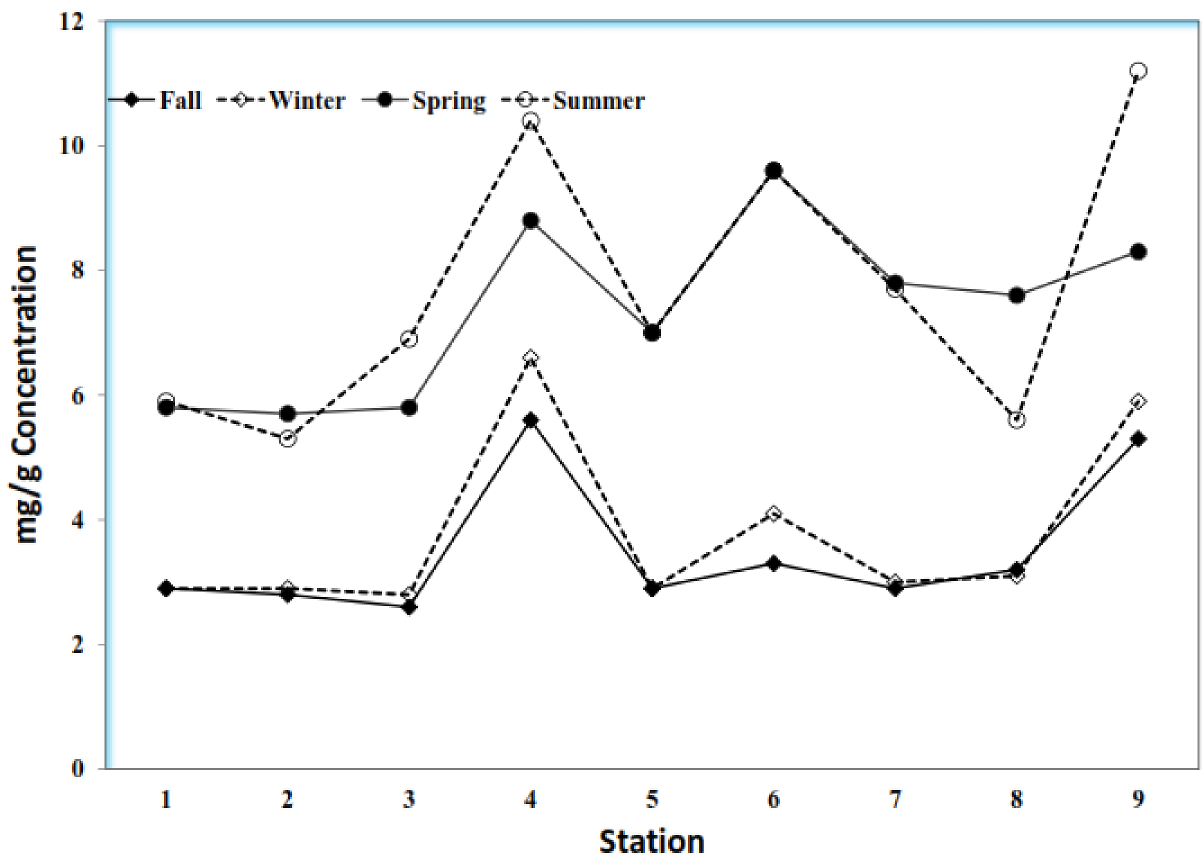

Figure 4 Concentration of copper from nine stations in different seasons during a year

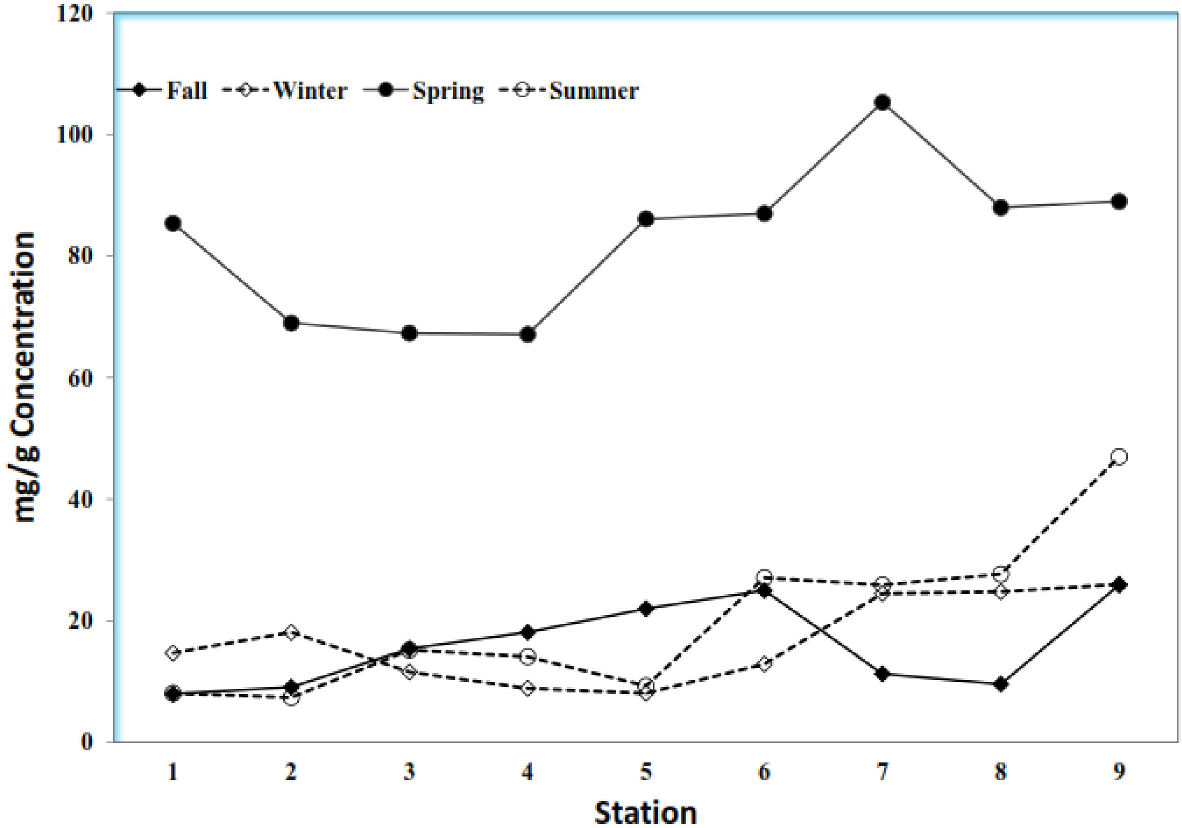

Figure 5 Concentration of nickel from nine stations in different seasons during a year

Principal component factor analysis

Sampling stations, are the primary variables (X_i, i = 1, ..., p) used in this study. As it was mentioned from each station 4 parameters of concentrations of heavy metals that were more important were measured. The results of the Kolmogorov-Smirnove tests suggest that some of the parameters which were measured at the nine stations are not normally distributed. In order to better articulate the parameters that have distributed nearly normal but have skewed distribution the median is used instead of average. So in this study, the median value is used for the parameters that were not normally distributed (Anderson and Sclove 1986). To implement these two techniques, after standardization of the parameters measured at the nine stations, the correlation matrix with the order of 9 equal to the primary variables (stations) was formed. This matrix that is a symmetric matrix is shown in Table 2.

Table 2 Correlation matrix with the order of 9

	Correlation Matrixa								
	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Station 8	Station 9
Station 1	1								
Station 2	0.994	1							
Station 3	0.972	0.991	1						
Station 4	0.987	0.997	0.992	1					
Station 5	0.992	1	0.992	0.999	1				
Station 6	0.791	0.849	0.898	0.878	0.861	1			
Station 7	0.872	0.919	0.955	0.938	0.927	0.988	1		
Station 8	0.743	0.811	0.873	0.837	0.822	0.992	0.974	1	
Station 9	0.642	0.72	0.794	0.753	0.734	0.974	0.935	0.989	1

According to Table 3 it is indicated that the numerical value of the first component is nearly 9. The second column shows that this component alone includes 91.19% of information of the primary variables and other 8 components, only express 8.8% of variation of the primary variables. Because of this in this part our discussion is summarized on the first component. The third column of the table shows the amounts listed as cumulative.

Table 3 Characteristics of components created from primary variables (stations)

	Initial Eigenvalues		
Components	Eigenvalues	% of Variance	% of Variance (Cumulative)
1	8.207	91.19	91.19
2	0.78	8.662	99.852
3	0.013	0.148	100
4	6.57E-16	7.30E-15	100
5	3.12E-16	3.47E-15	100
6	2.35E-16	2.61E-15	100
7	7.31E-17	8.13E-16	100
8	-2.01E-16	-2.24E-15	100
9	-3.39E-16	-3.77E-15	100

According to the matrix R elements (a_{ij}), which represents the correlations values between the i and j stations, all the a_{ij} elements in this matrix shows the correlation of each station with itself, that this value is equal to one. From solving the Equation (3), nine eigenvalues are achieved and for each eigenvalue, nine eigenvectors are achieved. Since the matrix R is a symmetric matrix the sum of the eigenvalues of this matrix is equal to its dimension then equivalent to 9.

Table 2 indicates the amount of eigenvectors to form each component. It is understood from the above table that to forming first component we should multiply the values of measured parameters at the station A0.94,measured parameters at the station Bin 0.90, measured parameters at the station Cin0.84, measured parameters at the station Din 0.87, measured parameters at the station Ein0.89, measured parameters at the station Fin0.53,measured parameters at the station G in 0.65,measured parameters at the station H in0.47 and measured parameters at the station Iin 0.34. Then these values for each parameter (e.g. Pb) should be separately summed together in these stations, and should be replaced instead of that parameter. According to the argumentation that has been taken, from eigenvectors obtained in Table 3, the first component is made by the following equation:

$$Z_1 = 0.94X_1 + 0.90X_2 + 0.84X_3 + 0.87X_4 + 0.89X_5 + 0.53X_6 + 0.65X_7 + 0.47X_8 + 0.34X_9 \quad (3)$$

As the equation shows, the first component does not have the same coefficients for all stations that this means that in the formation of the first component the share of all stations is not the same.

Determination of principal stations

Although in the method of principal component analysis it is revealed that one component alone shows 93.7% of the total variance in the principal variables (stations), but this method will offer no information about which stations describes dominant changes (principal stations). In this study, factor analysis was used to determine the principal stations. In the method, with using VARIMAX rotation on the coefficient component matrix, eigenvalues (multipliers of stations at each component formation) were obtained, that these information is shown in Table 4.

Table 4 Characteristics of principal components with VARIMAX rotation (eigenvectors) In the PCFA

	Components								
	1	2	3	4	5	6	7	8	9
Station 1	0.937	0.348	0.017	-2.49E-09	-1.08E-08	-6.45E-09	-9.78E-09	-1.35E-09	2.94E-09
Station 2	0.895	0.445	-0.011	1.95E-10	2.06E-09	-3.41E-09	1.32E-08	3.54E-09	-9.51E-10
Station 3	0.836	0.544	-0.072	1.08E-08	2.50E-09	1.22E-10	1.72E-09	7.25E-10	-1.97E-09
Station 4	0.871	0.49	0.038	4.39E-09	1.71E-08	-1.53E-09	9.42E-10	1.39E-09	9.21E-10
Station 5	0.886	0.464	0.007	1.13E-09	6.35E-10	1.62E-08	-5.14E-10	1.88E-09	-5.19E-12
Station 6	0.528	0.847	0.063	1.94E-08	3.33E-09	9.43E-10	3.62E-10	2.99E-09	-8.91E-10
Station 7	0.647	0.762	0.016	5.82E-09	2.67E-09	2.51E-09	2.59E-09	1.47E-08	1.45E-10
Station 8	0.465	0.884	-0.044	-5.39E-09	-4.31E-11	-2.55E-10	-1.83E-09	-3.29E-10	1.19E-08
Station 9	0.335	0.942	-0.009	-5.32E-09	3.92E-11	2.32E-10	1.79E-09	-3.41E-09	-7.66E-09

In this way, principal parameters, are the parameters which at least one of their coefficients that is used to form the relevant factors, have relatively high amount. Determination of this coefficient depends on the investigation conditions and the extent of the study area and parameters used. As the study area has a greater extent, the less amount of this coefficient should be considered. But for issues that are simple and small, usually the higher value of this coefficient is selected.

In this study, because of the extent of the study area and a few parameters, this criterion was equal to 0.8. The rest of the stations are principal stations. According to the criteria considered and Table 4 it is determined that only in the station G the value of this coefficient between each of the factors considered for this station is less than 0.8 and As a result, this station is a Substation and the other stations are principal stations.

REFERENCE

1. Aghili V, (2009) Persian Gulf: the environment, sustainable development. Journal of Human Geography 4:119-125.[In Persian]
2. Anderson TW, Sclove SL (1986) The statistical analysis of data. The ScientificPress, 2nd Ed, California, USA.
3. Camdevyren H, DemyrN, Kanik A, Keskyn S (2005)Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. Ecol Model 181:581–589.doi:10.1016/j.ecolmodel.2004.06.043
4. Dadolahi Sohrab A, Nazarizadeh DehkordiM (2013) Heavy Metals Contamination in Sediments from the North of the Strait of Hormuz. Journal of the Persian Gulf (Marine Science) 4(11):39-44.
5. Delshab H, Farshchi P, Mohammadi M, Moattar F (2016) Preliminary Assessment of Heavy Metal Contamination in Water and Wastewater from Asaluyeh Port (Persian Gulf). Iranian Journal of Science and Technology Transactions A: Science, doi:10.1007/s40995-016-0031-x
6. Förstner U (1979) Sources and sediment associations of heavy metals in polluted coastal regions. Physics and Chemistry of the Earth 11:849-866.doi:10.1016/0079-1946(79)90078-8
7. Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L (2000)Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Res 34:807–816.doi:10.1016/S0043-1354(99)00225-0
8. HemmatpourS, HashemiA(2011)Techniques to reduce the principal component analysis (PCA) and parameterized delimiters analysis (RDA) in the ranking of seismic attributes. Earth and Space Physics Journal 37: 217-227.[In Persian]
9. JafariM (2005)Oil pollution port of Khorramshahr. Master's Thesis, Khorramshahr University of Science and Technology.[In Persian]
10. Kikuchi T, Furuichi T, Hai HT, Tanaka S, (2009) Assessment of Heavy Metal Pollution in River Water of Hanoi, Vietnam Using Multivariate Analyses. Bull Environ Contam Toxicol 83(4): 575-582. doi:10.1007/s00128-009-9815-4
11. Legates DR, McCabeG J (1999) Evaluating the use of –fit measures in hydrologic and hydroclimatic model validation. Water Resour 35: 233-241. doi:10.1029/1998WR900018.
12. LotfiH, MosaviH, KhayambashiS(2010) Persian Gulf environment and it's protection. Quarterly Journal of Human Geography, 3(1):1-9.[In Persian]

13. Lu WZ, Wang WJ, Wang XK, Xu ZB, Leung AYT (2003) Using improved neural network to analyze RSP, NO_X and NO_2 levels in urban air in Mong Kok, Hong Kong. Environ Monit Assess87: 235–254. doi:10.1023/A:1024819309108
14. Manly BFJ (1986) Multivariate statistical methods. A primer. Chapman and Hall, London.
15. Mohiuddin KM, Ogawa Y, Zakir HM, OtomoK, Shikazono N (2011) Heavy metals contamination in water and sediments of an urban river in a developing country. Environ. Sci. Technol 8(4): 723-736. doi:10.1007/BF03326257
16. MohseniM(2002) Identify and quantify heavy metals nickel, cobalt, cadmium and vanadium in sediments in coastal waters West of Bandar Abbas. Master Thesis, Islamic Azad University, North Tehran Branch.[In Persian]
17. MoralesMM, Martı P, Llopis A, Campos L, Sagrado S (1999) An environmental study by factor analysis of surface seawaters in the Gulf of Valencia (Western Mediterranean). Analytica Chimica Acta 394(1):109-117.doi:10.1016/S0003-2670(99)00198-1
18. NooriR, KerachiyanR, Khodadadi DarbanA, ShakibaieniyaA(2007) Assessment of river water quality monitoring stations using principal component analysis and factor analysis Case study: Karun River. Water and Wastewater 63: 60-69 .[In Persian]
19. Noori R, Sabahi MS, Karbassi AR, Baghvand A, Taati Zadeh H (2010) Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. Desalination 260(1): 129-136.doi:10.1016/j.desal.2010.04.053
20. Noori R, Karbassi A, Khakpour A, Shahbazbegian M, Mohammadi Khalf Badam H, Vesali-Naseh M (2012) Environ Model Assess 17: 411-420. doi:10.1007/s10666-011-9302-2
21. RezaieA(2003) Determine the accumulation of heavy metals in surface sediments khor khoran.Master's Thesis, Khorramshahr University of Science and Technology.[In Persian]
22. Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A. Kouimtzis T (2003) Assessment of the surface water quality in Northern Greece. Water Res 37(17): 4119-4124. doi:10.1016/S0043-1354(03)00398-1
23. TahmasebiM(2011) The effects of heavy metals and environmental monitoring using barnacles in Bandar Imam Khomeini. Master's Thesis.[In Persian]
24. Vega M, Pardo R, Barrado E, Debán L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res32(12):3581-3592.doi:10.1016/S0043-1354(98)00138-9
25. Zare Garizi A, Sheikh V, Sadoddin A (2011) Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods. Int. J. Environ. Sci. Tech 8 (3): 581-592. doi:10.1007/BF03326244