Geopersia **9** *(2), 2019, PP. 293-304* *DOI: 10.22059/GEOPE.2019.263512.648406*

Evaluation of soil pollution sources using multivariate analysis combined with geostatistical methods in Zanjan Basin, Iran

Mehdi Boroumandi[1], Mashalah Khamehchiyan[1]*, Mohammad Reza Nikoudel[1], Mohsen Mohammadzadeh[2]
[1] *Department of Engineering Geology, Faculty of Basic Sciences, Tarbiat Modares University, Teharn, Iran*
[2] *Department of Statistics, Faculty of Basic Sciences, Tarbiat Modares University, Teharn, Iran*
**Corresponding author, e–mail: khamechm@modares.ac.ir*
(*received: 14/08/2018 ; accepted: 23/02/2019*)

Abstract

The increase of heavy metals concentration in soils is potentially threatening the environment and human health. In this paper, multivariate analysis methods such as Positive Matrix Factorization (PMF), Principal Component Analysis (PCA) and Cluster Analysis (CA) combined with geostatistical method were employed to identify the potential sources of soil pollution. A collection of 103 samples were obtained from surface soils of different types of lithology and landuse in Zanjan Basin, Iran. The concentration of As, Bi, Cd, Co, Cr, Cu, Pb, Fe, Mo, Ni, Zn, Se and Hg beside of physical and chemical properties were measured. The results showed a strong effect of anthropogenic sources on the enrichment of heavy metals especially, Zn, Pb, Cd, As and Cu in soils. From the results of PMF and PCA, the four–factor model showed the optimized solution for this study. One of the factors is related to the background concentration, another one is associated with agricultural activities and the other two are associated with industrial activities and industrial waste. The PMF method in comparison with the other common methods in multivariate analysis presents physically acceptable and more reasonable results because of non–negative condition for factors and weighting of the variables.

Keywords: *Soil Pollution, Heavy Metals, Multivariate Analysis, Positive Matrix Factorization, GIS*

Introduction

Pollution of soils by heavy metal is a growing concern, because of the potentially undesirable effects of heavy metals on the environment and human health. In recent decades, by the rapid development of urbanization and industrialization, landuse changes, the increment of wastes, agricultural activities and unsuitable wastes management, soil pollution has changed to become one of the main environmental concerns (Liang *et al*., 2017b; Liang *et al*., 2017c). Therefore, attention of researchers was drawn to the problem of soil pollution, to prevent additional environmental hazards and study applicable approaches of soil remediation (Hao *et al*., 2016; Sharma *et al*., 2016; Turekian & Wedepohl, 1961; Udayakumar *et al*., 2014; Zhang *et al*., 2016; Liu *et al*., 2016). Also in Iran, industrial and agricultural activities, landuse changes and inappropriate waste management have changed soil pollution to one of the environmental concerns.

Heavy metals of soils are naturally occurring elements of the Earth's crust; therefore their concentrations tend to remain low, and complex pedological and geological processes can explain most of the soil dissimilarities (Bhuiyan *et al*., 2010; Wang *et al*., 2015; Zhang *et al*., 2013; Zhao *et al*., 2010). In recent decades, concentrations of heavy metals in soils have exceeded natural concentrations from pedogenesis due to increase in anthropogenic inputs, even at a regional scale (Facchinelli *et al*., 2001; Sharma *et al*., 2016). Anthropogenic inputs such as industrial development, urbanization, pollution from metallurgical activities, waste disposal, influx of agricultural fertilizers and changes in land management systems have significant influences on the changes of soil properties (Zhang *et al*., 2008; Zhao *et al*., 2010; Harris *et al*., 2011; Lin *et al*., 2011; Olubunmi & Olorunsola, 2010).

Understanding each possible source of heavy metals and the degree of correctness of the apportioned outcomes is a precondition for developing pollution control strategies (Chen *et al*., 2013; Huang *et al*., 2015; Sharma *et al*., 2016; Zhang *et al*., 2008; Zhao *et al*., 2010). Application of multivariate statistical methods for identifying pollutant have been developed in recent decades, for example, the chemical mass balance (CMB) model, principal component analysis (PCA), cluster analysis (CA), positive matrix factorization (PMF) and multiple linear regression (MLR). In contrast to CMB models, PCA, CA and PMF models do not need information on the potential sources of their profiles theoretically (Chen *et al*., 2013; González–Macías *et al*., 2013). Apart from the mentioned methods, geostatistical methods can be used for estimation of polluted areas, according to the calculation of unbiased approximation of heavy metal concentrations in areas without samples (Amini *et al*., 2005; Hussain *et al*., 2015; Zhang *et al*., 2008). Also, a combination of multivariate statistics with geostatistical methods is used for source identification of soil pollution and their spatial distribution (Liang *et al*., 2016; Liang *et al*., 2017a). Moreover, if the researcher has a general knowledge of the probable pollutants, multivariate analysis methods can provide more accurate results even when exact sources information are missing. These characteristics are the key advantages of multivariate methods when compared with other receptor models.

Multivariate statistical analyses, like PCA and CA have generated acceptable results in soil pollution studies (Davis *et al*., 2009; Facchinelli *et al*., 2001; Li, 2009; Mitchell *et al*., 2016; Reimann *et al*., 2007; Turekian & Wedepohl, 1961; Zhang *et al*., 2016; Zhao *et al*., 2010; Ali & Malik, 2011). Though, its application is dependent on restrictions such as management of uncertainties and handling of noisy data (Paatero & Tapper, 1994), the possible occurrence of negative factor loadings is problematic to interpret in terms of certain positive physical variables such as masses, concentrations and other physicochemical parameters (Vaccaro *et al*., 2007). In the early 1990s, Paatero and Tapper (1994) developed PMF model. It basically controls such restrictions by using empirical uncertainties in the data matrix as well as by limiting the solutions to non–negative values (Paatero & Tapper, 1994). Similar to PCA, PMF also has the advantage of being a posteriori method, which does not depend on prior information of sources by direct measurement or from emission inventories (Vaccaro *et al*., 2007). PMF has been recently used in the examination of temporary data such as atmospheric pollution by particulate matter (Kim *et al*., 2003; Pekey *et al*., 2012; Polissar *et al*., 1998)) and study of wet deposition (Anttila *et al*., 1995). PMF has been effectively used to spatially distribute data sets to assign the sources in sediments (Chen *et al*., 2013) and soils (Comero *et al*., 2012; Vaccaro *et al*., 2007; Wang *et al*., 2009; Guan *et al*., 2018; Liang *et al*., 2017a).

In this paper, multivariate analysis including PCA, CA and especially PMF were used to determine the main pollutant of soil in Zanjan Basin, Iran. Source identification technique based on multivariate methods and support of geostatistical methods, have been used for determination of potential soil pollution sources at this regional scale. The results of this research can help in determining the potential sources of heavy metals from industrialization and agricultural development at the study area.

Materials and Methods

Study area

Zanjan Basin is in the central part of Zanjan province, northwest of Iran (Fig. 1). The area is susceptible to soil pollution because of industrial towns, cities, and villages in this basin. Zinc and copper industrial town where smelting and refining are done and other industrial towns are located in this basin.

Zanjan province has several mines and related industrial activities due to natural condition and lithological properties (Khamehchiyan *et al*., 2011). The most industrial towns of the province are active in this basin. Inappropriate waste management around industrial towns is another reason for increase in heavy metals in soil and environment.

The geology of this basin (Fig. 1) includes different lithological units that formed outcrop in the mountains in western and eastern parts of the basin. These mountains in the western parts include mainly igneous lithology comprised of andesite, tuff, granite and granodiorite. In eastern parts of the basin, there are also tuff and andesite and sedimentary rocks include dolomite and limestone. The northern part of this basin is mainly comprised of hilly topography consisting of conglomerates and marls. The central area includes terraces and recent alluvia which are mainly comprised of weathering of surrounding rocks (Stocklin & Eftekharnezhad, 1969). Most of the samples were collected from this area. This basin has different landuses. Except for industrial towns and urban areas, most of the areas of this basin include ranges and dry–farming. The southeastern part of the basin and some areas around Zanjan city has agricultural activities.

Soil sampling

Collection of 103 samples was done in the study area. Sampling design is a combination of systematic and judgmental sampling. This design helped for complete coverage of the study area and also soil pollutants were considered. The rocky outcrops excluded and in remained areas distance of samples are about five km or less. The density of samples near to industrial towns is higher than other areas, to evaluate the radius of pollution around of these states.

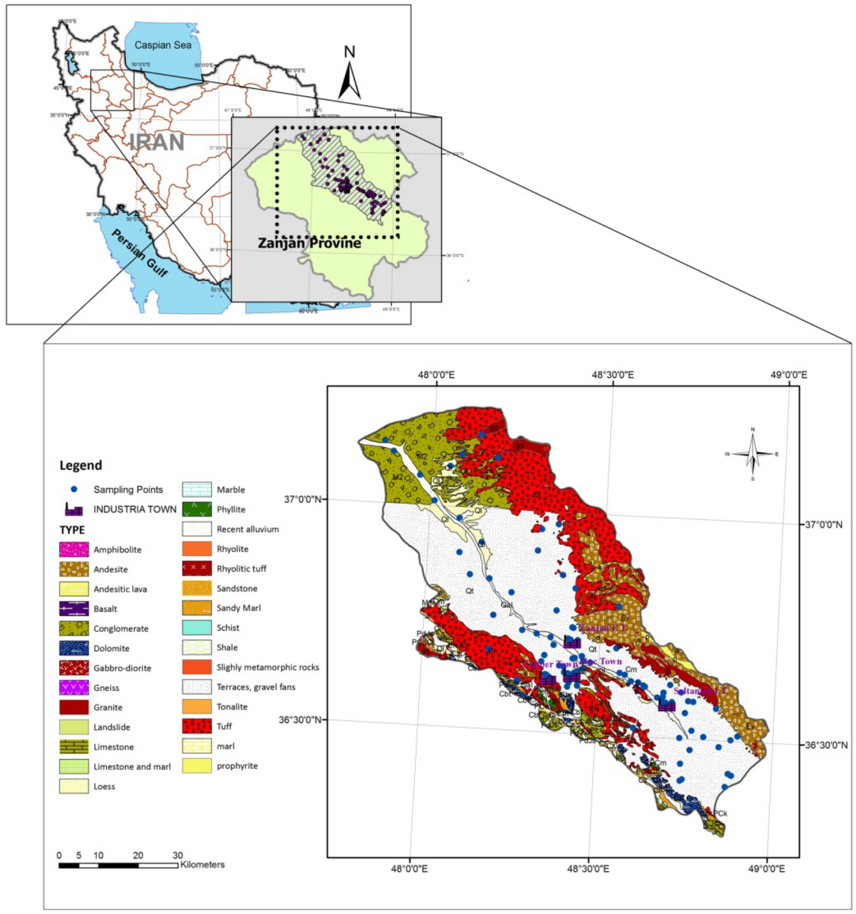

Figure 1. Location of the study area in Iran and the lithology map of the basin.

Samples were obtained from 0–30 cm depth, and they are composed of four sub–samples from the vortexes a square block with 20 m width. Sub–samples mixed to get a bulk sample. A stainless steel spatula was used for collection of samples and they kept in plastic bags. Sampling sites are illustrated in the lithological (Fig. 1) of the study area. Samples were obtained from different geological condition and also different types of landuse. Duplicate samples were obtained in some points to decrease the uncertainty of sampling. Samples were transferred to the laboratory for physical and chemical analysis.

Analysis

The soil samples were air–dried and then ground for suitable meshes (2 mm sieve) for chemical analysis. For sample preparation after weighing 0.8 g of dried sample, 5 ml 65% HNO_3 added and then placed in TOPwave Microwave Digestion (Analyticgena®) for 25 minutes with 40 bar pressure. The digested solutions were made to 50 ml with deionized water for analysis by Atomic Absorption Spectrometer (AAS) using ContrAA 700 instrument. All the soil samples were analyzed by AAS for total As, Bi, Cd, Co, Cr, Cu, Pb, Fe, Mo, Ni, Zn, Se and Hg. The levels of Bi and Cd after digestion were measured by graphite furnace of AAS. Hg and As were measured by the hybrid method of AAS and the other heavy metals were determined by the flame of AAS. Duplicate samples were measured in the same procedure to evaluate the precision and bias of analysis. The main physico–chemical parameters such as grain size, pH, cation exchange capacity (CEC), the percentage of total neutralizing value (TNV), organic matter and EC were also measured. Sieving performed based on ASTM D 422–63. Values of pH were measured in H_2O and 0.01 $CaCl_2$ solutions with a soil/solution ratio of 1:5 following Conyers and Davey, 1988. EC measured with a conductivity meter in a saturated paste of soil and water (Rhoades, 1996). Organic matter contents were measured by ignition of samples at a temperature of

450ºC (Davies, 1974). CEC of samples were measured by Bower method described by Chapman 1965. TNV measured by Bernard calcimeter method developed by Hulseman, 1966.

Positive matrix factorization

Basic Equation

PMF was developed by Paatero and Tapper in 1994, which controls the limitations of previous methods by using non–negativity constraints to find more physically understandable factors (Comero *et al*., 2012). In comparison with PCA, PMF is not sensitive to missing values (MV), below detection limit (BDL) and outlier data, because it is a weighted least–squared model and individual error estimations give optimal weights for scaling of data (Paatero & Tapper, 1994). So, in PMF, non–representative data can be handled by reducing their significance and skewed distributions could be correctly weighted rather than normalized. Also, mathematical algorithm of PMF prevents the existence of negative factor scores, which can appear from PCA analysis, permitting more actual realistic answers (Comero *et al*., 2012). Error estimation of observed values allows the user to manage outlier data by using the estimates as weighting parameter (Sofowote *et al*., 2008).

The algorithm of the PMF has been comprehensively defined (Paatero *et al*., 2002; Paatero & Tapper, 1994). The principle of PMF algorithm starts from the basic mass balance equation is described by the following equation:

$$\boldsymbol{x_{ij}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{p}} \boldsymbol{g_{ik} f_{ki}} + \boldsymbol{e_{ij}} \quad \boldsymbol{i = 1, \dots, m; j = 1, \dots, n; k = 1, \dots, p} \tag{1}$$

where xij is the ij–th elements of the matrix of input data, p is the number of determined factors, g_{ik} and f_{kj} are the elements of the factor scores and factor loadings matrices, respectively and e_{ij} are the residuals (the difference between input data and predicted values) (Comero, 2011; Paatero *et al*., 2002; Liang *et al*., 2017.a).

The PMF calculates elements based on a weighted least–squared problem which minimizes the so–called object function Q, defined in Paatero (1997) and specified by the simplified equation (Comero, 2011):

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{e_{ij}}{s_{ij}} \right)^2 \tag{3}$$

where S_{ij} is error estimates of data values. Typically, environmental data sets can contain BDL and/or missing values. To make use of their information content, suitable estimates for their values and uncertainties must be determined (Comero, 2011).

The essential input data for PMF are the measured values, X, and uncertainties of the measured values, E. Measured values and uncertainties for PMF should be positive values (González–Macías *et al*., 2013). PMF 5.0 released by U.S. Environmental Protection Agency (Norris *et al*., 2014) was used to solve the problem of this research. This model developed based on basic model.

Data pretreatment

Missing values could be replaced by the geometric mean of the values, and their associated uncertainties set at four times of this geometric mean concentration (Kim & Hopke, 2005). According to the studies of Kim & Hopke (2005), Farnham *et al*., (2002) and Polissar *et al*., (1998), the measured concentrations BDL were substituted by half of the BDL values, and their associated uncertainties were set at 5/6 of the BDL values. Fortunately, there is no missing value in these data.

The variables were categorized as strong, weak or very weak based on the signal to noise ratio (S/N) range recommended by Reff *et al*., (2007). As the ratio of S/N for Hg is very low due to BDL of many samples, it is categorized as very weak data and it was omitted. The other heavy metals have high S/N ratios and considered as a strong category.

Results and discussion

General characteristics of parameters

Descriptive statistics of the concentration of elements and physicochemical variables at sampling points are presented in Tables 1 and 2. The concentration of Se is BDL, so the results of these parameters were omitted. There is no missing value (MV) in the results. Distributions of most of the parameters are strongly positively skewed, except physical parameters, Fe, Cr, Ni, and Co. Zn, Cu and Cd are highly skewed and it is expectable because of great industrial activities related to these heavy metals.

PCA and CA

Cluster analysis (CA) was used for clustering of variables. CA was performed on variables using the Ward's method based on Euclidian distance to categorize variables, in order to find clusters which show similar behavior. Dendrogram of variables (Fig. 2) shows three main clusters were distinguished.

The first cluster includes Cd, Pb, Zn, and As related to industrial activities in this basin with their main activities related to zinc and lead. The second cluster that contains Ni, Cr, Co and Fe appears to be related to the background concentration of these metals in the study area.

Table 1. Descriptive statistics of the concentration of elements

Elements (mg/kg)	Mean	Median	Min	Max	SD	Skewness	CV (%)
Bi	0.68	0.51	0.01	5.23	0.73	4	107
Cd	0.54	0.36	0.01	4.89	0.74	4.4	137
Co	20.3	19.2	12.3	45.3	5	2.5	25
Cu	59.1	38	20.7	763.4	89.9	6	152
Pb	76.9	27.3	4.3	975	150.5	3.8	196
Ni	53.6	52.6	20	98.8	16.6	0.4	31
Zn	425.4	125.4	62.7	16260	1654.9	8.9	389
Fe	54258.6	52030	35150	81885	10831.7	0.5	20
Cr	43	41.6	22.2	80.7	11.5	0.8	27
Mo	0.3	0.3	0.1	1	0.1	2.2	37
V	62.4	55.9	30.8	184.1	22.1	2.4	35
As	12.3	10.7	1.4	43.6	6.2	2.7	50
Hg	*	*	<DL	5.8	*	*	*

* It cannot be calculated because of BDL data, DL: Detection Limit

Table 2. Descriptive statistics of physicochemical parameters

Physicochemical parameters	Mean	Median	Min	Max	SD	Skewness	CV (%)
EC (µS/cm)	745.6	411	2.2	3980	791.5	1.8	106.2
CEC (mEQ/100g)	27.1	22.6	10.3	77.9	13	2.2	47.9
OM (%)	1.9	1.5	0.3	6.8	1.4	1.5	70.6
TNV (%)	11.7	11.5	0.6	34.1	5.7	0.4	48.8
Clay (%)	7.9	7	1	22	4.5	0.9	56.3
Silt (%)	35.4	35	9.5	56	10.5	0	29.6
Sand (%)	56.8	58	26	89	13.9	-0.1	24.5
pH	8.1	8.1	7.1	8.8	0.3	-0.3	3.6

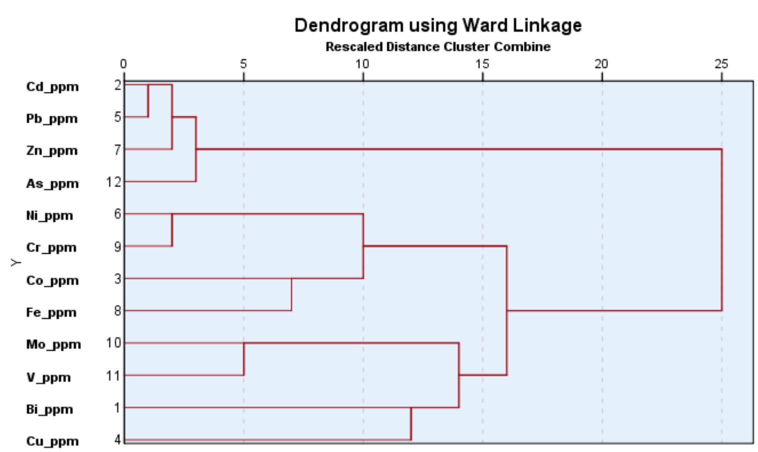

Figure 2. Dendrogram of cluster analysis of variables

The third cluster contains Mo, V, Bi, and Cu which can be related to another type of lithological background in this area.

PCA with varimax rotation was employed for heavy metal concentrations of samples. Principal components with Eigenvalue larger than 1 (Kaiser Criterion) were selected. Four PCs were extracted which explains about 74% of the total variation (Table 3). By using geostatistical methods, spatial maps of principal components scores were produced and presented in Fig. 3.

The first PC, including 31% of the total variance, was positively associated with cadmium, lead, zinc, and arsenic. PC1 could be attributed to anthropogenic activities relating to the industrial activity associated with Zn and Pb mines at the study area. As shown in Fig. 3, the high scores of PC1 were around the industrial towns; therefore, this PC is related to industrial towns located at the central part of the basin.

The PC2, explaining 18% of the total variance, is described by high–positive loadings in Co, Cr, and Ni. These elements were found to have the natural concentration in soils. The high values of this PC are distributed in the eastern part of the study area which igneous rocks outcropped. Vanadium, molybdenum, and iron showed greater cooperation in the third component (PC3). This PC can also be explained by the natural background of elements.

The PC4, explaining 10% of the total variance, was correlated mainly with Bi, however, Cu loaded negatively. For all components, there are negative values for some variables that cannot be interpreted as physical behavior.

PMF

Factor loading of PMF

Data matrix and the number of sources were inputted to PMF model. The concentration values were used for the measured data and the Formula (4) (Polissar *et al*., 2001) was used as the total uncertainty given to each measured data.

$$x_{ij}= \sqrt{a_j u_{ij}^2 + b_j DL_{ij}^2} \tag{4}$$

where u_{ij} is analytical uncertainties, DL is the method detection limit, and a and b are scaling factors, both determined by trial and error.

BDL values were substituted by half of the detection limit values, and their total uncertainties were set as five sixth of the detection limit values (Chen *et al*., 2013; Polissar *et al*., 2001).

Number of factors

To determine the number of sources, it is essential to evaluate different numbers of sources and find the best one with the most physically significant results. Furthermore, since rotational obscurity exists in factor analysis modeling, PMF was run many times with different FPEAK values to determine the range within which the objective function Q value remains relatively fixed (Kim *et al*., 2004).

Table 3. Principal component analysis of variables

Parameter	Rotated Component			
	1	2	3	4
Bi	.122	-.144	-.075	.774
Cd	.923	.069	-.035	.069
Co	.530	.615	.172	-.217
Cu	.111	-.219	-.067	-.688
Pb	.903	.065	-.029	-.055
Ni	.196	.899	-.117	.055
Zn	.862	.093	.076	-.103
Fe	-.130	.381	.620	-.127
Cr	-.133	.858	.183	.114
Mo	.292	-.135	.776	.052
V	-.175	.063	.891	.024
As	.833	-.076	-.091	.135
Eigenvalue	3.601	2.181	1.884	1.191
Variance explained (%)	30.011	18.177	15.700	9.922
Cumulative % variance	30.011	48.189	63.888	73.810

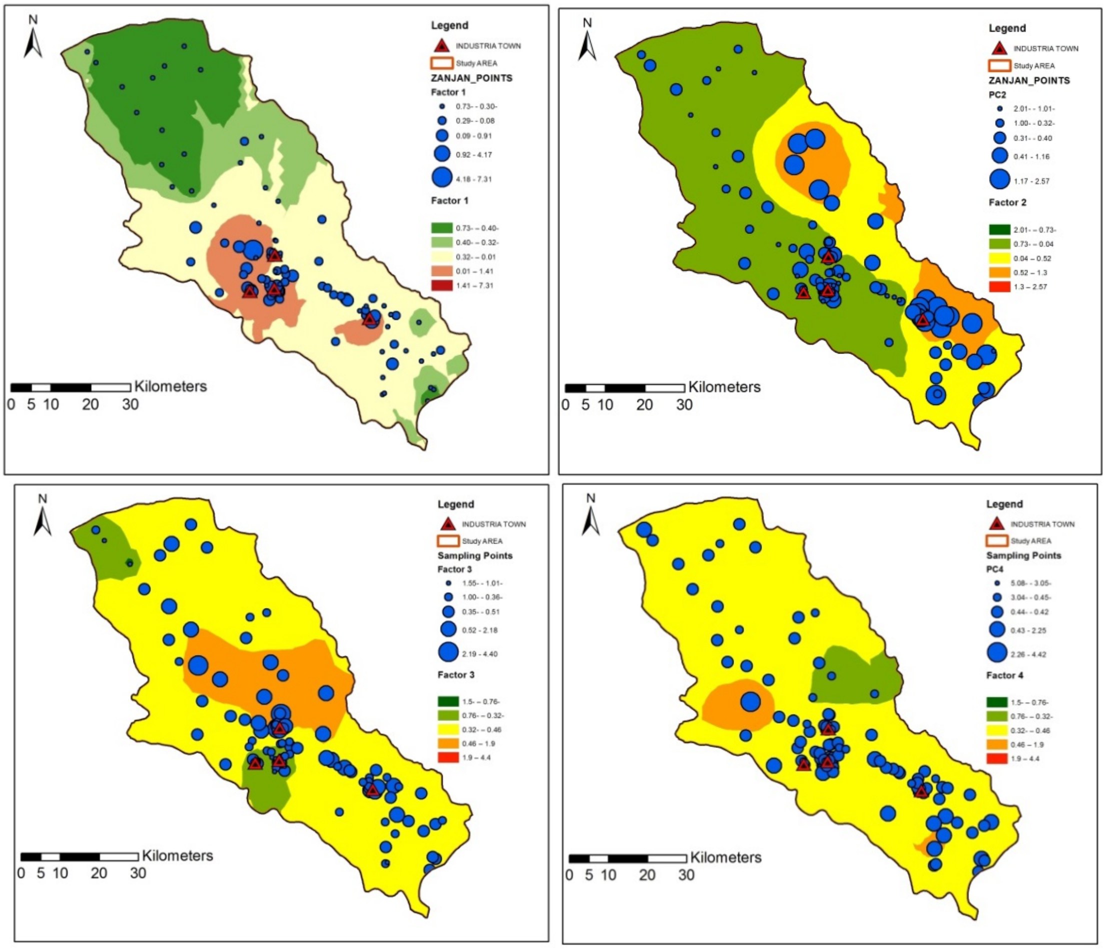

Figure 3. Spatial distribution of 4 PCs of principal component analysis

Initial runs of PMF were conducted by two to ten factors in order to specify the optimal number of factors, which determined 4 factors as the best solution for this problem.

Physical interpretation

Fig 4 shows the PMF results and displays the proportion of each variable (As, Bi, Cd, Co, Cu, Cr, Fe, Mo, Ni, Pb, V, and Zn) in factors determined with the model. The spatial distribution maps of factors were prepared by applying geo–statistical interpolation of the G factor scores (Fig. 5). These maps can be used in the interpretation of factors.

Factor 1 (Natural influence)

The first factor showed high proportions of Fe, Ni, Cu, Co, Cr, Mo, V and As. This factor could be linked with a component controlled by parent lithology. According to geological map of the study area (Fig. 1), higher scores of this factor seen around igneous rocks (Eastern part of the study area) may increase natural heavy metal concentration by weathering of parent material. So, this factor can be interpreted as natural processes and background concentration of elements in soil. Also, a strong correlation between Ni and Cr (based on Pearson correlation) at this area matches the spatial patterns of soil parent material, indicating that the sources of Cr and Ni are predominately geochemical (Zhao *et al*., 2010). This factor can compare PC2 and PC3 of PCA. High values of these factors have been distributed in east–northern part of the study area. The lower value of this factor is realized in industrial areas.

Factor 2 (Anthropogenic sources)

The second factor determined by PMF is described by Pb and As variation which could be interpreted to represent industrial activities. The main sources of Pb in the environment consist of manufacturing activities and atmospheric deposition.

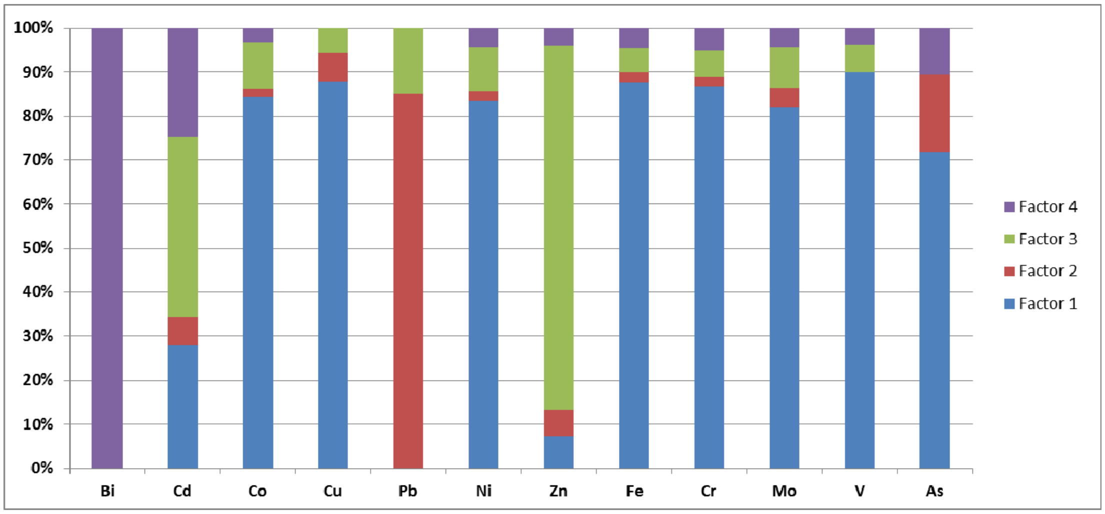

Figure 4. Factor loadings of the resolved problem by PMF

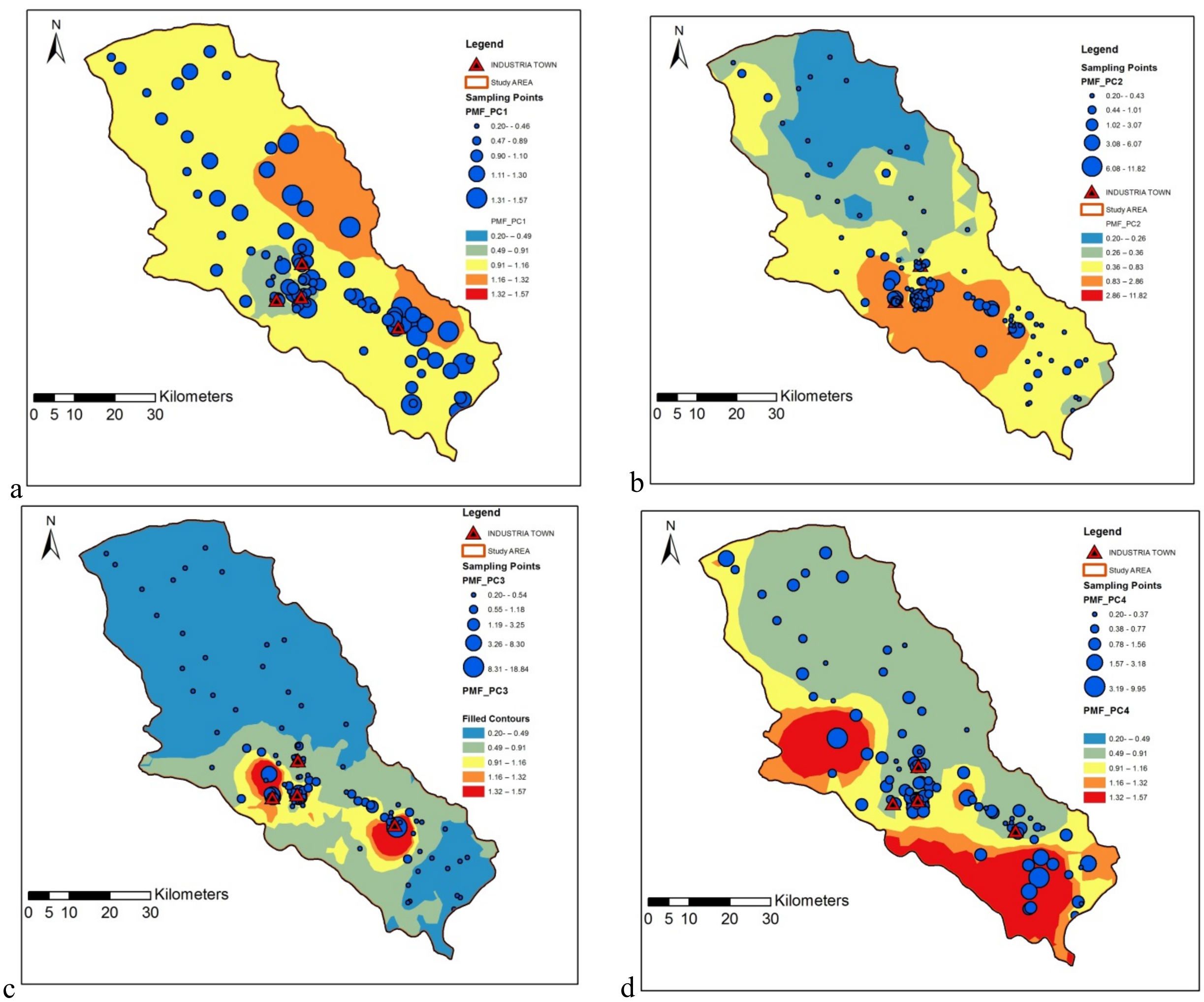

Figure 5. Spatial distribution of resolved problem by PMF for a) Factor 1 (Natural influence); b) Factor 2 (Anthropogenic Sources); c) Factor 3 (Anthropogenic Sources); d) Factor 4 (Agricultural and industrial influence).

Mining, disposal, and incineration of municipal and industrial wastes, fossil fuel processing and combustion, wood preservation, pesticide production and application are the anthropogenic activities that discharge arsenic into the environment (Huang & Conte, 2009), and most of them occur in the study area. PC1 of PCA analysis is same as this factor.

Factor 3 (Anthropogenic sources)
The 3rd factor is characterized by Zn and Cd concentrations and less strongly by Pb. Zn is a very common soil pollutant and its sources consists of domestic wastewater, coal–burning power plants, and manufacturing activities including metals, atmospheric fallout, and fertilizer and cement production (González–Macías *et al*., 2013). The main source of anthropogenic Zn in the environment is industrial states which have activities of refining zinc and lead. Also, unmanaged disposal of industrial waste from metallurgical activities are the main sources of Cd and As around industrial towns located in this area. Higher scores of factor 3 in the map (Fig 5.c) around industrial town can verify this statement. This factor can be used to compare the PC1 of PCA, which in both spatial maps, have higher values distributed around industrial towns.

Factor 4 (Agricultural and industrial influence)
The fourth factor is dominated by Bi and Cd. Other elements have low proportion in this factor. Bismuth occurs in native form and in minerals such as bismite. The production of metallic bismuth is related to lead and copper refining. Also, Bismuth compounds are used as pesticides which are highly insoluble (Krieger, 2001). So, Bi has remained in agricultural areas. This factor can be associated with agricultural activities and some industrial activities, especially in southeastern part of this basin. The spatial map of factor 4 shows higher value in the agricultural area (Fig. 5.d).

Conclusion

Soil has complex media and due to this, recognition of physical, chemical and biological behavior for better interpretation and understanding of its environment is necessary. Understanding the amount of heavy metals in the environment is important to mankind and wildlife health. Source apportionment studies can be used as a fairly precise, rapid and cost–effective method for identifying pollution sources and their relative contributions to the pollution. Soil pollutant determination is one of the essential parts of soil pollution studies. If the main sources of soil pollution are determined, then control and prevention is easier and have lower costs in the treatment and remediation processes.

In the present study, PMF, PCA and CA and using geostatistics were applied to evaluate the main sources of heavy metal pollution. Descriptive statistics of data in the study area indicates significant variation of soil heavy metal concentrations in the study area. Results show a strong influence of anthropogenic sources on the enrichment of heavy metals, especially Zn, Pb, Cd, As, and Cu in soils of the study area.

PCA and CA were done for heavy metal concentrations. PCA results showed 4 principal components which are potential sources for soil pollution. But negative values in the results posed problems in the interpretation of the results, especially for PC–4. PMF allows controlling of non–representative data and outlier data to decrease their importance by using specific error estimates. Four factors were determined. One of them is related to background components and natural influence of elements in soils. Prepared map for this factor showed higher values in the eastern part of the study area which igneous rocks are dominant. This subject can be helpful in the management and landuse strategies of the study area. Two factors were associated with industrial activities in this area. One of them was mainly controlled by Zn and Cd and the other one is related to Pb and As which is associated with anthropogenic sources including smelting and refining of Zn and Pb, industrial activities and inappropriate waste disposal. So, decision–makers can use these results in remediation and cleanup programs, and future plans. The last one is controlled by agricultural activities.

Application of the PMF approach produced interesting results, besides, GIS–based technique was successfully applied on the positive scores produced. Therefore, the combination of a geostatistical method by PMF results leads to classified map which is very useful in the interpretation of results and determination of pollutant.

Acknowledgements

Productive comments and helpful suggestions from the reviewers are acknowledged, which have helped improve this manuscript noticeably.

References

Ali, S. M., Malik, R. N., 2011. Spatial distribution of metals in top soils of Islamabad City, Pakistan. Environmental monitoring and assessment, 172 (1–4): 1–16.

Amini, M., Afyuni, M., Fathianpour, N., Khademi, H., Flühler, H., 2005. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. Geoderma, 124 (3): 223–233.

Anttila, P., Paatero, P., Tapper, U., Järvinen, O., 1995. Source identification of bulk wet deposition in Finland by positive matrix factorization. Atmospheric Environment, 29 (14): 1705–1718.

Bhuiyan, M. A., Parvez, L., Islam, M. A., Dampare, S. B., Suzuki, S., 2010. Heavy metal pollution of coal mine–affected agricultural soils in the northern part of Bangladesh. Journal of hazardous materials, 173 (1): 384–392.

Chapman, H.D., 1965. Cation–exchange capacity 1. Methods of soil analysis. Part 2. Chemical and microbiological properties, (methods of soil and), pp.891–901

Chen, H., Teng, Y., Wang, J., Song, L., Zuo, R., 2013. Source apportionment of trace element pollution in surface sediments using positive matrix factorization combined support vector machines: application to the Jinjiang River, China. Biological trace element research, 151 (3): 462–470.

Comero, S., 2011, Source identification of environmental pollutants using chemical analysis and Positive Matrix Factorization. PhD Thesis.

Comero, S., Servida, D., De Capitani, L., Gawlik, B. M., 2012. Geochemical characterization of an abandoned mine site: a combined positive matrix factorization and GIS approach compared with principal component analysis. Journal of Geochemical Exploration, 118: 30–37.

Conyers, M.K. Davey, B.G., 1988. Observations on some routine methods for soil pH determination. Soil Science, 145 (1): 29–36.

Davis, H. T., Aelion, C. M., McDermott, S., Lawson, A. B., 2009. Identifying natural and anthropogenic sources of metals in urban and rural soils using GIS–based data, PCA, and spatial interpolation. Environmental Pollution, 157(8): 2378–2385.

Davies, B.E., 1974. Loss–on–ignition as an estimate of soil organic matter 1. Soil Science Society of America Journal, 38 (1): 150–151.

Norris, G. A., Duvall, R., Brown, S. G., Bai, S., 2014. EPA Positive Matrix Factorization (PMF) 5.0 fundamentals and User Guide Prepared for the US Environmental Protection Agency Office of Research and Development, Washington, DC. DC EPA/600/R–14/108.

Facchinelli, A., Sacchi, E., Mallen, L., 2001. Multivariate statistical and GIS–based approach to identify heavy metal sources in soils. Environmental pollution, 114 (3): 313–324.

Farnham, I. M., Singh, A. K., Stetzenbach, K. J., & Johannesson, K. H., 2002. Treatment of nondetects in multivariate analysis of groundwater geochemistry data. Chemometrics and Intelligent Laboratory Systems, 60 (1): 265–281.

González–Macías, C., Sánchez–Reyna, G., Salazar–Coria, L., Schifter, I., 2014. Application of the positive matrix factorization approach to identify heavy metal sources in sediments. A case study on the Mexican Pacific Coast. Environmental monitoring and assessment, 186(1): 307–324.

Guan, Q., Wang, F., Xu, C., Pan, N., Lin, J., Zhao, R., Luo, H., 2018. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, northwest China. Chemosphere, 193: 189–197.

Hulseman, J., 1966. An inventory of marine carbonate materials. Journal of Sedimentary Petrology ASCE, 36 (2): 622–625.

Hao, L., Tian, M., Zhao, X., Zhao, Y., Lu, J., Bai, R., 2016. Spatial distribution and sources of trace elements in surface soils, Changchun, China: Insights from stochastic models and geostatistical analyses. Geoderma, 273: 54–63.

Harris, P., Brunsdon, C., & Charlton, M., 2011. Geographically weighted principal components analysis. International Journal of Geographical Information Science, 25 (10): 1717–1736.

Huang, S., Conte, M. H., 2009. Source/process apportionment of major and trace elements in sinking particles in the Sargasso Sea. Geochimica et Cosmochimica Acta, 73 (1): 65–90.

Huang, Y., Li, T., Wu, C., He, Z., Japenga, J., Deng, M., Yang, X., 2015. An integrated approach to assess heavy metal source apportionment in peri–urban agricultural soils. Journal of hazardous materials, 299: 540–549.

Hussain, R., Khattak, S. A., Shah, M. T., Ali, L., 2015. Multistatistical approaches for environmental geochemical assessment of pollutants in soils of Gadoon Amazai Industrial Estate, Pakistan. Journal of Soils and Sediments, 15(5): 1119–1129.

Khamehchiyan, M., Nikoudel, M.R. Boroumandi, M., 2011. Identification of hazardous waste landfill site: a case study from Zanjan province, Iran. Environmental earth sciences, 64 (7): 1763–1776.

Kim, E., Hopke, P. K., 2005. Identification of fine particle sources in mid–Atlantic US area. Water, Air, and Soil Pollution, 168 (1–4): 391–421.

Kim, E., Hopke, P. K., Edgerton, E. S., 2003. Source identification of Atlanta aerosol by positive matrix factorization. Journal of the Air & Waste Management Association, 53 (6):731–739.

Kim, E., Hopke, P. K., Edgerton, E. S., 2004. Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization. Atmospheric Environment, 38(20): 3349–3362.

Krieger, R. (Ed.)., 2001. Handbook of Pesticide Toxicology, Two–Volume Set: Principles and Agents (Vol. 1). Academic Press.

Li, J., He, M., Han, W., Gu, Y., 2009. Analysis and assessment on heavy metal sources in the coastal soils developed from alluvial deposits using multivariate statistical methods. Journal of Hazardous Materials, 164 (2): 976–981.

Liang, J., Zhong, M., Zeng, G., Chen, G., Hua, S., Li, X., Yuan, Y., Wu, H., Gao, X., 2016. Risk management for optimal land use planning integrating ecosystem services values: a case study in Changsha, Middle China. Sci. Total Environ. 579:1675–1682

Liang, J., Feng, C., Zeng, G., Gao, X., Zhong, M., Li, X., Fang, Y., 2017a. Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. Environmental Pollution, 225: 681–690.

Liang, J., Feng, C., Zeng, G., Zhong, M., Gao, X., Li, X., He, X., Li, X., Fang, Y. and Mo, D., 2017b. Atmospheric deposition of mercury and cadmium impacts on topsoil in a typical coal mine city, Lianyuan, China. Chemosphere, 189: 198–205

Liang, J., Zhong, M., Zeng, G., Chen, G., Hua, S., Li, X., Yuan, Y., Wu, H. Gao, X., 2017c. Risk management for optimal land use planning integrating ecosystem services values: A case study in Changsha, Middle China. Science of the Total Environment, 579: 1675–1682

Liu, D., Li, Y., Ma, J., Li, C., Chen, X., 2016. Heavy metal pollution in urban soil from 1994 to 2012 in Kaifeng City, China. Water, Air, & Soil Pollution, 227 (5): 154.

Lin, Y. P., Cheng, B. Y., Chu, H. J., Chang, T. K., Yu, H. L., 2011. Assessing how heavy metal pollution and human activity are related by using logistic regression and kriging methods. Geoderma, 163(3): 275–282.

Mitchell, K. N., Gómez, M. S. R., Barrera, A. L. G., Flores, L. Y., de la Torre, J. A. F., González, F. J. A., 2016. Evaluation of Environmental Risk of Metal Contaminated Soils and Sediments Near Mining Sites in Aguascalientes, Mexico. Bulletin of environmental contamination and toxicology, 97 (2): 216–224.

Olubunmi, F. E., & Olorunsola, O. E., 2010. Evaluation of the status of heavy metal pollution of sediment of Agbabu bitumen deposit area, Nigeria. European Journal of Scientific Research, 41 (3): 373–382.

Paatero, P., 1997. Least squares formulation of robust non–negative factor analysis. Chemometrics and intelligent laboratory systems, 37(1): 23–35.

Paatero P, Hopke PK, Song X–H, Ramadan Z., 2002. Understanding and controlling rotations in factor analytic models. Chemometrics and Intelligent Laboratory Systems, 60: 253–264

Paatero, P., & Tapper, U., 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5 (2): 111–126

Pekey, H., Pekey, B., Arslanbaş, D., Bozkurt, Z. B., Doğan, G., Tuncel, G., 2013. Source apportionment of personal exposure to fine particulate matter and volatile organic compounds using positive matrix factorization. Water, Air, & Soil Pollution, 224(1): 1–11.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., Sisler, J. F., 1998. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. Journal of Geophysical Research: Atmospheres, 103(D15): 19045–19057.

Polissar, A. V., Hopke, P. K., Poirot, R. L., 2001. Atmospheric aerosol over Vermont: chemical composition and sources. Environmental science & technology, 35(23): 4604–4621.

Reimann, C., Arnoldussen, A., Englmaier, P., Filzmoser, P., Finne, T. E., Garrett, R. G., Nordgulen, Ø., 2007. Element concentrations and variations along a 120–km transect in southern Norway–Anthropogenic vs. geogenic vs. biogenic element sources and cycles. Applied Geochemistry, 22(4): 851–871.

Rhoades, J.D., 1996. Salinity: Electrical conductivity and total dissolved solids. Methods of Soil Analysis Part 3—Chemical Methods, (methodsofsoilan 3): 417–435.

Sharma, S. K., Mandal, T. K., Jain, S., Sharma, A., Saxena, M., 2016. Source Apportionment of PM2. 5 in Delhi, India Using PMF Model. Bulletin of environmental contamination and toxicology, 97 (2): 286–293.

Sofowote, U. M., McCarry, B. E., Marvin, C. H., 2008. Source apportionment of PAH in Hamilton Harbour suspended sediments: comparison of two factor analysis methods. Environmental Science & Technology, 42(16): 6007–6014.

Stocklin, J. & Eftekharnezhad, J., 1969. Zanjan Quadrangle map, with explanatory text. Geol. Survey. Iran, scale 1:250,000.

Turekian, K. K., Wedepohl, K. H., 1961. Distribution of the elements in some major units of the earth's crust. Geological Society of America Bulletin, 72(2): 175–192.

Udayakumar, P., Jose, J. J., Krishnan, K. A., Kumar, C. R., Manju, M. N., Salas, P. M., 2014. Heavy metal accumulation in the surficial sediments along southwest coast of India. Environmental Earth Sciences, 72 (6): 1887–1900.

Vaccaro, S., Sobiecka, E., Contini, S., Locoro, G., Free, G., Gawlik, B. M., 2007. The application of positive matrix

factorization in the analysis, characterisation and detection of contaminated soils. Chemosphere, 69 (7): 1055–1063.
Wang, D., Tian, F., Yang, M., Liu, C., Li, Y. F., 2009. Application of positive matrix factorization to identify potential sources of PAHs in soil of Dalian, China. Environmental Pollution, 157 (5): 1559–1564.
Wang, Q., Xie, Z., Li, F., 2015. Using ensemble models to identify and apportion heavy metal pollution sources in agricultural soils on a local scale. Environmental Pollution, 206: 227–235.
Zhang, C., Li, Z., Yang, W., Pan, L., Gu, M., Lee, D., 2013. Assessment of metals pollution on agricultural soil surrounding a lead–zinc mining area in the Karst region of Guangxi, China. Bulletin of environmental contamination and toxicology, 90(6): 736–741.
Zhang, C., Wu, L., Luo, Y., Zhang, H., Christie, P., 2008. Identifying sources of soil inorganic pollutants on a regional scale using a multivariate statistical approach: role of pollutant migration and soil physicochemical properties. Environmental Pollution, 151 (3): 470–476.
Zhang, Z., Juying, L., Mamat, Z., 2016. Sources identification and pollution evaluation of heavy metals in the surface sediments of Bortala River, Northwest China. Ecotoxicology and environmental safety, 126: 94–101.
Zhao, Y., Wang, Z., Sun, W., Huang, B., Shi, X., Ji, J., 2010. Spatial interrelations and multi–scale sources of soil heavy metal variability in a typical urban–rural transition area in Yangtze River Delta region of China. Geoderma, 156 (3–4): 216–227.