Caspian J. Env. Sci. 2015, Vol. 13 No.4 pp. 333~347

CJES
Caspian Journal of Environmental Sciences

[Research]

Application of multivariate statistics and geostatistical techniques to identify the spatial variability of heavy metals in groundwater resources

F. Khanduzi, A. Parizanganeh*, A. Zamani

Department of Environmental Sciences, Faculty of Science, Environmental Science Research Laboratory, University of Zanjan, Zanjan, Iran

** Corresponding author's E-mail: h_zanganeh@znu.ac.ir*

(Received: Feb. 29.2015 Accepted: July. 22.2015)

ABSTRACT
The performance of geostatistical and spatial interpolation techniques were investigated for estimation of spatial variability of heavy metals and water quality mapping of groundwater resources in Ramiyan district (Golestan province, Iran). 24 spring/well water samples were collected and the concentration of heavy metals (Ni, Co, Pb, Cd and Cu) was determined using differential pulse polarography. Multivariate and geostatistical methods have been applied to differentiate the influences of natural processes and human activities as the sources of heavy metal pollutants in groundwater across the study area. The results of the cluster analysis and factor analysis show that Ni and Co are grouped in the factor F1, whereas, Pb and Cd in F2 and Zn and Cu in F3. The probability of presence of elevated levels for the three factors was predicted by utilizing the most appropriate Variogram Model, whilst the performance of methods, was evaluated using mean absolute error, mean bias error and root mean square error. The spatial structure results show that the variograms and cross-validation of the six variables can be modeled with three methods, namely, the radial basis fraction, inverse distance weight and ordinary kriging. Moreover, the results illustrated that radial basis fraction method was the best due to its highest precision and lowest error. The geographic information system can fully display spatial patterns of heavy metal concentrations in groundwater resources of the study area.

Key words: Groundwater, Heavy metals, Geostatistical, Multivariate statistics, Interpolation, Spatial mapping.

INTRODUCTION
Water is the basic requirement for all life on earth and an increase in the population and urbanization necessitates growth of agricultural and industrial sectors, increasing demand for fresh water. When surface water is not available; the alternative is to depend on groundwater (GW) (Subramani *et al.,* 2012). A variety of natural and human factors, affects the quality and use of water resources. Heavy metals are among the major pollutants of these sources (Marcovecchio *et al.,* 2007). Many human activities, such as agriculture, mining and the combustion of fossil fuels, release heavy metals into the environment. Thereby, with an increase in their concentration and a decrease in the capacity of soils towards heavy metals, these leach into the soil solution and GW and then they accumulate in living tissues among people through the food chain (Mantovi *et al.,* 2003; Lei *et al.,* 2008), in addition to being sensitive indicators for monitoring changes in the aqueous environment. In environmental monitoring, such as groundwater quality investigations, the collected data may harbor significant uncertainty, including complex or extremely complicated variations in the observed values of measurable characteristics, of the investigated medium or pollution sources in time and space (Yeh *et al.,* 2006). Geostatistics, is a spatial statistical technique used in environmental monitoring, which is applied to analyze and map distributions of pollutant concentrations and their spatial and

temporal variations. It is more widely used to analyze the collected data from groundwater resources (Yu *et al.,* 2003; Yeh *et al.,* 2006; Nas & Berkta, 2006; Khodapanah & Sulaiman, 2009; Uyan & Cay, 2010; Amin et al., 2010; Belkhiri *et al.,* 2011; Sarukkalige, 2012). Furthermore, the application of different multivariate statistical techniques helps in the interpretation of complex data matrices, for a better understanding of water quality of the studied systems. These methods allow identification of possible factors/sources which influence the water systems and offer a valuable tool for a reliable management of water (Shrestha & Kazama, 2007; Iscen *et al.,* 2008; Ogunribido & Kehinde–Philips, 2011; Li *et al.,* 2012; Bajpayee et al., 2012). Multivariate geostatistical methods combine the advantages of geostatistical techniques and multivariate analysis, while incorporating spatial or temporal correlations and multivariate relationships to detect and map the varied sources of spatial variation on different scales (Smyth & Istok, 1989; Einax & Soldt, 1999; Yeh *et al.,* 2006; Zheng *et al.,* 2008; Lin *et al.,* 2009). Excavation of coal mines, agricultural activities and development of industrial parks in Ramiyan, in Golestan Province (Iran), provoke evaluation of contaminations resulting from these activities. The lack of a systematic investigation of the probable contamination by heavy metals in Ramiyan, urges an assessment of the quality of groundwater sources in this area.

The aquifer is the main source for drinking and irrigation critical for the local residents. 24 well/spring samples were collected and analyzed by voltametric method for determination of such heavy metals. The presence and concentration of heavy metals were determined and the results were compared to the maximum contaminant level, specified by WHO and the Institute of Standards and Industrial Research of Iran (ISIRI). This study aims at investigating the contents of Cu, Ni, Zn, Cd, Pb and Co in the groundwater resources of Ramiyan, including the analysis of their spatial distribution as well as unveiling their possible sources by integrating multivariate statistical and geostatistical methods.

MATERIALS AND METHODS

Site description

Golestan Province is located at the southeast of the Caspian Sea in Northern Iran. The study area is Ramiyan district, with an area of 780.73 km^2 situated between 54° 45´ and 55° 15´ east longitude and 36° 48´ and 37° 12´ north latitude. The main activity carried out in this area is agriculture and the main crops grown are wheat, oilseeds, rice and garden products (Mosaedi & Gharib, 2008). Due to the presence of coal mines, industrial and mining activities have also been developed across the study area.

Sample collection

The samples were collected for the assessment of groundwater pollution with heavy metals from twenty four stations (wells/springs) in the study area (Fig. 1, Table 1). The sampling was carried out in summer 2012 and three replicate samples from each station were selected for analysis. The glassware and vessels were treated in 10% (v/v) nitric acid solution for 24 h and washed with distilled and de-ionized water. The samples were collected in polypropylene containers, labeled and a few drops of HNO_3 (ultrapure grade) of pH < 2 were added immediately, to prevent the loss of metals, bacterial and fungal growth. These were then stored in a refrigerator.

Multivariate and geostatistical analysis

The multivariate analysis provides techniques, such as the Principle Component Analysis (PCA), Factor Analysis (FA) and Cluster Analysis (CA) for classifying the inter-relationship of measured variables (Zamani *et al.,* 2012). The CA was performed on the data, by utilizing the ward method and squared euclidean distance characteristic. Multivariate geostatistical methods combine the advantages of geostatistical techniques and multivariate analysis, whereas, the geostatistical techniques have been applied to illustrate the incorporating spatial or temporal correlations and multivariate relationships, in order to map the various sources of spatial variation on divergent scales (Faccinelli *et al.,* 2001). Geostatistics is presented as a collection of techniques for solving estimation problems involving spatial variables. It includes a variety of tools such as interpolation, integration and differentiation of hydro-geologic parameters to produce the prediction surface and other derived characteristics from measurements at known locations (Sahoo & Jha, 2014).

The first step in the geostatistical estimation, is a provision of a model that can facilitate the computation of semivariogram value for any possible sampling intervals.

The most commonly used models are the Spherical, Exponential, Gaussian and Pure Nugget effect (Isaaks & Srivastava, 1989).

The semivariogram plays a fundamental role in the analysis of geostatistical data by employing the Kriging Method. Prior to performing Kriging, a valid semivariogram model has to be selected and the model parameters have to be estimated (Pang *et al.,* 2009). An experimental semivariogram is calculated as follows:

$$\gamma(h)=\frac{1}{2n(h)}\sum_{i=1}^{n}\left[Z(x_i)-Z(x_i-h)\right]^2 \qquad (1)$$

Where, $\gamma(h)$ denotes the semivariogram, h is the spatial interval, which is designated as lag; $n(h)$ is the observed paired data, when the h interval $Z(x_i)$ and $Z(x_i-h)$ are the measured values, when the Z(x) values are as xi+h, respectively. Valid models which are commonly fitted to the experimental semi variograms include the spherical, Gaussian and exponential functions. These are characterized by a sill, which represents the covariance accounted for by the model and a range that signifies the extent of spatial correlation. The value of the semi variograms is referred to as the nugget effect, where the model approaches the abscissa. These significant geostatistical parameters can indicate the spatial variation and relativity of regionalized variables under a certain scale (Yang *et al.,* 2009).

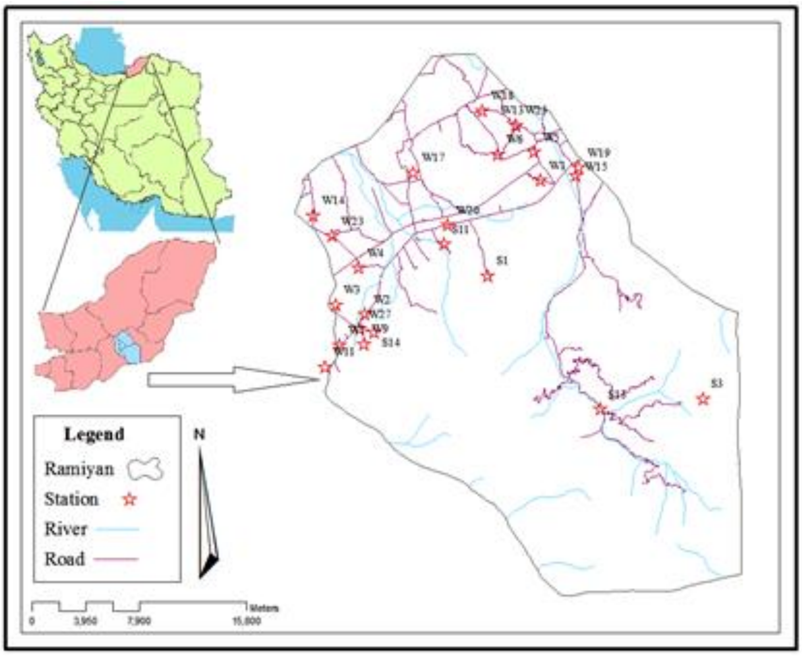


Fig 1. Location map of Ramiyan and the sampling points.

Interpolation methods

Kriging Method was used as estimating tool in sustainable management of groundwater. It is a geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown areas (Sahoo & Jha, 2014). This technique is an exact interpolation estimator, which is used to detect the best linear unbiased estimate. The optimum linear unbiased estimator must have a minimum variance of error of estimation (Einax & Soldt, 1999; Ahmadi & Sedghamiz, 2008). In order to estimate the values of some locations which are not sampled, it is necessary to solve the following linear equation:

$$Z^*(x) = \sum_{i=1}^{n} W_i Z(x_i) \tag{2}$$

Z^* denotes the estimate of the unknown value $Z(x_i)$ and W_i are the weights of known neighboring points x_i.

Kriging is an estimating method that is stable on weighty mobile average coincident. This estimator is known as a best unbiased linear estimator. Spherical, circular, Gaussian and exponential functions are available models when the Kriging method is ordinary (Nas, 2009). Goovaerts describes the detail of the method (Goovaerts, 1997).

Because it uses statistical models, it allows a variety of map outputs, including predictions, prediction standard errors, probability, and quantile maps. Among the various forms of Kriging, ordinary Kriging has been used widely as a reliable estimation method (Nas, 2009). In interpolation with the Inverse Distance Weighted (IDW) method, a weight is attributed to the point to be measured. In other words weight is the function of inverse distance and closer points have more influence in estimating unknown points (Eslami *et al.*, 2013). The amount of this weight depends on the distance of the point to another unknown point. These weights are controlled on the bases of power ten.

So, with an increase of power, the effect of the points (that are farther) diminishes, whilst a lesser power distributes the weights more uniformly between neighboring points. In this method the distance between the points counts, so that, the points of equal distance have equal weights (Balakrishnan *et al.*, 2011). The weight factor is determined based on the distance between the data points as follows:

$$W_i = \frac{D_i^{-\alpha}}{\sum_{i=1}^{n} D_i^{-\alpha}} \tag{3}$$

Where W_i designates the weight of point D_i which is the distance between point i and the unknown point, α which is the weight on the bases of power ten and n is the number of data points (Karandish & Shahnazari, 2014). Kriging in geostatistics is similar to inverse distance weighting except that the weights are based not only on the distance between the measured sampling points but also on the overall spatial arrangement among the sampling points. The basic assumption in kringing is that the sampling points that are close to each other are similar than those that are away. Kriging is regarded as an optimal spatial interpolation method, which is a type of weighted moving average (Gorai & Kumar, 2013). The Radial Basis Functions (RBF) Methods are a series of exact interpolation techniques, where the surface must go through for each measured sample value. The basis of each function has a different shape and results in a slightly different interpolation surface (Kazemi Poshtmasari *et al.*, 2012). RBF Methods predict values that can vary above the maximum or below the minimum of the measured values. For all RBF Methods, there is a parameter that controls the smoothness of the resulting surface. The estimated values of the methods are based on a mathematical function that minimizes the overall surface curvature, generating surfaces that are quite smooth. The differences among them are slight, so the generated surfaces are almost similar. A formula f, which minimizes the following factor [eq. (4)], is an example of the RBF technique and more specifically of the exact SP line method (Karydas *et al.*, 2009):

$$A(f) + \sum_{i=1}^{n} W_i^2 \left[f(x_i) - y(x_i)\right]^2 \tag{4}$$

Where $y(x_i) = z(x_i) + \varepsilon(x_i)$ signifies the source of random error, z is the measured value of an attribute at point x_i and epsilon is the associated random error. The term $A(f)$ represents the smoothness of the function f and the second term represents its proximity to the data (Karydas *et al.,* 2009).

Evaluation criteria

The adequacy and validity of the developed semivariogram models was tested satisfactorily by a technique called cross-validation. The idea of cross-validation consists of removing a datum at a time from the data set and reestimating this value from remaining data by using different variogram models. The interpolated and actual values are compared, and the model that yields the most accurate predictions is retained (Burrough & McDonnell, 1998; Karimi Nezhad *et al.,* 2012 ;). In this paper, to compare the applied Interpolation methods, a cross validation was performed by utilizing the Mean Bias Error (MBE), Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) of the statistical parameters. When MAE and MBE shift to zero, the applied method simulates the fact well. Finally, we used the RMSE to evaluate the model performances in the cross-validation mode. Each of these measures is such 'dimensioned' that, it expresses an average interpolator error in the units of the variable of interest. The smallest RMSE indicates the most accurate predictions. This method was recently adopted by many researchers (Twomey & Smith, 1996; Willmott & Matsuura, 2006; Kazemi Poshtmasari *et al.,* 2012; Karandish & Shahnazari, 2014). These parameters are calculated according to the following equation Nos. (5 to 7):

$$MAE = \frac{\sum_{i=1}^{n} \left| Z^*(x_i) - Z(x_i) \right|}{n} \quad (5)$$

$$MBE = \frac{\sum_{i=1}^{n} (Z^*(x_i) - Z(x_i))}{n} \quad (6)$$

$$RMSE = \frac{\sum_{i=1}^{n} \sqrt{(Z^*(x_i) - Z(x_i))^2}}{n} \quad (7)$$

Where Z (xi) is the observed value at point xi, Z*(xi) is the predicted value at point x and N denotes the number of samples.

RESULTS AND DISCUSSION

The extent of heavy metal contamination

The results of the analysis of target metal ions i.e., Co, Ni, Zn, Cd and Pb in samples from 24 wells/springs under study are given in Table (2). The results show that Co, Ni, Pb and Cd are evident in 100% of the samples and Zn and Cu are detected in 96% and 88% of the samples, respectively. The concentration of investigated metals (in µg/L) in the samples were found to be below their MCL and in the ranges of 5.69 - 92.44 for Zn, 1.23 -7.06 for Pb, 0.14-8.40 for Cu, 0.01-0.99 for Cd, 1.23 -21.79 for Ni and 0.49 -7.79 for Co. The geographical location of the sampling stations and the average concentrations of metals at each station are shown in Table (1).

Classification survey of heavy metals by the Cluster Analysis Method

Two main groups of elements have been determined using the Cluster Analysis Method, one group includes Ni and Co and the other comprises of Pb, Cd, Zn and Cu (Fig. 2).

Principal component analysis and factor analysis

The major objective of the Factor Analysis (FA) is to reduce the contribution of less significant variables so as to further simplify even more of the data structure given by the PCA. This goal can be achieved by rotating the axis defined by the PCA and the construction of new variables, which are also called Varifactors (Shrestha & Kazama, 2007). Prior to such analysis, the raw data is commonly normalized to avoid misclassifications, due to the varied order of magnitude and range of variation of the analytical parameters (Tabachnick & Fidell, 2007). This process reduces the dimensionality of data by a linear combination of original data, to generate new latent variables which are orthogonal and uncorrelated to each other (Nkansah *et al.,* 2010). According to the results of the Eigen values in Table (3), three factors are extracted from the available data set, which accounts for over 82.07% of all the data variation. The common factors were extracted by means of the maximum-likelihood method with the Varimax-rotation.

Nickel and cobalt, contained in the first factor, are typical emitted elements of electronic plants. The second factor includes cadmium and lead elements which are emitted by the agricultural activities and the metallurgical plant.

The third factor is loaded with zinc and copper, which are emissions of batteries, pigments and fungicides. The heavy metal grouping has been explored in plotting the first three principle components generated from these parameters (Fig. 3).

Table 1. GPS location and concentration of heavy metals in sampling stations.

Station	X	Y	Zn (µg/L)	Pb (µg/L)	Cd (µg/L)	Cu (µg/L)	Ni (µg/L)	Co (µg/L)
W_1	331881	4103244	21.29	2.87	0.04	ND	2.48	6.20
W_2	318661	4093812	36.48	3.64	0.15	4.68	5.33	1.72
W_3	316518	4094590	19.22	7.06	0.37	3.41	2.69	1.29
W_4	318301	4097255	64.68	7.03	0.99	ND	2.99	1.78
W_5	331486	4105353	34.60	5.88	0.24	5.12	1.71	1.51
W_6	328781	4105267	12.85	2.92	0.09	1.86	2.29	2.32
S_1	327797	4096396	36.77	2.54	0.08	1.75	2.28	2.11
W_7	316718	4091641	6.73	3.26	0.08	3.06	7.48	2.18
S_2	343356	4086870	56.52	3.42	0.04	3.72	2.00	2.31
W_8	318492	4091655	15.11	5.56	0.08	7.33	6.98	1.81
W_9	315558	4090055	32.65	4.18	0.09	6.60	5.07	2.03
W_{10}	330180	4107361	52.07	6.46	0.2	7.43	1.89	1.12
W_{11}	315050	4101192	19.68	4.59	0.13	2.47	4.64	2.25
W_{12}	334472	4103554	18.46	3.14	0.03	0.14	4.43	0.81
W_{13}	322519	4104116	ND	2.98	0.15	1.27	5.12	1.77
W_{14}	327726	4108520	17.88	5.50	0.05	3.66	1.98	0.01
W_{15}	334669	4104096	50.23	5.49	0.05	2.16	2.84	1.19
W_{16}	324922	4100196	12.62	3.89	0.07	5.79	19.30	3.65
W_{17}	316424	4099638	10.61	3.06	0.05	2.54	3.98	2.77
S_3	324628	4098775	16.15	5.04	0.08	ND	3.29	2.77
W_{18}	330192	4107354	62.28	3.40	0.09	3.22	4.12	1.44
S_4	335731	4086350	18.93	3.28	0.04	1.48	2.81	1.32
S_5	319249	4092456	44.84	3.21	0.07	1.83	3.50	2.48
W_{19}	318221	4092774	31.49	3.14	0.07	4.58	5.50	1.82

Table 2. Summary of statistics of heavy metal contents in water samples (µg/L).

Metal	Ni	Co	Pb	Cd	Zn	Cu
Detected (%)	100%	100%	100%	100%	96%	88%
Min	1.23	0.49	1.23	0.01	5.69	0.14
Max	21.79	7.79	7.06	0.99	92.44	8.40
Mean	4.38	1.99	4.21	0.12	30.55	4.05
Standard deviation	3.65	1.22	1.94	0.17	18.77	14.26
WHO Standard	70	-	10	3	3000	1000
ISIRI Standard	70	-	10	3	3000	2000

Table 3. Rotated component matrix of three-factor model.

Variable	Component		
	F_1	F_2	F_3
Ni	0.90	-0.07	0.11
Co	0.84	-0.21	-0.10
Pb	-0.15	0.86	0.21
Cd	-0.08	0.88	-0.04
Zn	-0.52	-0.16	0.72
Cu	0.29	0.31	0.82
Eigen Value	2.21	1.57	1.12
Variance (%)	36.99	26.27	18.79
Cumulative (%)	36.99	63.27	82.07

†Extraction method: Principle component analysis. Rotation method: Varimax with Kaiser Normalization.

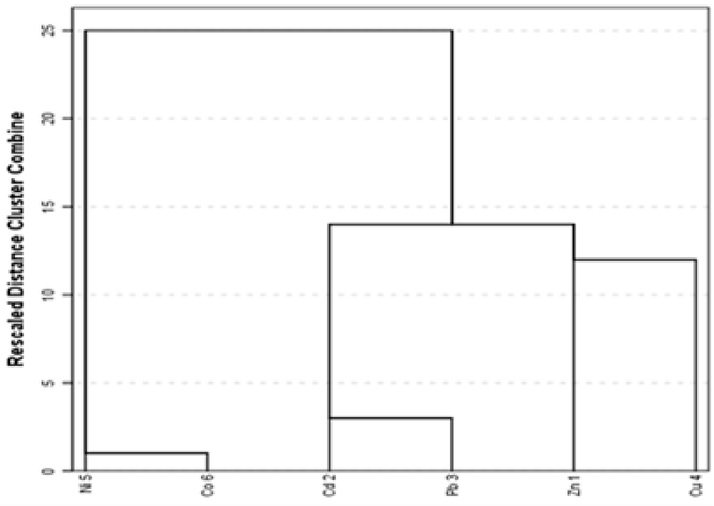


Fig. 2. Dendrogram of heavy metal concentrations in groundwater samples.

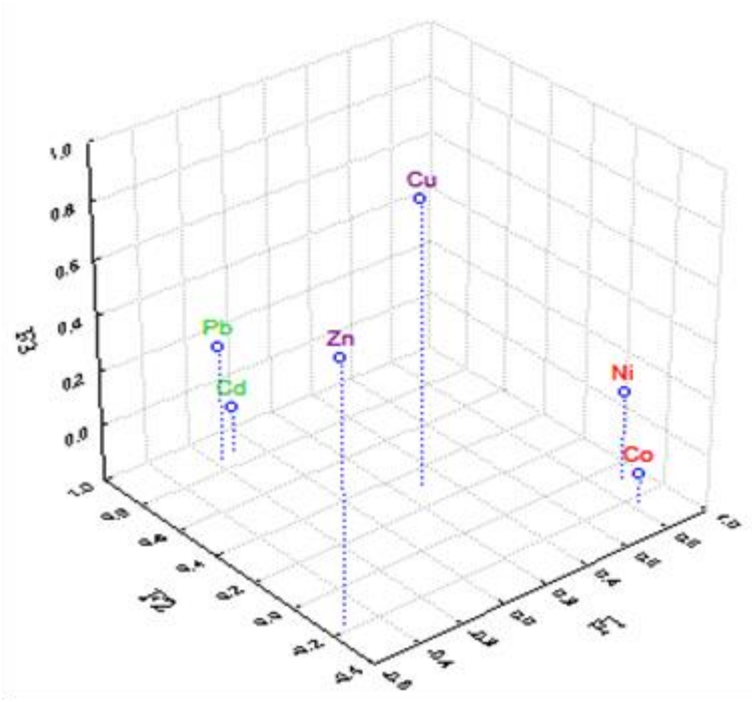


Fig. 3. Component plot in rotated space for heavy metals (Factor loading, factor 1 vs. factor 2 vs. factor 3, Rotation: varimax normalized, extraction: principle component).

Spatial structure analysis

The geostatistical analysis is to be assumed that the distribution behavior of the metal ions in the sampling stations is normal. The random and normal distribution assumptions were checked by the (K-S) (Kolmogorov-Smirnov) Methods. Alternatively, the homogeneity and normal distribution in the data, can be achieved by transforming the obtained data to another mathematically presentation, which lowers the difference between the data. This can be achieved by using the logarithmic form of data. The normality of heavy metal data set was checked by the Kolmogorov–Smirnov Test. It is often observed that environmental variables are lognormal (McGrath *et al.,* 2004), and data transformation is necessary to normalize such data sets. The normality tests of the six heavy metals for the 24 samples were performed as described by K-S test. It was detected that only Cu and Zn were in accordance with the normal distribution using K-S ($p>0.05$) before data transformation. To further normalize the data logarithmic transformation was utilized (Table 4).

After the logarithmic transformation of the original data, a normal distribution can be obtained. Thus, the following calculations must be performed on the logarithms of the data. After normalizing the data Semivariogram parameters were generated for each theoretical model.

Then, the confidence level of all variograms was evaluated using the ratio of nugget variance to sill which is regarded as a criterion for classifying the spatial dependence of ground water quality parameters. If this ratio is less than 25%, then the variable has strong spatial dependence; if the ratio is between 25 and 75%, the variable has moderate spatial dependence and the ratio greater than 75%, represents weak spatial dependence (Taghizadeh *et al,* 2008).

The most appropriate theoretical model was selected, which was based on highest R2 and lowest RSS (Table 5).

Table 4. Normal distribution behaviors of heavy metal concentration.

Metal	N	Mean	Std. Deviation	Kolmogorov-Smirnov Z	Asymp. Sig. (2-tailed)
Ni	74	4.38	3.65	1.90	0.001
Co	74	1.99	1.22	1.80	0.003
Pb	74	4.21	1.94	1.72	0.005
Cd	59	0.12	0.17	2.52	0.00
Zn	72	30.53	18.77	1.20	0.11
Cu	48	4.05	2.25	1.06	0.20
log Ni	74	0.55	0.25	0.67	0.75
log Co	74	0.24	0.21	0.90	0.38
log Pb	74	0.58	0.17	1.14	0.14
log Cd	59	-1.70	0.34	1.25	0.08
log Zn	72	1.39	0.28	0.86	0.43
log Cu	48	0.52	0.31	0.76	0.69

Table 5. Summary of the most appropriate models for different heavy metals of GW.

Heavy metals		Transformation	Best-fit model	Nugget (C_0)	Sill (C_0+C)	Proportion (C_0/ C_0+C)×100	R^2	RSS
F_1	Ni	Log-normal	Exponential	0.024	0.276	8.69	0.196	0.100
	Co	Log-normal	Exponential	0.029	0.162	17.90	0.194	0.033
F_2	Pb	Log-normal	Exponential	0.015	0.060	25.00	0.154	0.0032
	Cd	Log-normal	Exponential	0.095	0.412	23.05	0.052	0.256
F_2	Zn	Log-normal	Gaussian	0.0910	9.647	0.943	0.099	109.6
	Cu	Log-normal	Gaussian	1.195	4.805	24.86	0.179	57.23

The attributes of the semivariograms for each factor are summarized in Table (5). Semivariograms show that the first and second factors are appropriate with the Exponential

Model, whereas, the third factor fits well with the Gaussian Model. The values of R2 illustrate that the semivariogram models give good descriptions of the spatial structure of the heavy metals of groundwater. The nugget/sill ratios can be regarded as the criterion to classify the spatial dependence of data sets (Liu *et al.,* 2009). The ratio of nugget to sill (RNS) can be used to express the extent of spatial autocorrelations of environmental factors, for example, groundwater heavy metal concentrations, in this study. A low RNS indicates the strong spatial autocorrelations of heavy metal concentrations in groundwater sources, while a high RNS indicates that random effects play an important role in spatial heterogeneity of heavy metals (Zheng *et al.,* 2008). The RNS of six heavy metals demonstrate weak spatial correlations for all factors. Cross-validation permits the determination as to which model provides the best predictions (Adhikary *et al.,* 2012).

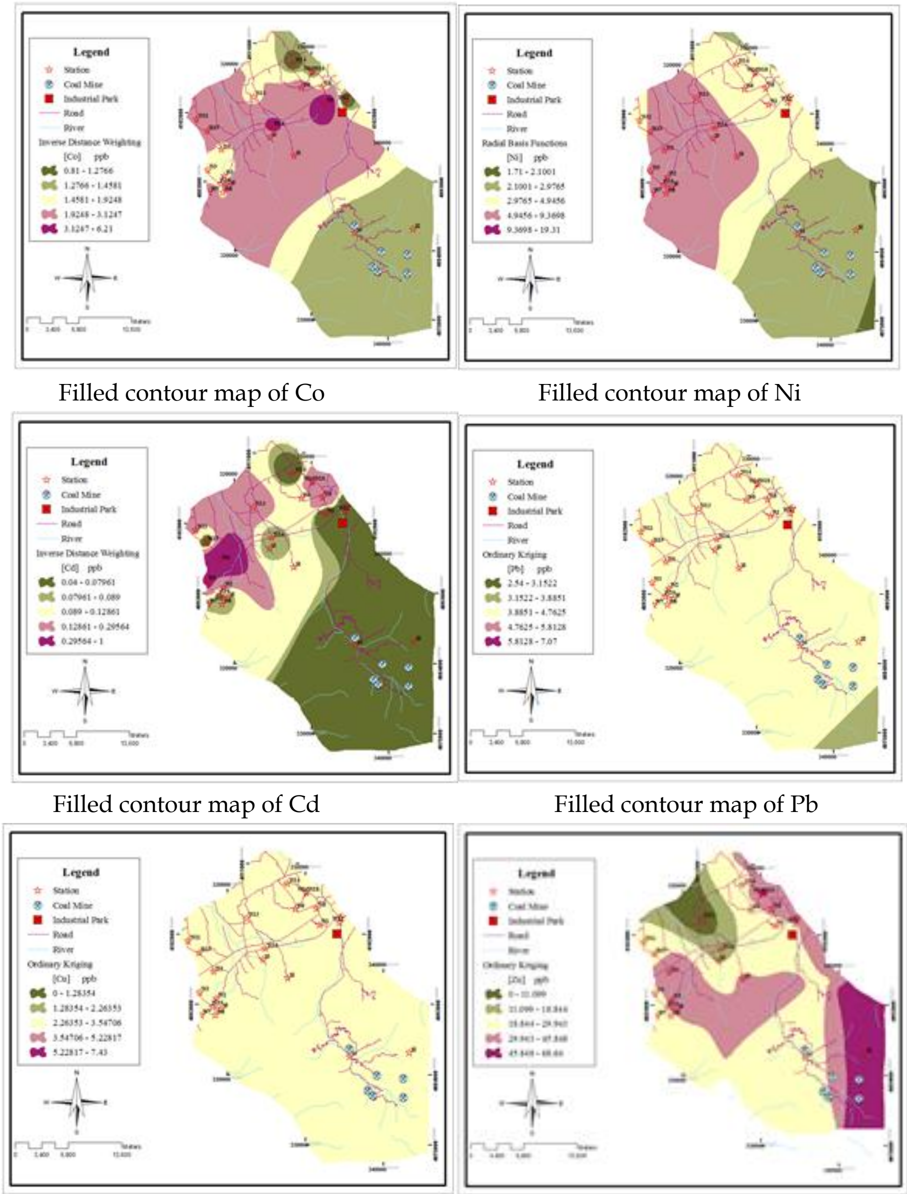
Table 6. Geostatistical analyses of heavy metals in groundwater (Ramiyan area).

Heavy Metal	Method	Model	Cross validation		
			MBE	MAE	RMSE
Ni	OK	Exponential	0.426	2.587	3.958
	IDW	1	0.240	2.216	3.828
		2	0.388	2.650	4.377
		3	0.268	2.877	4.812
	RBF	SP line with Tension	0.041	2.242	3.687
		Multi-quadric	0.308	3.636	4.352
		Inverse Multi-quadric	-0.011	2.218	3.628
Co	OK	Exponential	-0.018	0744	1.190
	IDW	1	-0.083	0.661	1.125
		2	-0.030	0.632	1.143
		3	-0.074	0.728	1.176
	RBF	SP line with Tension	0.002	0.727	1.132
		Multi-quadric	-0.027	0.739	1.173
		Inverse Multi-quadric	-0.041	0.695	1.107
Pb	OK	Exponential	-0.050	1.258	1.436
	IDW	1	-0.027	1.453	1.715
		2	0.013	1.617	1.799
		3	0.027	1.703	1.843
	RBF	SP line with Tension	0.168	1.382	1.618
		Multi-quadric	-0.052	1.604	1.808
		Inverse Multi-quadric	-0.010	1.312	1.440
Cd	OK	Exponential	-0.003	0.103	0.197
	IDW	1	-0.003	0.109	0.199
		2	-0.015	0.098	0.194
		3	-0.026	0.092	0.191
	RBF	SP line with Tension	0.000	0.111	0.196
		Multi-quadric	0.020	0.109	0.199
		Inverse Multi-quadric	0.004	0.107	0.202
Zn	OK	Gaussian	0.008	15.82	18.57
	IDW	1	2.927	18.01	20.56
		2	3.268	19.14	21.64
		3	2.402	19.33	23.34
	RBF	SP line with Tension	1.039	15.86	18.02
		Multi-quadric	-0.801	17.73	20.07
		Inverse Multi-quadric	0.024	17.21	17.15
Cu	OK	Gaussian	-0.056	1.961	1.006
	IDW	1	0.225	2.278	2.609
		2	0.247	2.401	2.878
		3	0.426	2.675	3.039
	RBF	SP line with Tension	0.202	2.020	2.417
		Multi-quadric	0.228	2.846	3.125
		Inverse Multi-quadric	-0.047	1.944	2.267

The applicability of different semivariogram models is tested by cross-validation and best model is selected (Table 6). In this study, ordinary kriging (OK), IDW and RBF were utilized to estimate six heavy metal concentrations.

Comparisons between different methods were carried out by the MAE, MBE, and RMSE statistical parameters. In this research, the Radial Basis Functions Method (Inverse Multiquadric Model) was found to be the most suitable method for the estimation of Ni mapping. Whereas, statistics for the geostatistical method also show that Ordinary Kriging for Pb (Exponential Model), Zn and Cu (Gaussian Model); the Inverse Distance Weighted method for Co (power 2) and Cd (power 3) provides a much better estimation for results of concentrations, than the other methods (Table 6).

After plotting the values of heavy metal concentrations of groundwater for various sample locations, drinking water quality maps for heavy metal concentrations, can be drawn to demonstrate locations, where the water is almost clean or to some extent at risk (Fig. 4).



Filled contour map of Co

Filled contour map of Ni

Filled contour map of Cd

Filled contour map of Pb

Filled contour map of Cu

Filled contour map of Zn

Fig. 4. Filled contour maps of heavy metals in sampling groundwater.

CONCLUSIONS

Due to the complexity and a large variation of environmental data sets, the application of geostatistical and multivariate statistical methods is recommended.

The main objective of this study was to determine the best estimators for providing heavy metals maps in ground water resources in Ramyian district. The application of multivariate statistical and geostatistical methods were performed on six heavy metals and three principal components were identified, so as to represent the variability of heavy metals in groundwater sources. From the spatial distributions of 6 heavy metals, it was evident that the parent materials and anthropogenic factors played important roles in heavy metal concentrations of GW in Ramiyan. The effects of these two factors varied with that of the heavy metals. The results of the Cluster Analysis (CA) and Factor Analysis (FA) on the heavy metals, showed that Ni and Co was grouped in factor F1, Pb and Cd in F2 and Zn and Cu in F3. The probability of the presence of elevated levels of the heavy metals studied in the groundwater was predicted by using the best-fit semivariogram model. The performance of methods was evaluated by utilizing the Mean Average Error (MAE), Mean Bias Error (MBE), and Root Mean Square Error (RMSE). Moreover, results showed that Radial Basis Functions (RBF), Inverse distance weighted (IDW) and Ordinary Kriging (OK) methods were the best methods employed to estimate the Ni; Co and Cd; Pb, Zn and Cu mappings, respectively. The Geographic Information System (GIS) can fully display the spatial patterns and relationships among landscape indices and heavy metal concentrations, in the groundwater of this area of study. Application of different multivariate statistical techniques interprets complex data matrices and better understanding of water quality. Although the concentrations of investigated metals in the collected samples were found to be below their maximum contaminant level values reported by WHO and ISIRI but the source of heavy metals contamination should be investigated specially in hot points within the studied area.

ACKNOWLEDGMENTS

Sincere gratitude to Rural Water and Wastewater Company (Golestan Province, Iran) for partial financial support (Grant Number 4987). The authors gratefully acknowledge Younes Khosravi's contribution to this work.

REFERENCES

Adhikary PP, Dash, ChJ, Chandrasekharan, H, Rajput, TB, S & Dubey, SK, 2012, Evaluation of groundwater quality for irrigation and drinking, using GIS and geostatistics in a peri-urban area of Delhi-India. *Arabian Journal of Geosciences,* 5: 1423 - 1434.

Ahmadi, SH, Sedghamiz, A, 2008, Application and evaluation of kriging and cokriging methods on groundwater depth mapping. *Environmental Monitoring and Assessment,* 138: 357 - 368.

Amin, MM, Ebrahimi, A, Hajian, M, Iranpanah, N & Bina, B, 2010, Spatial analysis of three agrichemicals in groundwater of Isfahan using GS+. *Journal of Environmental Health Science and Engineering,* 7: 71 - 80.

Bajpayee, S, Das, R, Ru, B, Adhikari, K & Chatterjee, PK, 2012, Assessment by multivariate statistical analysis of ground water geochemical data of Bankura, India, *International Journal of Environmental Science,* 3: 870 - 880.

Balakrishnan, P, Saleem, A & Mallikarjun, ND, 2011, Groundwater quality mapping using geographic information system (GIS): A case study of Gulbarga City, Karnataka, India. *African Journal of Environmental Science and Technology,* 5: 1069 - 1084.

Belkhiri, L, Boudoukha, A & Mouni, L, 2011, A multivariate Statistical Analysis of Groundwater Chemistry Data. *International Journal of Environmental Research,* 5: 537 - 544.

Burrough, PA, & McDonnell, RA, 1998, *Principles of Geographical Information Systems*. Oxford University Press. pp. 16 - 34.

Einax, JW & Soldt, U, 1999, Geostatistical and multivariate statistical methods for the assessment of polluted soils-merits and limitations. *Chemometrics and Intelligent Laboratory Systems,* 46: 79 - 91.

Eslami, H, Dastorani, J, Javadi, MR & Chamheidar, H, 2013, Geostatistical Evaluation of Ground Water quality Distribution with GIS (Case Study: Mianab-Shoushtar Plain). Bulletin of Environment, Pharmacology and Life Sciences, 3: 78 - 82.

Faccinelli, A, Sacchi, E & Mallen, L, 2001, Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. *Environmental Pollution,* 114: 313 - 324.

Goovaerts, P, 1997, Geostatistics for natural resources evaluation, Oxford University Press, New York, Page 483.

Gorai, AK, Kumar, S, 2013, Spatial Distribution Analysis of Groundwater Quality Index Using GIS: A Case Study of Ranchi Municipal Corporation (RMC) Area, *Geoinformatics and Geostatistics,* 1: 1 - 11.

Isaaks, EH, & Srivastava, RM, 1989, *An Introduction to applied geostatistics.* Oxford University Press- New York, Page 561.

Iscen, CCF, Ozgür, OE, Semra, SI, Naime, NA, Veysel, VY & Seyhan, SA, 2008, Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environmental Monitoring and Assessment,* 144: 269 - 76.

Karandish, F & Shahnazari, A, 2014, Appraisal of the geostatistical methods to estimate Mazandaran coastal ground water quality. *Caspian Journal of Environmental Sciences,* 12: 129 - 146.

Karimi Nezhad, MT, Mir Ahmad, F & Mohammadi, Kh, 2012, Assessment of nitrates contamination in topsoils of Babagorgor watershed, western Iran. ARPN. *Journal of Agricultural and Biological Science,* 7: 792 - 797.

Karydas, ChG, Gitas, IZ, Koutsogiannaki, E, Lydakis-Simantiris, N & Silleos, GN, 2009, Evaluation of spatial interpolation techniques for mapping agricultural topsoil properties in Crete. *European Association of Remote Sensing Laboratories, EARSeLe Proceedings,* 8: 26 - 39.

Kazemi Poshtmasari, H, Tahmasebi Sarvestani, Z, Kamka, B, Shataei, Sh, Sadeghi, S, 2012, Comparison of interpolation methods for estimating pH and EC in agricultural fields of Golestan province (north of Iran). *International Journal of Agriculture and Crop Sciences,* 4:157-167.

Khodapanah, L & Sulaiman, WNA, 2009, Groundwater Quality Assessment for Different Purposes in Eshtehard District, Tehran, Iran. *European Journal of Scientific Research,* 36: 543 - 553.

Lei, D, Yongzhang, Z, Jin, M, Yong, L, Qiuming, Ch, Shuyun, X, Haiyan, D, Yuanhang, Y & Hongfu, W, 2008, Using Multivariate Statistical and Geostatistical Methods to Identify Spatial Variability of Trace Elements in Agricultural Soils in Dongguan City. Guangdong, China. *Journal of China University of Geosciences,* 19: 343 - 353.

Li, J, Wang, Y, Xie, X & Su, Ch, 2012, Hierarchical cluster analysis of arsenic and fluoride enrichments in groundwater from the Datong basin, Northern China. *Journal of Geochemical Exploration,* 118: 77 - 89.

Lin, YP, Teng, TP & Chang, TK, 2002, Multivariate analysis of soil heavy metal pollution and landscape pattern in Changhua County in Taiwan. *Landscape and Urban Planning,* 62: 19 - 35.

Liu, X, Zhang, W, Zhang, M, Ficklin, DL & Wang, F, 2009, Spatiotemporal variations of soil nutrients influenced by an altered land tenure system in China. *Geoderma,* 152: 23 - 34.

Mantovi, P, Bonazzi, G & Maestri, E, 2003, Accumulation of Copper and Zinc from

Liquid Manure in Agricultural Soils and Crop Plants. *Plant and Soil,* 250: 249 - 257.

Marcovecchio, JE, Botte, SE, Freije, RH, 2007, *Heavy Metals, Major Metals, Trace Elements, in Handbook of Water Analysis.* L M Nollet, 2nd Ed. CRC Press, London, Page 483.

McGrath, D, Zhang, Ch & Carton, OT, 2004, Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. *Environmental Pollution,* 127: 239 – 248.

Mosaedi, A & Gharib, M, 2008, Investigation of flood characteristics in Gharahchy basin of Ramian. *Journal of Agricultural Science and Natural Resource,* 14: 203 - 214.

Nas, B & Berkta, A, 2006, Groundwater contamination by nitrates in the city of Konya, (Turkey): A GIS perspective. *Journal of Environmental Management,* 79: 30 - 37.

Nas, B, 2009, Geostatistical approach to assessment of spatial distribution of groundwater quality, *Polish Journal of Environmental Studies,* 18: 1073 - 1082.

Nkansah, K, Dawson-Andoh, B & Slahor, J, 2010, Rapid characterization of biomass using near infrared spectroscopy coupled with multivariate data analysis: Part 1 yellow-poplar (*Liriodendron tulipifera* L.). *Bioresource Technology,* 101: 4570 - 4576.

Ogunribido, THT, & Kehinde–Philips, OO, 2011, Multivariate statistical analysis for the assessment of hydrogeochemistry of groundwater in Agbabu area, S.W. Nigeria. In proceeding of Environmental Management Conference. Federal University of Agriculture: Abeokuta, Nigeria, pp 424-433.

Pang, Su, Li, TX, Wang, YD, Yu, HY & Li, X, 2009, Spatial Interpolation and Sample Size Optimization for Soil Copper (Cu) Investigation in Cropland Soil at County Scale Using Cokriging. *Agricultural Sciences in China,* 8: 1369 - 1377.

Sahoo, S & Jha, MK, 2014, Analysis of spatial variation of groundwater depths using geostatistical modeling. *International Journal of Applied Engineering Research,* 9: 317 - 322.

Sarukkalige, R, 2012, Geostatistical Analysis of Groundwater Quality in Western Australia, IRACST, *Engineering Science and Technology,* 2: 790 - 794.

Shrestha, S & Kazama, F, 2007, Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling and Software,* 22: 464 - 475.

Smyth, JD, Istok, JD, 1989, Multivariate geostatistical analysis of groundwater contamination by pesticide and nitrate. *Water Resources Research Institute.* Oregon State University, Corvallis- Oregon, Page 1-48.

Subramani, T, Krishnan, S & Kumaresan, PK, 2012, Study of Groundwater Quality with GIS Application for Coonoor Taluk in Nilgiri District. *International Journal of Modern Engineering Research,* 2: 586 - 592.

Tabachnick, BG, & Fidell, LS, 2007, *Using Multivariate Statistics*: Allyn and Bacon, London, pp 56-110.

Taghizadeh Mehrjardi, R, Zareian Jahromi, M, Mahmodi, Sh & Heidari, A, 2008, Spatial Distribution of Groundwater Quality with Geostatistics (Case Study: Yazd-Ardakan Plain). *World Applied Sciences Journal,* 4: 9 - 17.

Twomey, JM, & Smith, AE, 1996, *Validation and Verification Chapter 4 in Artificial Neural Networks for Civil Engineers:* ASCE Press, pp 1-29.

Uyan, M, & Cay, T, 2010, Geostatistical methods for mapping groundwater nitrate concentrations. In proceeding of the 3rd International Conference on Cartography and GIS: 15-20 June, Nessebar, Bulgaria, pp 68-85.

Willmott, CJ & Matsuura, K, 2006, on the use of dimensioned measures of error to evaluate the Performance of spatial interpolators. *International Journal of Geographical Information Science,* 20: 89 - 102.

Yang, M, Liu, SL, Yang, ZF, Sun, T & Beazley, R, 2009, Multivariate and geostatistical analysis of wetland soil salinity in nested areas of the Yellow River Delta. *Australian Journal of Soil Research,* 47: 486 - 497.

Yeh, MSh, Lin, YP & Chang, L, 2006, Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms. *Environmental Geology,* 50: 101 - 121.

Yu, WH, Harvey, CM & Harvey, CF, 2003, Arsenic in groundwater in Bangladesh: A geostatistical and epidemiological framework for evaluating health effects and potential remedies. *Water Resources Research,* 39: 1146-1163.

Zamani, AA, Yaftian, MR & Parizanganeh, AH, 2012, Multivariate statistical assessment of heavy metal pollution sources of ground water around a lead and zinc plant. *Journal of Environmental Health Science and Engineering,* 9: 1-10.

Zheng, YM, Chen, TB & He, JZ, 2008, Multivariate Geostatistical Analysis of Heavy Metal in Top soils from Beijing. *China Journal of Soils & Sédiments,* 8: 51 - 58.

کاربرد آمار چند متغیره و روش‌های زمین آمار در تعیین توزیع فضایی فلزهای سنگین در منابع آب زیر زمینی

ف. خاندوزی، ع. پری زنگنه* ، ع. زمانی

گروه علوم محیط زیست، دانشکده علوم، دانشگاه زنجان، زنجان، ایران

(تاریخ دریافت: ۹۳/۱۲/۱۰   تاریخ پذیرش: ۹۴/۴/۳۱)

چکیده

این تحقیق با هدف تعیین کارآیی روش‌های درونیابی برای تهیه‌ی نقشه پهنه‌بندی کیفیت آب زیرزمینی و هم‌چنین تخمین روش‌های آمار چند متغیره در تحلیل توزیع فضایی فلزهای سنگین در آب‌های زیرزمینی شهرستان رامیان (گلستان-ایران) انجام شده است. فلزهای سنگین (نیکل، کبالت، سرب، کادمیم ، مس) در ۲۴ نمونه‌ی آب چشمه/ چاه اندازه‌گیری شد. روش‌های آماری چند متغیره و زمین آمار برای متمایز نمودن منابع آلاینده انسان‌زاد و طبیعی به فلزهای سنگین در منطقه‌ی مورد مطالعه استفاده شد. نتایج تحلیل خوشه‌ای و فاکتوری نشان می‌دهد که نیکل و کبالت در فاکتور اول، سرب و مس در فاکتور دوم و روی و مس نیز در فاکتور سوم قرار دارند. احتمال حضور سطح‌های تخمین زده شده برای سه عامل با استفاده از مدل واریاگرام پیش‌بینی شد، و عملکرد روش پیش‌بینی شده با استفاده از خطای متوسط مطلق، متوسط انحراف خطا و ریشه میانگین مجذور بررسی شد. نتیجه ساختار مکانی نشان داد که روش واریوگرام و اعتبار متقابل شش متغیر با استفاده از سه روش به نام‌های کسر پایه شعاعی، معکوس وزن فاصله و کریجینگ معمولی مدل می‌شود. با این وجود نتایج نشان داد که روش کسر پایه شعاعی بهترین مدل با بالاترین دقت و کم‌ترین خطا بود. هم‌چنین سیستم اطلاعات جغرافیایی به طور کامل می‌تواند در تحلیل فضایی مقدار فلزهای سنگین، در آب‌های زیرزمینی منطقه مورد مطالعه استفاده شود.

* مولف مسئول