

# Estimation of spatial distribution of heavy metals in groundwater using interpolation methods and multivariate statistical techniques; its suitability for drinking and irrigation purposes in the Middle Black Sea Region of Turkey

Hakan Arslan · Nazlı Ayyıldız Turan

Received: 20 November 2014 / Accepted: 7 July 2015 / Published online: 24 July 2015  
© Springer International Publishing Switzerland 2015

**Abstract** Monitoring of heavy metal concentrations in groundwater potentially used for drinking and irrigation is very important. This study collected groundwater samples from 78 wells in July 2012 and analyzed them for 17 heavy metals (Pb, Zn, Cr, Mn, Fe, Cu, Cd, Co, Ni, Al, As, Mo, Se, B, Ti, V, Ba). Spatial distributions of these elements were identified using three different interpolation methods [inverse distance weighing (IDW), radial basis function (RBF), and ordinary kriging (OK)]. Root mean squared error (RMSE) and mean absolute error (MAE) for cross validation were used to select the best interpolation methods for each parameter. Multivariate statistical analysis [cluster analysis (CA) and factor analysis (FA)] were used to identify similarities among sampling sites and the contribution of variables to groundwater pollution. Fe and Mn levels exceeded World Health Organization (WHO) recommended limits for drinking water in almost all of the study area, and some locations had Fe and Mn levels that exceeded Food and Agriculture Organization (FAO) guidelines for drip irrigation systems. Al, As, and Cd levels also exceeded WHO guidelines for drinking water. Cluster analysis classified groundwater in the

study area into three groups, and factor analysis identified five factors that explained 73.39 % of the total variation in groundwater, which are as follows: factor 1: Se, Ti, Cr, Mo; factor 2: Ni, Mn, Co, Ba; factor 3: Pb, Cd; factor 4: B, V, Fe, Cu; and factor 5: AS, Zn. As a result of this study, it could be said that interpolation methods and multivariate statistical techniques gave very useful results for the determination of the source.

**Keywords** Interpolation methods · Spatial variation · Heavy metal pollution · Factor analysis · Drinking water

## Introduction

Heavy metal pollution of groundwater represents an important environmental problem. Some heavy metals that are essential for human health may cause various health problems if present in concentrations that exceed permissible limits. In order to determine the suitability of groundwater for drinking and agricultural and industrial use, heavy metal levels must be evaluated. Heavy metals as well as pesticides may cause problems due to their generally non-biodegradable nature (Eugenia et al. 1996). Numerous studies have been conducted examining heavy metal contamination of groundwater and surface water in different parts of the world (Alam and Umar 2013; Kanmani and Gandhimathi 2013; Madhulakshmi et al. 2012; Tiwari et al. 2013). For example, Haloi and Sarma (2012) examined groundwater in the Barpeta District of Assam, India, and found Fe, Mn, and Pb concentrations to be above permissible

---

H. Arslan (✉)  
Department of Agricultural Structures and Irrigation, Faculty  
of Agriculture, Ondokuz Mayıs University, 55139 Samsun,  
Turkey  
e-mail: harslan55@yahoo.com.tr

N. Ayyıldız Turan  
Seventh Regional Directorate of State Hydraulic Works (DSI),  
55070 Samsun, Turkey  
e-mail: nazaytu@dsi.gov.tr

limits for drinking water. Kumari et al. (2014) evaluated groundwater quality in industrial areas of Ghaziabad, India, and found 12 elements to exceed prescribed standards. Similarly, Yılmaz et al. (2010) found heavy metal contamination of groundwater in Turkey posed a significant risk for drinking as well as irrigation.

Interpolation methods used to estimate unknown values from data observed at known locations can successfully map the spatial distribution of various hydrochemical parameters (Quyung et al. 2013; Shan et al. 2013). Techniques such as inverse distance weighting (IDW), ordinary kriging (OK), and radial basis functions (RBF) have been extensively used in groundwater quality and pollution studies (Rabah et al. 2011). Yao et al. (2014) used eight different interpolation methods to investigate the spatial distribution of groundwater levels and found kriging methods to produce the best models. Noori et al. (2013) used several spatial interpolation methods (kriging, cokriging, RBF, IDW) to map the spatial distribution of groundwater levels in Iran, evaluating the performance of each method by RMSE, MAE and  $R^2$ . Hua et al. (2009) compared IDW and OK estimates of groundwater depths in the Shule River Basin, China, and concluded that OK was more efficient than IDW based on the lower RMSE values of the former.

Multivariate techniques are very useful for identifying relationships between variables. Methods such as cluster analysis (CA) and factor analysis (FA) have been used increasingly in studies conducted to measure and monitor groundwater. Whereas CA can be used to illustrate the overall similarity of variables in a data set and link similar groups (Massart and Kaufmann 1983), FA can identify similarities between factors as well as the most important factors contributing to data structure. CA and FA have been widely used to study heavy metals in groundwater (Brahman et al. 2013; Monjerezi et al. 2012). In their study evaluating groundwater quality in coastal areas in China using FA, Huang et al. (2013) identified four factors affecting groundwater quality, namely seawater intrusion and As contamination; water-rock interaction, surface water recharge, and acidic precipitation; heavy metal pollution from industry; and agricultural pollution and sewage intrusion.

Several recent studies have combined the use of multivariate analysis and GIS techniques in order to evaluate groundwater (Lu et al. 2012; Arslan 2013). In addition to measuring heavy metal concentrations in the Çarşamba

Plain in northern Turkey, the present study evaluated three interpolation methods and used the optimal method to map the spatial distribution of each heavy metal in the study area. Then, using multivariate statistical analysis, it classified the study area into groups according to groundwater pollution by heavy metals and identified the genetic origin of these pollutants.

## Materials and methods

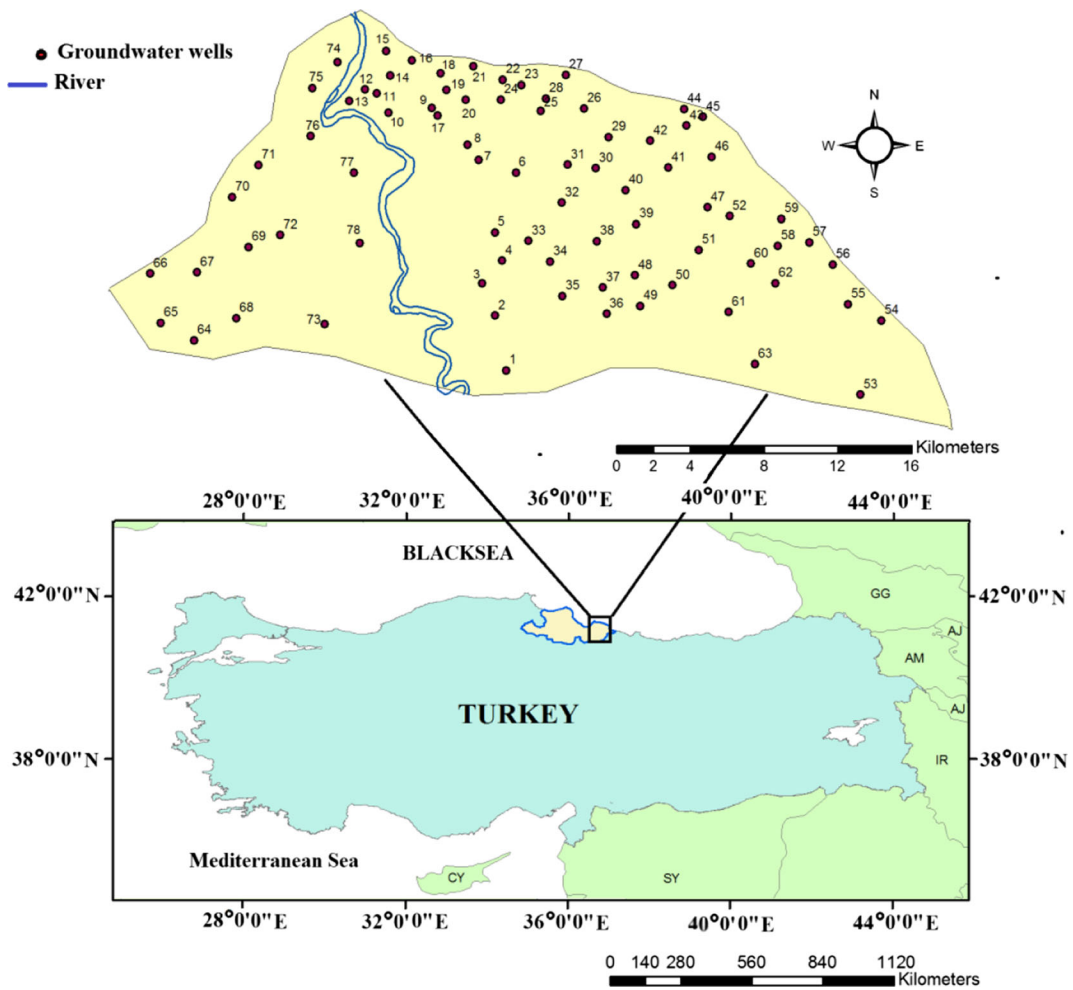
### Study area

The study was conducted in the Middle Black Sea Region of Northern Turkey in Çarşamba Plain, one of Turkey's largest plains ( $41^{\circ} 11' - 41^{\circ} 23' E$ ,  $36^{\circ} 30' - 37^{\circ} 00' N$ ) (Fig. 1). Elevation of the research site ranges from 0.0–18.0 m above sea level, with slope gradients of approximately 1 %. The climate is semi-humid, with an average annual precipitation of about 700 mm and an average annual temperature of 17 °C. Cropping patterns in the study area vary considerably. The dominant crop was hazelnut. Wheat, tomatoes, and rice are grown in the irrigation season, whereas cabbage and leek are grown in rainy seasons, and corn is grown as both a primary and secondary crop. There are several industries, including copper and fertilizer factories over the study area.

Çarşamba Plain was formed by the alluviums brought by Yeşilirmak River. The aquifer system lies over the Quaternary deposits under unconfined conditions. The unconfined aquifer is within a wedge of unconsolidated gravels and sands that thicken towards the coast (Fig. 2). The fill of Çarşamba Plain consists of detrital Quaternary sediments that contain sand, silt, and clay that have little vertical or lateral continuity and frequent lateral changes in facies with 20–110 m thick and increase from south to north. The aquifer of Çarşamba Plain is composed of alluvium deposit and eosin-aged volcanic rock units. The aquifer has available discharges of between 2 and 32 l/s. The aquifer has varying hydraulic gradient and the highest gradient is 0.0025. Groundwater depths ranged from about 3 to 10 m below ground surface (DSI 1994).

### Groundwater sampling and analysis

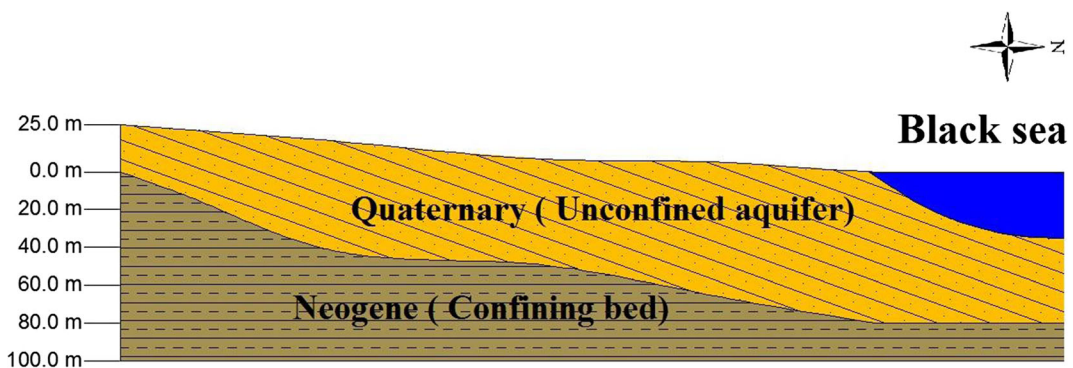
Groundwater samples were collected from 78 water wells in July 2012. Sampled wells were homogeneously



**Fig. 1** Study area and sample locations

distributed over the study area as to represent the entire site. Well locations are shown in Fig. 1. Each sampling location was recorded with a global positioning system (GPS). Pumps were operated for 15 min. Prior to sample collection. Water samples

were analyzed for concentrations of 17 heavy metals [lead (Pb), zinc (Zn), chromium (Cr), manganese (Mn), iron (Fe), copper (Cu), cadmium (Cd), cobalt (Co), nickel (Ni), aluminum (Al), arsenic (As), molybdenum (Mo), selenium (Se), boron (B), titanium



**Fig. 2** Geological units of Çarşamba Plain aquifer (modified from DSI 1994)

(Ti), vanadium (V), barium (Ba)] using an Agilent 7500a inductively coupled plasma mass spectrometry (ICP-MS) at the General Directorate of Hydraulic Works (DSI), Department of Technical Research and Quality Control.

#### Interpolation methods for heavy metals

Spatial interpolation refers to the estimation of values of a particular attribute at unsampled locations using existing information from other observation points. By converting data from observation points to continuous fields, the spatial patterns of sampled measurements can be compared with the spatial patterns of other entities. The three most widely used interpolation methods are inverse distance weighting (IDW), radial basis functions (RBF), and ordinary kriging (OK) (Sun et al. 2009). IDW estimates values at unsampled points by using a linear combination of values at sampled points weighted by an inverse function of the distance from the point of interest to the sampled points using the following formula:

$$Z = \left[ \frac{\sum_{i=1}^n (Z_i/d_i^m)}{\sum_{i=1}^n (1/d_i^m)} \right] \quad (1)$$

where  $Z$  is the estimated value,  $Z_i$  is the measured sample value at point  $i$ ,  $d_i$  is the distance between  $Z$  and  $Z_i$ , and  $m$  is the weighting power that defines the rate at which weights fall off with  $d_i$ , with a typical  $m$  value of 1–5 (Keshavarzi and Sarmadian 2012). RBF comprises a series of interpolation methods in which the estimated surface must pass through all measured sample values. There are five different basis functions: thin-plate spline (TPS), spline with tension (SPT), completely regularized spline (CRS), multiquadric function (MQ), and inverse multi-quadric function (IMQ) (Xie et al. 2011). Ordinary kriging, the most common interpolation method used in geostatistical studies (Guler et al. 2014), resembles IDW in that it uses a linear combination of weights at known points to estimate the value at an unknown point. Other types of kriging include simple kriging (SK), universal kriging (UK), and indicator kriging (IK). OK uses a linear combination of measured values whose

weights are determined based on their spatial correlation to produce estimated values using the following formula:

$$Z = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

Where  $Z$  is the estimated value,  $Z(x_i)$  is the measured value at  $x_i$ ,  $\lambda_i$  is the weight assigned to the residual of  $Z(x_i)$ , and  $n$  is the number of the data used at known locations in a neighborhood. This study compared the results of IDW (using weighting powers of 1, 2, and 3), RBF (using five different basis functions), and OK interpolation in estimating heavy metal concentrations and then used the best method to evaluate the spatial distribution of each heavy metal in the study area. Data was evaluated using the software ARCGIS 10.0 with Geostatistical Analyst Extensions.

This study compared the results of IDW (using weighting powers of 1, 2, and 3), RBF (using five different basis functions), and OK interpolation in estimating heavy metal concentrations and factor scores of each factor and then used the best method to evaluate the spatial distribution of each heavy metal and factor scores in the study area. Data was evaluated using the software ARCGIS 10.0 with Geostatistical Analyst Extensions.

#### Cross validation

Cross-validation is performed to assess the best method of interpolation (Yao et al. 2014). Several techniques can be used to judge the relationships between observed and predicted values and determine the best method (Li and Heap 2011).

Root-mean-square error (RMSE) and mean absolute error (MAE) were used to evaluate the predictive performance of different techniques, with the smallest RMSE and MAE indicating the most accurate predictions. MAE and RMSE are calculated using the following formulas:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Z_i - \tilde{Z}| \quad (3)$$

and

$$RMSE = \sqrt{\frac{\sum (Z_i - Z)^2}{n}} \tag{4}$$

Where  $Z_i$  is the predicted value,  $Z$  is the observed value, and  $n$  is the number of observations.

### Multivariate statistical analysis

Multivariate statistical analysis is widely used to identify and evaluate surface water and groundwater data (Chidambaram et al. 2012; Narany et al. 2014; Vieira et al. 2012). Multivariate techniques make it possible to simplify, organize, and classify data to draw out useful meaning.

Cluster analysis is used to group data into hierarchies based on similarities or dissimilarities. Yadav et al. (2013) and Hossain et al. (2013) successfully used cluster analysis to classify groundwater samples according to similarities. In the present study, CA was applied to group groundwater samples for Pb, Zn, Cr, Mn, Fe, Cu, Cd, Co, Ni, Al, As, Mo, Se, B, Ti, V, and Ba content.

Factor analysis is used primarily to reduce the contribution of less significant variables in order to further simplify data structure. In a study evaluating groundwater in a region of Taiwan affected by Blackfoot disease, for example, Liu et al. (2003) defined factor loading as “strong,” “moderate,” and “weak,” corresponding to absolute values of >0.75, 0.50–0.75, and 0.30–0.50, respectively; factors with eigenvalues >1 explained more total variation in the data than individual groundwater quality variables, and factors with eigenvalue <1 explained less total variation than individual variables. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS), version 20.

## Results and discussion

### Heavy metals in groundwater

Descriptive statistics related to groundwater heavy metal concentrations are given in Table 1. Coefficient of variation (CV) was the most important factor in describing the variability of groundwater properties. Data was ranked according to amount of variation as “low

**Table 1** Descriptive statistics for heavy metal concentrations in groundwater in the study area

Parameter (mg/l)	Min	Max	Mean	CV	WHO guideline value	Skewness	Kurtosis	K-S p* value	Data distribution
Pb	0.0001	0.0093	0.0006	241.1	0.010	4.76	28.55	0.000	Non-normal
Zn	0.0010	0.1265	0.0142	113.3	5.000	4.83	32.41	0.000	Non-normal
Cr	0.0011	0.0121	0.0031	83.10	0.050	6.14	47.40	0.006	Non-normal
Mn	0.0044	2.4820	0.3396	119.7	0.100	2.58	12.19	0.000	Non-normal
Fe	0.1711	5.3790	0.9978	109.2	0.300	2.32	8.20	0.000	Non-normal
Cu	0.0010	0.0761	0.0080	132.6	2.000	4.04	24.31	0.000	Non-normal
Cd	0.0008	0.0073	0.0002	461.9	0.003	6.17	42.04	0.000	Non-normal
Co	0.0009	0.0088	0.0006	227.3	0.050	4.34	22.81	0.000	Non-normal
Ni	0.0003	0.0516	0.0042	152.4	0.070	5.60	41.46	0.000	Non-normal
Al	0.0010	0.6357	0.0770	166.7	0.200	2.63	9.40	0.000	Non-normal
As	0.0001	0.0634	0.0079	182.5	0.010	2.54	8.45	0.000	Non-normal
Mo	0.0002	0.0051	0.0004	224.3	0.070	3.29	14.20	0.000	Non-normal
Se	0.0002	0.0091	0.0008	161.4	0.040	8.41	73.14	0.000	Non-normal
B	0.0100	0.9700	0.1694	103.8	2.40	2.30	8.68	0.000	Non-normal
Ti	0.0002	0.0040	0.0009	92.60	–	8.60	75.26	0.004	Non-normal
V	0.0003	0.0066	0.0020	64.00	–	1.65	5.66	0.001	Non-normal
Ba	0.0127	0.3465	0.1139	76.80	0.700	0.72	2.64	0.004	Non-normal

\* $p < 0.05$

variability” (CV,  $\leq 15\%$ ), “moderate variability” (CV, 15–35 %), or “high variability” (CV,  $>35\%$ ) (Wilding 1985). All heavy metals in the study area were found to exhibit high variability.

Groundwater heavy metal constituents were examined in relation to WHO standards for drinking water (2011). Concentrations of zinc (Zn), chromium (Cr), copper (Cu), cobalt (Co), molybdenum (Mo), vanadium (V), and titanium (Ti) and selenium (Se) in the study area were far below the WHO limits for drinking water, nickel (Ni), boron (B), and barium (Ba) concentrations in all samples were also below the established limits. Lead (Pb) concentrations ranged between 0.00–0.0093 mg/l (mean: 0.0013 mg/l), which are also below WHO limits for drinking water, although some samples are close to these limits. Manganese (Mn) concentrations ranged between 0.0044–2.4820 mg/l (mean, 0.3396 mg/l), with many sites exceeding WHO limits for drinking water. Iron content (Fe) varied considerably, from 0.1711 to 5.3790 mg/l (mean, 0.9978), with many samples above the WHO limit of 0.3 mg/l. Similarly, Al content measured at some wells greatly exceeded the WHO recommended limit of 0.2 mg/l for drinking water. Arsenic (As) and cadmium (Cd) concentrations also exceeded guidelines in some areas.

## Comparison of interpolation methods

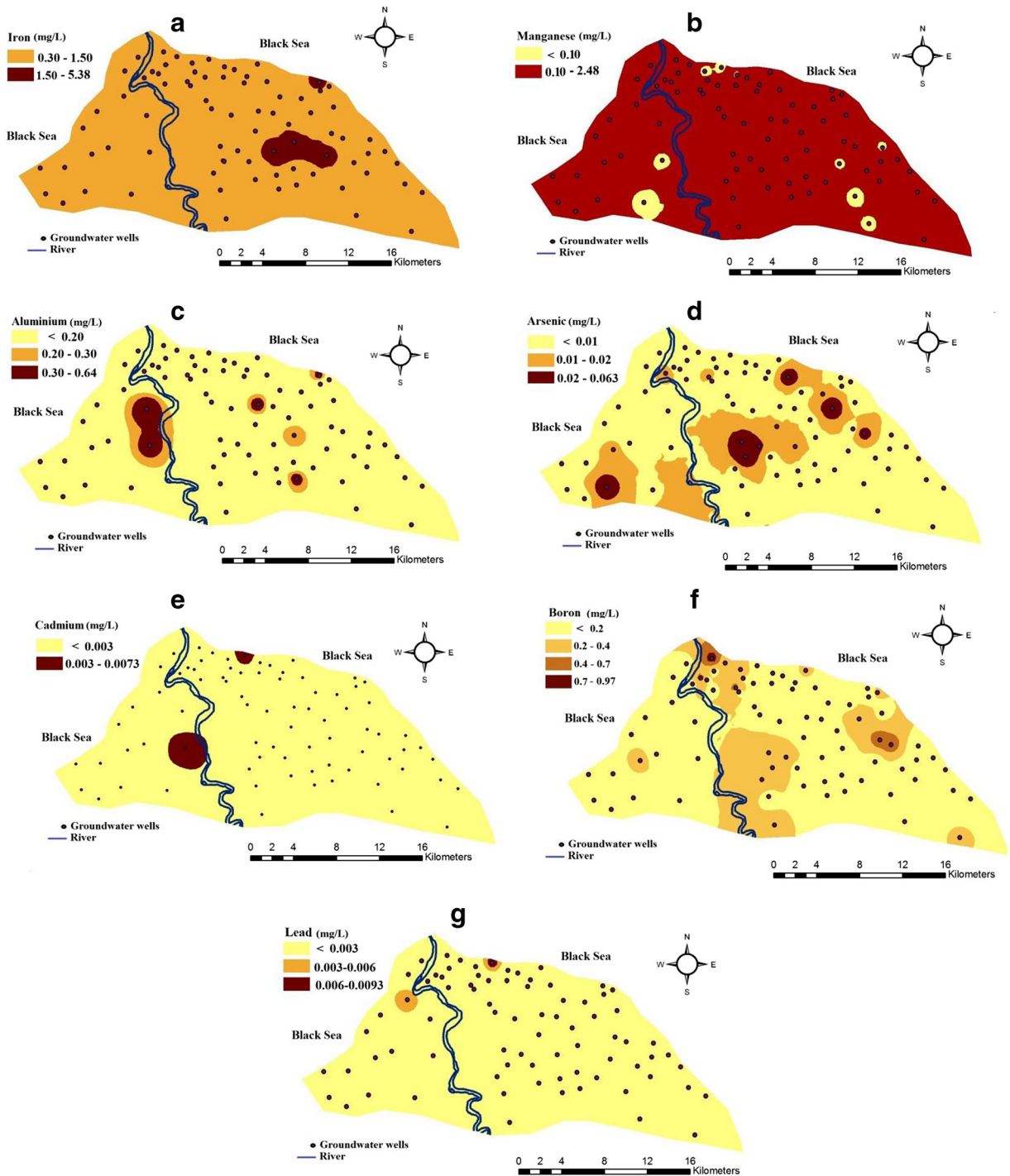
Maps were prepared only for those heavy metals whose concentrations were close to or above the recommended limits for drinking and irrigation water (Fe, Mn, Al, As, Cd, B, Pb). The spatial variability of heavy metals was assessed using inverse distance weighting (IDW) raised to powers of 1, 2, and 3; radial basis functions (CRS, ST, MQ, IMQ, and TPS) and ordinary kriging. Kriging methods work best with data with a normal distribution (Xie et al. 2011). However, a Kolmogorov-Smirnov test showed the data for heavy metals concentrations in groundwater in the study area were not normally distributed ( $P < 0.05$ ); therefore, values were log-transformed prior to calculation of semi variance.

The RMSE and MAE were used to compare the performance of interpolation methods. The method that yields the smallest value of RMSE and MAE is the best (Table 2). However, if MAE is not at the lowest value when RMSE is at the lowest, the most accurate method is the one which has the lowest RMSE value. In addition, MAE values are considered to determine the best method when RMSE values are equal.

IDW-1 had the lowest RMSE values for Mn, Al, and B, whereas Pb was best estimated using IDW-2, Cd using IDW-3, As using RBF-IMQ, and Fe using OK.

**Table 2** Accuracy of methods used to interpolate groundwater heavy metal content

Heavy metal	Cross validation	IDW-1	IDW-2	IDW-3	RBF-CRS	RBF-ST	RBF-MQ	RBF-IMQ	RBF-TPS	OK
Fe	RMSE	1.170	1.210	1.260	1.210	1.190	1.330	1.130	1.520	1.15
	MAE	0.840	0.860	0.880	0.840	0.840	0.960	0.816	1.170	0.740
Mn	RMSE	0.431	0.448	0.486	0.442	0.438	0.491	0.432	0.575	0.465
	MAE	0.292	0.299	0.320	0.295	0.294	0.322	0.292	0.387	0.354
Al	RMSE	0.131	0.134	0.138	0.133	0.132	0.141	0.131	0.156	0.135
	MAE	0.086	0.086	0.090	0.086	0.086	0.092	0.087	0.106	0.094
As	RMSE	0.014	0.015	0.015	0.015	0.015	0.016	0.014	0.018	0.015
	MAE	0.009	0.009	0.009	0.009	0.009	0.010	0.008	0.012	0.009
Cd	RMSE	0.00100	0.00100	0.00100	0.00102	0.00102	0.00107	0.00101	0.00113	0.00103
	MAE	0.00340	0.00320	0.00300	0.00360	0.00360	0.00380	0.00370	0.00440	0.00380
B	RMSE	0.000170	0.000180	0.000190	0.000180	0.000190	0.000200	0.000180	0.000240	0.000170
	MAE	0.000125	0.000127	0.000129	0.000126	0.000125	0.000139	0.000124	0.000167	0.000127
Pb	RMSE	0.001440	0.001430	0.001450	0.001470	0.001460	0.001600	0.001430	0.001800	0.001450
	MAE	0.000770	0.000770	0.000790	0.000820	0.000810	0.000920	0.000780	0.001130	0.000790



**Fig. 3** Interpolated maps for heavy metals **a** Fe using OK, **b** Mn using IDW-1, **c** Al using IDW-1, **d** As using RBF-IMQ, **e** Cd using IDW-3, **f** B using IDW-1, **g** Pb using IDW-2

Xie et al. (2011) found that the greater the weighting power of IDW, the greater the RMSE of interpolation. As Table 2 shows, RBF-TPS had the highest RMSE and

MAE values for all heavy metals and should be considered unsuitable for interpolating heavy metal groundwater concentrations.

### Spatial distribution of heavy metals

Best-fit methods (Fig. 3a–g) were used to interpolate the spatial distribution of heavy metals. Problematic areas for Fe, Mn, Al, As, Cd, B, and Pb were mapped using geographic information systems (GIS). Other recent studies have also used GIS in identifying spatial and temporal distribution of groundwater and soil properties (Sun et al. 2009; Varouchakis and Hristopulas 2013).

Using OK, Fe was mapped into three categories for drinking and irrigation (Fig. 3a). Fe concentrations were above 0.17 mg/l throughout the study area (Table 1), with the highest concentrations mainly in the eastern part of the area. Normally, iron concentrations in drinking water must be below 0.3 mg/l; however, as Fig. 2a indicates, only 1 % of the study area had Fe concentrations below 0.3 mg/l, whereas the remaining 99 % should be considered unsuitable for drinking due to high Fe concentrations. In terms of irrigation, the recommended maximum concentration of soluble iron in irrigation water is 5 mg/l, whereas concentrations above 1.5 mg/l are considered to represent a severe clogging hazard for drip irrigation emitters, and concentrations between 0.1 and 1.5 mg/l represent a moderate clogging

hazard (FAO 1994). Fe concentrations ranged from 0.17 to 1.5 mg/l in close to 88 % of the study area, representing a moderate hazard for drip irrigation, whereas close to 12 % of the study area had a Fe value above 1.5 mg/l, representing a severe hazard. These areas thus require another type of irrigation system, such as sprinkler or furrow irrigation. Finally, the iron concentration of one well (site 38) was above 5 mg/l, making it unsuitable for any irrigation use.

Using IDW-1, Mn was mapped into two categories for drinking and irrigation (Fig. 3b). Permissible limits for Mn in drinking water and irrigation water are normally less than 0.1 and 2 mg/l, respectively. Accordingly, water in 97 % of the study area was found unsafe for drinking, whereas only one well (site 7) was found unsafe for irrigation. However, while Mn concentrations in most of the study area were below the level of plant toxicity (2 mg/l), groundwater in 97 % of the area had Mn concentrations above 0.1 mg/l, making it a moderate hazard for drip irrigation systems (FAO 1994).

Using IDW-1, Al was mapped into three categories for drinking (Fig. 3c). As the map shows, groundwater in a portion of the center of the study area amounting to 7 % of the total is unsuitable for drinking.

**Table 3** Rotated factor pattern of five factors after varimax rotation

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Se	<i>0.981</i>	0.000	0.049	-0.078	-0.066
Ti	<i>0.967</i>	-0.028	0.044	-0.155	-0.049
Cr	<i>0.950</i>	0.113	0.017	0.139	0.104
Mo	<i>0.563</i>	0.047	0.277	0.252	-0.099
Ni	0.048	<i>0.874</i>	0.253	0.08	-0.051
Mn	-0.128	<i>0.831</i>	-0.100	-0.136	-0.090
Co	0.328	<i>0.642</i>	0.640	0.071	-0.028
Ba	0.118	<i>0.623</i>	-0.337	-0.07	-0.095
Pb	-0.013	-0.104	<i>0.901</i>	-0.086	0.020
Cd	0.518	0.033	<i>0.751</i>	0.024	0.018
B	0.224	0.146	-0.225	<i>0.752</i>	0.466
Va	-0.036	-0.012	0.411	<i>0.734</i>	0.317
Fe	-0.099	0.333	0.057	-0.592	0.407
Cu	0.489	-0.046	0.153	<i>0.514</i>	0.228
Al	0.299	0.049	0.262	-0.499	0.133
As	0.002	-0.119	-0.103	0.025	<i>0.803</i>
Zn	-0.026	-0.085	0.125	0.100	<i>0.585</i>
Eigenvalue	4.510	2.640	2.100	1.930	1.300
% total variance	26.55	15.53	12.36	11.32	7.630
Cumulative % variance	26.55	42.08	54.44	65.76	73.39

Italic values indicate strong (<0.75) and moderate (0.50–0.75) loadings, respectively



Using RBF-IMQ, As was mapped into three categories for drinking (Fig. 3d). Different parts of the study area amounting to 26 % of the total were found to exceed the WHO limit of 0.01 mg/l for drinking water.

Using IDW-3, Cd was mapped into two categories for drinking (Fig. 3e). For the majority of the area, cadmium does not represent a problem; however, 2 % of the study area registered levels above the 0.001 mg/l WHO guideline.

Using IDW-1, B was mapped into four categories for drinking and irrigation (Fig. 3f). Boron is essential for drinking and irrigation water. However, the concentration of boron in irrigation water should not exceed 0.7 mg/l, with concentrations ranging from 0.7 to 3.0 mg/l considered to represent a severe to moderate

hazard for irrigation (FAO 1994). Excessive boron concentrations were observed in only 1 well (site 15) in the northern part of the study area.

Using IDW-2, Pb was mapped into three categories for drinking (Fig. 3g). Pb concentrations in all samples were below the permissible limits, with the highest Pb value found at site 21.

Factor analysis

Correlations among hydrochemical constituents of the groundwater samples were examined by factor analysis (Li and Zhang 2010; Guler et al. 2013) using each of the heavy metals tested as variables. Five factors were found to explain 73.39 % of the total variance in the

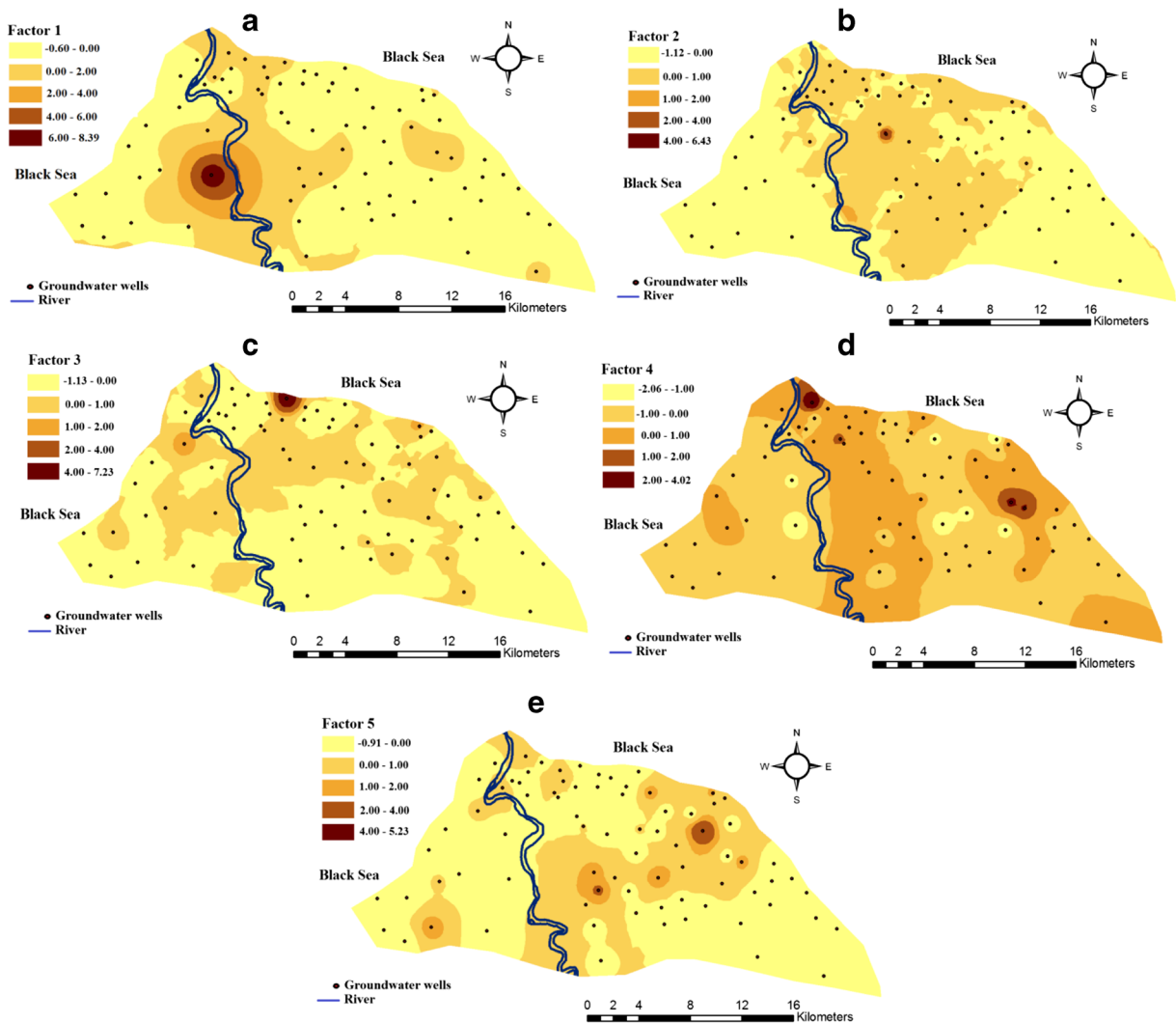


Fig. 4 a–e Distribution of scores for five factors

data set. Eigenvalues, percentages of variance, and cumulative percentages for the five factors identified are given in Table 3.

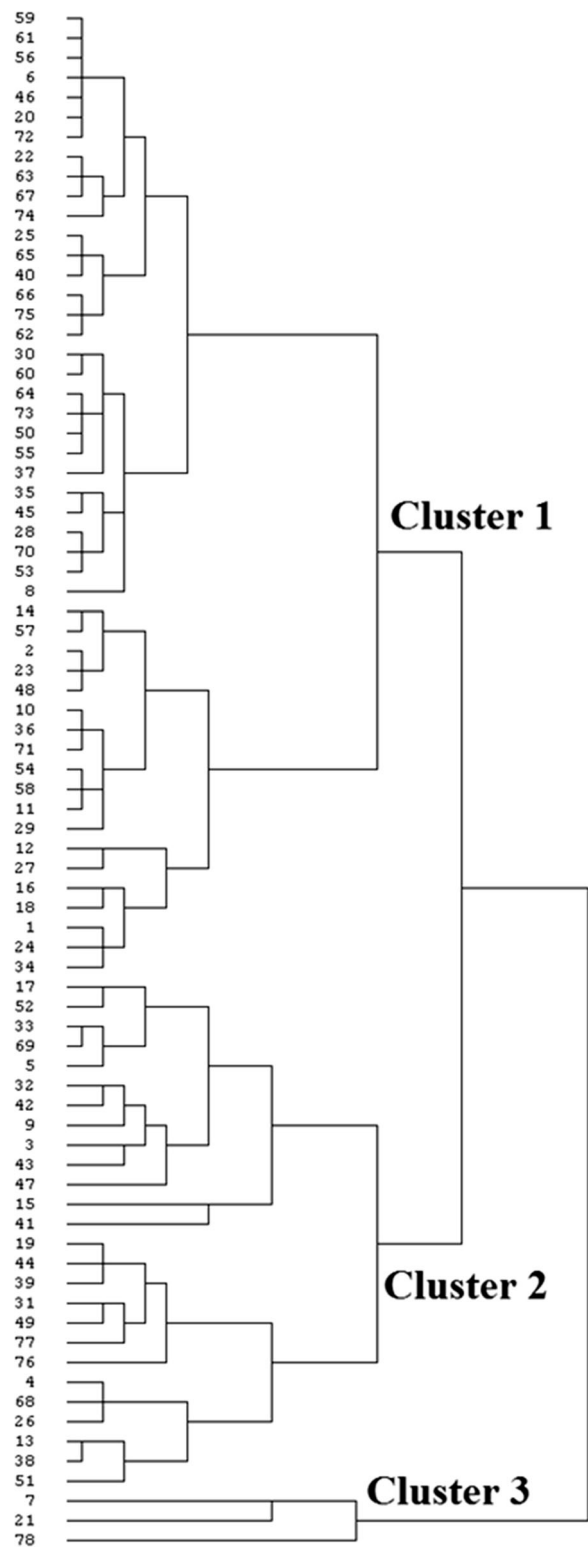
Three different weighing powers of IDW method, five different functions of RBF method and ordinary kriging method were tested. IDW-1 had the lowest RMSE and MAE values for factor 1, factor 3, and factor 5, whereas factor 2 and factor 4 were best estimated using IDW-2. Best-fit methods were used to interpolate the spatial distribution of factor scores. Spatial distribution of the scores for factors 1, 2, 3, 4, and 5 are given in Figs. 4a–e, respectively, with lighter colors indicating lower values and darker colors indicating higher values. Heavy metals in groundwater originate from a variety of sources. Cu, Mo, and Cd are common components of pesticides, insecticides, herbicides, and fertilizers used in the study area (Kumbur et al. 2008). Whereas Cd originates primarily from anthropogenic sources (e.g., agriculture or industry); As and Pb may originate from urban and Industrial activities such as energy production, mining, manufacturing processes, and waste incineration; and Ba, Fe, Co, Mn, Ni, and Cr may originate from pedogenic sources (Li and Zhang 2010; Demirel 2007).

Factor 1 explained 26.55 % of the total variance (eigenvalue, 4.51) and included strong positive loadings on Se, Ti, and Cr and moderate loading on Mo. This factor was ascribed to predominantly anthropogenic and industrial sources. Cr is a marker of paint and metal industrial waste, and high Mo concentrations may also be related to industrial activities (Demirel 2007). As Fig. 4a shows, factor 1 is generally the highest in the middle of the region.

Factor 2 explained 15.53 % of total variance (eigenvalue, 2.64), with strong loadings on Ni and Mn and moderate loadings on Co and Ba. Hence, this factor was attributed to mixed origins, including pedogenic (Mn, Ni, Co) and geogenic (Ba) sources (Krisha et al. 2009). As Fig. 4b shows, the highest score for factor 2 was at site 7.

Factor 3 accounted for 12.36 % of the total variance (eigenvalue, 2.10), with strong loadings on Pb and Cd. Pb is mainly an indicator of agrochemical and industrial waste (Li et al. 2008). This factor can be interpreted as relating to variations in agrochemical sources. As Fig. 4c shows, the highest score for factor 3 was at site 21.

Factor 4 explained 11.32 % of the total variance (eigenvalue, 1.93), with a strong loading on B, moderate



**Fig. 5** Dendrogram showing the clustering of heavy metals parameters of groundwater

**Table 4.** Mean values (mg/l) for heavy metals for each group

Heavy metal	Cluster 1 (n = 49)	Cluster 2 (n = 26)	Cluster 3 (n = 3)
Pb	0.00030	0.00071	0.00355
Zn	0.01135	0.01941	0.01522
Cr	0.00237	0.00410	0.01919
Mn	0.34463	0.27067	0.85415
Fe	0.72137	1.46012	1.50683
Cu	0.00544	0.01109	0.02179
Cd	0.00001	0.00015	0.00398
Co	0.00028	0.00057	0.00721
Ni	0.00317	0.00387	0.02276
Al	0.03645	0.14149	0.17804
As	0.00228	0.01917	0.00137
Mo	0.00015	0.00070	0.00249
Se	0.00048	0.00119	0.02888
B	0.00013	0.00026	0.00012
Ti	0.00063	0.00143	0.03770
V	0.00173	0.00253	0.00292
Ba	0.12270	0.09426	0.13960

loadings on Va and Cu, and negative moderate loading on Fe (Fig. 4d). Huang and Jin (2008) showed that long-term use of chemical fertilizers contributes to the accumulation of Cu in agricultural soil. This factor was interpreted as relating to agricultural run-off and industrial effluents. As Fig. 4d shows, the highest score for factor 4 was at site 15.

Factor 5 accounted for 7.63 % of total variance (eigenvalue, 1.30), with a strong positive loading on As and moderate loading on Zn. This factor was interpreted as influenced by industrial effluents. As

Fig. 4e shows, excessive arsenic represented the greatest problem in the northeastern part of the study area (sites 4, 38, and 41) and the smallest problem in the south, west, and southeastern regions (sites 28, 56, and 34).

Cluster analysis

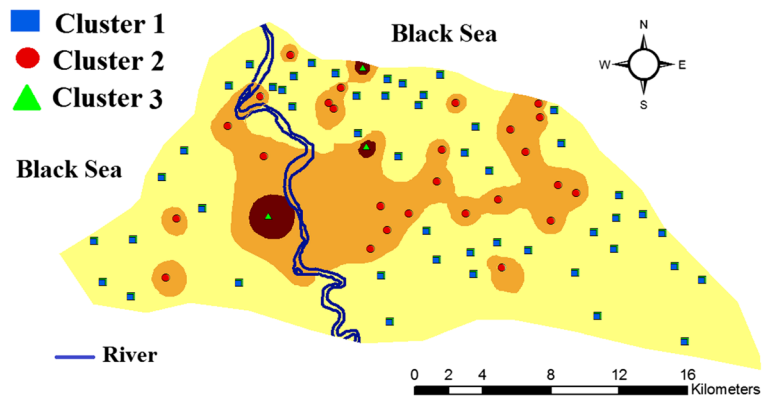
Hydrochemical data was classified by cluster analysis (CA) into 17 dimensional spaces represented by the dendrogram shown in Fig. 5, and groundwater wells were distributed into three groups according to significant differences in groundwater heavy metal concentrations (Table 4). When evaluated in light of existing drinking water guidelines, clusters 1, 2, and 3 could be classified, respectively, as “polluted,” “highly polluted,” and “very highly polluted.” IDW, RBF, and OK methods were separately tested to create the spatial distribution of groups, and IDW-2 yielded the lowest RMSE and MAE values. Therefore, spatial distribution map was created according to IDW-2.

Cluster 1 comprised the 49 monitoring wells with the lowest loading score for Fe and moderate loading for Mn. In general, cluster 1 had the lowest values for groundwater heavy metals in the study area and included wells located on the outer limits of the area (Fig. 6).

Cluster 2 comprised the 26 wells located in the central part of the Çarşamba Plain that had moderate loading for Fe and Mn and high loading for As.

Cluster 3 comprised the three wells (7, 21, and 78) with the worst groundwater quality in the study area. Located in different parts of the plain, these wells had

**Fig. 6** Spatial distribution of groups formed by cluster analysis



the highest loadings for Fe, Mn, Cd, and Al. The reason for this is thought to be agricultural and industrial activities. Thus, it would be useful to control the pesticide and fertilizer usage in these areas.

## Conclusion

This study clearly demonstrated the usefulness of interpolation methods and multivariate statistical analysis for identifying groundwater heavy metals. Optimal interpolation methods for estimating the spatial distribution of heavy metals in the Çarşamba Plain in northern Turkey were found to vary, which are as follows: IDW-1: best fit for Mn, Al, and B; IDW-2: best fit for Pb; IDW-3: best fit for Cd; OK: best fit for Fe; and RBF-IMQ: best fit for As. Maps illustrating problematic areas for drinking and drip irrigation were drawn using the best interpolation method for each heavy metal. Anthropogenic activities, agriculture among them, have led to high levels of groundwater pollution. The leading mineral pollutants in the study area include Fe, Mn, Al, As, and Cd, while B and Zn represent potential pollutants. Whereas As, Al, and Cd concentrations exceed permissible limits at certain locations within the study area; alarmingly, approximately 99 % of the study area suffers from Fe and Mn pollution. In addition to interpolation, data obtained from 78 groundwater wells were subjected to multivariate statistical analysis, with cluster analysis grouping wells into three clusters (“polluted,” “highly polluted,” “very highly polluted”) and factor analysis/principal component analysis identifying five main factors explaining 73.59 % of the total variance of the 17 variables tested. Results indicated that groundwater in the study area cannot be used as drinking water due to high levels of Fe, Mn, and As. Moreover, groundwater cannot be used for drip irrigation due to very high levels of iron and magnesium.

**Acknowledgments** This study was supported by Ondokuz Mayıs University, Scientific Research Programs under the project no. of PYO. ZRT. 1901.12.005.

## References

- Alam, F., & Umar, R. (2013). Trace elements in groundwater of Hindon-Yamuna interfluvial region, Baghpat District, western Uttar Pradesh. *Journal of the Geological Society of India*, 81, 422–428.
- Arslan, H. (2013). Application of multivariate statistical techniques in the assessment of groundwater quality in seawater intrusion area in Bafra Plain, Turkey. *Environmental Monitoring and Assessment*, 185(3), 2439–2452.
- Brahman, K. D., Kazi, T. G., Afridi, H. I., Naseem, S., Arain, S. S., & Ullah, N. (2013). Evaluation of high levels of fluoride, arsenic species and other physicochemical parameters in underground water of two sub districts of Tharparkar, Pakistan: a multivariate study. *Water Research*, 47, 1005–1020.
- Chidambaram, S., Karmegam, U., Prasanna, M. V., & Sasidhar, P. (2012). A study on evaluation of probable sources of heavy metal pollution in groundwater of Kalpakkam region, South India. *The Environmentalist*, 32(4), 371–382.
- Demirel, Z. (2007). Monitoring of heavy metal pollution of groundwater in a phreatic aquifer in Mersin-Turkey. *Environmental Monitoring and Assessment*, 132, 15–23.
- DSI (1994) Hydrogeological investigation report of Çarşamba Plain, Ankara.
- Eugenia, G. G., Vicente, A., & Rafael, B. (1996). Heavy metals incidence in the application of inorganic fertilizers and pesticides to rice forming soils. *Environmental Pollution*, 92, 19–25.
- FAO (1994). Water quality for agriculture, FAO Irrigation and Drainage Paper 29 rev. 1, Rome, 174 pp.
- Guler, C., Kaplan, V., & Akbulut, C. (2013). Spatial distribution patterns and temporal trends of heavy-metal concentrations in a petroleum hydrocarbon-contaminated site: Karaduvar coastal aquifer (Mersin, SE Turkey). *Environmental Earth Sciences*, 70(2), 943–962.
- Guler, M., Arslan, H., Cemek, B., & Ersahin, S. (2014). Long-term changes in spatial variation of soil electrical conductivity and exchangeable sodium percentage in irrigated mesic ustifluents. *Agricultural Water Management*, 135, 1–8.
- Haloi, N., & Sarma, H. P. (2012). Heavy metal contaminations in the groundwater of Brahmaputra flood plain: an assessment of water quality in Barpeta District, Assam (India). *Environmental Monitoring and Assessment*, 184(10), 6229–6237.
- Hossain, M. G., Selim, A. H. M., Lutfun-Nessa, M., & Ahmed, S. S. (2013). Factor and cluster analysis of water quality data of the groundwater wells of Kushtia, Bangladesh: implication for arsenic enrichment and mobilization. *Journal of the Geological Society of India*, 81(3), 377–384.
- Hua, Z., Debai, M., Cheng, W. (2009). Optimization of the spatial interpolation for groundwater depth in Shule River Basin. *Environmental Science and Information Application Technology*. ESIAT 2009.
- Huang, S. W., & Jin, J. Y. (2008). Status of heavy metals in agricultural soils as affected by different patterns of land use. *Environmental Monitoring and Assessment*, 139, 317–327.
- Huang, G., Sun, J., Zhang, Y., Chen, Z., & Liu, F. (2013). Impact of anthropogenic and natural processes on the evolution of groundwater chemistry in a rapidly urbanized coastal area. South China. *Science of the Total Environment*, 463–464(1), 209–221.
- Kanmani, S., & Gandhimathi, R. (2013). Investigation of physicochemical characteristics and heavy metal distribution

- profile in groundwater system around the open dump site. *Applied Water Science*, 3, 387–399.
- Keshavarzi, A., & Sarmadiani, A. (2012). Mapping of spatial distribution of soil salinity and alkalinity in a semi-arid region. *Annals of Warsaw University of Life Sciences Land Reclamation*, 44(1), 3–14.
- Krishna, A. K., Satyanarayanan, M., & Govil, P. K. (2009). Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: a case study from Patancheru, Medak District, Andhra Pradesh, India. *Journal of Hazardous Materials*, 167, 366–373.
- Kumari, S., Singh, A. K., Verma, A. K., & Yaduvanshi, N. P. S. (2014). Assessment and spatial distribution of groundwater quality in industrial areas of Ghaziabad, India. *Environmental Monitoring and Assessment*, 186(1), 501–514.
- Kumbur, H., Ozsoy, H. D., & Ozer, Z. (2008). Mersin ilinde tarımsal alanlarda kullanılan kimyasalların su kalitesi üzerine etkilerinin belirlenmesi. *Ekoloji*, 17, 54–58.
- Li, J., & Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics*, 6, 228–241.
- Li, S. Y., & Zhang, Q. F. (2010). Spatial characterization of dissolved trace elements and heavy metals in the upper Han River (China) using multivariate statistical techniques. *Journal of Hazardous Materials*, 176, 579–588.
- Li, S., Xu, Z., Cheng, X., & Zhang, Q. (2008). Dissolved trace elements and heavy metals in the Danjiangkou Reservoir, China. *Environmental Geology*, 55, 977–983.
- Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *The Science of the Total Environment*, 313, 77–89.
- Lu, K. L., Liu, C. W., & Jang, C. S. (2012). Using multivariate statistical methods to assess the groundwater quality in an arsenic-contaminated area of Southwestern Taiwan. *Environmental Monitoring and Assessment*, 184(10), 6071–6085.
- Madhulakshmi, L., Rauma, A., & Kannan, N. (2012). Seasonal distribution of some heavy metal concentrations in groundwater of Virudhunagar District, Tamilnadu, South India. *Electronic Journal of Environmental, Agricultural and Food Chemistry*, 11(2), 32–37.
- Massart, D. L., & Kaufman, L. (1983). *The interpretation of analytical chemical data by the use of cluster analysis*. New York: Wiley.
- Monjerezi, M., Vogt, R. D., Gebru, A. G., Saka, F. D. K., & Aagaard, P. (2012). Minor element geochemistry of groundwater from an area with prevailing saline groundwater in Chikhwawa, lower Shire Valley (Malawi). *Physics and Chemistry of the Earth Parts A/B/C*, 50(52), 52–63.
- Narany, T. S., Ramli, M. F., Aris, A. Z., Sulaiman, W. N. A., & Fakharian, K. (2014). Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol–Babol Plain, Iran. *Environmental Monitoring and Assessment*, 186(9), 5797–5815.
- Noori, S. M. S., Ebrahimi, K., Liaghat, A. M., & Hoorfar, A. H. (2013). Comparison of different geostatistical methods to estimate groundwater level at different climatic periods. *Water and Environment Journal*, 27(1), 10–19.
- Quyang, Y., Zhang, J., & Parajuli, P. (2013). Characterization of shallow groundwater quality in the Lower St. Johns River Basin: a case study. *Environmental Science and Pollution Research*, 20(12), 8860–8870.
- Rabah, F. K. J., Ghabayen, S. M., & Salha, A. A. (2011). Effect of GIS interpolation techniques on the accuracy of the spatial representation of groundwater monitoring data in Gaza strip. *Journal of Environmental Science and Technology*, 4, 579–589.
- Shan, Y., Tysklind, M., Hao, F., Quyang, W., Chen, S., & Lin, C. (2013). Identification of sources of heavy metals in agricultural soils using multivariate analysis and GIS. *Journal of Soils and Sediments*, 13(4), 720–729.
- Sun, Y., Kang, S., & Li, F. (2009). Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. *Environmental Modelling & Software*, 24, 1163–1170.
- Tiwari, R. N., Mishra, S., & Pandey, P. (2013). Study of major and trace elements in groundwater of Birsinghpur Area, Satna District Madhya Pradesh, India. *International Journal of Water Resources and Environmental Engineering*, 5(7), 380–386.
- Varouchakis, E. A., & Hristopulo, D. T. (2013). Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environmental Monitoring and Assessment*, 185(1), 1–19.
- Vieira, J. S., Pires, J. C. M., Martins, F. G., Vilar, V. J. P., Boaventura, R. A. R., & Botelho, C. M. S. (2012). Surface water quality assessment of Lis River using multivariate statistical methods. *Water, Air, and Soil Pollution*, 223, 5549–5561.
- WHO (2011). *The guidelines for drinking-water quality* (fourth ed., ). Geneva: World Health Organization.
- Wilding, L. P. (1985). Spatial variability: its documentation, accommodation and implication to soil surveys. In D. R. Nielsen, & J. Bouma (Eds.), *Soil spatial variability* (pp. 166–194). Wageningen: Pudoc.
- Xie, Y., Chen, T., Lei, M., & Yang, J. (2011). Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: accuracy and uncertainty analysis. *Chemosphere*, 82, 468–476.
- Yadav, P., Singh, B., Mor, S., & Garg, V. K. (2013). Quantification and health risk assessment due to heavy metals in potable water to the population living in the vicinity of a proposed nuclear power project site in Haryana, India. *Desalination and Water Treatment*. doi:10.1080/19443994.2013.833877.
- Yao, L., Huo, Z., Feng, S., Mao, M., Kang, S., Chen, J., Xu, J., & Steenhuis, T. S. (2014). Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China. *Environmental Earth Sciences*, 71, 1911–1924.
- Yılmaz, T., Seçkin, G., & Sarı, B. (2010). Trace element levels in the groundwater of Mediterranean coastal plains—the case of Silifke, Turkey. *CLEAN – Soil Air Water*, 38(3), 221–224.