Integration and analysis of geological, geochemical, and remote sensing data of south of Neyshabur using principal component analysis

Hamed Fazliani [a, *], Abolghasem Kamkar-Rouhani [a], Alireza Arab-Amiri [a]

[a] *Faculty of Mining, Petroleum and Geophysics, Shahrood University of Technology, Shahrood, Iran*

ABSTRACT

Lack of the existence of known mineral prospects in the preliminary stages of mineral exploration is the main problem of data-driven mineral potential modeling methods. On the other hand, applying the expert's knowledge and judgment in different stages of mineral potential modeling is the main difficulty of knowledge-driven mineral potential modeling methods. In addition, other difficulties in these methods can be mentioned such as determination of important variables, weighting to various classes of maps or information layers, and so on. Hence, the accuracy of the results of the knowledge-driven modeling methods is highly dependent on the amount of knowledge and experience of the expert. In this study, the principal component analysis (PCA) has been introduced as a knowledge-driven method with the least reliance on the expert's knowledge for mineral potential modeling. In this method, the expert's knowledge is only used in the interpretation of the results obtained from the modeling and is not considered in the first stages of mineral potential modeling and the definition of the conceptual model. In the introduced method, the interpretation of the results is conducted based on the positive and negative coefficients of variables in the eigenvalues table. Using these coefficients, it is determined that each principal component (PC) is associated with what type of mineralization. An advantage of this introduced method is to identify various types of mineralization in the area of interest using just one modeling effort. In order to evaluate the efficiency of this method, a region including two geological maps of Kadkan and Shamkan in the south of Neyshabur, northeast of Iran was selected. Two mineralization types including podiform chromite and epithermal gold-antimony mineralization types have been identified using the proposed method that presents more precise results than those of conventional univariate and multivariate geochemical studies.

Keywords: *Mineral potential modeling, Principal component analysis, Podiform chromite deposit, Epithermal gold deposit, Kadkan, and Shamkan area*

1. Introduction

The principal component analysis (PCA) is a classic analytical method that converts multivariate observations into several nondependent components to be ranked based on their variances [1]. In this process, every new component continually estimates the maximum possible value of the total variance. Therefore, the new computed components possess the total variance of the original data [2]. This method has been successfully applied almost in all different branches of mineral exploration. Many researchers have used the PCA method in exploration studies, including examination of geological units [3, 4], geochemical data processing [1, 5, 6, 7], geophysical data processing [8, 9, 10, 11], and remote sensing processes [12, 13]. Also in some exploration studies, PCA has been used to reduce the dimensions of the data. For example, in a study conducted in the same area [14], 28 geochemical variables have been initially reduced using PCA, and then the fuzzy logic method has been used to model the epithermal gold-antimony deposits. In addition, the PCA method has been rarely used in the integration of various exploration data. In a few studies on the volcanic belt of Central Iran [15, 16, 17], the PCA method has been used as a data-driven approach to integrating the exploration data. In these studies, the PCA method is first applied to the data of known mineralized targets in the study region, and then, the results are used to determine the factors of integration formula and to detect the altered

and mineralized zones in the region.

In this study, the PCA method has been introduced as a knowledge-driven method for mineral potential modeling. Because, in the primary stages of an exploration project, there are not usually a lot of known mineral prospects to use in the data-driven approaches. Besides, excessive use of knowledge and experience of the expert is one of the problems of knowledge-driven methods of mineral potential modeling, which makes the results non-unique and error-prone. In the proposed method, it has been tried to use the knowledge and experience of the expert only in the stage of interpreting the obtained results. Also, in all common methods of mineral potential modeling, the conceptual model prepares based on the type of mineralization sought. As a result, the type of mineralization must first be determined. But, in the proposed method there is no need to prepare a conceptual model. As a result, there is no need to know basic information about the type and characteristics of the mineralization in the region. Finally, in all the conventional mineral potential modeling methods, each modeling process results in identifying one specific mineralization in the area of interest. But one of the aims of the proposed method is to identify all types of mineralization in the region by carrying out just once the modeling process.

To introduce this method, all exploration data available in the Kadkan and Shamkan prospecting areas in the northeast of Iran, including geological, remote sensing, and stream geochemical data have been used. For this purpose, after performing the necessary preprocessing, the PCA technique was directly applied to the data. Then, the obtained

---

results were interpreted according to the positive and negative coefficients of variables in the eigenvalues table. Using these coefficients, it was determined whether each output component from the PCA indicates mineralization or not, and the type of mineralization was also determined. Finally, the location of known mineral prospects in the region was used to compare the results of the proposed method with those of conventional univariate and multivariate geochemical studies.

2. Location and geology of the study region

The study region is located in the south of Neyshabur, northeast of Iran. The region can be seen in two 1:100,000 geological maps of Kadkan and Shamkan areas with a total surface area of 4800 km², situated between longitudes of 58° to 59° E and latitudes of 35°30′ to 36° N. From a tectonic point of view, the region is located in the southern part of the fractured Sabzevar Zone between the Lut Block in the south and the Binalood Zone (eastern Alborz) in the north. The southern border of the region is the Darouneh Fault. The region is arid to semiarid and is formed of small hills and low mountains in the north and flat plains to hills in the middle and high mountains of elevation around 3000 m in south [18].

The oldest geological units in the region are sedimentary, pyroclastic, volcanic, plutonic, and ophiolitic rocks of the Upper Cretaceous age that become thick in the Cenozoic. The most important lithologies involved in the mineralization are intrusive and ophiolitic rocks [19, 20].

The intrusive rocks of the region have been formed within the Pyrenean orogeny. This event in the early Oligocene is formed of the granitoid intrusive rocks after the Eocene volcanism. These suites are the main agent of metallic mineralization and hydrothermal alteration in the region. The felsic intrusive rocks include acidic to alkaline rocks of granitic, granodioritic, and dioritic types that have extended in the north and southwest of the region. The intrusion process has affected the host rocks through thermal alteration and has produced contact metamorphism [19, 20].

The ophiolitic facies are formed before the Upper Cretaceous age in a narrow ocean between the sub-continent in the east of Central Iran and parts of it obducted on the continental crust through the rotation of the sub-continent during the Laramide orogeny. Overall, these ophiolitic rocks extended from southeast to northwest and are mostly extended in the east and southwest of the region [19, 20]. The map of the main geological units in the study region is shown in Fig. 1.
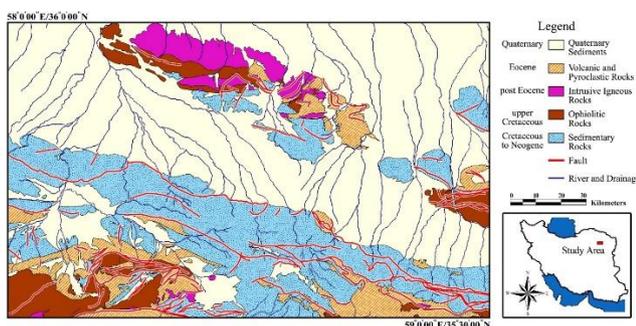

Fig. 1. The geological map of the main units and faults of the study region (modified from [19, 20]).

3. Dataset

Since the purpose of this study was to investigate the possibility of using the PCA for the integration of exploration data and mineral potential modeling, all available exploration data in the study region have been used for this purpose. These data sets include geological, remote sensing, and stream sediment geochemical data.

3.1. Geological data

The data, used to study the geological conditions of the region, were extracted from the Kadkan and Shamkan 1:100,000 geological maps published by the Geological Survey of Iran in 1998. In order to extract the lithological units from these maps, first, the geological maps were digitized, and then, the digitized units were divided into five main units including Quaternary plains, sedimentary rocks, volcanic rocks, ophiolite units, and intrusive suites.

Since the information in the geological maps is qualitative, it should be converted into quantitative data to be used in statistical analyses. Two different methods have been used to achieve this purpose. In the first method, the geological units are scored as 1 or 0 based on their presence or absence in a specific cell. This method is the common approach of preparing the binary evidential maps [21]. In the second method, map cells are scored based on the distance between each cell and a certain geological unit. In this approach, the geological data is no longer discrete (0 and 1), and in addition, the statistical distribution of the data changes from the Bernoulli distribution to the Gaussian distribution, which is one of the requirements for most statistical analyses [22]. In this study, the inverse distance weighting method was used to score each cell based on the distance from the desired geological unit. In this method, the score of each cell is calculated from the following equation [21]:

$$S_i = \frac{1}{d_i} \tag{1}$$

Where d_i is the distance of the central point of the cell from the desired geological unit. Using this method, the geological qualitative data is converted into quantitative data that could be used in statistical analysis. Fig. 2a shows the map of intrusive suites in the region. Fig. 2b also shows the scoring based on the distance from the intrusions.
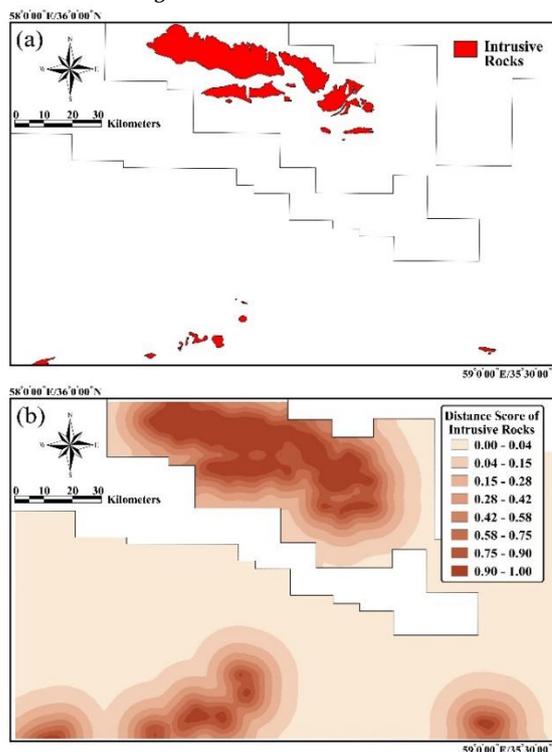

Fig. 2. (a) The map of the intrusive suites in the study region, **(b)** scoring based on the distance from the intrusive rocks.

3.2. Remote sensing data processing

Considering the fact that most of the ore bodies are surrounded by alteration zones, the presence of altered rocks is a valuable index for exploring the ore deposits [23]. Multispectral images can be used to detect the altered rocks in the regions that have outcropped lithologies because the spectral reflection of these rocks differs from that of unaltered rocks [23]. Hydroxyl-bearing minerals, such as kaolinite,

montmorillonite, illite, and other clay minerals, as well as iron-oxide minerals including hematite, limonite, and goethite, are predominant alteration products in various alteration zones [24]. Detecting these minerals using satellite images can lead to discrimination of altered zones and exploration of new mineral prospects.

In this study, the Landsat 7 ETM+ images (Acquisition date: May 22, 2001) were used. These ETM+ images were processed using the band ratio method by the ENVI 5.1 software to detect the altered zones. In order to process the ETM+ images, two-band ratios of 3/1 and 5/7 are usually used to discriminate the hydroxyl and ferric iron oxide minerals, respectively [25, 26].

3.2.1. Band ratios 5/7 and 3/1 for the ETM+ sensor

The highest spectral reflection of the hydroxyl-bearing minerals belongs to the wavelength of band 5 of the ETM+ sensor. On the other hand, these minerals have a low spectral reflection in the wavelength of band 7 [23]. In order to create the band ratio of 5/7, first, the atmospheric correction was applied to the images by the FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) method [27] in ENVI 5.1 software. Then, the digital number of each pixel in band 5 was divided into the digital number of that exact pixel in band 7. The spectral reflection of unaltered rocks in band 7 is identical to that of band 5. Therefore, the band ratio 5/7 for these rocks is approximately equal to unity (1.00). However, altered rocks have low reflections in band 7, and therefore, the band ratio of 5/7 for the altered rocks is much more than unity. Hence, altered and unaltered rocks can be discriminated against based on the difference in their band ratio values [23]. Fig. 3a shows the density sliced band ratio of the 5/7 image in the study region, in which the high values of the band ratio 5/7 are shown in red.

The ferric-iron-oxide minerals also have their lowest reflection in band 1 and their highest reflection in band 3. Therefore, the altered rocks that bear these minerals will show the maximum value for band ratio 3/1 [23]. Fig. 3b shows the density sliced band ratio 3/1 image in the study region, in which the high values of the band ratio 3/1 are shown in red.

3.2.2. Removal of the effects of vegetation and limestone rocks

The vegetation and limestone rocks possess the same spectra that interfere with the images that show clay minerals (images of band ratio 5/7) and create errors in the results. In this study, the normalized difference vegetation index (NDVI) was used to remove the vegetation effect. The value of the NDVI is calculated from equation (2) [28]:

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2}$$

where *NIR* and *Red* are pixel values from the near-infrared and red (visible) bands, respectively. This method first measures the NDVI value for the area of interest, and then, subtracts it from the values of band ratio 5/7 for each pixel. In order to remove the effect of limestone rocks, the area of these rocks was extracted from the geological map, and their pixel values were substituted by the mean value of their surrounding pixels.

3.2.3. Resizing the satellite data cells

The pixel size in ETM+ satellite images is 30×30 m, which does not match the cell size of other data (e.g., the cell size of geochemical data is about 1.5×1.5 km). Therefore, in order to transfer the remote sensing data into the PCA analysis, the cell size should be extended to match the geochemical cells.

This task was conducted by the ENVI 5.1 software by averaging the digital numbers of the existing satellite pixels within each geochemical cell (Fig. 4). Afterward, the data of each cell as well as its geographic coordinates were extracted to be used in the PCA analysis.

3.2.4. Calculation of the photolineament factor

The patterns of local and regional fractures, as the controllers of ore bodies, are very important. The comparison between the lineation pattern and mineralization events has proved that the ore bodies tend

to occur along the lineaments and to focus at the intersections [23]. In this study, the photolineament factor was used to determine the fracture pattern in the region (Fig. 3c). The value of the photolineament factor is calculated from equation (3) [29]:

$$PF = \frac{a}{A} + \frac{b}{B} + \frac{c}{C} \tag{3}$$

where *PF* is the photolineament factor, *a* is the total length of lineation per cell, *b* is the number of intersections per cell, *c* is the number of lineations per cell, and the capital letters represent the mean values of the same parameters in all of the cells.
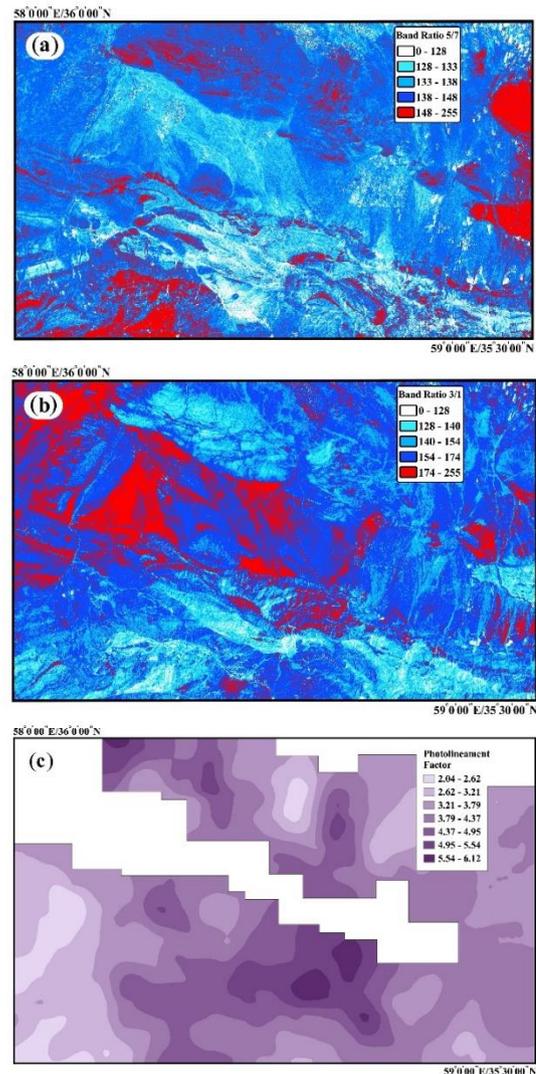

Fig. 3. (a) The density sliced band ratio 5/7 image, (b) the density sliced band ratio 3/1 image (the high values of the band ratios are shown in red), and (c) photolineament factor map of the study region.
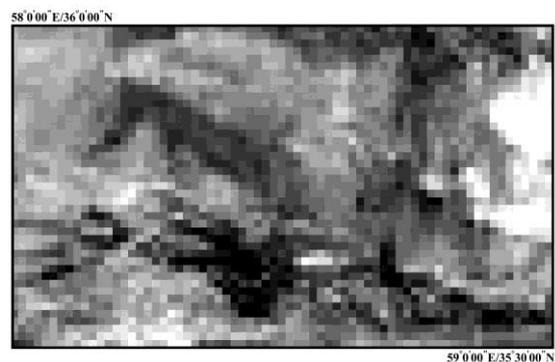

Fig. 4. The band ratio of 5/7 in the study region after resizing the pixels.

3.3. Stream sediment geochemical data

The geochemical data, used in this study, were obtained from the geochemical studies of the Kadkan and Shamkan geological maps. These studies were carried out by Jiangxi, a Chinese company [18, 30]. A systematic sampling grid with an approximate spacing of 1.5×1.5 km² was applied to the areas having outcrops. Afterward, in each cell 2, 3, or 4 sub-samples from first- and second-order streams were taken, and a composite sample, obtained from the combination of these sub-samples, was assigned to the sampled cell. The samples were chemically analyzed for 28 elements including Ag, As, Au, B, Ba, Be, Bi, Co, Cr, Cs, Cu, F, Hg, Li, Mo, Nb, Ni, P, Pb, Rb, Sb, Sn, Sr, Th, U, V, W, and Zn. In order to get the best results, different analytical methods have been used for different elements. The concentrations of Sb, As and Hg have been detected by the atomic fluorescence spectrometry (AFS) method. Also, the concentration of Au, B, Cr, and Mo have been determined using electrospray ionization (ESI), atomic emission spectroscopy (AES), X-ray fluorescence (XRF), and catalytic polarography (POL) methods, respectively. Ultimately, the concentration of Cu, Co, and other elements have been detected by the inductively coupled plasma atomic emission spectroscopy (ICP) method. [18, 30]

3.3.1. Removal of the lithologic effect from geochemical data

The concentration of geochemical samples, taken from different rocks or materials, is influenced by the syngenetic component (element background in source rock) of those specimens. In mineral exploration, one generally aims to identify the epigenetic component associated with the mineralization, but sometimes the lithologic effect is so high that it completely covers the epigenetic component [31]. Hence, before analyzing the geochemical data, the lithologic effect should be eliminated. Since the stream sedimentary samples of this study were taken from the tributaries and the results of their chemical analyses were assigned to their systematic sampling cells, the removal of the lithologic effect based on the rocks in each cell was conducted. Considering the existence of six main geological units in the study region and based on the presence of different rock units in the sampling cells, the samples were categorized into 42 different categories. Eventually, the lithological effect was eliminated in all cells by dividing each data by the median value of the category and calculating the enrichment factor in each category. Equation (4) shows how to calculate the enrichment factor [31]:

$$EF_i = \frac{c_i}{c_m} \qquad (4)$$

where EF is the enrichment factor, C_i is the concentration of the element in i^{th} sample and C_m is the median value of each element in each category.

3.3.2. Geochemical data selection

An effective factor, which reduces the accuracy of the PCA results, is the multiplicity of the variables. Furthermore, the large number of variables can increase the number of PCs and makes it harder to interpret the data. Therefore, eliminating low-impact variables before conducting the PCA would increase the accuracy of the analysis and would provide better results. Since there are many elements in the stream sediment geochemical data, the low-impact variables were identified using the maximum enrichment factor method and were eliminated from the model. The value of this factor can be calculated from equation (5) for each element.

$$EFM = \frac{c_{MAX}}{c_m} \qquad (5)$$

where EFM is the maximum enrichment factor, C_{MAX} is the maximum concentration of the element in the study region, and C_m is the median value of the concentration of that element. This factor can show the element enrichment intensity compared to the central statistic parameter. Statistical parameters of 28 studied geochemical elements, and their maximum enrichment factors, are shown in Table 1.

Table 1. Statistical parameters and maximum enrichment factors of the stream sediment geochemical data.

Variable	Count	Mean	Minimum	Maximum	EFM
Zn	1721	59.48	34.78	245.06	4.12
Pb	1721	12.26	1.70	37.30	3.04
Ag	1721	0.06	0.03	0.17	2.64
Cr	1721	1110.24	37.80	18437.20	16.61
Ni	1721	196.08	18.78	1193.00	6.08
Bi	1721	0.13	0.07	0.32	2.40
Cu	1721	42.37	14.73	248.08	5.85
As	1721	9.21	2.10	482.00	52.35
Sb	1721	0.39	0.10	12.10	30.76
Co	1721	24.54	8.22	123.89	5.05
Sn	1721	1.36	0.77	3.00	2.20
Ba	1721	261.26	77.04	699.82	2.68
V	1721	163.48	38.76	940.03	5.75
Sr	1721	383.00	68.28	1970.00	5.14
Hg	1721	0.01	0.00	0.31	29.34
W	1721	0.60	0.17	11.10	18.44
B	1721	37.00	7.16	216.20	5.84
Be	1721	1.39	0.62	3.13	2.26
Mo	1721	0.69	0.20	4.00	5.77
Li	1721	24.50	8.17	77.11	3.15
Au	1721	1.16	0.30	18.54	16.01
Rb	1721	35.04	3.30	95.00	2.71
P	1721	0.05	0.02	0.16	2.91
Cs	1721	4.44	2.26	11.01	2.48
Nb	1721	7.76	3.30	19.90	2.56
Th	1721	5.70	1.00	13.30	2.33
U	1721	1.25	0.27	5.52	4.42
F	1721	302.38	100.00	853.00	2.82

In order to reduce the number of variables and to eliminate the low-impact geochemical variables, only those elements were used in the PCA that their maximum enrichment factors were above 5.

3.4. Preprocessing and data correction

The PCA method is sensitive to outlier values and without eliminating these values, the results are incorrect [32]. Therefore, before performing the main processing, the outlier values were identified for all variables by the boxplot method [33] and were replaced by the maximum values of those variables. For example, the Tukey boxplots of Cr, Ni, Cu, Sb, and Au are shown in Fig. 5.
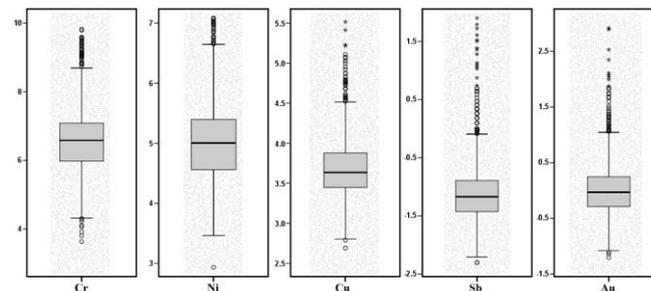

Fig. 5. Tukey boxplots of Cr, Ni, Cu, Sb, and Au after transforming by a logarithmic function.

Furthermore, the Gaussian distribution of the input data is one of the main requirements that should be considered for most multivariate statistical methods [22]. However, the distribution for most of the geochemical data, used in this study, was close to lognormal. Therefore, in order to minimize the skewness effect of the distribution and to transform the data into a normal distribution, the raw data were transformed through a logarithmic function.

Moreover, the PCA method is sensitive to the scale of initial variables by which they have been measured. Even if the measurements of variables have been conducted based on one unit, but the range of a variable is much bigger than the others, the first PC is affected by the variable of question. A practical solution for this issue is data standardization. This helps all initial variables to have the same value in the analysis [34]. In this study, data standardization was conducted

based on their statistical properties using equation (6) [34, 35]:

$$y_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i} \qquad (6)$$

where μ_i is the mean value of the ith variable and σ_i is the variance of that variable. The main feature of this method is that each variable would have a mean of zero and a variance of one after the transformation.

4. Mineral potential modeling using the PCA method

After selecting the variables and performing the required preprocessing, for mineral potential modeling in the study region, the PCA technique was applied to the data using the IBM SPSS Statistics 22 software package. Afterward, the results were plotted using ArcMap 10.3 software and were exported as maps. In order to model the mineral potential in the study region, 18 variables including geological, remote sensing, and geochemical variables were used. These variables have been listed in Table 2.

Because mineralization-related data (e.g., geochemical anomalies) accounts for a small proportion of the total variance of all input data, they may not be among the first components obtained from PCA. Therefore, the elimination of the last output components may lead to the loss of information on some mineralization in the area. Hence, the number of output components in this study was equal to the number of input variables, and useful components were selected in the results interpretation stage.

Table 2. The variables used in mineral potential modeling of the study region.

No.	Type	Variable
1	Geological Data	Intrusive Rocks
2		Ophiolitic Rocks
3	Remote Sensing Data	Photolineament Factor (PF)
4		Hydroxyl Bearing Minerals (ETM+ 5/7 Band Ratio)
5		Ferric Iron (ETM+ 3/1 Band Ratio)
6	Geochemical Data	Cr
7		Ni
8		Co
9		V
10		Cu
11		Mo
12		W
13		Au
14		Sb
15		As
16		Sr
17		Hg
18		B

In Table 3, the eigenvalues obtained from the PCA method are shown. Furthermore, a 100% stacked column chart of different analyzed PCs is demonstrated in Fig. 6. As shown in Fig. 6, the share of different variables in all PCs is balanced. This indicates that the PCs are not affected by a specific variable, and there are a number of variables that affect each PC.

Table 3. Eigenvalues of the variables in principal components analysis of the study region.

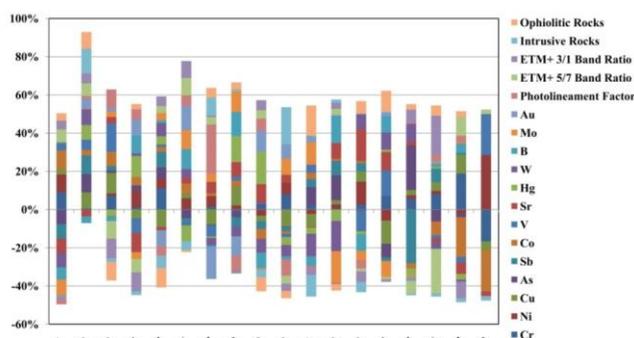| Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18
Cr	0.75	0.03	0.35	0.06	0.46	0.01	0.04	-0.03	0.06	0.19	0.05	0.01	0.05	0.11	0.06	0.09	0.19	-0.07
Ni	0.77	-0.02	0.06	0.43	0.21	0.21	0.14	0.06	-0.08	0.14	-0.05	0.04	0.22	-0.09	0.02	0.02	-0.02	0.12
Cu	0.23	0.48	0.52	-0.23	-0.38	0.09	-0.20	0.28	-0.07	-0.21	-0.14	0.04	0.10	-0.19	0.03	0.05	0.10	-0.02
As	-0.63	0.54	0.07	0.14	0.25	-0.17	-0.01	-0.08	0.01	-0.01	0.20	0.24	-0.01	-0.11	0.25	-0.07	-0.02	0.00
Sb	-0.63	0.55	0.11	0.01	0.34	-0.07	0.00	-0.14	-0.10	-0.05	0.09	0.14	0.08	-0.03	-0.30	0.08	0.01	0.00
Co	0.81	0.16	0.36	0.22	0.10	0.20	0.04	0.08	-0.13	-0.03	0.07	0.02	0.02	0.00	-0.03	-0.08	-0.21	-0.09
V	0.32	0.31	0.72	-0.39	0.06	-0.07	-0.15	0.06	0.02	-0.04	0.09	0.00	-0.20	0.21	0.00	-0.02	-0.03	0.09
Sr	-0.62	-0.18	0.16	-0.50	-0.07	0.11	0.15	0.21	0.24	0.10	-0.01	0.15	0.31	0.15	0.02	-0.02	-0.06	-0.01
Hg	0.05	0.43	-0.17	0.55	0.18	-0.31	-0.02	0.38	0.42	-0.10	-0.07	-0.11	-0.03	0.02	-0.04	0.01	-0.03	0.00
W	-0.62	0.47	-0.01	0.07	0.29	0.17	-0.10	-0.14	-0.18	-0.14	-0.25	-0.29	0.14	0.14	0.10	-0.04	-0.01	0.00
B	-0.53	-0.20	-0.11	0.47	0.06	0.39	-0.02	0.35	-0.20	-0.04	-0.11	0.26	-0.16	0.13	0.00	0.02	0.05	0.00
Mo	-0.68	0.02	0.14	-0.18	0.16	0.37	0.11	0.30	-0.02	0.20	0.25	-0.30	-0.09	-0.13	-0.01	0.01	0.01	0.00
Au	-0.15	0.28	0.09	0.42	-0.35	0.47	-0.43	-0.28	0.27	0.18	0.07	0.01	0.01	0.03	-0.02	0.01	0.00	0.00
ETM+ 5/7 BR	0.54	0.37	-0.45	-0.35	0.15	0.34	0.10	0.00	0.11	-0.10	0.01	0.06	-0.03	-0.02	-0.07	-0.26	0.10	0.01
ETM+ 3/1 BR	0.38	0.30	-0.50	-0.50	0.21	0.33	0.02	-0.04	0.13	-0.09	-0.07	0.06	-0.10	-0.02	0.07	0.23	-0.09	0.00
PF	-0.24	0.09	0.47	0.27	-0.21	0.21	0.66	-0.22	0.16	-0.20	-0.03	-0.02	-0.08	0.00	0.01	0.03	0.03	0.00
Intrusive Rocks	0.00	0.73	-0.09	-0.10	-0.27	-0.17	0.24	0.03	-0.10	0.46	-0.24	0.03	-0.10	0.01	-0.01	-0.02	-0.02	-0.01
Ophiolitic Rocks	0.32	0.49	-0.47	0.14	-0.42	-0.04	0.13	0.10	-0.18	-0.09	0.33	-0.06	0.12	0.17	0.03	0.06	0.03	0.00

Fig. 6. 100% stacked column chart indicates the share of the variables in different PCs.

4.1. Interpretation of the results

After modeling, the results are interpreted based on the eigenvalues obtained from the PCA method. The eigenvalues obtained for each PC are equivalent to the loading values of different variables and indicate the effect of each variable on that PC. As a result, using these values, it can be determined that each PC indicates a geological event or ore-forming process.

In the study region, first, the results of the modeling were interpreted based on the eigenvalues (Table 3) obtained from the PCA method. Afterward, in each case, the results were validated using the known mineral prospects in the study region. These known mineral prospects include the active mines or the outcropped mineralized locations and were extracted from the mineral prospecting studies [36, 37] and the active mines report [38] prepared by the Geological Survey of Iran.

Finally, the results of the modeling were compared with those of the studies that were conducted only on the stream sediment geochemical data, to determine the effect of using multiple data.

4.2. Chromium and associated elements mineralization

As shown in Table 3, Co, Ni, Cr, alterations (hydroxyl-bearing minerals and ferric iron oxides), V, and ophiolitic rocks have the highest weights in PC1, respectively. The weights of geochemical elements Co, Ni, Cr, and V in PC1 are 0.81, 0.77, 0.75, and 0.33, respectively. These elements are associated with podiform chromite deposits [39, 40]. It should be noted that almost all Iranian chromite deposits are podiform [41].

Also, podiform chromite deposits are always found in dunite,

serpentinite, or peridotite rocks in ophiolitic sequences [39, 42], which is indicative of the high weight of the ophiolitic rocks (0.32) in PC1.

Furthermore, the ETM+ band ratios of 5/7 and 3/1 (representing hydroxyl-bearing minerals and ferric iron oxides) have weights of 0.54 and 0.38, respectively. Therefore, there are two effective variables in PC1. This can be explained by the presence of serpentinites and serpentine alterations in the ophiolitic rock sequence. Although the spectral curve of serpentinite is relatively flat, there is a small spectral absorption at about 0.45 μm (ETM+ band 1) due to the presence of ferric iron, and a relatively more apparent spectral absorption at about 2.3 μm (ETM+ band 7) due to the presence of OH [43]. Therefore, serpentinite has relatively high values in the ETM+ band ratios 5/7 and 3/1. On the other hand, as mentioned before, chromite deposits are always hosted by dunite, serpentine, or peridotite. Hence, the presence of serpentine alteration can be a signature of the presence of these deposits [44].

According to the above discussion, PC1 can be considered as the representative of the chromium mineralization in the performed modeling on the study region. In Fig. 7, the results of PC1 are compared with those of univariate and multivariate stream sediment geochemical studies of the region. It should be noted that in each map, the range of the class is $\bar{X} + 1S$, $\bar{X} + 2S$ and $\bar{X} + 3S$, where $\bar{X}$ is mean and S is the standard deviation of the variable.
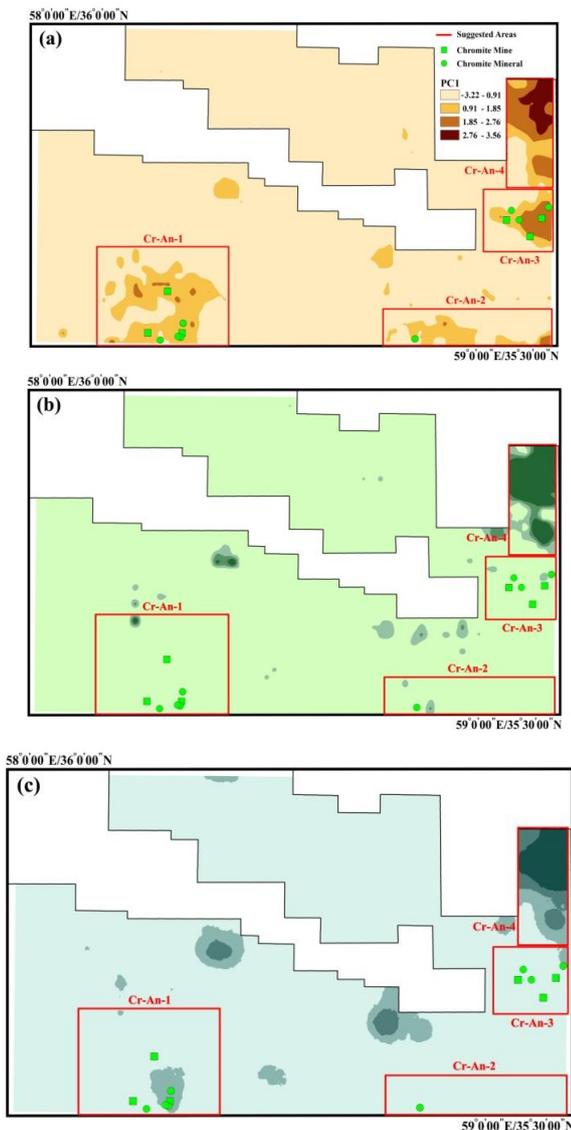

Fig. 7. (a) Map of PC1 scores (the present study), (b) map of Cr concentration values [18, 30], and (c) map of PC2 scores from the PCA on only the stream sediment geochemical data [45].

Fig. 7a shows the known Cr prospects. These known mineral prospects include active chromite mines and the outcropped mineralized zones. As shown in Fig. 7a, the locations of known Cr prospects in the study region, completely agree with the modeling performed by the PCA method. Accordingly, in other areas where PC1 shows anomalies, even if there are no known Cr prospects, it can be considered as the presence of Cr deposits. Therefore, these areas can be studied for further mineral prospecting. In Fig. 7a, the areas Cr-An-1 to Cr-An-4 were identified as the zones having Cr-mineralization potential in the study region.

To compare the results obtained from the mineral potential modeling using the PCA method with the results of univariate stream sediment geochemical studies, the map of Cr concentration values [18, 30] was used, which is shown in Fig. 7b. As shown in Fig. 7b, there are no anomalous Cr zones in Cr-An-1 and Cr-An-3 areas, and there are only small anomalies within the Cr-An-2 area.

Fig. 7c shows the map of PC2 scores obtained from the PCA performed only on the stream sediment geochemical data of the region [45]. As the map shows, performing the PCA only on the geochemical data cannot identify the anomalies in the Cr-An-2 and Cr-An-3 areas. In addition, in Cr-An-1, the identified anomaly is not very strong. Besides, as a result of applying the PCA method only on the geochemical data, many zones in the central parts of the region are considered anomalies, while these anomalies do not match with known mineral prospects in the region.

As a result, according to the known mineral prospects in the region, it can be stated that the modeling conducted by the PCA of all the available data (geological, stream sediment geochemical, and remote sensing data) leads to more accurate results in identifying the promising mineral zones of chromium and associated elements, compared to univariate and multivariate geochemical studies.

4.3. Au-Sb Mineralization

According to Table 3, in PC2 of the mineral potential modeling of the study region, the variables of distance from Intrusive Rocks, Sb, As, ophiolitic rocks, Cu, W, Hg, hydroxyl-bearing minerals, V, ferric iron oxides, and Au have the highest weights, respectively. The existing geochemical elements in the eigenvalues of PC2 include Sb, As, Cu, W, Hg, V, and Au, which are 0.55, 0.54, 0.48, 0.47, 0.43, 0.31 and 0.28, respectively. Almost all of these elements are associated with epithermal gold deposits [46].

Epithermal deposits are generally formed in shallow parts of hydrothermal systems (depths less than 1.5 km and temperatures below 300 °C). Hydrothermal fluids, mainly through replacing or filling the veins and cavities [47], affect the calc-alkaline to alkaline rocks of the volcanic arc on the convergent margins [48] and form the epithermal ore deposits.

Among the mentioned chemical elements, the study of the geochemical halos of Sb, As, Cu, and Hg as pathfinder elements of epithermal mineralization are common [46, 48]. Au, Sb, As, and Cu may also be found as epithermal ore deposits [47].

Although epithermal ore deposits are generally associated with volcanic calc-alkaline rocks in subduction-derived magmatic arcs [49, 50], they are also considered as the distal part of high-temperature hydrothermal systems [47]. As a result, these deposits could be spatially related to the intrusive rocks that form hydrothermal systems [51, 52]. Therefore, in the conducted modeling, the high weight of intrusive bodies in PC2 can be justified. In the table of PC2 eigenvalues, the intrusive bodies have the highest value, namely 0.73, showing a high correlation of intrusive rocks with epithermal gold and antimony mineralization. On the other hand, the presence of ophiolitic units with a weight of 0.49 among the important variables needs to be explained. Since this variable is not correlated with epithermal deposits, it seems that the spatial correlation of ophiolitic units with upward movement of the continental crust, which is the center of the formation of an epithermal deposit, is the reason for their weight.

Other effective variables on PC2 are the variables that show the alterations (hydroxyl-bearing and ferric iron oxide minerals). In the

eigenvalues table of this PC, the 5/7 band ratio of the ETM+ sensor (detecting the hydroxyl-bearing minerals) weights 0.37, and the 3/1 band ratio of the ETM+ sensor (detecting the ferric iron oxide minerals) weights 0.30.

It can be stated that the presence of hydrothermal alteration minerals is the key and discriminating feature of epithermal deposits, and usually, these minerals present distinct zoning around the epithermal deposits [48]. The alterations around these deposits include propylitic, argillic, and advanced argillic [47]. These alterations extensively bear hydroxyl minerals including kaolinite, illite, and alunite [47]. The presence of these minerals can justify the high value of the 5/7 band ratio of the ETM+ sensor in PC2.

Moreover, iron oxide minerals are present in the alteration zones around the epithermal deposits. These minerals form through supergene weathering and the oxidation of highly sulfidic rocks in advanced argillic zones and include jarosite and oxide-hydroxide iron minerals [47]. In some epithermal mineral deposits, part of the ore body is likely to be exposed to deep supergene oxidation (deeper than 300m). This process replaces the sulfide content with iron oxide minerals [53]. Accordingly, the high weight of the 3/1 band ratio of the ETM+ sensor in PC2 is explained through the presence of these minerals. Based on the above discussion and the modeling carried out in this study, it can be said that PC2 represents promising or prospect locations for epithermal gold and antimony mineralization in the study region. Fig. 8 compares the results of PC2 with those stream sediment geochemical univariate and multivariate analyses.

Fig. 8a shows the PC2 score map and the known mineral prospects of Au and Sb in the study region. The prospects of these ores include the active gold and antimony mines and the outcropped mineralized zones. As can be seen from Fig. 8a, the known Au and Sb prospects completely conform to the PCA modeling, which proves the correctness of the modeling. Therefore, it is possible to explore Au and Sb deposits in other areas that have anomalous PC2 values. Considering Fig. 8a, four areas, Au-An-1 to Au-An-4, were introduced as the likely mineralized Au and Sb areas.

To compare the results of the PCA mineral potential modeling with those of stream sediment geochemical univariate analyses, the maps of Au and Sb concentration values [18, 30] were used, which are shown in Fig. 8b and 8c respectively. As shown in Fig. 8b, Au is not anomalous in Au-An-3 and has a small anomalous area in Au-An-1. Moreover, Au values over the sedimentary units in central parts of the region are not correlated with the known outcropped mineralized locations, which can be due to the pollution or the usual errors of gold sampling. Besides, the map of Sb values (Fig. 8c) does not show any anomalous area in Au-An-1 and has a small anomalous area in Au-An-3.

Fig. 8d shows the PC3 score map of the PCA, which was conducted on the stream sediment geochemical data [45]. As it shows, the PCA modeling of only the stream sediment geochemical data does not detect any anomalies in the Au-An-1 zone, and the anomaly in Au-An-3 is also very weak. Moreover, the anomalous location in the southeast part of the region resulted from applying the PCA method on the geochemical data does not agree with the known mineral prospects.

Based on the above discussion and the investigation of the known mineral prospects, it can be concluded that the modeling conducted by the PCA method on all the existing datasets in the study region (stream sediment geochemical data, as well as the geological and remote sensing data) provides better results than those obtained from the univariate and multivariate geochemical studies.

5. Conclusion

After applying necessary preprocessing and basic processing analyses on the available exploration data from the study region, the mineral potential modeling by the PCA method was conducted using 18 variables from geological, remote sensing, and stream sediment geochemical investigations (Table 2). The results of mineral potential modeling by the PCA method (Table 3) were interpreted based on the

eigenvalues of initial variables in each PC. Finally, the results of the conducted mineral potential modeling were validated using the known mineralized outcrops, and also, these results were compared with conventional univariate and multivariate geochemical studies. Consequently, more precise results were obtained using the proposed method in comparison with univariate and multivariate geochemical studies.
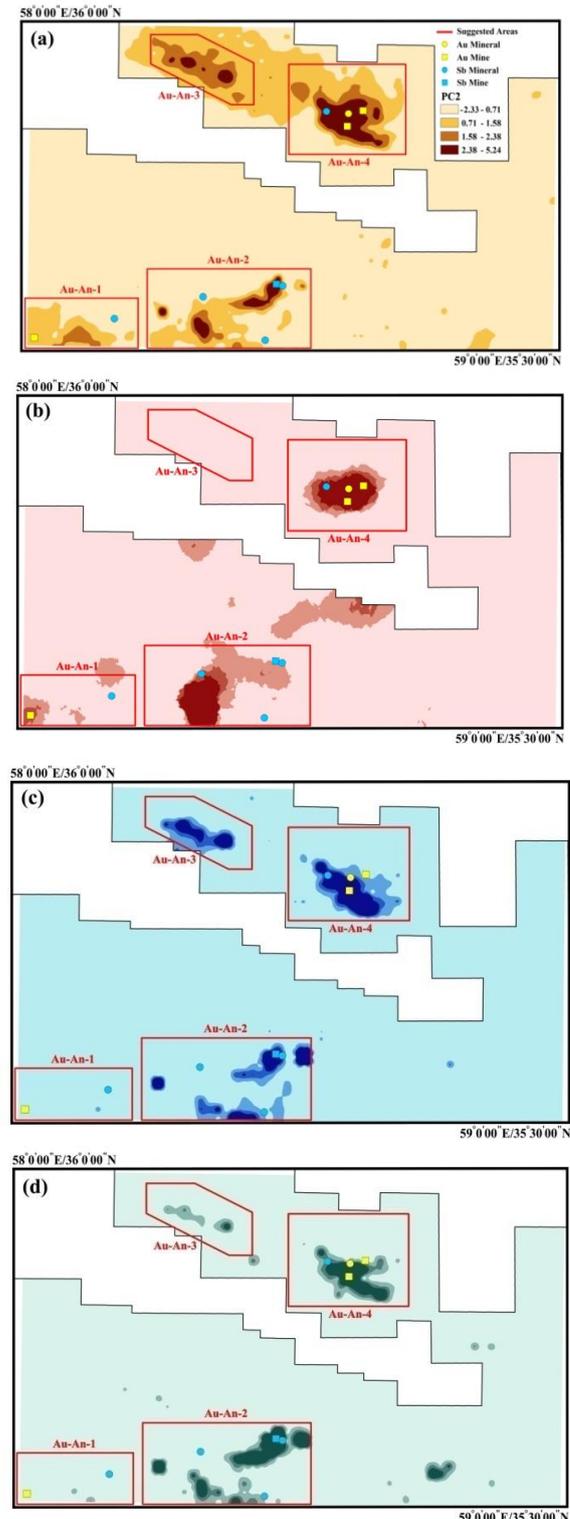

Fig. 8. (a) Map of PC2 scores (the present study), (b) map of Au concentration values [18, 30], (c) map of Sb concentration values [18, 30], and (d) map of PC3 scores from the PCA on only the stream sediment geochemical data [45].

Considering the known mineral prospects of the study region (Fig. 7a and 8a), the PCA was proved to be an effective method for mineral potential modeling in the early exploration stages. The proposed method does not suffer from the common problems of knowledge-driven and data-driven methods of mineral potential modeling, because it does not use the expert knowledge and experience as well as the known mineralized zones. Applying the expert knowledge and experience in knowledge-driven methods provides non-unique and erroneous results. Besides, the necessity of using the known mineralized outcrops in the data-driven methods makes their application difficult. It should be noted that information on the mineral prospects is not accessible in the primary stages of mineral exploration in an area.

Comparing the results of the PCA method with those of univariate and multivariate geochemical data (Fig. 7b, 7c, 8b, 8c, and 8d) showed that the PCA results were more correlated with the known mineralized zones and can be more effective in the exploration of promising mineral deposits. Considering the lithological and alteration information, which are spatially correlated to most mineral deposits, a general conclusion is that the proposed method increases the exploration chance in the promising areas.

In the proposed method, there is no need for basic information on existing mineralization types in the study region that is a big advantage over the conventional methods of mineral potential modeling. In all common mineral potential modeling methods, first, the type of mineralization should be determined [21], and then, based on the mineralization features, the effective variables and their weights are determined. However, as it was shown in this study, the proposed method was successfully capable of determining the presence of two different mineralization models including a podiform chromite deposit and an epithermal gold-antimony deposit in the study region through finding the spatial relationship of variables without using the basic mineral information of the region.

In conventional mineral potential modeling methods, each modeling process will result in one specific known mineralization in the area of interest [21]. In the proposed method, however, all mineralization types in the region are determined by carrying out modeling just once. Furthermore, applying the expert knowledge on the modeling process is significantly decreased in the proposed method, and the expert knowledge is only used in the interpretation stage. The interpretation of the results is also made considering the positive and negative correlations of variables in the eigenvalues table. These relationships show what type of mineralization each principal component (PC) represents.

A large number of variables used in the PCA method increases the number of effective variables in the exported PCs and makes the interpretation process difficult. It is necessary to conduct further detailed investigations on detecting the less effective variables and optimizing the number of input variables in the analyses.

Acknowledgment

REFERENCES

[1] Yang, J., Cheng, Q. (2015 a). A comparative study of independent component analysis with principal component analysis in geological objects identification, Part I: Simulations. *Journal of Geochemical Exploration*, 149, 127–135.

[2] Davis, J. C. (2002). *Statistics and Data Analysis in Geology*. 3rd ed, JohnWiley & SonsInc, New York.

[3] Cheng, Q., Bonham-Carter, G., Wang, W., Zhang, S., Li, W., Xia, Q. (2011). A spatially weighted principal component analysis for multi-element geochemical data for mapping locations of felsic intrusions in the Gejiu mineral district of Yunnan, China. *Computers & Geosciences*, 37 (5), 662–669.

[4] Wang, W., Zhao, J., Cheng, Q. (2011). Analysis and integration of geo-information to identify granitic intrusions as exploration targets in southeastern Yunnan District, China. *Computers & Geosciences*, 37, 1946–1957.

[5] Zuo, R. (2011). Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal modeling in the Gangdese Belt, Tibet (China). *Journal of Geochemical Exploration*, 111, 13–22.

[6] Yang, J., Cheng, Q. (2015 b). A comparative study of independent component analysis with principal component analysis in geological objects identification, Part II: A case study of Pinghe District, Fujian, China, *Journal of Geochemical Exploration*, 149, 136–146.

[7] Ueki, K., and Iwamori, H. (2017). Geochemical differentiation processes for arc magma of the Sengan volcanic cluster, Northeastern Japan, constrained from the principal component analysis. *Lithos*, 290, 60–75.

[8] Davydenko, A. Y., Grayver, A. V. (2014). Principal component analysis for filtering and leveling of geophysical data. *Journal of Applied Geophysics*, 109, 266–280.

[9] Li, Q., Dehler, S. A. (2015). Inverse spatial principal component analysis for geophysical survey data interpolation. *Journal of Applied Geophysics*, 115, 79-91.

[10] Wang, J., Xiaohong, M., Li, F. (2015). Improved curvature gravity gradient tensor with principal component analysis and its application in edge detection of gravity data. *Journal of Applied Geophysics*, 118, 106–114.

[11] Zhang, Q., Peng, C., Lu, Y., Wang, H., Zhu, K. (2018). Airborne electromagnetic data levelling using principal component analysis based on flight line difference. *Journal of Applied Geophysics*, 151, 290–297.

[12] Ghosh, T., Basu S., Hazra S. (2014). Geological mapping of the Schuppen belt of north-east India using geospatial technology. *Journal of Asian Earth Sciences*, 79, 97–111.

[13] Mandeng, E. P. B., Bidjeck, L. M. B., Wambo, J. D. T., Taku, A., Betsi, T. B., Ipan, A. S., Nfada, L. T., Dieudonné, L. B. (2018). Lithologic and structural mapping of the Abiete–Toko gold district in southern Cameroon, using Landsat 7 ETM+/SRTM. *Comptes Rendus Geoscience*, 350, 130–140.

[14] Fazliani, H., Kamkar-Rouhani, A., Arab-Amiri, A. R. (2019). Fuzzy logic and principal component analysis for mineral potential modeling of epithermal gold-antimony deposits in southern Neyshabur. *11th Symposium of the Iranian Society of Economic Geology*. Ahvaz, Iran.

[15] Ranjbar, H., Hassanzadeh, H., Torabi, M., Ilaghi, O. (2001). Integration and analysis of airborne geophysical data of the Darrehzar area, Kerman Province, Iran, using principal component analysis. *Journal of applied geophysics*, 48, 33–41.

[16] Honarmand, M., Ranjbar, H., Moezifar, Z. (2002). Integration and analysis of airborne geophysics and remote sensing data of sar-cheshmeh area, using directed principal component analysis. *Exploration and Mining Geology*, 11, 43–48.

[17] Ranjbar, H., Honarmand, M., Moezifar, Z. (2003b). Integration and analysis of airborne geophysics and remote sensing data for exploration of porphyry copper deposits in the Central Iranian Volcanic Belt. *Map Asia Conference*.

[18] Dezhong, H., Delian, L., Shuigen, X. (1995a). *Explanatory text of geochemical map of Shamkan, stream sediment survey*. Geological Survey of Iran Press.

[19] Naderi Mighan, N. (1998). *Geological map of Shamkan (1:100000)*. Geological Survey of Iran Press.

[20] Naderi Mighan, N. (1999). *Geological map of Kadkan (1:100000)*. Geological Survey of Iran Press.

[21] Carranza, E. J. M., (2008). *Geochemical anomaly and mineral prospectivity mapping in GIS, handbook of exploration and environmental geochemistry*. Vol. 11, Elsevier, Amsterdam.

[22] Rencher, A. C. (2002). *Method of Multivariate Analysis*. second Edition, Wiley series in probability and statistics.

[23] Sabins, F. F. (1999). *Remote sensing for mineral exploration*. Remote Sensing Enterprises, 1724 Celeste Lane, Fullerton, CA 92833, USA, 157–183.

[24] Ranjbar, H., Honarmand, M., Moezifar, Z. (2003a). Analysis of ETM+ and airborne geophysical data for exploration of porphyry type deposits in the Central Iranian Volcanic Belt, using fuzzy classification. Remote sensing for environmental monitoring, GIS applications, and geology III. *SPIE Conference*, Barcelona, Spain.

[25] Stern, R. J. (1999). Mineral exploration with satellite remote sensing imagery: examples from the Neoproterozoic Arabian-Nubian Shield. *11th International Conference of the Geological Society of Africa*.

[26] Volesky, J. C., Stern, R. J., Johnson, P. R. (2003). Geological control of massive sulfide mineralization in the Neoproterozoic Wadi Bidah shear zone, southwestern Saudi Arabia, inferences from orbital remote sensing and field studies. *Precambrian Research*, 123, 235–247.

[27] Adler-Golden, S., Berk, A., Bernstein, L. S., Richtsmeier, S., Acharya, P. K., Matthew, M. W., Anderson, G. P., Allred, C., Jeong, L. Chetwynd, J. H. (1998). FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. *In Summaries of the seventh JPL airborne earth science workshop.* 9-14.

[28] Costa, L., Nunes, L., Ampatzidis, Y. (2020). A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. *Computers and Electronics in Agriculture*, 172.

[29] Hardcastle, K. C. (1995). Photolineament Factor: A new computer-aided method for remotely sensing the degree to which bedrock is fractured. *Photogrammetric Engineering & Remote Sensing*, 61(6), 739– 747.

[30] Dezhong, H., Delian, L., Shuigen, X. (1995b). *Explanatory text of geochemical map of Kadkan, stream sediment survey*. Geological Survey of Iran Press.

[31] Beus, A. A., Gregorian, S. V. (1975). *Geochemical exploration methods for mineral deposits. Applied Pub.* Ud., Wilmette.

[32] Filzmoser P., Hron K., Reimann C. (2009). Principal component analysis for compositional data with outliers. *Environmetrics*, 20, 621–632.

[33] Salomão, G. N., Figueiredo, M. A., Dall'Agnol, R., Sahoo, P. K., de Medeiros Filho, C. A., da Costa, M. F., Angélica, R. S. (2019). Geochemical mapping and background concentrations of iron and potentially toxic elements in active stream sediments from Carajás, Brazil − implication for risk assessment. *Journal of South American Earth Sciences*, 92, 151–166.

[34] Bishop, M. (1996). *Neural Networks for Pattern Recognition*. MIT Press.

[35] Theodoridis, S., Koutroumbas, K. (1999). *Pattern Recognition*, Academic Press.

[36] Dorri, M. B., Sadeghi, Kh. (2008). *Regional exploration on the Kadkan 1:100000 geological map*. Geological Survey of Iran Press.

[37] Heydari, E., Manaf Nejad, M. S. (2009). *Regional exploration on the Shamkan 1:100000 geological map*. Geological Survey of Iran Press.

[38] Azmi, H. (2011). *Mineral prospecting over an area of 500 km² in different locations of Khorasan Razavi province*. Geological Survey of Iran Press.

[39] Mosier, D. L., Singer, D. A., Moring, B. C., Galloway, J. P. (2012). *Podiform chromite deposits database and grade and tonnage models*, U.S. Geological Survey Scientific Investigations Report 2012–5157, 45 p. and database.

[40] Constantinou, G. (1980). Metallogenesis associated with Troodos ophiolite. In Panayiotou, A. (Ed.), Ophiolites. *Proceedings of the International Ophiolite Symposium*, Nicosia, Cyprus, 663–674.

[41] Yaghubpur, A. and Hassan Nejad A. A. (2006). The spatial distribution of some chromite deposits in Iran, Using Fry Analysis. *Journal of Sciences*, Islamic Republic of Iran, 17(2), 147–152.

[42] Wells, F. G., Cater, F. W., Jr., Rynearson, G. A. (1946). Chromite deposits of Del Norte County, California. *California Division of Mines Bulletin*, 134, 1–76.

[43] Abrams, M. J., Rothery, D. A., Pontual, A. (1988). Mapping in the Oman Ophiolite using enhanced Landsat Thematic Mapper images. *Tectonophysics*, 151, 387–401.

[44] Lipin, B. R. (1984). Chromite from the Blue Ridge Province of North Carolina. *American Journal of Science*, 284, 507–529.

[45] Fazliani, H., Rahimi Pour, Gh. R., Ranjbar, H. (2007). Investigation of mineral zoning in the Kadkan and Shamkan geological sheets using the stream sediment geochemical data. *26th Geoscience Conference*. Geological Survey of Iran, Tehran.

[46] Hedenquist, J. W., Izawa, E., Arribas, A., White, N. C. (1996). *Epithermal Gold Deposits: Styles, Characteristics, and Exploration*. Society of Resource Geology of Japan.

[47] Simmons, S. F., White, N. C., JOHN, D. A. (2005). *Geological Characteristics of Epithermal Precious and Base Metal Deposits*. in: Hedenquist, J. W., Thompson, J. F. H., Goldfarb, R. J., and Richards, J. P., eds., Economic Geology 100th Anniversary Volume: The Economic Geology Publishing Company, 485–522

[48] Taylor, B. E. (2007). *Epithermal gold deposits*. in Goodfellow, W. D., ed., Mineral Deposits of Canada: A Synthesis of Major Deposit-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods: Geological Association of Canada, Mineral Deposits Division, Special Publication No. 5, 113–139.

[49] Sawkins, F. J. (1990). *Metal deposits in relation to plate tectonics*. Berlin, Springer-Verlag, 461.

[50] Sillitoe, R. H., Hedenquist, J. W. (2003). *Linkages between volcano-tectonic settings, ore fluid compositions, and epithermal precious metal deposits*. Society of Economic Geologists Special Publication 10, 315–343.

[51] Bethke, P. M., Rye, R. O., Stoffregen, R. E., Vikre, P. G. (2005). Evolution of the magmatic-hydrothermal acid-sulfate system at Summitville, Colorado: Integration of geological, stable isotope, and fluid inclusion evidence. *Chemical Geology*, 215, 281–315.

[52] Arancibia, G., Matthews, S. J., Cornejo, P., Perez de Arce, C., Zuluaga, J. I., and Kasanevan, S. (2006). $^{40}Ar/^{39}Ar$ and K-Ar

geochronology of magmatic and hydrothermal events in a classic low-sulphidation epithermal bonanza deposit; El Peñon, Northern Chile. *Mineralium Deposita*, 41, 505–516.

[53] Harvey, B., Myers, S., Klein, T. (1999). Yanacocha gold district, northern Peru. *Pacific Rim Congress*, Bali, Indonesia, Australasian Institute of Mining and Metallurgy, Proceedings, 445–459.