International Journal of Mining and Geo-Engineering

IJMGE

IJMGE 53-2 (2019) 123-131

DOI: 10.22059/ijmge.2019.251554.594714

Optimizing grade classification model of mineralized zones using a learning method based on harmony search algorithm

Saeid Esmaeiloghli [a, *], Seyed Hassan Tabatabaei [a], Ahmad Reza Mokhtari [a]

[a] *Department of Mining Engineering, Isfahan University of Technology (IUT), Isfahan, Iran*

Article History:

Received: 30 January 2018,

Revised: 02 February 2019

Accepted: 05 February 2019.

ABSTRACT

The classification of mineralized areas into different groups based on mineral grade and prosperity is a practical problem in the area of optimal risk, time, and cost management of exploration projects. This paper presents a new approach for optimizing the grade classification model of an orebody. The machine learning technique was integrated with a metaheuristic algorithm called Harmony Search (HS) to obtain a proper model for the spatial distribution of the grade classes, while improving the computational cost of the traditional classification methods. The HS is an algorithm inspired by the simulation of the process where a composer tries to harmonize a piece of music. By interpolating the dataset of Cu and Mo concentrations in the surface rock samples taken from the Kooh-Panj mineral prospect, the grid data of the two elements were extracted. Five popular indices were used to estimate the true number of groups in the dataset, which in turn determined two classes as the optimal number of groups. Harmony Search Learning (HSL) was used to classify the grid dataset of Cu and Mo. A comparison of the results of the proposed approach with those of the conventional *k*-means clustering suggested that the use of the HSL method significantly reduced the cost function of the problem (up to 13%). The comparison of the mineralization class derived from the HSL and *k*-means clustering with the borehole locations proved that the results of the HSL were more successful in the accurate estimation of the economic mineralization class identified by the exploratory excavations. In this respect, the HSL technique significantly improved the *k*-means performance by 25%. Furthermore, the results of the HSL were more consistent with the lithological units and the alteration zones involved in ore-forming processes. The use of HS-based learning technique rectified the disadvantages arising from the typical clustering methods regarding the entrapment in local optimums. It also led to the extraction of weak mineralization signals, numerically laid in boundary conditions. This approach can be extended to more than two geochemical variables and can be a valuable tool for the classification of mineralized areas to design and optimally manage mineral exploration projects.

Keywords : *Geochemical data, Grade classification, Harmony Search, Kooh-Panj Cu-Mo deposit, Machine learning*

1. Introduction

One of the most critical strategies concerning the interpretation of geological phenomena related to minerals, on one hand, and the plan for designing the mining operation, on the other hand, is the identification of homogenous spatial domains, which introduce specific geochemical attributes such as hydrothermal alterations or mineralization zones. A conventional method for detecting such spatial dependencies includes the comparison of univariate geochemical maps and the investigation of their compliance with the geological structures in the region. If the number of elements is low, the given approach is accompanied by uncertainty; however, if the number is high, the case gets complicated [1]. The available tools for the simultaneous investigation of such variables include the machine learning methods. The unsupervised aspect of these methods is known as *clustering* or *unsupervised classification*. The main objective of such analyses is the classification of multivariate observations into a number of multivariate homogenous groups, i.e. mapping the observations onto spaces called *centroid* of the group. These centroids represent the information of the entire group, offering an appropriate viewpoint about the spectrum of data [2].

In geochemical data mining, a cluster analysis is a fundamental method in discovering valuable knowledge such as spatial distribution of geo-objects, their geochemical characteristics, the distinction among mineralization targets, and the relationship between geophysical and geochemical anomalies [3]. The main idea of mineral data classification is that the geochemical dataset is a mixture of different processes and phenomena, and machine learning can classify the data into different groups, each of which is related to a specific mechanism. The data of each process is usually located in a particular region with individual grade allocations and may be associated with specific geological and structural properties. When such a classification is made, designing the detailed exploration operations and planning for mining operations are carried out with higher reliability [2].

Due to the lack of big training datasets and unavailability of a priori knowledge about the statistical nature of the processes involved in mineral concentration, unsupervised learning methods play a significant role in discovering knowledge from the exploration datasets. Given the different types of data and applications, various kinds of unsupervised classification methods have been developed, which can generally be categorized as *partitioning* and *hierarchical* methods [4]. The hierarchical methods are commonly used for analyzing exploration datasets and investigating inter-elemental relationships, while the partitioning methods seek to classify the observations (samples) and investigate the spatial distribution of classes and their dependence on the geology of the region [1]. The hierarchical methods have high

* Corresponding author. Tel: +98-9195688303, P.O.Box: 8415683111. *E-mail address:* s.esmaeiloghli@mi.iut.ac.ir (S. Esmaeiloghli).

computational complexity, whereas the partitioning-based methods are more commonly used for big datasets as well as pattern recognition applications [5]. In the same vein, to analyze the mineral data, the partitioning methods have proved their greater information worth, compared to the classification of variables, and have widely been used in various fields, including anomaly separation from background [6–9], recognition of volcanic occurrence patterns [10], classification of joints and fractures [11–14], recognition of hydrothermal alterations [15], and identification of environmental pollutions [6, 16].

There are various classification techniques in the category of the partitioning approaches, such as Gustafson-Kessel [17], fuzzy c-means (FCM) [18], Self-Organizing Map (SOM) [19], and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [20]. Nevertheless, the most practical method in this group is the *k*-means clustering technique [21]. In this method, the samples are assigned to one of the clusters based on the minimum distance from the centers. Thus, *k* number of clusters is formed, each containing a number of samples. Thereafter, by calculating the cluster means, the new center of each cluster is updated, and this process continues until the results converge to the best classification. The *k*-means technique has been widely used for geochemical data processing. Meshkani et al. (2011) [22] used this method for classifying Pb-Zn deposits on a regional scale. Clare and Cohen (2001) [9] as well as Abbaszadeh et al. (2014) [7] employed the *k*-means clustering analysis for the separation of geochemical anomalies from background. Shi and Zeng (2014) [16] used this method to map chemical contaminations zoning. Ghannadpour et al. (2013, 2015) [23, 24] and Ghannadpour and Hezarkhani (2015) [25] employed this method to interpret the geochemical behavior of trace elements in a porphyry deposit. Caciagli (2015) [26] used the *k*-means technique to separate geological processes and geochemical modeling of these processes in an epithermal gold deposit. Abbaszadeh et al. (2015) [15] detected the hydrothermal alterations using *k*-means clustering. Zaremotlagh and Hezarkhani (2016) [27] used this approach as a preprocessing method to classify the REE geochemical data in an iron mine. Morshedy (2016) [28] examined the capabilities of this method for spatial zoning of iron grades. Hood and Cracknell (2017) [29] used the *k*-means analysis to classify the altered rocks in a Cu-Au mine. Mohammadi et al. (2018) [30] employed this technique for multivariate estimation of Au concentrations. Zhou et al. (2018) [31] used the *k*-means method to identify the potential zones of mineralization out of stream sediment geochemical data.

Given the mixing effect of the geochemical datasets, the traditional cluster analysis methods have numerous limitations regarding the processing of big and complex datasets [3]. Partitioning-based classification methods have extensive applications due to their speed and simplicity in knowledge discovery of mineral datasets. However, the initial conditions of the problem (the initial position of the cluster centers) have a great impact on the performance of the algorithm, causing convergence of the final solutions to a local optimum and inability to find the global optimum. This problem is primarily reflected in geochemical datasets more due to the high dimensions of features and samples, generating incorrect and unreal results [9, 32–35]. Over the recent years, researchers have tried to use metaheuristic algorithms to overcome the entrapment of clustering responses in local optimums. However, some of these algorithms have downsides, including low-quality results and lower convergence speed such as the genetic algorithm, while some have a complex structure and lower convergence rate such as particle swarm optimization. This research aims to optimize the classification process of the surface geochemical samples using a metaheuristic algorithm so that the results can represent the real classes in the mineralization area. For this purpose, the Harmony Search (HS) algorithm was used to optimize the unsupervised learning process in the geochemical dataset. The HS algorithm was first introduced by Geem et al. (2001) [36], inspired by the improvisation of musicians. Although the algorithm has been recently introduced and applied, it has been extensively utilized in solving many optimization problems, including optimization of fundamental problems, optimization of the economical load of dispatching problems in power systems, optimization of water distribution networks, improvement of control systems, etc. [37]. The results of Harmony Search Learning (HSL) were compared with the results obtained from the conventional *k*-means clustering both quantitatively, in terms of improving the cost function and intensification of population contrast, and qualitatively, considering the spatial distribution of classes.

2. Study area and dataset

The techniques discussed in this paper were implemented on a geochemical dataset collected from the Kooh-Panj porphyry Cu-Mo deposit. This deposit is located on the Urmia-Dokhtar Magmatic Arc (UDMA), southwest of Kerman Province, Iran. UDMA has a calc-alkaline affinity, and it is the host of many Iranian porphyry copper deposits such as Sarcheshmeh, Meiduk, and Darrehzar [38]. The Kooh-Panj area is mainly covered with volcanic units, including andesite-dacite, andesite agglomerate, tuff, and andesitic dikes, as well as diorite and quartz diorite (Fig. 1) [39–41]. Six major hydrothermal alteration zones can be identified horizontally and vertically around the quartz diorite porphyry stoke of the Kooh-Panj deposit. These zones from the oldest and deepest to the youngest and most superficial are potassic, potassic overlapped with sericitic, propylitic, argillic, sericitic, and advanced argillic [38]. The porphyry Cu-Mo mineralization is the most important mineral occurrence in the Kooh-Panj area, hosted by the quartz diorite porphyry body [40].

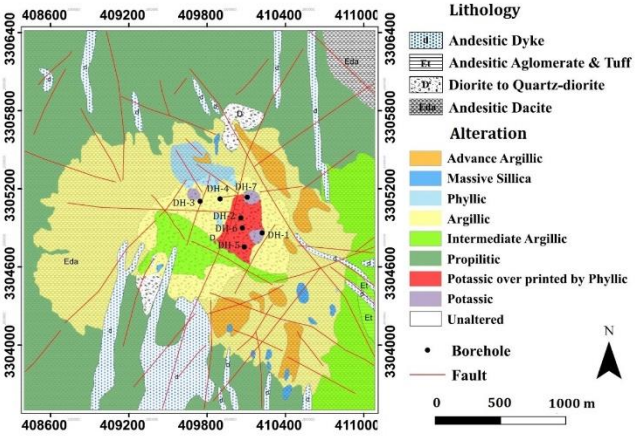

Fig. 1. Simplified geology and alteration map of the Kooh-Panj porphyry Cu-Mo mineralization area [38].

Table 1. Descriptive statistics of Cu and Mo data in the samples taken from the Kooh-Panj deposit area.

	Cu	Mo
Sample size	612	612
Unit	ppm	ppm
Mean	111.47	6.90
Standard deviation	171.21	15.06
Coefficient of variation	0.15	0.21
Skewness	4.97	3.61
Kurtosis	29.56	15.03
Minimum	5.90	0.075
1st quartile	43.52	0.50
Median	59.95	0.95
3rd quartile	100.65	4.50
Maximum	1400.00	110.00

Fig. 2 displays the scatter plot and histograms of Cu-Mo. As seen, the two elements were relatively correlated (the Pearson product-moment correlation coefficient was $r = 0.50$), and their frequency distribution pattern was almost the same (strongly skewed distributions with a positive sign). To produce the continuous final map, the Cu and Mo datasets were separately interpolated using the Kriging method, in a

regular grid with 100 × 100 nodes (cell dimension of 25 m × 25 m). As an output, a dataset containing 10000 geochemical square cells was obtained in terms of the concentrations of Cu and Mo. It should be noted that there were seven exploratory boreholes in this area, indicated with DH codes in Fig. 1. Data from these drills included Cu and Mo analyses in 1550 core samples, merely used in the evaluation of the research results.

The surface geochemical observations in the Kooh-Panj mineralization area were used as the dataset of this study. The dataset contained 612 rock samples, chemically analyzed for 43 geochemical elements, by using a combination of ICP-OES and ICP-MS methods. These samples were taken from a 2400 m × 2400 m square area with a grid spacing of 100 m. The primary mineralization elements (Cu and Mo) were selected as the target elements of the study. Initially, preprocesses operations such as the estimation of censored data and outlier corrections were performed on the dataset. The statistical summary of both elements is described in Table 1.

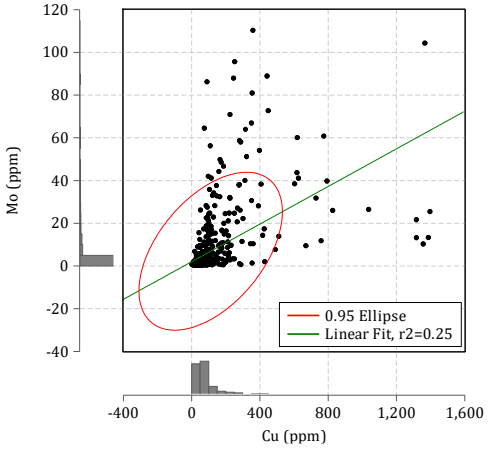

Fig. 2. Scatter plot and frequency distribution of Cu and Mo data in samples taken from the Kooh-Panj deposit area.

3. Harmony Search Learning (HSL)

k-means clustering, traditionally used for classifying geochemical data, tries to appoint multivariate observations into k predetermined groups, so that each group contains n_k observations with the centroid of each cluster as x_k [42]. The classification begins with a random set of k observations as the group centroids. Each observation is allocated to the closest center depending on the distance to the centroids (mostly the Euclidean distance). Now, each x_k is transferred to the cluster mean in the feature space of the data, and then the appointment of each member to the new centroid is reviewed. This process continues until the new centers are not displaced. The mathematical expression of this convergence is stated as within-group distance w, which should be minimized as [9]:

$$w = \sum_{i=1}^{k}\sum_{i=1}^{n_k} \mathrm{dist}^2(x_{ik}, x_k) \quad (1)$$

This function ensures the maximum similarity of the samples to the centroid of the relevant class and the maximum discrimination among the components of distinct classes. The *k*-means problem is strongly influenced by the multivariate geometry of the geochemical dataset [9]. Thus, in this research, the HSL algorithm was used to explore the global optimum solution and to achieve reliable results.

Similar to the optimization problem trying to find a global optimum based on the objective function of the problem, performing music seeks to find a beautiful harmony based on aesthetic sciences. In the music sciences, harmony refers to auxiliary frequencies, which add to the dominant frequencies of the song and sometimes to the beautification. A composer always tries to achieve the best aesthetic state of the song by improvising different harmonies. He performs this using his harmony memory repository, which is created as a result of listening to different pieces. For this purpose, he tries to achieve the optimal form of music by applying minor changes to various harmonies in his memory. In this way, every musician plays each pitch of music within a possible range, all forming a harmony vector altogether. If all pitches aesthetically produce an acceptable harmony, this experience is stored in the memory of the composer, allowing them to create a better harmony in the future. Likewise, in the engineering optimization problems, each decision variable is first assigned a certain value within a permissible range, developing a solution vector altogether. If all the variables produce an acceptable solution based on the objective function, this experience is stored in the variable memory and the probability of achieving better responses in future iterations increases [43].

In terms of aesthetics, the piece played with a specific musical instrument can be assessed using three parameters, including frequency (pitch), resonance (quality), and oscillation range (loudness). The resonance is mainly specified as a waveform, by the harmonic content of harmony and through modulating sound signals. Nevertheless, the generated harmonies are mainly dependent on the pitch of the sound or the frequency spectrum of different instruments. Different notes have various frequencies. For example, the note A4 has the principal frequency of $f_0 = 440$ Hz. Therefore, when adjusting the pitch, we tried to change the frequency [43]. When composers create a piece of music, they have three options as follows [37]:

1) Playing a famous piece, exactly as it is in his memory;
2) Playing a song similar to a famous piece, with some minor modifications;
3) Creating a new piece with new notes.

Formulating the three above items is an optimization process, developed by Geem et al. (2001) [36] through the following steps:

1. Generating Harmony Memory (HM): random vectors of $\{x_1^i, x_2^i, \dots, x_n^i\}$, $i = 1, \dots,$ HMS, each of which used as a solution for the problem, are stored in a HM matrix:

$$\mathrm{HM} = \begin{bmatrix} x_1^1 & \cdots & x_n^1 \\ \vdots & \ddots & \vdots \\ x_1^{\mathrm{HMS}} & \cdots & x_n^{\mathrm{HMS}} \end{bmatrix} \begin{bmatrix} w(x^1) \\ \vdots \\ w(x^{\mathrm{HMS}}) \end{bmatrix} \quad (2)$$

where n represents the number of variables; $w(\cdot)$ shows the objective function of the learning problem (Eq. 1) per harmony vector of $\{x_1^i, x_2^i, \dots, x_n^i\}$, and HMS is the Harmony Memory Size [44]. The HM matrix plays the role of famous pieces of music that the composer intends to play them exactly the same, and each row of HM is a harmony [37].

2. Generating a new harmony: with a probability of Harmony Memory Consideration Rate (HMCR), the new solution of $\{x_1', x_2', \dots, x_n'\}$ is selected from HM matrix. The factor HMCR is defined as the probability of selecting the new component x_j' out of the members stored in the HM, and ensures that the best harmonies are not removed from the memory during the optimization process. In this way, with the probability of $1 - \mathrm{HMCR}$, the new solution is randomly generated from the search space [44]. It is obvious that $0 \le \mathrm{HMCR} \le 1$, so that the $\mathrm{HMCR} \to 0$, then the new harmonies are randomly selected from the search space. On the other hand, as $\mathrm{HMCR} \to 1$, the probability of approving a new harmony from HM is high. The value of this parameter establishes a balance between the components of *exploitation* (high quality of the solutions taken from HM) and *exploration* (a search in the space of local solutions) [37].

3. Frequency modification (pitch adjustment): in the music sciences, pitch adjustment denotes changing the structure and content of frequencies, which equals the production of neighboring solutions in the optimization process. If the new component x_j' is taken from HM, it is mutated with the probability of Pitch Adjustment Rate (PAR) as much as fw, and is replaced with $x_j'' = x_j' + fw.u(-1,1)$. The fw represents fret width (bandwidth), and describes the maximum changes in the pitch adjustment [43]. The term $u(-1,1)$ expresses the positive and negative changes of fw in the form of a uniform statistical distribution. It is obvious that $0 \le \mathrm{PAR} \le 1$, explaining the probability of mutations in the elements taken from HM.

4. Updating HM: the harmony generated from the previous step is compared with the worst harmony of the memory (x^{worst}) in terms of the cost function. If $x_j^{'}$ or $x_j^{''}$ is better than x^{worst}, then it replaces [44].

5. Termination of the optimization: the termination condition of the algorithm is usually achieved by a certain amount of iterations (it_{max}). By creating loops in the previous steps, the HS approaches the optimal value and improves the solutions [44].

Algorithm 1 presents the pseudo-code of the algorithm. In terms of computational complexity, although the *k*-means technique is faster and consumes less memory compared to HSL, the optimization due to the use of Harmony Search algorithm can lead to a significant improvement in the clustering results.

Algorithm 1. The pseudo-code of the Harmony Search algorithm.

STEP 1. Initialize the problem
Load the data related to the optimization problem
Define the objective function w %% the objective function defined in Eq. 1
Set the HS parameters (HMS, HMCR, PAR, fw, it_{max})
STEP 2. Initialize the harmony memory
Construct the vectors of the harmony memory, HM = $\{x^1, x^2, \dots, x^{HMS}\}$
Identify the worst vector in the HM, $x^{worst} \in \{x^1, x^2, \dots, x^{HMS}\}$
STEP 3. Improvise a new harmony vector
$x^{'} = \phi$ %% new harmony vector
for $i = 1, \dots, n$ **do** %% n is the number of variables.
if $R(0,1) \leq$ HMCR **then** %% R is a uniform random number generator.
begin
$x_i^{'} \in \{x_i^1, x_i^2, \dots, x_i^{HMS}\}$ %% harmony memory consideration
if $R(0,1) \leq$ PAR **then**
$x_i^{'} = y_{i,j\pm m}$ %% $x_i^{'} = y_{i,j}$, pitch adjustment
end
else
$x_i^{'} \in \boldsymbol{X}_i$ %% random consideration
end if
end for
STEP 4. Update the harmony memory
if $w(x^{'}) < w(x^{worst})$ **then**
Include $x^{'}$ to the HM
Exclude x^{worst} from HM
STEP 5. Check the termination condition
while (the termination criterion specified by it_{max} is not provided)
Repeat **STEP 3** and **STEP 4**

4. Results and discussion

A Laptop hardware with an operating system of Windows 7 (32 Bit), with 1.66 GHz processor, Core i2, and RAM 2 GB was used to run the HSL program on the grid data of Cu and Mo obtained from the Kooh-Panj deposit. The HSL algorithm was implemented in the MathWorks MATLAB environment. For the objective function of the problem (Eq. 1), an independent module, named fitness, was developed and called as a function in the program interface (with the name of HarmonySearch). The input dataset included the estimated grades of Cu and Mo in 10000 geochemical cells with the dimensions of 25 m × 25 m. To prevent any deviations in the results, the feature vectors of the input matrix should have the same scale [45]. This process was provided through $z_i = (x_i - x_{min})/(x_{max} - x_{min})$ standardization, and the grid data of Cu and Mo was scaled between 0 and 1. After a transposition of the input matrix, the standardized data was loaded into a 2 × 10000 matrix in the HarmonySearch interface.

The essential parameters of HSL algorithm and their calibrated values are provided in Table 2. To implement the HSL, the actual number of classes (k) should first be estimated in the current data space. In this study, five practical indices were used to determine the number of groups in the dataset. These indices combined the available information about between-group separation and within-group compactness, and determined the optimal number of classes [46]. To implement the indices, the Euclidean distance criterion was considered as a measure of distance and the generation of the non-similarity matrix. Furthermore, the possible scenarios of cluster numbers (upper and lower bounds of class numbers) were considered in the interval [2, 10] in order to extract the optimal solution from this range. As shown in Table 3, all the employed indices detected two classes as the optimal number of clusters in the dataset. Meanwhile, the performance of Dindex was based on knee points of the index values and peak points of the second differences of the index values, whose results are graphically shown in Fig. 3.

Table 2. The calibrated parameters of the HSL algorithm for classifying the grid data of Cu and Mo in the Kooh-Panj deposit area.

Parameter	Value	Description
NC	2	Number of classes (k)
maxiter	50	Maximum number of iterations (it_{max})
npop	100	Harmony memory size (HMS)
HMCR	0.90	Harmony memory consideration rate (HMCR)
PAR	0.20	Pitch adjustment rate (PAR)
BW	0.95	Bandwidth (fw)
BW_RF	0.95	Bandwidth reduction factor (fret width damp ratio)

Table 3. The optimal number of groups and the numerical values of the indices performed on the dataset of Cu and Mo in the Kooh-Panj deposit area.

Index	Value of the indices									Optimum number of groups
	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	
Duda index [47]	0.989	1.822	1.502	2.180	1.441	1.497	1.297	2.097	1.752	2
Silhouette coef. [48]	0.813	0.740	0.737	0.689	0.706	0.704	0.615	0.535	0.496	2
Gap statistic [49]	1.023	0.516	0.438	0.168	0.018	-0.104	-0.149	-0.358	-0.364	2
Dindex* [50]	0.068	0.054	0.046	0.041	0.039	0.035	0.032	0.030	0.027	2
Dunn index [51]	0.021	0.005	0.010	0.010	0.010	0.012	0.005	0.003	0.001	2

The HSL algorithm was run across 50 iterations with a memory size of 100 (Table 2). Although choosing a large memory size results in a diminished convergence rate, it allows for achieving a more extensive range of local solutions. Low HMCR values cause slowness of the algorithm convergence because only a limited number of harmonies are selected during the optimization process. On the other hand, high rates make it possible to choose and use just the harmonies contained in the HM. This event, in turn, causes the objective function to be entrapped in the local optimum. The suitable range for variations in the HMCR has been proposed to be [0.75,0.95] [36], and this study employed 0.90. The pitch adjustment rate in the HS is responsible for controlling the selection pressure among the exploration and exploitation parameters. High values of PAR enhance the variety in HSL algorithm. However, excessive growth of this parameter causes the HS to behave as a random search method. In this study, a value of 0.2 was used to calibrate PAR. The bandwidth parameter fw was adjusted as 0.95 of the input range. Furthermore, a factor of 0.95 was used as the bandwidth reduction factor (fret width damp ratio) to linearly reduce the bandwidth of frequency

changes in the subsequent iterations. This value increased the convergence rate in the final stages of the algorithm.

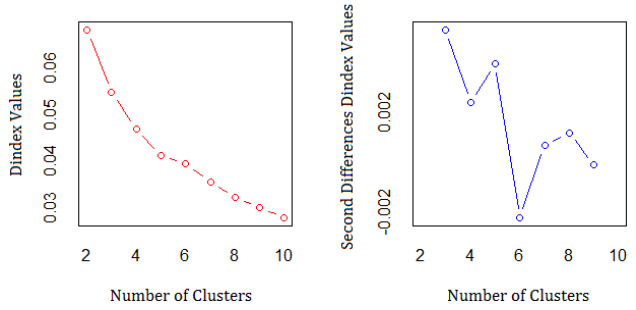

Fig. 3. Plots of Dindex values (left) and the second differences Dindex values (right) against the cluster numbers of the dataset of Cu and Mo in the Kooh-Panj deposit area.

By running the HSL under the conditions provided in Table 2, bivariate pixels of Cu and Mo were classified in Groups A and B. Fig. 4 indicates the variations in the minimum and mean cost of HSL across 50 iterations of the algorithm. As can be observed, the algorithm reached the stall zone of convergence almost at the 25th iteration. To verify the efficiency of HSL-based partitioning, the performance of this technique was compared with that of the traditional *k*-means method. Accordingly, through scripting in the MathWorks MATLAB, the value of the cost function, mentioned in Eq. 1, was calculated using *k*-means clustering, and the data were partitioned into classes A and B. The summary statistics of the classes obtained from the running of the two techniques were separately provided for Cu and Mo data in Table 4. The results confirmed that each of the subpopulations had different statistical properties. As can be seen in Table 4, the classes separated by the HSL method had a lower mean and standard deviation values in comparison to *k*-means clustering. Based on the probability distribution function of Cu and Mo in populations A and B (Fig. 5), it was evident that, despite the reduction and closeness of mean values of the elements in the HSL, this method developed greater density contrast compared to the *k*-means technique. This suggested a higher resolution of the HSL method in classifying mineral grade populations.

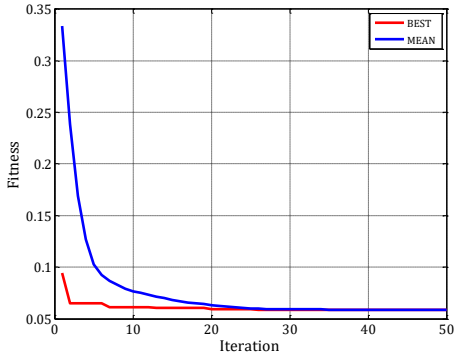

Fig. 4. The minimum and mean convergence plots of the cost function of the HSL algorithm for classifying the grid data of Cu and Mo in the Kooh-Panj deposit area.

Table 4. Statistical parameters of the grid data of Cu and Mo in the classes separated by HSL and *k*-means methods.

	Class A				Class B			
	Cu (km)	Cu (HSL)	Mo (km)	Mo (HSL)	Cu (km)	Cu (HSL)	Mo (km)	Mo (HSL)
Sample size	1060	1701	1060	1701	8940	8299	8940	8299
Unit	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ppm
Mean	436.49	342.04	39.16	31.14	76.36	67.90	3.31	2.18
Std	293.24	266.40	17.70	17.59	51.97	37.01	5.18	3.12
Min	81.25	55.79	9.59	3.30	5.90	5.90	0.075	0.075
Median	372.70	254.84	34.91	25.66	59.37	57.11	0.89	0.79
Max	1409.95	1409.95	98.14	98.14	444.37	378.28	28.75	17.41

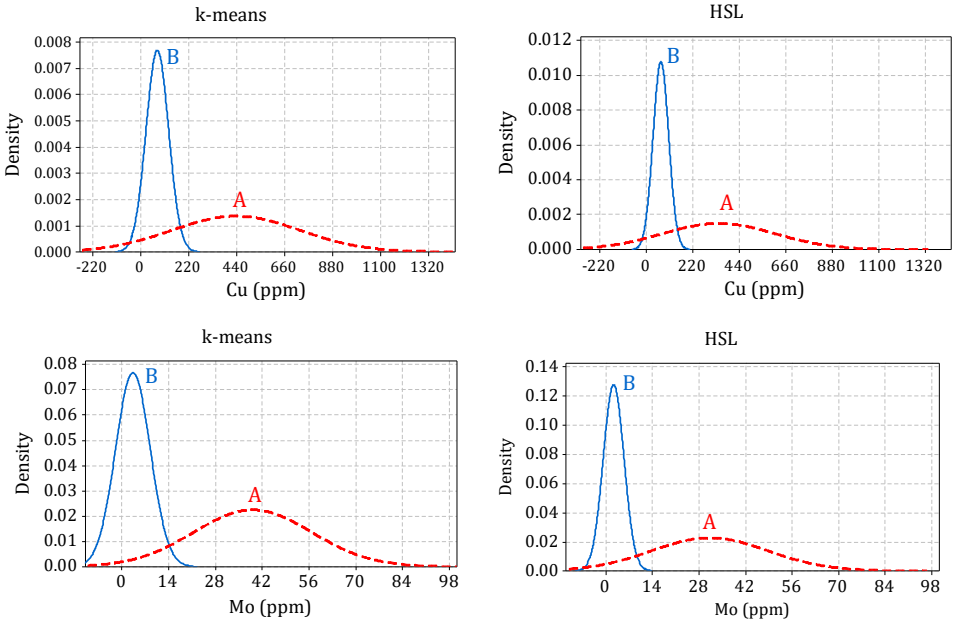

Fig. 5. Probability density function plots of the Cu and Mo data in the classes separated by HSL and *k*-means methods.

Spatial distribution of the classes obtained from the HSL and *k*-means algorithms is demonstrated in Fig. 6. This figure shows that the geochemical cells of each population were spatially distributed in a meaningful way. It was expected that clustering the geochemical cells into groups A and B categorized the mineralization map of the region into two economic and non-economic components, respectively. Accordingly, the HSL method significantly expanded the area of economic mineralization compared to *k*-means clustering. This subsequently led to the identification of weaker Cu-Mo mineralization signals, such as those observed in the marginal regions of the economic mineralization zone.

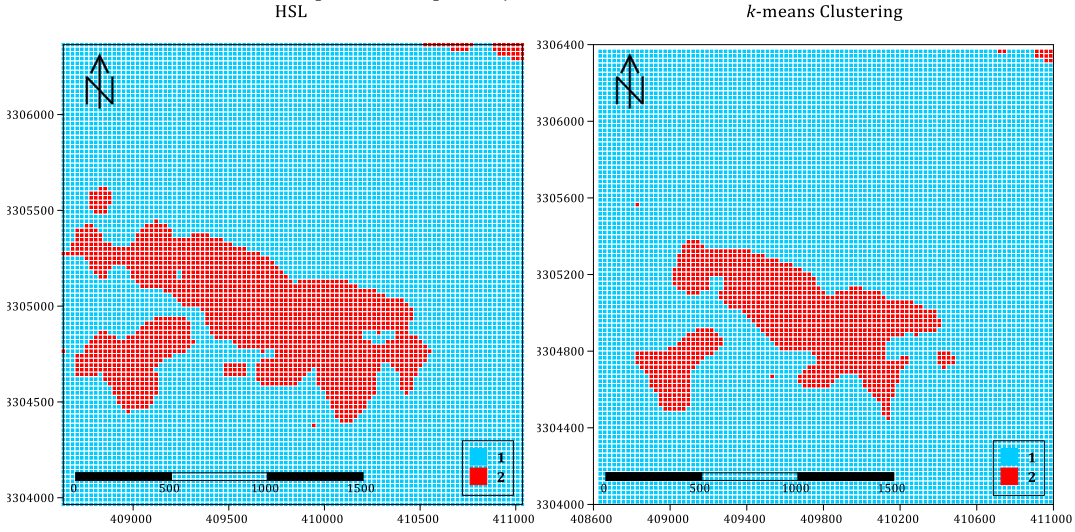

Fig. 6. The Cu-Mo classification map derived from (a) HSL and, (b) *k*-means clustering in the Kooh-Panj deposit area. The color of the classes is in line with Fig. 5.

Obviously, by optimizing the classification process, the value of the cost function decreases as described in Eq.1. Considering two classes for clustering 10000 standardized Cu and Mo data, both the HSL and *k*-means methods were run for 20 iterations, and the cost value of each technique was recorded in each run. The statistical characteristics of the results of these two ways are summarized in Table 5. As can be seen in the table, the worst cost of the HSL was still better than the best cost of *k*-means clustering. Furthermore, the mean cost of the HSL was far less than that of *k*-means, explicitly confirming the better performance of the HS-based learning method. Regarding the mean costs resulting from the two techniques, due to the use of the HSL, the rate of improvement in clustering can be estimated to be around 13%.

Table 5. Statistical characteristics related to the cost values of the Cu and Mo data classification using HSL and *k*-means clustering.

	k-means	HSL
Number of iterations	20	20
Minimum cost	673.59	587.73
Maximum cost	673.87	588.79
Mean cost	673.70	587.99

Table 4 also represents population a separated by the HSL and *k*-means methods as an economic mineralization class in the Kooh-Panj area. To investigate the spatial coherence of this zone with the geological units of the Kooh-Panj porphyry system, the spatial distribution map of class A was plotted by both HSL and *k*-means methods, corresponding against the geological map of the deposit (Fig. 7). As can be observed in the figure, although the centers of the zones identified by both methods were similar and corresponded to the diorite unit of the area, the economic zone identified by HSL was significantly wider than the region detected by *k*-means. The HSL technique also identified weak anomalous signals in the margin of the system. Investigating the characteristics shown in Fig. 7 revealed that zone A, derived from HSL, covered a wider area of diorite and quartz diorite units altered by potassic and argillic zones, compared to the findings obtained *k*-means. Seven exploratory boreholes were drilled in the Kooh-Panj deposit area, many of which tended to cut the economic orebody, and others only identified the waste. Table 6 listed the excavated boreholes along with the detected mineralization class in each borehole. In this study, the condition of economic mineralization was that the mean value at the upper 50 m of the boreholes had a minimum of Cu ≥ 0.2% and Mo ≥ 30 ppm. Accordingly, as can be observed in Fig. 7, the superiority of the HSL technique to *k*-means clustering became evident in identifying the mineralization class of DH-4 and DH-7. These boreholes were in the boundary conditions regarding the mineral content. The *k*-means method classified both boreholes as non-economic cases, while the HSL correctly recognized them as the drilling targets by passing them through the local optimum of the problem. As shown in Table 7, for proper identification of the deep mineralization class, the *k*-means and HSL methods had a total accuracy of 57% and 71%, respectively. Therefore, the use of HS-based learning improved the performance of the traditional *k*-means approach by 25%.

Table 6. The quality of mineralization detected by the exploratory boreholes based on the mean concentrations of Cu-Mo at the upper 50 m level of the Kooh-Panj deposit area.

	DH-1	DH-2	DH-3	DH-4	DH-5	DH-6	DH-7
Mean of Cu (%)	0.10	0.33	0.25	0.20	0.09	0.26	0.13
Mineralization class of Cu	N*	E*	E	E	N	E	N
Mean of Mo (ppm)	10.15	34.72	48.77	31.12	25.72	28.20	34.16
Mineralization class of Mo	N	E	E	E	N	N	E

*** E: Economic**
*** N: Non-economic**

Table 7. The success rate of the *k*-means and HSL methods in determining the correct non-economic and economic Cu-Mo mineralization zones in the Kooh-Panj deposit area.

	Count	Correct detection of the k-means	Correct detection of the HSL
Non-economic boreholes of Cu	3	2	1
Non-economic boreholes of Mo	3	1	1
Success rate		50%	33%
Economic boreholes of Cu	4	3	4
Economic boreholes of Mo	4	2	4
Success rate		62%	100%
Overall success rate		**57%**	**71%**

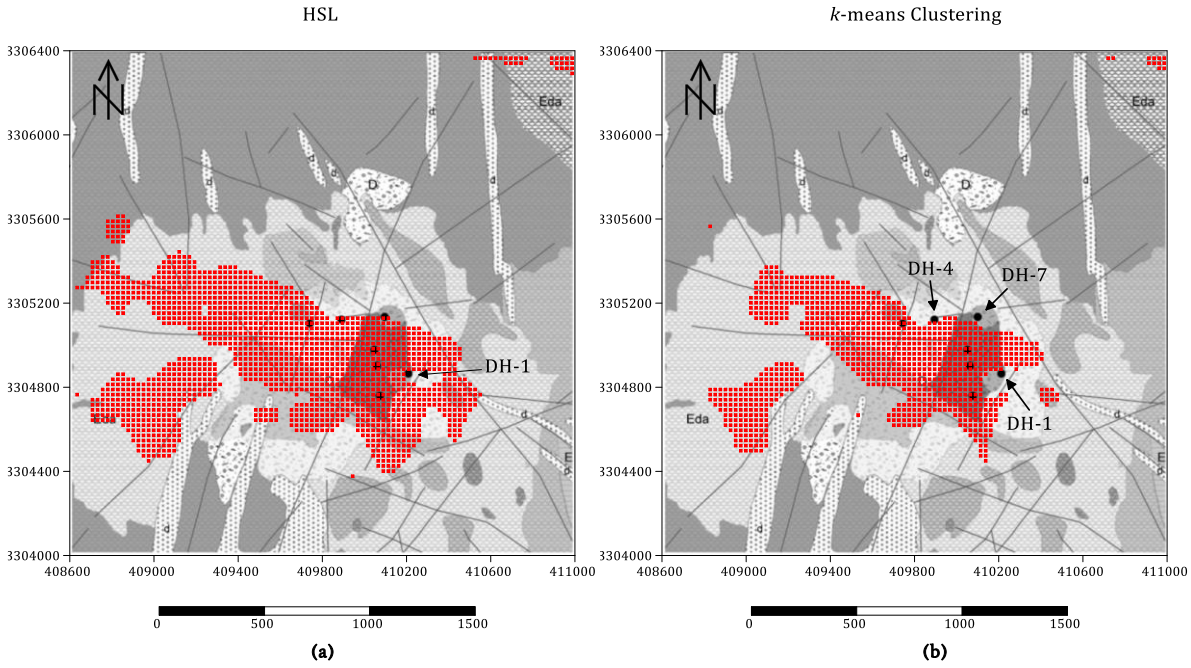

Fig. 7. The economic Cu-Mo mineralization map derived from (a) HSL and, (b) *k*-means clustering, and its relation to the geological structures in the Kooh-Panj deposit area. The symbols are in line with Fig. 1.

5. Conclusion

Similar to other metaheuristic algorithms, such as the genetic algorithm and the particle swarm optimization, the HS algorithm is also a random search method. This algorithm does not require the initial range of information, such as the gradient of the objective function. Furthermore, unlike other population-based stochastic algorithms, the HS uses only one search memory, which leads to its simplification and enhances its efficiency to cope with local minimums. The HS algorithm has a higher rate of convergence compared to similar algorithms, such as the genetic algorithm. Another advantage of this method is that, unlike the genetic method which uses two solution vectors in each generation,, in order to create a new solution the HS algorithm employs all the available solutions in the memory. This feature enhances the flexibility of the algorithm in searching for better solutions.

Despite their extensive use for classifying homogeneous mineral fields, traditional machine learning methods, such as *k*-means clustering, involve many disadvantages, which are often due to entrapment of the objective function in local optimums. However, metaheuristic algorithms are more successful in achieving the global optimum as they enjoy intelligent computing. Therefore, hybridizing the traditional learning problems with a metaheuristic optimizer decreases the entrapment probability of the solutions in the local minimum. This can help to achieve real solutions and to discover correct spatial relationships in the mineral classification model. Theoretically, the application of the HSL for partitioning the Cu-Mo data in the Kooh-Panj deposit area improved the cost function of the learning problem compared to the typical *k*-means clustering (up to 13%). Practically, the use of the HSL also led to more accurate extraction of the mineralization patterns in the region. By preventing the objective function from approaching the local minimums, this technique was able to identify the weak signals of the economic mineralization in the marginal areas of the Cu-Mo mineralization zone. Meanwhile, the use of *k*-means for this purpose led to the incomplete modeling of the mineralization zone and the loss of the part of the economic mineralization class. The success rate of the *k*-means method in the correct identification of deep mineralization classes was improved by 25% using the HSL method. The strategy introduced in this research can be expanded to more than two geochemical variables. The expansion of HSL and its usage, instead of other computational algorithms based on the objective function, can lead to the modification of the mineral grade classification model and optimal alterations in the exploration design for detailed operations.

REFERENCES

[1] Templ, M., Filzmoser, P., & Reimann, C. (2008). Cluster analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry*, *23*(8), 2198–2213.

[2] Ellefsen, K. J., & Smith, D. (2016). Manual hierarchical clustering of regional geochemical data using a Bayesian finite mixture model. *Applied Geochemistry*, *75*(1), 200–210.

[3] Meng, H. D., Song, Y. C., Song, F. Y., & Shen, H. T. (2011). Research and application of cluster and association analysis in geochemical data processing. *Computers & Geosciences*, *15*(1), 87–98.

[4] Rousseeuw P. J., & Kaufman, L. (1990). Finding Groups in Data. Wiley Online Library.

[5] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *31*(3), 264–323.

[6] Vriend, S. P., van Gaans, P. F. M., Middelburg, J., & de Nijs, A. (1988). The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: examples from Portugal. *Applied Geochemistry*, *3*(2), 213–224.

[7] Abbaszadeh, S., Rahimipour, G., & Najmadini, M. (2014). Recognition of Cu-Porphyry Mineralization Areas by Using One and multivariate integration methods on drainage geochemical data in Ghale Askar area, Kerman province. *Journal of Analytical and Numerical Methods in Mining Engineering*, *3*(6), 69–82.

[8] Shiva, M., Aryafar, A., & Zaremotlagh, S. (2011). Fuzzy c-means cluster analysis, a robust multivariate technique in stream sediment geochemical exploration, a case study in Eastern part of Iran, Birjand. *Journal of Geology and Mining Research*, *3*(1), 1–6.

[9] Clare, A. P., & Cohen, D. R. (2001). A comparison of unsupervised neural networks and k-means clustering in the analysis of multi-element stream sediment data. *Geochemistry: Exploration, Environment, Analysis*, *1*(2), 119–134.

[10] Corsaro, R. A., Falsaperla, S., & Langer, H. (2013). Geochemical pattern classification of recent volcanic products from Mt. Etna, Italy, based on Kohonen maps and fuzzy clustering. *International Journal of Earth Sciences*, *102*(4), 1151–1164.

[11] Hammah, R. E., & Curran, J. H. (1998). Fuzzy cluster algorithm for the automatic identification of joint sets. *International Journal of Rock Mechanics and Mining Sciences*, *35*(7), 889–905.

[12] Sirat, M., & Talbot, C. J. (2001). Application of artificial neural networks to fracture analysis at the Äspö HRL, Sweden: fracture sets classification. *International Journal of Rock Mechanics and Mining Sciences*, *38*(5), 621–639.

[13] Marcotte, D., & Henry, E. (2002). Automatic joint set clustering using a mixture of bivariate normal distributions. *International Journal of Rock Mechanics and Mining Sciences*, *39*(3), 323–334.

[14] Jimenez-Rodriguez, R., & Sitar, N. (2006). A spectral method for clustering of rock discontinuity sets. *International Journal of Rock Mechanics and Mining Sciences*, *43*(7), 1052–1061.

[15] Abbaszadeh, S., Ansari, A., & Rahimipour, G. R. (2015). Determination of Cu porphyry hydrothermal alterations using multivariate analysis methods on lithogeochemical data in Hararan area, Kerman province. *Journal of Analytical and Numerical Methods in Mining Engineering*, *5*(10), 15–26.

[16] Shi, W., & Zeng, W. (2014). Application of k-means clustering to environmental risk zoning of the chemical industrial area. *Frontiers of Environmental Science & Engineering*, *8*(1), 117–127.

[17] Gustafson, D. E., & Kessel, W. C. (1979). Fuzzy clustering with a fuzzy covariance matrix. in *Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on*, 761–766.

[18] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2), 191–203.

[19] Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

[20] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. in *Kdd*, *96*(34), 226–231.

[21] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(14), 281–297.

[22] Meshkani, S. A., Mehrabi, B., Yaghubpur, A., & Alghalandis, Y. F. (2011). The application of geochemical pattern recognition to regional prospecting: A case study of the Sanandaj–Sirjan metallogenic zone, Iran. *Journal of Geochemical Exploration*, *108*(3), 183–195.

[23] Ghannadpour, S. S., Hezarkhani, A., & Sabetmobarhan, A. (2015). Some statistical analyses of Cu and Mo variates and geological interpretations for Parkam porphyry copper system, Kerman, Iran. *Arabian Journal of Geosciences*, *8*(1), 345–355.

[24] Ghannadpour, S. S., Hezarkhani, A., & Farahbakhsh, E. (2013). An investigation of Pb geochemical behavior respect to those of Fe and Zn based on k-Means clustering method. *Journal of Tethys*, *1*(4), 291–302.

[25] Ghannadpour, S. S., & Hezarkhani, A. (2015). Lead and zinc geochemical behavior based on geological characteristics in Parkam Porphyry Copper System, Kerman, Iran. *Journal of Central South University*, *22*(11), 4274–4290.

[26] Caciagli, N. (2015). Multielement Geochemical Modelling for Mine Planning: Case Study from an Epithermal Gold Deposit. in *International Workshop on Compositional Data Analysis*, 45–61.

[27] Zaremotlagh, S., & Hezarkhani, A. (2016). A geochemical modeling to predict the different concentrations of REE and their hidden patterns using several supervised learning methods: Choghart iron deposit, bafq, Iran. *Journal of Geochemical Exploration*, *165*, 35–48.

[28] Morshedy, A. H. (2016). Deposit zoning based on the spatial distribution of ore grade using self-organizing map clustering algorithm-Case study: Choghart iron deposit. *Iranian Journal of Mining Engineering*, *11*(32), 73–86.

[29] Hood, S. B., & Cracknell, M. J. (2017). Relating altered rocks to equivalent protoliths using clustering and classification of geochemical data at the Minto copper-gold mine, Yukon, Canada. in *SEG 2017 Ore Deposits of Asia: China and Beyond*.

[30] Mohammadi, N. M., Hezarkhani, A., & Maghsoudi, A. (2018). Application of K-means and PCA approaches to estimation of gold grade in Khooni district (central Iran). *Acta Geochimica*, *37*(1), 102–112.

[31] Zhou, S., Zhou, K., Wang, J., Yang, G., & Wang, S. (2018). Application of cluster analysis to geochemical compositional data for identifying ore-related geochemical anomalies. *Frontiers of Earth Science*, *12*(3), 491–505.

[32] Selim, S. Z., & Ismail, M. A. (1984). K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(1), 81–87.

[33] Garey, M. R., Johnson, D., & Witsenhausen, H. (1982). The complexity of the generalized Lloyd-max problem (corresp.). *IEEE Transactions on Information Theory*, *28*(2), 255–256.

[34] Cheung, Y. M. (2003). k*-Means: A new generalized k-means clustering algorithm. *Pattern Recognition Letters*, *24*(15), 2883–2893.

[35] Hamerly, G., & Elkan, C. (2002). Alternatives to the k-means algorithm that find better clusterings. in *Proceedings of the eleventh international conference on Information and knowledge management*, 600–607.

[36] Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation*, *76*(2), 60–68.

[37] Wang, X., Gao, X. Z., & Zenger, K. (2015). An introduction to harmony search optimization method. Springer.

[38] Gonbadi, A. M., Tabatabaei, S. H., & Carranza, E. J. M. (2015). Supervised geochemical anomaly detection by pattern recognition. *Journal of Geochemical Exploration*, *157*, 81–91.

[39] Khosravi, A. (2007). Geological and Alteration Map of Kuh Panj Copper Deposit. *Exploration Department of National Iranian Copper Industries Company, International Report*.

[40] Roshani, P., Mokhtari, A. R., & Tabatabaei, S. H. (2013). Objective based geochemical anomaly detection—application

of discriminant function analysis in anomaly delineation in the Kuh Panj porphyry Cu mineralization (Iran). *Journal of Geochemical Exploration*, *130*, 65–73.

[41] Mokhtari, A. R. (2014). Hydrothermal alteration mapping through multivariate logistic regression analysis of lithogeochemical data. *Journal of Geochemical Exploration*, *145*, 207–212.

[42] Roy, A. (1981). Application of cluster analysis in the interpretation of geochemical data from the Sargipalli lead-zinc mine area, Sundergarh district, Orissa (India). *Journal of Geochemical Exploration*, *14*, 245–264.

[43] Yang, X. S. (2009). Harmony search as a metaheuristic algorithm. in *Music-inspired harmony search algorithm*, Springer, 1–14.

[44] Gao, X. Z., Govindasamy, V., Xu, H., Wang, X., & Zenger, K. (2015). Harmony search method: theory and applications. *Computational Intelligence and Neuroscience*, *2015*, 39.

[45] Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis, *5*(8). Prentice hall Upper Saddle River, NJ.

[46] Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(i06).

[47] Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis, *3*, Wiley New York.

[48] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

[49] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

[50] Lebart, L., Piron, M., & Morineau, A. (2000). Statistique exploratoire multidimensionnelle. Dunod.

[51] Dunn†, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*(1), 95–104.