Geochemical Pattern Recognition for Cu-Au Deposit Based on Self-Organizing Map (SOM) and Fuzzy K-means Clustering (FKMC) in Meshginshahr, NW of Iran

Aynur Nasseri[1*]
[1]Assistant Professor, Department of Mining Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran
Corresponding Author E-mail: aynur.nasseri@gmail.com & Ay.Nasseri@iau.ac.ir

Abstract
Mapping the mineralized zones and providing an appropriate distribution pattern of elements for characterizing geochemical system and targeting potentially promising areas of Cu-Au mineralization by utilizing an adequate technique and establishing an optimized exploration tool is the main object of this study in Meshginshahr, NW of Iran. In this respect 144 stream sediments samples were collected and analyzed for Au, Ba, Bi, Cd, Ce, Co, Cr, Cu, Hg, Mo, Ag, As, Sn, Sb, W and Pb. In this study, self-organizing map (SOM) and Fuzzy K-means clustering (FKMC) approaches with the aim of pattern recognition were employed. The SOM as a dimension reduction approach was introduced to recognize geochemical dispersion patterns with high certainty while preserving the originality of data.. During data processing, SOM appropriate structure with a pattern including six clusters was selected and the related elements distribution model was extracted. Results represent two significant sets of elements in clusters for anticipating the mechanism of distribution. In this target pattern, copper and pertaining trace elements formation are localized in the north of the area. Also, Au Anomalies and its associated elements are mostly elongated from NW to SW of the area. To evaluate the SOM results, a comparative study was carried out with the results obtained from Fuzzy K-means clustering (FKMC). FKMC performance showed the proper compliance with the SOM results with respect to the relationship between the elements and their corresponding membership's probabilities in different clusters. The results illustrated higher performance of the approaches in characterizing geochemical pattern and detecting the element paragenetic sequence in the area for locating the exploration targets..

Keywords: geochemical pattern recognition, elements distribution, Self-Organizing Maps (SOMs), Fuzzy K-means clustering (FKMC) , Meshginshahr.

1-Introduction

Recognition of geochemical dispersion patterns and paragenetic interrelationship between different elements and the separation of their populations can be effective to more accurately identify of anomalies of ore-related elements in the geochemical environments. The proper mapping of the genetic accumulation of the elements and associated minerals is an important problem in the prospecting of metals in an area. In most cases, due to the large scale of data and their empirical analysis, the lack of proper models for describing the source, deposition, and the concentration of different elements in rocks leads to conceal prone areas and influence the factors associated with mineralization components in the area. Therefore, it is essential to have geochemical models to illustrate the nature and origin of the surface expression of mineralization. Such models attempt to express the spatial and genetical relationship between geochemical dispersion process and to apply them for optimizing of exploration procedures (Butt, 2005). Thus, this paper has been introduced with the aim of enhancing the optimal recognition of geochemical patterns with constraining the prospecting uncertainty in order to map the mineral potentials in the Meshginshahr, Ardebil, NW of Iran. In this respect, integrating and analyzing of different spatial data using various modeling techniques is essential. Nowadays, different modeling techniques have been used in mineral exploration programs (Grunsky and Agterberg, 1992; Bellehumeur et al., 1994;Grunsky, 2010), among them, the techniques associated with the identification of various exploration patterns are of great importance. In this respect, it can be mentioned to the Bayesian decision-making theory (Porwal et al., 2003), non-linear Kernel methods(Al-Anazi and Gates, 2010; Schölkoph et al., 2000; Zuo and Carranza, 2011), Neural networks based methods (Porwal et al., 2003) and Multivariate statistical methods (David and Woussen ,1973; Cazes ,1970; Reimann et al., 2002; Lindqvist et al.,1987; Javid, et al.,2015; Nasseri, et al., 2015). As well as different methods also have developed for identifying patterns and clustering variables to reduce the dimensions of the data matrix. It is essential to note that some aforementioned methods, despite their desirability, adhere to the statistical assumptions about data distribution which affects the nature of the exploration data and the final results. Therefore, in order to overcome this problem, a self-organizing map neural network (SOM) (Kohonen, 1997, 2001) was proposed for recognizing geochemical patterns in the area. Also, the application of this method is important in overcoming the limitations of neural network input variables, in improving the accuracy of predictions in exploratory environments and in recognizing & analyzing the nonlinear multi-dimensional spaces in order to propose an optimal hybrid method in identifying the geochemical patterns. In this respect, the assessment of the elements dispersion, enrichment, and trend of their linear and nonlinear regional variations in the area, was illustrated by using the proposed method.

The Self-Organizing Map (SOM) is a fairly well-known neural network and indeed one of the most popular unsupervised learning algorithms which has been introduced by Finnish Professor Teuvo Kohonen in the early 1980s. The method performs a non-linear projection of multi-dimensional data onto a two-dimensional map. The mapping is topology-preserving from an input space onto the 2-D grid of map units. It means that the more alike two data samples are in the input space, the closer they will appear together on the final map. The resulting maps comprehensively visualize natural groupings and relationships in the data. The method as a remarkable tool in information visualization has a simple basic implementation, capability in data structure maintaining, reliable results and the algorithm scales exceptionally well (Schatzmann, et al., 2003).

From the SOM training point of view, it can express that the SOM consists of a regular, usually two-dimensional grid of map units. Each unit i is presented by a prototype vector $m_i = [m_{i1},...,m_{id}]$, where d is the input vector dimension. The units are connected to adjacent ones by a neighborhood relation. The SOM is trained iteratively and at each training step, a sample vector x is randomly selected from the input data set. Distances between x and all the prototype vectors are computed. The best matching unit (BMU), which is denoted here by b, is the map unit with prototype closest to x

$$\|x - m_b\| = \min\{\|x - m_i\|\} \qquad (1)$$

Then, the prototype vectors are updated. The BMU and its topological neighbors are moved closer to the input vector in the input space. The updated rule for the prototype vector of unit i is:

$$m_i(t+1) = m_i(t) + \alpha(t)h_{bi}(t)[x - m_i(t)],$$

$$h_{bi}(t) = \exp\left(-\frac{\|\tau_b - \tau_i\|^2}{2\sigma^2(t)}\right) \qquad (2)$$

$t : time$, $\alpha(t) : adoption\ coffecient$,

h_{bi} : *neighborhood* ker*nel centered on the winner unit*

τ_b, τ_i : *positiones of neurons b and i on the SOM grid*

For a discrete dataset and fixed neighborhood kernel, the error function of SOM can be shown to be:

$$E = \sum_{i=1}^{N}\sum_{j=1}^{M} h_{bj} \|x_i - m_j\|^2 \quad ,$$

N : *number of training sample* , (3)

M : *number of map units*

Neighborhood kernel h_{bi} is centered at unit b, which is the BMU of vector x_i, and evaluated for unit j .If neighborhood kernel value is one for the BMU and zero elsewhere, this leads to minimization of the error function (Vesanto and alhoniemi, 2000).

SOM has been successfully applied in a broad spectrum of research areas ranging from speech recognition to financial analysis and mineral exploration. For instance, it has been used to solve the prediction problems in neural networks (Kaski, 1997, Pastukhov, A.A. and Prokofiev, A.A., 2016). It also as a most reliable clustering method in spatial data analysis has widely been applied in various fields such as image processing (Bação, et al., 2005), precipitation estimation (Kalteh, et al., 2008 , 2007; Liu, et al., 2011; Nourani, et al., 2013; Hsu, et al., 2002), biology and ecology (Chon, 2011; Park et al. 2006), underground water investigation (Peeters, et al. 2007; Lin and Chen. 2005), in seismic attributes assessment and interpretation (Strecker and Uden. 2002), well logging (Baldwin, et al. 1990; Nasseri et al., 2017), in agriculture (Jianwen, and Bagan. 2005), hydrology (Herbst, et al. 2009; Kalteh, et al., 2008; Hsu, et al., 2002), catchment basin classification (Ley,et al. 2011; Céréghino, et al., 2001), remote sensing hyperspectral image processing(Lin, et al., 2005; Lindqvist, et al., 1987; Lin, et al., 2011; Lee, et al., 2006; Martinez, et al., 2001; Matteoli, et al., 2010; Neagoe, et al., 2002; Patil, et al., 2011; Tasdemir, et al., 2009; Toivanen, et al., 2003; Villmann, et al., 2003), landslide zoning (Hentati et al., 2010), metal distribution evaluation in soil and sediments (Löhr et al., 2010), mineral exploration (Caneiro, et al., 2012; Fraser, et al., 2012,2006; Marroquin, et al., 2005; Cracknell, et al.2015,2014; Fraser, et al., 2007; Choi, et al., 2014; Brehme, et al., 2017; Sun, et al., 2009) and most other fields. Despite of SOM applications in various fields, there are still limited implementations in earth sciences. Therefore, this study intends to address the shortcomings of the research history in applying of the SOM method in mineral exploration and expresses its practical application in the field of geochemical exploration for mapping promising areas of probable copper mineralization in the Meshgin shahr area of Ardebil province. For assessing the effectiveness of SOM in pattern recognition in the study area, we employed Fuzzy k-means clustering to the geochemical data.

Fuzzy K-means clustering creates homogeneous groups of clusters described by a set of quantitative variables. If the dataset contains multiple groups which are too close, it is possible to introduce a coefficient of fuzziness that allows each observation to be linked to each group with a probability of membership. Accordingly, the cluster with high membership will be significant. Moreover, in this model using fuzzy distribution of elements within clusters define all elements' properties without ignoring any. Also, the fuzzy distribution detects element behavior patterns faster than normal clustering of elements. Additionally, it recognized elements with multi patterns behavior.

Basically, Fuzzy k-means is a generalization of the classical k-means (De Gruijter and McBratney, 1988) where each observation is associated to each cluster with a probability μ_{ij}. A starting point is chosen by associating the K centers to k observations (randomly or not). Then the distances between the observations and the centers are computed. Next, the membership probably $\mu_{i,j}$ is computed for each observation i and each center j, and the each center C_j is updated using the membership probability and the fuzzy coefficient m as follows:

$$\mu i,j = \frac{\overline{\frac{1}{\omega_i d(X_i, C_j)}}^{\frac{1}{m-1}}}{\sum_{i=1}^{k} \overline{\frac{1}{\omega_i d(X_i, C_j)}}^{\frac{1}{m-1}}} ,$$

$$Cj = \frac{\sum_{i=1}^{N}\sum_{j=1}^{k} \omega_i \mu_{i,j}^{m} X_i}{\sum_{i=1}^{N}\sum_{j=1}^{k} \omega_i \mu_{i,j}^{m}} \tag{4}$$

The higher the coefficient is, the fuzzier the borders of the clusters are. When m value is 1 or close to 1, it means the point is closet to cluster center and more weight is given to that point.

Previous studies using the fuzzy k-means clustering algorithm have been applied to petrophysical logs from boreholes in the Sudbury area to characterize rock types (Mahmoodi, 2016). However, geological differences (mineralogy, textures, alteration) amongst different lithologies are important in placing data into clusters, collecting samples and observing the geological changes helps in understanding why there is heterogeneity in the data. Furthermore, since there can be an overlap of chemical properties amongst different lithologies, the samples can belong to more than one cluster and this leads to reduce the accuracy of predicting the rock type from the cluster information. Therefore, this study used the fuzzy k- means algorithm to overcome the problems and to identify how chemical changes can affect the measurements and how the related patterns can be recognized accurately.

In this study it was attempted to combine the positive aspects of SOM and Fuzzy k-means clustering to avoid forcing every elements and samples into a single cluster along with considering the linear and nonlinear relationship between behavior features of elements, deduce feature values similarity within elements. Thereby, the epigenetic related elements with multiple behavior patterns can be highlighted.

1-1 Study area

The study area is situated in Meshginshahr 1:100000 sheet with a longitude of (47°30'18") to (47°36'28") and latitude of (38°14'35") to (38°19'3") in Ardebil province, NW of Iran. (Fig. 1). According to the division of structural - sedimentary zones of Iran (Aghanabati, 2004), Meshginshahr sheet is located in the northwest of Alborz- Azerbaijan zone. According to lithological similarity of Arasbaran-Garadagh (Iran NW-East Azarbaijan) and the presence of similar indices and deposits in the area, it is essential to study more accurately and recognize optimal exploration patterns for targeting mineral potentials in the area. Therefore, it is important to evaluate the major rock units that are exposed in the area and are involved in the evolution of the rocks of the area. The igneous, pyroclastic and Cenozoic sediments cover more than 95% of the area and from the old to the new are Eocene, Oligocene and Quaternary alluvial deposits. Eocene volcanic rocks, pyroclastic and sedimentary rocks are expanding in the northeastern and western parts of the study area and their composition is as trachyandesite, trachy basalt and trachyte and the parts with acidic composition have less expansion.

1-2-Sampling

The sampling points on the geological map are presented in Fig. 1. Choosing the sampling environment and appropriate parts of the stream sediment for analyzing is a crucial problem in achieving the best possible contrast. The main objective is to increase the contrast and reduce the natural variation within the communities of background and anomalies with low cost and a good level of confidence. A total of 144 stream sediment samples were collected, especially from silica and clay components and prepared up to -120 mesh for the analysis of As, Au, Ba, Bi, Cd, Ce, Co, Cr, Cu, Hg, Mo, Pb, Sb, Sn and W elements and mapping secondary halos. The analytical method for the elements Au, Pb, Ag, Sn is the emission spectrograph (Es); for Hg, Bi, Sb, As, Atomic fluorescence (AF); polarography (POL) for Mo, W; atomic absorption (AA) for Cd; ICP-OES for other elements and oxides which was done in Karaj Applied Research Center.

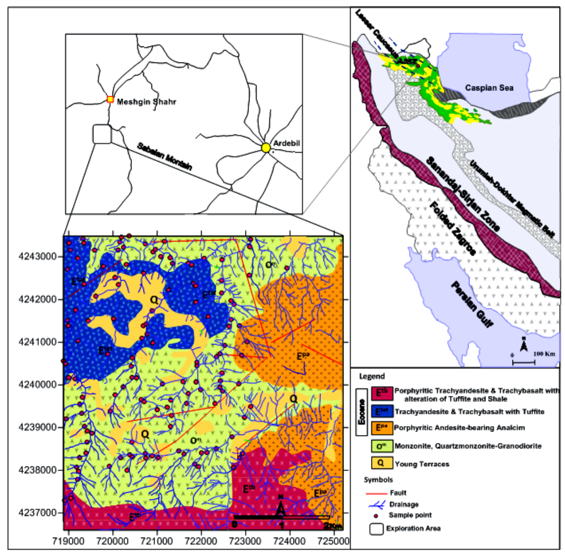

Fig.1. Simplified geological map of the studied area along with the sampling point location

2- Materials & Methods

In this study, to evaluate the performance of the proposed methods for pattern recognition of geochemical data in Meshginshahr, the spatial variables and geochemical elements were considered as input neurons, whereas the partition of input vectors into clusters were presented as output neurons in the mapping space. Every input neuron is connected to every output neuron with weighted links. For assessing the geochemical dispersion patterns and vectoring exploration targets in the area, five steps were undertaken. At first, all variables were standardized between (0-

1) based on the Z-score. Considering that, the optimal number of clusters is not known from the beginning of data analysis, therefore, clustering based on SOM was carried out independently and each time by taking a number of different classes. In the second stage, the data was divided into training and testing data. Then, organizing of the nonlinear relation between geochemical data in patterns was made based on Self- Organizing Maps (SOMs) (Fig.2). Accordingly, K-means clustering technique (Vesanto et al., 2000) was employed to classify the SOM topography into related conceptual models. Finally, the geochemical dispersion model was constructed to illustrate the nature and origin of the surface expression of mineralization in the area (Fig.8). According to (Kohonen et al., 1997), during the process, the algorithm provides a projection of the multi-dimensional data into a two-dimentional map and preserves the topology of this input data space. The presented SOM was trained with different numbers of map units and the optimum map size was selected based on the R-Squared(RS) index (Fig.3). The RS index measures the dissimilarity of clusters and has the values from 0 to 1 where 0 shows there are no differences among the clusters and 1 indicates significant differences among them. Noted that, if the resulting map size is small, it might not explain some important differences in the data that should be detected. Conversely, if the map size is too big, the differences are too small (Vesanto, et al, 2000). Considering the above mentioned context, the results of SOM performance on the data are presented in Figs (5-8) and the optimal patterns were shown in six main clusters indicating the individual geochemical dispersion pattern.

Ultimately, the SOM and FKMC results were compared. Based on clustering accuracy and capability of the implemented methods in discriminating optimal geochemical pattern, the best option was proposed (Figs .9-10). Noted that, pattern recognition techniques are able to characterize the rocks and link the geochemical and geological data quantitatively. Here, the fuzzy k-means algorithm as an unsupervised pattern recognition technique can group data into clusters based on properties measured. In order to implement the FKMC on the data, all geochemical stream sediments samples which control the lithology and environments was undertaken. Normally, The grater the geological anisotropy, the higher the number of clusters. With respect to this, Figure (4) shows the flowchart of FKMC in the present geochemical data.

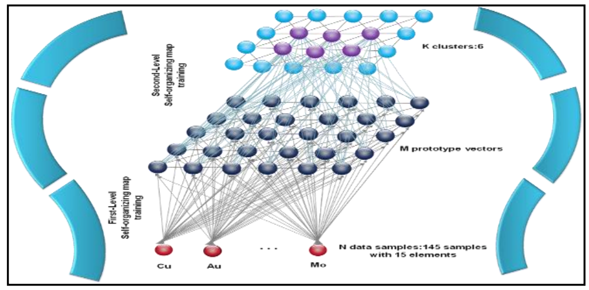

Fig.2. Topology of SOM network for geochemical pattern recognition in the studied area

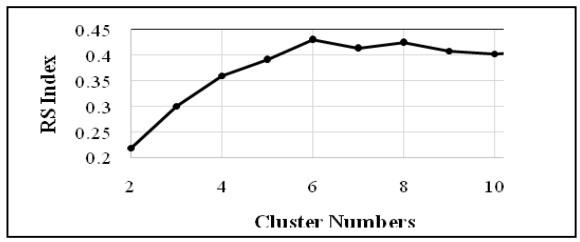

Fig.3. RS Index variations based on the number of clusters in the data

At first, the data was prepared and pre-processed. Also, a reasonable method was selected (from Cosine, Jaccard and Euclidean methods) to calculate the distance between elements and centers. Then optimal values for class number and fuzzy component were determined. To obtain centroid and membership based on the mentioned equations, it was attempted to apply the iterations until the centers or memberships was constant to within some small value. Next, we initialized the membership as random values using a distribution that satisfies all conditions. Then, centers were calculated and afterward memberships were recalculated according to the new centers. If the new memberships did not change compared to the older or had a small difference, the clustering process is ended. Otherwise, the recalculation of new centers and memberships are continued. The process repeats until the algorithm converges to a point where the relative change in objective function is lesser than 0.001 and saves the best memberships and centers that resulted from the optimum random initiation corresponding to the least objective function (Figs9-10, Tables 3-4).

3-Results and Discussions

To demonstrate the effectiveness of the proposed model-based clustering method, the performance of the SOM algorithm on discriminating of geochemical dispersion patterns was tested the relevant results are presented in Figs.2-7. Based on SOM visualization(Figs 5-6), the data were classified into six clusters associated with the specific geochemical behavior. Figure 4 presents “unified distance matrix” U-matrix which indicates the closeness between adjacent nodes on the map and Figure 5 which shows a color-temperature scale where the cooler colors separate adjacent nodes that are closer (similarity) and the hotter colors indicate larger Eulidean separations (difference).

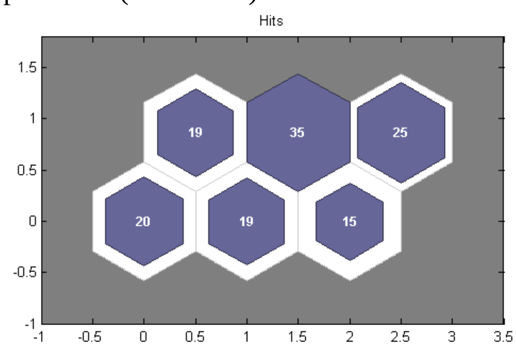

Fig.5. U-matrix representation with white hexagons sized proportional to the number of input samples falling on the each node

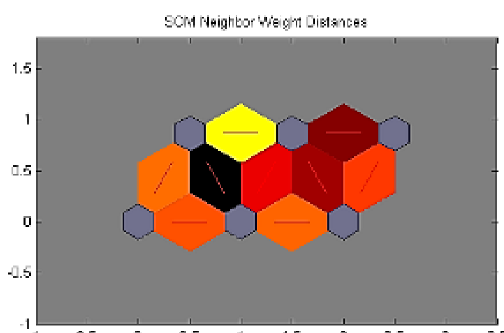

Fig.6. Shows the color-coding of the nodes which can be used in distribution patterns

About the optimal number of clusters, it is essential to express that the number of clusters depends on the precision, interpretation overview and the discrimination quality of clusters. In this respect, the separation quality is controlled by RS index (Fig.5). Accordingly, the next interpretations were performed assuming an optimal pattern with six clusters on the data. Subsequently, Figures 7-8 show the elements dispersion model which provides a proper context for interpreting the geochemical expression of mineralization by characterizing the principal factors influencing elements dispersion in the area. In cluster 1, which is elongated the N-NW of the region and to some extent in the middle part of the studied area, an enrichment of Ba, Co and Cu elements are visible.The 2nd cluster has been located in the center and northwest part of the studied area (Fig.6) and contains the anomalies of As, Au elements. There is a relatively good compliance between the anomalies and the position of points in the second and sixth clusters. Also, 6th cluster has expanded in the south and southwest of the studied area (Fig.7). Based on Table 2 and Figure 7, W, Sn, Pb, Mo, Ag elements show significant anomalies in this cluster. With regard to the geochemical distribution pattern of Ag in the cluster, it is concluded the fluctuation range of Ag concentration in the whole region is approximately the same.

The distribution pattern of Pb, Sn, Ag, Mo, Au, Bi, Sb, and W elements on the 6th cluster is shown in Fig. 7, respectively. The enhancement of their anomalies, except for Sn, is strong in the sixth cluster and has a south-southwest trend.

Paragenetic sequence of elements which were explained in different clusters has been presented in Table 1. Accordingly, in the first cluster, the group of Ba, Co and Cu reflect ore-forming elements associated with the base metals.

Table 1: SOM-based clusters and their pertinent elements

-based clustering SOM	SOM -based clustering
Cluster 1: Ba, Co, Cu	Cluster 4:Ce, Cr, Co
Cluster 2: As, Au, Hg, Ba	Cluster 6: Ag, Mo, Pb, Sn, W, Au, Bi, Sb
Cluster 3: Bi, Sb, As, Cr, Hg	

The explained elements in the second cluster show the group of As, Hg, To, Au elements which expressing the relative enrichment of Au and the relevant trace elements. Also, the dispersion characteristics of Cr, Ce and Co elements in the fourth cluster was due to lithology and syngenetic effects. The enrichment of the oriented elements in the sixth cluster, indicates the geochemical system of Ag, Mo, Pb, Sn, W, Au, Bi, Sb, elements which are associated with the acidic solutions resulting from the penetration of igneous masses into the acidic volcanic rocks of the area.

In order to evaluate and confirmation of the results of SOM in geochemical pattern recognition, it was attempted to compare its performance with the results obtained from FKMC in the area (Figs.9-10).

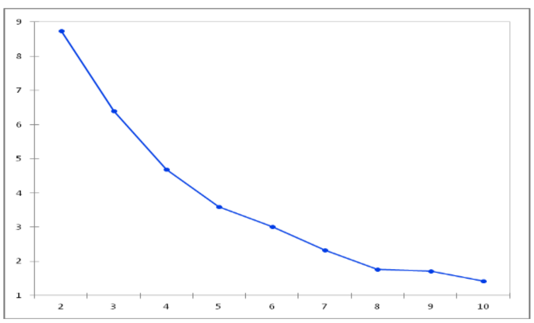

Fig .9. a. Showing the objective function variations vs number of clusters

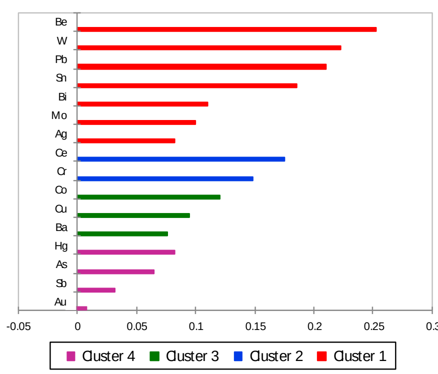

Fig 9. b. Presenting the classification of geochemical elements based on silhouette coefficient

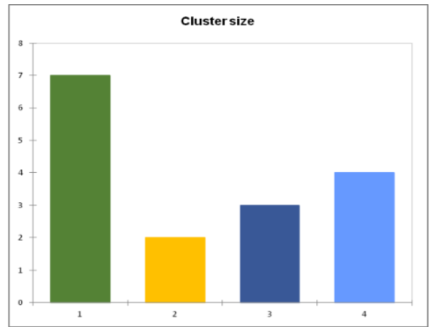

Fig.9.c. Showing clusters size resulted from FKMC

So, in this study, it was attempted to apply FKMC as a fuzzy-based approach which has a remarkable ability in recognizing distribution pattern and identify the epigenetic components. In this respect, the most reliable cluster structure which was compatible with geology, geochemical characteristics and paragenesis of minerals in the study area, was extracted. In this regard, for computing the FKM algorithm, the coefficient of fuzziness (m) was 1.05. As mentioned before, this parameter is variable and chosen in advance from 1 to infinity but in most application of FKMC, a value between 1 to 2 has provided appropriate results. Besides this, the results showed that increasing in this coefficient provides greater stability in the classification. In view of all conditions in this clustering, we had to select the optimal pattern. For this, m equal to 1.05 represented a good compromise with respect to entropy and stability of the classification. Here, another point of concern is the optimal fuzzy classification which is obtained by minimization of the object function to satisfy the conditions. The clustering objective function is computed depending on the choice of clustering distance. Here, with respect to the comparison of the distances Jaccard Index distance was more adoptable to the studied data. About the number of clusters, with respect to the behavior of geochemical data in the area, it could be chosen from two to six clusters. The four cluster alternative exhibits the optimal results in term of objective function, stability in classification, silhouette coefficient (Fig. 10 a, b) and average membership in each cluster.

The stop conditions of this clustering are the iterations and convergence during the process.

Here, the calculations are stopped when the maximum number of iterations has been exceeded of the considered value and the algorithm running will continue until the convergence reach to the threshold of convergence. Figure 10 shows the cluster size and expresses the number of elements for each cluster. Accordingly, the first cluster is assigned to make a significant percent of data corresponding to a considerable extent of W, Pb, Mo, Sn, Bi, Be, Ag mineralization in the area (Table 2). The related distribution pattern has been depicted in figure (10.a). Consequently, the placement of Ag, Mo, Pb in this cluster and their associations can be a sign of hydrothermal mineralization. Also, in this pattern the geochemical system of Bi, W, Pb, Sn, Mo that are associated with acidic rocks and it seems that in both of the mentioned clustering methods, Mo is not only distinct from Cu, but also its accumulation has been mostly affected by acidic solutions.

The fourth cluster is related to Au local enrichments and has the second priority of halos extent in the area especially in center and western parts of area (Fig. 10.c). The placement of Au, Sb and As in this cluster indicates the association of these elements which is commonly observed in Au bearing hydrothermal deposits especially in the Au bearing vein deposits.

The pattern related to third cluster expressed the association of Ba, Co, and Cu where copper enrichments seem to be occurs in relation to secondary processes and mineralization and are mostly concentrated in the north of the area. (Fig. 10.b).

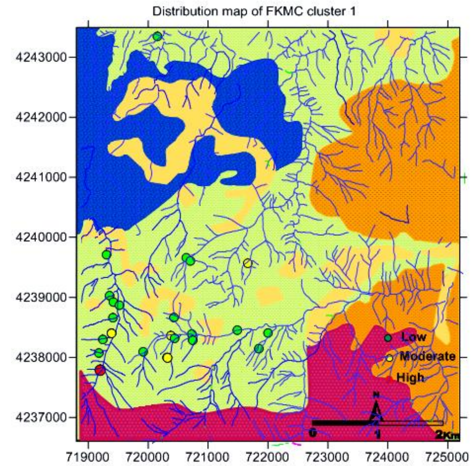

Fig. 10. a. Geochemical distribution pattern of FKMC cluster1

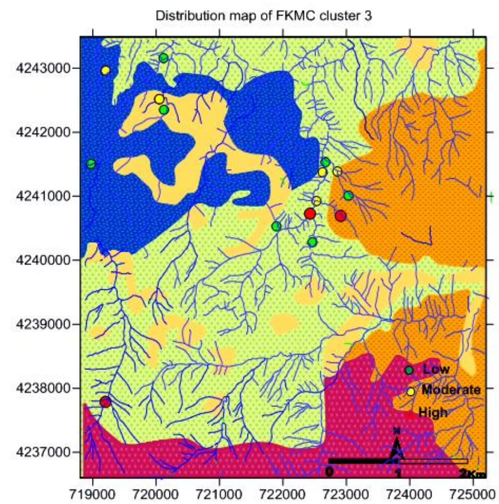

Fig. 10. b. Geochemical distribution pattern of FKMC cluster 3

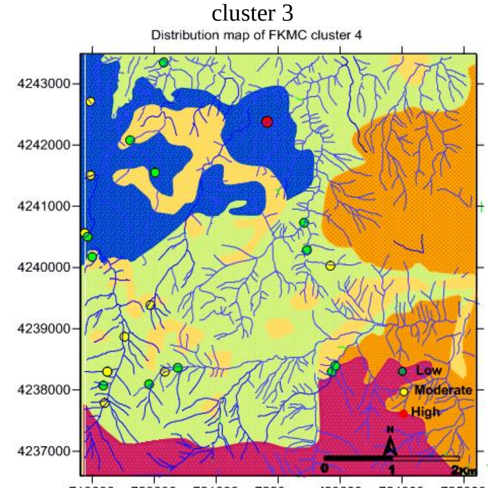

Fig . 10.c. Geochemical distribution pattern of FKMC cluster 4

The second cluster with low size shows the relation of Cr, Ce elements whose distributions were probably influenced by lithology and syngenetic effects.

Since an increasing in number of classes would result in decreasing the average membership, therefore, the proper and optimal results in accordance with the reality were achieved by four clusters. Subsequently, Table 3 shows the membership probabilities of each element in FKM clusters and highlighted the oriented elements with high membership probability in related clusters. The table confirmed that the discriminations that could be seen in classification correspond well to field geochemical behaviors.

Table 3. Membership probabilities of assigned elements in related clusters

Membership probabilities				
Variables	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Au	0.015	0.002	0.002	**0.981**
Pb	**1.000**	0.000	0.000	0.000
Ag	**0.997**	0.000	0.002	0.001
Sn	**1.000**	0.000	0.000	0.000
Hg	0.000	0.000	0.000	**1.000**
As	0.000	0.000	0.000	**1.000**
Sb	0.000	0.000	0.000	**1.000**
Bi	**1.000**	0.000	0.000	0.000
W	**1.000**	0.000	0.000	0.000
Mo	**0.997**	0.001	0.001	0.001
Ba	0.000	0.000	**1.000**	0.000
Cu	0.002	0.003	**0.993**	0.002
Be	**1.000**	0.000	0.000	0.000
Ce	0.000	**1.000**	0.000	0.000
Co	0.000	0.000	**1.000**	0.000
Cr	0.003	**0.983**	0.013	0.001

Comparing the results of two above mentioned clustering methods reveals the appropriate compliance of them in geochemical pattern recognition but the higher capability of the SOM method in single-element halos enhancement while the FKMC shows a behavior similar to composite halos and provides higher accuracy in recognizing the patterns originates from various geological processes. Therefore, the results of both methods are complementary and are of considerable importance in mapping the mineralization potential of the area and constrain the uncertainty in detecting appropriate exploration patterns.

Ultimately, it was attempted to validate the results of above mentioned methods in the area with the outcome of alterations obtained from remote sensing (Fig.11).

Well correlation of the ore-related elements patterns resulted from the methods with the major alteration zones containing Fe-bearing and clay minerals in the area shows that most of the major anomalies trend correspond to iron minerals in the northwest-southwest direction.

4-Conclusions

In this paper characterizing the distribution pattern of elements for a large volume of geochemical data have undertaken by SOM and FKMC methods. Due to the advantages of the SOM method as mentioned, it was applied to preserve the nature of the data and recognizing the geochemical distribution patterns with high certainty. The summary of elements distribution pattern and tables derived from the SOM method shows that clusters 1, 2, and 6 of this method are important from mineralization trend point of view in the area. Cluster 1 is important in terms of Cu and associated elements Ba, Co anomalies which affected by hydrothermal solutions. The results suggests a lack of high erosion and indicates the potential for major elements anomalies in depth. The focus of these elements is mainly in the northern part of the area. The concentration of Au in the area is low and the dispersion is highlighted in the second cluster. Au anomalies are weak but patterned and dispersed in both major rocks (intrusive and volcanic) from northwest to southwest.

In order to assess the results of the SOM-based clustering, comparison was made with the results of the fuzzy k-means clustering (FKMC) that revealed well correlation among the results of the methods and favorability of the area for Cu-Au mineralization. On the other hands, FKMC not only confirmed the SOM performance in single-element distribution pattern recognition, but also, highlighted multiple behaviors of patterns and thereby, it enhanced the geochemical composite halos by considering the inter-relationship of elements and their different membership probabilities in the area.

Finally, this study suggests the higher capability of the SOM and FKMC methods in identifying elements associations and its competence for optimal discriminating of Paragenetic sequence. The resulted clean ore-related clusters enable the exploratory planning for various purposes and reduce uncertainty, leading prospecting programs to more reliable promising areas with higher accuracy and precision.

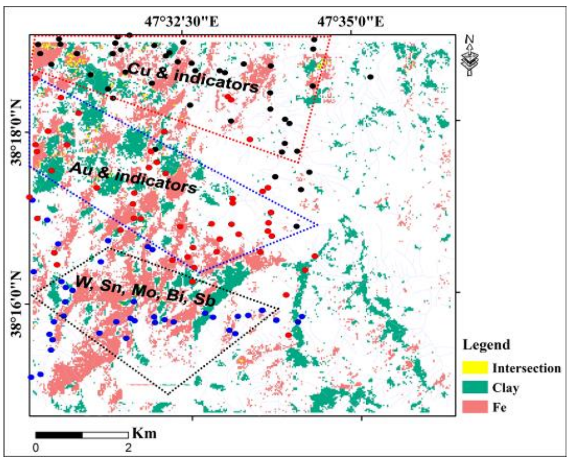

Fig.11. Shows the assigning of elements to specific clusters discriminated by two methods and their compliance with the alteration of the area resulted from remote sensing

5- Acknowledgement

The authors would like to acknowledge their sincere thanks to editor and anonymous reviewers for their constructive comments.

6- References

[1] Aghanabati, A. (2004). "Geology of Iran," Geological survey of Iran.

[2] Al-Anazi, A., and Gates, I. D. (2010). On the Capability of Support Vector Machines to Classify Lithology from Well Logs. Natural Resources Research **19**, 125-139.

[3] Bação, F., Lobo, V., and Painho, M. (2005). The self-organizing map, the Geo-SOM, and relevant variants for geosciences. Computers & Geosciences **31**, 155-163.

[4] Baldwin, J. L., Bateman, R. M., and Wheatley, C. L. (1990). Application of a neural network to the problem of mineral identification from well logs. The Log Analyst **31**.

[5] Brehme, M., Bauer, K., Nukman, M., and Regenspurg, S. (2017). Self-organizing maps in geothermal exploration–A new approach for understanding geochemical processes and fluid evolution. *Journal of Volcanology and Geothermal Research* **336**, 19-32.

[6] Butt, C. (2005). Geochemical dispersion, process and exploration models. *Regolith Expression of Australian Ore Systems. CRC LEME*, 81-106.

[7] Carneiro, C. d. C., Fraser, S. J., Crósta, A. P., Silva, A. M., and Barros, C. E. d. M. (2012). Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. *Geophysics* **77**, K17-K24.

[8] Cazes, P. (1970). Application de l'analyse des données au traitement de problèmes géologiques.

[9] Céréghino, R., Giraudel, J., and Compin, A. (2001). Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecological Modelling* **146**, 167-180.

[10] Choi, B.-Y., Yun, S.-T., Kim, K.-H., Kim, J.-W., Kim, H. M., and Koh, Y.-K. (2014). Hydrogeochemical interpretation of South Korean groundwater monitoring data using self-organizing maps. *Journal of Geochemical Exploration* **137**, 73-84.

[11] Chon, T.-S. (2011). Self-organizing maps applied to ecological sciences. *Ecological Informatics* **6**, 50-61.

[12] Cracknell, M., Reading, A., and McNeill, A. (2014). Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps. *Australian Journal of Earth Sciences* **61**, 287-304.

[13] Cracknell, M., Reading, A. M., and De Caritat, P. (2015). Multiple influences on regolith characteristics from continental-scale geophysical and mineralogical remote sensing data using Self-Organizing Maps. *Remote Sensing of Environment* **165**, 86-99.

[14] David, M., and Woussen, G. (1973). Correspondence analysis, a new tool for geologists. *In* "Proc. Mining Pribram", Vol. 1, pp. 41-65.

[15] De Gruijter, J., and McBratney, A. (1988). A modified fuzzy k-means method for predictive classification.

[16] Didon, J., Gemaime,Y.M. (1976). Le Sabalan Volkan plioquartermair del Azerbaijan oriental (Iran), Etude qeologique et petrographique del edfic et de son environmental regional, Grenoble, , France.

[17] Fraser, S., and Dickson, B. (2006). Data mining geoscientific data sets using self organizing maps. *Mastering the Data Explosion in the Earth and Environmental Sciences, Extended Abstracts*, 5-7.

[18] Fraser, S., and Dickson, B. (2007). A new method for data integration and integrated data interpretation: Self-organising maps. *In* "Proceedings of exploration", Vol. 7, pp. 907-910. Citeseer.

[19] Fraser, S. J., Wilson, G. A., Cox, L. H., Cuma, M., Zhdanov, M. S., and Vallée, M. A. (2012). Self-organizing maps for pseudo-lithological classification of 3D airborne electromagnetic, gravity gradiometry and magnetic inversions. *ASEG Extended Abstracts* **2012**, 1-4.

[20] Gorsevski, P. V., Gessler, P. E., and Jankowski, P. (2003). Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of geographical systems* **5**, 223-251.

[21] Grunsky, E., and Agterberg, F. (1992). Spatial relationships of multivariate data. *Mathematical Geology* **24**, 731-758.

[22] Grunsky, E. C. (2010). The interpretation of geochemical survey data. *Geochemistry: Exploration, Environment, Analysis* **10**, 27-74.

[23] Gustavsson, N. (1976). LITHOLOGICAL CLASSIFICATION OF TILLS BY DISCRIMINANT ANALYSIS.

[24] Harris, D., and Pan, G. (1999). Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis. *Natural Resources Research* **8**, 93-109.

[25] Harris, J., Wilkinson, L., and Bernier, M. (2001). Analysis of geochemical data for mineral exploration using aGIS - A case study from the Swayze greenstone belt, Northern Ontario, Canada. *Geological Society, London, Special Publications* **185**, 165-200.

[26] Heil, J., Häring, V., Marschner, B., and Stumpe, B. (2019). Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils. *Geoderma* **337**, 11-21.

[27] Herbst, M., Gupta, H. V., and Casper, M. (2009). Mapping model behaviour using self-organizing maps. *Hydrology and Earth System Sciences* **13**, 395-409.

[28] Hsu, K. l., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B. (2002). Self- organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. *Water Resources Research* **38**, 38-1-38-17.

[29] Iran, G. s. a. M. e. o. (2008). "Geochemical exploration report of Meshgin shahr (2)1:25000 sheet." Kan Azin engineering company.

[30] Javid, F., Mohammadzadeh, M. J., and Nasseri, A. (2015). Optimization of Geochemical Patterns Based on Multivariate Methods in the Duzduzan Area, E-Azerbaijan. *In* "Proceedings of the 24th international Mining congress of Turey(IMCET)", pp. 360-369, Antalya/TURKEY.

[31] Jianwen, M., and Bagan, H. (2005). Land-use classification using ASTER data and self-organized neutral networks. *International Journal of Applied Earth Observation and Geoinformation* **7**, 183-188.

[32] Kalteh, A. M., and Berndtsson, R. (2007). Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP). *Hydrological sciences journal* **52**, 305-317.

[33] Kalteh, A. M., Hjorth, P., and Berndtsson, R. (2008). Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software* **23**, 835-845.

[34] Kaski, S. (1997). Data exploration using self-organizing maps. *In* "Acta polytechnica scandinavica: Mathematics, computing and management in engineering series no. 82". Citeseer.

[35] Kesler, S. E., Chryssoulis, S. L., and Simon, G. (2002). Gold in porphyry copper deposits: its abundance and fate. *Ore Geology Reviews* **21**, 103-124.

[36] Klose, C. D. (2006). Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data. *Computational Geosciences* **10**, 265-277.

[37] Kohonen, T., Huang, T. S., and Schroeder, M. R. (2001). "Self-Organizing Maps," Springer Berlin Heidelberg.

[38] Kohonen, T., Kaski, S., and Lappalainen, H. (1997). Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural computation* **9**, 1321-1344.

[39] Lee, S., and Lathrop, R. G. (2006). Subpixel analysis of Landsat ETM/sup+/using self-organizing map (SOM) neural networks for urban land cover characterization. *IEEE transactions on geoscience and remote sensing* **44**, 1642-1654.

[40] Ley, R., Casper, M., Hellebrand, H., and Merz, R. (2011). Catchment classification by runoff behaviour with self-organizing maps (SOM). *Hydrology & Earth System Sciences* **15**.

[41] Lin, G. F., and Chen, L. H. (2005). Time series forecasting by combining the radial basis function network and the self- organizing map. *Hydrological Processes: An International Journal* **19**, 1925-1937.

[42] Lindqvist, L., Lundholm, I., Nisca, D., Esbensen, K., and Wold, S. (1987). Multivariate geochemical modelling and integration with petrophysical data. *Journal of Geochemical Exploration* **29**, 279-294.

[43] Liu, C., Hu, T., Ge, Y., and Xiong, H. (2012). Which distance metric is right: An evolutionary k-means view. *In* "Proceedings of the 2012 SIAM International Conference on Data Mining", pp. 907-918. SIAM.

[44] Liu, Y., and Weisberg, R. H. (2011). A review of self-organizing map applications in meteorology and oceanography. *Self-Organizing Maps: Applications and Novel Algorithm Design*, 253-272.

[45] Marroquín, I. D., Brault, J.-J., and Hart, B. S. (2009). A visual data-mining methodology for seismic facies analysis: Part 1—Testing and comparison with other unsupervised clustering methods. *Geophysics* **74**, P1-P11.

[46] Martínez, P., Gualtieri, J., Aguilar, P., Pérez, R., Linaje, M., Preciado, J., and Plaza, A. (2001). Hyperspectral image classification using a self-organizing map. *In* "Summaries of the X JPL Airborne Earth Science Workshop".

[47] Nasseri, A., and Mohammadzadeh, M. J. (2017). Evaluating distribution pattern of petrophysical properties and their monitoring under a hybrid intelligent based method in southwest oil field of Iran. *Arabian Journal of Geosciences* **10**, 9.

[48] Nasseri, A., Mohammadzadeh, M. J., and Tabatabaei Raeisi, S. H. (2015). Fracture enhancement based on artificial ants and fuzzy c-means clustering (FCMC) in Dezful Embayment of Iran. *Journal of Geophysics and Engineering* **12**, 227-241.

[49] Neagoe, V.-E., and Ropot, A.-D. (2002). Concurrent self-organizing maps for pattern classification. *In* "Proceedings First IEEE International Conference on Cognitive Informatics", pp. 304-312. IEEE.

[50] Nourani, V., Baghanam, A. H., Adamowski, J., and Gebremichael, M. (2013). Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *Journal of hydrology* **476**, 228-243.

[51] Park, Y.-S., Tison, J., Lek, S., Giraudel, J.-L., Coste, M., and Delmas, F. (2006). Application of a self-organizing map to select representative species in multivariate analysis: a case study determining diatom distribution patterns across France. *Ecological Informatics* **1**, 247-257.

[52] Pastukhov, A. A., and Prokofiev, A. A. (2016). Kohonen self-organizing map application to representative sample formation in the training of the multilayer perceptron. *St. Petersburg Polytechnical University Journal: Physics and Mathematics* **2**, 134-143.

[53] Patil, J. K., and Kumar, R. (2011). Advances in image processing for detection of plant diseases. *Journal of Advanced Bioinformatics Applications and Research* **2**, 135-141.

[54] Peeters, L., Bação, F., Lobo, V., and Dassargues, A. (2007). Exploratory data analysis and clustering of multivariate spatial hydrogeological data by means of GEO3DSOM, a variant of Kohonen's Self-Organizing Map.

[55] Porwal, A., Carranza, E., and Hale, M. (2003). Artificial neural networks for mineral-potential mapping: a case study from Aravalli Province, Western India. *Natural resources research* **12**, 155-171.

[56] Porwal, A., and Carranza, E. J. M. (2015). Introduction to the Special Issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. Elsevier.

[57] Rantitsch, G. (2000). Application of fuzzy clusters to quantify lithological background concentrations in stream-sediment geochemistry. *Journal of Geochemical Exploration* **71**, 73-82.

[58] Reddy, R., and Bonham-Carter, G. (1991). A decision-tree approach to mineral potential mapping in Snow Lake area, Manitoba. *Canadian Journal of Remote Sensing* **17**, 191-200.

[59] Rendu, J. (1976). Bayesian decision theory applied to mineral exploration and mine valuation. *In* "Advanced Geostatistics in the Mining Industry", pp. 435-445. Springer.

[60] Riou, R., Dupuy, C., and Dostal, J. (1981). Geochemistry of coexisting alkaline and calc-alkaline volcanic rocks from northern Azerbaijan (NW Iran). *Journal of Volcanology and Geothermal Research* **11**, 253-275.

[61] Schatzmann, J., and Ghanem, M. (2003). Using self-organizing maps to visualize clusters and trends in multidimensional datasets. *Department of Computing Data Mining Group, Imperial College, London*, 132.

[62] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001).

Estimating the support of a high-dimensional distribution. *Neural computation* **13**, 1443-1471.
[63] Strecker, U., and Uden, R. (2002). Data mining of 3D poststack seismic attribute volumes using Kohonen self-organizing maps. *The Leading Edge* **21**, 1032-1037.
[64] Sun, X., Deng, J., Gong, Q., Wang, Q., Yang, L., and Zhao, Z. (2009). Kohonen neural network and factor analysis based approach to geochemical data pattern recognition. *Journal of Geochemical Exploration* **103**, 6-16.
[65] Tasdemir, K., and Merényi, E. (2009). Exploiting data topology in visualization and clustering of self-organizing maps. *IEEE Transactions on Neural Networks* **20**, 549-562.
[66] Toivanen, P. J., Ansamäki, J., Parkkinen, J., and Mielikäinen, J. (2003). Edge detection in multispectral images using the self-organizing map. *Pattern Recognition Letters* **24**, 2987-2994.
[67] Ultsch, A., and Herrmann, L. (2005). The architecture of emergent self-organizing maps to reduce projection errors. *In* "Esann", pp. 1-6.
[68] Vesanto, J. (2000). Neural network tool for data mining: SOM toolbox. *In* "Proceedings of symposium on tool environments and development methods for intelligent systems (TOOLMET2000)", pp. 184-196. Citeseer.
[69] Vesanto, J., and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on neural networks* **11**, 586-600.
[70] Villmann, T., Merényi, E., and Hammer, B. (2003). Neural maps in remote sensing image analysis. *Neural Networks* **16**, 389-403.
[71] Zuo, R., and Carranza, E. J. M. (2011). Support vector machine: a tool for mapping mineral prospectivity. *Computers & Geosciences* **37**, 1967-1975.

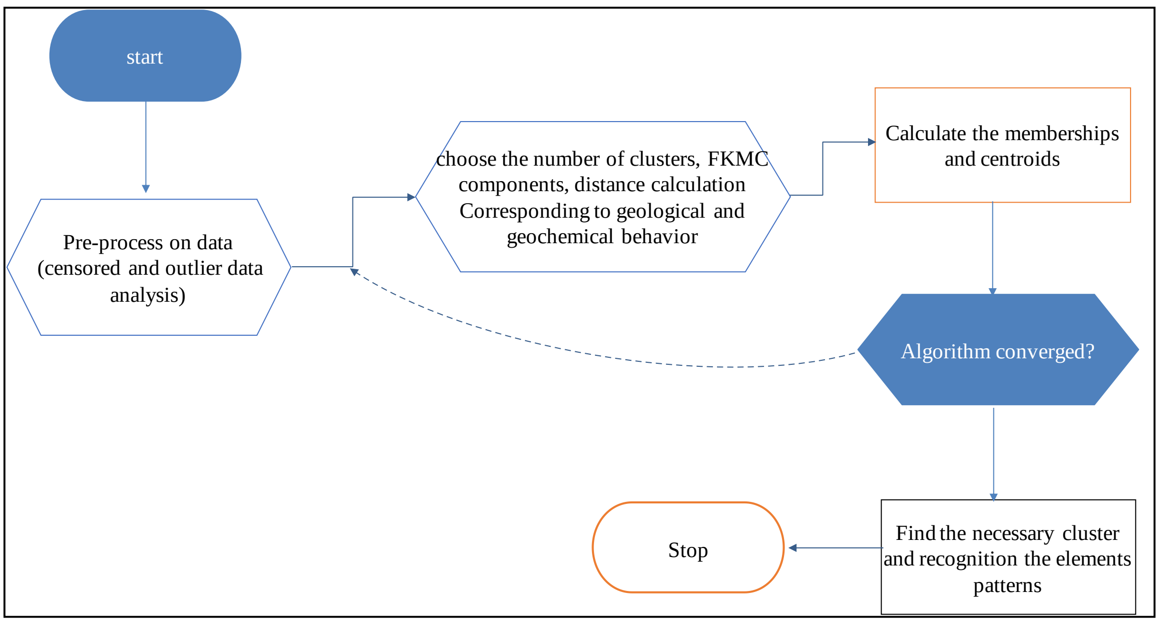

Fig.4. Flowchart illustrating the proposed fuzzy k-means algorithm applied in the area.

Table 2. Displays some of the important sets of clusters in FKMC

Cluster	Size	Within-class	Minimum distance to centroid	Maximum distance to centroid	Average distance to centroid	The elements in each cluster						
Cluster 1	7	2.224	0.458	0.693	0.556	Be	W	Pb	Sn	Bi	Mo	Ag
Cluster 2	2	0.222	0.333	0.000	0.333	Ce	Cr					
Cluster 3	3	0.723	0.420	0.482	0.488	Co	Cu	Ba				
Cluster 4	4	1.508	0.544	0.559	0.610	Hg	As	Sb	Au			

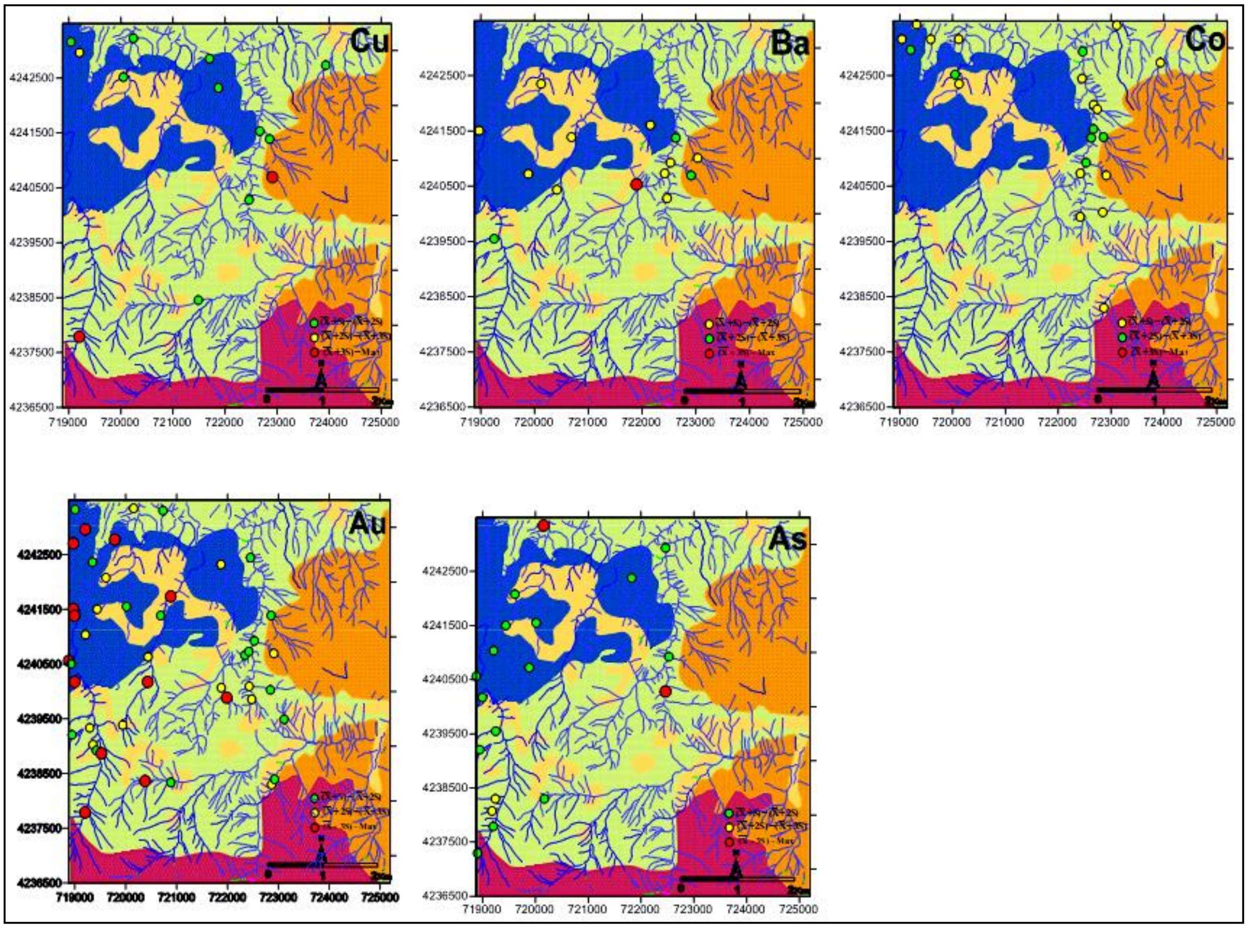

Fig.7. Geochemical distribution pattern of elements in the first cluster resulting from the SOM method

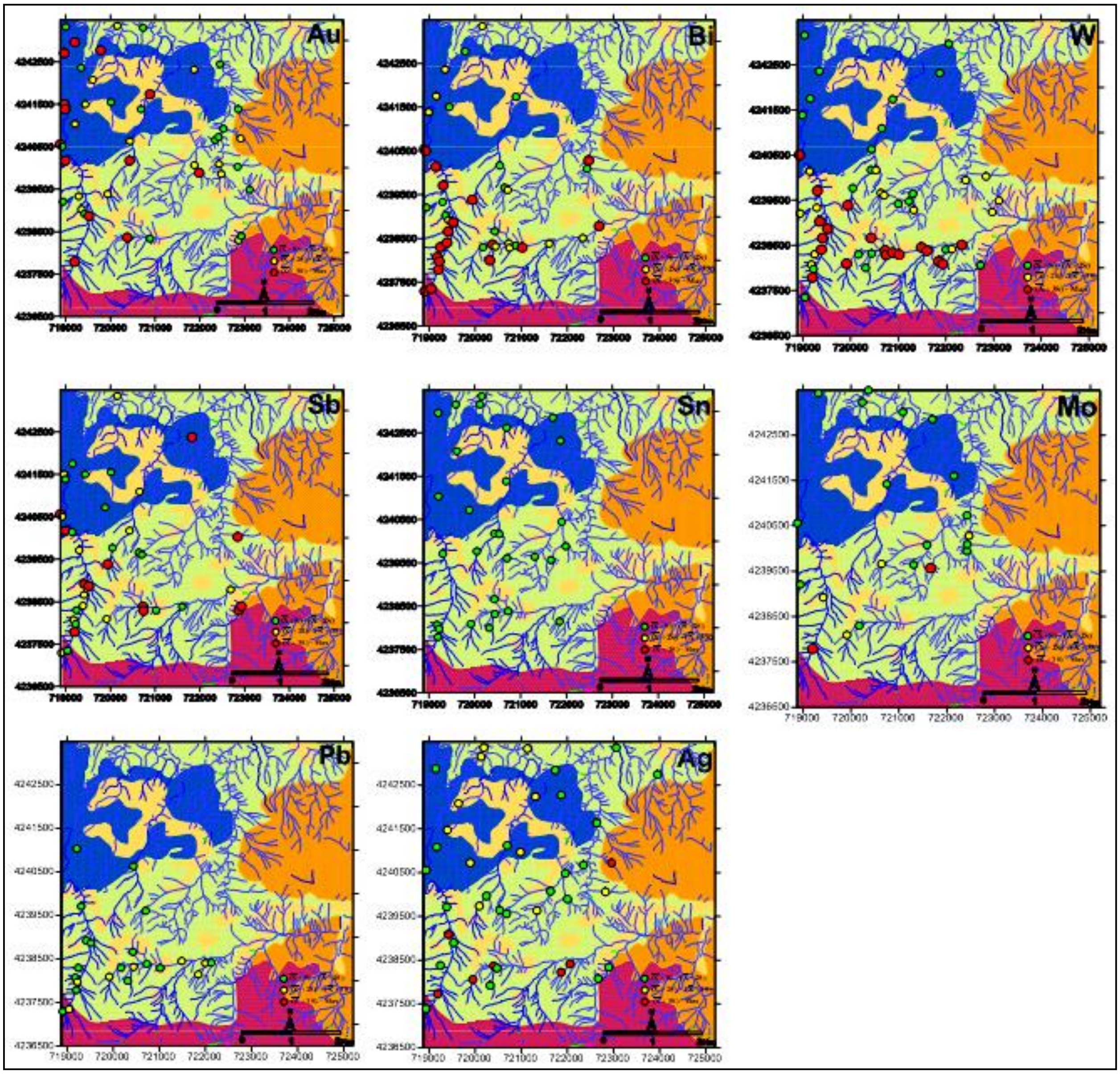

Fig.8. Geochemical distribution pattern of the elements justified in the sixth cluster resulting from the SOM method