Application of Fuzzy C-mean Clustering Algorithm (FCM) in determination of Pb and Zn elements anomalies of Razan, Hamedan Province, Iran

Mehrdad malekmohammadi

ABSTRACT: The determination of high potential regions, according to the segregation of the real anomalies, from those of false types of mineralization, has a syngenetic component. In this regard, the Fuzzy C-Means analysis is considered as a valid method that specifies the real high potential regions through elimination of syngenetic component based on the residual values. Two hundred and thirty five stream sediment samples were analyzed using ICP 14-element analysis to determine the high potential regions of Pb and Zn in Hamedan province. After receiving the results, for data processing, the FCM method was applied. The calculations pertinent algorithm is written in Excel Software using Visual Basic Language. The data entered the program as input was those normalized using Cox and Box Method. This algorithm calculates the centers of clusters and considering the membership degree of each sample to the cluster center, the elements residual values were achieved. Finally, through applying P.N method on the residual values achieved through FCM method, a regions was determined between the longitudinal coordinates of X=332000 & X=335000 and latitudinal coordinates Y=3923000 & y=3926000 for Pb and Zn.

KEY WORDS: Stream sediment, Geochemistry, Clustering, Fuzzy, Razan.

1. Introduction
The C-mean fuzzy analysis is one of the multi-variable statistical methods utilized in geochemical explorations through elimination of syngenetic component in high potential regions. The method aims at designating data to several clusters. The result of such clustering is the allocation of a set of data to a p-dimensional space. This method is based on data similarity matrix. It means that first, the n×n similarity matrix of the existing data should be calculated by non-fuzzy cluster methods and next a fuzzy operator should be applied on the resulted matrix as illustrated in Equation (1).

\[ [S], [S] = [S^2] = [V, (s_{k}, s_{k})], \]
\[ [S^2], [S] = [S^3], \]
\[ S^{n-1} = [S^n], \]

In which, V is the maximum value selector and ^ is the minimum value selector and [S] is the matrix of similarity. Utilizing from such operators continues until the time the resulted matrix becomes convergent. No changes will be made in its entries due to continuation of the calculation stages. The number of iteration steps to achieve the ultimate convergent matrix considering the dimension of matrix (n) follows the following Equation (2).

\[ [\log_2 n] + 1 \]

2. FCM Analysis Algorithm
a- At the first step, for each sample, regarding to each cluster, a membership degree was selected. Random Equation is used for this selection. Note that sum of the membership degrees of a sample regarding all clusters should be equal to one.

b- Using Equation 3, a cluster center is calculated for each cluster.
\[
C_{ij} = \frac{\sum (\mu_k i)^s x_{ki}}{\sum (\mu_k i)^s}, \quad (3)
\]

\(C_{ij}\) is the value of variable \(i^{th}\) from cluster center \(j^{th}\), \(\mu_k\) is the membership degree of sample \(k^{th}\) to cluster \(i^{th}\), \(X_{ki}\) is the value of variable \(j^{th}\) in sample \(k^{th}\); \(q\) is the Fuzzy degree of variable \(j^{th}\) in sample \(k^{th}\).

c- After calculation of cluster centers, new membership degrees regarding new cluster centers are calculated based on Euclidian distance through Equations 4 and 5.

\[
\mu_k = \frac{(d_k^2)^{-1}}{\sum_{k=1}^{n} (d_k^2)^{-1}} \quad (4)
\]

\[
(d_k^2) = \sum_{j=1}^{m} (X_{kj} - C_{ij})^2 / S_j \quad (5)
\]

\(S_j\) is the standard deviation of variable \(j^{th}\).

d- The last step, is the calculation of objective Equation through Equation 6.

\[
J_p = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_k i (d_ik)^q \quad (6)
\]

e- Steps B and D should be iterated until the contrast between two subsequent steps of calculation of \(J_p\) becomes less than the accuracy required. To control the calculations accuracy, proper number of clusters, and degree of Fuzzification, \(H\) and \(F\) control parameters could be applied, that are defined according to Equations 7 and 8.

\[
H = -\sum_{i=1}^{c} \sum_{k=1}^{n} \frac{\mu_k i \log(\mu_k i)}{n} \quad 0 \leq H \leq \log(C) \quad (7)
\]

\[
F = \sum_{i=1}^{c} \sum_{k=1}^{n} \frac{\mu_k i}{n} \quad \frac{1}{C} \leq F \leq 1 \quad (8)
\]

\(F\) (partition coefficient) introduces somehow, the ratio of intra cluster variance to inter cluster variance. The value should be near 1 otherwise, the classification entropy should be near zero.

3. Geological setting

OM\textsuperscript{ant-tr}

This unite contain trachyandesite and dacite to dacitic lavas with minor trachyandesitic lithic crystal tuff as interlayer. In petrological investigation the texture of rocks are porphyry, the mineralogical investigation show that this unit contain euhedral to subhedral plagioclase phenocrysts in form of andesite-oligoclase, euhedral to subhedral phenocrysts of clinopyroxen, augite and diopsid. In this unite the matrix of rocks contain plagioclase, albite with flow texture, cryptocrystalline and recrystallized quartz, a major of chlorite, a minor of carbonate,apatite, Fe oxide, opaque mineral and a minor of K- feldspar.

OM\textsuperscript{an}

This unite contain calcareous vitric lithic tuff, recrystalized lithic tuff, silicified and carbonated crystal lithic tuff. Rock texture is euhedral with andesitic to dacitic composition. Mineralogical investigation show that the plagioclase phenocrysts in form of oligoclase-albite with clay minerals such as sercite, Ca carbonate, chlorite. In rock matrix, we can observe recrystallized silice, feldspar, chlorite, sericate and segment of hyaline.

OM\textsuperscript{sc}

This unite with minor volume contain red to gray conglomerate, locally with interlayer of limestone and olive green to marls.

Q\textsuperscript{1}

This unites spread in southern area and contains old high level terraces, gravel fans and conglomerate. The conglomerate is composited from the older rock.

Alteration zones

This unit is composited by stream of hydrothermal fluid in alignment of fractures in OM\textsuperscript{an} and OM\textsuperscript{ant-tr} units and caused silicification, kaolinitization, anhydrite and jarusit. These reactions also spread from the core toward the outside of the fractures. Petrological investigation illustrate that the basic texture of the rocks are somehow lost. The conglomerate in this unit contain fine and coarse segment with good maturity. The composition of conglomerate is plagioclase and clay minerals which substitute the amphiboles, Fe oxide and chlorite. Fig. 1
4. Application of fuzzy analysis on Razan’s Geochemical data.

In this article, the algorithm of these calculations is in Visual Basic Language, written in Excel. The data entered to these programs are those normalized using Cox and Box methods. This algorithm calculates the centers of clusters and determines the value of Residual of the analyzed elements considering the membership degree of each sample to the cluster centers. To determine the proper fuzzy degree, the allocated variance for different degrees of fuzzification, usually between 1.5 and 3, was calculated for various cluster numbers using software and the optimum value of fuzzy coefficient revealed to be 1.6.

5. Determination of parameters F and H

H (Classification Entropy) & F (Partition Coefficient) are the control parameters of cluster analysis validity. When H is minimal and F is maximal, the fuzzy clustering is in its best validity condition. According to these parameters, the different number of the clusters for H and F are calculated and the diagram of changes is plotted based on the number of the clusters so that the horizontal axis shows the number of clusters and the vertical axis shows the value of H and F. In this research, to calculate H and F based on different clusters number, an algorithm was written in visual basic language in Excel software, the data of which were normalized. This algorithm, calculates the H and F parameters based on different numbers of clusters from 2 to 10, and shows the result in form of a matrix. The scatter plot of the matrix was illustrated in Statistica software. With regard of Fig. 2 and table 1, partitioning the data
in 3 clusters was a proper way of segregation and it was concordant with the geological map. Thus, the final matrix, calculated based on Residual values with fuzzy degree 1.6.

Cluster 2: considering the Fig. 4, Andesit units influence a significant percentage of the existing samples in this cluster which is representative of andesitic units of the region. Such results are in complete concordance with the geological map of Razan. As it is illustrated in the Fig. 2, the samples belonging to this cluster are spread across the region. Such scattering could be the result of high expansion of trachyandesite units in a major part of the region.

Cluster3: considering Fig. 5, this cluster is pertinent to $Q^1$ unit (old high level terraces) and the other lithological units did not affect it. In this Fig., the highest concentration of the samples is in the lowest parts of the region and this is in concordance with the geological map. The achieved results revealed that the main lithologic parts of the region are partitioned in the clusters considering the composition and expansion. Application of this technique, the syngenetic component of the data will be neutralized and the existing samples will illustrate the anomaly region.

### Table 1 Value of H & F

<table>
<thead>
<tr>
<th>N.Cluster</th>
<th>H</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.905796</td>
<td>0.170468</td>
</tr>
<tr>
<td>11</td>
<td>0.893494</td>
<td>0.188084</td>
</tr>
<tr>
<td>10</td>
<td>0.895829</td>
<td>0.183846</td>
</tr>
<tr>
<td>9</td>
<td>0.891666</td>
<td>0.190465</td>
</tr>
<tr>
<td>8</td>
<td>0.892597</td>
<td>0.1895</td>
</tr>
<tr>
<td>7</td>
<td>0.901258</td>
<td>0.177101</td>
</tr>
<tr>
<td>6</td>
<td>0.896188</td>
<td>0.181061</td>
</tr>
<tr>
<td>5</td>
<td>0.89251</td>
<td>0.184906</td>
</tr>
<tr>
<td>4</td>
<td>0.895393</td>
<td>0.179731</td>
</tr>
<tr>
<td>3</td>
<td>0.903012</td>
<td>0.165013</td>
</tr>
<tr>
<td>2</td>
<td>0.919027</td>
<td>0.136415</td>
</tr>
</tbody>
</table>

6. Contour maps interpretation based on the sample membership degrees

Cluster 1: as it is illustrated in Fig.3, the samples with membership degree higher than 75% are mainly concentrated in the central part and northwest part of the region. This cluster is mostly influenced by trachyandesit magma and trachyandesit quartz. These units are expanded mainly in the northwest to southeast parts of the region and thus the results obtained from Razan’s geological map (1/100000) are in concordance with the contour map of samples membership degree. It is noteworthy that because of less expansion of trachyandesite to andesite, the influenced sample percentage by this unit, within this cluster is significantly lower.
7. Determination of residual values

Generally, the samples from the stream sediments show the chemical compound of the lithology and mineralization within the region. In the geochemical explorations, regarding the intervention and relation of the syngenetic components with probable mineralization, it is essential to eliminate the effect of this component from the data. In this method, after calculation the number of the clusters and their centers, the value of each variable in the center of each cluster will be equal with the weighted mean of the same variable in the existing samples of the cluster. Equation 9 is used to calculate this value.

\[ X_i = \mu A_i X_A + \mu B_i X_B + \ldots + \mu n_i X_n, \quad (9) \]

\( X_i \) stands for the value of variable in sample \( X^i \), \( \mu \) stands for the degree of membership to clusters; \( X_A \) stands for the value of variable \( X^A \) in the center of cluster \( A^A \).

Usually, the calculated value of each point is not the same as its measured value. This can be the result of various reasons such as contamination, calculation errors or the result of unreal neutralization of syngenetic components in the region. In this condition, the residual value will be calculated from Equation 10.

\[ \sigma_{k,i} = X_{k,i} - \sum_{j=1}^{n} \mu_{i,k} C_{ij} \quad (10) \]

8. Residual values map

After determination of residual value and separating the abnormal samples, the residual map is prepared. Since the syngenetic component has been neutralized in the maps, they clearly reveal the location of epigenetic components of mineralization that is the real anomaly location of the region. Fig.6 and 7 are the maps of Residual value of Pb and Zn respectively. Next the existing anomalies of the region are determined using the residual value of the elements through P,N method. To do so, a program has been written using the data which the residual values of different elements are normalized and next the anomalies of Pb and Zn are specified.
9. Conclusion
In the geochemical explorations implemented by stream sediment method, the variability has two main components including syngenetic and epigenetic. But the epigenetic component is considered as a useful explorative component for which we look. The C-mean fuzzy method is a multi variable statistical method that clarifies the high potential regions through elimination of syngenetic component. Applying this method on the data obtained from the stream sediment analysis, the data was partitioned into 3 clusters. Next, the residual value was achieved based on these three clusters with the fuzzification degree of 1.6. Finally through applying P.N method on the residual values, the anomaly of Pb and Zn was clarified to be between the longitudinal coordinates of x=332000 & x=335000 and y=3923000 and y=3926000.
Reference


[3]- Geological map of Razan (1/100000),Geological organization, Tehran, Iran, 1979.


