# Exploration Geochemistry Data-Application for Cu Anomaly Separation Based On Classical and Modern Statistical Methods in South Khorasan, Iran

Aref Shirazi
Amirkabir University of Technology (Tehran Polytechnic)
Mining and Metallurgical Engineering Department
Tehran, IRAN

Ardeshir Hezarkhani
Amirkabir University of Technology (Tehran Polytechnic)
Mining and Metallurgical Engineering Department
Tehran, IRAN

Adel Shirazy
Shahrood University of Technology
Faculty of Mining, Petroleum and Geophysics
Shahrood, IRAN

Abstract: The polymetal mining area is located 30 kilometers northwest of Birjand, South Khorasan Province of Iran. Considering the importance of recognizing the geochemical limit value for post-analysis studies, the limit value (= non-normative visualization) in the data of the stream was identified and described using the classic and modern statistical method. Sampling method in this area was lithogeochemical samples. Simple statistical methods, K-Means, K-Medoids, Fuzzy C-Mean (FCM), Self-Organized Map (SOM), have been used in this study. Anomaly maps are depicted in each method and separated from the background. Each method showed different anomalies, but the K-Mean and K-Medoids methods had similar responses.

Keywords: FCM; SOM; K-Means; Classical statistics; Anomaly separation; Exploration geochemistry; Copper

## 1. INTRODUCTION

Separation of anomalies from the background is one of the most important and key steps in geochemical exploration. There are several ways to identify and separate anomalous areas from the field. These methods include the classical statistics, K-Means, K-Medoids, Fuzzy C-Mean (FCM), Self-Organized Map (SOM) used in this study. In general, anomalous areas are also useful in surface surveys to determine the location of exploratory drilling. Also, by studying more precisely, the resulting maps can be used to structure the sampling grid for later, more coherent steps as well. [1,2]

## 2. Geolocation of the area

This area is located in the geographical location of 32° 58' 23" northern latitudes and 58° 56' 7" eastern longitudes in the south Khorasan province of, Iran. It is called Siojan. Siojan mining area is located 30 kilometers from Birjand city (see Figure 1 and 2).
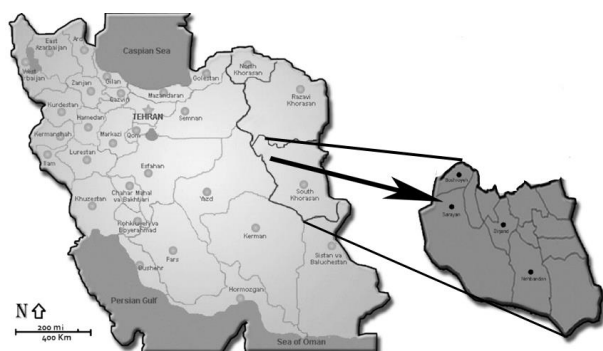


Figure. 1 South Khorasan Province Location and Siojan, Iran Map [3]





Figure. 2 Geolocation and access way of studied area.

### 2.1 Geological setting

This region is located in the northwest of the Lut zone in terms of the division of the structural states of the Iranian crust in the east of the Central Block of Iran. The Nehbandan

fault system, which covers all the Sistanian states, enters into the Lut zone in the northern part as a direction to the west. [4] The area generally consists of volcanic and intermediate cenozoic rocks in the form of andesitic masses, basaltic andesite, dacite and intrusive rhyodacite. In some cases, this formation has caused alteration and mineralization in the region. [5] Pyroclastic rocks such as altered acid tuffs, and cuttings with andesitic elements are also present in the region. Non-volcanic units such as conglomerate, sandstone, salt salts and young alluvium are found in the area. [6,7]

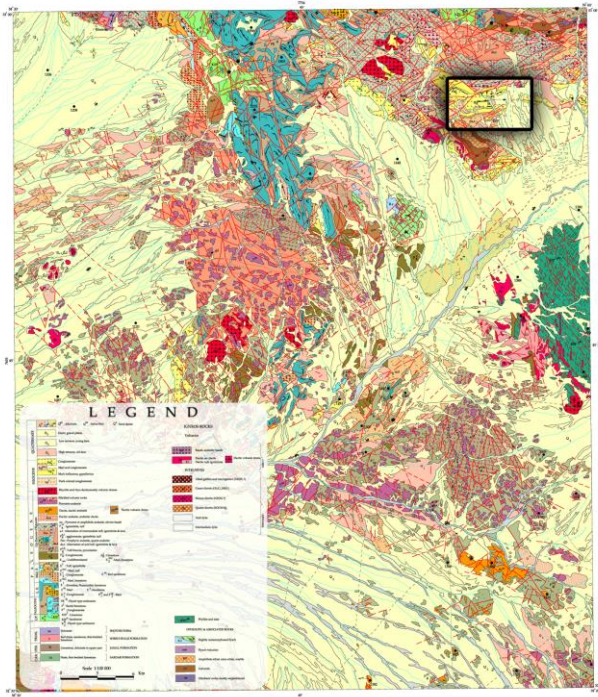You can see the location of Siojan area in geological map in figure 3.



Figure. 3 Geological map of Khusf on scale 1:100000, The Siojan area is shown in the map. [8]

## 3. Sampling

The area was covered with 120 samples of lithogeochemicals. The samples were analyzed as 44 elements. The sample preparation method was Aquaragia. And samples were analyzed by ICP-MS method. You can see the location of the samples in the UTM system in Figure 4.
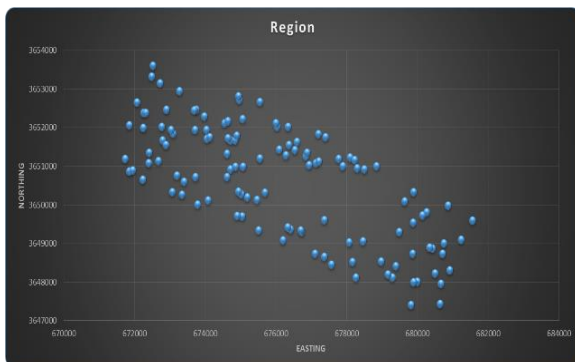


Figure. 4 locations of lithogeochemical samples in the area.

## 4. Geochemical Anomaly Separation Methods

Separation of geochemical anomalies from background has always been a major concern of exploration geochemistry. The search for methods that can make this analysis quantitative and objective aims not only at the reduction of subjectiveness but also at providing an automatic routine in exploration, assisting the interpretation and production of geochemical maps. [9,10]

## 4.1 Statistical Methods

In classical methods, anomalies are usually detected, regardless of the location of each instance, and only by formulating relationships. Commonly updated methods can be found in Formula 1. [1,9]

$$Anomaly = (\overline{X})Or(median) + 3(S)Or(MAD) \quad (1)$$

Here, x is the mean, median is the same. And S is the standard deviation and the MAD is the mean / median with the absolute value of the deviation difference [11,12], which has the following two formulas.

$$MAD = Mean|Xi - Mean|$$
$$MAD = Median|Xi - Median|$$

By placing the above values in Formula 1, we can reach different ranges.

## 4.2 K-Means and K-Medoids

The K-Means method, despite simplicity, is a basic method for many other clustering methods (such as fuzzy clustering). This method is a monolithic and flat method. For this algorithm, different shapes are expressed. But they all have repetitive routines that try to estimate the following for a fixed number of clusters [13]:

- Getting points as centers of clusters, which are actually the same average points belonging to each cluster.

- Assigning each data sample to a cluster that gives the data a minimum distance to the center of that cluster.

In this method, first, the number of clusters needed for points is randomly selected. Then the datas are attributed to one of these clusters according to the degree of similarity. And so new clusters are achieved. By repeating the same procedure, it is possible to calculate new centers for each replication by averaging the data, and the data are re-assigned to new clusters. This process continues as long as there is no change in the data. If this method is such that each cluster is displayed with one of the objects located near the center, then the " K-Medoids " is called. [14,15]

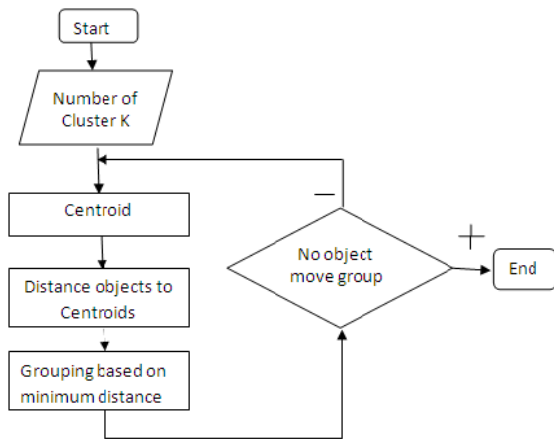Calculation flowchart of the method algorithm is shown in figure 5.

Figure. 5 Calculation flowchart of method algorithm. [16]

## 4.3 Fuzzy C-Means (FCM)

FCM is a separation clustering method that uses the Euclidean distance similarity criterion to measure the similarity of data and clusters. In other words, this algorithm identifies spherical clouds of points in a p-dimensional space. These clusters are approximately equal in size. Each cluster is displayed with its center. This mode of displaying clusters is also called a model or example, because it is often referred to as a representative of all the data assigned to the cluster. [1,9]

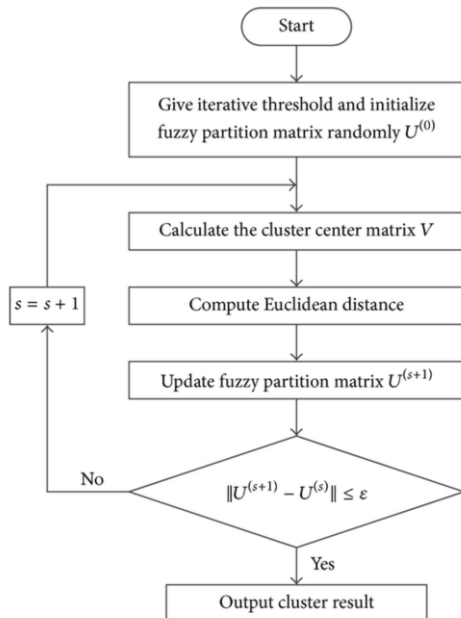Calculation flowchart of fuzzy C-Means clustering algorithm is shown in figure 6.



Figure. 6 Calculation flowchart of the Fuzzy C-Means clustering algorithm. [17]

As an meter for distance, Euclidean distance is used between a point and a sample. In order to select the center of the cluster, as the name of the algorithm finds, the mean value is used. To compute the center of the cluster, the sum of the degrees of the membership of each element is divided by the power of M in itself into the product of the power of the M

degree of membership. [18] The problem with this algorithm is that the algorithm cannot identify clusters of different shapes, sizes and densities. To identify other shapes instead of the identity matrix, we can use other matrixes to determine the distance. Such as a diameter matrix to detect elliptical clusters. The benefits of this algorithm are ease, which reduces computational time. In practice, with little repetition, it can reach a near-final solution. [19]

## 4.4 Self-Organized Map (SOM)

In the self-organization map, the competitive learning method is used for training. This method is based on specific characteristics of the human brain. The cells in the human brain are organized in different regions in different sensory regions, with rigorous and meaningful computational maps. For example, sensory inputs for touch, hearing, etc. are associated with a significant geometric arrangement in different regions. [20]

In this method that is called SOM or sometimes called SOFM (Self-Organizing Feature Map) Processor units are placed in the nodes of a one-dimensional grid, two-dimensional or more. Units are organized in a competitive learning process rather than input patterns. The place of the units set in the network is organized in such a way as to create a meaningful coordinate system on the network for input characteristics. Therefore, a self-organized map forms a topographic map of the input patterns in which the location of the units corresponds to the inherent characteristics of the input patterns. [21,22]

The competitive learning used in this grid is that in each step of the learning, the units compete to engage with each other, At the end of a competition stage, only one unit wins, which weighs in a different way than the weights of other units. This type of learning is called "uncontrolled learning." [23]

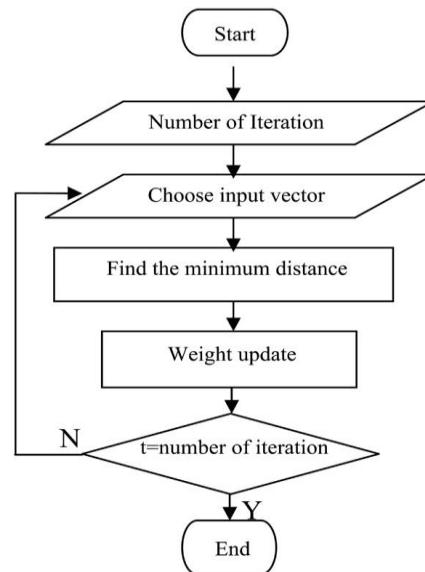Calculation flowchart of the self-organization map algorithm is shown in figure 7.



Figure. 7 Calculation flowchart of the Self-Organization Map (SOM) algorithm. [24]

## 5. General Map of Copper Concentration

In order to improve the understanding of the anomal areas, the main map of the dispersion of copper concentration in the region can be estimated as a preview. After analyzing the sample dispersion and copper concentration, a map was prepared with the estimation of other points by Kriging method [25,26].

In this map, we can see concentrations with respect to the color as well as the contours. (see Figure 5).
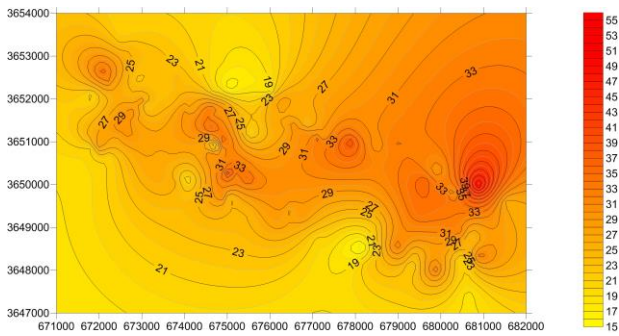


Figure. 5 Concentration map of copper element along the range with the conjugated estimator.

## 6. Maps of Anomaly Seperation Methods

In this section, the anomalies of the copper element in the study area, which have been obtained by different methods of anomaly seperation from the background, are mapped out.

## 6.1 Statistical Methods

As discussed in Section 4.1, the threshold can be obtained from three relative-like formulas. The copper anomaly map depicted using pure classical statistics (use of mean and standard deviation) is shown in Figure 6.
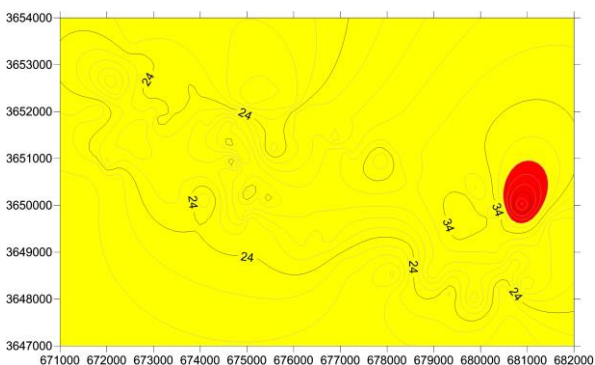


Figure. 6 Pure statistics (using mean and standard deviation) - The threshold value is 34.3 ppm.

The copper anomaly map depicted by using the updated statistical method of the first type (using MAD and mean) is shown in figure 7.
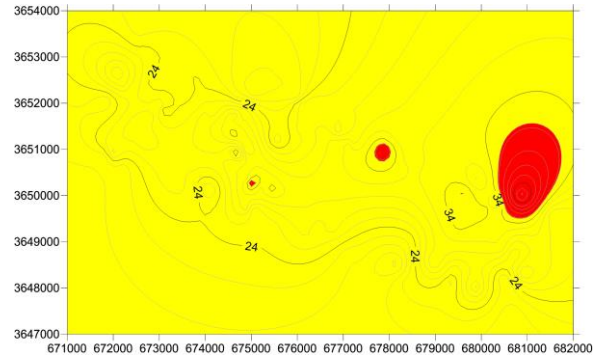


Figure. 7 The updated type I statistical method (using MAD and Mean) - The threshold value is 35.7 ppm.

In the following, the anomaly of the copper element derived from the statistical updated method of the second type (using MAD and Median) is shown in figure 8.
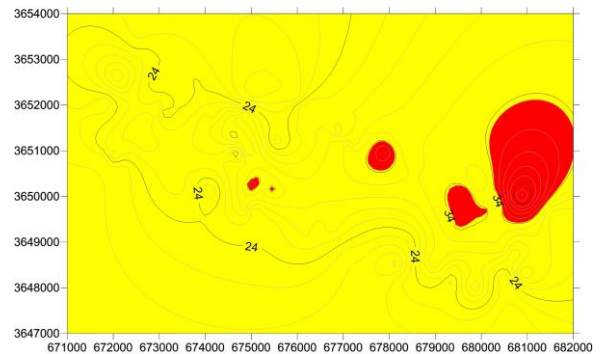


Figure. 8 The updated type I statistical method (using MAD and Mean) - The threshold value is 38.1 ppm.

As shown in the maps, The Copper threshold values were obtained for pure statistical methods, the updated type of statistical method, and the second type updated method were 34.3 ppm, 35.7 ppm and 38.1 ppm, respectively.

## 6.2 K-Means and K-Medoids

Due to the close approach of the two methods, the threshold value in both methods was estimated to be similar to each other. Figure 9 shows a map that confirms both methods. (Red color is anomaly and yellow color is background).
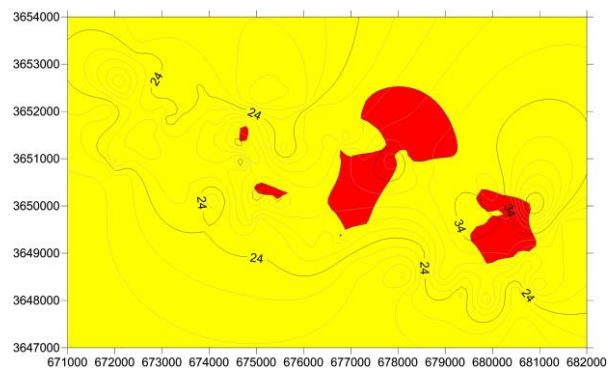


Figure. 9 Estimation of anomalies for the KaMeS and KaMoides method.

## 6.3 Fuzzy C-Means (FCM)

In the FCM method, due to the percentage probability of each sample being attributed to the anomaly, 80% was considered probable, since 80% could be a good possibility for diagnosis [27] and could reduce the wasting caution of much of the backgrounds. Figure 10 shows the map of the anomalous areas derived from the FCM method. (Red color is anomaly and yellow color is background).
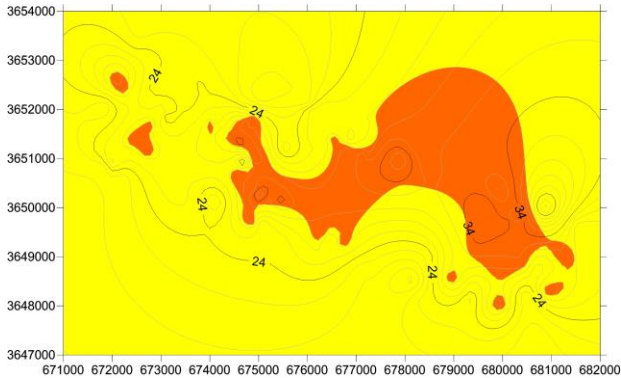


Figure. 10 Estimated anomalies by FCM method with consideration of 80% anomaly probability

## 6.4 Self-Organization Network Method

According to this method, the following map (Figure 11) emerges which represents two groups of anomalies (red) and a background (yellow).
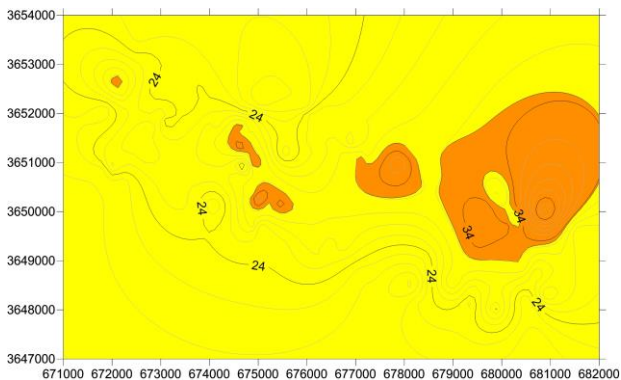


Figure. 11 Anomaly Estimation for SOM Method

## 7. Conclusion

The Siojan Polymetal Mineral Area is located in South Khorasan province, 30 km northwest of Birjand city. Considering the importance of recognizing the geochemical limit value for post-analysis studies, the limit value in the data of the lithogeochemical samples were determined using 5 statistical methods.

Methods such as classic statistics, K-Mean, K-Medoids, Fuzzy C-Mean clustering (FCM), Self-Organized Map (SOM) were used in this study. Anomalies were mapped in each method and the anomalies were seperated from the field. Each method showed different anomalies, but the K-Mean and K-Medoids methods had similar responses.

By matching all the maps, it can be seen that each method can be efficient and should not rely on a method.

It is recommended to use field studies and re-sampling by estimating the anomaly maps, the accuracy of each method for the region is estimated and the best method for further studies is to be selected.

## 4. REFERENCES

[1] Hasani Pak, A.A., Sharafodin, M. (2012). Exploration data analysis. University of Tehran Publication. Tehran.

[2] Govett, G. J. S., Goodfellow, W. D., Chapman, R. P., & Chork, C. Y. (1975). Exploration geochemistry-distribution of elements and recognition of anomalies. Journal of the International Association for Mathematical Geology, 7 (5-6), 415 - 446.

https://doi.org/10.1007/ BF02080498

[3] Emami, M. H. (1972). Geology and petrological investigation on Shahkuh volcanic rocks, south of Birjand, eastern Iran. University of Tehran, Tehran, Iran.

[4] Walker, R. T., & Khatib, M. M. (2006). Active faulting in the Birjand region of NE Iran. Tectonics, 25(4).

[5] Hoseinpoor, M. K., & Aryafar, A. (2016). Using robust staged R-mode factor analysis and logistic function to identify probable Cu-mineralization zones in Khusf 1: 100,000 sheets, east of Iran. Arabian Journal of Geosciences, 9(2), 157.

https://doi.org/10.1007/s12517-015-2266-9

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

[7] Ghorbani, M. (2002). An introduction to economic geology of Iran. Tehran, Iran. Geological survey and mineral explorations of Iran (in Persian).

[8] Vahdati Daneshmand, F. (1989). 1: 100,000 Geology Map of Khusf, Geological survey of Iran.

[9] de Mulder, E. F., Cheng, Q., Agterberg, F., & Goncalves, M. (2016). New and game-changing developments in geochemical exploration. Episodes, 39(1), 70-71. http://dx.doi.org/10.22059/ijmge.2014.53107

[10] Journel, A. G., & Huijbregts, C. J. (1978). Mining geostatistics. Academic press.

[11] Howell, D. C. (2014). Median absolute deviation. Wiley StatsRef: Statistics Reference Online.

[12] Twain, M., & Weather, N. E. (2004). Mean Absolute Deviation. Dynamic Portfolio Theory and Management, 235. https://doi.org/10.1287/mnsc.37.5.519

[13] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.

https://doi.org/10.1016/S0031-3203(02)00060-2

[14] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), 100-108. http://dx.doi.org/10.2307/2346830

[15] Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. Expert systems with applications, 36(2), 3336-3341. https://doi.org/10.1016/j.eswa.2008.01.039

[16] Sohaib, M., & Mushtaq, Q. (2013). Dimensional Reduction of Hyperspectral Image DataUsing Band Clustering and Selection Based on Statistical Characteristics of Band Images. International Journal of Computer and Communication Engineering, 2(2), 101. http://dx.doi.org/10.7763/IJCCE.2013.V2.148

[17] Wang, Z. T., Zhao, N. B., Wang, W. Y., Tang, R., & Li, S. Y. (2015). A fault diagnosis approach for gas turbine exhaust gas temperature based on fuzzy c-means clustering and support vector machine. Mathematical Problems in Engineering, 2015. http://dx.doi.org/10.1155/2015/240267

[18] Gary, A. C., Wakefield, M. I., Johnson, G. W., & Ekart, D. D. (2009). Application of fuzzy c-means clustering to paleoenvironmental analysis: example from the Jurassic, central North Sea, UK. Geologic Problem Solving with Microfossils: A Volume in Honor of Garry D. Jones. SEPM Special Publication, 93, 9-20. http://dx.doi.org/10.2110/sepmsp.093.009

[19] Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191-203. https://doi.org/10.1016/0098-3004(84)90020-7

[20] Carpenter, G. A., & Grossberg, S. (Eds.). (1991). Pattern recognition by self-organizing neural networks. MIT Press.

[21] Kohonen, T. (1998). The self-organizing map. Neurocomputing, 21(1-3), 1-6. http://dx.doi.org/10.1016/S0925-2312(98)00030-7

[22] Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9), 1464-1480. https://doi.org/10.1109/5.58325

[23] Kohonen, T., Hynninen, J., Kangas, J., & Laaksonen, J. (1996). Som pak: The self-organizing map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science.

[24] Anifah, L., Purnama, I. K. E., Hariadi, M., & Purnomo, M. H. (2013). Osteoarthritis classification using self-organizing map based on gabor kernel and contrast-limited adaptive histogram equalization. The open biomedical engineering journal, 7, 18. http://dx.doi.org/10.2174/1874120701307010018

[25] Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System, 4(3), 313-332. http://dx.doi.org/10.1080/02693799008941549

[26] Adhikary, P. P., Dash, C. J., Bej, R., & Chandrasekharan, H. (2011). Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India. Environmental monitoring and assessment, 176(1-4), 663-676.

https://doi.org/10.1007/s10661-010-1611-4

[27] Cheng, Q. (1999). Spatial and scaling modelling for geochemical anomaly separation. Journal of Geochemical exploration, 65(3), 175-194. https://doi.org/10.1016/S0375-6742(99)00028-X