

Estimation of geochemical elements using a hybrid neural network-Gustafson-Kessel algorithm

M. Jahangiri*, S.R. Ghavami Riabi and B. Tokhmechi

Faculty of Mining, Petroleum & Geophysics Engineering, Shahrood University of Technology, Shahrood, Iran

Received 12 March 2017; received in revised form 9 August 2017; accepted 5 November 2017

*Corresponding author: m.jahangiri@shahroodut.ac.ir (M. Jahangiri).

Abstract

Bearing in mind that lack of data is a common problem in the study of porphyry copper mining exploration, our goal was set to identify the hidden patterns within the data and to extend the information to the data-less areas. To do this, the combination of pattern recognition techniques has been used. In this work, multi-layer neural network was used to estimate the concentration of geochemical elements. From 1755 surface and boreholes data available, analyzed by ICP, 70% was used for training, and the rest for testing. The average accuracy of estimators for 22 geochemical elements when using all data was equal to 75%. Based on validation, the optimal number of clusters for the total data was identified. The Gustafson-Kessel (GK) clustering was used to design the estimator for the geochemical element concentrations in different clusters, and the clusters were selected for estimation. The results obtained show that using GK, the estimator's average accuracy increase up to 84%. The accuracy of the elements Zn, As, Pb, Mo, and Mn with low accuracies of 0.51, 0.62, 0.64, 0.65, and 0.68 based on all data were developed to 0.76, 0.86, 0.76, 0.80, and 0.71 with the clustered data, respectively. The mean square error using all the data was 0.079, while in the case of hybrid developed method, it decreased to 0.048. There were error reductions in Al from 0.022 to 0.012, in As, from 0.105 to 0.025, and from 0.115 to 0.046 for S.

Keywords: *Clustering Algorithm, Estimation Precision Improvement, Gustafson-Kessel, Geochemical Elements Estimation, Neural Network.*

1. Introduction

The main purpose of geochemical exploration studies in porphyry copper mines is to identify the primary and secondary geochemical halos, mineralized elements, tail, and above front halo elements. Estimation of element concentration is one of the methods used for evaluation of the behavior of elements in geochemical dispersion halos. Estimation of geochemical concentration is a challenge in mining engineering and geology [1]. Geochemical element distribution is the end product of diverse geochemical processes operating at a wide range of scales, and usually interacting with each other via various ways [2-4]. Exploring the geochemical patterns based on data analysis is a convenient and effective way to improve our understanding of the geochemical characteristics of a given studied area [5].

Estimation of element concentration is a common method for evaluation of mineralization in depth [6]. The precision of estimation is the most important parameter that depends on the method, and is identified by specific indices [7]. Different methods such as geo-statistics [8, 9], artificial neural networks (ANNs) [10-13], and fuzzy logic are used for this purpose [14, 15].

Numerous limitations exist in training and designing networks using neural networks. For instance, a relatively great amount of data from the studied area is required when using neural networks. In the network architecture, the weight of input data is one of the most important items that has a direct effect on performance. The weights are controlled by the neural network structure and learning algorithm parameters. The

parameters affecting the performance of neural network include the input type, number of hidden layers and their nodes, memory, and learning rate [16]. Researchers have attempted to combine ANNs with other optimization methods to increase the estimation precision and to optimize ANN for concentration estimation using the Levenberg–Marquardt (LM) algorithms and the genetic algorithm [16, 17].

Data mining and pattern recognition techniques (PRTs) have been widely used in the Earth sciences for effective, scalable, and flexible data analyses [18, 19]. PRTs have been used to classify the data into a number of various classes [20, 21]. PRTs have been used in the identification of geological information, hidden mineralization in the geochemical data, and determination of anomaly and background pattern determination since 1970s [22]. These methods have also been used in the examination of relationships between regional geochemical patterns and large deposits [23, 24].

The data with the highest sample correlation arranged within the same cluster and clustering algorithms has the ability to find these clusters. Using the data with a greater sample correlation has improved the relationship between the estimator input and output and increased the estimation accuracy. This condition was not considered in the earlier studies, and the precision of the estimation was reduced. In the present research work, the pattern recognition methods such as fuzzy clustering were used to increase the precision of estimating the concentration of geochemical elements.

In this work, multi-layer perceptron (MLP) was applied to identify non-linear relations between the input and output data. To increase the accuracy of neural network estimation, the Gustafson Kessel clustering method was used.

2. Geology of studied area

The Sonajil exploration area is located 17 Km from Heris, East Azarbaijan province, NW of Iran. Magmatic activities in the subduction zone including the intrusive and volcanic rocks were developed through the studied area. The studied area has a potential for copper-molybdenum-gold mineralization based on economic geology investigations [25].

Field investigations on Sonajil deposit show that host rocks including volcanic and Eocene volcano clastic rocks consist of basaltic flows (Q_v), andesite and hornblende andesite (E_{an}), mega-porphyry andesite (E_{am}), porphyry andesite (E_{ap}), microdiorite and microdioritic dikes (E_{im}), and granitoids (O_g). Porphyry andesite has covered an extensive part of the area. The Sonajil porphyry displays various alterations including potassic and phyllic ones in the surface [25] (Figure 1).

3. Geochemical data

The surface geochemical data consisted of 562 rock samples that were systematically collected with a distance of 100 m. Six exploration boreholes, a total of 2465 m drilling including 1193 rock samples were also utilized in this investigation (Figure 2).

The surface rock samples were analyzed by the ICP-MS method for 45 elements, and the borehole rock samples were analyzed for 23 elements. The same elements for the surface and borehole were selected for this research work, which are listed in Table 1.

4. Algorithms

4.1. Artificial neural networks (ANNs)

ANNs are based upon complex structures of human brain, which consists of millions of neural cells interacting with each other [26]. ANN identifies the relationships between the input and output variables through a group of processing units called neurons. They are capable of identifying the complex relationship between the input and output data [27].

The main feature of neural networks is based upon training samples. This feature produces a highly practical and computational model that is used in different research fields. In the condition that there is no accurate understanding of the nature, it becomes more applicable [28].

ANN consists of the input, hidden, and output layers. Data is used in the input layer, and all the other layers cooperate in processing to produce the final output [29].

Each ANN except for the input layer consists of five parts of input matrices (P), weight matrix (W), combination function (Σ), activation function (f), and output (a) [30] (Figure 3).

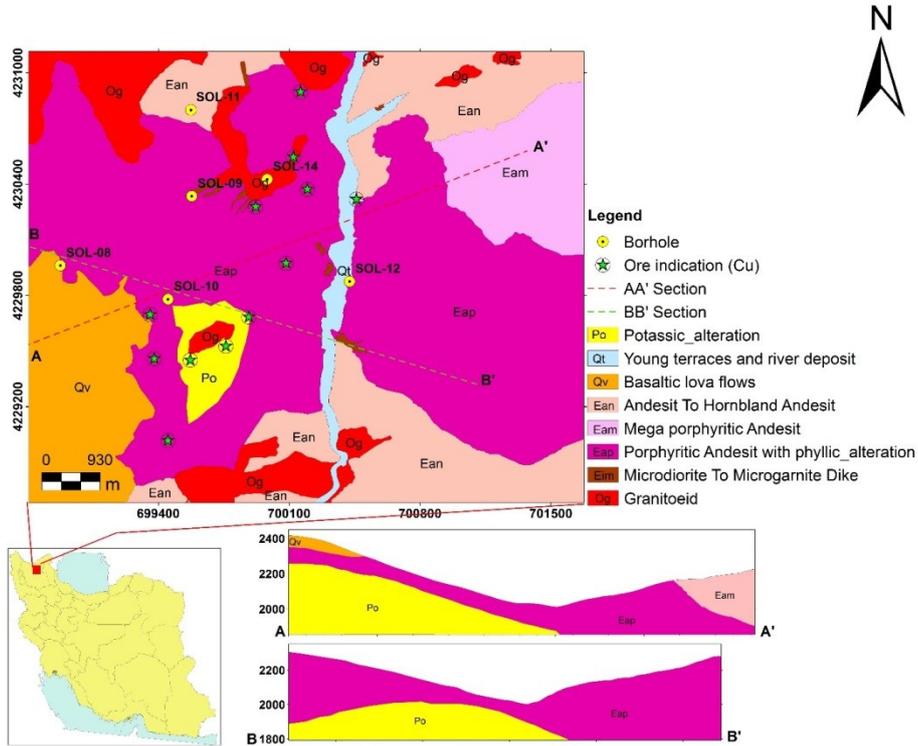


Figure 1. Geological map of eastern part of Sonajil-Heris exploration zone [Modified, 25].

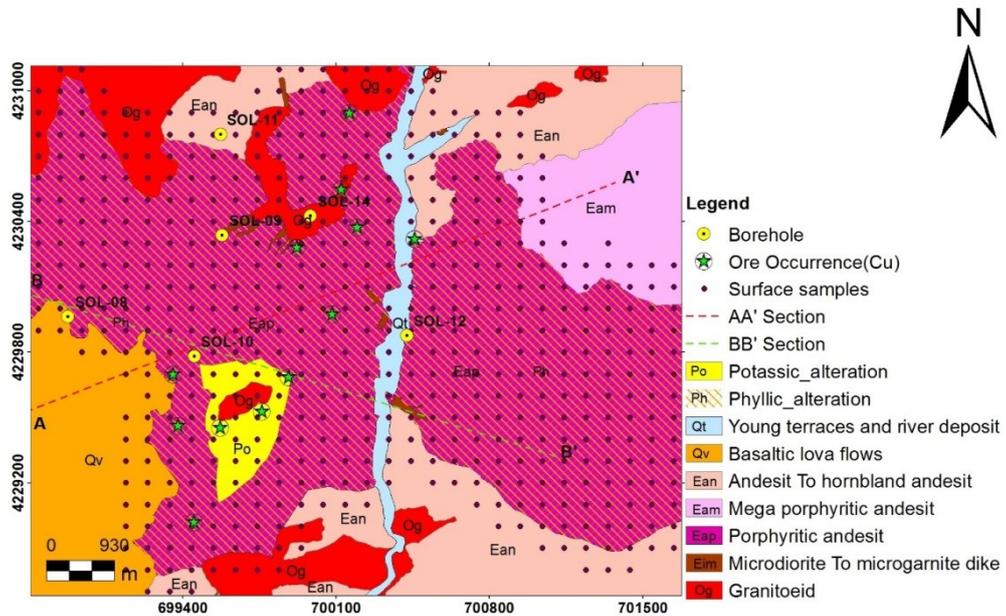


Figure 2. Position of surface samples and location of boreholes in studied area.

Table 1. Selected surface and borehole geochemical elements.

Al	As	Ba	Be	Ca	Co	Cu	Fe	K	La	Mg
Mn	Mo	Na	Ni	P	Pb	S	Sc	Sr	V	Zn

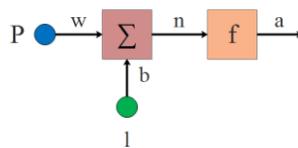


Figure 3. Different parts of ANN including input layer, hidden layer, and output layer.

4.2. Gustafson-Kessel clustering algorithm

The Gustafson-Kessel (GK) algorithm, as one of the soft clustering methods, could identify linear and elliptical clusters using an induction matrix. In the GK algorithm, each cluster is created based on one point and one matrix in which the point determines the center of the cluster, and the matrix defines its covariance. Hence, the GK algorithm has the ability of elliptical cluster separation, and unlike most clustering methods, it has no limitation for creating circular clusters [31].

Many fuzzy clustering algorithms are based upon optimization of one scheme, and they tend to minimize an objective function such as $J(V, U)$, which shows data clustering connection error [32].

$$J(V, U) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^\beta d_{ik}^2 \quad (1)$$

where u_{ik} is the membership degree of point data x_k to the model of pattern i (center of cluster), $U=[u_{ij}]$ is the segmentation matrix with dimension of $c \times N$, $V=[v_i]$ is the initial sample matrix with dimension of $c \times q$, and d_{ik} is the distance between the point data K and sample i . The parameter $\beta > 1$ is the exponential weight, which controls the segmentation fuzzification.

The GK algorithm is able to identify not only the spherical clusters but also the elliptical cluster. This algorithm is an expansion of the standard algorithm FCM using the adaptive distance norm to identify clusters from various geometrical shapes in dataset.

The adaptive distance for each separate cluster is derived from Equation (2), and the inductive norm matrix ($S_i=1 \dots c$) is calculated from Equation (3):

$$d_{ik}^2 = \|x_k - v_i\|_{S_i}^2 = (x_k - v_i)^T S_i (x_k - v_i) \quad (2)$$

$$S_i = [\rho_i \det(F_i)]^{1/q} F_i^{-1} \quad (3)$$

where q is the number of initial data features, ρ_i is the volume of cluster i , and F_i is the fuzzy covariance matrix that is calculated based on the following equation:

$$F_i = \frac{\sum_{k=1}^N (u_{ik})^\beta (x_k - v_i)(x_k - v_i)^T}{\sum_k (u_{ik})^\beta} \quad (4)$$

Minimization of the cost function $J(V, U)$, subject to the limitation of $\sum_{i=1}^c u_{ik} = 1$, is performed by

an iterative algorithm that optimizes center of clusters and membership degree [33]:

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^\beta x_k}{\sum_k (u_{ik})^\beta}, i = 1 \dots c, k = 1 \dots N \quad (5)$$

and

$$u_{ik} = \frac{1}{\sum_{j=1}^c (d_{ik} / d_{jk})^{2/\beta-1}}, i = 1 \dots c, k = 1 \dots N \quad (6)$$

4.3. Validation indices

In order to estimate the number of clusters, validation indices are utilized, which determine the compression or concentration and the separation. The first group, compression indices, exclusively use clustering memberships, while the second one, separation indices, are used to evaluate memberships in connection with the data. For the first group, Partition Coefficient (PC) and Classification Entropy (CE) are often calculated [34]. For the second group, Xie-Beni, XB [35], Partition, SC, and Separation (S) index are applied for validation [36].

(a) PC index has been developed to measure the overlap of clusters by Bezdek [34], as follows:

$$PC = \frac{1}{N} \sum_{i=1}^m \sum_{k=1}^N \mu_{ik}^2 \quad (7)$$

in which, N denotes the number of data and μ_{ik} is the membership of K point data at cluster i .

(b) CE index, which measures the amount of fuzziness, is defined as follows [34]:

$$CE = \frac{1}{N} \sum_{i=1}^m \sum_{k=1}^N \mu_{ik} \log_a \mu_{ik} \quad (8)$$

When the number of clusters are evaluated, the PC and CE Indices reach zero and one, respectively, for the best cluster number. The PC and CE Indices are sensitive to the β parameter.

When the PC Index approaches $1/\beta$, the results become more ambiguous. A continuous decrease in the PC index, with respect to β , is the main weakness for it and has no direct dependency on the data. However, the PC index shows how much the clusters overlap. The CE index has the same disadvantages. As the number of clusters rises, the PC index reduces, while the CE index increases. Estimations of both the PC and CE indices relatively denote how many clusters overlap.

(c) The Xie-Beni index is defined to determine the ratio of the total variation within

clusters and separation of clusters based on the following equation [35]:

$$XB = \frac{\sum_{i=1}^m \sum_{k=1}^N \mu_{ik}^\beta \|x_k - u_i\|^2}{N \cdot \min_{i,k} \|x_k - u_i\|^2} \quad (9)$$

x_k denotes the studied sample and u_i is the center of cluster.

The Xie-Beni index is focused on the feature condensation and separation. As more clusters are partitioned from each other, the Xie-Beni index adopts the least value [34].

(d) The SC index is a proportion of the total condensation to separate clusters from each other, which is defined by Equation (10):

$$SC = \sum_{i=1}^m \frac{\sum_{j=1}^N \mu_{ij}^\beta \|x_j - u_i\|^2}{N_i \cdot \sum_{k=1}^m \|u_k - u_i\|^2} \quad (10)$$

The more SC index is reduced, the better clustering is derived. This index is useful when different clusterings having the same clusters are compared with each other [36, 37].

The XB and SC indices exhibit the amount of noise among clusters and the resolution of separation, respectively. In other words, with the highest XB index and lowest SC index, the best separation of the cluster is performed.

(e) The separation index S uses a minimum separation distance for segmentation verification, and is defined as follows [36]:

$$S = \frac{\sum_{i=1}^m \sum_{k=1}^N \mu_{ik}^2 \|x_k - u_i\|^2}{N \cdot \min_{i,k} \|u_i - u_k\|^2} \quad (11)$$

Index S represents an optimum segmentation. The more the S index, the better the clusters are separated. When prediction of cluster number is unknown, the validation indices are useful.

5. Estimator design

5.1. Estimator design based on whole data

The MLP method is used to estimate the geochemical element concentrations. To design estimators, all the available geochemical

information including 1755 surface and borehole rock samples for 22 elements was utilized in this work. The estimator was designed based on 70% of training and 30% of test data (Figure 4). The hidden layers were chosen to be 24 in number based on trial and error and maximum accuracy of estimators. In the input network, X, Y, and Z were considered as the coordinates of samples, and in the output network, concentrations of different geochemical elements were estimated (Figure 5). The results obtained from estimator design were calculated and shown in Figure 6. The results obtained indicated that neural network-based estimator reached the acceptable accuracies of 0.95, 0.89, 0.87, 0.87, 0.84, 0.83, and 0.81 for the elements Al, Na, K, Ba, S, Be, and Cu, respectively. In addition, except for Zn, As, Pb, Mo, and Mn with the respective accuracies of 0.51, 0.62, 0.64, 0.65, and 0.68, other elements showed a satisfactory accuracy. The average accuracy of 22 elements was about 75%.

5.2. Estimator design using clustered data

Identification of optimum clusters based on validation indices and then clustered data using GK algorithm was investigated in the next phase of study. For this purpose, 70% of the data was used for training and 30% for testing (Figure 7).

According to the indices for determining optimized cluster number (Figures 8 and 9), all indices converge to an approximate constant value after four clusters.

Based on the clustering data, there are 416, 369, 532, and 438 data in the first, second, third, and fourth clusters, respectively (Table 2).

Distribution of clusters in 3D space of coordinate axes is shown in Figure 10.

–The 3D diagram of the clustered data shows that the clusters are well-separated, and that similar samples are within the same cluster.

The results of estimator design by clustered data with GK for geochemical elements are illustrated in Figure 11.

The estimator accuracy has generally increased, and the average accuracy has reached 84%, which is a 9% increase.

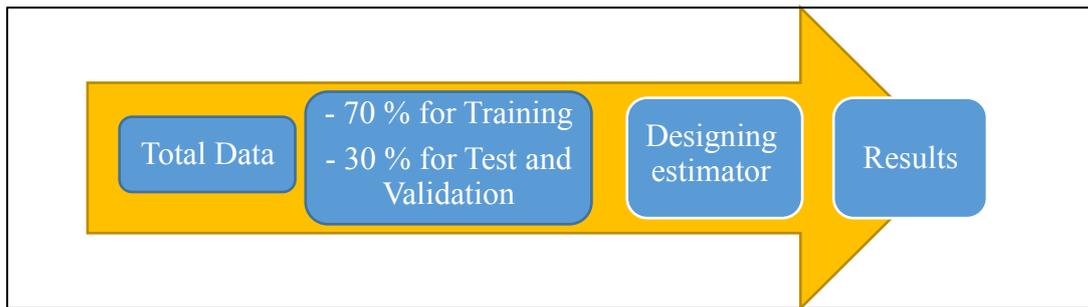


Figure 4. Estimator design using whole geochemical data.

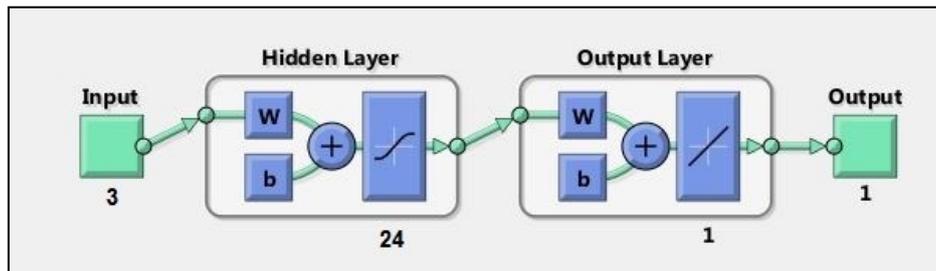


Figure 5. Estimator design model for MLP neural network.

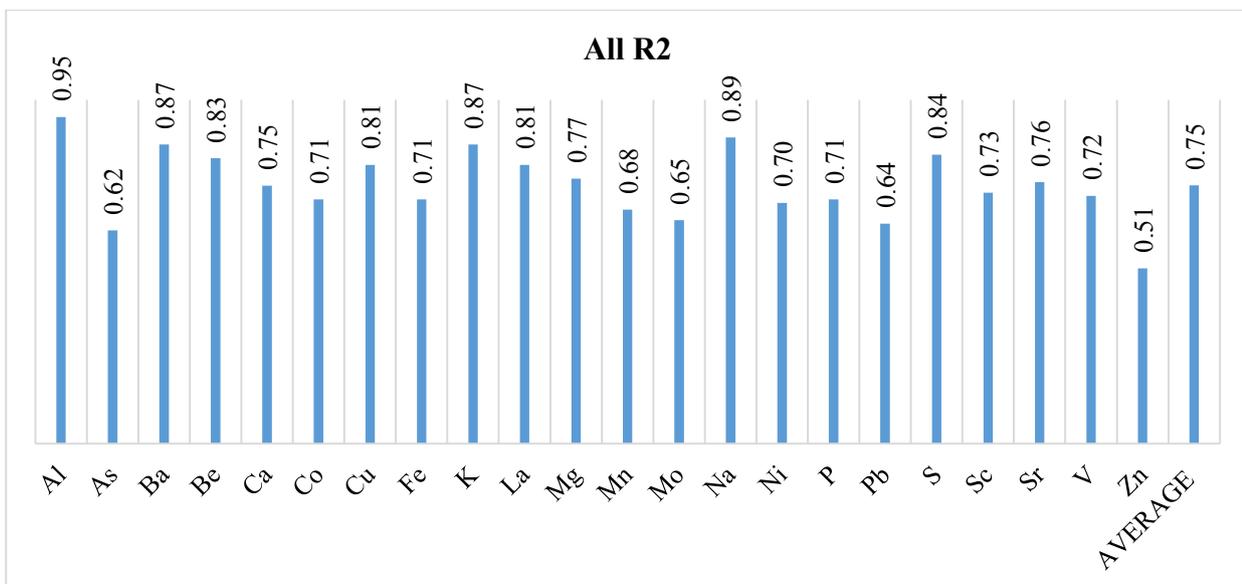


Figure 6. Estimator accuracy bar chart for all data.

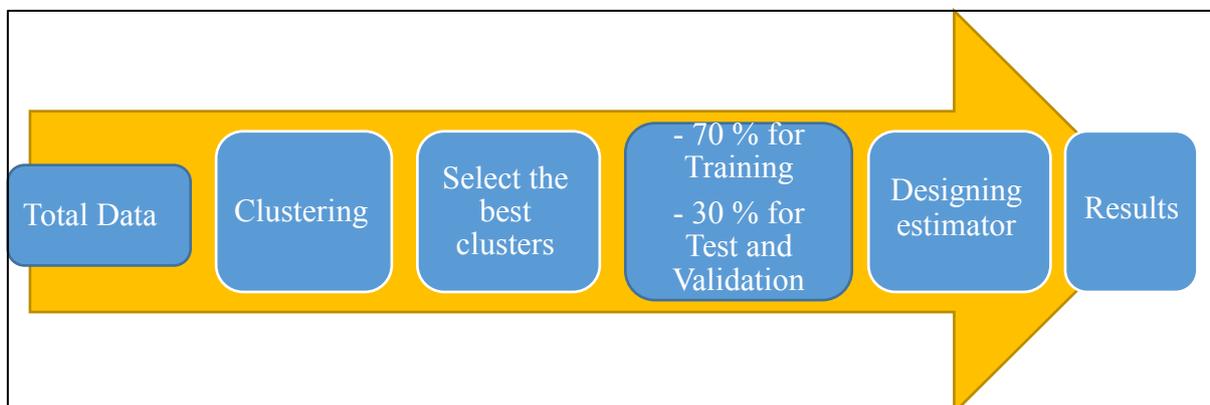


Figure 7. Estimator design using clustering data for geochemical elements.

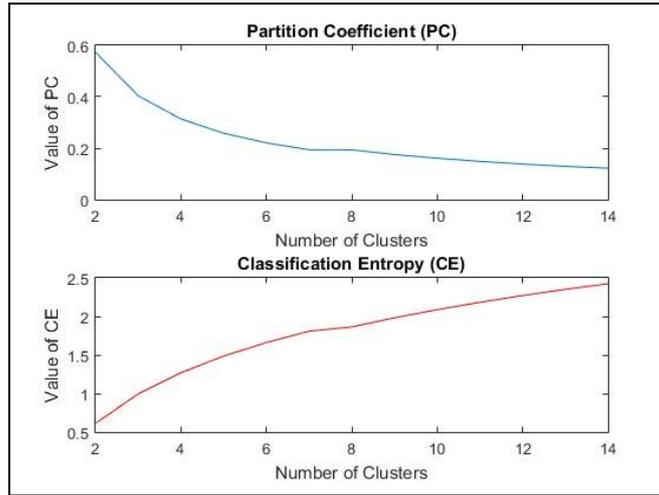


Figure 8. PC and CE validation indices.

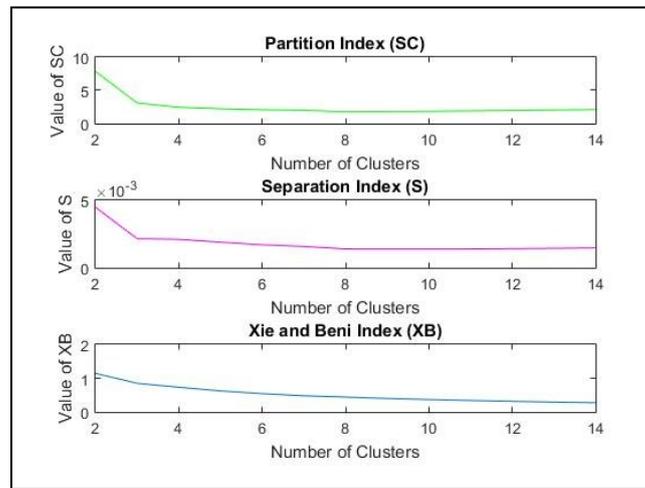


Figure 9. SC, S, and XB validation indices.

Table 2. Data clustering results using GK clustering algorithm.

Clustering algorithm	Total data	Number of data in different clusters			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gustafson kessel	1755	416	369	532	438

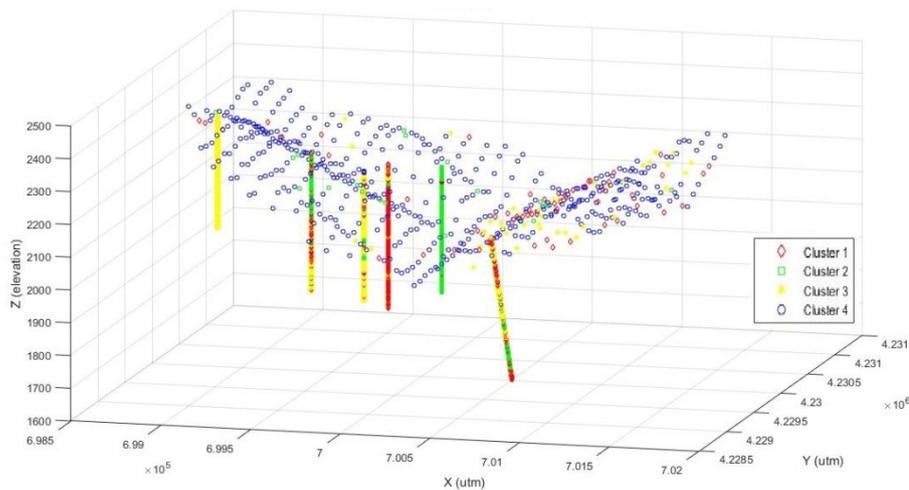


Figure 10. Distribution of clustered samples using Gustafson-Kessel method in 3D space.

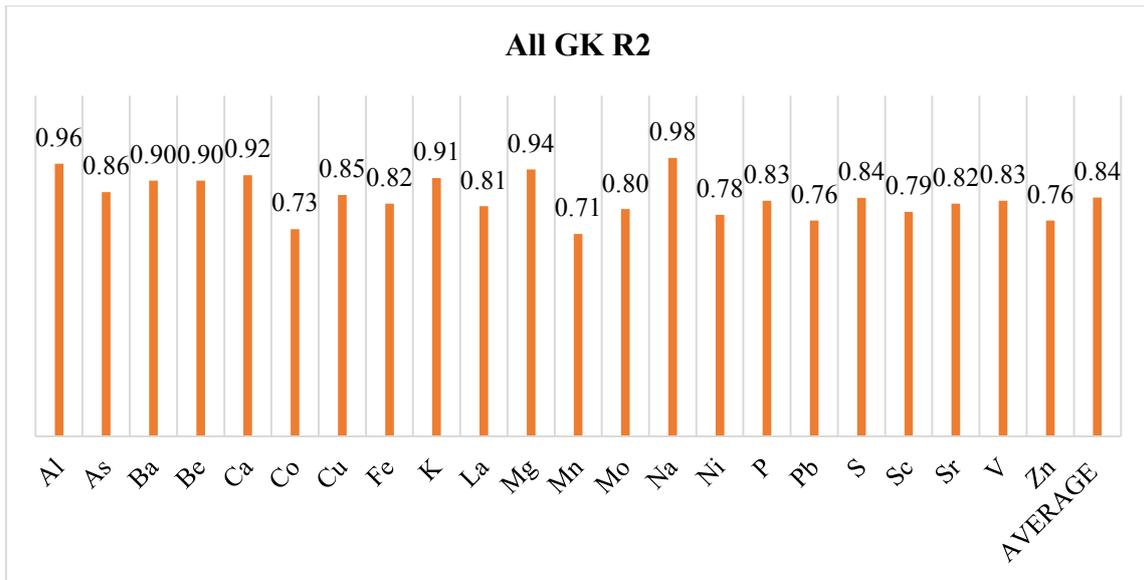


Figure 11. Estimator accuracy bar chart using GK clustered data.

6. Comparing accuracy of estimators and estimation error comparison

6.1. Comparing accuracy of estimators

The comparison of results from two estimators shows that the average accuracy of estimators when using clustered data increases dramatically from 75 to 84 percent (Figure 12).

The accuracy of the elements Zn, As, Pb, Mo, and Mn with low accuracies of 0.51, 0.62, 0.64, 0.65, and 0.68 is based upon all data, and increases to 0.76, 0.86, 0.76, 0.80, and 0.71 with clustered data. In between, the accuracy of zinc with unacceptable value reaches an acceptable value of

76% with clustering. The results of different element estimators based on all data and clustered data with GK method are demonstrated in Table 3.

Clustering algorithm not only increases the estimation accuracy of elements but also develops the estimation validity. For a better understanding, the regression graph of Zn and Al estimators with the highest and the lowest accuracy estimation are shown in Figure 13. While using clustering algorithm, since inter-cluster data has a better coordination, estimation accuracy increases and causes regression slope to approach $y=x$.

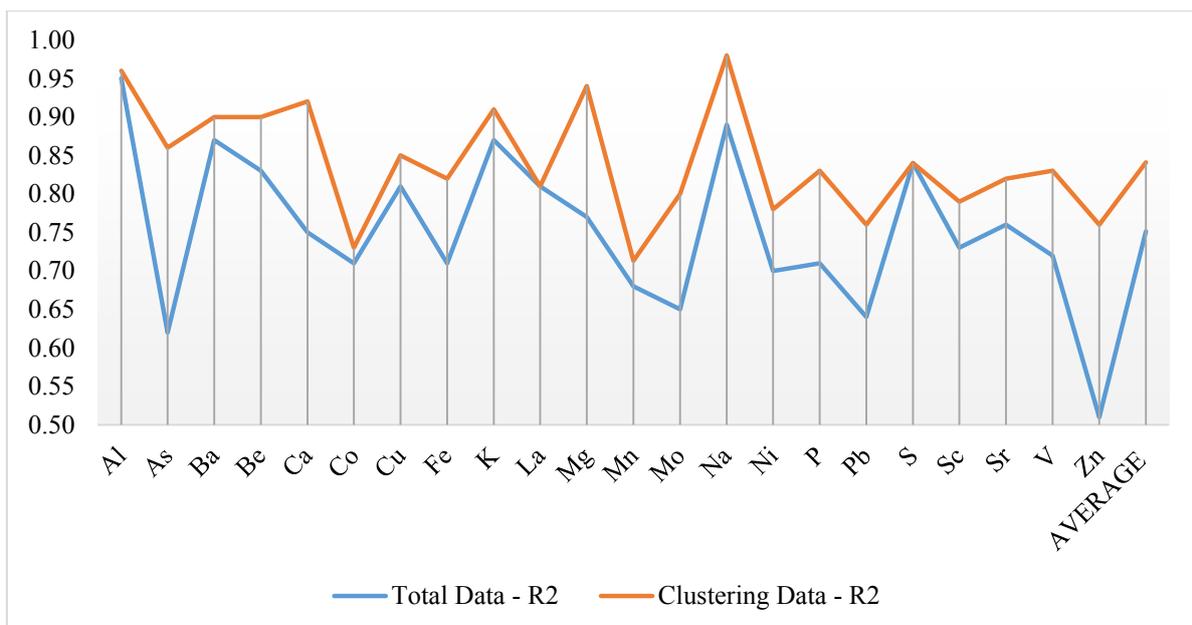


Figure 12. Comparison of accuracy of estimators (1) using all data and (2) using GK clustered data.

Table 3. Accuracy of estimating concentration of geochemical elements (1) using all data and (2) clustering algorithm.

Total data					Gustafson kessel Clusterin method				
Variable	All R2	Training R2	Validation R2	Test R2	Variable	All R2	Training R2	Validation R2	Test R2
Al	0.95	0.96	0.94	0.94	Al	0.96	0.96	0.96	0.96
As	0.62	0.64	0.57	0.60	As	0.86	0.88	0.82	0.82
Ba	0.87	0.86	0.85	0.89	Ba	0.90	0.90	0.91	0.93
Be	0.83	0.83	0.82	0.82	Be	0.90	0.91	0.88	0.87
Ca	0.75	0.77	0.76	0.67	Ca	0.92	0.95	0.88	0.86
Co	0.71	0.72	0.68	0.71	Co	0.73	0.74	0.67	0.74
Cu	0.81	0.81	0.82	0.86	Cu	0.85	0.85	0.82	0.90
Fe	0.71	0.71	0.72	0.70	Fe	0.82	0.84	0.79	0.78
K	0.87	0.87	0.86	0.85	K	0.91	0.92	0.91	0.89
La	0.81	0.82	0.78	0.74	La	0.81	0.82	0.73	0.83
Mg	0.77	0.78	0.77	0.72	Mg	0.94	0.96	0.92	0.86
Mn	0.68	0.70	0.60	0.64	Mn	0.71	0.72	0.80	0.62
Mo	0.65	0.66	0.62	0.64	Mo	0.80	0.81	0.79	0.75
Na	0.89	0.89	0.88	0.89	Na	0.98	0.99	0.96	0.96
Ni	0.70	0.70	0.70	0.69	Ni	0.78	0.80	0.73	0.76
P	0.71	0.72	0.73	0.65	P	0.83	0.86	0.82	0.70
Pb	0.64	0.65	0.62	0.60	Pb	0.76	0.76	0.73	0.76
S	0.84	0.84	0.86	0.82	S	0.84	0.81	0.86	0.92
Sc	0.73	0.73	0.73	0.74	Sc	0.79	0.81	0.76	0.75
Sr	0.76	0.76	0.71	0.80	Sr	0.82	0.82	0.80	0.88
V	0.72	0.71	0.76	0.71	V	0.83	0.86	0.88	0.74
Zn	0.51	0.53	0.54	0.43	Zn	0.76	0.80	0.77	0.61
Average	0.75	0.76	0.74	0.73	Average	0.84	0.85	0.83	0.81

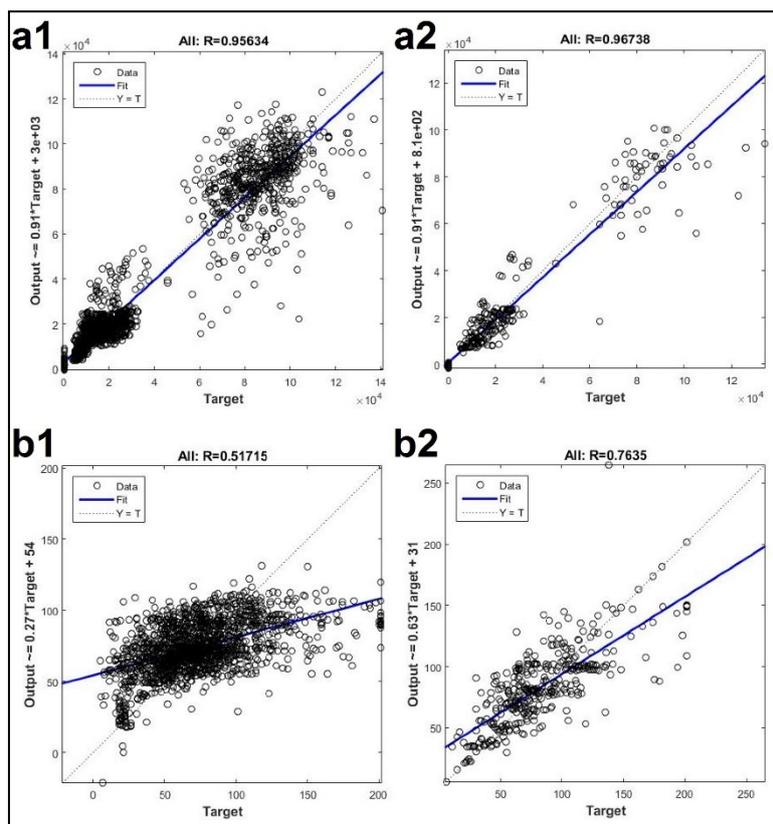


Figure 13. Regression graph of actual and estimated data; x-axis represents actual data and y-axis shows estimated data. (a.1) Regression graph of Al using all data, and (a.2) clustered data, (b.1) regression graph of Zn using all data, and (b.2) clustered data.

The comparison of concentration maps of the actual and estimated values shows that the results obtained correspond well with reality. For example, the surface geochemical maps were drawn for the actual and estimated concentration of the elements Al, Cu, and Zn (Figure 14). A

homogeneous distribution of Cu concentration is realized when using clustered data in comparison with the actual data. The geochemical spatial distribution pattern was achieved using an estimated map based on the surface and borehole data.

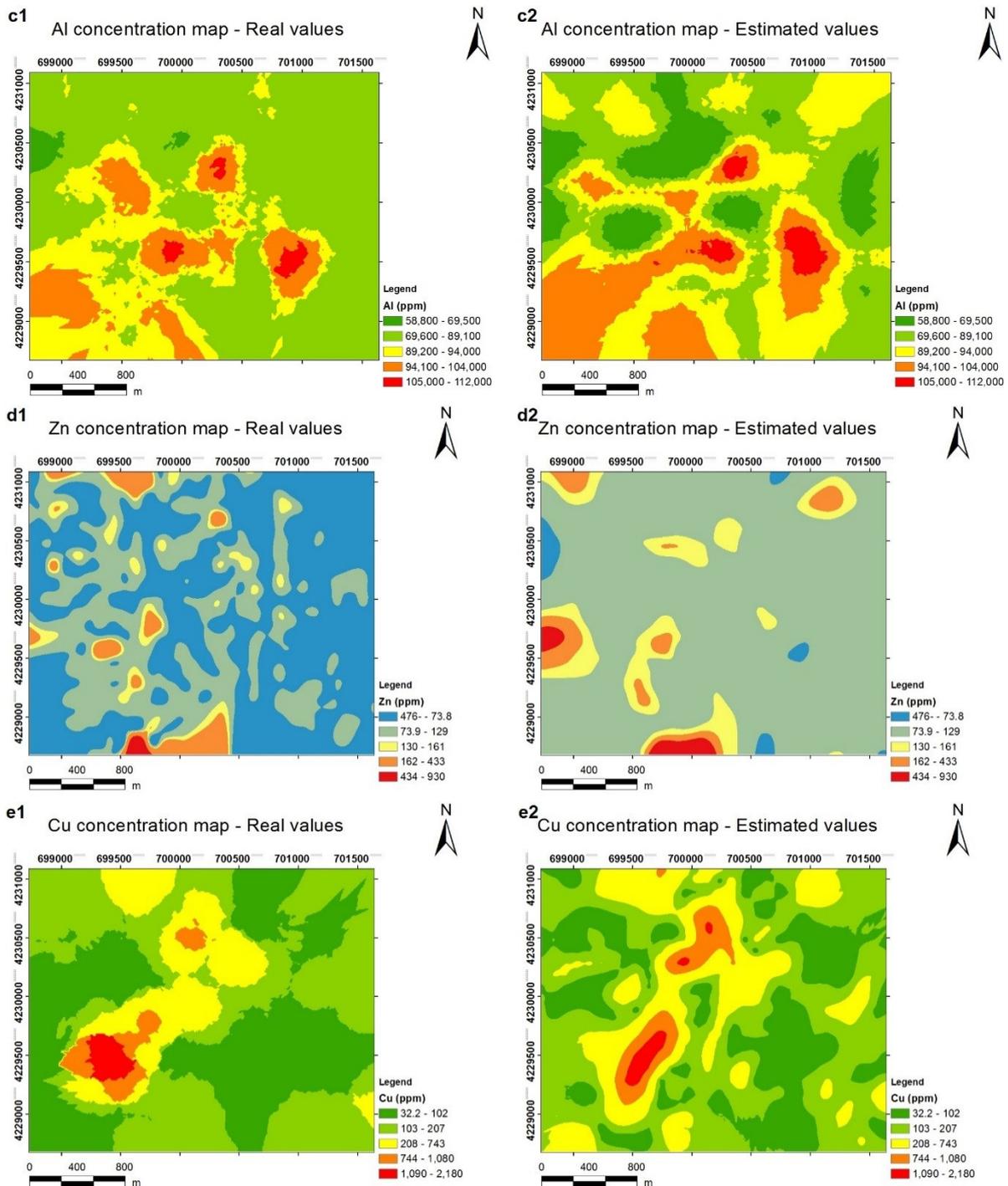


Figure 14. Comparison between actual and estimated surface concentration, (c.1) concentration map for Al with actual data, and (c.2) estimated data, (d.1) concentration map for Zn with actual data, and (d.2) estimated data, (e.1) concentration map for Cu with actual data, and (e.2) estimated data.

6.2. Estimation error comparison

Estimation error is also a parameter in evaluation of estimators' performance, which must be reduced to minimum. Data range varies between -1 and 1.

For estimation error analysis, MSE was used. According to the results from different design schemes, the mean error using all data for estimator design was 0.079. This error was higher than the case of using clustering algorithm. The average estimation error of 22 elements using GK clustering algorithm reduced to 0.048. Therefore, clustering algorithm increases the estimation

accuracy or reduces the estimation error. The results of estimator error of geochemical elements are shown in Figure 15.

Exploring error plots reveals a reduce in AI from 0.022 to 0.012, in As, from 0.105 to 0.025, and from 0.115 to 0.046 in S.

The performance graph and distribution histogram of error for As element show that the estimator performance for clustered data is followed by error reduction and symmetrical histogram (in Figure 16). Generally, estimation error for all the elements decreases by the clustering algorithm.

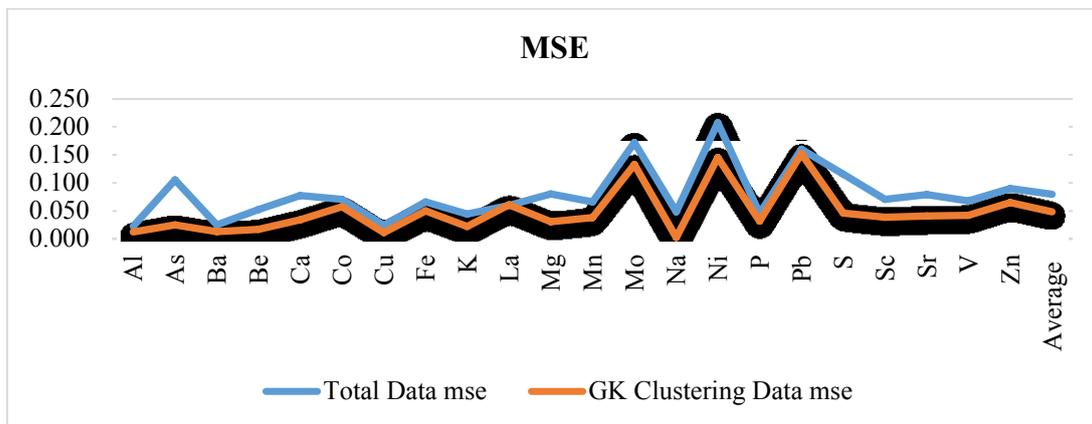


Figure 15. Estimation error comparison for using all data and clustering method.

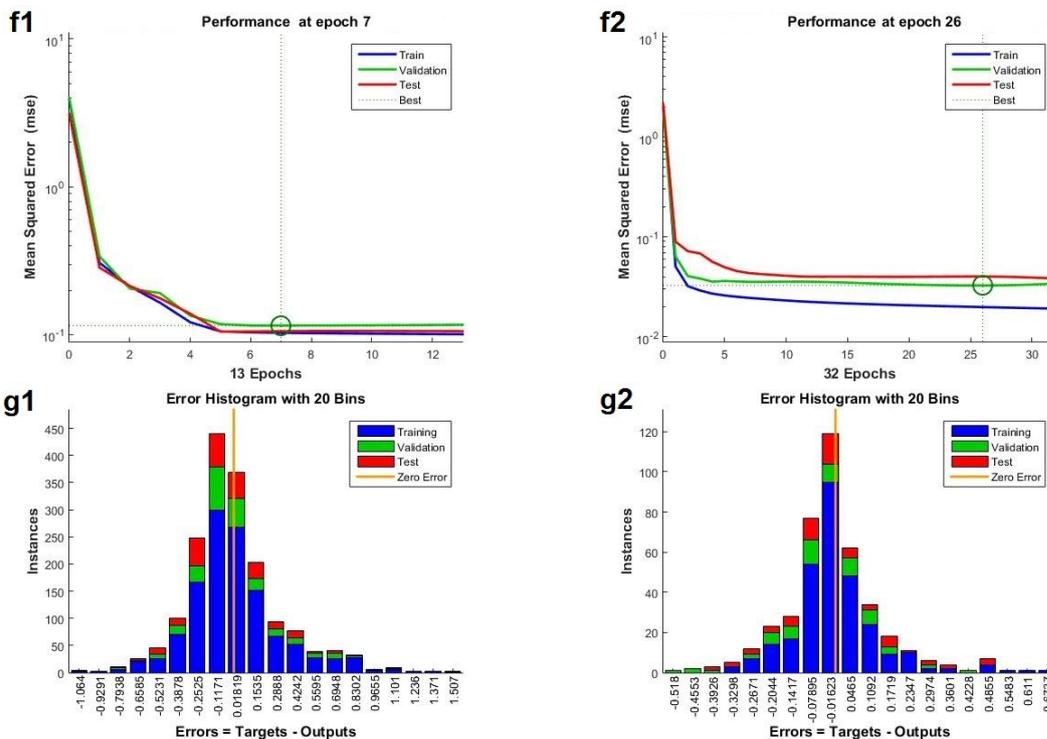


Figure 16. Error variation plot and error distribution histogram for element As, (e.1) error plot when using all data, (e.2) error plot when using clustered data, (f.1) error histogram when using all data, (f.2) error histogram when using clustered data.

7. Conclusions

In order to identify the pattern for element distribution concentrations and to estimate the concentration, multi-layer neural networks can be applied. The results obtained from estimation using neural networks, in the case of using all data, showed that they adopted a proper performance in element concentration estimation. This method did not produce satisfactory results for some elements such as Zn, As, Pb, Mo, and Mn.

By clustering the available data and selecting the best cluster for estimation design, accuracy of estimators was improved. Clustering made different data with the same characteristics in the same clusters, and the correlation of data in clusters reached a maximum value. The performance of neural network in identification governing pattern and estimation accuracy were developed. The average accuracy for the 22 studied elements reached 84% when using clustered data, which showed 9% increase in comparison with estimation using all data. Elemental accuracy increased in the case of using clustered data.

In addition, a method other than estimation accuracy could improve the estimation error. The results obtained exhibited that the estimation error when using GK clustering algorithm decreased from 0.079 to 0.048 in comparison with all the data used. Finally, using the estimated concentrations of the geochemical elements at different depth levels, it is possible to study the changes of geochemical halos. Using the results obtained, it is possible to identify the hidden mineralization, surface erosion, and deep extension of mineralization.

References

[1]. Tahmasebi, P. and Hezarkhani, A. (2012). A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation. *Computers and Geosciences*. 42: 18-27.

[2]. Zuo, R., Carranza, E.J.M. and Wang, J. (2016). Spatial analysis and visualization of exploration geochemical data. *Earth-Science Reviews*. 158: 9-18.

[3]. Zuo, R. and Wang, J. (2016). Fractal/multifractal modeling of geochemical data: a review. *Journal of Geochemical Exploration*. 164: 33-41.

[4]. Cheng, Q. (2012). Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas. *Journal of Geochemical Exploration*. 122: 55-70.

[5]. Wang, J., Zuo, R. and Caers, J. (2017). Discovering geochemical patterns by factor-based cluster analysis. *Journal of Geochemical Exploration*. 181: 106-115.

[6]. Jalloh, A.B., Kyuro, S., Jalloh, Y. and Barrie, A.K. (2016). Integrating artificial neural networks and geostatistics for optimum 3D geological block modeling in mineral reserve estimation: A case study. *International Journal of Mining Science and Technology*. 26 (4): 581-585.

[7]. Journel, A.G. and Huijbregts, C. (1978). *Mining Geostatistics*. Academic Press. London 600 P.

[8]. Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Network*. 2 (5): 359-366.

[9]. Rendu, J.M. (1979). Kriging, logarithmic Kriging, and conditional expectation: comparison of theory with actual results, Proc. 16th APCOM Symposium. Tucson, Arizona, pp. 199-212.

[10]. Koike, K. and Matsuda, S. (2003). Characterizing content distributions of impurities in A limestone mine using a feed forward neural network: *Nat. Resour. Res.* 12 (3): 209-223.

[11]. Samanta, B., Bandopadhyay, S. and Ganguli, R. (2004). Data segmentation and genetic algorithms for sparse data division in Nome placer gold grade estimation using neural network and geostatistics. *Mining Exploration Geology*. 11 (1-4): 69-76.

[12]. Jozanikohan, G., Norouzi, G.B., Sahabi, F., Memarian, H. and Moshiri, B. (2015). The application of multilayer perceptron neural network in volume of clay estimation: Case study of Shurijeh gas reservoir, Northeastern Iran. *Journal of Natural Gas Science and Engineering*. 22: 119-131.

[13]. Rooki, R., Doulati Ardejani, F., Aryafar, A. and Bani Asadi, A. (2011). Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran. *Environmental Earth Sciences*. 64 (5): 1303-1316.

[14]. Tutmez, B. (2005). Reserve estimation using fuzzy set theory. Unpublished Ph.D dissertation. Hacettepe University. Ankara. 168 P.

[15]. Tutmez, B., Tercan, E.A. and Kaymak, U. (2007). Fuzzy modeling for reserve estimation based on spatial variability. *Mathematical Geology*. 39 (1): 87-111.

[16]. Tahmasebi, P. and Hezarkhani, A. (2009). Application of optimized neural network by genetic algorithm. IAMG09Stanford University, California.

[17]. Mahmoudabadi, H., Izadi, M. and Menhaj, M.B. (2009). A hybrid method for grade estimation using genetic algorithm and neural networks. *Computational Geosciences*. 13: 91-101.

- [18]. Yousefi, M. and Nyk€anen, V. (2016). Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping. *J. Geochem. Explor.* 164: 94-106.
- [19]. Zaremotlagh, S. and 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. *J. Geochem. Explor.* 165: 35-48.
- [20]. Webb, A.R. (2002). *Statistical Pattern Recognition*. Second edition JohnWiley & Sons, Ltd., England, pp. 169-202.
- [21]. Duda, R.A., Hart, P.E. and Stork, D.G. (2012). *Pattern Classification*. John Wiley & Sons.
- [22]. Cheng, Q. (2004). Application of Weights of Evidence Method for Assessment of Flowing Wells in the Greater Toronto Area, Canada. *Natural Resources Research*. 13 (2): 77-86.
- [23]. Cheng, Q. (2008). Special Issue of Mathematical Geosciences for the 33rd IGC, Journal of China University of Geosciences. 19 (4): 307-308.
- [24]. Zhijing, W., Qiuming, C., Deyi, X. and Yaosong, D. (2008). Fractal modeling of sphalerite banding in Jinding Pb-Zn deposit, Yunnan, southwestern China. *Journal of China University of Geosciences*. 19 (1): 77-84.
- [25]. Hezarkhani, A. (2003). Exploration of Sonajil copper deposit, Iranian company of copper, Northwestern report exploration.
- [26]. Amato, M., Bitella, G., Rossi, R., Gomez, J.A., Lovelli, S. and Gomes, J.J. (2009). Multi-electrode 3D resistivity imaging of alfalfa root zone. *Europ. J. Agronomy*. 31: 213-222.
- [27]. Mahmoudabadi, H., Izadi, M. and Menhaj, M.B. (2009). A hybrid method for grade estimation using genetic algorithm and neural networks. *Computational Geosciences*. 13 (1): 91-101.
- [28]. Asadi, R., Mustapha, N., Sulaiman, N. and Shiri, N. (2009). New Supervised Multilayer feed forward Neural Network model for Accelerate Classification with high Accuracy. *European Journal of Scientific Research*. 33 (1): 163-178.
- [29]. Altrock, C.V. (1995). *Fuzzy logic & Neurofuzzy applications explained*. Prentice Hall. 384 P.
- [30]. Ghaffari, A., Abdollahi, H., Khoshayand, M.R., Soltani, I., Dadgar, A. and Tehrani, M. (2006). Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International Journal of Pharmaceutics*. 327 (1): 126-138
- [31]. Lesot, M.J. and Kruse, R. (2008). Gustafson-Kessel-like clustering algorithm based on typicality degrees, Book title: *Uncertainty and Intelligent Information Systems*. Publisher: World scientific Publishing Company. pp. 117-130.
- [32]. Gustafson, E. and Kessel, W. (1978). Fuzzy clustering with a fuzzy covariance matrix. In: *IEEE Conference on Decision and Control*. pp. 761-766.
- [33]. Serir, L., Ramasso, E. and Zerhouni, N. (2012). Evidential evolving Gustafson- Kessel algorithm for online data streams partitioning using belief function theory. *International Journal of Approximate Reasoning*. 53 (5): 747-768.
- [34]. Bezdek, J.C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Plenum Press.
- [35]. Xie, X.L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. pp. 841-847.
- [36]. Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P., Silbiger, M.L., Arrington, J.A. and Murtagh, R.F. (1996). Validity-guided (Re) clustering with applications to image segmentation, *IEEE Transactions on Fuzzy Systems*. 4: 112-123.
- [37]. Balasko, B., Abonyi, J. and Feil, B. (2003). *Fuzzy Clustering and Data Analysis Toolbox*.

تخمین عناصر ژئوشیمیایی با استفاده از الگوریتم ترکیبی شبکه عصبی - گوستافسون کسل

محرم جهانگیری*، سید رضا قوامی ریایی و بهزاد تخم‌چی

دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه صنعتی شاهرود، ایران

ارسال ۲۰۱۷/۳/۱۲، پذیرش ۲۰۱۷/۱۱/۵

* نویسنده مسئول مکاتبات: m.jahangiri@Shahroodut.ac.ir

چکیده:

با توجه به اینکه کمبود داده یک مشکل رایج در مطالعات اکتشافی معادن مس پورفیری است، در این پژوهش هدف تعیین الگوهای مخفی در داده‌ها و گسترش اطلاعات به مناطق ناشناخته است. برای انجام این کار، ترکیبی از تکنیک‌های تشخیص الگو استفاده شده است. در این پژوهش، شبکه عصبی چند لایه برای تخمین غلظت عناصر ژئوشیمیایی مورد استفاده قرار گرفت. از مجموع ۱۷۵۵ داده سطحی و عمقی موجود که با روش ICP مورد آنالیز قرار گرفته است، ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای آزمون استفاده شده است. میانگین دقت تخمین‌گرهای طراحی شده برای ۲۲ عنصر ژئوشیمیایی در حالت استفاده از کل داده‌ها برابر با ۷۵ درصد است. بر اساس اعتبار سنجی، تعداد مطلوب خوشه‌ها برای کل داده‌ها شناسایی شد. الگوریتم خوشه‌بندی گوستافسون-کسل برای طراحی تخمین‌گر غلظت عناصر ژئوشیمیایی در خوشه‌های مختلف مورد استفاده قرار گرفت و خوشه‌ها برای طراحی تخمین‌گر انتخاب شدند. نتایج به دست آمده نشان می‌دهد که با استفاده از گوستافسون-کسل دقت متوسط تخمین‌گرها تا ۸۴ درصد افزایش می‌یابد. دقت عناصر روی، آرسنیک، سرب، مولیبدن و منگنز با دقت‌های کم ۰/۵۱، ۰/۶۲، ۰/۶۴، ۰/۶۵ و ۰/۶۸ بر اساس تمام داده‌ها به ترتیب به ۰/۷۶، ۰/۸۶، ۰/۷۶، ۰/۸۰ و ۰/۷۱ بر اساس داده‌های خوشه‌بندی شده افزایش داده شد. میانگین خطای مربع با استفاده از تمام داده‌ها ۰/۰۷۹ بود، در حالی که با استفاده از روش ترکیبی توسعه یافته، به ۰/۰۴۸ کاهش یافت. کاهش خطا در آلومینیوم از ۰/۰۲۲ به ۰/۰۱۲ در آرسنیک از ۰/۱۰۵ به ۰/۰۲۵ و از ۰/۱۱۵ به ۰/۰۴۶ برای گوگرد بود.

کلمات کلیدی: الگوریتم خوشه‌بندی، بهبود دقت تخمین، گوستافسون-کسل، تخمین عناصر ژئوشیمیایی، شبکه عصبی.
