Global Journal of Computer Sciences: Theory and Research

Volume 12, Issue 2, (2022) 78–92 www.gjcs.eu

Geological domaining at Sungun porphyry copper deposit using cluster analysis

Meysam Nikfarjam [a**], Amirkabir University of Technology, Faculty of Mining Engineering, Tehran, Iran

Ardeshir Hezarkhani [b], Amirkabir University of Technology, Faculty of Mining Engineering, Tehran, Iran

Farhad Aziz-Afshari [c], National Iranian Copper Industries Co. (NICICO), Sungun Copper Mine, East-Azarbaijan, Iran

Suggested Citation:

Nikfarjam, M., Hezarkhani, A., & Aziz-Afshari, F. (2022). Geological domaining at Sungun porphyry copper deposit using cluster analysis. *Global Journal of Computer Sciences: Theory and Research*. 12(2), 78–92. https://doi.org/10.18844/gjcs.v12i2.7284

Received from March 11, 2022; revised from June 24, 2022; accepted from September 04, 2022.
Selection and peer review under the responsibility of Assist. Prof. Dr. Ezgi Pelin Yildiz, Kafkas University, Turkey.

Abstract

Geometallurgical sampling has an important role in the geometallurgical programme. There are some critical aspects to be considered in geometallurgical sampling. Cluster analysis (CA) is one of the most popular methods that aids in creating domains. This study aims to use CA to create Geological domaining at the Sungun porphyry copper deposit (SPCD). Geological domains of SPCD were defined using PCA and *K*-means clustering with one hot encoding algorithm. Categorical data were encoded with one hot method and then PCA was used to condense the large new data set to its relevant features. In this research, we have exploited two validity indices to define the optimum number of clusters, namely the silhouette index and the elbow method. Therefore, clustering was performed using four clusters and geological domains were partitioned. The richest domain in this deposit is the second cluster in which the average grades of Cu and Mo are higher than in other clusters.

Keywords: Cluster analysis, geological domains, *K*-means, PCA, Sungun deposit;

* *ADDRESS FOR CORRESPONDENCE: Meysam Nikfarjam, Amirkabir University of Technology, Faculty of Mining Engineering, Tehran, Iran.
E-mail address: Meysam47@aut.ac.ir/Tel.: +98-935-6575603

Introduction

Geometallurgy as a multidisciplinary approach provides new opportunities for optimising mine planning by integrating geology, metallurgy, mineral economics, geotechnical engineering and geoenvironmental parameters [1]–[3]. The definition of geological domains is the first step in geometallurgical modelling that is vital in taking representative geometallurgical samples [4]–[8]. It is important to note that there is no specific relationship between geological and geometallurgical domains, although geological domains are used to collect geometallurgical representative samples and define their domains [4], [9], [10].

Geometallurgical sampling has an important role in the geometallurgical programme. There are some critical aspects to be considered in geometallurgical sampling. Sample representativity and spatial distribution in the domains are the key factors in this procedure; therefore, it is important to define geological domains correctly [11]. Cluster analysis (CA) is one of the most popular methods that aid in creating domains. CA is a multivariate statistical method and unsupervised machine learning approach that groups objects based on similarity within a cluster and dissimilarity between other clusters. The finer cluster is one with greater homogeneity within a group and greater dissimilarity between other groups [12]. Among a large number of clustering approaches, *K*-means with one hot encoding algorithm and PCA were used to handle clustering with mixed data types [13].

Purpose of the study

Geometallurgical sampling has an important role in the geometallurgical programme. There are some critical aspects to be considered in geometallurgical sampling. CA is one of the most popular methods that aid in creating domains. This study aims to use CA to create geological domaining at the Sungun porphyry copper deposit (SPCD). Geological domains of SPCD were defined using PCA and *K*-means clustering with one hot encoding algorithm.

Materials and methods

This study conducted research using the case study method.

Case study: SPCD

SPCD is located in the north-western region of Iran, East Azarbaijan province (Figure 1). The oldest exposed rock of the SPCD is represented by a set of Cretaceous limestone and shale interlayers. The Upper Eocene volcanic breccia and sandstone overlie the Cretaceous sedimentary rocks. The diorite/granodioritic to quartz-monzonitic stock intrudes the Eocene volcanic-sedimentary and Cretaceous carbonate rocks [14]. Porphyry stocks I and II are the two types of porphyry in SPCD. Porphyry stock II, which causes hydrothermal activities and related mineralisations, has a composition ranging from quartz monzonite through granodiorite to granite [15]–[20]. Hydrothermal activities from this stock cause the formation of potassic and propylitic alterations that were followed by phyllic and argillic alterations.

Furthermore, contact metamorphism has transformed carbonate rocks into skarn and related mineralisation occurred at the eastern and northern margins of the stock. There are three mineralisation zones of porphyry stock: leached, supergene and hypogene. Additionally, three main

rock types are controlling the mineralisation: Sungun porphyry stock (SP), skarn mineralisation (SK) and late-injected dykes (DK) [21].

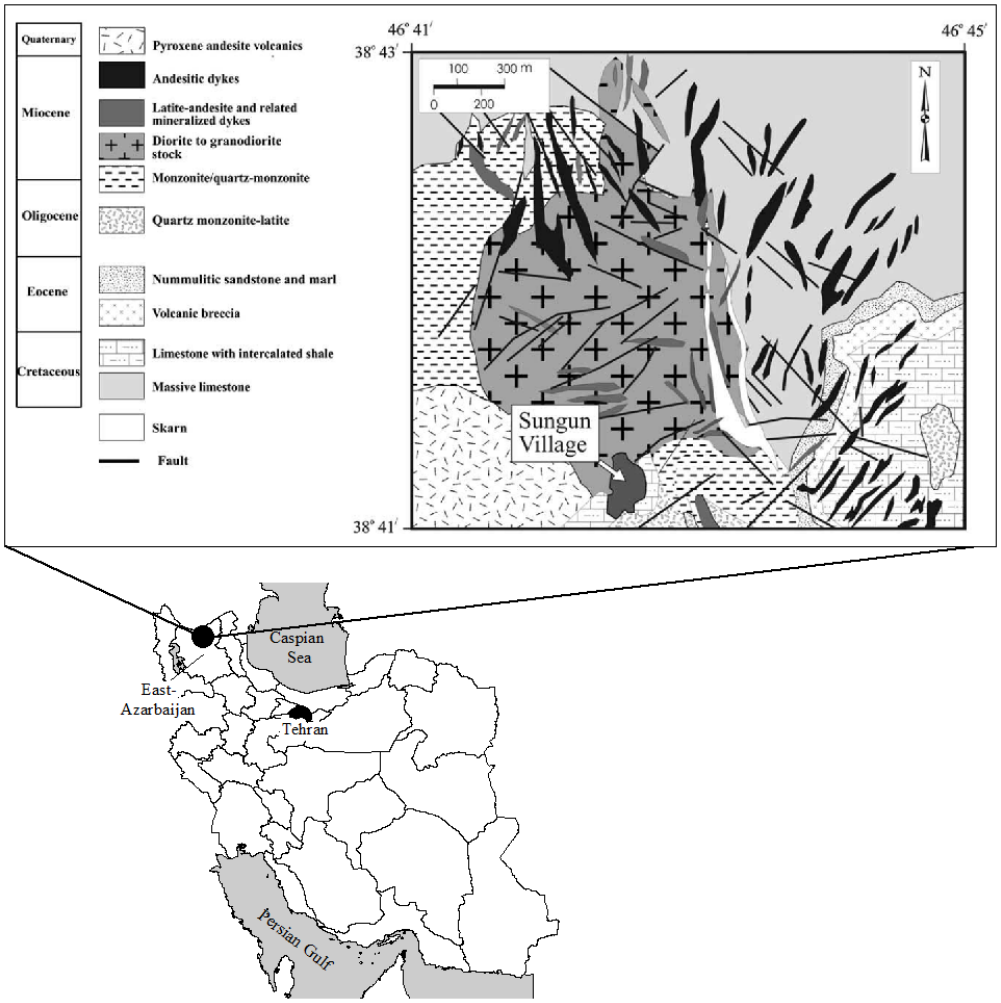


Figure 1. Location map and geological map of SPCD [20]

Analysis

In the machine learning world, CA is referred to as unsupervised learning and forms part of statistical learning. CA is a technique in which observations are grouped into subgroups in a way that they are as similar as possible to each other within a subgroup and as different as possible from the observations belonging to another subgroup, given a similarity or dissimilarity metric [12], [13], [22].

In geological domaining, the data set is defined as a set of samples (observations) with several dimensions of attributes that each attribute comprises continuous or categorical variables. By CA, this data set is grouped into different subsets where samples within a group are similar and groups are well separated [9], [22]. One of the conventional CA methods commonly used in clustering techniques and efficiently used for large data is the *K*-means algorithm. However, this method is not suitable for data that contains categorical variables. This problem happens when the cost function in *K*-means is

calculated using the Euclidian distance that is only suitable for numerical data, but *K*-mode is only suitable for categorical data only, not mixed data types [23], [24]. In a data set containing mixed categorical and numerical variables, the main challenge is finding a way to perform clustering algorithms. Facing these problems, the *K*-means algorithm with one hot encoding categorical variables was used to cluster the mixed data types.

K-means with one hot encoding

Although *K*-means is a simple and quite efficient method, it is not suitable for all types of data. To make it suitable for mixed data types, it needs to be encoded. One hot encoding is a process of creating a new binary column for each categorical value in the data set [13]. In *K*-means clustering, the initial number of *K* that is a user-specified parameter of clusters desired will be defined. Then, according to the minimum distance between a centroid and a data point, the algorithm partitions the data set into *K* clusters and in which each point assigns to the closest centroid. This algorithm iterates the assignment and updates steps until the value of the centroid does not vary. To assign a point to the closest centroid, the minimum distance between a centroid and a data point is calculated. There are several types of proximity measures, including Euclidean distance, Cosine and City block, and several other techniques that can be used [12]. Euclidean distance as a distance measure is used in this study. In *K*-means clustering, let X be d-dimensional data that have been partitioned into *k* clusters C_1, ... C_k. Assume that *X* has been partitioned into *k* clusters C_1, ... C_k. Then, within-cluster variability is defined as follows:

$$W = \sum_{v=1}^{k} \sum_{\{X_i \in C_v\}} \Delta(X_i, c_v)^2 \quad (1)$$

where X_i is belonging to C_υ, c_υ represents the centre of cluster C_υ and Δ is the distance between X_i and C_υ [13].

Cluster validity indices

Some clustering methods such as *K*-means require determining the correct number of clusters [25]–[28]. In this research, we have exploited two validity indices to define the optimum number of clusters, namely the silhouette index (SI) and the elbow method. The SI method, proposed previously [26], [29]–[33], calculates the difference between within-cluster and between-cluster distances as follows:

$$SI_{i,K} = \left\{ \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right\} \quad (2)$$

Where $a(i)$ is the within-cluster average distance and $b(i)$ is the between-cluster distance (lowest average distance). These parameters represent cohesion and separation, respectively. Objects with SI close to 1 represent good classification, while those with SI near to −1 denote misclassification. When SI equals zero, an object is located in the vicinity of the decision boundary between two neighbouring clusters [27], [33]–[35]. To determine the overall performance of a cluster solution, one must average

the SI over the complete data set. In this method, the optimal number of clusters in different *K*-values is that which maximises the average SI value.

In the elbow method, the inertia that expresses the cohesion of the clusters measures the distance between an observation and the centroid of the cluster to which it has been assigned, summed up over all observations. The *k*-means method tries to minimise inertia. The visual clues in the elbow plot enable us to select the correct number of clusters. The elbow signs the point where the line represents its maximum curvature. Before we reach this point, an increase in the number of clusters helps to reduce the sum of squared errors [35]–[38].

Results

Data set

To define the geological domains in SPCD, assay data and borehole loggings formed our data set. This data set, which is a mix of numerical and categorical data, includes measurements of Cu, CuO, Fe and Mo as numerical and lithology, alteration and mineralisation, alteration and porphyry mineralisation zones (supergene, hypogene, leached and oxidation) as categorical data (Table 1). In addition, sample coordinates (*X*, *Y* and *Z*) were used as a spatial distribution of samples in clustering. The data set includes 20,409 samples from 315 drill holes (Figure 2). Cu and Mo were used to determine the high-grade and low-grade zones of deposit. Also, CuO and Fe contents affect the processing procedure. Therefore, these parameters must be considered in the feed such that the CuO/Cu_T content must be less than 10% and Fe content less than 5%.

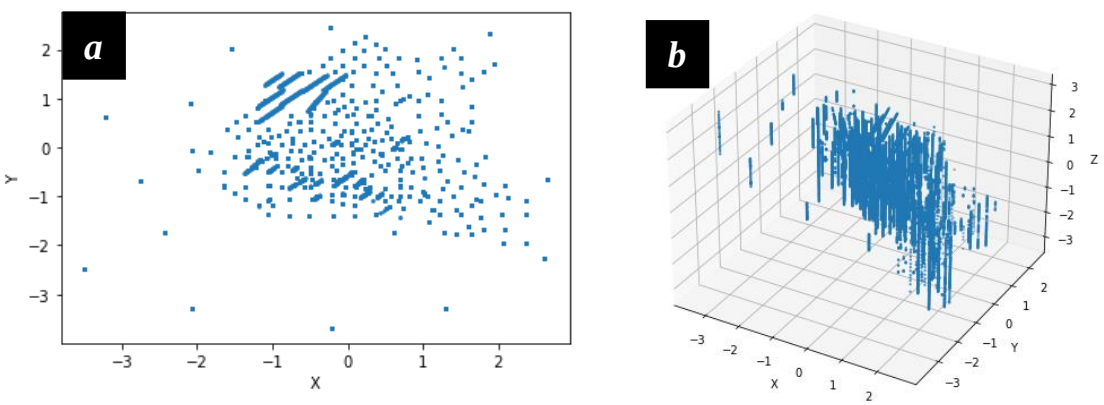


Figure 2. Location map of the data. (a) Plan view of drill hole's collars. (b) 3d view of the samples

Because of the existence of different scales of variables in our data set and their influence on the results of clustering algorithms, the numerical data were standardised using the *z*-score method. *Z*-score is the most common approach to rescale variables with the mean and standard deviation of the original data. In this approach, the mean and variance of all transformed variables are 0 and 1, respectively [25], [29]. Descriptive statistics of assay data before standardisation and the categorical variables used in the clustering procedure are presented in Tables 1 and 2, respectively.

Table 1. Descriptive statistics of numerical (assay) raw data

Statistical parameters	Cu (%)	CuO (%)	Fe (%)	Mo (ppm)
Average	0.48	0.03	3.82	107.5
Median	0.36	0.01	3.36	37
Mod	0.010	0.001	3.700	10
Std	0.47	0.12	2.09	172.1
Min	0.0003	0.0001	0.1602	0.0743
Max	8.94	8.10	29.90	2,737.7
Count	20,409	20,409	20,409	20,409

Table 2. Categorical variables content

Mineralisation zone	Alteration	Lithology
LEA	Fresh-Hem-SK	SP
OXI-SUP	POT	SK
HYP	QSE	HFS
SK	ARG	DK
HFS	PRP	
DK		

SP: Sungun porphyry, SK: Skarn, HFS: Hornfels, DK: Dykes, LEA: Leached zone, OXI-SUP: Oxidation and supergene zone, HYP: Hypogene zone, Hem: Hematite alteration, POT: Potassic alteration, QSE: Quartz–sericite–pyrite (or phyllic) alteration, ARG: Argillic alteration and PRP: Propylitic alteration.

Domaining by PCA and K-means with one hot encoding

After the pre-processing stage, data were prepared for CA. In *K*-means with one hot encoding method, before standardisation data, a new column for each categorical variable in the data set was considered and 1 or 0 was assigned based on the existence of categorical value in the data. All the variables were then converted to numerical data and standardisation operations were performed on them. Also, a PCA was used to reduce the large dimension of a new data set. The variance of each component and the scree plot are shown in Figure 3. The first PC explains 19.02% of the total variation of the original data and the second and third explain 17.43% and 10.08%, respectively. Together, the three PCs explain about 46.5% of the total variation and encompass most of the features. Additionally, eigenvalues can help us retain the number of PCs. The first three PCs have eigenvalues >2 and thus provide a good approximation of the variance of the data set and retain for further analysis. Even though the first seven components contribute 74.46% and have eigenvalues >1, it will be difficult to visualise. Instead of plotting and interpreting all component pairs and examining existing clusters, we used the K-means method. The input data in this method is 7 data components that reflect the essence of the primary data instead of 22 primary data columns.

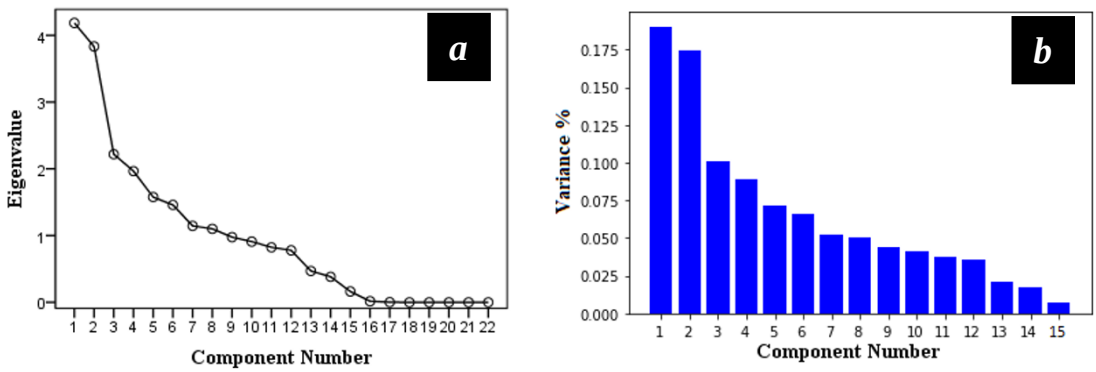
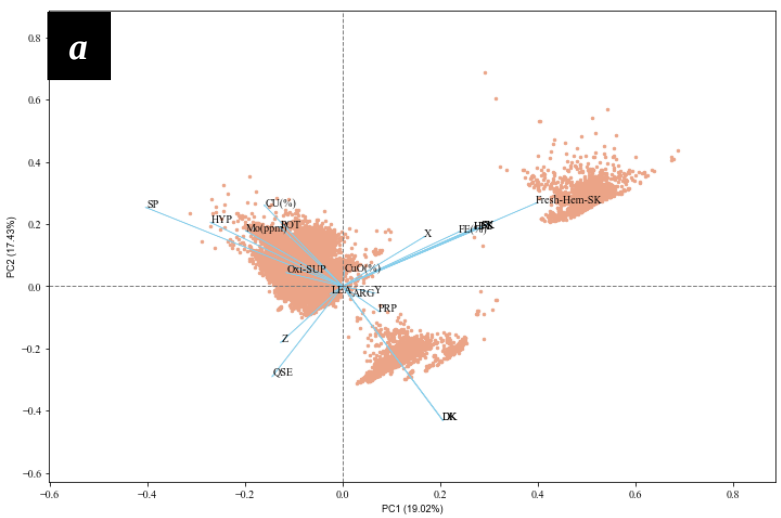


Figure 3. Component number. (a) The scree plot. (b) The variance of each component in PCA

In biplots, the PCs loadings and scores were plotted and the relationships between variables and observations are shown in Figure 4. In the biplot diagram, the small angle between the two vectors indicates the strong association between the two variables. Also, the longer length of components shows a higher amount of variance contributed. From the biplot, we can see that variables Cu, Mo, HYP, SP, PO and Oxi-SUP are highly associated and form a cluster. Similarly, Fe, SK, Fresh-Hem-SK, and X are highly associated and form another cluster (conditions are highly similar but different from other clusters). Similarly, PRP, ARG, Y, DK and Z, and QSE form two other clusters (Figure 4).



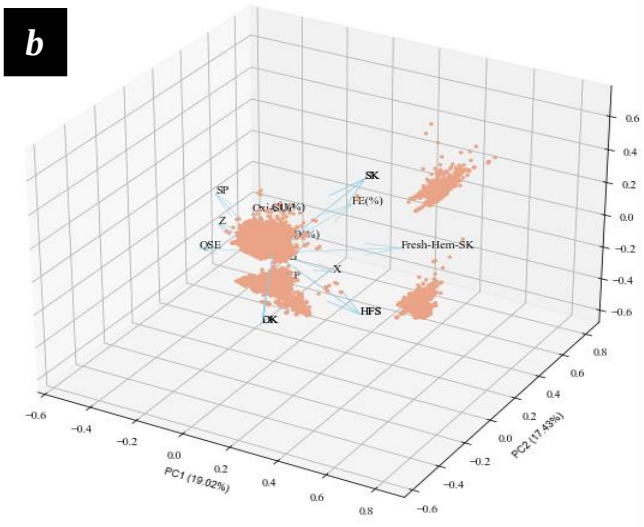


Figure 4. (a) Biplot of the first two components. (b) First three components

In the following, the correlation matrix of loadings was plotted to identify which variables have the largest effect on each component. The first PC is relatively correlated with eight of the original variables. The first PC increases with increasing Fe and SK, HFS and DK scores and decreases with SP and HYP. This component can be viewed as skarn mineralisation and dykes. The second PC increases with Cu, SP and HYP values, decreasing DK and QSE. This component can be viewed as a measure of Cu mineralisation in the hypogene zone. The influences of variables on other components are shown in Figure 5.

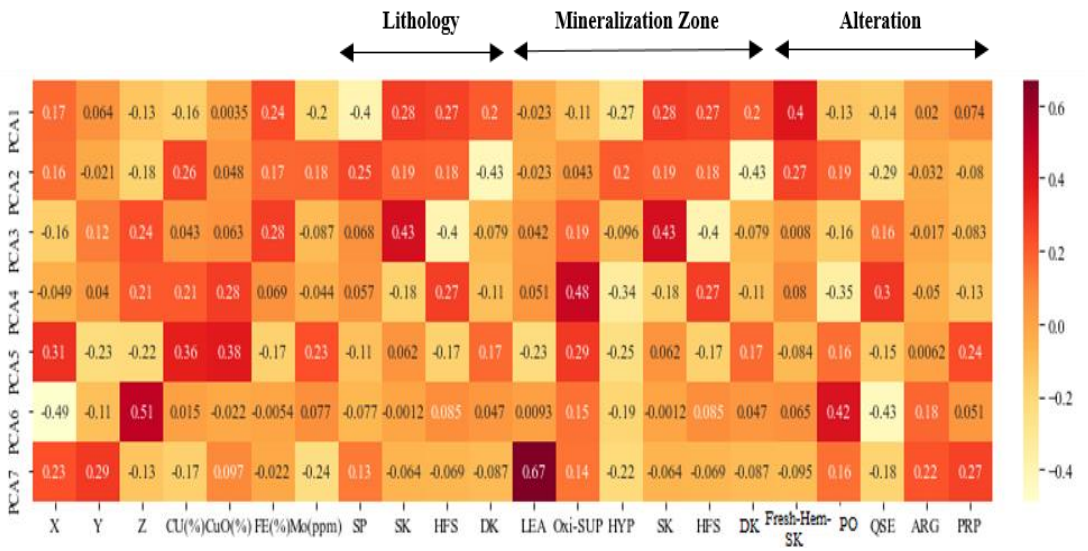


Figure 5. Heat map of PCs and feature correlation analysis

SP: Sungun porphyry, SK: Skarn, HFS: Hornfels, DK: Dykes, LEA: Leached zone, OXI-SUP: Oxidation and supergene zone, HYP: Hypogene zone, Hem: Hematite alteration, POT: Potassic alteration, QSE: Quartz–sericite–pyrite (or phyllic) alteration, ARG: Argillic alteration, PRP: Propylitic alteration.

Instead of interpreting all components and examining existing clusters, we used the *K*-means clustering method. The input data in this method, instead of 22 primary data columns, is 7 data components that reflect the essence of the primary data. The first step in this method is to find the optimum number of clusters. To choose the correct number of clusters, elbow and SI methods were used on the new data set. In the elbow method that enables us to pick the appropriate number of clusters based on the visual clues in a chart, it was found that $k = 4$ is appropriate for clustering. In the SI method, the silhouette score returns the mean ratio of the nearest-cluster distance to the intra-cluster distance. A higher score indicates a better clustering quality. The chart shows a maximum amount of score, at $k = 4$ clusters (Figure 6).

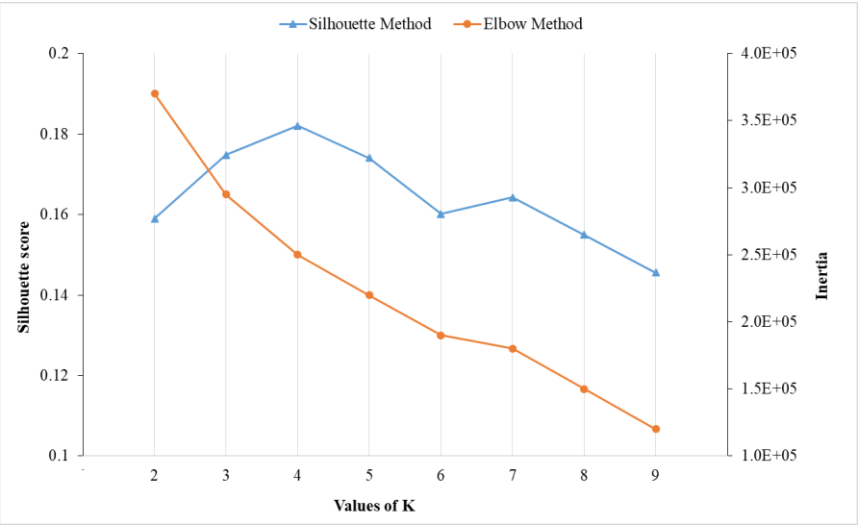


Figure 6. Silhouette score and inertia amounts versus number of clusters

We performed our analysis to determine the number of clusters and confirmed by two different methods that the most appropriate number of clusters is 4. So, we used K-means clustering with 4 clusters and the results are shown in Figure 7. The 4 resulting clusters can be described as follows:

Cluster 1: This cluster contains 680 members and indicates copper mineralisation in the hornfels unit, which is either fresh or contains iron-oxide alteration (Table 3). In this cluster, the average grade of copper is medium and the grade of iron is less than 10% (allowable limit). This cluster is located in the northern and eastern parts of SPCD (Figure 7).

Cluster 2: This cluster has 14,504 members and indicates the rich ore in hypogene and supergene zones with SP rocks (Table 3). The average grades of copper and molybdenum are high and the iron grade is low. This cluster has the highest frequency and is located at the centre of the deposit (Figure 7).

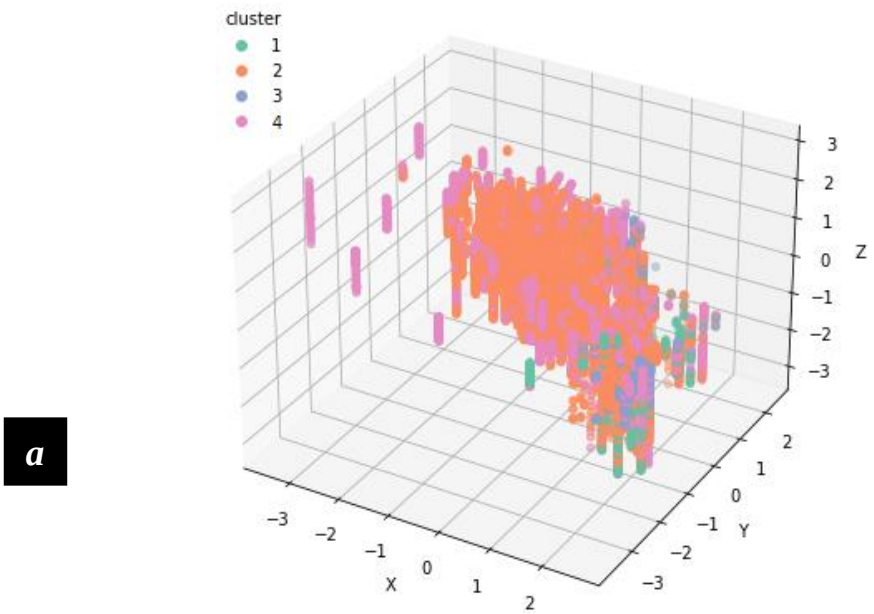
Cluster 3: This cluster includes 606 members and indicates copper mineralisation in the skarn zone. The average grades of both copper and iron are high (Table 3). This cluster, like cluster 1, is situated in the northern and eastern parts of the SPCD (Figure 7).

Cluster 4: This cluster with 4,619 members indicates the barren dykes of SPCD. The average grades of copper, molybdenum and iron are low (Table 3). This cluster is scattered around the deposit in this area (Figure 7).

Table 3. Characteristics of variables in *K*-means clustering with one hot encoding

Variable	Cluster 1 (*n* = 680)		Cluster 2 (*n* = 14,504)		Cluster 3 (*n* = 606)		Cluster 4 (*n* = 4,619)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Cu (%)	0.48	0.53	0.60	0.44	0.50	0.78	0.09	0.19
CuO (%)	0.05	0.44	0.04	0.08	0.04	0.08	0.03	0.12
Fe (%)	6.27	3.11	3.63	1.57	10.05	3.85	3.21	1.29
Mo (ppm)	48.78	115.12	145.35	188.68	15.06	55.02	9.28	31.76
Lithology	HFS		SP		SK		DK	
Zone	HFS		Most HYP Medium SUP Less LEACH		SK		DK	
Alteration	Fresh-Hem-SK		Most QSE Medium POT		Fresh-Hem-SK		Most QSE Medium PRP	

SP: Sungun porphyry, SK: Skarn, HFS: Hornfels, DK: Dykes, LEA: Leached zone, OXI-SUP: Oxidation and supergene zone, HYP: Hypogene zone, Hem: Hematite alteration, POT: Potassic alteration, QSE: Quartz–Sericite–Pyrite (or phyllic) alteration, ARG: Argillic alteration, PRP: Propylitic alteration.



a

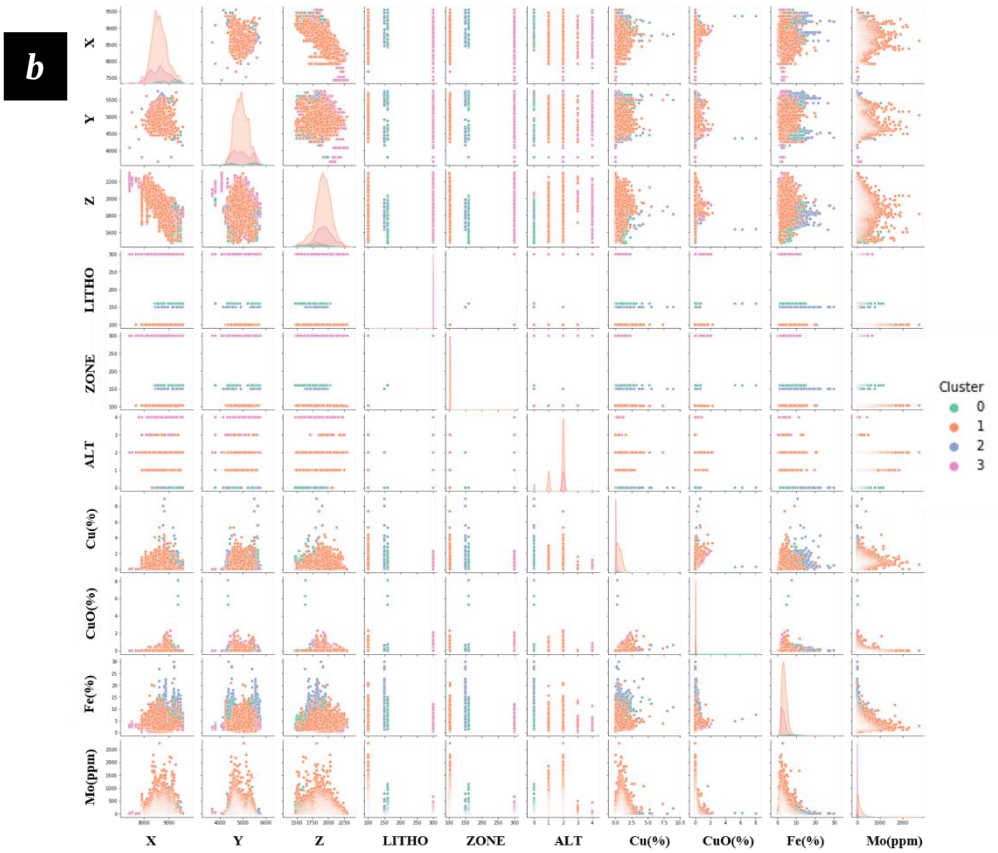


Figure 7. (a) Resulting clusters from the *K*-means with one hot encoding method in 3d space. (b) Resulting clusters from the *K*-means with one hot encoding method in variables scatter matrix. LITHO (100 = SP: Sungun porphyry, 150 = SK: Skarn, 160 = HFS: Hornfels, 300 = DK: Dykes); ZONE (101 = LEA: Leached zone, 102 = OXI-SUP: Oxidation and supergene zone, 103 = HYP: Hypogene zone, 150 = SK: Skarn, 160 = HFS: Hornfels, 300 = DK: Dykes); ALT (0 = Fresh-Hem-SK: Fresh Hematite Alteration or Skarn, 1 = POT: Potassic alteration, 2 = QSE: Quartz–Sericite–Pyrite (or phyllic) alteration, 3 = ARG: Argillic alteration, 4 = PRP: Propylitic alteration).

Conclusion

CA is a multivariate approach that groups data objects into different homogenous regions. In a data set containing mixed categorical and numerical variables, the main challenge is finding a way to perform clustering algorithms. *K*-means with one hot encoding clustering algorithm were used to solve this problem. The PCA approach was used to condense the large new data set to its relevant features. The input data in the *K*-means method, instead of 22 primary data columns, was 7 data components that reflect the essence of the primary data.

In this research, we have exploited two validity indices to define the optimum number of clusters, namely the SI and the elbow method. These methods concluded that the most appropriate number of clusters for our data set is 4. Therefore, clustering was performed using four clusters and geological domains were partitioned. The richest domain in this deposit is the second cluster in which the average grades of Cu and Mo are higher than in other clusters. By comparing the clustering results with core loggings and assay analysis, it can be concluded that this method was able to identify the richest part of the deposit, which has a high grade of copper and molybdenum and a low grade of iron and is mainly located in the hypogene zone of the deposit. In addition, the dykes, skarn and poor mineralisation domains are identified by this method.

Acknowledgements

This research is part of a Ph.D. project supported by the National Iranian Copper Industries Co. (NICICO). The authors would like to thank all the staff of SPCD for their assistance in providing the data and doing the sampling.

References

[1]. S. Dominy, L. O'Connor, and A. Parbhakar-Fox, "Geometallurgy-a route to more resilient mine operations," *Minerals*, vol. 8, no. 12, p. 560, 2018. Available: https://doi.org/10.3390/min8120560.

[2]. M. A. Parian, "Development of a geometallurgical framework for iron ore – a mineralogial approach to particle-based modelling," Doctoral thesis, Luleå University of Technology, Luleå, Sweden, 2017. Available: https://www.semanticscholar.org/paper/Development-of-a-geometallurgical-framework-for-A-Parian/6674bdf71f4ee3c1b4b14d1ba316f4078b2f6149

[3]. V. Lishchuk and M. Pettersson, "The mechanisms of decision-making when applying geometallurgical approach to the mining industry," *Miner Econ,* vol. 34, pp. 71–80, 2021. https://doi.org/10.1007/s13563-020-00220-9

[4]. B. Rajabinasab and O. Asghari, "Geometallurgical domaining by cluster analysis: Iron ore deposit case study," *Nat Resour Res,* vol. 28, no. 665–684, 2018. https://doi.org/10.1007/s11053-018-9411-6

[5]. T. Romary, F. Ors, J. Rivoirard, and J. Deraisme, "Unsupervised classification of multivariate geostatistical data: Two algorithms," *Computers & Geosciences*, vol. 85, pp. 96–103, 2015. https://doi.org/10.1016/j.cageo.2015.05.019

[6]. F. Fouedjio, "A hierarchical clustering method for multivariate geostatistical data." *Spatial Statistics*, vol. 18, pp. 333–351, 2016. https://doi.org/10.1016/j.spasta.2016.07.003

[7]. R. Baumgartner, et al., "Building a geometallurgical model for early-stage project development—A case study from the canahuire epithermal Au–Cu–Ag deposit, Southern Peru," presented at Proceedings of the First AusIMM International Geometallurgy Conference, Brisbane, Australia, pp. 53–60, 2011. Available: https://www.researchgate.net/publication/285403898_Building_a_geometallurgical_model_for_early-stage_project_development_-_A_case_study_from_the_Canahuire_epithermal_Au-Cu-Ag_deposit_Southern_Peru

[8]. A. Bye, "Case studies demonstrating value from geometallurgy initiatives," presented at Proceedings of the First AusIMM International Geometallurgy Conference, Brisbane, Australia, pp. 53–60, 2011. Available: https://espace.library.uq.edu.au/view/UQ:317243

[9]. E. Sepúlveda, P. A. Dowd, and C. Xu, "Fuzzy clustering with spatial correction and its application to geometallurgical domaining," *Math Geosci,* vol. 50, pp. 895–928, 2018. https://doi.org/10.1007/s11004-018-9751-0

[10]. G. J. Harbort, K. Lam, and C. Sola, "The use of geometallurgy to estimate comminution parameters within porphyry copper deposits," presented at Proceedings of the International Geometallurgy Conference, Brisbane, Australia, Australasian Institute of Mining and Metallurgy: Melbourne, Australia, pp. 217–230, 2013. Available: https://www.researchgate.net/publication/285918428_The_use_of_Geometallurgy_to_Estimate_Comminution_Parameters_within_Porphyry_Copper_Deposits

[11]. S. Dominy, L. O'Connor, H. Glass, and S. Purevgerel, "Strategic and tactical geometallurgical application in an underground high-grade narrow-vein gold operation," presented at Proceedings of the 28th International Symposium on Mine Planning and Equipment Selection – MPES, pp. 165–174, 2020. Available: https://link.springer.com/chapter/10.1007/978-3-030-33954-8_20

[12]. P. N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining,* 2nd ed., Pearson, 2018. Available: https://www.pearson.com/us/higher-education/program/Tan-Introduction-to-Data-Mining-2nd-Edition/PGM214749.html

[13]. I. Koch, *Analysis of Multivariate and High-Dimensional Data,* New York: Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139025805

[14]. A. Hezarkhani, A. E. Williams, and C. H. Gammons, "Factors controlling copper solubility and chalcopyrite dissolution in the Sungun porphyry copper deposit, Iran," *Miner Deposita,* vol. 34, pp. 770–783, 1999. https://doi.org/10.1007/s001260050237

[15]. A. A. Calagari, "Stable isotope (S, O, H and C) studies of the phyllic and potassic–phyllic alteration zones of the porphyry copper deposit at Sungun, East Azarbaidjan, Iran," *Journal of Asian Earth Sciences,* vol. 21, pp. 767–780, 2003. https://doi.org/10.1016/S1367-9120(02)00083-4

[16]. A. Hezarkhani, "Physicochemical controls on alteration and copper mineralization in the sungun porphyry copper system, Iran," Doctoral thesis, University of McGill, Quebec, CA, 1997. Available: https://escholarship.mcgill.ca/concern/theses/79407z97c

[17]. A. Hezarkhani, "Specific physico-chemical conditions (360°C) for chalcopyrite dissolution/porphyry system in the Reagan porphyry copper porphyry system, Iran," *Amirkabir Journal of Sciences and Technology*, vol. 13, no. 52, pp. 668–687, 2002. Available: https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=34681

[18]. A. Hezarkhani, "Hydrothermal evolution in the Raigan porphyry copper system based on fluid inclusion studies (Kerman, Iran): the path to an uneconomic deposit," *Amirkabir Journal of Sciences and Technology*, vol. 15, pp. 74–84, 2003. Available: https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=5056

[19]. O. Asghari and A. Hezarkhani, "Applying discriminant analysis to separate the alteration zones within the Sungun porphyry copper deposit," *Journal of Applied Scences,* vol. 24, pp. 4472–4486, 2008. https://doi.org/10.3923/jas.2008.4472.4486

[20]. A. Hezarkhani, "Petrology of the intrusive rocks within the Sungun porphyry copper deposit, Azerbaijan, Iran," *Journal of Asian Earth Sciences,* vol. 27, no. 3, pp. 326–340, 2016. https://doi.org/10.1016/j.jseaes.2005.04.005

[21]. A. Hezarkhani, E. Anthony, and A. E. Williams, "Controls of alteration and mineralization in the Sungun porphyry copper deposit, Iran; evidence from fluid inclusions and stable isotopes," *Journal of Economic Geography,* vol. 93, no. 5, pp. 651–670, 1998. https://doi.org/10.2113/gsecongeo.93.5.651
[22]. S. I. Cevik and J. M. Ortiz, "Machine learning applied in mineral resource sector: An overview," Predictive Geometallurgy and Geostatistics Lab, Queen's University, Annual Report, pp. 106–129, 2020. Available: https://qspace.library.queensu.ca/bitstream/handle/1974/28545/2020-07-Cevik-MLoverview.pdf
[23]. Z. Huang, "Extensions to the K-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery,* vol. 2, no. 3, pp. 283–304, 1998. https://doi.org/10.1023/A:1009769707641
[24]. L. Tran, L. Fan, and C. Shahabi, "Clustering mixed-type data with correlation-preserving embedding," in *Database Systems for Advanced Applications. DASFAA 2021*, Cham: Springer, vol. 12682, 2021. https://doi.org/10.1007/978-3-030-73197-7_2
[25]. A. C. Rencher, *Methods of multivariate analysis*, 2nd ed., New Jersey: Wiley, 2003. Available: https://www.wiley.com/en-us/Methods+of+Multivariate+Analysis%2C+2nd+Edition-p-9780471461722
[26]. N. Sahli and L. Belaid, "The incarnation of global citizenship education in an Algerian secondary education textbook," *GJFLT*, vol. 12, no. 1, Feb., pp. 01–10, 2022.
[27]. Z. Shams, "The relationship between emotional intelligence and self-concept and educational achievement of Tehran students," *IJLT*, vol. 11, no. 1, Jan., pp. 28–33, 2019.
[28]. C. Giaconi, N. D. Bianco, I. D'Angelo, S. Halwany, and S. A. Capellini, "Cultural accessibility of people with Intellectual disabilities: A pilot study in Italy," *JESET*, vol. 7, no. 1, Nov., pp. 16–26, 2021.
[29]. A. B. Mohammed, "Comparative analysis of clustering techniques in the Internet of Things," *GJIT*, vol. 10, no. 2, Oct., pp. 106–112, 2020.
[30]. P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7
[31]. S. D. Anastasiadou, "Comparison of multivariate pattering methods in group/cluster identification regarding the science of educational research: Implicative Statistical Analysis vs. L' Analysee Factorielle des Correspondances," *PROSOC*, vol. 6, no. 1, May., pp. 238–245, 2019.
[32]. P. Bayram, E. E. Celep, S. Kostak, and Z. K. Öztürk, "Two-stage multi-objective renewable energy optimisation," *GJPAAS*, vol. 15, no. 15, Jul., pp. 16–28, 2022.
[33]. S. Camelia, "The Danube Delta - a unique place in Europe," *WJER*, vol. 10, no. 1, May., pp. 09–16, 2020.
[34]. R. S. Jawarneh, "Physical education teachers' attitudes towards the use of the Internet in education," *IJIRE*, vol. 9, no. 2, Sep., pp. 98–111, 2022.
[35]. N. Al-Qahtani and W. A. Al-Salem, "Repercussions of Covid-19 crisis on students' remote learning in Saudi universities," *WJET*, vol. 13, no. 4, Oct., pp. 651–671, 2021.
[36]. A. Celik, "Analysis of 5th, 6th and 7th grade students' attitudes and perceptions regarding helpfulness value," *CJES*, vol. 14, no. 4, Dec., pp. 565–579, 2019.
[37]. B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster analysis,* 5th ed., Chichester: Wiley, 2011. Available: https://www.wiley.com/en-us/Cluster+Analysis%2C+5th+Edition-p-9780470978443

[38]. E. Umargono, J. Suseno, and S. K. V. Gunawan, "K-means clustering optimization using the elbow method and early centroid determination based-on mean and median," presented at Proceedings of the International Conferences on Information System and Technology - CONRIST, pp. 234–240, 2020. https://doi.org/10.2991/assehr.k.201010.019