

Domain Delineation Using Geological Data, Variogram Analysis, and Clustering Algorithms

Farzaneh Khorram^{1,2,3*}, Amin Hossein Morshedy⁴

¹Facultad de Ingeniería, Universidad Andres Bello, Santiago, Chile

²Department of Mining Engineering, Universidad de Chile, Santiago, Chile

³Advanced Mining Technology Center, Universidad de Chile, Santiago, Chile

⁴Mining Engineering Department, Yazd University, Yazd, Iran

Email: *farzaneh.khorram@unab.cl

How to cite this paper: Khorram, F., & Morshedy, A. H. (2025). Domain Delineation Using Geological Data, Variogram Analysis, and Clustering Algorithms. *Journal of Geoscience and Environment Protection*, 13, 31-47.

<https://doi.org/10.4236/gep.2025.131003>

Received: November 15, 2024

Accepted: January 7, 2025

Published: January 10, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Domaining is a crucial process in geostatistics, particularly when significant spatial variations are observed within a site, as these variations can significantly affect the outcomes of spatial modeling. This study investigates the application of hard and fuzzy clustering algorithms for domain delineation, using geological and geochemical data from two exploration campaigns at the eastern Kahang deposit in central Iran. The dataset includes geological layers (lithology, alteration, and mineral zones), geochemical layers (Cu, Mo, Ag, and Au grades), and borehole coordinates. Six clustering algorithms—K-means, hierarchical, affinity propagation, self-organizing map (SOM), fuzzy C-means, and Gustafson-Kessel—were applied to determine the optimal number of clusters, which ranged from 3 to 4. The fuzziness and weighting parameters were found to range from 1.1 to 1.3 and 0.1 to 0.3, respectively, based on the evaluation of various hard and fuzzy cluster validity indices. Directional variograms were computed to assess spatial anisotropy, and the anisotropy ellipsoid for each domain was defined to identify the model with the highest level of anisotropic discrimination among the domains. The SOM algorithm, which incorporated both qualitative and quantitative data, produced the best model, resulting in the identification of three distinct domains. These findings underscore the effectiveness of combining clustering techniques with variogram analysis for accurate domain delineation in geostatistical modeling.

Keywords

Domaining, Hard and Fuzzy Clustering, Spatial Anisotropy, Kahang Deposit

1. Introduction

In earth sciences, domaining plays a crucial role in enhancing the accuracy and

reducing the uncertainty of subsurface modeling. The primary rationale behind domaining is to partition a site into domains where the variables of interest exhibit homogeneous variations and can be represented by stationary spatial random fields. Understanding the nature of the boundaries between neighboring domains is a vital aspect of this process, as it influences whether data from one domain can be effectively used for modeling another. Boundaries can be categorized as hard (distinct and non-interdependent), gradational (allowing domains to merge seamlessly), or fuzzy (gradual merging up to a certain distance before becoming distinct) (Hossein Morshedy et al., 2015; Kubler et al., 2016).

Clustering, as an unsupervised pattern classification method, is extensively used to identify groups within datasets (Kaufman & Rousseeuw, 2009). The main objective of clustering is to maximize the similarity within each cluster while ensuring significant differences between clusters (Jain, 2010). Clustering methods can broadly be categorized into hard and fuzzy clustering. Hard clustering methods assign each data point to a single cluster, whereas fuzzy clustering methods allow data points to belong to multiple clusters with varying degrees of membership (Emmert-Streib et al., 2020). Both hard and fuzzy clustering methods utilize cluster validity indices to determine the optimal number of clusters and evaluate clustering quality (Oyelade et al., 2016; Khorram et al., 2021).

There is a wide variety of clustering algorithms, which can be classified into probability model-based and nonparametric algorithms. Probability model-based algorithms assume that the data follow a mixture probability model, while nonparametric algorithms rely on an objective function based on similarity or dissimilarity measures. Nonparametric algorithms can be further divided into hierarchical and partitional methods, with partitional algorithms being the most commonly used. These methods assume that the dataset can be represented by a finite number of cluster prototypes, each associated with its own objective function. To implement these algorithms, it is necessary to define the dissimilarity (or distance) between data points and cluster prototypes.

In general, clustering does not impose spatial contiguity or compactness in the determination of clusters. To explicitly incorporate the spatial component, Oliver and Webster (1989) integrated a spatial variogram model into the dissimilarity measure, using an isotropic exponential structure with parameters such as nugget effect, sill, and range. Bourgault et al. (1992) extended this approach by employing a multivariate (co)variogram to account for both spatial and attribute correlations in clustering. Fouedjio (2016) introduced the spatial component into the dissimilarity measure within the agglomerative hierarchical clustering algorithm, where the dissimilarity/similarity between two observations is not simply a Euclidean measure but a function of their spatial correlation. However, the effect of negative correlations in cross-variograms on this dissimilarity measure, particularly when spatial correlation is low, remains unclear. Romary et al. (2015) further incorporated the spatial component into the distance metric of a hierarchical clustering method. Their distance function accounts for spatial connectivity introduced by a moving neighborhood, allowing the user to define

attribute weights that are integrated into the distance measure. Additionally, the coordinates are treated as attributes in this framework.

In this study, the authors analyze various spatial geoscientific data from a copper deposit in Iran, including lithology, alteration, geochemical zoning, and values of key geochemical ore elements. We apply different clustering methods to determine the optimal domain structure for the region. Hard and fuzzy cluster validity indices are calculated to identify the ideal number of domains. To model the spatial structure of the data, directional variograms are computed for each domain. Finally, anisotropic search ellipsoids for all domains are compared to select the clustering method and dataset that produce the highest average anisotropic dissimilarity between domains.

2. Study Area: East Kahang Copper Deposit

The Kahang Cu-Mo porphyry deposit is located 73 km northeast of Isfahan and 10 km east of the town of Zefreh in central Iran. This porphyry system is part of the Urumieh-Dokhtar magmatic arc (UDMA), which trends northwest-southeast and hosts several of Iran's major copper-bearing deposits, including Sarcheshmeh, Sungun, Meiduk, Darehzar, Sarkuh, and Daralu (Haschke et al., 2010; Sun et al., 2015). The Kahang area is divided into three main sectors: eastern, central, and western. Detailed exploration drilling was conducted in the eastern anomaly to further evaluate its potential.

2.1. Regional Geology

In the eastern part of the Kahang deposit, the primary rock types include andesite, andesite porphyry, dacite porphyry, quartz monzonite, diorite, andesitic breccia, and quartz veins filled with iron oxides. The mineralization comprises chalcocite, covellite, chalcopyrite, pyrite, malachite, chalcantite, magnetite, hematite, jarosite, and goethite. The oldest igneous units are andesitic lavas and Eocene volcanic breccias, which exhibit propylitic alterations and are considered to have high potential for Cu-Mo mineralization. The alteration types in the region include silicification, phyllic, argillic, propylitic, and potassic. The dacitic unit, characterized by a porphyritic texture, is associated with quartz-sericite, argillic, and quartz-goethite alterations. A semi-deep section of the andesite porphyry is intruded by quartz monzonite and diorite stocks, which display phyllic and propylitic alterations. Geochemical zones within the deposit are classified into oxide, leach, supergene, and hypogene, extending from the surface to greater depths. This porphyry system is also influenced by two fault systems with NE-SW and NW-SE orientations (NICICO, 2011; Afzal et al., 2012; Afshooni et al., 2013). **Figure 1** shows the geological map along with the location of the eastern part of the Kahang deposit within the structural map of Iran.

2.2. Data sets

In the first phase of exploration, 2-meter core samples from 10 boreholes were

logged to obtain information on lithology, alteration, and mineral zones (Table 1). These samples were then assayed for Cu, Mo, and Ag grades. The boreholes varied in depth from 100 m to a maximum of 450 m, with dips ranging from approximately 60 degrees to vertical. In the subsequent detailed exploration phase, the number of boreholes increased to 36, including both vertical and directional holes, with depths ranging from 100 m to 700 m. The core samples were assayed

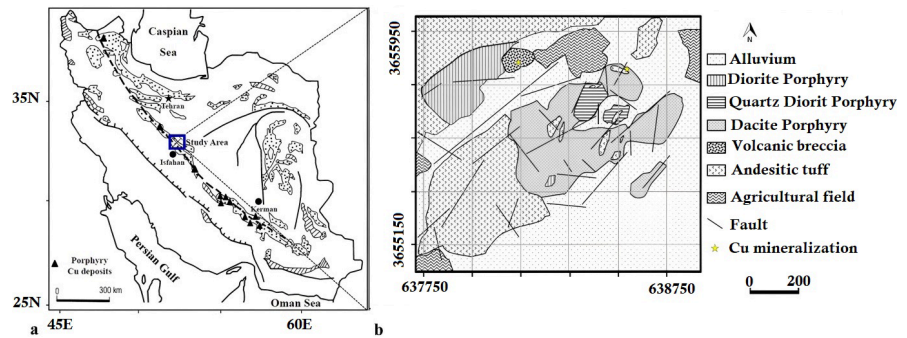


Figure 1. Geological map of the eastern part of the Kahang deposit (b), with its location (blue square) in the structural map of Iran (a). Adapted from Alavi.

Table 1. Summary of logged classes for lithology, alteration, and mineral zone in initial exploration campaign.

Lithology	Alteration	Mineral zone
Alluvium	Unaltered	Oxide
Tuff	Chloritic	Leach
Dacite	Quartz-sericitic	Transitional
Quartz diorite	Potassic	Supergene
Diorite	Phyllic	Hypogene
Andesite	Propylitic	
	Argillic	
	Silicification	

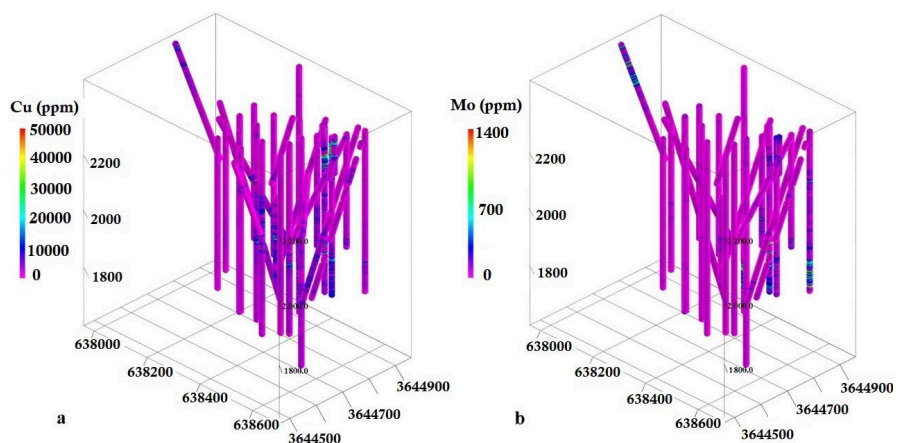


Figure 2. Three-dimensional distribution of ore grades in the 36 boreholes: (a) Cu; (b) Mo.

for Cu, Mo, Ag, and Au. For illustration, the distributions of copper and molybdenum grades are shown in **Figure 2**. The selected geochemical dataset includes less than 5% of values below the detection limit. Preprocessing involved outlier detection and removal, followed by data standardization to prepare it for clustering analysis.

3. Methodology

Clustering Methods

Clustering algorithms are essential tools for grouping data points based on similarity, and they play a crucial role in geological applications, particularly in the domain of rock type classification and geological modeling. In the context of rock types, clustering algorithms can help group rocks with similar characteristics, such as mineral composition, texture, and physical properties, which are essential for tasks like resource estimation, ore body modeling, and the optimization of mining operations (Zhu & Li, 2017).

K-means is one of the most well-known and widely used clustering algorithms. It partitions the data into a predefined number of clusters, K , by iteratively assigning each data point to the nearest centroid and recalculating the centroids based on the points assigned to each cluster. This process continues until the centroids no longer change. While K-means is fast and efficient, it assumes that clusters are spherical and equally sized, which can be a limitation for datasets with more complex structures. Moreover, the number of clusters K must be specified before running the algorithm, which can be a challenge when the optimal number of clusters is unknown (Xu & Wunsch., 2005).

In contrast, hierarchical clustering creates a tree-like structure called a dendrogram, where data points are merged or split at different levels based on their similarity. This method doesn't require the number of clusters to be predefined and allows for a more flexible analysis of the data. There are two main types: agglomerative (bottom-up) and divisive (top-down). In agglomerative clustering, each data point starts as its own cluster, and the algorithm merges the closest clusters iteratively. In divisive clustering, all points start in a single cluster, and the algorithm divides it recursively. This flexibility makes hierarchical clustering ideal for understanding the nested structure of data, but it can be computationally expensive for large datasets (Gan et al., 2020).

Affinity propagation takes a different approach by identifying exemplars, or representative points, that best define each cluster, without the need to predefine the number of clusters. The algorithm uses a message-passing mechanism, where data points send messages to one another to indicate how well they serve as exemplars. These messages are updated iteratively until the algorithm converges on a set of exemplars, which define the clusters. Affinity propagation is advantageous in cases where the number of clusters is uncertain and where clusters may have varying sizes (Frey & Dueck., 2007).

Self-organizing maps (SOM), an unsupervised learning method, is particularly

useful for visualizing high-dimensional data by mapping it onto a 2D grid. SOMs are inspired by neural networks and work by adjusting the positions of neurons in response to data patterns, creating a map where similar data points are grouped together. This method is particularly valuable for exploring and visualizing complex datasets and understanding the underlying structure. SOM is often used in cases where data visualization is as important as clustering (Miljković, 2017).

Fuzzy C-means offers a flexible alternative to K-means by allowing data points to belong to multiple clusters with varying degrees of membership. Instead of making a hard assignment of each point to a single cluster, fuzzy C-means computes the degree to which each data point belongs to each cluster, with membership values between 0 and 1. This approach is helpful when the boundaries between clusters are ambiguous, and data points may share characteristics with more than one cluster. Like K-means, fuzzy C-means iterates between assigning membership values and updating the cluster centers (Ikotun et al., 2023).

Finally, Gustafson-Kessel is an extension of fuzzy C-means that overcomes one of the limitations of fuzzy C-means: the assumption that clusters are spherical. Gustafson-Kessel adapts the covariance matrix for each cluster, allowing for the detection of elliptical clusters, which makes it more versatile when dealing with data that doesn't fit the spherical cluster model. This adaptation makes Gustafson-Kessel particularly effective for complex datasets with clusters of varying shapes and sizes (Ruspini et al, 2019; Gustafson & Kessel 1978).

The difference between hard and fuzzy clustering is associated with the assignment of data samples to clusters (Everitt et al., 2011; Xu & Tian, 2015). In hard clustering, a sharp boundary describes two domains that are in contact with each other but are otherwise unrelated, resulting in a perfect separation of the domains. In the fuzzy boundary, the domains can overlap up to a finite distance, while a gradational boundary can be seen as an extreme case where the overlap distance becomes infinitely large.

4. Discussion and Results

4.1. Comprehensive Method

In this study, four hard clustering algorithms and two fuzzy clustering algorithms are applied to define domains and inter-domain relationships in the eastern Kahang deposit (Table 2). These algorithms identify clusters based on optimality measures, such as minimizing intra-cluster distance or maximizing inter-cluster distance (Wang et al., 2022; Oyelade et al., 2016). To assess the clustering performance, cluster validity indices (Table 3) are employed to determine the optimal number of clusters and evaluate the quality of the results. For each clustering algorithm, the validity indices are calculated based on four distinct combinations of input data (Table 4), representing the four cases under investigation. These combinations include geological layers (lithology, alteration, and mineral zone), geochemical assay data (Cu, Mo, Ag, and Au grades) from borehole core samples collected during both explora-

tion phases, and the 3D coordinates of these samples.

Table 2. Clustering algorithms under consideration.

Clustering algorithm	Acronym	Nature	Reference
K-means	KM	Hard	MacQueen (1967)
Hierarchical complete linkage	HC	Hard	Defays (1977)
Self-organizing map	SOM	Hard	Kohonen (1998)
Affinity propagation	AP	Hard	Frey and Dueck (2007)
Fuzzy c-means	FCM	Fuzzy	Dunn (1974)
Gustafson-Kessel algorithm	GK	Fuzzy	Gustafson and Kessel (1978)

Table 3. Clustering algorithms under consideration.

Validity index	Use	Optimum	References
Dunn	Hard clustering	Max	Dunn (1974)
Davies-Bouldin	Hard clustering	Min	Davies and Bouldin (1979)
Silhouette	Hard clustering	Max	Rousseeuw (1987)
C	Hard clustering	Min	Hubert and Schultz (1976)
Calinski-Harabasz	Hard clustering	Max	Calinski and Harabasz
Hartigan	Hard clustering	Min	Hartigan (1975)
Krzanowski-Lai	Hard clustering	Max	Krzanowski and Lai (1985)
Partition coefficient	Fuzzy clustering	Max	Bezdek (1974)
Classification entropy	Fuzzy clustering	Min	Bezdek (1975)
Partition	Fuzzy clustering	Min	Xie and Beni (1991); Bensaid et al. (1996)
Separation	Fuzzy clustering	Min	Xie and Beni (1991); Bensaid et al. (1996)
Xie and Beni's	Fuzzy clustering	Min	Xie and Beni (1991); Bensaid et al. (1996)
Alternative Dunn	Fuzzy clustering	Min	Xie and Beni (1991)

Table 4. Four combinations of input geological and geochemical layers on the clustering algorithm. The indices 10 and 36 refer to the assays of the initial and advanced exploration campaigns, respectively.

Case	Coordinates	Lithology	Alteration	Mineral zone	Cu ₁₀	Mo ₁₀	Ag ₁₀	Cu ₃₆	Mo ₃₆	Ag ₃₆	Au ₃₆
1	YES	YES	YES	YES							
2	YES				YES	YES	YES				
3	YES	YES	YES	YES	YES	YES	YES				
4	YES							YES	YES	YES	YES

The implementation involves preparing the input data, applying the clustering

algorithms, and calculating the cluster validity indices for partitions with varying numbers of clusters within the integer range [2, 10]. This range corresponds to the minimum and maximum number of domains considered feasible for the eastern Kahang deposit (Figure 3). The resulting indices, computed for different combinations of input layers, are presented in Figure 4.

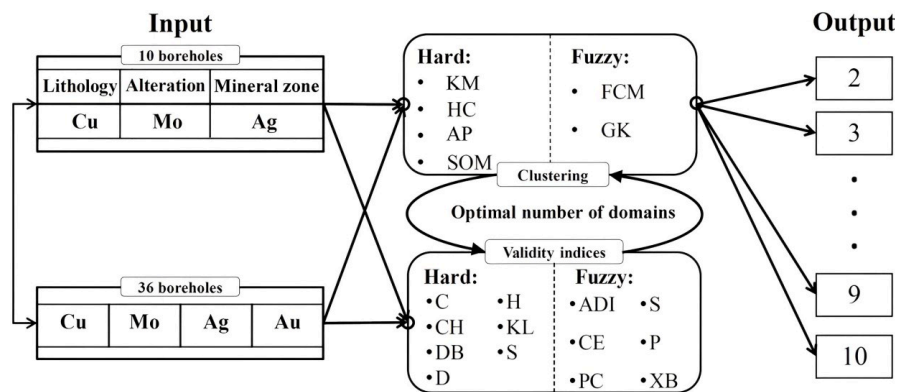
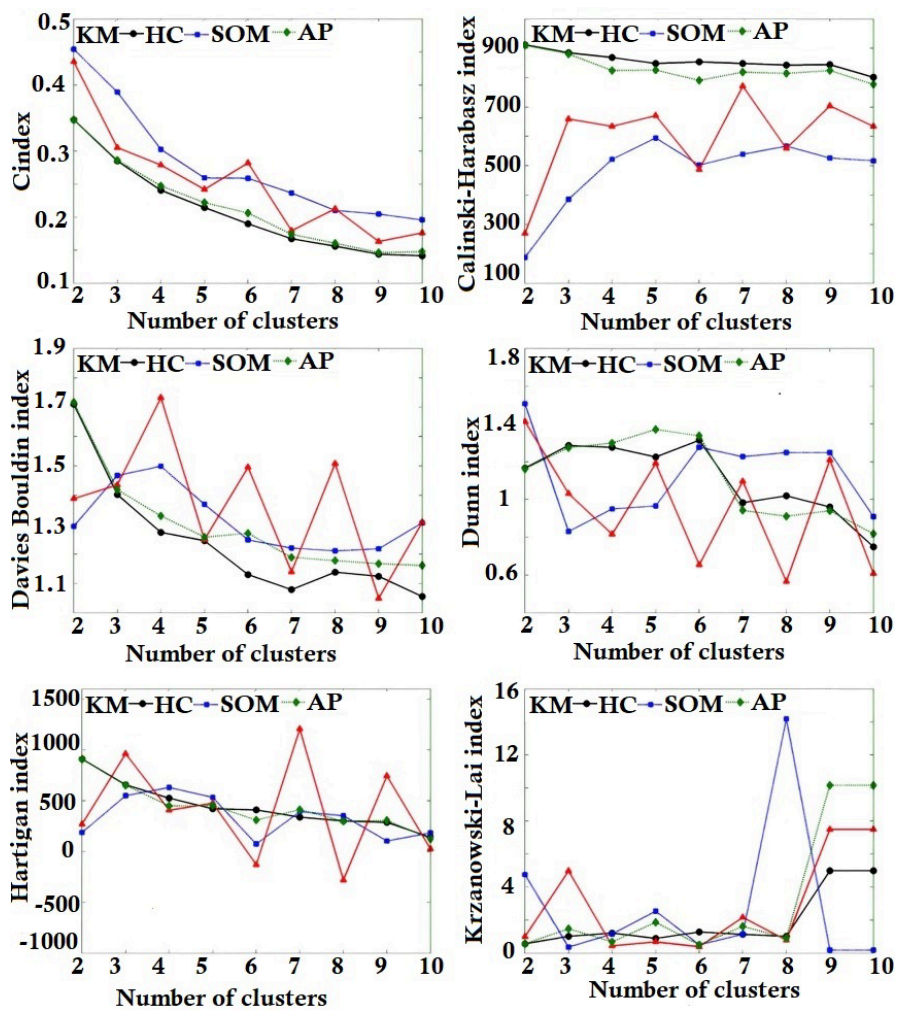


Figure 3. Steps for domaining from input layers to output of clustering consisting of a partition of the data into 2 to 10 domains.



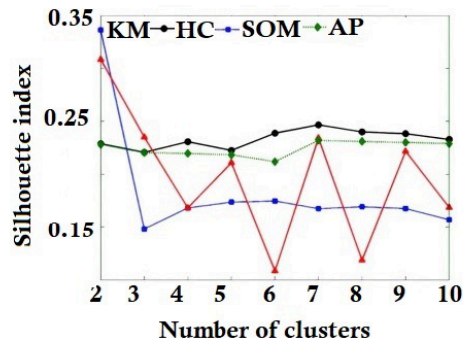


Figure 4. Validity indices computed for hard clustering algorithms: (a) C; (b) CH; (c) DB; (d) D; (e) H; (f) KL; (g) S.

In addition to the number of clusters, fuzzy partitioning algorithms are sensitive to other parameters. For example, in FCM clustering, the validity criteria can be modeled as a function of both the number of clusters and a weighting exponent (fuzziness parameter) within the range of 1 to 2, as shown in **Figure 5**. Similarly, in the modified Gustafson-Kessel algorithm, a tuning parameter (ranging from 0 to 1) must also be considered, as depicted in **Figure 6**.

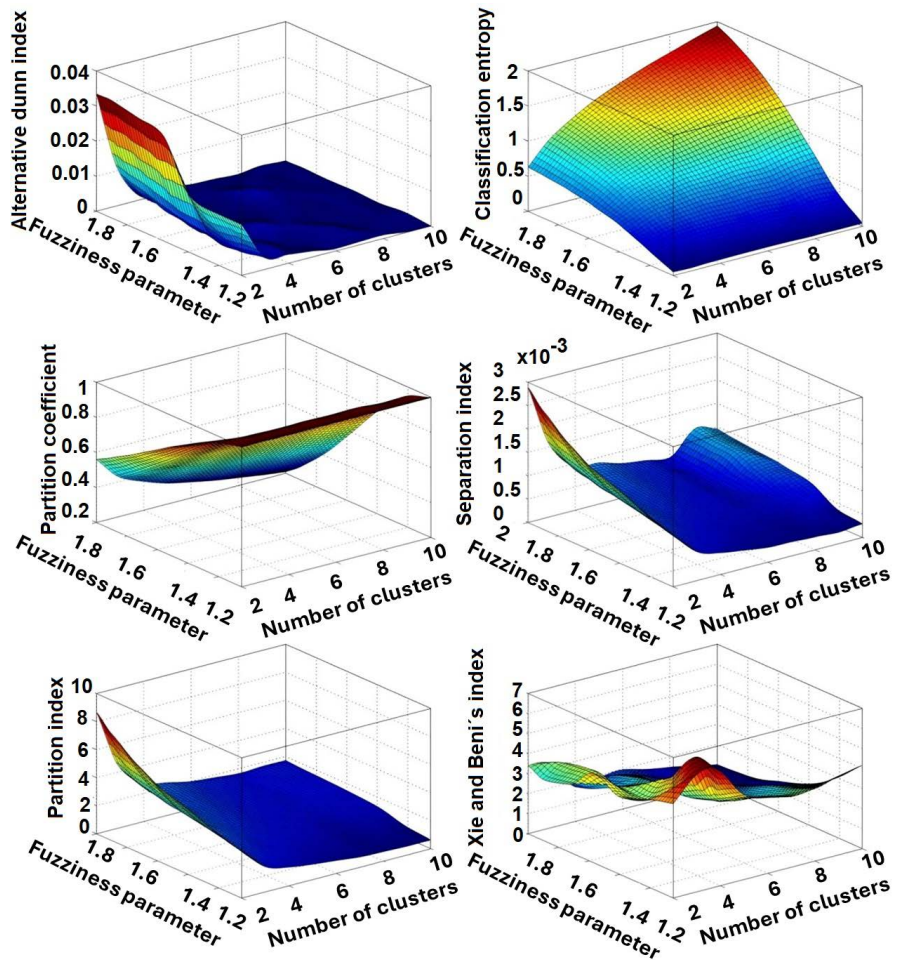


Figure 5. Validity indices computed for FCM clustering: a) ADI; b) CE; c) PC; d) S; e) P; f) XB.

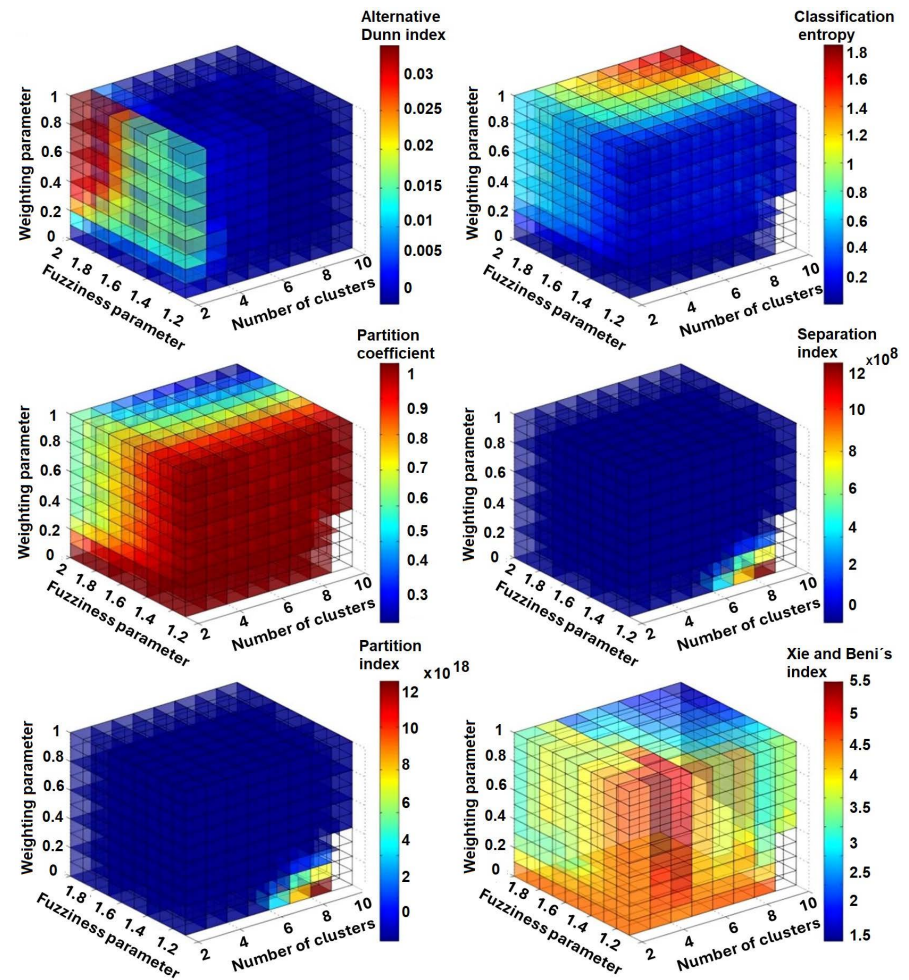


Figure 6. The cluster validity indices of GK clustering: a) ADI; b) CE; c) PC; d) S; e) P; f) XB.

Due to the direct or inverse relationship between each validity index and the number of clusters, a sharp decrease or increase in an index serves as a criterion for identifying optimal clustering parameters, such as the ideal number of clusters. Additionally, partitions with fewer clusters are preferred when the differences between validity indices are minimal (Hadiloo et al., 2018). Considering both the validity indices and expert geological judgment, the optimal number of domains ranges from 3 to 4, depending on the hard clustering algorithm and the combination of input data. In contrast, for the fuzzy clustering algorithms, the fuzziness and weighting parameters are selected within the ranges of 1.1 - 1.3 and 0.1 - 0.3, respectively, with the optimal number of domains consistently defined as 3.

4.2. Post Processing of Clustering Outputs

The objective of domaining is to partition the region into domains that exhibit the maximum statistical and spatial differences. A key tool for analyzing the spatial structure of a regionalized variable is the variogram, which quantifies the average squared difference of paired data points as a function of the distance between them.

For each clustering algorithm and resulting domain, variograms of the feature data are computed in various directions to capture anisotropy, which can be represented as an ellipsoid characterized by dimensional (range) and directional (orientation) parameters. The goal is to identify a domaining scheme with a high degree of dissimilarity between the anisotropy ellipsoids of each domain, as this indicates better spatial differentiation and discrimination of the data's spatial properties.

The outputs of the SOM and GK algorithms under cases 3 and 4, using the selected input data, are chosen to assess the domaining performance. For each data feature and domain, experimental variograms were computed for 24 directions and fitted with theoretical models. **Figure 7** illustrates the experimental variograms and their corresponding fitted models, highlighting the primary anisotropy direction (with the largest correlation range) for each domain obtained from the SOM and GK clustering algorithms. The correlation ranges of the directional variograms were used to calculate the anisotropy ellipsoid parameters, such as anisotropy factors (ratios of the major axis range to the semi-major axis range and the major axis range to the minor axis range) and rotation angles (azimuth, dip, and plunge). The anisotropy ellipsoid effectively summarizes the spatial data structure within each domain, as shown in **Figure 8**.

The Jensen-Shannon divergence (JS) is used to measure the dissimilarity between probability distributions, based on Jensen's inequality and Shannon entropy (Somani et al., 2017; Jiang et al., 2021). For each pair of domains, the Jensen-Shannon divergence is computed by considering the covariance and anisotropy ellipsoids (**Table 5**). The results show that, based on the average Jensen-Shannon divergence, the dataset from Case 3, which includes both qualitative (logging) and quantitative (assay) data combined with the SOM clustering algorithm, is more effective at distinguishing between domains.

In practice, once the borehole core samples have been clustered, translating these clusters into a comprehensive 3D model may involve various interpolation techniques, such as manual drawing, implicit modeling, kriging, or geostatistical simulation. These methods facilitate a nuanced and accurate representation of the clustered data in three-dimensional space, enhancing the depth of understanding in practical applications.

Table 5. The Jensen-Shannon divergences for each pair of domains obtained with SOM and GK as clustering algorithms

Data Set	Algorithm Domain	SOM			GK		
		1	2	3	1	2	3
Case 3	1		0.5598	0.6783		0.4747	0.0140
	2			0.2107			0.4241
	3		Average: 0.4829			Average: 0.3043	
Case 4	1		0.3231	0.1633		0.0056	0.0257
	2			0.1173			0.0162
	3		Average: 0.2012			Average: 0.0158	

Incorporating spatial coordinates into the clustering algorithms aims to generate clusters that exhibit greater spatial cohesion. Several alternative approaches to spatial clustering have been proposed, diverging from the traditional method of treating coordinates as feature variables. Instead, these approaches integrate coordinates as constraints or parameters within the calculation of similarity between data points (Romary et al., 2015; Fouedjio, 2016; Faraj & Ortiz, 2021). It is important to emphasize that the focus of this study lies in the analysis of the obtained clusters through variogram analysis. Our approach parallels the work of Faraj and Ortiz (2021), where clusters are delineated based on localized changes in the spatial correlation of the data.

Finally, incorporating geoscientific data (such as geophysical or geological surveys) offers an opportunity to complement the logging and assay information collected from the boreholes. There are two potential ways to leverage this supplementary data: (a) Integrating it directly into the clustering methodologies

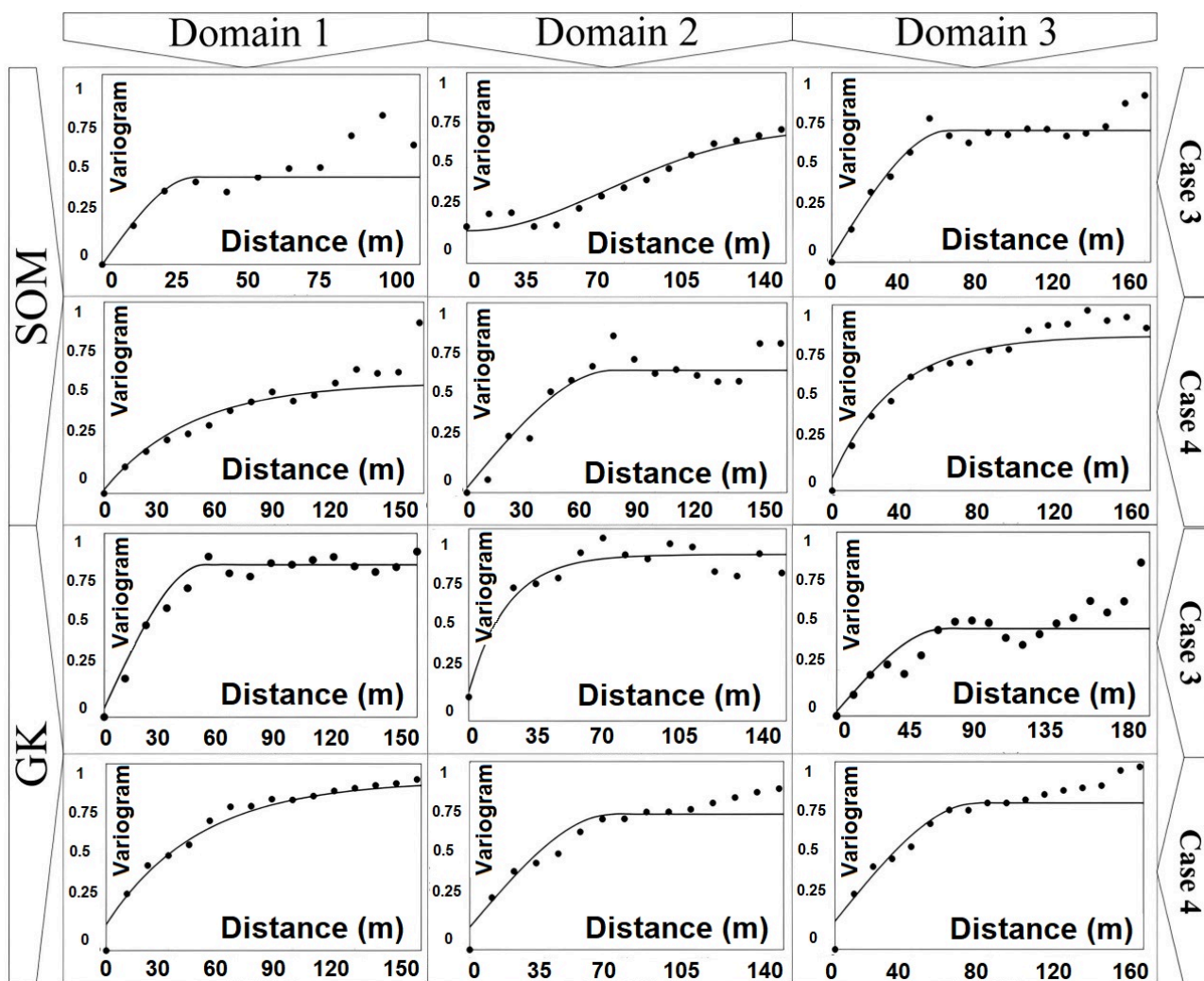


Figure 7. Experimental variograms (dots) with fitted models (solid lines) along the main anisotropy direction for different domains. The feature variable here is the copper grade.

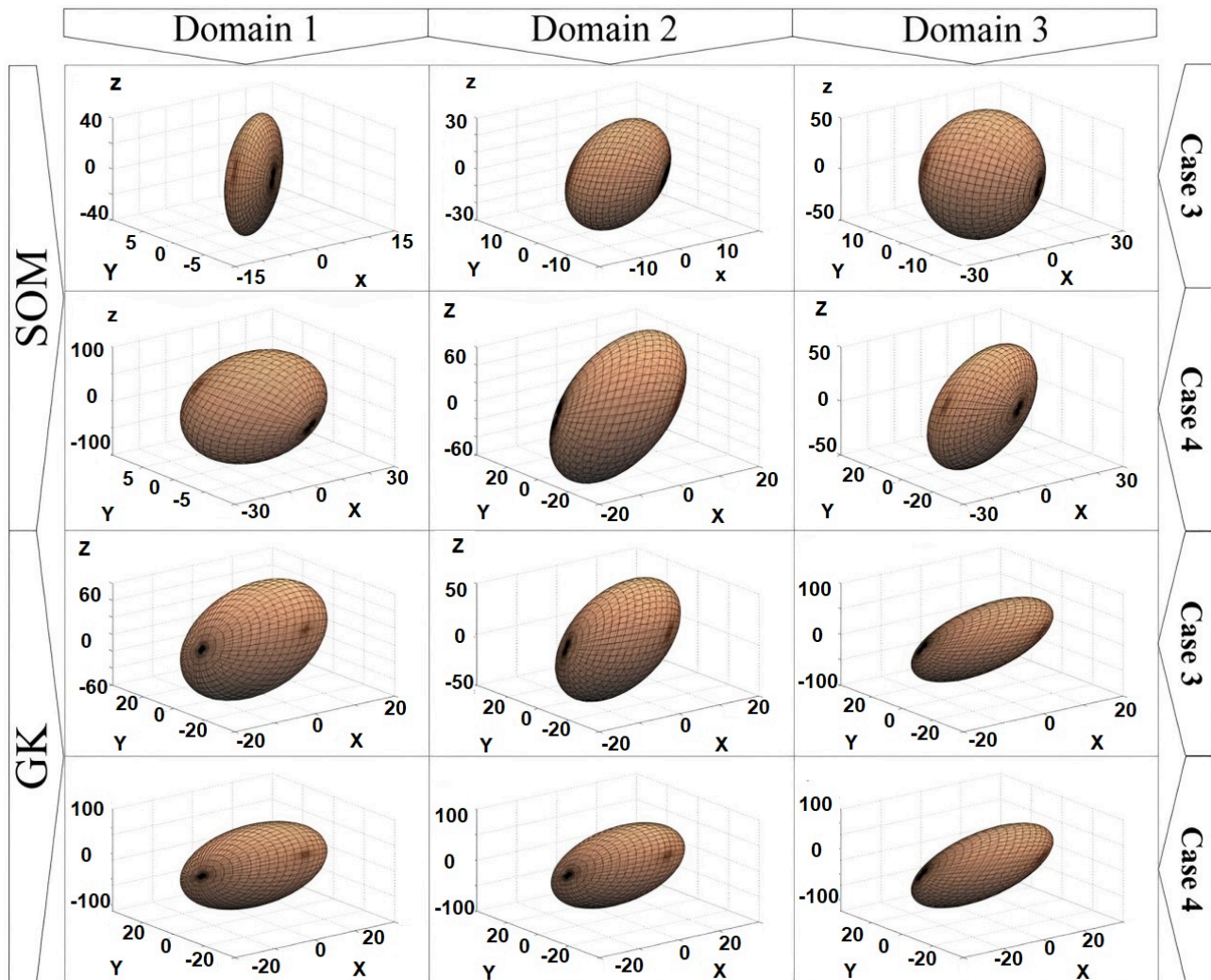


Figure 8. Representations of the anisotropy ellipsoids for each domain. The feature variable here is the copper grade.

as additional input, treating it similarly to other variables. However, this integration poses challenges, as geoscientific data may not share the same coordinate system as the borehole data, complicating the analysis setup. (b) Using this additional information to interpret the clustering outcomes obtained from borehole data. This approach allows for a correlation or alignment between clusters identified in the borehole data and interpretations derived from geophysical or geological surveys, aiding in the validation and enrichment of the clustered information by leveraging complementary datasets.

5. Conclusion

This paper presents a novel approach for delineating geological domains using data-driven methods, specifically clustering algorithms combined with variogram anisotropy modeling. Unlike conventional knowledge-based techniques, this approach leverages data layers to address their inherent limitations. The study focused on the eastern part of the Kahang copper deposit in central Iran,

incorporating geological (lithology, alteration, and mineral zones), geochemical (Cu, Mo, Ag, Au grades), and spatial data from boreholes across two exploration phases. The optimal number of domains was determined by evaluating several hard and fuzzy cluster validity indices. In the fuzzy clustering analysis, appropriate fuzziness and weighting parameters were identified within specified ranges. The results were validated through comparison with geological observations and expert assessments, with the SOM and GK algorithms yielding the best outcomes when using data from cases 3 and 4. Anisotropy modeling for spatial domaining involved assigning data to specific domains and calculating directional variograms to derive the dimensional and directional parameters of the anisotropy ellipsoid for each domain. The SOM algorithm, applied as a hard clustering method, significantly discriminated anisotropy parameters among the domains, as evidenced by a high average value of the Jensen-Shannon (JS) divergence (~0.5). Overall, the data-driven approach combining clustering algorithms with variogram anisotropy modeling proved effective in geological domaining, offering a more reliable and comprehensive understanding of the geological structure in the study area. The results of this study can be helpful in the guiding exploration activities, targeted drilling programs, anomaly detection, enhanced resource estimation, grade control, optimized mine design, economic evaluation, and geotechnical stability.

Acknowledgments

The authors acknowledge the support of the National Agency for Research and Development of Chile, through postgraduate study scholarship ANID-Subdirección de Capital Humano/Doctorado Nacional/2022-21220317 (F. Khorram) grant ANID AFB230001 (F. Khorram and X. Emery).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Afshooni, S. Z., Mirnejad, H., Esmaily, D., & Haroni, H. A. (2013). Mineral Chemistry of Hydrothermal Biotite from the Kahang Porphyry Copper Deposit (NE Isfahan), Central Province of Iran. *Ore Geology Reviews*, *54*, 214-232. <https://doi.org/10.1016/j.oregeorev.2013.04.004>
- Afzal, P., Alghalandis, Y. F., Moarefvand, P., Omran, N. R., & Haroni, H. A. (2012). Application of Power-Spectrum-Volume Fractal Method for Detecting Hypogene, Supergene Enrichment, Leached and Barren Zones in Kahang Cu Porphyry Deposit, Central Iran. *Journal of Geochemical Exploration*, *112*, 131-138. <https://doi.org/10.1016/j.gexplo.2011.08.002>
- Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A. et al. (1996). Validity-Guided (re)clustering with Applications to Image Segmentation. *IEEE Transactions on Fuzzy Systems*, *4*, 112-123. <https://doi.org/10.1109/91.493905>
- Bezdek, J. C. (1975). Mathematical Models for Systematics and Taxonomy. In *8th International Conference on Numerical Taxonomy* (pp. 143-166).

- Bezdek†, J. C. (1974). Cluster Validity with Fuzzy Sets. *Journal of Cybernetics*, 3, 58-73. <https://doi.org/10.1080/01969727308546047>
- Bourgault, G., Marcotte, D., & Legendre, P. (1992). The Multivariate (Co)variogram as a Spatial Weighting Function in Classification Methods. *Mathematical Geology*, 24, 463-478. <https://doi.org/10.1007/bf00890530>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1, 224-227. <https://doi.org/10.1109/tpami.1979.4766909>
- Defays, D. (1977). An Efficient Algorithm for a Complete Link Method. *The Computer Journal*, 20, 364-366. <https://doi.org/10.1093/comjnl/20.4.364>
- Dunn†, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, 4, 95-104. <https://doi.org/10.1080/01969727408546059>
- Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models with Big Data. *Frontiers in Artificial Intelligence*, 3, Article 4. <https://doi.org/10.3389/frai.2020.00004>
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). Wiley. <https://doi.org/10.1002/9780470977811>
- Faraj, F., & Ortiz, J. M. (2021). A Simple Unsupervised Classification Workflow for Defining Geological Domains Using Multivariate Data. *Mining, Metallurgy & Exploration*, 38, 1609-1623. <https://doi.org/10.1007/s42461-021-00428-5>
- Fouedjio, F. (2016). A Hierarchical Clustering Method for Multivariate Geostatistical Data. *Spatial Statistics*, 18, 333-351. <https://doi.org/10.1016/j.spasta.2016.07.003>
- Frey, B. J., & Dueck, D. (2007). Clustering by Passing Messages between Data Points. *Science*, 315, 972-976. <https://doi.org/10.1126/science.1136800>
- Gan, G., Ma, C., & Wu, J. (2020). *Data Clustering: Theory, Algorithms, and Applications, Second Edition*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611976335>
- Gustafson, D., & Kessel, W. (1978). Fuzzy Clustering with a Fuzzy Covariance Matrix. In *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes* (pp. 761-766). IEEE. <https://doi.org/10.1109/cdc.1978.268028>
- Hadiloo, S., Mirzaei, S., Hashemi, H., & Beiranvand, B. (2018). Comparison between unsupervised and supervised fuzzy clustering method in interactive mode to obtain the best result for extract subtle patterns from seismic facies maps. *Geopersia*, 8, 27-34.
- Hartigan, J. A. (1975). *Clustering Algorithms, Wiley Series in Probability and Mathematical Statistics* (p. 365). John Wiley & Sons Inc.
- Haschke, M., Ahmadian, J., Murata, M., & McDonald, I. (2010). Copper Mineralization Prevented by Arc-Root Delamination during Alpine-Himalayan Collision in Central Iran. *Economic Geology*, 105, 855-865. <https://doi.org/10.2113/gsecongeo.105.4.855>
- Hosseini Morshedy, A., Torabi, S. A., & Memarian, H. (2015). A New Method for 3D Designing of Complementary Exploration Drilling Layout Based on Ore Value and Objective Functions. *Arabian Journal of Geosciences*, 8, 8175-8195. <https://doi.org/10.1007/s12517-014-1754-7>
- Hubert, L., & Schultz, J. (1976). Quadratic Assignment as a General Data Analysis Strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190-241. <https://doi.org/10.1111/j.2044-8317.1976.tb00714.x>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means Clustering Algorithms: A Comprehensive Review, Variants Analysis, and Advances in the Era of Big Data. *Information Sciences*, 622, 178-210.

- <https://doi.org/10.1016/j.ins.2022.11.139>
- Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-means. *Pattern Recognition Letters*, 31, 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jiang, J., Chen, M., & Fan, J. A. (2021). Deep Neural Networks for the Evaluation and Design of Photonic Devices. *Nature Reviews Materials*, 6, 679-700. <https://doi.org/10.1038/s41578-020-00260-1>
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis* (p. 342). Wiley.
- Khorram, F., Asghari, O., Memarian, H., Morshedy, A. H., & Emery, X. M. (2021). The Fuzzy Classification of Geometallurgical Domains. *Bulletin of Geophysics and Oceanography*, 62, 467-484.
- Kohonen, T. (1998). The Self-Organizing Map. *Neurocomputing*, 21, 1-6. [https://doi.org/10.1016/s0925-2312\(98\)00030-7](https://doi.org/10.1016/s0925-2312(98)00030-7)
- Krzanowski, W. J., & Lai, Y. T. (1985). A Criterion for Determining the Number of Groups in a Data Set Using Sum-Of-Squares Clustering. *Biometrics*, 44, 23-34. <https://doi.org/10.2307/2531893>
- Kubler, S., Robert, J., Derigent, W., Voisin, A., & Le Traon, Y. (2016). A State-Of-The-Art Survey & Testbed of Fuzzy AHP (FAHP) Applications. *Expert Systems with Applications*, 65, 398-422. <https://doi.org/10.1016/j.eswa.2016.08.064>
- MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, 281-297.
- Miljković, D. (2017). Brief Review of Self-Organizing Maps. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1061-1066). IEEE. <https://doi.org/10.23919/mipro.2017.7973581>
- National Iranian Copper Industries Co (NICICO) (2011). *Summery Geological Report of Kahang Deposit* (p. 23). (In Persian)
- Oliver, M. A., & Webster, R. (1989). A Geostatistical Basis for Spatial Weighting in Multivariate Classification. *Mathematical Geology*, 21, 15-35. <https://doi.org/10.1007/bf00897238>
- Oyelade, J., Isewon, I., Oladipupo, F., Aromolaran, O., Uwoghiren, E., Ameh, F. et al. (2016). Clustering Algorithms: Their Application to Gene Expression Data. *Bioinformatics and Biology Insights*, 10, 237-263. <https://doi.org/10.4137/bbi.s38316>
- Romary, T., Ors, F., Rivoirard, J., & Deraisme, J. (2015). Unsupervised Classification of Multivariate Geostatistical Data: Two Algorithms. *Computers & Geosciences*, 85, 96-103. <https://doi.org/10.1016/j.cageo.2015.05.019>
- Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Ruspini, E. H., Bezdek, J. C., & Keller, J. M. (2019). Fuzzy Clustering: A Historical Perspective. *IEEE Computational Intelligence Magazine*, 14, 45-55. <https://doi.org/10.1109/mci.2018.2881643>
- Somani, G., Gaur, M. S., Sanghi, D., Conti, M., & Buyya, R. (2017). DDoS Attacks in Cloud Computing: Issues, Taxonomy, and Future Directions. *Computer Communications*, 107, 30-48. <https://doi.org/10.1016/j.comcom.2017.03.010>
- Sun, W., Huang, R., Li, H., Hu, Y., Zhang, C., Sun, S. et al. (2015). Porphyry Deposits and

- Oxidized Magmas. *Ore Geology Reviews*, 65, 97-131.
<https://doi.org/10.1016/j.oregeorev.2014.09.004>
- Wang, H., Wang, J., & Wang, G. (2022). A Survey of Fuzzy Clustering Validity Evaluation Methods. *Information Sciences*, 618, 270-297. <https://doi.org/10.1016/j.ins.2022.11.010>
- Xie, X. L., & Beni, G. (1991). A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 841-847.
<https://doi.org/10.1109/34.85677>
- Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2, 165-193. <https://doi.org/10.1007/s40745-015-0040-1>
- Xu, R., & WunschII, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16, 645-678. <https://doi.org/10.1109/tnn.2005.845141>
- Zhu, L., & Li, S. (2017). Application of Clustering Algorithms in Geological Data Analysis: A Review. *Mathematical Problems in Engineering*, 2017, 1-14.