Journal of Geochemical Exploration 292 (2027) 108261

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Comparison of unsupervised clustering methods in distinguishing mineralized catchment basins☆

Shahed Shahrestani [a], Emmanuel John M. Carranza [b,*], Ioan V. Sanislav [a]

[a] College of Science and Engineering, Economic Geology Research Centre (EGRU), James Cook University, Townsville, Australia
[b] Department of Geology, University of the Free State, Bloemfontein, South Africa

ARTICLE INFO

Keywords:
Copper mineralization
Stream sediment samples
K-means clustering
Fuzzy C-means clustering
Hierarchical clustering
Spectral clustering

ABSTRACT

The paper focuses on the use of unsupervised clustering methods to identify and characterize mineralized catchment basins in the Jebal Barez region in southeast Iran, known for its copper deposits. The study utilized a dataset of 658 stream sediment samples that were collected and analyzed for 20 elements. Different clustering approaches – K-means, fuzzy C-means, hierarchical, and spectral clustering – were employed to evaluate, compare and contrast their efficiency in distinguishing and characterizing mineralized catchment basins. To evaluate the performance of the clustering techniques, different clustering validation criteria such as the Caliński–Harabasz Index, Silhouette Index, and within-cluster sum of squares were utilized to determine the optimal number of clusters and to evaluate the outcomes of different clustering approaches. The study found that K-means clustering was the superior method for accurately distinguishing different clusters. Of the 16 known mineral occurrences in the study area, 11 were recognized by two K-means clusters. It was also found that K-means clusters 2 and 3 describe the most promising zones for copper mineralization. Based on these findings, the study demonstrated the potential of unsupervised clustering methods for effective exploration targeting in the region.

1. Introduction

The composition of stream sediment at a catchment outlet reflects source lithology, relief, climate and transport processes (Arribas et al., 2000; Shahrestani et al., 2020), so catchment basins are used as the zone of influence of each sample (Bonham-Carter and Goodfellow, 1984). In the sample catchment basin approach (SCBA), each sample represents only the incremental area between itself and the adjacent upstream samples, and the approach links sediment chemistry to the average composition of source rocks (Carranza and Hale, 1997; Spadoni et al., 2004). Several refinements followed. The extended sample catchment basin (ESCB) approach uses geomorphological boundaries and stream confluences to weight the spatial representativeness of samples (Spadoni, 2006). The true sample catchment basin approach (TSCBA) takes the whole upstream basin into account and delineated anomalous basins containing known mineralization more accurately than SCBA (Garousi Nezhad et al., 2017). Yousefi et al. (2013) defined drainage catchment basins from main rivers, independent of sample location and sampling density, and proposed the weighted drainage catchment basin (WDCB) approach, which weights each basin by the distribution of anomalous and background samples. They showed that WDCB anomalies relate more strongly to geological features than those from SCB modeling or contouring, especially for low-density sampling. Farahbakhsh et al. (2019) extended this with an enhanced WDCB (E-WDCB) that integrates multivariate scores such as factor scores into a catchment basin score. In this study, the multi-element clustering results are mapped on sample catchment basins.

Analysis of stream sediment elemental concentrations is a widely recognized and effective method for detecting and delineating geochemical anomalies related to mineralization (Shahrestani et al., 2020). To interpret stream sediment geochemical data, in addition to outlier detection techniques, cluster analysis is one of the most common exploratory data analysis tools. It groups multi-element observations into homogeneous clusters and reveals underlying patterns in the data (Ikotun et al., 2023). Compared with traditional statistical approaches, clustering is more suitable for identifying complex multi-element relationships in geochemical data (Zhang et al., 2023). Therefore, it plays an important role in mineral exploration by separating geochemical

☆ This article is part of a Special issue entitled: 'In honour of Prof. Renguang Zuo' published in Journal of Geochemical Exploration.

* Corresponding author.
E-mail address: ejmcarranza@gmail.com (E.J.M. Carranza).

https://doi.org/10.1016/j.gexplo.2026.108261
Received 26 May 2026; Received in revised form 3 October 2026; Accepted 21 October 2026
Available online 24 October 2026
0375-6742/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

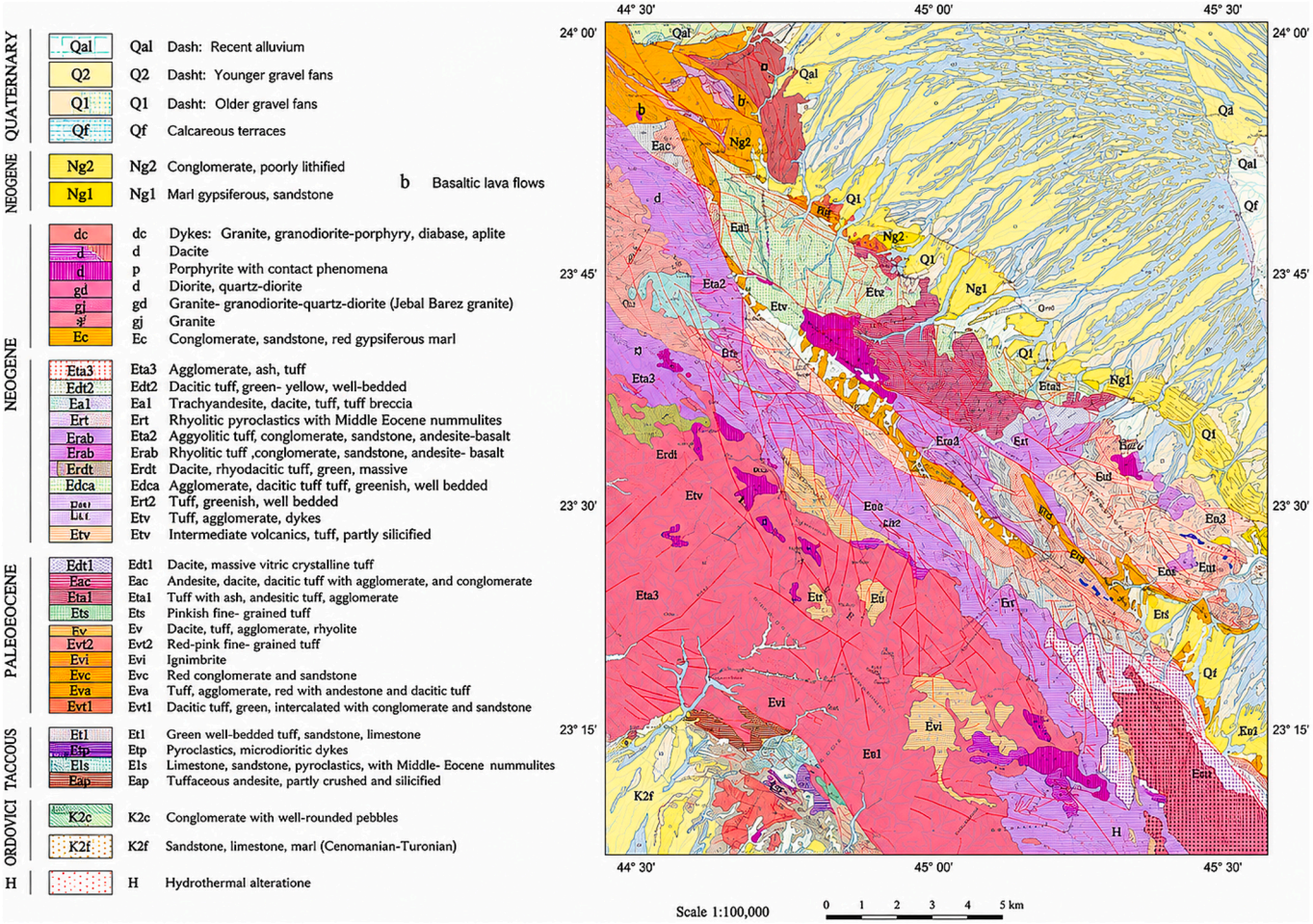


Fig. 1. Geological map of Jebal Barez with a scale of 1:100,000 (Valeh, 1972).

anomalies from background values (Ribeiro et al., 2023). When no prior information about potential anomalies exists, unsupervised clustering methods are particularly useful for analyzing multi-element geochemical data. These methods can divide data into groups that correspond to different geochemical processes or regions with distinct geological or anthropogenic influences, improving data interpretation (Ghezelbash et al., 2023; Cicchella et al., 2022).

Unsupervised clustering methods can be grouped into several main categories, including partitioning, hierarchical, density-based, probabilistic, and model-based approaches (Taha, 2023). Depending on whether the focus is on elements or samples, clustering can also be divided into R-mode, Q-mode, or combined R/Q-mode types (Geranian and Carranza, 2022). Many clustering techniques have been applied to delineate anomalous zones in geochemical exploration, such as genetic K-means clustering (Ghezelbash et al., 2020), standard K-means clustering (e.g., Bigdeli et al., 2022), hierarchical clustering (Ribeiro et al., 2023), grey wolf optimizer (Agarwal et al., 2018) combined with K-means (Daviran et al., 2024), OPTICS (Hajihosseinlou et al., 2024), deep embedded clustering (Hoseinzade and Bazoobandi, 2024), DBSCAN and Gaussian mixture models (Shirjang et al., 2025), and fuzzy C-means (FCM) clustering (Ghezelbash et al., 2025). Hajihosseinlou et al. (2024) reported that DBSCAN extracted anomaly clusters without a predefined cluster number and was more flexible than centroid-based methods. Bitaraf and Ghezelbash (2025) compared DBSCAN, SOM, K-means, K-medians and K-medoids on 800 stream-sediment samples from the Hashtjin district (NW Iran), after ILR and Box–Cox transformation, using PCA to select Cu, Au, Mo and As as clustering inputs. They found DBSCAN to be the best performer (success-rate AUC 84%), followed by SOM (76%), and K-series methods the weakest (62–69%), attributing this to sensitivity to outliers and skewed distributions. A weighted combination of the methods reached an AUC of 92% and reduced the strongest-anomaly area to about 1.85% of the region. Our study differs in using 20 elements and four standard methods (K-means, FCM, hierarchical and spectral), and evaluates them with internal validation indices as well as known occurrences.

However, the effectiveness of each clustering method in identifying potentially-mineralized zones from stream sediment geochemical data remains uncertain. Because clustering is an unsupervised approach, different clustering algorithms may produce inconsistent or even contradictory results. This raises an important question: which clustering technique is the most reliable for distinguishing between background and anomalous populations in regional stream sediment geochemical data, particularly when multiple statistical populations are expected? In this study, three clustering algorithms—including K-means, FCM, hierarchical, and spectral clustering—were applied to stream sediment geochemical data from the Jebal Barez region in southeast Iran, an area known for its copper deposits. The main objective of this study was to evaluate how effectively can each of these unsupervised clustering methods identify copper-mineralized catchment basins, especially in cases where single-element copper data show weak correlation with known copper occurrences such as in the study area.

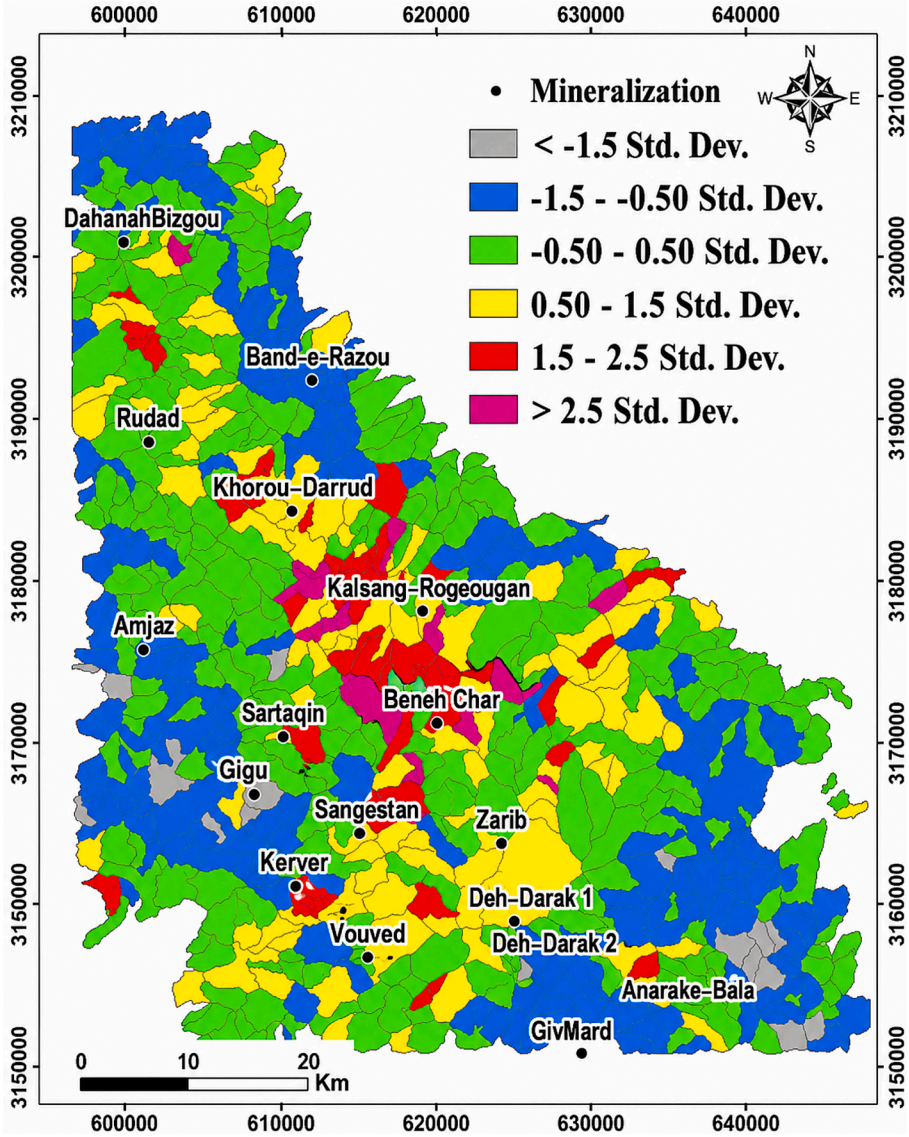


Fig. 2. Spatial distribution of raw Cu data classified based on standard deviation.

2. Materials and methods

2.1. Geological setting

The area under investigation lies within the Shar-e-Babak-Bam ore field in the southern region of the Central Iranian volcano–plutonic magmatic arc, named the Jebal Barez Plutonic Complex (JBPC), which has a NW–SE trend and covers an area of approximately 500 km^2 (Rasouli et al., 2016). The Sahand–Basman metallogenic province and Tertiary volcanic belt within the Central Iranian volcano–plutonic magmatic arc is significantly impacted by the Kuh Jebal Barez range, running from northwest to southeast, which passes through the area under study (e.g., Ziaii et al., 2019; Borna and Sodishoar, 2005). The calc-alkaline and occasional tholeiitic characteristics of the Eocene volcanic and intrusive suites in the Jebal Barez Mountains prior to the collisional period align with a subduction-related island arc context (Shahabpour, 2007).

The study area has numerous towering mountains, with the highest peak measuring an impressive 3741 m in altitude. In addition, the area has numerous perennial rivers. According to Valeh (1972), volcanic activity in the region began in the Late Cretaceous period and persisted until the Middle-Late Eocene epoch. Many granite and granodiorite outcrops can be found in the southern and eastern-southern parts of the area, while several dacite domes formed later in the region, now covering the volcanoclastic rocks (Fig. 1).

The Kerman Belt comprises two significant stages of intrusions: the first stage occurred during the Eocene–Oligocene period and is referred to as "Jebal Barez" while the second stage, which took place during the mid to late Miocene, involved dioritic to granodioritic porphyry intrusions commonly known as "Kuh Panj" (Sadigh et al., 2023; Hosseini et al., 2017). According to Hezarkhani et al. (1999) and Asadi (2013), copper mineralization in the Kuh Panj porphyry systems is mainly related to potassic alteration and shows limited association with phyllic alteration. In contrast, the Jebal Barez porphyry systems display significant phyllic alteration, which may be linked to copper mineralization.

Extensive studies of the porphyries in Jebal Barez have shown that they are primarily made up of plagioclase, quartz, K-feldspar, muscovite, and olivine diorite and granodiorite, with minor hornblende, biotite, and clinopyroxene. The mineralogical composition of the Jebal Barez porphyries has been described consistently in various investigations, including those by Asadi et al. (2014), Shafiei et al. (2009), and Dargahi et al. (2010). The granitoids of the Jebal Barez type, in their unweathered state, comprise oligoclase (32%), orthoclase (10%), quartz (5%),

Table 1
Silhouette index (SI), within-cluster sum of squares (WCSS), and Calinski–Harabasz (C-H) index values of individual clusters resulting from four clustering methods.

Num.Class	K-means			Num.Class	Hierarchical		
	SI	WCSS	C–H		SI	WCSS	C–H
2	0.55	187.40	30.21	2	0.52	203.11	27.12
3	0.37	157.38	36.97	3	0.26	181.22	21.76
4	0.36	137.19	37.17	4	0.26	157.57	28.38
5	0.31	127.47	50.68	5	0.30	140.27	31.64
6	0.35	114.38	59.00	6	0.28	127.38	33.11
7	0.33	109.64	50.93	7	0.27	118.84	38.32
8	0.30	104.03	48.62	8	0.24	113.24	38.11
9	0.28	100.80	55.72	9	0.23	108.55	42.50
10	0.27	97.47	69.54	10	0.23	103.98	47.12

Num.Class	Fuzzy C-means			Num.Class	Spectral		
	SI	WCSS	C-H		SI	WCSS	C–H
2	0.51	188.21	25.10	2	0.48	198.67	21.93
3	0.33	158.70	31.47	3	0.40	174.28	51.29
4	0.23	144.47	31.80	4	0.25	156.27	42.07
5	0.25	128.80	34.81	5	0.29	139.22	42.80
6	0.18	124.40	34.01	6	0.25	127.44	44.57
7	0.19	114.50	42.19	7	0.25	120.40	46.06
8	0.14	114.75	37.82	8	0.24	113.41	38.10
9	0.13	110.53	38.99	9	0.21	106.43	42.50
10	0.10	108.52	38.66	10	0.20	101.16	51.80

hornblende (6%), and biotite (5%), with a groundmass of aplitic quartz (24%) and orthoclase (16%). These rocks also contain small quantities of pyrite, chalcopyrite, magnetite, zircon, titanite, and apatite (Atapour and Aftabi, 2021). The copper mineralization in the Jebal Barez Mountains are typically associated with veins within faulted andesite, with some associations to dykes in granodiorite and diorite. The area is of significant importance due to several occurrences of copper mineralization such as Kerver, Vouved, Rudad, Sartaqin, Gigu, Amjaz, Band-e-Razou, DahanahBizgou, GivMard, Anarake–Bala, Khorou–Darrud, Kalsang–Rogeougan, Deh–Darak 1, Deh–Darak 2, Zarib, Beneh Char, and Sangestan (Ziaii et al., 2011; Bedakhshan and Sodishoar, 2000). In the same region, Ziaii et al. (2019) applied concentration-gradient modeling to 490 regional stream-sediment samples from Jebal-Barez, using the ratio of supra-ore (Pb × Zn) to sub-ore (Cu × Ag) gradients to distinguish blind from zone-dispersed mineralization. Their results indicated blind mineralization in the SW (Kerver) area, while the central zone, which hosts most of the known outcropping deposits, was interpreted as weak, zone-dispersed mineralization. Their local follow-up at Kerver included drilling. That study used a single-model, element-ratio approach, whereas the present study applies multi-element unsupervised clustering to all 20 elements.

2.2. Geochemical dataset

This study used data from 658 stream sediment samples collected by the Geological Survey and Mineral Exploration of Iran at a sampling density of about 1 samples 4 per km^2, and were analyzed for Ag, As, Au, Ba, Bi, Cd, Co, Cr, Cu, Hg, Ni, Pb, Sb, Sc, Sn, Sr, V, W, Y, and Zn. The geochemical dataset used in this study was obtained from a pre-existing regional geochemical survey and was not generated specifically for the present study. This dataset has previously been used in regional geochemical and mineral prospectivity studies, such as Ziaii et al. (2011) and Ziaii et al. (2019). The available published sources do not provide sufficient information to independently assess the original analytical QA/QC procedures, such as the use of duplicate samples, certified reference materials, blanks, or analytical precision. The original technical documentation for the regional survey is also not currently accessible to the authors. Therefore, no assumptions were made regarding the availability or performance of these QA/QC materials. The present study consequently treats the dataset as an existing regional geochemical dataset, and data preprocessing was conducted using the available analytical results and their suitability for the statistical analyses applied herein. This limitation is acknowledged when interpreting the results.

The summary statistics of the element data are presented in Supplementary Table S1. Histograms of element concentrations in stream sediments from the study area are presented in Supplementary Fig. S1. Regarding the compositional nature of multivariate geochemical data, the data studied here were subjected to centred log-ratio (clr) transformation to mitigate the closure problem inherent in compositional data (e.g., Templ et al., 2008). The original elemental concentrations exhibit variable distributional characteristics (Fig. S1). Several elements, including Au, Cd, Cr, Hg, Ni, Sb, and Y, show pronounced positive skewness, whereas other elements display relatively more symmetric distributions. Thus, normality of the original geochemical variables was not assumed. Given the compositional nature of multivariate geochemical data, the dataset was transformed using the centred log-ratio (clr) transformation. The clr transformation addresses the closure constraint inherent in compositional data and places the variables on a log-ratio scale. It also reduces the influence of the strong skewness observed in several of the original elemental distributions.

2.3. Data analysis methods

2.3.1. K-means clustering

K-means clustering is a popular and widely utilized unsupervised machine learning algorithm that partitions multivariate data points into K number of clusters, with the aim of minimizing the sum of squared distances between each data point and its assigned cluster centroid (WCSS) (Hartigan and Wong, 1979; Liu et al., 2022; Ikotun et al., 2023). For K-means clustering algorithm, the user must first specify K, and then select an appropriate distance metric and an initialization strategy for the cluster centroids. The algorithm iteratively assigns each data point to the nearest cluster centroid based on the selected distance metric, and then updates the cluster centroids by computing the mean of the assigned data points. This process is repeated until convergence is achieved, i.e., until cluster assignments no longer change significantly (Ghezelbash et al., 2019, 2023; Jain, 2010).

The optimal K for a dataset can be determined using various techniques, such as the elbow method or silhouette analysis (Ghezelbash et al., 2020). The choice of distance metric also depends on the nature of the data and the specific application. For high-dimensional data, Euclidean distance is the most widely used metric in K-means clustering (Jain, 2010; Ghezelbash et al., 2019). One advantage of K-means clustering over other methods, such as hierarchical clustering, is its computational efficiency, making it suitable for handling large datasets (Das et al., 2023). However, the algorithm has some limitations, such as sensitivity to the initial placement of centroids, and the assumption of equal variance within each cluster (Hartigan and Wong, 1979).

2.3.2. Fuzzy C-means clustering

FCM clustering is used to partition a dataset consisting of n data points in d-dimensional space (Setiawan et al., 2023). In this algorithm, instead of assigning data points to exclusive clusters as in K-means clustering, each data point has a membership value indicating its degree of belongingness to each cluster (Vriend et al., 1988; Zhang et al., 2023). The algorithm starts by randomly or arbitrarily initializing fuzzy cluster centers μ. These centers represent prototypes or centroids of clusters that one aims to form. The objective of FCM clustering is to minimize the fuzzy objective function, which is a weighted sum of squared distances between each data point and all cluster centers, where the membership values determine the weightings.

The algorithm iteratively updates the membership values of each data point based on a fuzzifier parameter m, which determines the

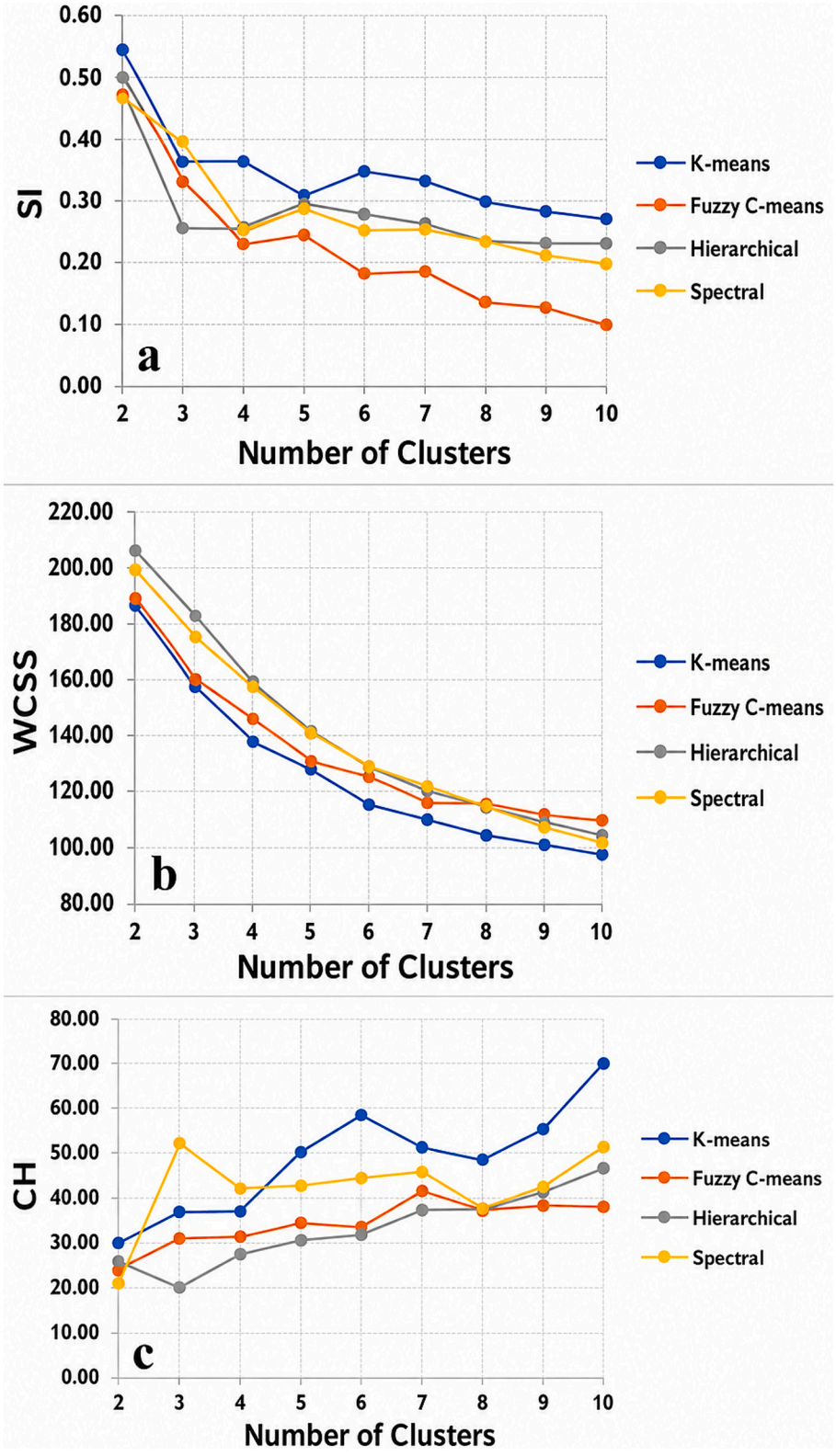


Fig. 3. Variation of Silhouette index (a), WCSS (b), and Caliński–Harabasz (C-H) index (c) values against the number of clusters.

fuzziness of a partition. The membership values are calculated by considering distances between data points and cluster centers and assigning higher membership values to data points closer to a corresponding cluster center. Once membership values are updated, cluster centers are recalculated using a weighted mean of the data points, where the membership values serve as weights. This allows each cluster center to be influenced by data points according to their degree of membership in that cluster. The iteration process continues until convergence, which occurs when membership values and cluster centers no longer change substantially or after a specified number of iterations. At the end of the algorithm, fuzzy clusters with their respective membership values and cluster centers are obtained (Daviran et al., 2020).

Table 2
Number of Cu occurrences associated with individual clusters determined by different clustering methods.

Cluster/clustering method	K-means	Fuzzy C-means	Hierarchical	Spectral
Cluster 1	2	2	0	2
Cluster 2	4	2	3	3
Cluster 3	7	4	0	4
Cluster 4	2	2	7	2
Cluster 5	1	3	2	5
Cluster 6	0	3	2	0
Cluster 7	*	0	2	0

2.3.3. Hierarchical clustering

One approach to organizing multivariate data, known as hierarchical organization, involves using a proximity matrix to create a dendrogram (Cabezas et al., 2023). This dendrogram serves as a hierarchical structure, with the entire multivariate data as the root, and subsequent leaf nodes corresponding to individual data points (Zhang et al., 2023; Carlsson and Mémoli, 2010). Cluster analysis, which groups together similar data points in multivariate data, can be performed in either a bottom-up (agglomerative) or top-down (divisive) manner. Agglomerative hierarchical clustering algorithms merge clusters of data points, starting from the leaf nodes, by calculating distances between clusters and merging two clusters with the minimum distance. This process results in a hierarchy that is updated with each merger until a single cluster containing all observations is reached. Divisive algorithms take the opposite approach and begin by treating an entire dataset as one cluster, iteratively splitting the data into smaller subsets until a termination criterion is met (Wegmann et al., 2021).

In order to measure the similarity between data points, a distance or proximity measure is required. The most widely used distance measure is the Euclidean distance. Alternative distance measures such as Manhattan, Minkowski, and Hamming distances may also be utilized. These alternative measures offer different ways to quantify the dissimilarity or similarity between data points, providing flexibility in capturing various aspects of multivariate data (Govender and Sivakumar, 2020).

2.3.4. Spectral clustering

Spectral clustering is a type of unsupervised machine learning algorithm that aims to cluster or group data points together based on their similarity or affinity (Khan et al., 2023). The key idea behind spectral clustering is to transform the data into a lower-dimensional space using the eigenvectors and eigenvalues of a matrix derived from the pairwise similarity or affinity matrix (Shang et al., 2023). In spectral clustering, the first step involves constructing the similarity matrix, which measures the similarity or affinity between each pair of data points. This matrix captures the relationships and provides a basis for further analysis. Next, a normalized Laplacian matrix is created from the similarity matrix (Spielman, 2007). The Laplacian matrix summarizes the connectivity of the data points and is used to reveal the geometric structure of the data. By performing eigen decomposition on the Laplacian matrix, the eigenvectors and eigenvalues are obtained. The eigenvectors serve as a transformation of the data into a reduced-dimensional space, and the eigenvalues indicate the importance of individual eigenvectors (e.g., Wu et al., 2015).

2.4. Clustering validation criteria

To assess the effectiveness of unsupervised clustering techniques in identifying and characterizing mineralized catchment basins, three methods – the Caliński–Harabasz (C-H) Index, Silhouette index, and WCSS – were implemented in this study to select the optimal number of clusters and to evaluate the outcomes of the different unsupervised clustering techniques described above. Consider a dataset $X \in R^{n\times m}$, which contains n points in a m-dimensional Euclidean space, where each point is represented as $x_i = (x_{i,1}, \ldots, x_{i,m})$. The objective is to partition X into K ($K \geq 2$) mutually disjoint and nonempty clusters. Assuming that k is pre-determined, one can represent this k-partition as $\{X_1, \ldots, X_k\}$ of a set $\{x_1, \ldots, x_n\}$, where each element of a set is assigned a cluster number using a surjection P: [1: n] onto → [1: k], where P(i) ∈ [1: k] represents the cluster number of the ith point.

The C-H index also known as the "variance ratio criterion," is expressed as (Caliński and Harabasz, 1974):

$$C-H = \frac{n-k\sum_{i=1}^{n}\left\|\mu-\mu_{C(i)}\right\|^2}{k-1\sum_{i=1}^{n}\left\|x_i-\mu_{C(i)}\right\|^2} \quad (1)$$

where $\mu_{C(i)}$ represents the centroid of a cluster containing the i-th point. The numerator calculates the between-cluster variance while the denominator calculates the within-cluster variance. It can demonstrate that maximizing the C-H index is equivalent to minimizing the unweighted WCSS in clustering tasks.

In 1987, Rousseeuw (1987) introduced the Silhouette index (SI), which refers to the average silhouette value across all samples. The SI is used to effectively assign samples to their respective clusters. It uses a_i to represent the average distance between sample i and other samples within the same cluster, while b_i represents the minimal arithmetic mean of the distance between sample i and samples in different clusters (Gagolewski et al., 2021), thus:

$$SI = \frac{1}{n}\sum_{i=1}^{n}\frac{b_i-a_i}{\max\{ai,bi\}} \quad (2)$$

A sample that is classified almost perfectly will have a SI value that is nearly 1. Conversely, a negative SI value indicates that a sample i is classified incorrectly. To identify the ideal value for K, one should select the value that has the highest overall average Silhouette score (Rousseeuw, 1987). The known Cu occurrences were not used in the clustering and were used only afterwards as independent data to compare how the clusters relate to mineralization. The number of recognized occurrences does not by itself measure prospectivity, because it ignores the size of the search space, i.e., the area that must be explored to capture them (Agterberg and Bonham-Carter, 2005; Carranza, 2008).

3. Results and discussion

3.1. Uni-element anomaly map of Cu

Elemental concentrations of stream sediment samples can be attributed to their respective sample catchment basins as upstream zones of influence per sample. In this regard, before applying multivariate techniques, the spatial distribution of the raw Cu data was analyzed to examine whether catchments with known copper occurrences could be distinguished effectively. Fig. 2 illustrates the spatial distribution of Cu raw data classified based on multiple standard deviation ranges. It is evident that out of all the known copper occurrences, only the Kalsang–Rogeougan and Band-e-Razou copper occurrences, and a copper occurrence in the southwestern part of the study area, coincide with the top two classes of the raw Cu data. Thus, significant factors that contribute to a certain portion of the variation in Cu concentration data may not always be attributable to known anomaly sources. This suggests that the contribution of lithology or hydrothermal alteration and other factors to the variation in elemental concentration must be considered in detecting mineralization-related geochemical anomalies. Additionally, geochemical anomalies may be reflected in multi-element relationships rather than just the main element of interest. This means that there may be a broader range of elements that show a distinct pattern or "footprint" in the spatial distribution of multi-element stream sediment geochemical data.

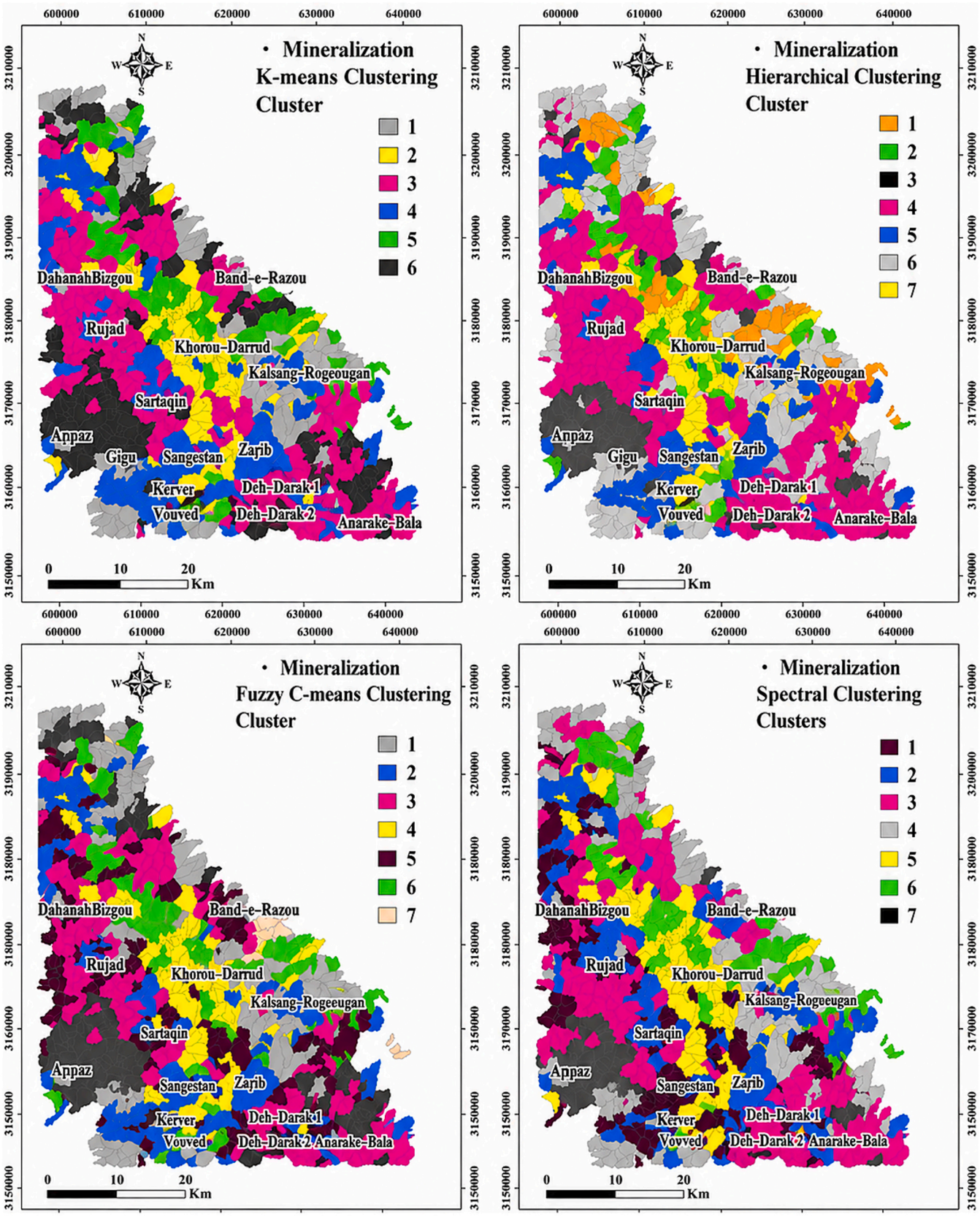


Fig. 4. Sample catchment basin-based spatial distributions of clusters determined by K-means, hierarchical, fuzzy C-means, and spectral clustering methods.

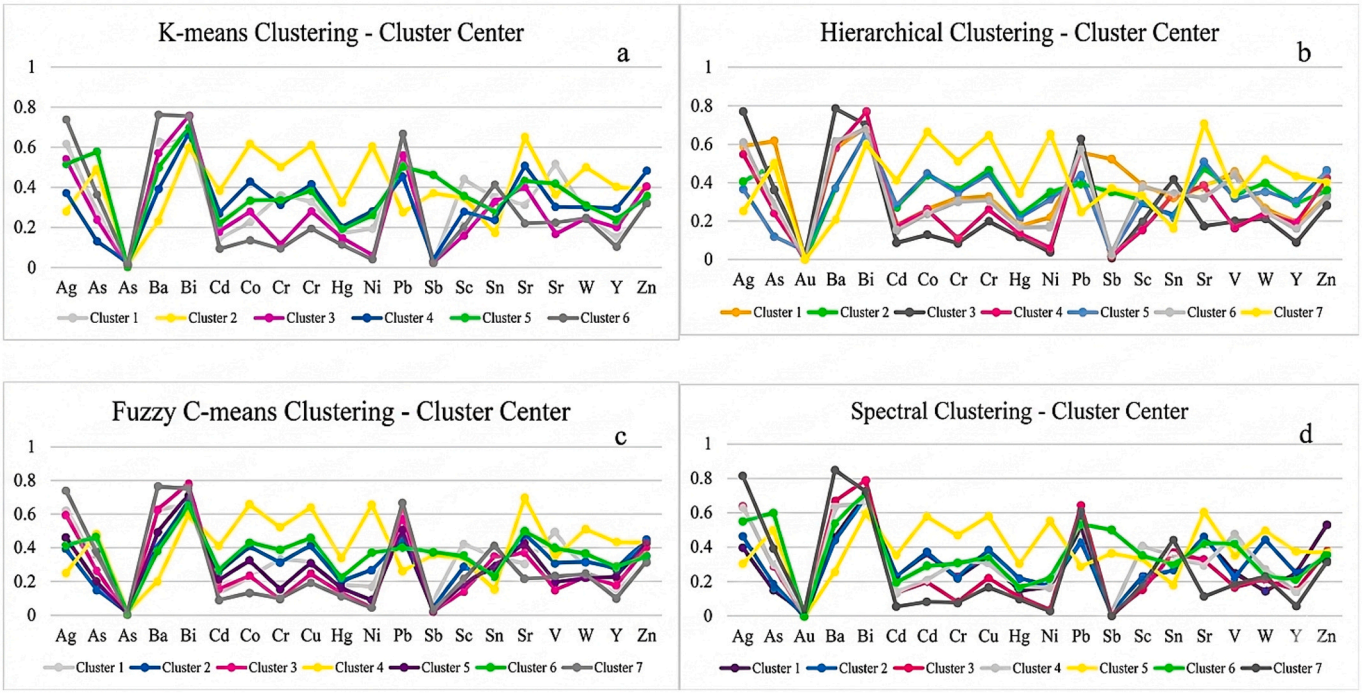


Fig. 5. Values of cluster centers for 20 elements resulting from K-means (a), hierarchical (b), fuzzy C-means (c), spectral clustering (d) methods.

3.2. Clustering

Prior to clustering, outlier values were identified and replaced by the 95th percentile value for each element, as there were insufficient outliers to merit further investigation. Next, all element concentrations were standardized to the range from 0 to 1 to simplify data analysis and interpretation. Due to the potential for variation in the results of different clustering techniques, several iterations were conducted to ensure convergence of the clustering outcomes. Lastly, each clustering method was executed with a pre-determined selection of clusters ranging from 2 to 10. For each iteration of the clustering process, the SI, WCSS, and C-H index were calculated to provide a comprehensive evaluation of the clustering outcomes (Table 1). After applying the clustering algorithms, the optimal numbers of clusters for K-means, FCM, hierarchical, and spectral clustering methods were determined as 6, 7, 7, and 7, respectively (Table 1, Fig. 3).

Based on the SI, WCSS, and C-H index in Table 1, the K-means clustering had the highest SI and C-H index. This indicates that K-means clustering was superior in accurately distinguishing between different clusters. However, the high number of variables and compact variations in the elemental concentrations, as shown in Supplementary Fig. S2, makes it challenging to effectively separate possible clusters. Additionally, many elements exhibited normal distribution, which may not contribute sufficient information when conducting clustering methods. With regards to WCSS, K-means clustering exhibited a lower average value during iterations in each specific cluster number and in different numbers of classes among all the elements.

Of the 16 mineral occurrences, two K-means clusters identified 11 mineral occurrences, while hierarchical clustering identified 10 mineral occurrences (Table 2). FCM clustering showed relatively even distribution of the known mineral occurrences among the 7 clusters. Two spectral clusters are associated with 9 out of the 16 known mineral occurrences. The mineral occurrences Amjaz and Gigu consistently appear together in a single cluster yielded by all clustering methods used, indicating the unique geochemical properties of stream sediments and their corresponding upstream lithological units in the area. The different clustering methods used revealed distinct spatial distributions of clusters in the southwest, central, and eastern parts of the study area. This result is attributed to the geological variation across the region. Notably, only the central zone exhibited the presence of mineral occurrences, namely Band-e-Razou, Korreh–Darrud, and Kalsang–Rogeougan. After analyzing the results of the different clustering methods, it is apparent that a distinct cluster is present in the central zone, and it is distinctly different from the other clusters (as shown in Fig. 5). This zone is characterized by high values (or enrichment) of Cd, Co, Cr, Cu, Ni, and Sr, and relatively lower values of Pb, Sn, Ag, and Ba. Based on these findings, the study area can be classified into two main geological/geochemical zones.

Previous studies in the Jebal Barez region have highlighted the Kerver area in the southwest as a promising location for exploring blind mineralization (Safari et al., 2016; Ziaii et al., 2011). Borna and Sodishoar (2005) utilized the zonality method on a local scale to investigate blind porphyry copper deposits in the Kerver1 and Kever2 areas. Hence, it is possible that some mineralization in the Jebal Barez region may be blind, and may not necessarily be identified through the analysis of surficial stream sediment geochemical anomalies.

According to Ziaii et al. (2011), lithology is a major controlling factor for the occurrence of copper deposits in the study area. Eight of the 15 copper deposits, which are the same deposits studied by Ziaii et al. (2011), are hosted by granite–granodiorite–quartz–diorite. The remaining deposits are hosted by rhyolitic pyroclastics (2 deposits), porphyrites (2 deposits), silicified intermediate volcanics and tuffs (1 deposit), dacites (1 deposit), and agglomerate–ash-tuff (1 deposit). In this study, the spatial distributions of several elements captured by the clustering methods mainly reflect controls by lithology, implying that removal of background variation due to lithology using methods like the sample catchment basin approach (e.g., Carranza and Hale, 1997) would significantly improve the recognition of mineralization-related geochemical anomalies. However, the mineralized zones throughout the study area have a diverse range of copper concentrations and erosion levels. As such, resulting secondary haloes cannot be specifically classified based solely on copper concentration in stream sediments, which can be observed in the map of the raw Cu data. The usefulness of a clustering result for exploration depends both on the proportion of known occurrences it captures and on the size of the search space it defines: a method capturing many occurrences within a large area

reduces exploration risk less than one capturing a similar number within a smaller area (Carranza, 2008; Yousefi and Carranza, 2015). Here, the comparison of methods is based on the number of recognized occurrences (Table 2), and the search space of each favourable cluster was not quantified. For prospecting of undiscovered deposits, the most relevant areas are catchments within K-means clusters 2 and 3 that contain no known occurrence, since they share the geochemical signature of catchments hosting known occurrences. The Kerver area, proposed in earlier work as a target for blind mineralization (Safari et al., 2016; Ziaii et al., 2011), should be assessed in the same context. These catchments could be prioritized, for example, by their distance to the cluster center or by combining them with geological, structural and alteration data. With only 16 known occurrences, the results provide a relative comparison of methods, not a probability of discovery. Based on the combined clustering results, known Cu occurrences, geochemical characteristics, spatial distribution, and geological controls, K-means clusters 2 and 3 were identified as favourable geochemical targets for further exploration of Cu mineralization. These clusters contain 11 of the 16 known Cu occurrences and are associated with geochemical patterns and geological units considered favourable for Cu mineralization. However, these clusters should be considered exploration targets rather than confirmed mineralized zones, as the clustering analysis is unsupervised and is based primarily on regional stream-sediment geochemistry. Catchments in K-means clusters 2 and 3 cover approximately 36% of the study area and contain 11 of the 16 known occurrences, i.e., about 1.9 times more occurrences than expected for a random selection of the same area. Catchments in these clusters that contain no known occurrence are the candidate targets for undiscovered mineralization (Fig. 4). They are located in the northwest, north-central, far-northern and southeastern parts of the area, in addition to the Kerver area proposed earlier as a blind target. These areas share the geochemical signature of the catchments hosting known occurrences (Fig. 5) and should be prioritized with geological, structural and alteration data, since the favourable clusters still cover a large part of the area.

4. Conclusion

This study show that unsupervised clustering methods are valuable for highlighting mineralization-related stream sediment geochemical anomalies in the study area. By utilizing these techniques to analyze stream sediment geochemical data, mineralized catchment basins can be identified and characterized. This study found that, as uni-element anomaly mapping is not always effective, multivariate clustering approaches provide better capability for detecting and distinguishing anomalous zones associated with mineralization. Among the four clustering methods used in this study, K-means clustering was found to be the most effective. This method was able to accurately distinguish between different clusters, resulting in the identification of two promising zones for exploration of copper mineralization. These promising zones, described by clusters 2 and 3 determined by K-means clustering, warrant further exploration. The study also highlighted the importance of considering multi-element data, particularly in cases where univariate analyses are not effective. Overall, the study demonstrates the potential of unsupervised clustering methods for identifying and characterizing mineralized catchment basins using stream sediment geochemical data. The findings suggest that these techniques can be particularly useful in cases where uni-element mapping is not effective in the identification of new exploration targets.

CRediT authorship contribution statement

Shahed Shahrestani: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Emmanuel John M. Carranza:** Writing – review & editing, Methodology, Formal analysis. **Ioan V. Sanislav:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gexplo.2026.108261.

Data availability

The authors do not have permission to share data.

References

Agarwal, A., Chandra, A., Shalivahan, S., Singh, R.K., 2018. Grey wolf optimizer: a new strategy to invert geophysical data sets. Geophys. Prospect. 66 (6), 1215–1226.
Agterberg, F.P., Bonham-Carter, G.F., 2005. Measuring the performance of mineral-potential maps. Nat. Resour. Res. 14 (1), 1–17.
Arribas, J., Critelli, S., Le Pera, E., Tortosa, A., 2000. Composition of modern stream sand derived from a mixture of sedimentary and metamorphic source rocks (Henares River, Central Spain). Sediment. Geol. 133 (1–2), 27–48.
Asadi, S., 2013. Geochemistry of Selected Productive and Barren Intrusions in Shahr Babak Copper Complex, Urumiyeh–Dokhtar Volcano-Magmatic Belt. Unpublished Ph.D. thesis. Shiraz University, Iran, p. 222.
Asadi, S., Moore, F., Zarasvandi, A., 2014. Discriminating productive and barren porphyry copper deposits in the southeastern part of the central Iranian volcano-plutonic belt, Kerman region, Iran: a review. Earth Sci. Rev. 138, 25–46.
Atapour, H., Aftabi, A., 2021. Petrogeochemical evolution of calcalkaline, shoshonitic and adakitic magmatism associated with Kerman Cenozoic arc porphyry copper mineralization, southeastern Iran: a review. Lithos 398, 106261.
Bedakhshan, G.H., Sodishoar, P., 2000. Geochemical exploratory exploration in block 2 (Jebal Barez, Sabzevaran, Hana). In: National Iranian Copper Industries Company (NICICO) Reports.
Bigdeli, A., Maghsoudi, A., Ghezelbash, R., 2022. Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran. J. Geochem. Explor. 233, 106923.
Bitaraf, A., Ghezelbash, R., 2025. Evaluation of SOM, DBSCAN, and K-series clustering algorithms for CuAu anomaly exploration in sample catchment basins: a case study from Hashtjin district, NW Iran. J. Geochem. Explor. 107959.
Bonham-Carter, G.F., Goodfellow, W.D., 1984. Autocorrelation structure of stream sediment geochemical data: interpretation of Zn and Pb anomalies, Nahanni River area, Yukon-Northwest Territories, Canada. In: Verly, G., David, M., Journel, A., Marechal, A. (Eds.), Geostatistics for Natural Resources Characterization, Vol. 2. Reidel Dordrecht, pp. 817–829.
Borna, B., Sodishoar, P., 2005. Primary exploration of copper in Kerver. In: National Iranian Copper Industries Company (NICICO) Reports.
Cabezas, L.M., Izbicki, R., Stern, R.B., 2023. Hierarchical clustering: visualization, feature importance and model selection. Appl. Soft Comput. 141, 110303.
Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. Theory Methods 3 (1), 1–27.
Carlsson, G.E., Mémoli, F., 2010. Characterization, stability and convergence of hierarchical clustering methods. J. Mach. Learn. Res. 11 (Apr), 1425–1470.
Carranza, E.J.M., 2008. Geochemical anomaly and mineral prospectivity mapping in GIS. In: Handbook of Exploration and Environmental Geochemistry, vol. 11. Elsevier, Amsterdam.
Carranza, E.J.M., Hale, M., 1997. A catchment basin approach to the analysis of reconnaissance geochemical-geological data from Albay Province, Philippines. J. Geochem. Explor. 60 (2), 157–171.
Cicchella, D., Ambrosino, M., Gramazio, A., Coraggio, F., Musto, M.A., Caputi, A., Avagliano, D., Albanese, S., 2022. Using multivariate compositional data analysis (CoDA) and clustering to establish geochemical backgrounds in stream sediments of an onshore oil deposits area. The Agri River basin (Italy) case study. J. Geochem. Explor. 238, 107012.
Dargahi, S., Arvin, M., Pan, Y., Babaei, A., 2010. Petrogenesis of post-collisional A-type granitoids from the Urumieh–Dokhtar magmatic assemblage, Southwestern Kerman, Iran: constraints on the Arabian–Eurasian continental collision. Lithos 115 (1–4), 190–204.
Das, J.N., Tiwari, M.K., Sinha, A.K., Khanzode, V., 2023. Integrated warehouse assignment and carton configuration optimization using deep clustering-based evolutionary algorithms. Expert Syst. Appl. 212, 118680.
Daviran, M., Ghezelbash, R., Maghsoudi, A., 2024. GWOKM: a novel hybrid optimization algorithm for geochemical anomaly detection based on Grey wolf optimizer and K-means clustering. Geochemistry 84 (1), 126036.
Daviran, M., Maghsoudi, A., Cohen, D.R., Ghezelbash, R., Yilmaz, H., 2020. Assessment of various fuzzy c-mean clustering validation indices for mapping mineral prospectivity: combination of multifractal geochemical model and mineralization processes. Nat. Resour. Res. 29, 229–246.

Farahbakhsh, E., Chandra, R., Eslamkish, T., Müller, R.D., 2019. Modeling geochemical anomalies of stream sediment data through a weighted drainage catchment basin method for detecting porphyry Cu-Au mineralization. J. Geochem. Explor. 204, 12–32.
Gagolewski, M., Bartoszuk, M., Cena, A., 2021. Are cluster validity measures (in) valid? Inf. Sci. 581, 620–636.
Garousi Nezhad, S., Mokhtari, A.R., Rodsari, P.R., 2017. The true sample catchment basin approach in the analysis of stream sediment geochemical data. Ore Geol. Rev. 83, 127–134.
Geranian, H., Carranza, E.J.M., 2022. Mapping of regional-scale multi-element geochemical anomalies using hierarchical clustering algorithms. Nat. Resour. Res. 31 (4), 1841–1865.
Ghezelbash, R., Daviran, M., Maghsoudi, A., Ghaeminejad, H., Niknezhad, M., 2023. Incorporating the genetic and firefly optimization algorithms into K-means clustering method for detection of porphyry and skarn Cu-related geochemical footprints in Baft district, Kerman, Iran. Appl. Geochem. 148.
Ghezelbash, R., Daviran, M., Maghsoudi, A., Hajihosseinlou, M., 2025. Density based spatial clustering of applications with noise and fuzzy C-means algorithms for unsupervised mineral prospectivity mapping. Earth Sci. Inf. 18 (2), 217.
Ghezelbash, R., Maghsoudi, A., Carranza, E.J.M., 2019. Mapping of single-and multi-element geochemical indicators based on catchment basin analysis: application of fractal method and unsupervised clustering models. J. Geochem. Explor. 199, 90–104.
Ghezelbash, R., Maghsoudi, A., Carranza, E.J.M., 2020. Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm. Comput. Geosci. 134, 104335.
Govender, P., Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). Atmos. Pollut. Res. 11 (1), 40–56.
Hajihosseinlou, M., Maghsoudi, A., Ghezelbash, R., 2024. A comprehensive evaluation of OPTICS, GMM and K-means clustering methodologies for geochemical anomaly detection connected with sample catchment basins. Geochemistry 84 (2), 126094.
Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. Appl. Stat. 28 (1), 100–108 p.105538.
Hezarkhani, A., Williams-Jones, A.E., Gammons, C.H., 1999. Factors controlling copper solubility and chalcopyrite deposition in the Sungun porphyry copper deposit, Iran. Mineral. Deposita 34, 770–783.
Hoseinzade, Z., Bazoobandi, M.H., 2024. Deep embedded clustering: delineating multivariate geochemical anomalies in the Feizabad region. Geochemistry 84 (4), 126208.
Hosseini, M.R., Hassanzadeh, J., Alirezaei, S., Sun, W., Li, C.Y., 2017. Age revision of the Neotethyan arc migration into the southeast Urumieh-Dokhtar belt of Iran: geochemistry and U–Pb zircon geochronology. Lithos 284, 296–309.
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. Inf. Sci. 622, 178–210.
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31 (8), 651–666.
Khan, S., Khan, O., Azam, N., Ullah, I., 2023. Improved spectral clustering using three-way decisions. Inf. Sci. 641, 119113.
Liu, Y., Tu, W., Zhou, S., Liu, X., Song, L., Yang, X., Zhu, E., 2022. Deep graph clustering via dual correlation reduction. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, pp. 7603–7611. No. 7. June.
Rasouli, J., Ghorbani, M., Ahadnejad, V., Poli, G., 2016. Calk-alkaline magmatism of Jebal-e-Barez plutonic complex, SE Iran: implication for subduction-related magmatic arc. Arab. J. Geosci. 9, 1–22.
Ribeiro, B.O.L., Barbuena, D., de Melo, G.H.C., 2023. Geochemical multifractal modeling of soil and stream sediment data applied to gold prospectivity mapping of the Pitangui Greenstone Belt, northwest of Quadrilátero Ferrífero, Brazil. Geochemistry 83 (2), 125951.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Sadigh, S., Mirmohammadi, M., Asghari, O., Porwal, A., 2023. Spatial distribution of porphyry copper deposits in Kerman Belt, Iran. Ore Geol. Rev. 105251.
Safari, S., Ziaii, M., Ghoorchi, M., 2016. Integration of singularity and zonality methods for prospectivity map of blind mineralization. Int. J. Min. Geo-Eng. 50 (2), 189–194.
Setiawan, K.E., Kurniawan, A., Chowanda, A., Suhartono, D., 2023. Clustering models for hospitals in Jakarta using fuzzy c-means and k-means. Proc. Comput. Sci. 216, 356–363.
Shafiei, B., Haschke, M., Shahabpour, J., 2009. Recycling of orogenic arc crust triggers porphyry Cu mineralization in Kerman Cenozoic arc rocks, southeastern Iran. Mineral. Deposita 44, 265–283.
Shahabpour, J., 2007. Island-arc affinity of the Central Iranian volcanic belt. J. Asian Earth Sci. 30 (5–6), 652–665.
Shahrestani, S., Mokhtari, A.R., Fatehi, M., 2020. The use of unmixing technique in stream sediment geochemical exploration. J. Geochem. Explor. 208, 106339.
Shang, R., Kong, J., Wang, L., Zhang, W., Wang, C., Li, Y., Jiao, L., 2023. Unsupervised feature selection via discrete spectral clustering and feature weights. Neurocomputing 517, 106–117.
Shirjang, M., Maghsoudi, A., Ghezelbash, R., 2025. Analysis of clustering methods for geochemical anomaly identification through weighted sample catchment basins. Geochemistry 126337.
Spadoni, M., 2006. Geochemical mapping using a geomorphologic approach based on catchments. J. Geochem. Explor. 90 (3), 183–196.
Spadoni, M., Cavarretta, G., Patera, A., 2004. Cartographic techniques for mapping the geochemical data of stream sediments: the “sample catchment basin” approach. Environ. Geol. 45 (5), 593–599.
Spielman, D.A., 2007. Spectral graph theory and its applications. In: 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07). IEEE, pp. 29–38. October.
Taha, K., 2023. Semi-supervised and un-supervised clustering: a review and experimental evaluation. Inf. Syst. 102178.
Templ, M., Filzmoser, P., Reimann, C., 2008. Cluster analysis applied to regional geochemical data: problems and possibilities. Appl. Geochem. 23 (8), 2198–2213.
Valeh, N., 1972. Geological Map of Iran Sheet7647-Jebal-Barez, Scale 1:100,000. Geological Survey of Iran, Tehran.
Vriend, S.P., van Gaans, P.M., Middelburg, J., De Nijs, A., 1988. The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: examples from Portugal. Appl. Geochem. 3 (2), 213–224.
Wegmann, M., Zipperling, D., Hillenbrand, J., Fleischer, J., 2021. A review of systematic selection of clustering algorithms and their evaluation. arXiv Preprint arXiv: 2106.12792.
Wu, K., Shi, P., Pan, D., 2015. An approach to automatic summarization for Chinese text based on the combination of spectral clustering and LexRank. In: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). IEEE, pp. 1350–1354. August.
Yousefi, M., Carranza, E.J.M., 2015. Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. Comput. Geosci. 74, 97–109.
Yousefi, M., Carranza, E.J.M., Kamkar-Rouhani, A., 2013. Weighted drainage catchment basin mapping of geochemical anomalies using stream sediment data for mineral potential modeling. J. Geochem. Explor. 128, 88–96.
Zhang, Z., Al-Bahrani, M., Ruhani, B., Ghalehsalimi, H.H., Ilghani, N.Z., Maleki, H., Ahmad, N., Nasajpour-Esfahani, N., Toghraie, D., 2023. Optimized ANFIS models based on grid partitioning, subtractive clustering, and fuzzy C-means to precise prediction of thermophysical properties of hybrid nanofluids. Chem. Eng. J. 144362.
Ziaii, M., Carranza, E.J.M., Ziaei, M., 2011. Application of geochemical zonality coefficients in mineral prospectivity mapping. Comput. Geosci. 37 (12), 1935–1945.
Ziaii, M., Safari, S., Timkin, T., Voroshilov, V., Yakich, T., 2019. Identification of geochemical anomalies of the porphyry–Cu deposits using concentration gradient modelling: a case study, Jebal-Barez area, Iran. J. Geochem. Explor. 199, 16–30.