Journal of Geochemical Exploration 258 (2024) 107393

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches

Mahsa Hajihosseinlou [a], Abbas Maghsoudi [a,*], Reza Ghezelbash [b]

[a] Department of Mining Engineering, Amirkabir University of Technology, Tehran, Iran
[b] School of Mining Engineering, College of Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
Sample Catchment Basins
Machine learning
Clustering
DSBCAN
Mean-shift
MVT

ABSTRACT

Cluster analysis can be used to organize samples and generate ideas regarding the multivariate geochemistry of given dataset. Traditional clustering techniques have the drawbacks of high computational complexity and poor adaptability to big data. Hence, there has recently been much focus on creating better clustering algorithms. Although many clustering algorithms have been applied, and some produce notable clustering results, the performance efficiency of algorithms is often highly dependent on the values the user chooses for the parameters. Currently, density-based spatial clustering of applications with noise clustering (DBSCAN) is widely utilized in image processing, bioinformatics, and social network analysis owing to its ability to detect clusters of various shapes. Even though partitional clustering techniques may be effective when the number of clusters K is known in advance, they cannot implement non-convex clustering and rapidly return to a local optimum. This study proposes the concept of DBSCAN clustering for stream sediment geochemical data. In this respect, the geochemical data collected from Varcheh district, SW Iran, were processed using the clr transformation before applying DBSCAN. Then, PCA was used to minimize the dimension of variables and specify the mineralization-related elements. In the following, one of the PCs connected with mineralization (PC2) was chosen for further analysis. DBSCAB, Mean-shift and Fuzzy K-means algorithms were used to monitor the multi-element geochemical anomalies linked to MVT Pb—Zn deposits in the study area. According to Davies-Bouldin and Silhouette as two validation metrics, it can be deduced that the three SCB models are advantageous, however, the model generated by DBSCAN is preferable to the model generated by Mean-shift and Fuzzy K-means.

1. Introduction

In this study, chemical surveys of active stream sediments were used as an exploratory technique (Li et al., 2022; Zhou and Yang, 2023). Since stream sediments result from erosion (Chen et al., 2023) and weathering processes (Jia and Zhou, 2023), they represent the drainage system's original catchment area (Cheng, 2007; Yilmaz et al., 2020; Qiu et al., 2023). In contrast to interpolated maps, catchment-based stream sediment geochemical maps depict anomalies with higher positive spatial connections to known mineral resources in the study region (Carranza, 2010; Nezhad et al., 2017; Ghezelbash et al., 2019a). Classifying geochemical samples that show the influence of various geochemical processes and environments (Li et al., 2023) is beneficial in exploratory geochemistry (Tian et al., 2020). Multiple techniques have been developed to detect and monitor geochemical anomalies (Liu et al., 2018; Ghezelbash et al., 2019a). There are two primary groups into which these algorithms might be classed. The first group is derived from the statistical analysis of geochemical data using a frequency distribution with mean plus two standard deviations (Kürzl, 1988), frequency-based analysis (Ghezelbash et al., 2019c; Thiombane et al., 2019), and fractal-multifractal models (Zuo and Wang, 2016; Ghezelbash and Maghsoudi, 2018; Daviran et al., 2020; Akbari et al., 2023) are several instances. Due to the multi-stage nature of ore-forming mechanisms (Xu et al., 2022) and the inherent complexity of geological structures (Yu et al., 2021; Ren et al., 2022), the statistical distribution of geochemical data is highly sophisticated (He et al., 2021). Therefore, most conventional techniques have drawbacks in analyzing multivariate geochemical data with a complicated distribution (Luo et al., 2022). The second group comprises machine learning algorithms (MLAs) that significantly impact geosciences owing to their impressive pattern recognition and insight discovery capabilities within massive volumes of earth system data (Zhou et al., 2022; Khorshidi et al., 2023; Yin et al., 2023a, 2023b,

* Corresponding author.
E-mail address: a.maghsoudi@aut.ac.ir (A. Maghsoudi).

https://doi.org/10.1016/j.gexplo.2024.107393
Received 21 October 2023; Received in revised form 12 December 2023; Accepted 15 December 2023
Available online 5 January 2024

2023c; Dong et al., 2023b). Algorithms based on machine learning have been applied to map mineral prospectivity (Ghezelbash et al., 2019b; Yao and Jiangnan, 2021; Daviran et al., 2022) and pinpoint geochemical anomalies (Xiong and Zuo, 2016; Zuo and Xiong, 2018; Ghezelbash et al., 2019b; Wang et al., 2020). Supervised and unsupervised machine learning methods have been successfully developed to highlight intrinsic geochemical patterns. Supervised learning algorithms, such as support vector machines (Ghezelbash et al., 2023a, 2023b), random forest (Baudron et al., 2013; Wang et al., 2019; Daviran et al., 2021), artificial neural networks (Ghezelbash et al., 2020a, 2020b; Guérillot and Bruyelle, 2020), and LightGBM (Hajihosseinlou et al., 2023) provide computers with the ability to categorize objects, problems, or situations based on human-labeled data (Xie et al., 2021; Yin et al., 2023a, 2023b, 2023c; Dong et al., 2023a). Conversely, unsupervised machine learning algorithms such as K-means (Chen et al., 2017), restricted boltzmann machines (Aryafar and Moeini, 2017), isolation forest (Zhang et al., 2022), and one-class SVM (Xiong and Zuo, 2020) have been applied to isolate geochemical communities and group data into similar groups. Separating geochemical anomaly populations from background often requires using clustering techniques, which are unsupervised approaches (Grunsky, 2010; Ghezelbash et al., 2023b). However, the structure of the data, the type of analysis to perform, and the size of the dataset are all crucial in selecting clustering algorithms (Prades, 2018; Yin et al., 2023a, 2023b, 2023c). Various clustering techniques are available, including Partitioning Clustering (Barioni et al., 2014), Density-Based Clustering (Kriegel et al., 2011), Distribution Model-Based Clustering (Kriegel et al., 2005), Hierarchical Clustering (Nielsen, 2016), and Fuzzy Clustering (Yang, 1993).

This is crucial during exploratory and assessment data analysis when researchers seek to uncover hidden characteristics without prior knowledge (Daviran et al., 2024; Hajihosseinlou et al., 2024). Many scholars have used the concepts of internal homogeneity and external separation to characterize a cluster (Hancer and Karaboga, 2017). This means that patterns within the same cluster should exhibit resemblance to one another, whereas patterns in distinct clusters should differ from each other. (Michaud, 1997; Wu et al., 2022). Although many different approaches have been suggested to clustering, most current algorithms require at least some prior knowledge of the data to be grouped. Therefore, their effectiveness is often highly dependent on user-specified input. Common methods, such as K-means, require a specified number of clusters as input (Kodinariya and Makwana, 2013), which is sometimes difficult to estimate. This study used the DBSCAN clustering technique to find a solution to this challenge. Some significant benefits over existing clustering techniques are demonstrated by density-based spatial clustering of applications with noise (DBSCAN) (Kriegel et al., 2011). Initially, it requires no predetermined number of clusters, and it recognizes outliers as noise, unlike other algorithms that bundle them together regardless of their dissimilarity. In contrast to DBSCAN, K-means (a popular clustering method) typically only uses clusters with a spherical form. Moreover, Mean-shift is another example of a density-based technique.

This case study uses the Varcheh region of western Iran to showcase the application of DBSCAN and Mean-shift to identify Pb, Zn and Ba geochemical anomalies of stream sediment samples connected to sample catchment basins (SCBs) of Pb, Zn and Ba in the Varcheh area in western Iran. This study aimed to describe and compare the application of the DBSCAN and Mean-shift algorithm and their effectiveness in detecting geochemical anomalies. Also, Fuzzy k-means, a commonly used variant of the k-means algorithm, was applied to validate the performance of both Mean-shift and DBSCAN methods.

Furthermore, this paper envisages how the proposed methods could be included in an exploration information system to facilitate the interpretation of geochemical data and to generate the ensuing evidence layers for mineral exploration targeting. While we achieved favorable results in applying DBSCAN, it is imperative to note that this approach also has specific limitations. The effectiveness of DBSCAN is impacted by the selection of parameters, specifically the epsilon (ε) and minimum points (MinPts). Determining the most appropriate values for these factors relies on the specific attributes of the data, and an incorrect selection might impact the clustering outcomes. Moreover, as the dimension of the data increases, DBSCAN's efficacy generally diminishes. High-dimensional spaces frequently encounter the curse of dimensionality (Braune et al., 2015), which restricts the algorithm's potential to establish significant clusters. Also, DBSCAN can classify outliers as noise or create small groups around them, particularly in datasets with different density levels. The susceptibility to noise may impact the overall clustering quality. Therefore, when using DBSCAN, it is crucial to understand certain constraints and factors that need careful consideration.

2. Methodology

Clustering is often defined as discovering patterns in datasets by grouping related items, and the generated groups are known as clusters (Tian et al., 2019). In order to be regarded as an effective technique, a clustering method must meet the criteria listed below: (a) parameters may be adjusted with limited domain expertise (Zheng et al., 2023), which is extremely helpful in dealing with big datasets (Luo et al., 2022), (b) identifying clusters with arbitrary shapes, and (c) optimal performance with big datasets (Ghezelbash et al., 2023a). The clustering procedure is shown in Fig. 1.

2.1. Density-based clustering

Clustering based on density (Campello et al., 2020) has been extensively studied. The basic concept behind density-based clustering is to build a structure for a given collection of data points that precisely represents the underlying density.

Density-based clustering differs from parametric cluster analysis methods such as Gaussian mixture models (GMMs) (e.g., McLachlan and Basford, 1988) in that the latter assume that the observed data are constructed by a combination of parametric distributions (generally considered to be Gaussian). Parametric methods are beneficial in several cases, but they make the unwarranted assumption that clusters have a convex (hyper-spherical or hyper-elliptical) form. K-means clustering (where k is the user-specified cluster number) is another algorithm that follows this pattern to produce convex cluster shapes. This method assumes that excellent clusters can be discovered by decreasing intra-cluster variance (also called cluster cohesion) and enhancing inter-cluster variance (Fahim et al., 2008). On the contrary hand, density-based clustering techniques do not rely on variance or parametric distributions. Therefore, they can identify clusters of any form, are robust against different types of noise and do not need expert knowledge to identify the ideal quantity of clusters (Nagpal and Mann, 2011; Bhuyan and Borah, 2013).

2.1.1. The theory behind the DBSCAN approach

Density-based spatial clustering of applications with noise (DBSCAN) is a commonly used clustering method based on the density concept introduced by Ester et al. (1996), which was developed to cluster multi-dimensional datasets, including both spatial and non-spatial data into clusters of any form when such datasets are exposed to noise. DBSCAN's primary goal is to cluster data points if their neighborhood of a defined radius (*Eps*) includes at least a minimum number of other data points (*MinPts*). In other words, the neighborhood's cardinality must be higher than the threshold. A random data point 'A' has an Eps-neighborhood, calculated as:

$$N_{Eps} = \{A \in D / distance(A, B) < Eps\} \quad (1)$$

Here, D represents a dataset of objects. If A's -neighborhoods include at least *MinPts*number of points, Athen is considered a core point.

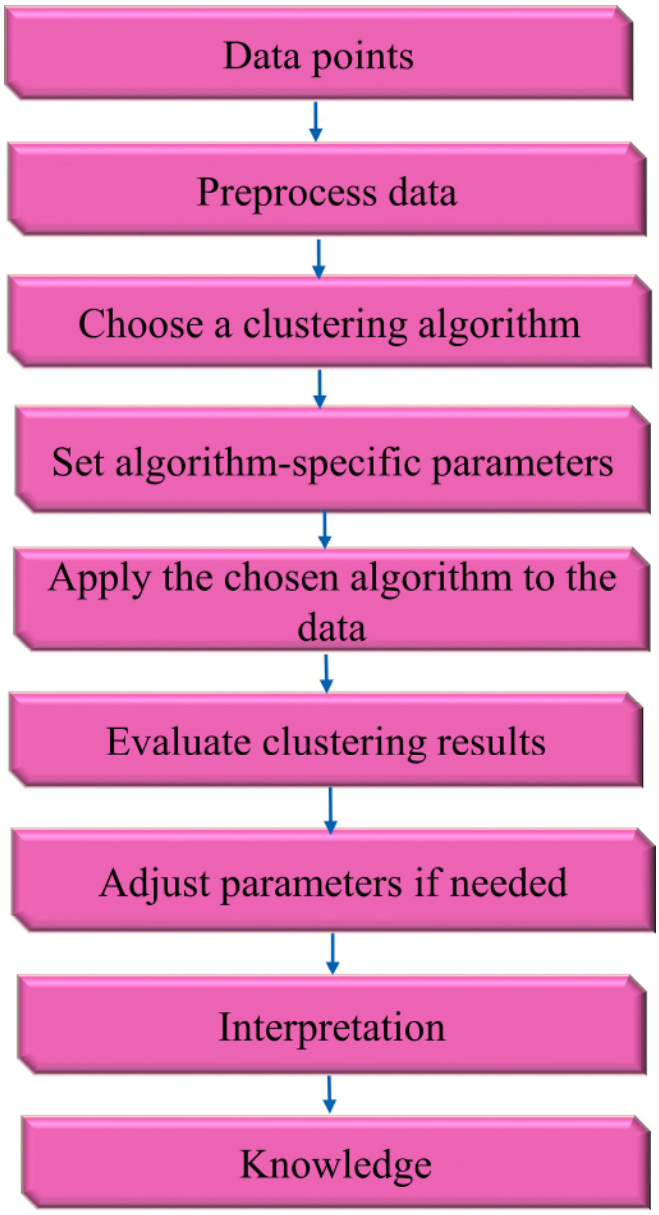

Fig. 1. The procedure for conducting cluster analysis.

$$N_{Eps}(A) > MinPts \tag{2}$$

Here, the parameters *Eps*and *MinPts*denote the radius of the neighborhood and minimal number of points in the Eps-neighborhood of a core point, respectively. Unless this requirement is met, the given point is non-core point.

2.1.2. Mathematical foundation of Mean-shift method

Mean-shift is a non-parametric clustering algorithm based on the concepts provided by Fukunaga and Hostetler (1975). It comprises mainly the three phases listed below. The first task is to randomly select a point p_i, $i \in \{1, 2, \ldots, n\}$ from the unlabeled data item $P = \{p_1, p_2, \ldots, p_n\}$ to serve as the center point y_j. The second is to locate all the data points inside the circle specified by y_j and $\frac{\varepsilon}{2}$ as the radius. It can be stated as:

$$\delta_\varepsilon(y_j) = \left\{p_i \| y_j - p_i\|_2^2 < \frac{\varepsilon}{2}\right\}, j = 1 \tag{3}$$

Determine the vector from the center point $y_j$to every item $p_i$in the set δ, assuming these points correspond to the cluster C_j; then sum these vectors to get the mean shift vector $MS\left(y_j\right)$ defined as:

$$MS(y_j) = \frac{\sum_{i=1}^{m} p_i G\left(\left\|\frac{y_j - p_i}{\varepsilon}\right\|^2\right)}{\sum_{i=1}^{m} G\left(\left\|\frac{y_j - p_i}{\varepsilon}\right\|^2\right)} - y_j \tag{4}$$

$p_i \in \delta_\varepsilon\left(y_j\right)$ and the Gaussian kernel function $G()$ are created to apply various weights to each data point. $\left|MS\left(y_j\right)\right|$ shifts the center point in the direction of the Mean-shift vector $MS\left(y_j\right)$ .Then, continue the preceding steps by $y_{j+1} = MS\left(y_j\right) + y_j (j = 1, 2, \ldots)$ until the iterative cycle $\left\{y_j\right\}_{j\geq 1}$ meets the condition $\left\|y_{j+1} - y_j\right\|^2 < \varepsilon$, where ε is a user-defined threshold value. Finally, repeat the previous two procedures until all points have been categorized. Fig. 2 depicts Mean-shift flowchart.

2.2. The mathematical basis of the Fuzzy k-means

One type of unsupervised machine learning is fuzzy clustering analysis, which uses the fuzzy theory to determine the level of uncertainty associated with each sample category (Tokushige et al., 2007). Fuzzy K-means is a modified version of the conventional K-means clustering technique that incorporates a degree of fuzziness or softness in assigning data points to clusters. The clustering algorithm calculates the degree of membership of each sample point to all cluster centers through boosting the objective function and identifying the optimal cluster center via various iterations. This process determines the category of the sample points and achieves the goal of classifying the sample data. The following is the usual expression of the objective function, which is minimized during the optimization phase of Fuzzy K-means:

$$Obj = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^{m} \|x_i - z_j\|^2 \tag{5}$$

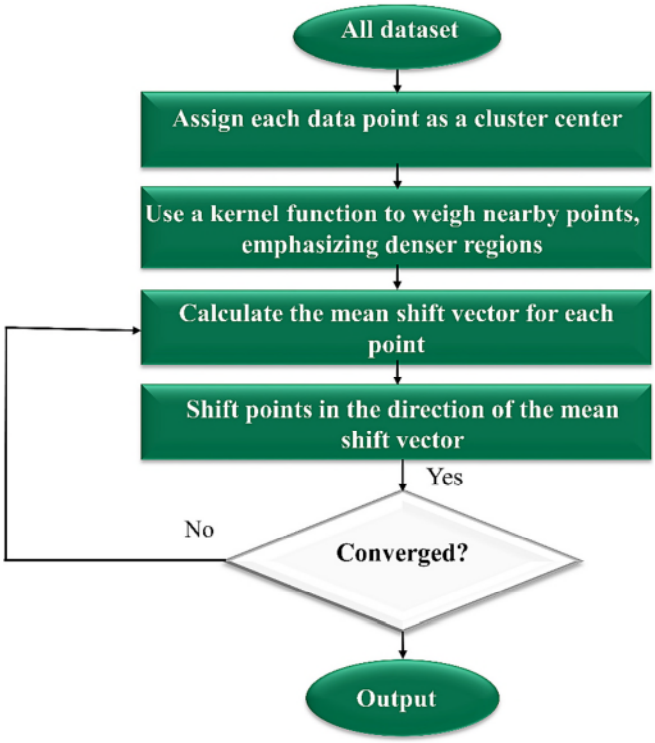

Fig. 2. Flowchart of Mean-shift framework.

u_{ij}represents the degree of membership of data point *i*in cluster *j*.The weighting exponent *w*is often set to 2 for Fuzzy k-Means. Lastly, z_j refers to the centroid of cluster *j*.

2.3. Davies-Bouldin index

David L. *Davies* and Donald W. Bouldin proposed the Davies-Bouldin index (DBI) in 1979 to assess various clustering algorithms. The score is computed as the ratio of distances inside clusters to the distances between clusters (Vergani and Binaghi, 2018). The score is described as the mean similarity of each cluster, C_i, and the one most similar to it, C_j. The similarity is specified for this index as a metric S_{ij} that compromises:

r_i, cluster diameter, the average distance between every cluster point and its centroid.

d_{ij}represents the distance between cluster centers *i*and. *j*.

a straightforward solution for constructing S_{ij} in a way that ensures it is nonnegative and symmetric, consider:

$$S_{ij} = \frac{r_i + r_j}{d_{ij}} \tag{6}$$

The Davies-Bouldin index is defined as follows.

$$DB = \frac{1}{n}\sum_{i=1}^{n} max S_{ij} \tag{7}$$

2.4. Silhouette index (SI)

The Silhouette is applied to assess and evaluate the distance between the generated clusters. This technique determines the proximity of each item in one cluster to those in another (Dudek, 2019). The Silhouette index (SI), in contrast to other performance assessment approaches, the evaluation of clustering results doesn't necessitate the use of a training set. The definition for the silhouette width $S(x_i)$at the location $A(x_i)$ is:

$$S(x_i) = \frac{B(x_i) - A(x_i)}{max\{B(x_i), A(x_i)\}} \tag{8}$$

where x_i is an object of cluster $.C_m$.

$A(x_i)$ represents the average distance between the element $x_i$and all other items in cluster C_m,and

$$B(x_i) = min\{D_r(x_i)\}, r \neq m \tag{9}$$

where $D_r(x_i)$ is the average distance between point $x_i$and all other points in cluster $C_r$for $r \neq j$.

According to Eq. (8), silhouette width might range between −1 and 1.

A negative number indicates that $A(x_i) > B(x_i)$, which is unfavorable; it implies that the degree of dissimilarity inside a given cluster is higher than between clusters. Where $A(x_i) < B(x_i)$, a positive value is generated, and the maximum width of the Silhouette is achieved at $S(x_i) = 1$ when $A(x_i) = 0$. A stronger positive $S(x_i)$value indicates that an element is more likely to be found in the appropriate cluster. Negative $S(x_i)$ elements more potentially be grouped in incorrect clusters (Yuan and Yang, 2019).

3. Study area and dataset construction

3.1. Geological setting

Varcheh district is located northwest of Iran's central province and is considered a part of the Sanandaj-Sirjan zone (SSZ) (Fig. 3). The SSZ, with a length of about 1500 km and a width of 150 to 250 km, was

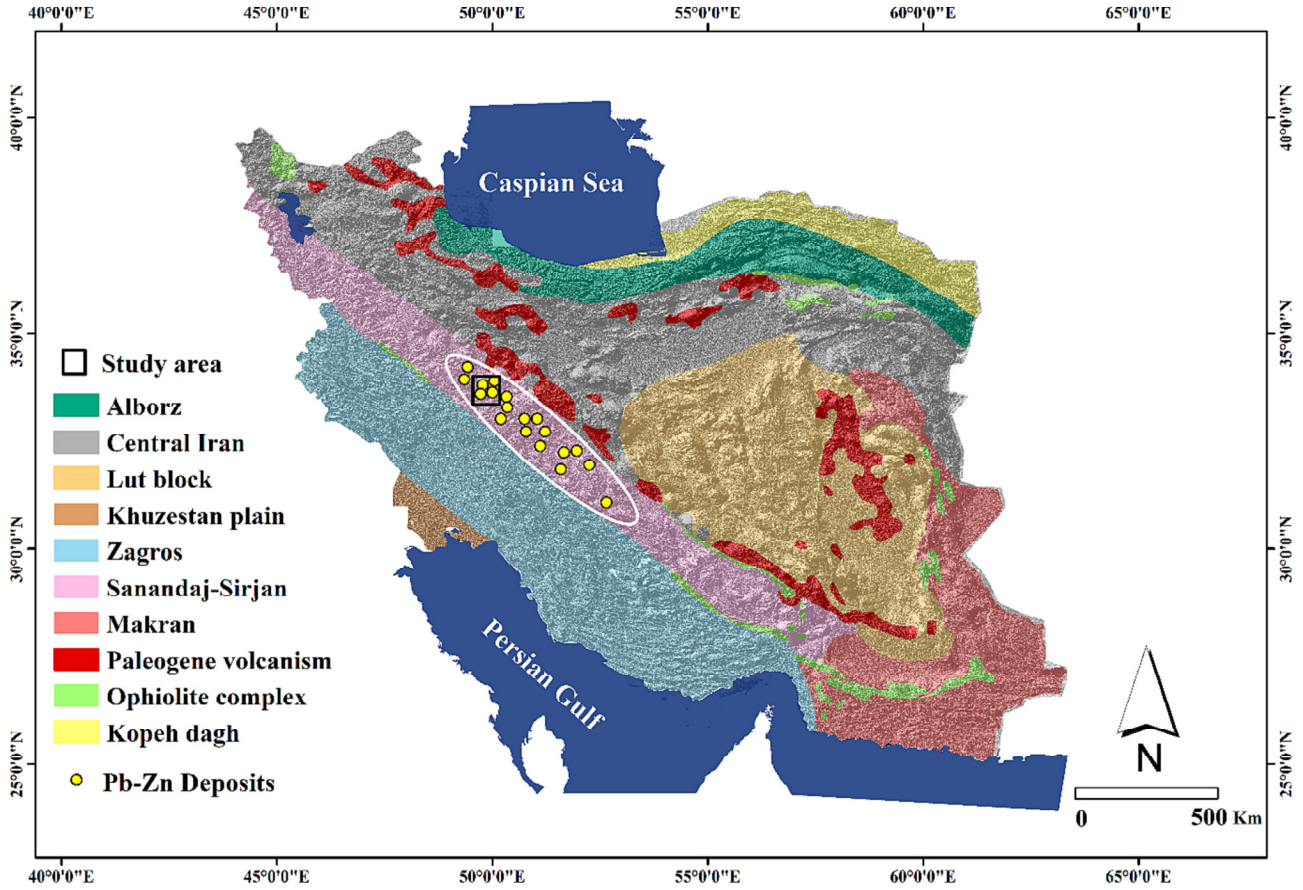

Fig. 3. The location of the study area and the position of the Malayer-Esfahan metallogenic belt on structural zones of Iran (Stocklin, 1968).

formed by the convergence of the African-Arabian and Eurasian plates (Ghasemi and Talbot, 2006; Ghalamghash et al., 2009; Ghazi and Moazzen, 2015). The most critical metamorphic processes in the SSZ are linked to tectonic movements that proceeded during the opening and closing of the Neotethys ocean.

The Malayer-Esfahan metallogenic belt, in which the shallow marine sediments of the Cretaceous sea are spread, contains many volcano-sedimentary Pb—Zn deposits is situated within the central section of the SSZ (Rajabi et al., 2019). It is generally acknowledged that sedimentary exhalative processes or carbonate-hosted, Mississippi Valley-type genesis are responsible for this Pb—Zn mineralization (Zarasvandi et al., 2014). The intrusion has reached the earth's surface in the form of small and scattered outcrops of Eocene among the limestone schists belonging to the Lower Cretaceous. Differences in morphology can be observed throughout the study area due to the presence of rocks with a wide range of diverse sources. Regional tectonics and volcanic activity have also significantly formed these areas (Ehya et al., 2010). The lithological composition of the intrusion is gabbro and monzogabbro, along with plagioclase, clinopyroxene, and opac minerals (ilmenite type) with ophitic and subophitic textures. Plagioclase is the most significant and common mineral of these rocks, accounting for around 12 % of the total mineral content. Also, the second most abundant mineral in these rocks is clinopyroxene. Additionally, the main alterations are sericitic, oralitization, and chloritization.

Geochemical researches reveal that Varcheh gabbro has special petrological features of alkaline rocks. These rocks are mantle-derived garnet peridotites developed at 100 to 105 km. Furthermore, the investigation of magmatic features indicates the insignificant role of crustal assimilation in extensional tectonics.

Emarat deposit, with proven reserves of 12.5 million tons and an average grade of 5 % zinc and 2 % lead, is considered to be the largest and most crucial zinc and lead deposit in the central portion of the Malayer-Esfahan belt. Also, the Moochan deposit, with an initial reserve of 300,000 tons and an average grade of 7.12 % zinc and 1.74 % lead, is located 2 km southeast of the Emarat deposit. Sphalerite is the most abundant sulphide mineral in both Emarat and Moochan deposits (Ehya et al., 2010), which is observed as 0.02 to 3 mm deposited in open space filled and veins. Based on these observations, sphalerite was formed in two stages. First-stage sphalerites have a high concentration of iron. Second-stage sphalerites are poor in iron. Considering the location of Emarat and Moochan deposits in the SSZ, the formation of these deposits can be attributed to tectonic events related to the convergence of the Arabian and Iranian plates and the closure of the Neotethys ocean (Ghasemi and Talbot, 2006). Since the construction of the other Pb—Zn deposits in the SSZ and the activity of the subduction event coincided, considering its role in the metallogenic of Pb—Zn deposits seems essential. This event, along with the compression of various stratified units, including Jurassic shale and sandstones, has caused the warming and dynamism of fossil waters and the release of carbon dioxide, silica, and metals during the diagenesis of clay minerals and pyroclastic rocks. The investigation of the Pb isotope in the Emarat deposit is consistent with the crustal origin of the metals (Fernández et al., 2000; Karimpour et al., 2017; Karimpour and Sadeghi, 2018).

3.2. Geochemical data

Stream sediment samples mainly include pyroclastic. Therefore, it is essential to sieve the sediments and collect the components of the correct size (Li et al., 2020). After transferring these samples to the laboratory, the preparation steps of the samples were carried out, including drying, removal of organic matter, and powdering. This research applied analytical data for ten elements (Ag, Ba, Co, Cr, Cu, Ni, Pb, Sr, V, and Zn) obtained from 1283 stream sediment samples representing catchment basins (Fig. 4a) that cover the total area. The samples were gathered by the Geological Survey of Iran (GSI) in the 1: 100,000 scale Varcheh map (Fig. 4b). Inductively coupled plasma optical emission spectrometry (ICP-OES) was performed to get these analytical results. According to the Howarth and Thompson (1976), using validated duplicate samples, the analytical precision for all elements was more than 10 %.

4. Results

4.1. Data preprocessing

Data mining and analysis need preprocessing before modeling (Zhang et al., 2022). The data relating to the elements were subjected to statistical analysis (Vahid et al., 2021), and various descriptive statistical measures were obtained and presented in Table 1. Generally, all geochemical datasets are compositional and considered closed number systems with a fixed sum per sample. Compositional data are often relative portions of a whole (Greenacre, 2021).

Compositional variables are dependent, and no variable may fluctuate independently of the others (Filzmoser et al., 2009). Also, the data closure issue must be considered even if there is only one component since the value of a chosen element indicates its proportions in the whole sample (Ghezelbash et al., 2019d). Consequently, geochemical data should be opened prior to analysis (Aitchison et al., 2005). A clr transformation was used in this research. For a D-part composition x, the clr transformation is described as:

$$clr(x) = \left[log\frac{x_1}{G(x)}...log\frac{x_D}{G(x)}\right], G(x) = \left(\prod_{i=1}^{D} x_i\right)^{\frac{1}{D}} \tag{10}$$

$G(x)$ is the geometric mean of the variable x. This function's inverse is known as the softmax function.

The ilr transformation depends on selecting a specific orthonormal based on the hyperplane in R^D produced by the clr transformation. The equation for the ilr transformation is:

$$ilr(x) = \sqrt{\frac{i}{i+1}}.log\frac{\sqrt[i]{\prod_{j=1}^{i} x_j}}{x_i + 1}, i = 1, 2, \ldots, D-1 \tag{11}$$

Since the new D - 1 variables are unrelated to the original variables, the data produced by the ilr transformation cannot be sensibly evaluated from an exploration viewpoint. This issue can be addressed by orthonormal back transforming the ilr coordinates to clr coefficients.

$$w_i = \left(\frac{i}{i+1}\right)^2.\left[\underbrace{\frac{1}{i}, \ldots \frac{1}{i}}_{\text{i elements}}, -1, 0, \ldots, 0\right], i = 1, 2, \ldots, D-1 \tag{12}$$

$W$is a (D - 1) D matrix with $w_i$representing row vectors in this case. Clr and ilr are related in the following way:

$$clr(x) = ilr(x).W, W = (w_1, w_2, \ldots, w_{D-1}) \tag{13}$$

It is also possible to generate a robust covariance matrix for the clr coefficients by:

$$\sum = W.\sum(K).W^T, K = (k_1, k_2, \ldots, k_{D-1}) \tag{14}$$

The Fig. 5a and b depict raw and transformed data. Based on the SCB method, we have created geochemical maps that display the original (Fig. 6) and clr-transformed (Fig. 7) values of three elements associated with Pb—Zn mineralization in the Varcheh district.

4.2. DBSCAN-based intelligent geochemical modeling

In the area of exploratory geochemistry, identifying geochemical associations is necessary for detecting geochemical patterns and mapping anomalies. In exploratory geochemistry, identifying geochemical associations is necessary for detecting geochemical patterns and mapping anomalies. In high-dimensional data, principal component analysis

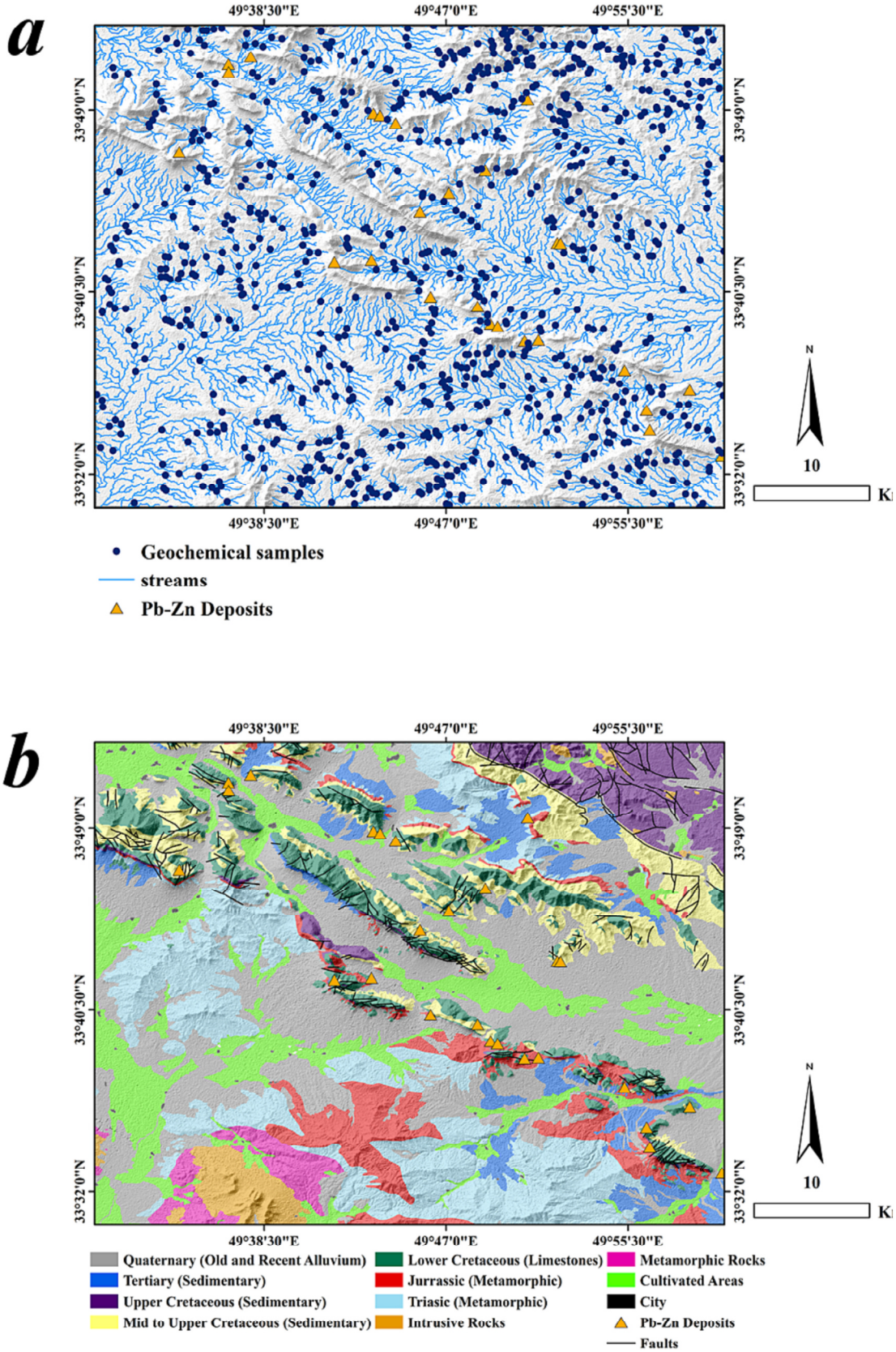

Fig. 4. Maps of (a) location of stream sediment samples and (b) geological units of Varcheh district.

Table 1
Description statistics of the stream sediment geochemical data.

Characteristic	Pb value	Zn value	Ba value
Mean	58.88	124.81	310.84
Median	30	91	290
Standard deviation	119.75	103.79	134.87
Maximum	1333	906	1333
Minimum	2.75	2.75	80
Skewness	7.02	2.69	2.47
Kurtosis	61	10.37	11.96

(PCA) is a tool for retaining the data set with the most variation while still providing dimension reduction. It reduces the number of dimensions and the data to be compressed by identifying general characteristics in multidimensional data (Maćkiewicz and Ratajczak, 1993). This research applied PCA on log-ratio transformed data to explore patterns in element relationships and trace material sources in Pb—Zn deposits generated by geochemical processes.

A graph known as the scree is used to estimate the number of principle components; it illustrates the relationship between the number of primary components and their eigenvalues. The eigenvalues on a screen plot are always shown in descending order, from greatest to smallest. Scree plots often have the same general form, rising sharply on the left, declining steeply, and then leveling out. Specifically, since the first component usually accounts for a significant part of the variance, the following factors explain a moderate proportion. Meanwhile, the latter components only account for a minor portion of the total variance. The scree plot criteria find the curve's "elbow" and choose all components before the line flattens. In this paper, the scree plot of eigenvalues (Fig. 8) isolated the four components responsible for 74.56 % of the overall variance. The first component explains the most variance, whereas the subsequent ones explain increasingly less variance (Table 2).

A matrix of size 1283 by 3 was created, which consisted of the clr-transformed values of elements associated with Pb—Zn mineralization. DBSCAN, an algorithm used for creating a geochemical anomaly map via SCBs, was fed this matrix as input. When the DBSCAN method is first run, it sets the cluster identification number to 0 and picks a random starting point that has not been allocated to a cluster or deemed noise. Then, computes the number of all the neighboring points within Eps distance of the primary selecting point. If this number exceeds MinPts, a cluster will construct around this point. Otherwise, this point will be identified as an outlier. The starting point and its neighbors will form a cluster labeled as the visited point. The DBSCAN algorithm iteratively continues this process for all neighbors until a point has fewer neighbors than MinPts. Therefore, this point is categorized as noise. After the cluster forms, its identification number will increase; DBSCAN picks a new point from the pool of unvisited data points to begin forming a new cluster. This process will be continued until every data point is assigned to a cluster or dismissed as noise. The basic premise of the DBSCAN clustering method is shown in Fig. 9.

MinPts should typically be larger than or equal to the number of dimensions in the dataset. So, MinPts = 2*dim is usually chosen, in which dim is equal to the dimensions of the data set if the data set contains over two dimensions (Sander et al., 1998). Once MinPts is selected, the value of ε should be found. This article describes an approach to automatically determining the best value. Each point's average distance to its k nearest neighbors was determined using this method, where k was whatever value chosen for MinPts. The average k-distances were then shown in increasing order on a graph depicting k-distances. The location of maximal curvature provided the best estimate for the value of ε. Ultimately, optimal results for DBSCAN were obtained by setting MinPts = 6 for the lowest number of points needed to create a cluster and Eps = 90 (Fig. 10) for the maximum distance allowed in the neighborhood between cluster points. Consequently, the DBSCAN-based SCB map (Fig. 11a) displaying several geochemical categories (cluster 1: background; cluster 2: weak anomaly; and cluster 3: strong anomaly) was generated.

4.3. Mean-shift-based intelligent geochemical modeling

In order to find the most reliable geochemical targets relevant to carbonate-hosted Pb—Zn deposits, the Mean-shift clustering technique was also used on the matrix created during the previous subsection. Mean-shift is an unsupervised clustering approach that seeks to identify blobs within a smooth sample density. The technique finds the average of points inside an area through updating candidates for centroids (generally known as bandwidth). After this step, the candidates undergo further processing to exclude near-duplicates and construct the ultimate collection of centroids. Therefore, unlike K-means, they are not required to manually choose the number of clusters. The Mean-shift algorithm depends highly on the kernel bandwidth. The density gradient has the most significant effect on the efficiency of the Mean-shift method. The Mean-shift algorithm may provide visually pleasing clustering if the bandwidth matrix is correctly selected to yield an acceptable kernel density gradient estimate. The final clusters seem different depending on the available bandwidth. In the current study, we first picked a minimal bandwidth. As a consequence, each point developed its distinct group. Conversely, we used a large bandwidth, producing single cluster including all the data. A manual approach to selecting the appropriate bandwidth for small, two-dimensional data sets could be feasible, but it will become more challenging as the data set grows. So rather than manually choosing the bandwidth, we used the estimate bandwidth function, a method provided by the Python sklearn package that employs a nearest-neighbor analysis. Lastly, geochemical samples of stream sediments associated with anomaly classes (clusters 3) were allocated to their relevant SCBs, and a mean-shift based geochemical anomaly map was generated (Fig. 11b).

4.4. Fuzzy K-means-based intelligent geochemical modeling

The matrix 1283 by 3, containing clr-transformed values of elements related to Pb—Zn mineralization generated in previous sections, served as the input for Fuzzy K-means to identify the most dependable geochemical targets associated with carbonate-hosted Pb—Zn deposits. Tuning the parameters of Fuzzy K-means involves a thoughtful approach to achieve optimal clustering results. The primary parameters to adjust are the number of clusters (K) and the fuzziness parameter (m). We expressly set the number of clusters to 3. We considered an equal number of clusters for Fuzzy K-means in alignment with the cluster counts specified for DBSCAN and Mean-shift, specifically three clusters each. This approach facilitates a straightforward comparison of the outcomes generated by all three models.

Additionally, fine-tuning the fuzziness parameter (m) to balance the degree of fuzziness in the memberships is essential. The number commonly used for the fuzziness parameter (m) in Fuzzy K-Means typically falls within the range of 1.5 to 2. We experimented with different values of m during parameter tuning to find the most suitable setting for a given dataset and clustering objective. Finally, we determined a fuzziness level of 2 for the memberships.

In this way, the model was constructed, and samples were labeled. In conclusion, stream sediment geochemical samples were linked to their corresponding SCBs, creating a geochemical anomaly map using the Fuzzy K-means method. Fig. 11c depicts the clustering result derived by Fuzzy K-means.

4.5. Evaluation

Cluster analysis depends on assessing clustering outcomes to determine the partition most accurately which represents the underlying data (Cheng et al., 2023). In the current study, the effectiveness of clustering performed for extracting geochemical anomaly classes has been

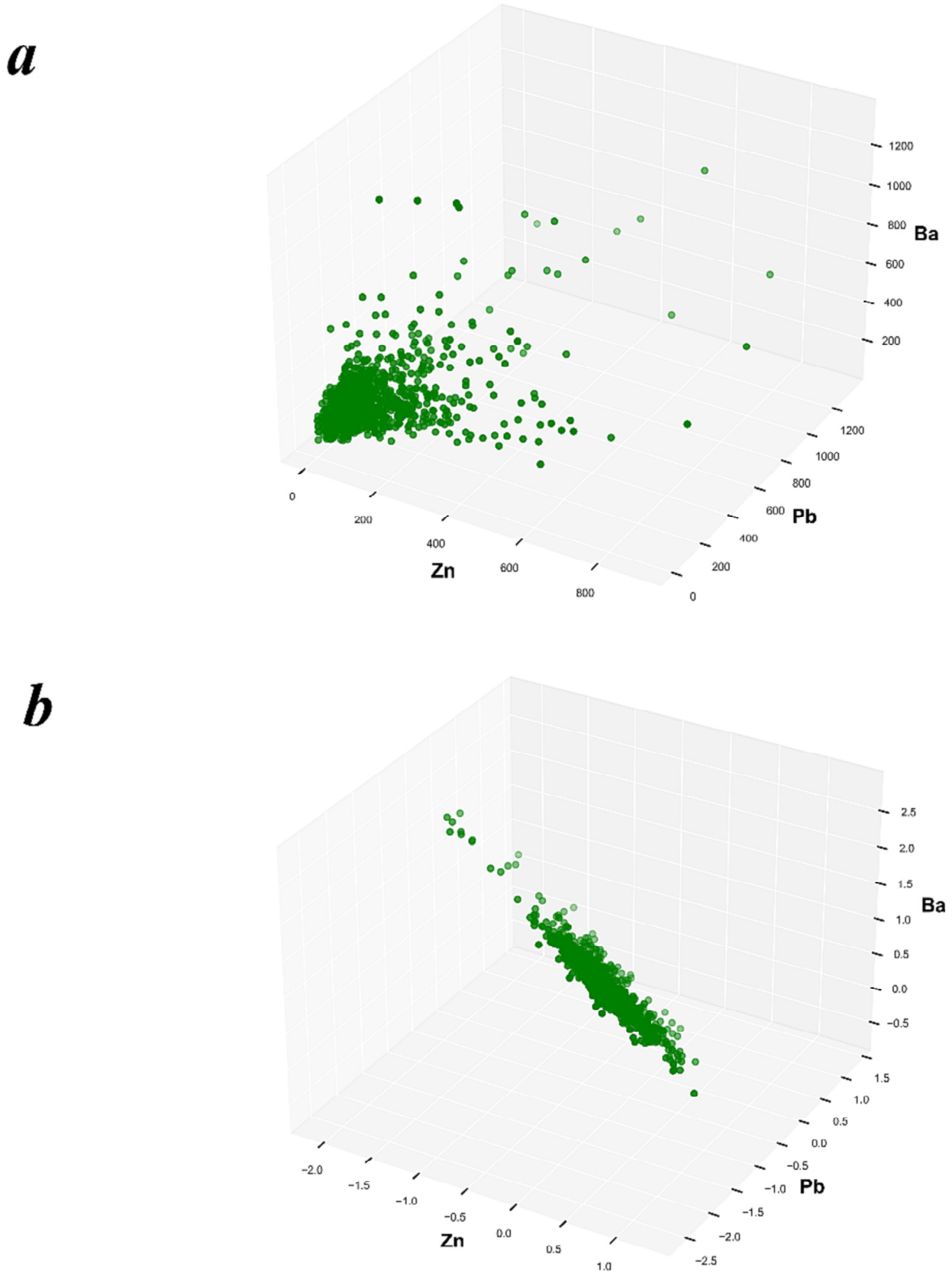

Fig. 5. The sample dataset (a) before clr transformation and (b) after clr transformation.

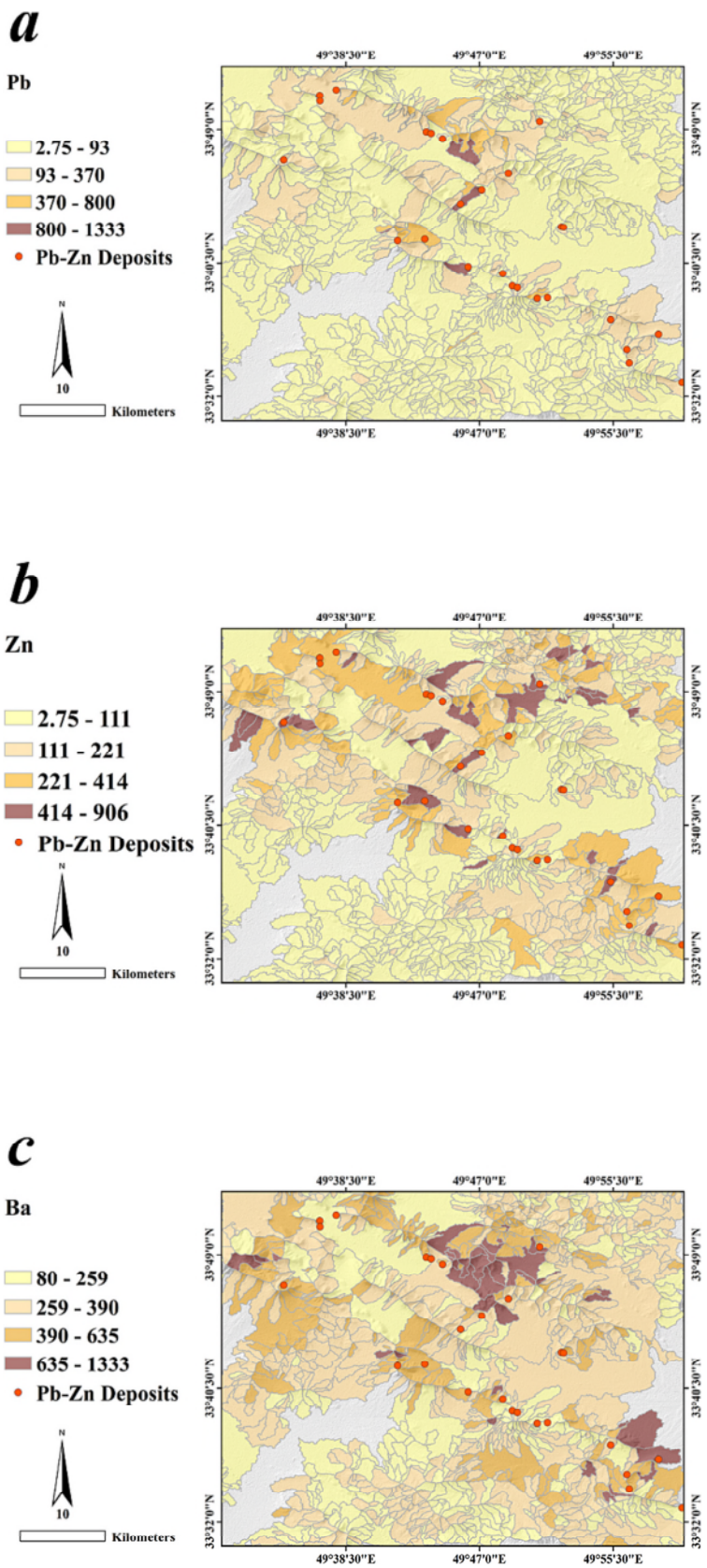

Fig. 6. SCB-based geochemical maps original values of three mineralization-related elements.

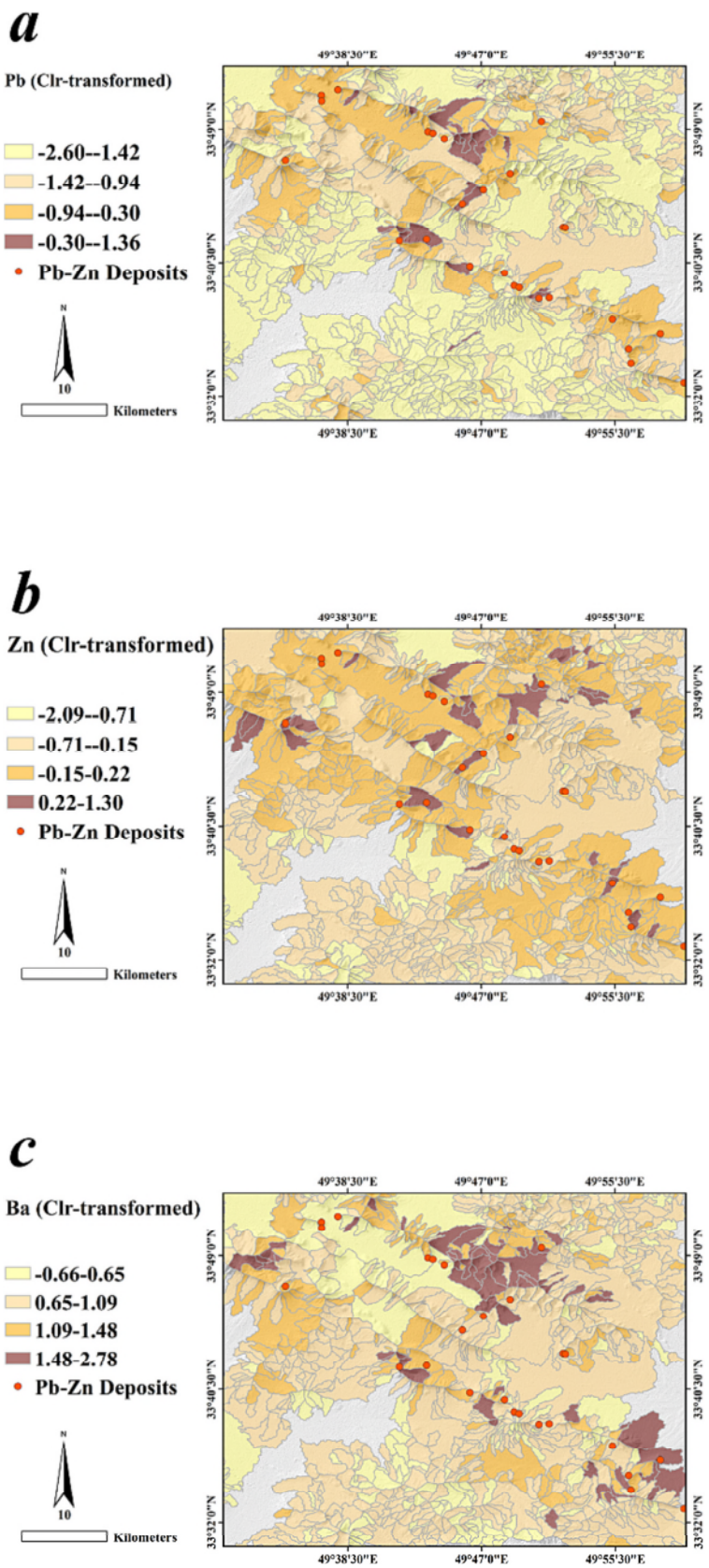

Fig. 7. SCB-based geochemical maps of clr-transformed values of three mineralization-related elements.

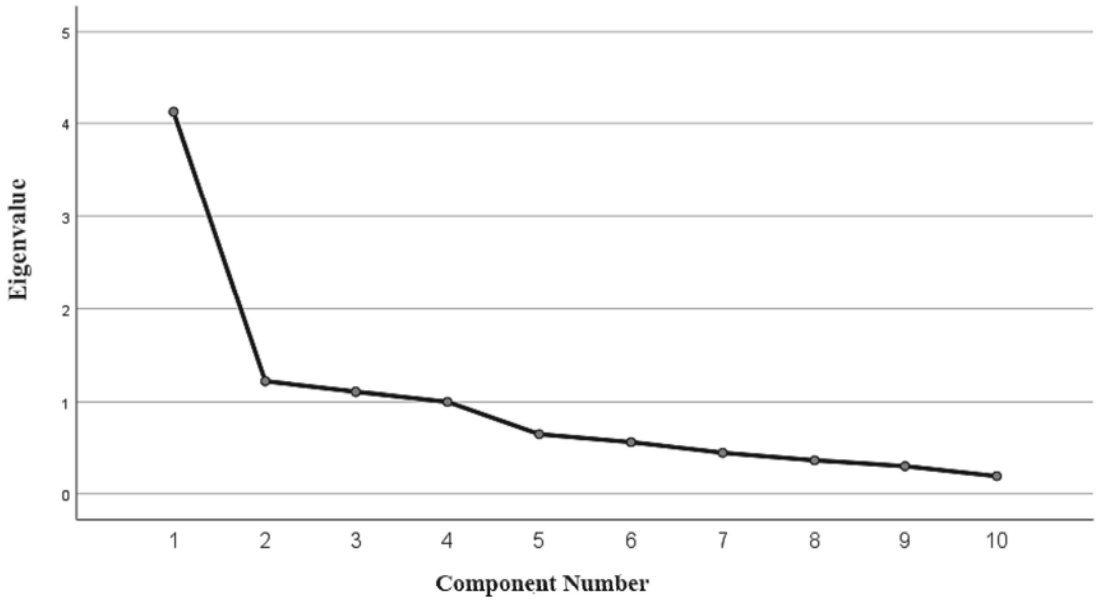

Fig. 8. Scree plot representing the eigenvalues accounted by PCA.

Table 2
Rotated componenet matrix of PCA.

Elements	PC1	PC2	PC3	PC4
Ag	0.012	−0.016	0.32	**0.984**
Ba	0.6	**0.51**	0.164	0.067
Co	**0.821**	0.024	0.366	0.01
Cr	**0.791**	0.250	−0.034	0.059
Cu	**0.662**	0.117	−0.084	0.192
Ni	**0.754**	0.42	−0.184	−0.18
Pb	0.47	**0.936**	−0.064	0.017
Sr	−0.007	0.044	**0.944**	−0.141
V	**0.827**	0.109	0.284	0.036
Zn	0.449	**0.606**	0.393	−0.056
Var. (%)	41.227	12.214	11.1	−0.043
Cum.Var. (%)	41.227	53.441	65.541	74.556

measured using the Davies-Boulding index (Wang et al., 2022).

Considering that a lower index can achieve a better clustering outcome and zero is the lowest score, the result demonstrates that the DBSCAN algorithm has a low value of DB score compared to the Mean-shift and Fuzzy k-means (Table 3). That means since the index has been minimized in DBSCAN, the most distinct geochemical anomaly clusters and the optimal partition have been portrayed by DBSCAN.

Similarly, the Silhouette is applied to assess and evaluate the distance between the generated clusters. This technique determines the proximity of each item in one cluster to those in another (Dudek, 2019).

The index has been employed on DBSCAN, Mean-shift, and Fuzzy K-means algorithms. For all models, the index value was positive and acceptable. A higher DBSCAN Silhouette score has implied (Table 3) that the geochemical samples were more appropriately clustered compared to those clustered using Mean-shift.

5. Discussion

Several variables influence the regional distribution of geochemical elements. Therefore, conducting an analysis and identification of prominent geochemical indicators that correspond to the desired deposit type within a specific research region is crucial to facilitate further exploration efforts. In this context, the paper has used multi-element connections of Pb-Zn-Ba as indicators for the identification and exploration of MVT Pb—Zn anomalies. Because they are essential geochemical traces for mineralization and have a promising spatial connection with the deposit type sought, it should be noted that since geochemical datasets often consist of several components and the closure effect is an intrinsic property of geochemical data, it is essential to highlight that we employed clr transformation to open the data before performing any analysis. Fig. 5 clearly illustrates how closed data were opened. Besides, we demonstrated the successful implementation of density-based clustering approaches for precisely defining the geochemical anomalies associated with Pb—Zn mineralization systems. The DBSCAN model was used in our study to identify multivariate geochemical anomaly patterns associated with SCBs in the Varcheh District, situated in the Malayer-Esfahan metallogenic belt, that has significant importance in the exploration of MVT deposits inside Iran. This article shows that SCB-based models are more efficient for mapping geochemical anomalies. Additionally, we proposed the Mean-shift clustering model as an alternative approach. DBSCAN was sensitive to the choice of parameters, notably the radius (epsilon) and the minimum number of points necessary to create a cluster (minPts). However, the Mean Shift was less susceptible to parameter adjustment since it automatically changes the bandwidth throughout the clustering process. They constructed fairly accurate anomaly detector models, suggesting that they may assist in identifying the subtle patterns that lie within geochemical information. In addition, DBSCAN efficiently deals with outliers, or data points that do not fit into any cluster (noise points), as part of its clustering procedure. As a result, DBSCAN gives clear cluster labels, such as a label for noise points, facilitating the interpretation of outliers. Mean-shift, however, detects cluster centers without explicitly labeling data points as outliers. Consequently, the interpretation of outliers may necessitate additional analysis.

Compared to the geological map shown in Fig. 4b, the anomalous groups (Fig. 11a, b, c) demonstrate a positive spatial association. This correlation is seen mainly in lower Cretaceous strata, recognized as significant geological variables contributing to the creation of MVT Pb—Zn deposits. Furthermore, the DBSACAN, Mean-shift and Fuzzy K-means algorithms have identified multivariate geochemical anomalies, particularly those belonging to strong anomaly classes, which tend to cluster in close proximity to the recognized MVT Pb—Zn deposits.

In this research, despite the favorable outcomes yielded by these two methods in detecting geochemical anomalies, it was imperative to employ an additional straightforward and widely used approach that has previously demonstrated acceptable results in geochemical clustering. The K-means method is frequently utilized by researchers,

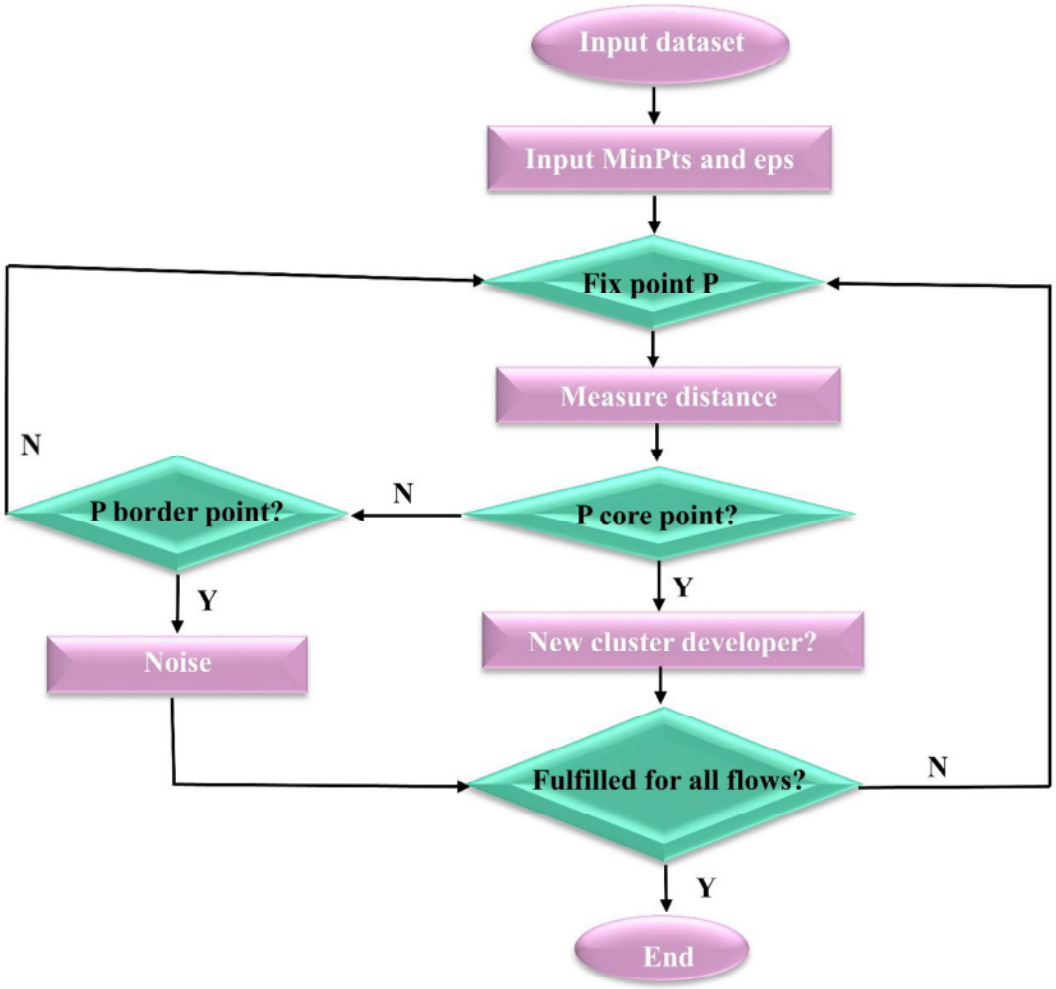

Fig. 9. Flowchart of DBSCAN clustering algorithm.

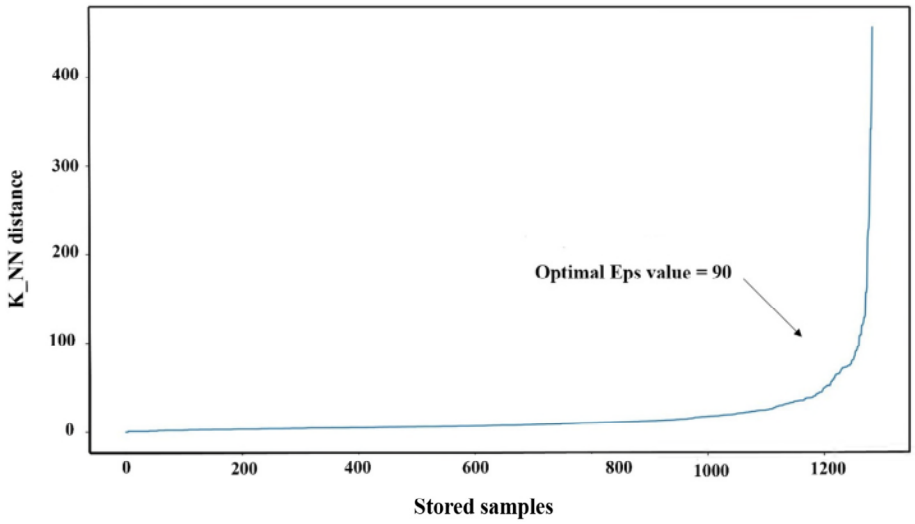

Fig. 10. K-nearest neighbor distance plot.

delivering satisfactory results in anomaly separation from the background (Ghezelbash et al., 2020a). In addition to K-means, derivatives of this method are also efficient. Hence, we used the Fuzzy K-means method to corroborate the results of the above two methods in this study. Essentially, this method was employed to validate the outcomes of the Mean-shift and DBSCAN, as the promising performance of the K-means method and its derivatives has been previously proven.

The primary unsupervised methods utilized in prior studies are

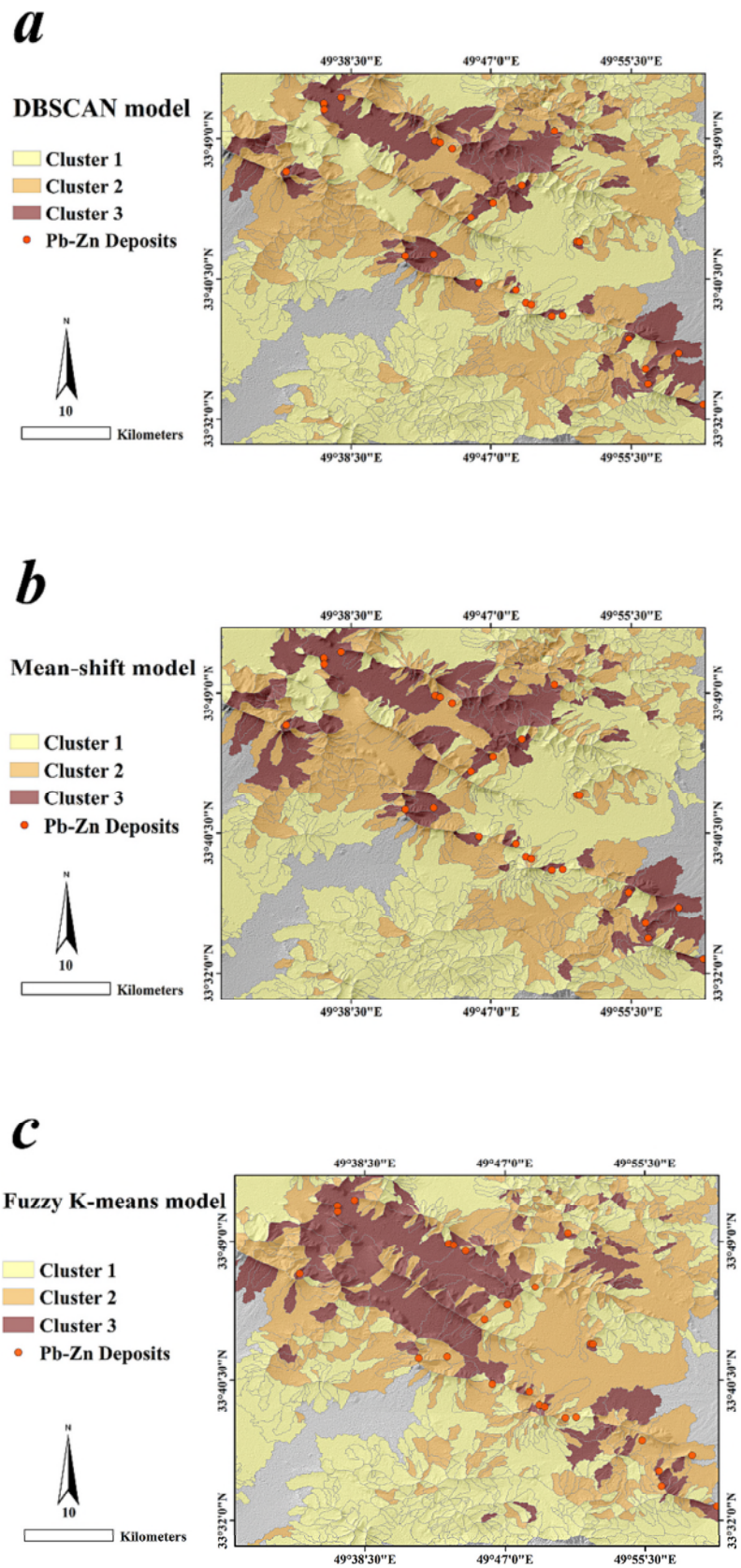

Fig. 11. SCB-based geochemical anomaly map derived by (a) DBSCAN, (b) Mean-shift, and (c) Fuzzy K-means algorithm.

Table 3
DBI and SI values of the single models.

Model	Davies-Bouldin index (DBI)	Silhouette index (SI)
DBSCAN	0.16	0.88
Mean-shift	0.24	0.69
Fuzzy K-means	0.36	0.55

distance-based, such as K-means or its optimized variants. In this paper, we demonstrated the effectiveness of density-based techniques in mapping geochemical anomalies. In exploratory sampling, more geochemical samples are collected from geologically complex locations featuring mineralization, alteration, etc. This results in a higher sampling density around mineral occurrences likely belonging to a specific class. Density-based methods similarly classify and group closely located samples. Consequently, this approach emerges as a practical tool for accentuating geochemical anomalies. Following this, a comparative analysis of the methods was conducted, and metrics evaluation (Table 3), coupled with the resultant map, affirmed the superiority of the DBSCAN method over the Mean-shift and Fuzzy K-means method.

6. Conclusion

Anomaly identification is vital when the abnormal behavior of a geodataset set gives essential information about the mineralization system. In this study, we employed the DBSCAN algorithm to recognize anomalies of stream sediment samples in the Varcheh area. In the experimental assessment, we compared the outcomes of the DBSCAN algorithm to those of the Mean-shift method. DBSCAN, as a density-based clustering algorithm, has proved helpful in identifying clusters in massive datasets of varying form and size. However, DBSCAN is a density-based clustering technique that shares certain features with Mean-shift, offering significant improvements. Compared to alternative clustering techniques, DBSCAN has several benefits. Unlike Mean-shift, which places outliers in a cluster regardless of their dissimilarity, this method recognized outliers as noise.

Furthermore, it detects arbitrarily sized and shaped clusters successfully. Therefore, the DBSCAN findings surpassed the Mean-shift results. This study shows that the DBSCAN algorithm can find anomalies since the investigation of the DBSCAN model and lithological factors affecting the occurrence of Pb—Zn mineralization in the surveyed region showed a significant correlation between the strong and weak anomaly categories and the Cretaceous formations. These formations were found to be the primary hosts of Pb—Zn deposits in the Varcheh district.

Moreover, comparing the outcomes of Fuzzy K-means, known for its consistent performance in geochemical clustering, indicated that these two density-based models predicted the highest Pb—Zn deposits within the smallest anticipated area compared to Fuzzy K-means in such a way that 81 % and 76 % of the Pb—Zn deposits, in 11 % and 15 % of the study district, respectively, are predicted by the strong anomaly classes of the SCBs derived through DBSACN and Mean-shift models.

CRediT authorship contribution statement

Mahsa Hajihosseinlou: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Abbas Maghsoudi:** Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation. **Reza Ghezelbash:** Conceptualization, Data curation, Formal analysis, Validation, Writing – review & editing.

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.

Data availability

The data that has been used is confidential.

Acknowledgment

The authors would like to thank Prof. Stefano Albanese, Editor-in-Chief of Journal of Geochemical Exploration, for handling this manuscript and also anonymous reviewers which their constructive comments improved the quality of our paper.

References

Aitchison, J., Egozcue, J., J., 2005. Compositional data analysis: where are we and where should we be heading? Math. Geol. 37 (7), 829–850.
Akbari, S., Ramazi, H., Ghezelbash, R., 2023. Using fractal and multifractal methods to reveal geophysical anomalies in Sardouyeh District, Kerman, Iran. Earth Sci. Inform. 1–18.
Aryafar, A., Moeini, H., 2017. Application of continuous restricted Boltzmann machine to detect multivariate anomalies from stream sediment geochemical data, Korit, East of Iran. J. Mining Environ. 8 (4), 673–682.
Barioni, M.C.N., Razente, H., Marcelino, A.M., Traina, A.J., Traina Jr., C., 2014. Open issues for partitioning clustering methods: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4 (3), 161–177.
Baudron, P., Alonso-Sarría, F., García-Aróstegui, J.L., Cánovas-García, F., Martínez-Vicente, D., Moreno-Brotóns, J., 2013. Identifying the origin of groundwater samples in a multi-layer aquifer system with Random Forest classification. J. Hydrol. 499, 303–315.
Bhuyan, R., Borah, S., 2013, March. A survey of some density based clustering techniques. In: Proceedings of the 2013 Conference on Advancements in Information, Computer and Communication.
Braune, C., Besecke, S., Kruse, R., 2015. Density based clustering: alternatives to DBSCAN. Partitional Clustering Algorithms 193–213.
Campello, R.J., Kröger, P., Sander, J., Zimek, A., 2020. Density-based clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10 (2), e1343.
Carranza, E.J.M., 2010. Catchment Basin Modelling of Stream Sediment Anomalies Revisited: Incorporation of EDA and Fractal Analysis.
Chen, J.L., Peng, R.M., Li, S.Z., Chen, X.C., 2017. Self-organizing feature map neural network and K-means algorithm as a data excavation tool for obtaining geological information from regional geochemical exploration data. Geophys. Geochem. Explor. 5, 919–927.
Chen, W., Liu, W., Liang, H., Jiang, M., Dai, Z., 2023. Response of storm surge and M2 tide to typhoon speeds along coastal Zhejiang Province. Ocean Eng. 270, 113646.
Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol. Rev. 32 (1–2), 314–324.
Cheng, Y., Lan, S., Fan, X., Tjahjadi, T., Jin, S., Cao, L., 2023. A dual-branch weakly supervised learning based network for accurate mapping of woody vegetation from remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 124, 103499.
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.
Daviran, M., Maghsoudi, A., Ghezelbash, R., Pradhan, B., 2021. A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach. Comput. Geosci. 148, 104688.
Daviran, M., Parsa, M., Maghsoudi, A., Ghezelbash, R., 2022. Quantifying uncertainties linked to the diversity of mathematical frameworks in knowledge-driven mineral prospectivity mapping. Nat. Resour. Res. 31 (5), 2271–2287.
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 1–13.
Dong, W., Yang, Y., Qu, J., Xiao, S., Li, Y., 2023a. Local information enhanced graph-transformer for hyperspectral image change detection with limited training samples. IEEE Trans. Geosci. Remote Sens. 16, 1–14.
Dong, W., Zhao, J., Qu, J., Xiao, S., Li, N., Hou, S., Li, Y., 2023b. Abundance matrix correlation analysis network based on hierarchical multihead self-cross-hybrid attention for hyperspectral change detection. IEEE Trans. Geosci. Remote Sens. 61, 1–13.
Dudek, A., 2019, September. Silhouette index as clustering evaluation tool. In: Conference of the Section on Classification and Data Analysis of the Polish Statistical Association. Springer, Cham, pp. 19–33.
Ehya, F., Lotfi, M., Rasa, I., 2010. Emarat carbonate-hosted Zn–Pb deposit, Markazi Province, Iran: a geological, mineralogical and isotopic (S, Pb) study. J. Asian Earth Sci. 37 (2), 186–194.
Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. August). A density-based algorithm for discovering clusters in large spatial databases with noise. kdd. 96 (34), 226–231.
Fahim, A.M., Saake, G., Salem, A.M., Torkey, F.A., Ramadan, M.A., 2008. K-means for spherical clusters with large variance in sizes. Int. J. Comput. Inform. Eng. 2 (9), 2923–2928.

Fernández, F.G., Both, R.A., Mangas, J., Arribas, A., 2000. Metallogenesis of Zn-Pb carbonate-hosted mineralization in the southeastern region of the Picos de Europa (central northern Spain) province: Geologic, fluid inclusion, and stable isotope studies. Econ. Geol. 95 (1), 19–40.
Filzmoser, P., Hron, K., Reimann, C., 2009. Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Sci. Total Environ. 407 (23), 6100–6108.
Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory 21 (1), 32–40.
Ghalamghash, J., Nédélec, A., Bellon, H., Abedini, M.V., Bouchez, J.L., 2009. The Urumieh plutonic complex (NW Iran): A record of the geodynamic evolution of the Sanandaj–Sirjan zone during Cretaceous times–Part I: Petrogenesis and K/Ar dating. J. Asian Earth Sci. 35 (5), 401–415.
Ghasemi, A., Talbot, C.J., 2006. A new tectonic scenario for the Sanandaj–Sirjan Zone (Iran). J. Asian Earth Sci. 26 (6), 683–693.
Ghazi, J.M., Moazzen, M., 2015. Geodynamic evolution of the Sanandaj-Sirjan zone, Zagros orogen, Iran. Turkish J. Earth Sci. 24 (5), 513–528.
Ghezelbash, R., Maghsoudi, A., 2018. Comparison of U-spatial statistics and C–A fractal models for delineating anomaly patterns of porphyry-type Cu geochemical signatures in the Varzaghan district, NW Iran. Comptes Rendus Geoscience 350 (4), 180–191.
Ghezelbash, R., Maghsoudi, A., Carranza, E.J.M., 2019a. 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., 2019b. Performance evaluation of RBF- and SVM-based machine learning algorithms for predictive mineral prospectivity modeling: integration of SA multifractal model and mineralization controls. Earth Sci. Inf. 12 (3), 277–293.
Ghezelbash, R., Maghsoudi, A., Daviran, M., 2019c. Combination of multifractal geostatistical interpolation and spectrum–area (S–A) fractal model for Cu–Au geochemical prospects in Feizabad district, NE Iran. Arabian J. Geosci. 12 (5), 1–14.
Ghezelbash, R., Maghsoudi, A., Daviran, M., 2019d. Prospectivity modeling of porphyry copper deposits: recognition of efficient mono-and multi-element geochemical signatures in the Varzaghan district, NW Iran. Acta Geochimica 38, 131–144.
Ghezelbash, R., Maghsoudi, A., Carranza, E.J.M., 2020a. Optimization of geochemical anomaly detection using a novel genetic K-means clustering (GKMC) algorithm. Comput. Geosci. 134, 104335.
Ghezelbash, R., Maghsoudi, A., Carranza, E.J.M., 2020b. Sensitivity analysis of prospectivity modeling to evidence maps: Enhancing success of targeting for epithermal gold, Takab district, NW Iran. Ore Geol. Rev. 120, 103394.
Ghezelbash, R., Maghsoudi, A., Shamekhi, M., Pradhan, B., Daviran, M., 2023a. Genetic algorithm to optimize the SVM and K-means algorithms for mapping of mineral prospectivity. Neural Comput. & Applic. 35 (1), 719–733.
Ghezelbash, R., Daviran, M., Maghsoudi, A., Ghaeminejad, H., Niknezhad, M., 2023b. 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, 105538.
Greenacre, M., 2021. Compositional data analysis. Annual Review of Statistics and its Application 8, 271–299.
Grunsky, E.C., 2010. The Interpretation of Geochemical Survey Data.
Guérillot, D., Bruyelle, J., 2020. Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation. Comput. Geosci. 24 (2), 697–707.
Hajihosseinlou, M., Maghsoudi, A., Ghezelbash, R., 2023. A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm. Nat. Resour. Res. 1–22.
Hajihosseinlou, M., Maghsoudi, A., Ghezelbash, R., 2024. Stacking: a novel data-driven ensemble machine learning strategy for prediction and mapping of Pb-Zn prospectivity in Varcheh district, West Iran. Expert Syst. Appl. 237, 121668.
Hancer, E., Karaboga, D., 2017. A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number. Swarm and Evolutionary Computation 32, 49–67.
He, M.Y., Dong, J.B., Jin, Z., Liu, C.Y., Xiao, J., Zhang, F., Deng, L., 2021. Pedogenic processes in loess-paleosol sediments: Clues from Li isotopes of leachate in Luochuan loess. Geochim. Cosmochim. Acta 299, 151–162.
Howarth, R.J., Thompson, M., 1976. Duplicate analysis in geochemical practice. Part II. Examination of proposed method and examples of its use. Analyst 101 (1206), 699–709.
Jia, B., Zhou, G., 2023. Estimation of global karst carbon sink from 1950s to 2050s using response surface methodology. Geo-spatial Inform. Sci. 1–18.
Karimpour, M.H., Sadeghi, M., 2018. Dehydration of hot oceanic slab at depth 30–50 km: KEY to formation of Irankuh-Emarat PbZn MVT belt, Central Iran. J. Geochem. Explor. 194, 88–103.
Karimpour, M.H., Malekzadeh Shafaroudi, A., Esmaeili Sevieri, A., Saeed, S., Allaz, J.M., Stern, C.R., 2017. Geology, mineralization, mineral chemistry, and ore-fluid conditions of Irankuh Pb-Zn mining district, south of Isfahan. J. Econ. Geol. 9 (2), 267–294.
Khorshidi, M., Ameri, M., Goli, A., 2023. Cracking performance evaluation and modelling of RAP mixtures containing different recycled materials using deep neural network model. Road Materials and Pavement Design 1–20.
Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of Cluster in K-Means Clustering. Int. J. 1 (6), 90–95.
Kriegel, H.P., Kroger, P., Pryakhin, A., Schubert, M., 2005, November. Effective and efficient distributed model-based clustering. In: Fifth IEEE International Conference on Data Mining (ICDM'05). IEEE, p. 8.
Kriegel, H.P., Kröger, P., Sander, J., Zimek, A., 2011. Density-based clustering. Wiley interdisciplinary reviews: data mining and knowledge discovery 1 (3), 231–240.
Kürzl, H., 1988. Exploratory data analysis: recent advances for the interpretation of geochemical data. J. Geochem. Explor. 30 (1–3), 309–322.
Li, J., Wang, Z., Wu, X., Xu, C.Y., Guo, S., Chen, X., 2020. Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index. J. Hydrometeorol. 21 (5), 891–908.
Li, J., Wang, Y., Nguyen, X., Zhuang, X., Li, J., Querol, X., Do, V., 2022. First insights into mineralogy, geochemistry, and isotopic signatures of the Upper Triassic high-sulfur coals from the Thai Nguyen Coal field, NE Vietnam. Int. J. Coal Geol. 261, 104097.
Li, W., Wang, W., Sun, R., Li, M., Liu, H., Shi, Y., Fu, S., 2023. Influence of nitrogen addition on the functional diversity and biomass of fine roots in warm-temperate and subtropical forests. For. Ecol. Manage. 545, 121309.
Liu, Y., Zhou, K., Carranza, E.J.M., 2018. Compositional balance analysis for geochemical pattern recognition and anomaly mapping in the western Junggar region, China. Geochem.: Explor., Environ., Anal. 18 (3), 263–276.
Luo, J., Wang, G., Li, G., Pesce, G., 2022. Transport infrastructure connectivity and conflict resolution: a machine learning analysis. Neural Comput. Appl. 34 (9), 6585–6601.
Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (PCA). Comput. Geosci. 19 (3), 303–342.
McLachlan, G.J., Basford, K.E., 1988. Mixture Models: Inference and Applications to Clustering, vol. 38. M. Dekker, New York.
Michaud, P., 1997. Clustering techniques. Futur. Gener. Comput. Syst. 13 (2–3), 135–147.
Nagpal, P.B., Mann, P.A., 2011. Comparative study of density based clustering algorithms. Int. J. Comput. Appl. 27 (11), 421–435.
Nezhad, S.G., 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.
Nielsen, F., 2016. Hierarchical clustering. In: Introduction to HPC with MPI for Data Science. Springer, Cham, pp. 195–211.
Prades, C., 2018. Geostatistics and Clustering for Geochemical Data Analysis.
Qiu, D., Zhu, G., Bhat, M.A., Wang, L., Liu, Y., Sang, L., Sun, N., 2023. Water use strategy of nitraria tangutorum shrubs in ecological water delivery area of the lower inland river: based on stable isotope data. J. Hydrol. 624, 129918.
Rajabi, A., Mahmoodi, P., Rastad, E., Niroomand, S., Canet, C., Alfonso, P., Yarmohammadi, A., 2019. Comments on "Dehydration of hot oceanic slab at depth 30–50 km: Key to formation of Irankuh-Emarat Pb-Zn MVT belt, Central Iran" by Mohammad Hassan Karimpour and Martiya Sadeghi. J. Geochem. Explor. 205, 106346.
Ren, C., Yu, J., Liu, S., Yao, W., Zhu, Y., Liu, X., 2022. A plastic strain-induced damage model of porous rock suitable for different stress paths. Rock Mech. Rock. Eng. 55 (4), 1887–1906.
Sander, J., Ester, M., Kriegel, H.P., Xu, X., 1998. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data Min. Knowl. Discov. 2, 169–194.
Stocklin, J., 1968. Structural history and tectonics of Iran: a review. AAPG bulletin 52 (7), 1229–1258.
Thiombane, M., Di Bonito, M., Albanese, S., Zuzolo, D., Lima, A., De Vivo, B., 2019. Geogenic versus anthropogenic behaviour and geochemical footprint of Al, Na, K and P in the Campania region (Southern Italy) soils through compositional data analysis and enrichment factor. Geoderma 335, 12–26.
Tian, H., Huang, N., Niu, Z., Qin, Y., Pei, J., Wang, J., 2019. Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm. Remote Sens. (Basel) 11 (7), 820.
Tian, H., Pei, J., Huang, J., Li, X., Wang, J., Zhou, B., Wang, L., 2020. Garlic and winter wheat identification based on active and passive satellite imagery and the google earth engine in northern China. Remote Sens. (Basel) 12 (21), 3539.
Tokushige, S., Yadohisa, H., Inada, K., 2007. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. Comput. Stat. 22, 1–16.
Vahid, R., Farnood Ahmadi, F., Mohammadi, N., 2021. Earthquake damage modeling using cellular automata and fuzzy rule-based models. Arab. J. Geosci. 14, 1–14.
Vergani, A.A., Binaghi, E., 2018, July. A soft davies-bouldin separation measure. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, pp. 1–8.
Wang, Z., Zuo, R., Dong, Y., 2019. Mapping geochemical anomalies through integrating random forest and metric learning methods. Nat. Resour. Res. 28 (4), 1285–1298.
Wang, J., Zhou, Y., Xiao, F., 2020. Identification of multi-element geochemical anomalies using unsupervised machine learning algorithms: a case study from Ag–Pb–Zn deposits in North-Western Zhejiang, China. Appl. Geochem. 120, 104679.
Wang, H., Zhang, X., Jiang, S., 2022. A laboratory and field universal estimation method for tire–pavement interaction noise (TPIN) based on 3D image technology. Sustainability 14 (19), 12066.
Wu, M., Ba, Z., Liang, J., 2022. A procedure for 3D simulation of seismic wave propagation considering source-path-site effects: Theory, verification and application. Earthquake Engineering & Structural Dynamics 51 (12), 2925–2955.
Xie, X., Xie, B., Cheng, J., Chu, Q., Dooling, T., 2021. A simple Monte Carlo method for estimating the chance of a cyclone impact. Nat. Hazards 107 (3), 2573–2582.
Xiong, Y., Zuo, R., 2016. Recognition of geochemical anomalies using a deep autoencoder network. Comput. Geosci. 86, 75–82.
Xiong, Y., Zuo, R., 2020. Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. Comput. Geosci. 140, 104484.
Xu, Z., Li, X., Li, J., Xue, Y., Jiang, S., Liu, L., Sun, Q., 2022. Characteristics of Source Rocks and Genetic Origins of Natural Gas in Deep Formations, Gudian Depression, Songliao Basin, NE China. ACS Earth and Space Chemistry 6 (7), 1750–1771.

Yang, M.S., 1993. A survey of fuzzy clustering. Math. Comput. Model. 18 (11), 1–16.
Yao, M., Jiangnan, Z., 2021. Advances in the application of machine learning methods in mineral prospectivity mapping, 40(1), pp. 132–141.
Yilmaz, H., Ghezelbash, R., Cohen, D.R., Sari, R., Sönmez, F.N., Maghsoudi, A., 2020. Comparison between the geochemical response of BLEG and fine fraction stream sediments to mineralization in the Eastern Black Sea region, Turkey. J. Geochem. Explor. 217, 106609.
Yin, L., Wang, L., Li, J., Lu, S., Tian, J., Yin, Z., Zheng, W., 2023a. YOLOV4_CSPBi: Enhanced land target detection model. Land 12 (9), 1813.
Yin, L., Wang, L., Li, T., Lu, S., Tian, J., Yin, Z., Zheng, W., 2023b. U-Net-LSTM: Time series-enhanced lake boundary prediction model. Land 12 (10), 1859.
Yin, L., Wang, L., Li, T., Lu, S., Yin, Z., Liu, X., Zheng, W., 2023c. U-Net-STN: a novel end-to-end lake boundary prediction model. Land 12 (8), 1602.
Yu, J., Zhu, Y., Yao, W., Liu, X., Ren, C., Cai, Y., Tang, X., 2021. Stress relaxation behaviour of marble under cyclic weak disturbance and confining pressures. Measurement 182, 109777.
Yuan, C., Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. J 2 (2), 226–235.
Zarasvandi, A., Sameti, M., Sadeghi, M., Rastmanesh, F., Pourkaseb, H., 2014. The Gol-e-Zard Zn-Pb Deposit, Lorestan Province, Iran: a Metamorphosed SEDEX Deposit. Acta Geologica Sinica-English Edition 88 (1), 142–153.
Zhang, S., Carranza, E.J.M., Xiao, K., Wei, H., Yang, F., Chen, Z., Xiang, J., 2022. Mineral prospectivity mapping based on isolation forest and random forest: Implication for the existence of spatial signature of mineralization in outliers. Nat. Resour. Res. 31 (4), 1981–1999.
Zheng, H., Fan, X., Bo, W., Yang, X., Tjahjadi, T., Jin, S., 2023. A multiscale point-supervised network for counting maize tassels in the wild. Plant Phenomics 5, 0100.
Zhou, G., Yang, Z., 2023. Analysis for 3-D morphology structural changes for underwater topographical in Culebrita Island. Int. J. Remote Sens. 44 (7), 2458–2479.
Zhou, G., Wang, Z., Li, Q., 2022. Spatial negative co-location pattern directional mining algorithm with join-based prevalence. Remote Sens. (Basel) 14 (9), 2103.
Zuo, R., Wang, J., 2016. Fractal/multifractal modeling of geochemical data: a review. J. Geochem. Explor. 164, 33–41.
Zuo, R., Xiong, Y., 2018. Big data analytics of identifying geochemical anomalies supported by machine learning methods. Nat. Resour. Res. 27 (1), 5–13.