



Rotation-based outlier detection for geochemical anomaly identification in stream sediment multivariate data

Shahed Shahrestani¹ · Ioan Sanislav¹ · Hosein Fereydooni²

Received: 28 October 2024 / Accepted: 15 February 2025
© The Author(s) 2025

Abstract

This research explores the use of the rotation-based outlier detection (ROD) method for identifying geochemical anomalies in a multivariate stream sediment dataset from Iran, targeting porphyry and vein-type Cu mineralization. Geochemical datasets often present challenges for outlier detection methods like local outlier factor (LOF) and k-nearest neighbor (KNN), which rely on distance or density metrics and require parameter tuning (e.g., neighborhood size k). High-dimensional feature spaces further complicate their application. ROD, in contrast, offers a parameter-free, rotation-based approach that effectively analyzes geometric relationships between samples in subspaces, mitigating the curse of dimensionality. This makes ROD particularly suited to high-dimensional geochemical datasets, where complex relationships between elements (due to lithology or mineralization) are critical for identifying anomalies. This study compares ROD with LOF and KNN using two subsets of geochemical variables (Ag, As, Au, Bi, Co, Cr, Cu, Mo, Ni, Pb, Sb, Zn; and Ag, As, Au, Cu, Mo, Sb) and evaluates its performance based on the receiver operating characteristic (ROC) analysis and the number of known mineral occurrences detected in anomaly class. ROD outperforms LOF and KNN, capturing 78% (14 out of 18) of known Cu-bearing mineral occurrences. Moreover, ROD shows better conformity between 10% of highest outlier scores and Cu-mineralization sites. Rotation cost function in ROD, evaluated using the median absolute deviation (MAD), enhances its ability to detect outliers by focusing on orientation rather than distance, and by reducing noise misclassification. In addition, the parameter-free design of ROD and improved handling of high-dimensional data makes it a promising tool for geochemical exploration, as it captures unique mineralization-related signals that might be missed by traditional methods.

Keywords Outlier detection · Geochemical anomaly · LOF · KNN · ROD · Cu mineralization

Introduction

Multivariate geochemical data resulting from rock chips, soil, or stream sediment sampling can serve as potential datasets for unsupervised outlier detection methods (e.g., Shahrestani and Sanislav 2025a). Unsupervised machine learning models have proven highly effective for anomaly detection, primarily because they do not require labeled

samples (e.g., Shahrestani and Carranza 2024). Outlier detection methods rank samples based on their deviations from normal behavior, reflecting signals originating from anomalous sources. These methods are classified into different groups based on the techniques employed to isolate outlier samples, including statistical approaches, proximity-based methods, density-based techniques, clustering methods, ensemble approaches, and machine learning algorithms (e.g., Shahrestani and Carranza 2024; Zimek et al. 2012). In the context of geochemical exploration, there is an increasing trend toward using unsupervised outlier detection approaches, such as restricted Boltzmann machines (RBMs) (Chen 2015), one-class support vector machines (OCSVMs) (Chen and Wu 2017; Shahrestani and Carranza 2024), dictionary learning (Chen and Sui 2022), deep autoencoders (Xiong and Zuo 2016), variational autoencoders (VAEs) (Zhao et al. 2023; Esmailoghli et al. 2023), geo-HGAN (Ding et al. 2024), model averaging

Communicated by: Hassan Babaie

✉ Ioan Sanislav
ioan.sanislav@jcu.edu.au

¹ College of Science and Engineering, Economic Geology Research Centre (EGRU), James Cook University, Townsville, Australia

² Department of Earth Sciences, Carleton University, Ottawa, ON, Canada

(Wang and Zuo 2022), isolation forest (IF) (Chen and Wu 2019), bat-optimized OCSVM models (Chen et al. 2019), k-nearest neighbor methods (Chen et al. 2021a), local outlier factor (LOF) (Puchhammer et al. 2024; Shahrestani and Carranza 2024), distance anomaly factors (Chen et al. 2021b), density-based spatial clustering of applications with noise (DBSCAN) (Hajihosseini et al. 2024a), ordering points to identify the clustering structure (OPTICS) (Hajihosseini et al. 2024b), concentration–volume (C–V) fractal modeling (Afzal et al. 2011, 2016), the U value-number fractal (U–N) and the U value-area fractal (U–A) methods (Ghannadpour and Hezarkhani 2022), generative adversarial networks (GANs) (Zhang and Zuo 2021) and self-organizing maps (Bigdeli et al. 2022), to name but a few.

From another perspective, outlier detection techniques can be categorized into global and local approaches (e.g., Liu et al. 2025; Rayana et al. 2016). Global detection algorithms analyze the entire dataset, making them effective when outliers exhibit significant deviations from the overall data distribution (e.g., Puchhammer et al. 2024). In contrast, local detection algorithms focus on specific subsets of objects (e.g., Zhao et al. 2019a; Schubert et al. 2014). The choice between global and local outlier detection methods can greatly impact performance, as the data structure plays a pivotal role in determining the most suitable approach. For instance, global methods are highly effective when anomalies differ markedly from the majority of the data, whereas local methods excel at identifying anomalies within dense clusters of normal behavior.

The integration of deep learning techniques in geochemical anomaly detection has garnered significant attention in recent years, offering advanced methods for identifying mineralization-associated anomalies. Zhang et al. (2021) introduced a method combining deep convolutional neural networks (DCNNs) with pixel pair features, enhancing anomaly detection. Huang et al. (2022) applied Bayesian convolutional neural networks to quantify uncertainty in anomaly predictions, improving geological interpretations. Cao et al. (2023) optimized a random forest model using a competitive mechanism and beetle antennae search to advance anomaly detection. To address class imbalance in training datasets, Shayilan and Chen (2024) proposed a SMOTified extreme learning machine (ELM) model, merging the synthetic minority oversampling technique (SMOTE) with ELMs. Guo and Chen (2024) improved SMOTE by integrating it with generative adversarial networks (GANs), creating a SMOTified-GAN hybrid to enhance anomaly detection and the performance of ELMs in mineralization detection.

Unsupervised deep learning methods for anomaly detection are categorized into reconstructive and generative approaches (Wang and Chen 2025). Reconstruction-based methods, such as autoencoders, are effective in identifying

anomalies through reconstruction errors (Xiong and Zuo 2016). Various methods, including continuous restricted Boltzmann machines (CRBMs) (Chen and Murray 2003), deep autoencoders (DAE) (Hinton and Salakhutdinov 2006), and variational autoencoders (VAEs) (Doersch 2016), have been shown to be effective for detecting multivariate geochemical anomalies. Chen et al. (2014) applied CRBMs to model probability distributions in complex geological environments, successfully identifying geochemical anomalies related to polymetallic mineralization. Similarly, Xiong and Zuo (2016) used DAEs to identify complex multivariate distributions linked to mineralization, while Chen and Shayilan (2022) distinguished gold mineralization-related anomalies using dictionary learning techniques. Furthermore, Luo et al. (2023) enhanced the interpretability of VAE models by integrating geological loss terms. Despite their success, deep autoencoder networks (DANs) have faced challenges when recognizing multi-element geochemical anomalies, as they disregard spatial structures (e.g., Zuo and Xu 2024). Recent advancements, such as the framework developed by Soltani et al. (2024), integrate deep learning with spatial machine learning processors to incorporate local neighborhood information and spatial non-stationarities in anomaly detection.

Spatial-based deep learning methods, such as convolutional neural networks (CNNs) (LeCun et al. 1998), convolutional autoencoders (CAEs) (Masci et al. 2011), and graph neural networks (Zuo and Xu 2023), have emerged to capture spatial relationships among geochemical variables. These methods overcome the fixed-size image limitations in CNNs and CAEs by allowing more flexible, graph-based modeling of geochemical data. Physics-constrained models further enhance deep learning interpretations by embedding prior geological knowledge and physical laws into frameworks, improving alignment with geological processes (e.g., Zuo et al. 2023). For example, Xiong et al. (2022) incorporated geological constraints into variational autoencoders (VAEs) to reveal nonlinear influences on mineralization. While reconstruction-based methods use reconstruction errors for anomaly detection, they may fail to capture anomaly complexities in geochemical data (Wang and Chen 2025). Recent advancements in unsupervised deep learning, such as deep autoencoders (DAE), GANs, and graph attention networks, are addressing these challenges. Li et al. (2024) developed a 1D-CNN model that achieved 99.18% accuracy in deposit type classification, while DAGMM (Wang and Chen 2025) merges dimensionality reduction with probabilistic modeling, enhancing geochemical anomaly detection. Wang et al. (2023) constructed an AlexNet-based CNN for automatic lithological identification from stream sediment data, further demonstrating the potential of deep learning techniques in geochemical exploration.

In geochemical exploration, several key factors should be considered when selecting an outlier detection method. First,

it is important to choose methods that are robust to high-dimensional feature spaces or at least evaluate whether they suffer from the "curse of dimensionality." As dimensionality increases, the distinction between the nearest and farthest distances becomes less pronounced, weakening the effectiveness of distance-based metrics (e.g., Park 2023; Beyer et al. 1999). Second, the time complexity of the chosen method is crucial. For example, the angle-based outlier detection method (ABOD; Kriegel et al. 2008) is robust to dimensionality but becomes computationally expensive when applied to large datasets (Shahrestani and Sanislav 2025b). Third, consideration must be given to the number of parameters required by an outlier detection method and its sensitivity to those parameters. For instance, the accuracy of local outlier factor (LOF; Breunig et al. 2000) scores is highly sensitive to the choice of the number of neighbours (e.g., Shahrestani and Sanislav 2025b). Fourth, high feature density in most multivariate geochemical datasets can render certain criteria, such as within-class distances from the center, ineffective in some clustering-based outlier detection methods. Fifth, in many multivariate geochemical datasets, not all samples are analyzed for every element, leading to missing data. The outlier detection method should address this issue during the scoring process. Sixth, interpretability of the resulting outlier scores varies significantly among different outlier detection methods (Li et al. 2022). Finally, the effectiveness of an outlier detection method in delineating geochemical anomalies across different scales, from regional surveys to local investigations, should be assessed. In regional multivariate datasets, for example, samples are often collected from diverse geological settings, and the outlier detection algorithm should effectively capture these intrinsic sources of variation (e.g., Shahrestani and Carranza 2024).

These considerations highlight that the application of outlier detection methods in mineral exploration remains in its early stages, indicating a significant need for implementing various detection strategies based on different criteria for identifying geochemical anomalies. In response to this need, this paper investigates the effectiveness of an outlier detection method known as rotation-based outlier detection (ROD) (Almardeny et al. 2020). In ROD, the dataset is first divided into successive 3D subspaces, and each subspace is then rotated around the geometric median to determine outlier scores. This approach is easy to implement and is parameter-free, requiring no assumptions about the underlying distribution. To evaluate the efficiency of the ROD method in detecting geochemical anomalies, the resulting scores are compared with those from two well-established local outlier detection methods, LOF and KNN, which are applied to a multivariate regional geochemical dataset from a Cu-bearing mineralization province in southeastern Iran. The performance of ROD is compared with LOF and KNN, as these methods are widely used in anomaly detection tasks

and represent two key approaches. LOF, a density-based method, identifies outliers by comparing the local density of data points, while KNN, a distance-based method, detects outliers based on the distance to the nearest neighbors. Both methods are commonly applied to multivariate datasets, making them suitable benchmarks for evaluating the effectiveness of new outlier detection techniques. Moreover, the well-understood behaviors and limitations of LOF and KNN, such as sensitivity of LOF to parameter choices and challenges of KNN in high-dimensional spaces, are considered. By comparing ROD—an innovative, parameter-free method that rotates data vectors about the geometric median—the aim is to demonstrate how unique approach of ROD addresses these challenges.

Materials and methods

Case study

The Iranian Plate is situated at the boundary between the Eurasian Plate to the north and the Gondwana Plate to the south (Aminzadeh and Rahimpour 2021). The geotectonic history of the Iranian Plate has been shaped by the opening and closing of the Paleo-Tethys and Neo-Tethys Oceans (Bagheri and Stampfli 2008; Masoodi et al. 2013). It is believed that the Iranian Plate separated from the Eurasian Plate during the Late Ordovician to Early Silurian, forming part of the passive margin of the Paleo-Tethys Ocean (Natašin and Şengör, 2005). The subduction of the oceanic crust of the Paleo-Tethys beneath the Eurasian Plate led to the rifting of Gondwana and the spreading of the Neo-Tethys Ocean in the Early Triassic (Stampfli and Borel 2002). The Neo-Tethys Ocean, located between the Iranian and Arabian Plates, played a pivotal role in shaping the tectonic framework of the region.

The subduction of the Neo-Tethys crust beneath the southern margin of the Iranian Plate resulted in the formation of three main tectonic belts: the Urumieh-Dokhtar Magmatic Belt (UDMB), the Sanandaj-Sirjan Zone, and the Zagros Folding and Thrusting Belt, which developed between the Early Jurassic and Eocene (Ghasemi and Talbot 2006; Bagheri and Stampfli 2008). The Urumieh-Dokhtar Magmatic Arc, a key feature of the region, extends over 1700 km along the NW–SE direction and is known for hosting numerous porphyry copper deposits associated with Oligo-Miocene intrusive bodies (e.g., Richards et al. 2012).

The Baft district, located within the Kerman Cenozoic Magmatic Arc (KCMA) of the Urumieh-Dokhtar Magmatic Belt (UDMB), covers an area of approximately 1250 km² and presents a complex geological setting (Safari et al. 2015). The basement of the district consists of Paleozoic metamorphic rocks, overlain by a Cretaceous mélangé, with

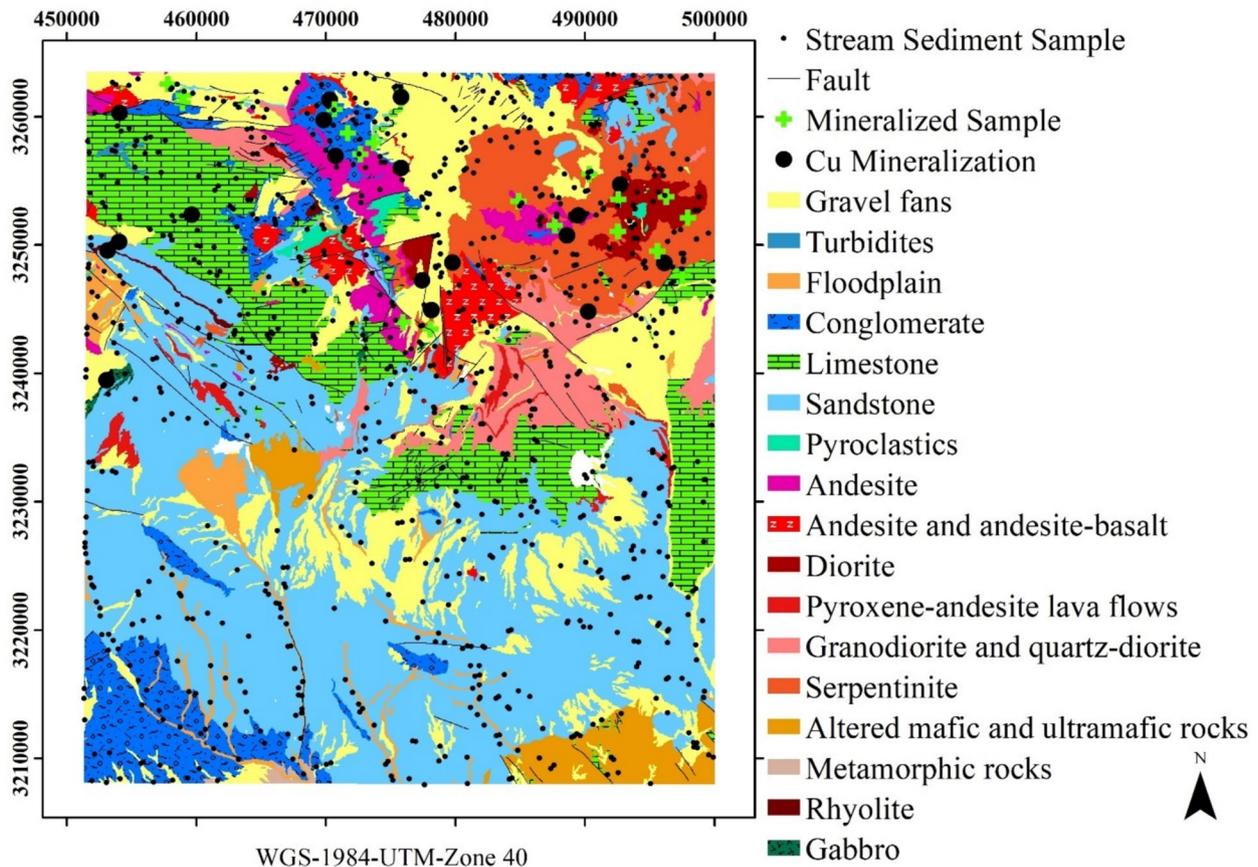


Fig. 1 Simplified geological map of Baft area with the scale of 100,000 (after Srdic et al. 1972)

Eocene volcanic and sedimentary formations (Fig. 1). Miocene intrusive rocks, including the Lalezar granitoid and Rabor-Lalezar Magmatic Complex (RLMC), intrude the Eocene volcanic rocks (Shafiei 2010; Srdic et al. 1972). Oligocene–Miocene volcanic and sedimentary sequences are also present, with Quaternary deposits covering the region (Ghasemzadeh et al. 2022).

The northern parts of the Baft district are dominated by Eocene volcanic rocks, including andesites and dacites, which are intruded by Upper Miocene quartz monzodiorite and granodiorite, part of the Lalezar granitoids or RLMC. These granitoids exhibit high-K calc-alkaline characteristics and range from gabbrodiorites to granites (Moghadam et al. 2018). Mineral compositions of the most felsic rocks in the region are characterized by the presence of Na-plagioclase, quartz, alkali feldspar, biotite, and hornblende, while gabbro-diorite rocks contain plagioclase (Ca-rich), hornblende, biotite, and clinopyroxene (Srdic et al. 1972; Dimitrijevic 1973).

The region is economically significant due to its rich porphyry copper deposits, with notable mineralization at locations such as Bid Khan, Ghaleh Asghar, and Harij. These

deposits are primarily associated with Miocene intrusive bodies (e.g., Dimitrijevic 1973). Additionally, vein-type mineralization is observed at sites such as Goghar, Ab Bahri, and Paynegin (Aghazadeh et al. 2015; Mohebi et al. 2015; Ghasemzadeh et al. 2019). The mineralization in these areas is linked to the hydrothermal alteration associated with the Miocene granitoid intrusions (Ghasemzadeh et al. 2022).

Geochemical data and preprocessing

A geochemical survey was conducted in the Baft district, during which 911 stream sediment samples were systematically collected, with an average sampling density of one sample per 1 km². These samples were analyzed by geological survey of Iran using the inductively coupled plasma—optical emission spectrometry method (ICP-OES), while Au contents were determined using fire assay followed by atomic absorption spectroscopy (AAS) (Ghasemzadeh et al. 2022). Analytical reliability was ensured through the application of duplicate analyses, following the methodology of Thompson and Howarth (1976), which demonstrated

a precision of better than 10% for most analyzed elements (Ghasemzadeh et al. 2019).

In addition to using mineral occurrences as validation points for regional geochemical anomalies, several rock samples were collected from upstream areas of stream sediment sites that exhibited relatively high concentrations of individual chemical elements, as indicated by the stream sediment geochemical survey (Table 1). For the multivariate geochemical analysis, 12 elements were selected, including the primary economic metal (Cu) and pathfinder elements (Ag, As, Au, Bi, Cu, Mo, Pb, Sb, Zn) due to their association with mineralization processes (e.g., Jamali 2017; Ghasemzadeh et al. 2019), along with three geologically influenced elements (Cr,

Co, Ni) (Table 2). To account for the compositional constraints inherent in geochemical data, the dataset was subjected to centered log-ratio (clr) transformation, as proposed by Aitchison (1986).

Outlier detection methods

Rotation-based outlier detection approach (ROD)

Stream sediment geochemical surveys play a crucial role in identifying geochemical anomalies associated with anomalous sources such as mineral deposits and understanding geochemical dispersion patterns. Elements in stream sediments can vary widely in concentration due to diverse upstream sources and roles of weathering, erosion, transportation, and deposition (e.g., Shahrestani et al. 2024; Najafian et al. 2023). These processes produce complex multivariate datasets characterized by inherent heterogeneities and anomalies that conventional statistical methods struggle to capture effectively. In this context, the rotation-based outlier detection (ROD) method offers significant advantages by considering the relationships among multivariate features, enabling the detection of subtle anomalies embedded within high-dimensional datasets. The concept of the ROD method can initially be illustrated using a three-dimensional feature space and subsequently generalized for higher dimensions, which is common in most outlier detection tasks. Consider three collinear vectors, \vec{x} , \vec{y} , and \vec{z} in R^3 , which form a parallelepiped with \vec{x} and \vec{y} serving as the base (Fig. 2). The volume of this parallelepiped can be expressed as follows (Almardeny et al. 2020):

$$\text{volume} = |(\vec{x} \times \vec{y}) \cdot \vec{z}| = (\|\vec{x}\| \|\vec{y}\| \sin \gamma) \cdot \|\vec{z}\| \cos \theta \tag{1}$$

where, $\gamma = \angle(\vec{x}, \vec{y})$ represents the angle between vectors \vec{x} and \vec{y} , while θ denotes the angle between \vec{z} and the height (h) that is perpendicular from \vec{z} to the base (Fig. 2a). According to the figure, vectors \vec{y} and \vec{z} can be generated by

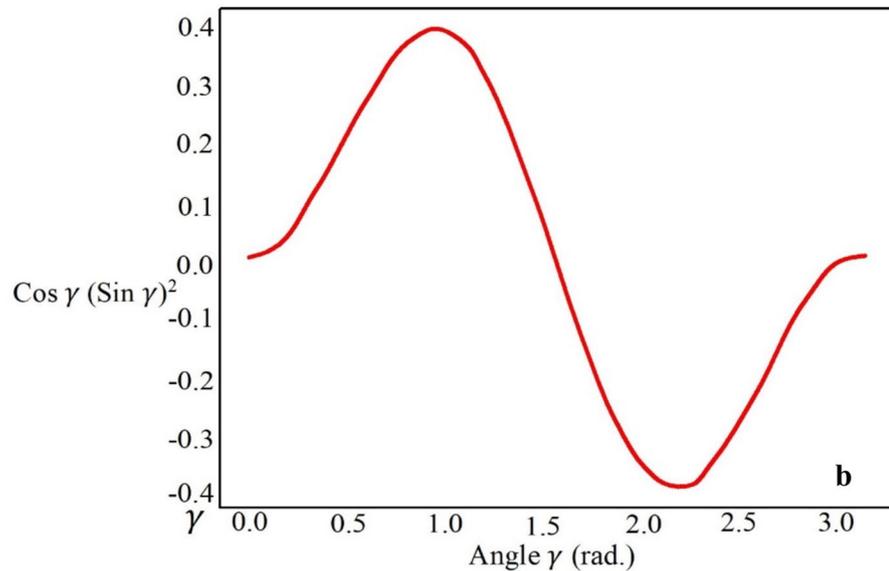
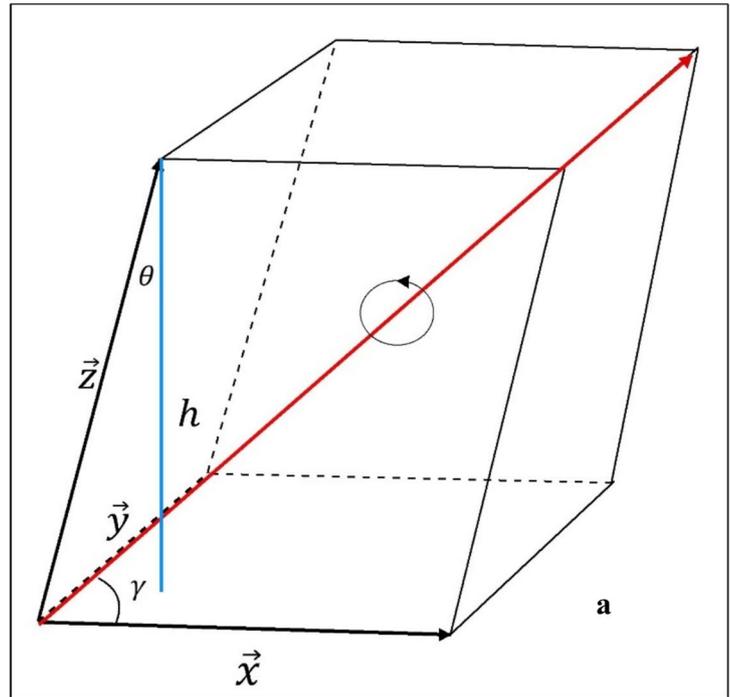
Table 1 Result of sample analysis for Au, Cu, Pb, and Zn in mineralized samples in anomaly checking stage

No	Au (ppb)	Cu (ppm)	Pb (ppm)	Zn (ppm)
1	0.75	61	13	1108
2	0.75	2380	21	108
3	1	71	1129	968
4	1	74	1403	1126
5	3	86	13440	11622
6	5	65	1887	1600
7	13	134	17370	13350
8	18	1100	150	1840
9	20	941	1239	669
10	20	943	13	55
11	44	2824	230	797
12	71	5200	8	36
13	122	36	177	60
14	134	3450	107	405
15	184	74	121	61
16	201	186	194	93
17	476	3230	609	2140
18	561	6400	788	184

Table 2 Summary statistics of the raw values for the 12 elements included in the data processing procedure from the geochemical dataset of the Baft area

	Min	Max	Mean	Range	St.Dev	Skewness	Kurtosis
Ag	0.01	29.40	0.13	29.39	1.03	25.94	724.56
As	1.20	743.00	19.36	741.80	29.24	17.33	415.12
Au	0.00	0.25	0.00	0.25	0.01	15.21	310.21
Bi	0.02	2.10	0.19	2.08	0.21	4.51	26.54
Co	10.20	89.30	24.47	79.10	9.29	2.08	6.15
Cr	24.00	5810.00	330.62	5786.00	521.58	4.14	25.26
Cu	3.40	331.00	57.34	327.60	26.35	2.35	15.40
Mo	0.30	6.96	1.01	6.66	0.62	4.08	24.63
Ni	9.60	1225.00	83.58	1215.40	124.27	4.19	21.07
Pb	3.40	10000.00	45.49	9996.60	398.67	20.01	456.60
Sb	0.12	119.00	1.69	118.88	4.86	17.32	385.67
Zn	45.00	5640.00	112.47	5595.00	226.27	19.08	423.33

Fig. 2 a Parameters related to the parallelepiped created by three non-collinear vectors of \vec{x} , \vec{y} , and \vec{z} and rotation around the diagonal line (red line), γ is angle between \vec{x} , \vec{y} and θ is the angle between \vec{z} and the height (h) **(b)** $F^*(\gamma)$ function



counterclockwise rotating \vec{x} around a diagonal line extending from the origin to the upper-right corner using two different angles, θ'_1 and θ'_2 . Consequently, the volume of the parallelepiped is proportional to the magnitude of \vec{x} and the angle between \vec{x} and the rotation axis (Almardeny et al. 2020).

Another essential concept in the ROD method is the geometric median, which generalizes the median using the appropriate L1 estimator. This parameter remains unaffected by outliers, as it is not skewed by extreme values on either side of the

distribution (e.g., Cheng et al. 2024). For dimensions greater than two, the computation process becomes more complex; therefore, Vardi and Zhang (2000) proposed a modified version of the Weiszfeld algorithm to determine the geometric median (g_m) of a feature space $S = \{X_1, X_2, \dots, X_m\}$ in \mathbb{R}^d :

$$g_m \rightarrow T(g_m) = \left(1 - \frac{\mu(g_m)}{r(g_m)}\right)^+ \tilde{T}(g_m) + \min\left(1, \frac{\mu(g_m)}{r(g_m)}\right)(g_m) \tag{2}$$

where

$$\tilde{T}(g_m) = \left\{ \sum_{X_i \neq g_m} \frac{\mu_i}{\|g_m - X_i\|} \right\}^{-1} \sum_{X_i \neq g_m} \frac{\mu_i X_i}{\|g_m - X_i\|} \quad (3)$$

$$r(g_m) = \|\tilde{R}(g_m)\|, \tilde{R}(g_m) = \sum_{X_i \neq g_m} \frac{\mu_i(X_i - g_m)}{\|X_i - g_m\|} \quad (4)$$

$$\mu(g_m) = \begin{cases} \mu(p), & \text{if } g_m = X_p, p = 1, \dots, m \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

In the equations above, g_m in \mathbb{R}^d represents the geometric median only if it is a fixed point and satisfies the condition $(g_m) \leq \mu(g_m)$. Here, $\mu(g_m)$ indicates a weight variable at g_m , which can be either equal to $\mu(p)$ —the number of data vectors in S that have zero Euclidean distance to g_m during the iteration process—or equal to zero if $\mu(p)=0$. In the latter case, $T(g_m)$ is considered as a weighted average of S , resulting in $T(g_m) = \tilde{T}(g_m)$, as outlined in the Weiszfeld algorithm. It is also noteworthy that g_m is unique when the points are not collinear, and it remains invariant under Euclidean similarity transformations, such as translation, rotation, and reflection (Almardeny et al. 2020).

In a three-dimensional scenario, let $W = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ in \mathbb{R}^3 represent a series of vectors corresponding to data samples. If \vec{m} in \mathbb{R}^3 is the unit geometric median vector of W that defines the rotation axis, it can be demonstrated that for any vector \vec{v} in W that is independent of \vec{m} , the signed volume of the parallelepiped formed by rotating \vec{v} twice around \vec{m} —considering the right-hand rule and with $\theta_1 < \theta_2$ in the range of 0 to 2π —can be approximately represented as a cost function using the Rodrigues rotation formula (Almardeny et al. 2020):

$$F(\vec{v}, \gamma) = \|\vec{v}\|^3 (\cos \gamma \sin \gamma^2) \quad (6)$$

In the above equation, the differences between the vectors in the dataset relate to their magnitudes and the angle γ between \vec{v} and \vec{m} , indicating the degree of deviation from the median vector. For simplicity, $(\cos \gamma \sin \gamma^2)$ is denoted as $F^*(\gamma)$ (Fig. 2b). Since $F^*(\gamma)$ changes periodically between 0 and 2π , it is advisable to restrict the analysis to the range 0 to π , as the trigonometric function is evaluated only for the smallest angle between \vec{v} and \vec{m} . Additionally, for γ in $\{0, \pi/2, \pi\}$, $F^*(\gamma)$ is equal to zero, leading to three cases: $\gamma = F^*(\gamma) = 0$ only if $\vec{v} = \vec{m}$; $\gamma = \pi/2$, where $F^*(\gamma) = 0$ only if $\vec{v} \perp \vec{m}$; and $\gamma = \pi$, where $F^*(\gamma) = 0$ only if $\vec{v} \parallel \vec{m}$. The maximum and minimum values of $F^*(\gamma)$ are approximately 0.385 and -0.385 at the turning points (Fig. 1b), referred to as the threshold angles ω_1 and ω_2 , respectively.

Another characteristic of $F^*(\gamma)$ is that $F^*(\gamma \in [0, \pi/2]) = -F^*(\gamma \in [\pi/2, \pi])$, which ensures the uniqueness of the values by emphasizing points not lying in the same hyperplane as \vec{m} , thereby indicating the orientation of the parallelepiped. To preserve the proportionality of deviation characteristics among the data vectors, $F^*(\gamma)$ is adjusted to ensure it does not approximate γ in $(\omega_1, \pi/2] \cup (\omega_2, \pi]$. The angles are thus scaled down to $(0, \omega_1] \cup (\pi/2, \omega_2]$. Subtracting the geometric median from each vector in the dataset helps reveal the rotation cost of low-magnitude vectors that are collinear with higher-magnitude vectors while maintaining the relative positional relationships among the data points (Fig. 3) (Almardeny et al. 2020).

To determine the rotation outlier score, it is essential to isolate the ROD costs that are inconsistent with other costs (Fig. 3). The median absolute deviation (MAD) is chosen due to its robustness and efficiency (Leys et al. 2013):

$$\begin{aligned} \text{MAD} &= \text{median}(|X_i - \tilde{X}|) \\ M_i &= \frac{0.6745(X_i - \tilde{X})}{\text{MAD}} \end{aligned} \quad (7)$$

where \tilde{X} denotes the median of the dataset, and M_i represents the outlieriness score of the i th data sample. The constant value of 0.6745, introduced by Iglewicz and Hoaglin (1993), is included in M_i because, for a normal-like distribution, M_i approaches 1, as $E(\text{MAD}) = 0.6745$ for large datasets (Almardeny et al. 2020).

In high-dimensional scenarios ($d > 3$), outliers hidden within subspaces of complex manifolds often remain undetected (Müller et al. 2012). Data may originate from various mechanisms across individual dimensions or subsets, suggesting that exploring different perspectives can unveil outliers that remain concealed in a comprehensive attribute space (Kriegel et al. 2012). In high-dimensional feature spaces, authentic outliers may also become obscured due to the curse of dimensionality, which leads to data points becoming too sparse and increases noise levels (Parsons et al. 2004). Additionally, in such contexts, the concept of neighborhood loses its significance (Beyer et al. 1999), making methods reliant on the relative contrast of distances between data points increasingly unreliable (Aggarwal et al. 2001). To address a dataset W in $\mathbb{R}^{n \times d}$, ROD can be applied in the 3D subspaces $\{U_{ij}, \dots, U_{ll}\}$ that emerge from decomposing the entire feature space V into subsets representing different combinations of the feature space $\{s_{i \in I} \mid s_i = \{U_j \mid j \in [J]\}\}$. Consequently, the overall outlier score of V for each sample is determined by integrating the ROD scores from every 3D subspace (Fig. 3) (Almardeny et al. 2020).

To adapt the high-dimensional case to the standard three-dimensional approach, two considerations are made.

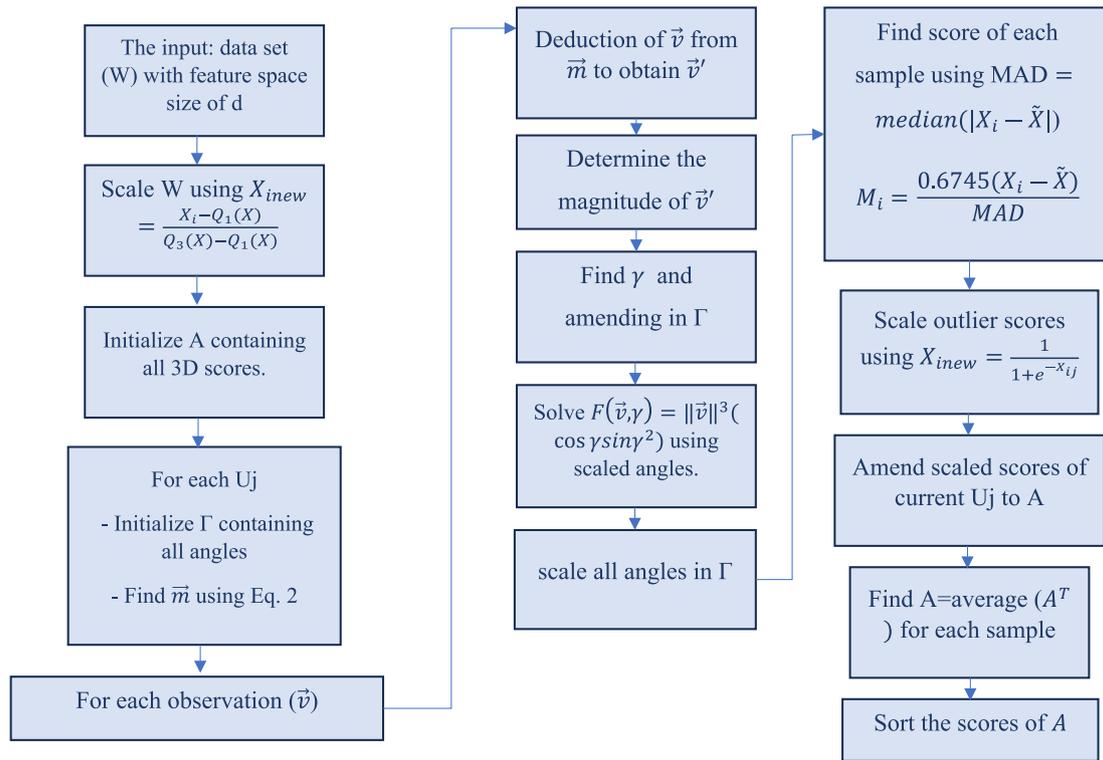


Fig. 3 Algorithm of the ROC approach (Almardeny et al. 2020)

In datasets containing features with high magnitudes, these features can dominate the algorithm, as the ROD cost correlates with vector magnitude for each sample. As a result, the MAD may vary significantly, masking some three-dimensional scores. To mitigate this issue, the data is scaled according to the quantile range, which is robust to outliers (Almardeny et al. 2020):

$$X_{inew} = \frac{X_i - Q_1(X)}{Q_3(X) - Q_1(X)} \tag{8}$$

where X_i represents the i th sample, and $Q_1(X)$ and $Q_3(X)$ denote the first and third quantiles of the dataset. In the second consideration, constraining the MAD within a predefined numerical range helps to correct false weights attributed by the average function to each 3D score. To address this, the results are scaled to the range [0,1] using the following formula:

$$X_{inew} = \frac{1}{1 + e^{-X_{ij}}} \tag{9}$$

In which X_{ij} is the MAD of the i th sample and j th 3D-subspace.

The LOF and KNN methods

The LOF method is a density-driven method which recently used to identify subtle and localized anomalies in geochemical studies (e.g., Puchhammer et al. 2024; Shahrestani and Carranza 2024). This technique evaluates the density of a data point in relation to its neighbors by calculating the local reachable density (ρ_k), which measures the proximity of nearby points (Breunig et al. 2000). An LOF score is calculated by comparing the density of a point to the average density of its k -nearest neighbors. Scores greater than 1 indicate potential outliers due to lower relative densities (Breunig et al. 2000). Unlike global approaches, LOF focuses on local density variations, making it effective for detecting both isolated and clustered anomalies within irregular geochemical distributions. Studies emphasize its effectiveness in complex geological environments where traditional methods often face challenges with uneven mineralization distribution (Shahrestani and Carranza 2024).

The KNN method is a widely utilized approach for outlier detection, particularly suitable for geochemical studies where identifying deviations from expected patterns is important (e.g., Chen et al. 2021c). This technique

determines the distance between a given data point and its k -nearest neighbors to assess its conformity within the dataset. Points with significantly larger distances compared to others are flagged as potential outliers, reflecting their distinctiveness within the data distribution (Angiulli and Pizzuti 2002). KNN is non-parametric, requiring no assumptions about the underlying data distribution, which enhances its versatility in handling diverse datasets. By examining local neighborhood relationships, KNN effectively detects anomalies in both clustered and dispersed datasets (Chen and Shayilan 2022).

Results and discussion

In this research, we utilized the Python implementation of the ROD algorithm (<https://github.com/yzhao062/pyod/blob/master/pyod/models/rod.py>; Almardeny et al. 2020; Zhao et al. 2019b) to determine the outlier scores for all stream sediment samples. Additionally, we calculated the scores from the LOF and KNN methods using the ‘scikit-learn’ Python library. Two criteria were employed to compare the efficiency of different outlier detection methods. In regional geochemical exploration, the spatial locations of known mineral occurrences, along with mineralized samples, can be leveraged to evaluate the efficiency of various outlier detection methods in delineating geochemical anomalies. The first criterion involved constructing the receiver operating characteristic (ROC) curve (Fawcett 2006) using the locations of known mineral occurrences and mineralized samples as true positive values, alongside randomly selected locations within the entire study area as false positive values (Wang and Zuo 2022; Nykänen et al. 2015). A greater area under the curve (AUC) indicates a higher efficiency of the outlier detection methods (e.g., Hastie et al. 2009). The second criterion focused on the number of detected known mineral deposits and mineralized samples in the anomaly classes of the interpolated multivariate geochemical map (e.g., Shahrestani and Sanislav, 2025a). Quantile classification was employed, dividing the interpolated outlier score maps into four quartiles: Q1, Q2, Q3, and Q4. Q1 encompasses the lowest 25% of the data, Q2 the next 25%, Q3 the following 25%, and Q4 the highest 25%. Since different outlier detection methods yield scores with varying distributions, this classification ensures that each quartile contains an equal proportion of data (e.g., Liu et al. 2024; Wang and Zuo 2022). When creating the KNN and LOF models for detecting geochemical anomalies in high-dimensional data, several parameters must be defined. The most critical of these is the value of k , which should be tailored to the characteristics of the dataset. Other parameters can typically be set to their default values. In regions with established data or “ground

truth,” the optimal k can be found through trial and error. However, for unexplored or greenfield areas lacking known mineral occurrences, selecting an appropriate k becomes more challenging (Chen et al. 2021c). Given the sensitivity of LOF and KNN to the number of neighboring samples (k), we calculated the scores of the LOF and KNN methods with varying k values ranging from 10 to 90 (approximately 10% of data points) to identify the k value at which LOF and KNN generate the most effective geochemical maps for capturing known mineralization (Fig. 4). From Fig. 4, two characteristics of the LOF and KNN methods are apparent. First, LOF exhibits greater sensitivity to k values compared to the KNN method, with the AUC significantly increasing as the number of neighbor samples grows. The second observation is the peak AUC value in both outlier detection methods when 70 samples are considered as the size of the neighborhood. Therefore, this optimal k value is selected for comparison of LOF and KNN with the ROD method.

After the optimal neighboring size was selected, the performances of LOF ($k=70$), KNN ($k=70$), and ROD were compared in delineating geochemical anomalies originating from anomalous sources. Figure 5 illustrates the anomaly maps based on outlier scores derived from LOF, KNN, and ROD. As shown in Fig. 6, which presents the ROC curves, a slight superiority of ROD over both LOF and KNN is demonstrated. However, comparable contrasts between hard data (i.e., mineralization) and randomly selected anomalous and non-anomalous points are exhibited in the anomaly maps produced by KNN and ROD. Additionally, mineralized samples were also considered validation points in this analysis (see Fig. 5). The same trend was observed when mineralized samples were used as validation points, revealing higher AUC values for ROD and an increase in the efficiency of LOF compared to ROD and KNN. This suggests that while a slight edge in overall performance is offered by ROD, improved efficacy of LOF is also shown when the validation dataset includes mineralized samples. These findings indicate that valuable insights into the delineation of geochemical anomalies are provided by all three methods, with ROD emerging as a particularly effective approach.

One important characteristic of outlier detection methods is their robustness to the number of dimensions. To examine this feature, two different datasets with varying feature space sizes were selected. In the first group, all trace elements, including Ag, As, Au, Bi, Co, Cr, Cu, Mo, Ni, Pb, Sb, and Zn, were considered. In the second group, Pearson correlation values were first calculated between all clr-transformed trace elements, and one element was selected from those with high correlation values. Consequently, six elements, namely Ag, As, Au, Cu, Mo, and Sb, were chosen for group 2.

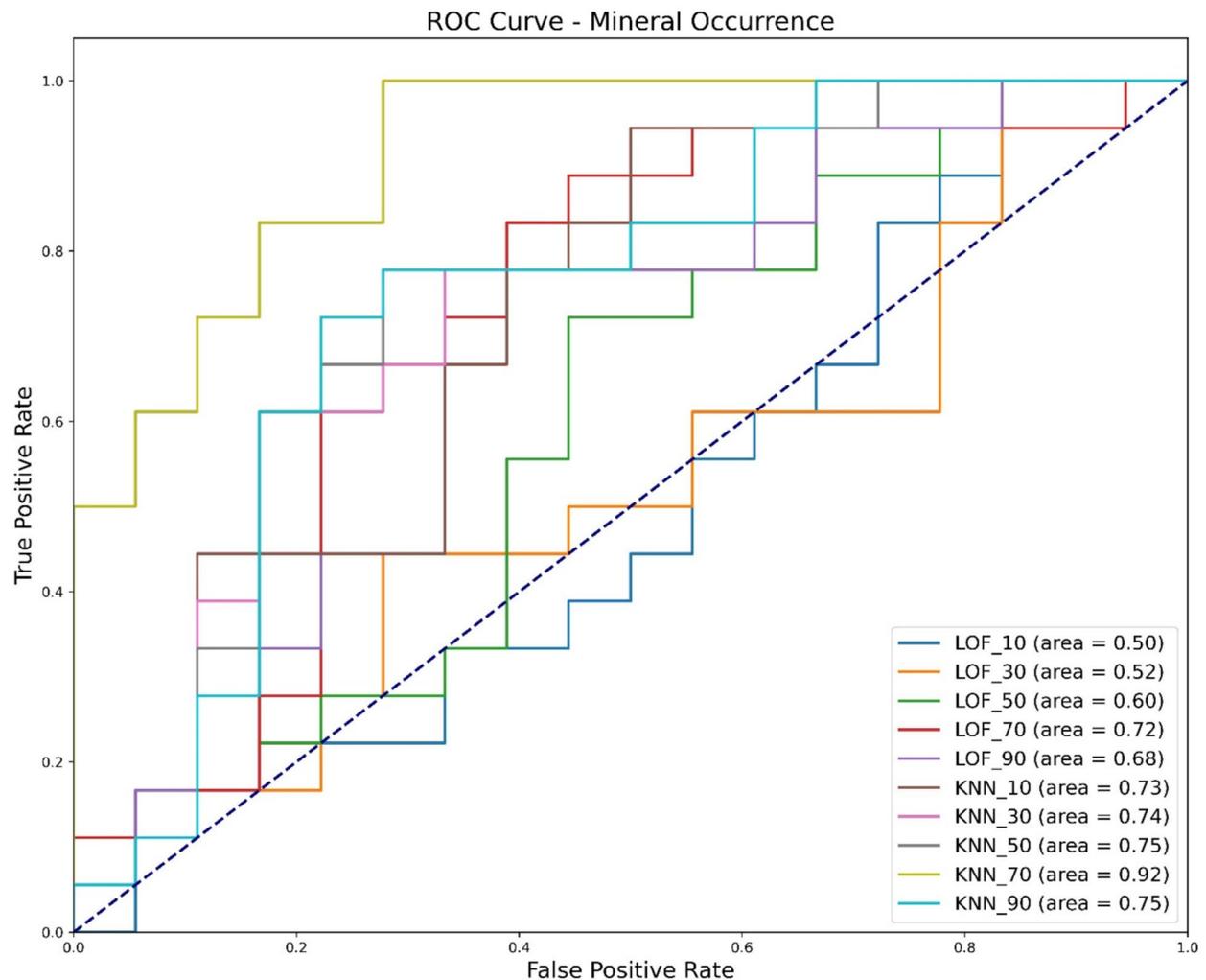


Fig. 4 ROC curves based on LOF and KNN scores considering various number of nearest neighbors implying more sensitivity of LOF relative to KNN

Three outlier detection methods were then applied to both groups of elements. In the context of geochemical exploration based on regional multivariate data, the number of known mineral occurrences in the anomaly class served as a benchmark for comparing the geochemical anomaly maps produced by different anomaly detection methods (see Fig. 5). The classified geochemical maps reveal three primary zones of geochemical anomalies. The first zone is located in the northern and northeastern parts of the study area, encompassing the majority of mineral deposits and aligning with the presence of volcanic rocks in these regions (Fig. 1). The second zone is situated in the western portion of the area and is likely associated with the influence of pyroxene-andesite lava flows and altered mafic and ultramafic rocks. The third zone is in the southeastern part of the study area, corresponding to altered mafic and ultramafic rocks. Table 3 was designed to compare the clr-transformed Cu data, LOF, KNN, and ROC applied to the first and

second groups of elements. The findings in Table 3 reveal that the single-element Cu anomaly map outperforms LOF in delineating known Cu mineral occurrences. Additionally, LOF was found to be more sensitive to the size of the input feature space compared to KNN. Moreover, the ROD outlier detection method demonstrated significant performance in detecting known mineral deposits compared to LOF and KNN across both dataset groups. The results suggest that adding a new dimension (i.e., a geochemical element) to the ROD procedure can enhance the effectiveness of the outlier detection practice. This is particularly relevant in geochemical exploration, where mineral occurrences typically emit unique multivariate geochemical signals that can be differentiated from non-anomalous sources. Consequently, an outlier detection method capable of managing high-dimensional data without compromising performance is of great interest.

Another characteristic of outlier detection methods is the spatial relationships between the highest outlier scores

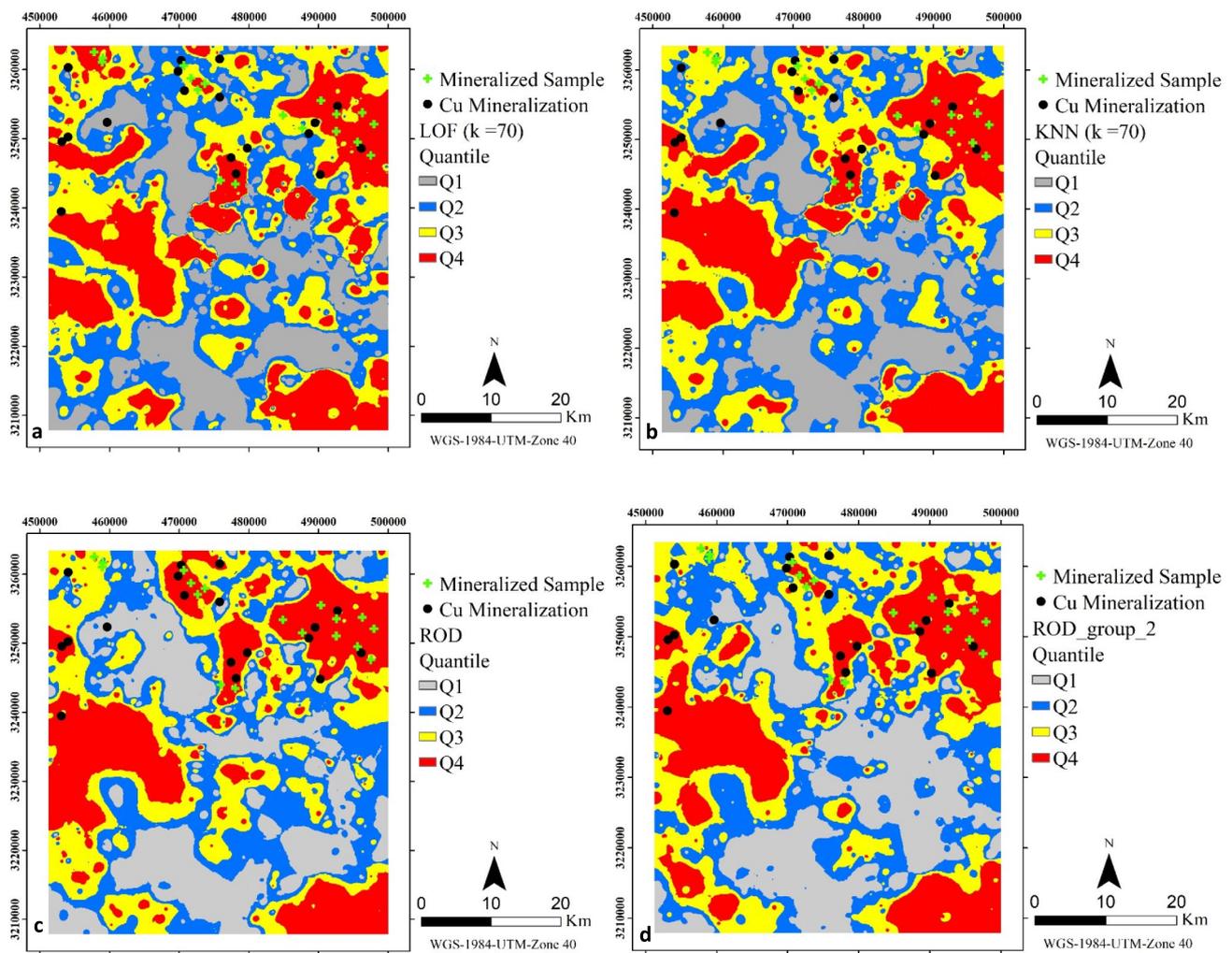


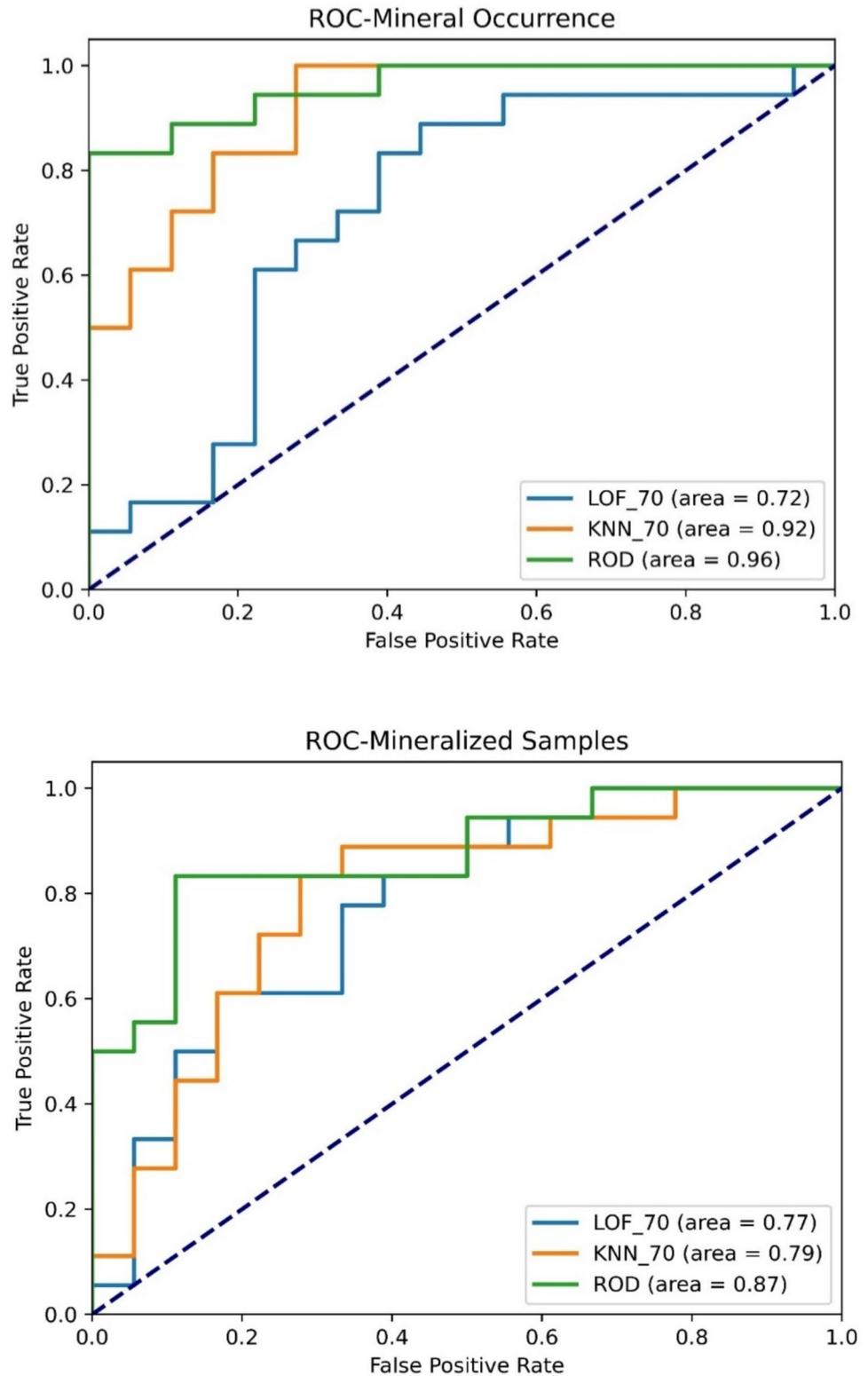
Fig. 5 Classified geochemical anomaly maps based on interpolated outlier scores emerging from (a) LOF, b KNN, c ROD (all trace elements), d ROD (group_2)

and the locations of validation points, which in this case are known mineral occurrences and mineralized samples. Figure 7 illustrates the locations of the top 10% of scores obtained from LOF, KNN, and ROD. Five zones are highlighted in Fig. 6 to facilitate the comparison of the spatial distribution of top outlier scores from the three methods. In zone 1, the top scores from KNN and ROD demonstrate spatial proximity to known Cu mineralization, with this proximity being more pronounced in ROD compared to KNN. In zone 2, only one top score stream sediment sample is observed relative to mineralization for both LOF and KNN, while no top score sample is identified in ROD. In zone 3, similar performance is noted across all three outlier detection methods. In zone 4, distinct patterns emerge for KNN and LOF in comparison to ROD, where, similar to zone 1, the top scores from ROD are closer to mineralization compared to those from KNN and LOF. In Zone

5, the absence of validation points makes it impractical to assume the superiority of any particular method.

The LOF and KNN methods are widely utilized in geochemical anomaly detection, particularly in mineral exploration. LOF, a density-based algorithm, is highly effective at identifying local anomalies within heterogeneous geological datasets by leveraging spatial neighborhood relationships (Puchhammer et al. 2024). Unlike global anomaly detection techniques such as isolation forest (iForest), LOF excels in analyzing stream sediment geochemical data with varying concentration levels, as it can detect deviations from localized geochemical patterns that might otherwise be overlooked (Puchhammer et al. 2024). Another advantage of LOF is its independence from covariance estimation, as it strictly relies on Euclidean distances, allowing for the application of isometric transformations in compositional data analysis (Puchhammer et al. 2024). Similarly, KNN, which

Fig. 6 ROC curves based on LOF, KNN, and ROD scores demonstrating the efficiency of ROD over LOF and KNN considering (up) mineral occurrences, and (down) mineralized samples



forms the basis of LOF, is a non-parametric method that identifies anomalies by comparing each sample to its closest neighbors. Its strength lies in its ability to capture complex nonlinear patterns in geochemical datasets without requiring

distributional assumptions, making it useful in both supervised and unsupervised anomaly detection scenarios (Chen et al. 2021c). Additionally, KNN is highly intuitive and can be adapted to different geological settings, particularly

Table 3 Number of mineral occurrences detected by single-variate clr-transformed Cu data, LOF, KNN, and ROD anomaly maps

Method	A	B	C
Cu univariate anomaly map	Q1	3	16.7
	Q2	1	5.6
	Q3	4	22.2
	Q4	10	55.6
LOF_70	Q1	1	5.6
	Q2	4	22.2
	Q3	8	44.4
	Q4	5	27.8
LOF_70_group_2	Q1	1	5.6
	Q2	3	16.7
	Q3	7	38.9
	Q4	7	38.9
KNN_70_	Q1	0	0.0
	Q2	3	16.7
	Q3	5	27.8
	Q4	10	55.6
KNN_70_group_2	Q1	0	0.0
	Q2	3	16.7
	Q3	6	33.3
	Q4	9	50.0
ROD	Q1	1	5.6
	Q2	1	5.6
	Q3	2	11.1
	Q4	14	77.8
ROD_group_2	Q1	1	5.6
	Q2	1	5.6
	Q3	4	22.2
	Q4	12	66.7

A: Threshold, B: Number of deposits in each class, C: Relative number of occurrences (%) in each class

where geochemical anomalies exhibit localized clustering (Puchhammer et al. 2024).

Despite their advantages, both LOF and KNN have significant limitations. A major drawback of LOF is its reduced effectiveness in high-density sampling areas, where geochemical values transition gradually toward mineralized zones. In such cases, samples located directly over mineralized deposits may not be flagged as anomalies, limiting the method applicability in certain geological environments (Puchhammer et al. 2024). Furthermore, LOF performance is highly sensitive to the choice of the nearest-neighbor parameter (k), with optimal results typically achieved when the number of input variables is limited to five to ten. Beyond this range, the reliability of distance-based calculations decreases due to the curse of dimensionality, which affects both LOF and KNN (Shahrestani and Sanislav 2025b). The

selection of k is particularly challenging in greenfield exploration, where no known mineral occurrences exist, making parameter optimization difficult (Chen et al. 2021c). Even in well-studied regions, determining the appropriate k value often requires a trial-and-error approach, relying on available ground truth data (Chen et al. 2021c). Additionally, the KNN framework underlying LOF is computationally intensive for large datasets, as it requires calculating pairwise distances between all samples, making it less efficient for regional-scale geochemical surveys (Aggarwal and Sathe 2017).

Based on the results, ROD is a promising outlier detection method for geochemical anomaly detection tasks due to apparent advantages over some well-established outlier detection methods including LOF and KNN. In geochemical exploration, where datasets often include multiple elements and complex relationships due to various sources of variability such as lithology or mineralization, ROD has demonstrated its effectiveness in improving the accuracy of anomaly detection. This research reveals a key strength of ROD to handle high-dimensional data. Geochemical datasets typically contain multiple elements that which of which may partially explain the geochemical patterns associated with mineralization. ROD performs well when additional dimensions are introduced. Unlike LOF and KNN, which are sensitive to high-dimensional data, ROD uses subspaces to minimize the curse of dimensionality. This makes ROD a proper candidate for geochemical exploration, where mineral occurrences release unique geochemical signals that may only be identifiable when the relationships between elements are considered. By applying rotation-based detection, ROD can better capture these complex signals and avoid the performance issues LOF and KNN encounter in higher-dimensional spaces.

Another advantage of ROD is its rotation cost function, which is particularly well-suited to detecting anomalies in geochemical data. The method uses a rotation-based approach to analyze the geometric relationships between samples, detecting anomalies based on their orientation in subspaces rather than relying solely on distance or density. This makes ROD highly effective in identifying geochemical outliers, where anomalies may not always be distinguished by their proximity to neighboring samples but rather by their overall deviation from expected geochemical patterns. ROD is less dependent on subjective parameters such as neighborhood size (k), which often require careful tuning in methods like LOF and KNN. The independence of ROD from this parameter provides a practical advantage over LOF and KNN, both of which require careful selection of k to achieve optimal results, particularly in the case of LOF. Furthermore, ROD, uses MAD in its evaluation of rotation costs, which enhances its ability to identify true anomalies while reducing the likelihood of classifying noisy points

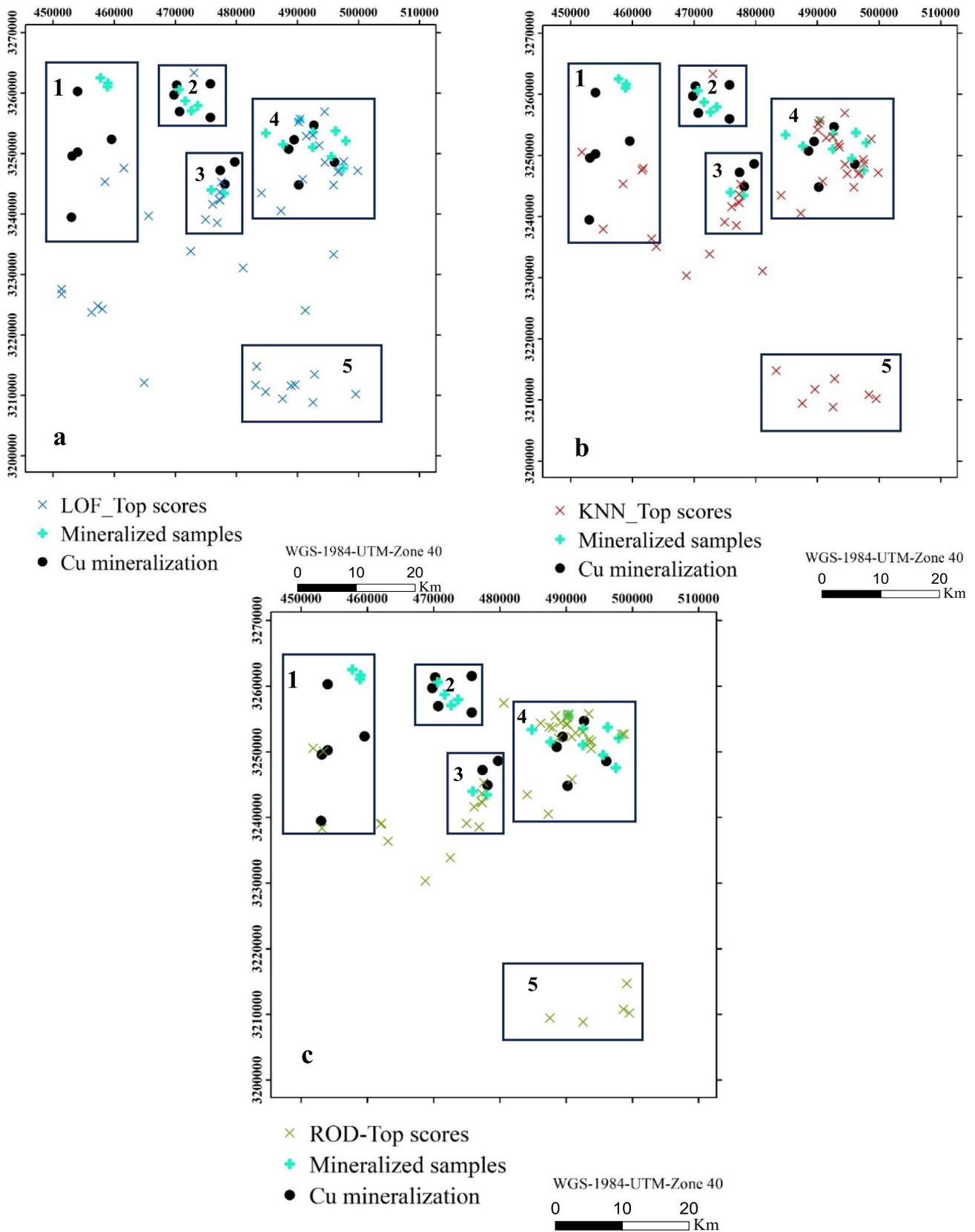


Fig. 7 Spatial distribution of top 10% outlier scores emerging from LOF (a), KNN (b), and ROD (c) relative to known Cu mineralization

as outliers. The ROD method has potential limitations in identifying geochemical anomalies. Its high computational complexity arises from breaking the entire attribute space into multiple 3D subspaces, making it inefficient as dimensionality increases, leading to long processing times with high-dimensional datasets. Additionally, ROD may struggle to detect anomalies that span across multiple dimensions since it focuses on localized subspaces rather than the overall dataset structure. While it does not require assumptions about statistical distributions, its reliance on the geometric median for rotation makes it sensitive to changes in data distribution, which can affect performance in skewed or multimodal datasets. Furthermore, selecting a limited number of 3D subspaces to enhance efficiency could reduce precision, resulting in missed anomalies if the chosen subspaces are not appropriate.

The ROD method is effective in high-dimensional geochemical datasets but does not explicitly consider spatial relationships or compositional associations between elements. While ROD excels at identifying statistical outliers by focusing on geometric relationships in feature space, it fails to account for the spatial context of the data or the geological processes influencing geochemical distributions. Consequently, ROD may struggle to distinguish between anomalies linked to geological processes and those arising from noise or data inconsistencies. However, identifying geochemical anomalies requires consideration of both spatial patterns and elemental associations, as these factors are crucial in differentiating mineralization-related anomalies from background variations (Zuo and Xu 2024). To address this need, a novel methodology involving a dual-branch deep learning model is proposed, which integrates a graph convolutional autoencoder (GCN-AE) for spatial feature extraction and a long short-term memory autoencoder (LSTM-AE) for spectral feature analysis. This approach captures both spatial relationships among geochemical samples and compositional variations within individual samples (Xu et al. 2024). In contrast, while ROD is effective in high-dimensional datasets, its lack of explicit incorporation of spatial context may limit its sensitivity to mineralization patterns that are spatially controlled.

Conclusions

This study highlights the distinct advantages of the ROD method in geochemical anomaly detection, showcasing its effectiveness compared to traditional techniques like LOF and KNN. ROD consistently demonstrates superior performance in high-dimensional datasets common in geochemical exploration, where complex interactions among multiple elements, influenced by factors such as lithology and mineralization, present significant challenges for conventional

methods. By addressing the curse of dimensionality through subspace analysis, ROD maintains strong performance even as additional variables are introduced, excelling in the identification of subtle geochemical signals associated with mineralization that other methods may overlook. A key advantage of ROD is its enhanced sensitivity, which allows for better detection of mineral occurrences by focusing on the intricate relationships between elements. This capability is evident in its superior AUC values and improved detection of known mineral occurrences in anomaly maps and ROC curves, effectively delineating anomalous areas while minimizing false positives. Such performance highlights the potential of ROD for advancing geochemical exploration. Despite its promising performance, there are opportunities for further improvement in the ROD method. One potential enhancement could be optimizing the selection of subspaces informed by geochemical domain knowledge. This approach may refine the identification of outliers associated with specific mineralization types, leading to improved accuracy. Additionally, future work aims to utilize ROD on a limited number of 3D subspaces of interest. This strategy is designed to speed up the running time of method while maintaining its precision, thereby enhancing its practicality for geochemical analysis.

Acknowledgements The authors present their gratitude to the Geological Survey of Iran for generously providing the regional stream sediment geochemical data from the Baft area, which made a significant contribution to this study.

Author Contribution S.S.—conceptualized and wrote the first draft and prepared figures and tables I.V.S. —supervision, draft review and funding acquisition H.F. —geological data, draft review, figures.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This research received no external funding.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afzal P, Alghalandis YF, Khakzad A, Moarefvand P, Omran NR (2011) Delineation of mineralization zones in porphyry Cu deposits by fractal concentration–volume modeling. *J Geochem Explor* 108(3):220–232
- Afzal P, Tehrani ME, Ghaderi M, Hosseini MR (2016) Delineation of supergene enrichment, hypogene and oxidation zones utilizing staged factor analysis and fractal modeling in Takht-e-Gonbad porphyry deposit, SE Iran. *J Geochem Explor* 161:119–127
- Aggarwal CC, Hinneburg A, Keim DA (2001) On the surprising behavior of distance metrics in high dimensional space. In: *Database Theory—ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings* 8. Springer Berlin Heidelberg, pp 420–434. https://doi.org/10.1007/3-540-44503-X_27
- Aggarwal CC, Sathe S (2017) *Outlier Ensembles: An Introduction*, 1st edn. Springer International Publishing AG, Springer Cham
- Aghazadeh M, Hou Z, Badrzadeh Z, Zhou L (2015) Temporal–spatial distribution and tectonic setting of porphyry copper deposits in Iran: constraints from zircon U–Pb and molybdenite Re–Os geochronology. *Ore Geol Rev* 70:385–406
- Aitchison J (1986) *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK) (Reprinted in 2003 with additional material by The Blackburn Press, London (UK), p 416
- Almardeny Y, Boujnah N, Cleary F (2020) A novel outlier detection method for multivariate data. *IEEE Trans Knowl Data Eng* 34(9):4052–4062
- Aminzadeh B, Rahimpour Y (2021) Geology, geochemistry and genesis of the Godar Sabz Mn deposit in the Baft region. *Iran Geochemistry* 81(4):125827
- Angiulli F, Pizzuti C (2002) August. Fast outlier detection in high dimensional spaces. In *European conference on principles of data mining and knowledge discovery*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp 15–27. https://doi.org/10.1007/3-540-45681-3_2
- Bagheri S, Stampfli GM (2008) The Anarak, Jandaq and Posht-e-Badam metamorphic complexes in central Iran: new geological data, relationships and tectonic implications. *Tectonophysics* 451(1–4):123–155
- Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999) When is “nearest neighbor” meaningful?. In *Database Theory—ICDT’99: 7th International Conference Jerusalem, Israel, January 10–12, 1999 Proceedings* 7. Springer Berlin Heidelberg, pp 217–235
- 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
- Breunig MM, Kriegel HP, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. pp 93–104
- Cao M, Yin D, Zhong Y, Lv Y, Lu L (2023) Detection of geochemical anomalies related to mineralization using the Random Forest model optimized by the Competitive Mechanism and Beetle Antennae Search. *J Geochem Explor* 249:107195
- Chen Y (2015) Mineral potential mapping with a restricted Boltzmann machine. *Ore Geol Rev* 71:749–760
- Chen H, Murray AF (2003) Continuous restricted Boltzmann machine with an implementable training algorithm. *IEE Proc-Vis, Image Signal Process* 150(3):153–158
- Chen Y, Shayilan A (2022) Dictionary learning for multivariate geochemical anomaly detection for mineral exploration targeting. *J Geochem Explor* 235:106958
- Chen Y, Sui Y (2022) Dictionary learning for integration of evidential layers for mineral prospectivity modeling. *Ore Geol Rev* 141:104649
- Chen Y, Wu W (2017) Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochem: Explor, Environ, Anal* 17(3):231–238
- Chen Y, Wu W (2019) Isolation forest as an alternative data-driven mineral prospectivity mapping method with a higher data-processing efficiency. *Nat Resour Res* 28(1):31–46
- Chen Y, Lu L, Li X (2014) Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *J Geochem Explor* 140:56–63
- Chen Y, Wu W, Zhao Q (2019) A bat-optimized one-class support vector machine for mineral prospectivity mapping. *Minerals* 9(5):317
- Chen Y, Sun G, Zhao Q (2021a) Detection of multivariate geochemical anomalies associated with gold deposits by using distance anomaly factors. *J Geochem Explor* 221:106704
- Chen Y, Wang S, Zhao Q, Sun G (2021b) Detection of multivariate geochemical anomalies using the bat-optimized isolation forest and bat-optimized elliptic envelope models. *Journal of Earth Science* 32(2):415–426
- Chen Y, Zhao Q, Lu L (2021c) Combining the outputs of various k-nearest neighbor anomaly detectors to form a robust ensemble model for high-dimensional geochemical anomaly detection. *J Geochem Explor* 231:106875
- Cheng G, Xiong Q, Lin R (2024) Online bootstrap inference for the geometric median. *Comput Stat Data Anal* 197:107992
- Dimitrijevic MD (1973) *Geology of Kerman region*. Geological Survey of Iran Report YU/52, p 247
- Ding L, Chen B, Zhu Y, Dong H, Chan G, Zhang P (2024) Geo-Hgan: Unsupervised anomaly detection in geochemical data via latent space learning. *Comput Geosci* 192:105703
- Doersch C (2016) Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908* (<https://arxiv.org/abs/1606.05908>)
- Esmailoghli S, Tabatabaei SH, Carranza EJM (2023) Infomax-based deep autoencoder network for recognition of multi-element geochemical anomalies linked to mineralization. *Comput Geosci* 175:105341
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874
- Ghannadpour SS, Hezarkhani A (2022) Delineation of geochemical anomalies for mineral exploration using combining U-statistic method and fractal technique: UN and UA models. *Appl Earth Sci* 131(1):32–48
- Ghasemi A, Talbot CJ (2006) A new tectonic scenario for the Sanandaj-Sirjan Zone (Iran). *J Asian Earth Sci* 26(6):683–693
- Ghasemzadeh S, Maghsoudi A, Yousefi M (2019) Application of geometric average approach for Cu-porphyry prospectivity mapping in the Baft area, Kerman. *Sci Q J Geosci* 29(113):231–230
- Ghasemzadeh S, Maghsoudi A, Yousefi M, Mihalasky MJ (2022) Information value-based geochemical anomaly modeling: A statistical index to generate enhanced geochemical signatures for mineral exploration targeting. *Appl Geochem* 136:105177
- Guo M, Chen Y (2024) A SMOTified-GAN-augmented bagging ensemble model of extreme learning machines for detecting geochemical anomalies associated with mineralization. *Geochemistry* 84(4):126156
- Hajihosseini M, Maghsoudi A, Ghezelbash R (2024a) Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches. *J Geochem Explor* 258:107393
- Hajihosseini M, Maghsoudi A, Ghezelbash R (2024b) A comprehensive evaluation of OPTICS, GMM and K-means clustering methodologies for geochemical anomaly detection connected with sample catchment basins. *Geochemistry* 126094

- Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction. New York: Springer, Vol. 2, pp 1–758
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Huang D, Zuo R, Wang J (2022) Geochemical anomaly identification and uncertainty quantification using a Bayesian convolutional neural network model. *Appl Geochem* 146:105450
- Iglewicz B, Hoaglin DC (1993) How to detect and handle outliers. ASQ Press, Milwaukee, WI
- Jamali H (2017) The behavior of rare-earth elements, zirconium and hafnium during magma evolution and their application in determining mineralized magmatic suites in subduction zones: constraints from the Cenozoic belts of Iran. *Ore Geol Rev* 81:270–279
- Kriegel HP, Schubert M, Zimek A (2008) Angle-based outlier detection in high-dimensional data. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 444–452
- Kriegel HP, Kröger P, Schubert E, Zimek A (2012) Outlier detection in arbitrarily oriented subspaces. In: 2012 IEEE 12th international conference on data mining. IEEE, pp 379–388
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Ley S, Ley C, Klein O, Bernard P, Licata L (2013) Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Soc Psychol* 49(4):764–766
- Li Z, Zhao Y, Hu X, Botta N, Ionescu C, Chen GH (2022) ECOD: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Trans Knowl Data Eng* 35(12):12181–12193
- Li S, Zhao Z, Miao T, Li X, Yu S, Zhang Y, Wei J, Zou K (2024) Discrimination of Pb-Zn deposit types using the trace element data of galena based on deep learning. *Ore Geol Rev* 170:106133
- Liu H, Zhang S, Wu Z, Li X (2025) Outlier detection using local density and global structure. *Pattern Recogn* 157:110947
- Liu Y, Xia Q, Duan J, Dai J, Wu S, Zhao Z (2024) Geochemical anomalies of critical metals in the Eastern Kunlun Orogenic Belt, China: Implications for nickel and cobalt mineral exploration. *Ore Geol Rev* 106168
- Luo Z, Zuo R, Xiong Y, Zhou B (2023) Metallogenic-factor variational autoencoder for geochemical anomaly detection by ad-hoc and post-hoc interpretability algorithms. *Nat Resour Res* 32(3):835–853
- Masci J, Meier U, Cireşan D, Schmidhuber J (2011) Stacked convolutional auto-encoders for hierarchical feature extraction. In: Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14–17, 2011, Proceedings, Part I 21. Springer Berlin Heidelberg, pp 52–59
- Masoodi M, Yassaghi A, Sadat MAAN, Neubauer F, Bernroider M, Friedl G, Genser J, Houshmandzadeh A (2013) Cimmerian evolution of the Central Iranian basement: Evidence from metamorphic units of the Kashmar-Kerman Tectonic Zone. *Tectonophysics* 588:189–208
- Moghadam MC, Tahmasbi Z, Ahmadi-Khalaji A, Santos JF (2018) Petrogenesis of Rabor-Lalehzar magmatic rocks (SE Iran): Constraints from whole rock chemistry and Sr-Nd isotopes. *Geochemistry* 78(1):58–77
- Mohebi A, Mirnejad H, Lentz D, Behzadi M, Dolati A, Kani A, Taghizadeh H (2015) Controls on porphyry Cu mineralization around Hanza Mountain, south-east of Iran: An analysis of structural evolution from remote sensing, geophysical, geochemical and geological data. *Ore Geol Rev* 69:187–198
- Müller E, Assent I, Iglesias P, Mülle Y, Böhm K (2012) Outlier ranking via subspace analysis in multiple views of the data. In: 2012 IEEE 12th international conference on data mining. IEEE, pp 529–538
- Najafian T, Mokhtari AR, Shahrestani S, Albanese S (2023) Is the pathway length of sediments relevant to assess the background value in stream sediment geochemical exploration? *J Geochem Explor* 253:107278
- Natašin BA, Şengör AC (2005) Late Palaeozoic to Triassic evolution of the Turan and Scythian platforms: the pre-history of the Palaeo-Tethyan closure. *Tectonophysics* 404(3–4):175–202
- Nykänen V, Lahti I, Niiranen T, Korhonen K (2015) Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geol Rev* 71:853–860
- Park CH (2023) A comparative study for outlier detection methods in high dimensional text data. *J Artif Intell Soft Comput Res* 14(1):23–38
- Parsons L, Haque E, Liu H (2004) Subspace clustering for high dimensional data: a review. *ACM SIGKDD Explor Newsl* 6(1):90–105
- Puchhammer P, Kalubowila C, Braus L, Pospiech S, Sarala P, Filzmoser P (2024) A performance study of local outlier detection methods for mineral exploration with geochemical compositional data. *J Geochem Explor* 258:107392
- Rayana S, Zhong W, Akoglu L (2016) Sequential ensemble learning for outlier detection: A bias-variance perspective. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). IEEE, pp 1167–1172
- Richards JP, Spell T, Rameh E, Razique A, Fletcher T (2012) High Sr/Y magmas reflect arc maturity, high magmatic water content, and porphyry Cu±Mo±Au potential: Examples from the Tethyan arcs of central and eastern Iran and western Pakistan. *Econ Geol* 107(2):295–332
- Safari HO, Bagas L, Bafti BS (2015) Structural controls on the localization of Cu deposits in the Kerman Cu metallogenic province of Iran using geoinformatic techniques. *Ore Geol Rev* 67:43–56
- Schubert E, Zimek A, Kriegel HP (2014) Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. *Data Min Knowl Disc* 28:190–237
- Shafiei B (2010) Lead isotope signatures of the igneous rocks and porphyry copper deposits from the Kerman Cenozoic magmatic arc (SE Iran), and their magmatic-metallogenetic implications. *Ore Geol Rev* 38(1–2):27–36
- Shahrestani S, Sanislav I (2025a) Mapping geochemical anomalies using angle-based outlier detection approach. *J Geochem Explor* 269:107635
- Shahrestani S, Sanislav I (2025b) How does dimensionality influence outlier detection effectiveness in multivariate geochemical data? insights from LOF and IF methods. *Earth Sci Inf* 18(1):1–15
- Shahrestani S, Mokhtari AR, Izadi M (2024) Enhancing geochemical background estimation using rock weathering. *Ore Geol Rev* 169:106063
- Shahrestani S, Carranza EJM (2024) Effectiveness of LOF, iForest, and OCSVM in detecting anomalies in stream sediment geochemical data. *Geochemistry: Exploration, Environment, Analysis geochem* 2024–009
- Shayilan A, Chen Y (2024) A SMOTified extreme learning machine for identifying mineralization anomalies from geochemical exploration data: a case study from the Yeniugou area, Xinjiang, China. *Earth Sci Inform* 17(2):1329–1343
- Soltani Z, Hassani H, Esmailoghli S (2024) A deep autoencoder network connected to geographical random forest for spatially aware geochemical anomaly detection. *Comput Geosci* 190:105657

- Srdic A, Dimitrijevic MN, Cvetic S, Dimitrijevic MD (1972) Geological Map of Baft. 1/100000 Series, Sheet 7348. Geological Survey of Iran
- Stampfli GM, Borel GD (2002) A plate tectonic model for the Paleozoic and Mesozoic constrained by dynamic plate boundaries and restored synthetic oceanic isochrons. *Earth Planet Sci Lett* 196(1–2):17–33
- Thompson M, Howarth RJ (1976) Duplicate analysis in geochemical practice. Part I. Theoretical approach and estimation of analytical reproducibility. *Analyst* 101(1206):690–698
- Vardi Y, Zhang CH (2000) The multivariate L 1-median and associated data depth. *Proc Natl Acad Sci* 97(4):1423–1426
- Wang X, Chen Y (2025) Unsupervised detection of multivariate geochemical anomalies using a high-performance deep autoencoder Gaussian mixture model. *J Geochem Explor* 271:107671
- Wang J, Zuo R (2022) Model averaging for identification of geochemical anomalies linked to mineralization. *Ore Geol Rev* 146:104955
- Wang Z, Zuo R, Yang F (2023) Geological mapping using direct sampling and a convolutional neural network based on geochemical survey data. *Math Geosci* 55(7):1035–1058
- Xiong Y, Zuo R, Luo Z, Wang X (2022) A physically constrained variational autoencoder for geochemical pattern recognition. *Math Geosci* 1–24
- Xiong Y, Zuo R (2016) Recognition of geochemical anomalies using a deep autoencoder network. *Comput Geosci* 86:75–82
- Xu Y, Shi L, Zuo R (2024) Geologically constrained unsupervised dual-branch deep learning algorithm for geochemical anomalies identification. *Appl Geochem* 174:106137
- Zhang C, Zuo R (2021) Recognition of multivariate geochemical anomalies associated with mineralization using an improved generative adversarial network. *Ore Geol Rev* 136:104264
- Zhang ZJ, Zuo RG, Xiong YH (2021) Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. *Appl Geochem* 130:104994
- Zhao Y, Nasrullah Z, Li Z (2019b) Pyod: A python toolbox for scalable outlier detection. *J Mach Learn Res* 20(96):1–7
- Zhao B, Zhang D, Tang P, Luo X, Wan H, An L (2023) Recognition of multivariate geochemical anomalies using a geologically-constrained variational autoencoder network with spectrum separable module—A case study in Shangluo District, China. *Appl Geochem* 156:105765
- Zhao Y, Nasrullah Z, Hryniewicki MK, Li Z (2019a) LSCP: Locally selective combination in parallel outlier ensembles. In: *Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*. pp 585–593. <https://doi.org/10.1137/1.9781611975673.66>
- Zimek A, Schubert E, Kriegel HP (2012) A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat Anal Data Min: the ASA Data Science Journal* 5(5):363–387
- Zuo R, Xu Y (2023) Graph deep learning model for mapping mineral prospectivity. *Math Geosci* 55(1):1–21
- Zuo R, Xu Y (2024) A physically constrained hybrid deep learning model to mine a geochemical data cube in support of mineral exploration. *Comput Geosci* 182:105490
- Zuo R, Xiong Y, Wang Z, Wang J, Kreuzer OP (2023) A new generation of artificial intelligence algorithms for mineral prospectivity mapping. *Nat Resour Res* 32(5):1859–1869

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.