



OPEN Analysis of closed and open numerical systems of geochemical data in spatial statistics environment in order to separate anomalous areas

Mirmahdi Seyedrahimi-Niaraq^{1✉} & Hossein Mahdianfar²

Geochemical data are expressed in closed numerical systems due to their non-normality and the presence of outliers. The specificity of such data makes it challenging to analyze them using standard statistical techniques. The U-modeling of log-transformed data represents a novel approach to geochemical anomaly separation. This method models the geochemical open or log-transformed data by the U-spatial statistics algorithm and has been used for the first time in this paper. In this research, Additive and Centered Logarithmic Transformations (ALR and CLR) were applied to data from the Doostbiglou region in Ardabil province, Iran, known for its copper-gold and molybdenum mineralization. After transforming the data into an open numerical system, the correlation between elements was calculated for both systems to compare the results. The output data were modeled using the U-spatial statistics method, and anomaly maps were subsequently generated. Validation and comparison of the results, considering field data obtained from the local and regional exploration, revealed that both models produced similar results in separating anomalous areas and showed a high degree of agreement with field data. However, the U-modeling of the ALR data more closely aligns with field observations and provides a more precise representation of the mineralization trend. Therefore, these new models are recommended for evaluating the spatial distribution of elements and determining the threshold value.

Keywords Closed and open numerical system, Additive logarithmic transformation, Centered logarithmic transformation, U-spatial statistics, Geochemical anomalous areas, Copper-gold and molybdenum mineralization

Chemical analysis of geochemical samples taken from different environments leads to the production of geochemical data, and the study of these data to identify geochemical anomalies is one of the essential issues in mineral resource exploration^{1–4}. In general, the quality and accuracy of the processing results performed by data mining methods depend on the quality of the data. Raw data is usually contains missing, noisy, inconsistent, incomplete, and outlier data. It is essential to perform preprocessing operations on the data because it is a crucial step that leads to improving the quality of raw data and increasing the efficiency of the results of the data processing stage. The data preprocessing step is the data preparation step to extract the most information in the final processing step⁵. Based on previous studies, it has been proven that statistical processing on log-transformed geochemical data produces more valuable results. Simply put, when a logarithmic transformation algorithm is applied to the data and then statistical processing is performed on them, better results and information about the distribution of geochemical elements in the study area are obtained^{3,6–8}.

The three most significant log-ratio transformations available for altering composition data are the Additive Log-Ratio Transformation (ALR), Centered Log-Ratio Transformation (CLR), and Isometric Log-Ratio Transformation (ILR). Because the logarithmic equations' coefficients are random variables that fluctuate between $-\infty$ and $+\infty$, multivariate statistical analysis can be carried out with ease^{9–12}. The process of ALR involves selecting one variable from the set of accessible variables, dividing the remaining variables by it, and

¹Department of Mining Engineering, Faculty of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran.

²Department of Mining Engineering, University of Gonabad, Gonabad, Iran. ✉email: M.seyedrahimi@uma.ac.ir

then determining their logarithm. Consequently, the variable will be eliminated from the dataset, but the system will still be open. Without reference to particular theories, this approach depends on the selection of the dividing variable, which is based on the experience and viewpoint of the individual^{13–16}. It is important to choose this variable carefully because there will be a spurious association if it is the rock-forming variable. Each variable's first logarithm is specified by the CLR technique, and this is then split into the variables' geometric mean; all variables must have the same unit. Although this procedure does not eliminate any variables from the dataset, it has a weakness in multivariate statistical analysis since the covariance matrix of the variables is irreversible. Therefore, this approach can be chosen over the earlier one^{17–20}. Due to its geometrical qualities, the inverse covariance matrix in the ILR can be determined and is preferred; however, this method is more complex than the other two methods, and there are various rules. D-1 variables from the primary D-dimensional space represent the method's output. This transformation reduces the dimensionality of the data, which makes the output hard to understand and prevents it from being used in the univariate analysis^{15,21–24}. Although logarithmic modifications frequently lessen skewness and bring the data closer to a normal distribution, they do not affect the data's composition¹³. Some geochemical study techniques, such as the U-statistics method, employ the distance between samples as a weight to separate anomalous from the background²⁵. With this procedure, the defined radius is used to calculate a range of different U-values. The concept of the maximum $|U|$ value is essential for optimizing anomaly separation and choosing the optimal U-value^{26–31}. Based on the U-histogram, the number of mineralization phases and geological changes that caused a particular concentration fluctuation is identified.

In this Article, by modeling the output data transformed to ALR and CLR methods with the structural method of U- spatial statistics, a new integrated method was introduced to find exploration targets in the gold-molybdenum and copper mineralization type. For this purpose, after placing the data in an open numerical system, the correlation between elements was also studied for these two systems in order to compare the results. The main objective of this research is to use the advantages of ALR and CLR algorithms to open exploration data and model them in a spatial statistics environment to optimally separate anomalous areas from the geochemical background and improve the results.

Geological setting

The study region is located within Azerbaijan's structural zone. The majority of its components are volcanic rocks from the Eocene. In the east of Ardabil Province, the Doostbiglou region is located approximately northwest of city of Meshginshahr. This region is situated in the Azerbaijan-Lesser Caucasus metallogenic belt³² and the western Alborz-Azerbaijan structural zone (Fig. 1), as per the zoning paradigm in Iranian geology³³. Pre-Cretaceous rock formations are often not extruded in the region, and the only sedimentary rocks found here are Quaternary travertine, formed in the mouths of hot springs which are widely distributed.

Magmatic activity in this region started at the end of the Cretaceous with the eruption of compounds in the andesite to trachy-andesite range, and continued with the granitoid intrusive intrusion. Following this period, with a brief break, subsequent eruptions resulted in the formation of the Sabalan volcanic system³⁵. Potassic rocks were found in the distant distances near the city of Kaleybar as intrusive masses of Nephelin Syenite and more (near the Urumieh Lake) on Saray Island extruded³⁶. The main reason for the presence of Cretaceous volcanics in this region is the tension status occurrence, which has dominated the area since the Laramide orogeny in the late Cretaceous³⁷.

Geological research in the region has revealed some intricate connections among the outcrops. Generally speaking, it is incredibly uncommon to observe the sharp contacts between the various lithology categories due to intense alteration processes, erosion, many collapses, massive landslides, and rock flows. However, several lithology units are recognized for the area examined by recent field geology operations (Fig. 2). According to field geology investigations, the majority of the region is composed of volcanic and volcanoclastic rocks from the Eocene (56–33 MA) that have been carved out by Oligomiocene subvolcanic and plutonic rocks (33–5.5 MA). The Eocene Rock strata were covered by alkali-feldsparoid-rich silica under saturated volcanics (tephritite, basanite, and latite) that contained zeolite and leucite, and lastly, ferricrete-red conglomerates from the Neogene epoch. Lastly, the original rock units were covered by nonconsolidated fan, glacial, and alluvial deposits of recent age (Quaternary).

There is some copper oxide and sulfide mineralization in this area. Additionally, the Doostbiglou area's phyllic and argillic zones exhibit intense pyrite mineralization through veinlets and scattered deposits. However, gypsum has replaced many of the oxidised stockwork veinlets. The primary factors of mineralization in this region are based on the mineralogy of phyllic and potassic zones. These alteration zones contain low-grade chalcopryrite (Cu sulphide) mineralization with a grade of roughly five ppm Au. Figure 3 depicts a volcanic breccia rock unit (E^{vb}) underlying the Neogene age ferricrete-conglomerates (Ng^c) in southwestern part of the area.

Materials and methods

Sampling and data

In the study area, 345 samples of silt stream sediments were collected from the area by Sarzamin-e-Jolgeha-e-Asemani Mining Company and sent to the laboratory for analysis using the ICP-Fire Assay method. After collecting the data, first the data was subjected to preliminary statistical analysis. Also, the Q-Q diagram of individual elements was used to check the out-of-order data. Among the elements, only As element had two outliers, which were used after correction. No outliers were found in the data for other elements, especially Au. In this study, the combination of Log-ratio Transformations with U-spatial statistics Method has been applied for detecting geochemical anomalies in regional scale in Doostbiglou area. In the first step, the transformations

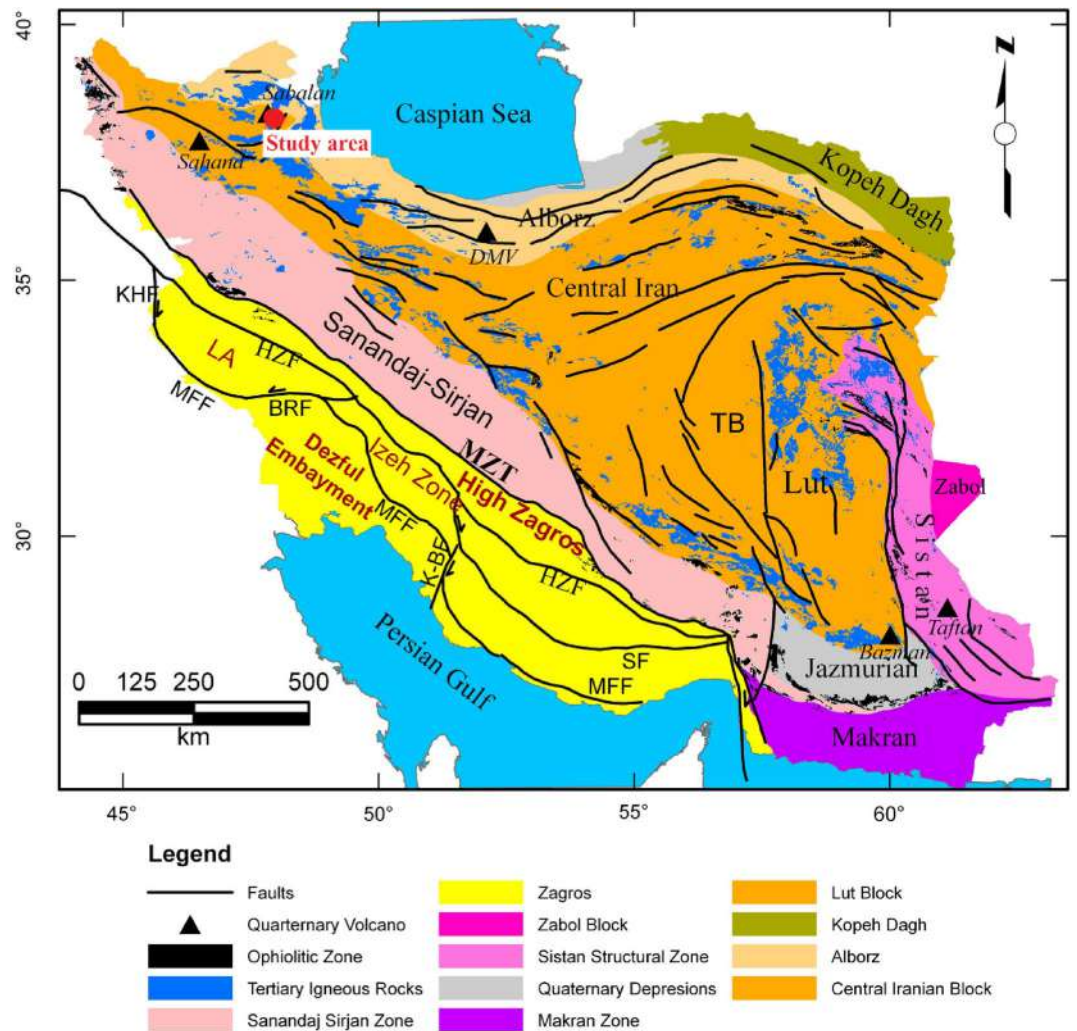


Fig. 1. The location of the Doostbiglou area on the structural geology model of Iran (Tectonic units' map of Iran taken from³⁴).

of ALR and CLR were performed to transform of Au, Cu and, Mo concentrations and so, the transformed values were modeled by U-spatial statistics Method.

Additive Log-Ratio transformation (ALR)

The additive log-ratio (ALR) transformation is a widely used for transforming data into a log-ratio space^{22,38}. This transformation involves choosing one variable from the set of variables, dividing the other variables by it, and then taking the logarithm. This way, the system is moved out of the closed state, but the selected variable is removed from the dataset. Therefore, care must be taken when selecting this variable, as if the selected variable is a major constituent of the rock, a spurious correlation will still be created. In geochemistry, trace or indicator elements such as Ti and Zr are used to select variables, as removing these elements from the dataset does not cause problems in mineral exploration analysis. Using this method, the direct relationship between the original data and their units of measurement is eliminated under this transformation^{13,39}. In this method, for values $X = (x_1, x_2, \dots, x_D)$, The following relationship holds:

$$alr(x) = \left[\ln \frac{x_1}{x_j}, \dots, \ln \frac{x_j - 1}{x_j}, \ln \frac{x_j + 1}{x_j}, \dots, \ln \frac{x_D}{x_j} \right]. \quad (1)$$

Centered Log-Ratio transformation (CLR)

In the CLR technique, each variable's first logarithm is given, and the logarithm is then divided by the variables' geometric mean (Eq. 2). In this case, every variable needs to have the same unit. One of the drawbacks of this approach is that the covariance matrix of the variables is irreversible (determinant = zero), even if no variables will be eliminated from the dataset in contrast to the prior way. For this reason, the data cannot be subjected to numerous multivariate statistical analyses¹³. Unlike the ALR method, one of the advantages of this method is that none of the variables are removed³.

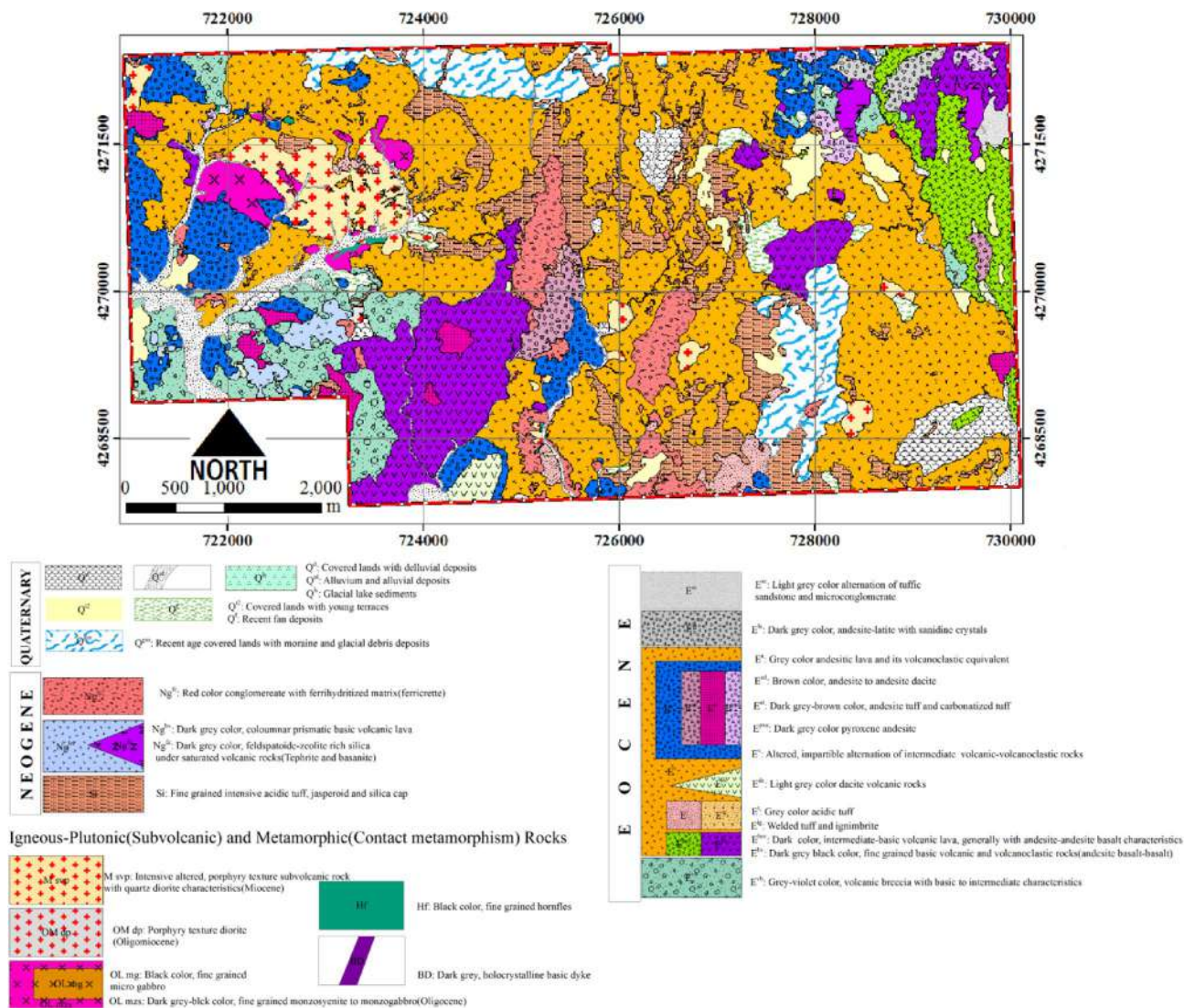


Fig. 2. A view from the study area lithology units³².



Fig. 3. Field view of volcanic breccia rock unit (E¹³) under the Neogene age ferricrete-conglomerates (Ng¹) in southwestern part of the area (view to North)³².

$$CLR(x)=\left[ln\frac{x_i}{\sqrt[D]{\prod_{i=1}^Dx_i}},\dots,ln\frac{x_D}{\sqrt[D]{\prod_{i=1}^Dx_i}}\right].\tag{2}$$

U-statistics modeling of Log-ratio transformed data

The spatial U-spatial statistics method is a structural method that separates anomalous samples from the geochemical background. The basis of this method is to calculate the weighted average of the samples within a moving window. In this method, the distance between samples is considered as the weight of the samples²⁷. A range of distinct U-values are obtained from these windows, which are the designated radiuses. The key idea for choosing the optimal U-value and maximizing anomaly separation is the maximum |U| value (U*)^{40,41}. Specific geological property variety is determined by the U* histogram. First, circles with radiuses ranging from 0 to 5000 m (r_{max}) were constructed, with radius changes of 10 m, to apply the U-spatial statistics approach and get the U-value for every sample location. According to⁴², the U-statistics value for each point I is determined as follows:

$$U_i(r)=\frac{\bar{x}_i(r)-\mu}{\sigma}.\tag{3}$$

For each sample point, the different U_i(r) is calculated by varying the r-value, where μ and σ represent the mean and standard deviation of the whole data, respectively. Accordingly, U_i(r) depends on r where every sample point is regarded as unknown, and many nearby samples inside the circles were used to compute associated U-values. Among all the surrounding samples, samples n₁ and n₂ are utilized to compute U_i for each search radius r. They belong to the background and anomaly populations, respectively. In light of this, it can be stated as follows^{29,42}and [43]:

$$U_i(r)=\frac{\sum_{j=1}^{n_1}w_j(r)x_j-\mu}{\sigma}+\frac{\sum_{k=1}^{n_2}w_k(r)x_k-\mu}{\sigma}.\tag{4}$$

If the average of the samples of the moving window of radius r is μ, and the averages of the background and anomalous populations are μ_A and μ_B, respectively, then the μ_B<μ<μ_A will be supported. Each sample's new features are computed using the ARL and CRL algorithms on the transformed data. In order to interpret each sample, absolute transformed values are computed in place of initial values (element concentration). This study examined the absolute transformed values of the concentrations of Au, Mo, and Cu. They are mappable and can thus be used as a geochemical anomaly mapping index. This was accomplished by applying U-value modeling to the transformed values relevant to the ARL and CRL.

Results

Correlation coefficient and cluster analysis in open and closed numerical systems

To analyze the correlation coefficient in open and closed systems, Spearman's correlation was used for data in the closed system due to non-normality. In contrast, while Pearson's correlation was applied to the results of additive log-ratio (ALR) and centered log-ratio (CLR) transformations. The corresponding results are presented in Tables 1 and 2, and 3. In the closed system, the correlation coefficient between Cu and Mo is 0.165, between Cu and Au is 0.45, and between Mo and Au is 0.43. In the open system using the ALR method, these correlations are 0.53, 0.68, and 0.56, respectively. Using the CLR method, they are 0.13, 0.31, and 0.31, respectively. As

	Au	Mo	Cu	Pb	Zn	Ag	Mn	Fe	As	Sb	Bi	Ca	Ba	Na	W	Sn
Au	1.000															
Mo	0.431**	1.000														
Cu	0.450**	0.165**	1.000													
Pb	0.459**	0.539**	0.207**	1.000												
Zn	0.190**	-0.007	0.220**	0.198**	1.000											
Ag	-0.018	-0.025	0.157**	-0.311**	-0.258**	1.000										
Mn	0.194**	-0.045	0.525**	0.072	0.295**	0.179**	1.000									
Fe	0.406**	0.475**	0.363**	0.392**	0.102	-0.102	0.127*	1.000								
As	0.224**	0.293**	0.276**	-0.061	-0.265**	0.425**	-0.031	0.312**	1.000							
Sb	0.049	0.112*	0.201**	0.100	-0.377**	0.342**	0.224**	0.159**	0.325**	1.000						
Bi	0.000	-0.007	-0.056	-0.030	0.115*	-0.080	-0.316**	0.109*	0.082	-0.186**	1.000					
Ca	-0.162**	-0.426**	0.056	-0.315**	0.004	0.098	0.524**	-0.287**	-0.263**	0.157**	-0.183**	1.000				
Ba	0.211**	0.286**	0.020	0.077	0.123*	-0.061	-0.324**	0.217**	0.357**	-0.233**	0.210**	-0.525**	1.000			
Na	0.046	-0.123*	0.191**	-0.214**	-0.114*	0.277**	0.451**	-0.003	0.169**	0.291**	-0.143**	0.463**	-0.319**	1.000		
W	0.069	0.401**	0.092	0.248**	0.075	0.048	-0.022	0.223**	0.050	0.205**	0.002	-0.215**	0.040	0.041	1.000	
Sn	-0.048	-0.166**	-0.250**	0.130*	0.422**	-0.535**	-0.294**	-0.189**	-0.437**	-0.467**	0.302**	-0.166**	0.210**	-0.428**	-0.066	1.000

Table 1. Spearman correlation coefficient of elements in samples (closed system).

	Au	Mo	Cu	Pb	Zn	Ag	Mn	Fe	As	Sb	Bi	Ca	Ba	Na	W	Sn
Au	1															
Mo	0.555**	1														
Cu	0.679**	0.529**	1													
Pb	0.613**	0.685**	0.646**	1												
Zn	0.507**	0.291**	0.509**	0.411**	1											
Ag	-0.085	-0.193**	-0.153**	-0.356**	-0.351**	1										
Mn	0.342**	0.190**	0.578**	0.347**	0.322**	-0.002	1									
Fe	0.567**	0.586**	0.687**	0.646**	0.469**	-0.177**	0.328**	1								
As	-0.048	0.087	-0.111*	-0.050	-0.143**	0.098	-0.158**	-0.043	1							
Sb	0.139**	0.157**	0.196**	0.236**	-0.055	0.191**	0.185**	0.485**	0.249**	1						
Bi	0.445**	0.471**	0.493**	0.422**	0.453**	-0.187**	0.050	0.782**	-0.049	0.261**	1					
Ca	0.144**	0.038	0.328**	0.165**	0.212**	0.019	0.477**	0.438**	-0.097	0.268**	0.320**	1				
Ba	0.575**	0.556**	0.633**	0.584**	0.501**	-0.177**	0.197**	0.884**	-0.040	0.344**	0.740**	0.395**	1			
Na	0.427**	0.392**	0.589**	0.482**	0.363**	-0.090	0.450**	0.680**	-0.095	0.215**	0.582**	0.623**	0.685**	1		
W	0.513**	0.556**	0.596**	0.573**	0.441**	-0.195**	0.201**	0.801**	-0.037	0.333**	0.666**	0.387**	0.796**	0.612**	1	
Sn	0.308**	0.101	0.430**	0.384**	0.439**	-0.261**	0.195**	0.479**	-0.061	0.233**	0.325**	0.352**	0.548**	0.484**	0.473**	1

Table 2. Pearson correlation coefficient of elements based on the results of ALR transformation.

	Au	Mo	Cu	Pb	Zn	Ag	Mn	Fe	As	Sb	Bi	Ca	Ba	Na	W	Sn
Au	1															
Mo	0.310**	1														
Cu	0.313**	0.128*	1													
Pb	0.363**	0.502**	0.189**	1												
Zn	0.356**	0.034	0.184**	0.169**	1											
Ag	-0.193**	-0.233**	-0.260**	-0.503**	-0.458**	1										
Mn	-0.037	-0.189**	0.222**	-0.063	0.051	0.004	1									
Fe	0.200**	0.416**	0.194**	0.467**	0.229**	-0.568**	-0.351**	1								
As	0.051	0.183**	-0.093	-0.120*	-0.134*	0.227**	-0.276**	-0.053	1							
Sb	-0.330**	-0.217**	-0.235**	-0.113*	-0.640**	0.358**	0.136*	-0.452**	0.078	1						
Bi	-0.022	-0.036	-0.142**	-0.137*	0.094	-0.083	-0.440**	-0.012	-0.004	-0.229**	1					
Ca	-0.273**	-0.451**	-0.086	-0.335**	-0.079	-0.015	0.553**	-0.346**	-0.416**	0.081	-0.149**	1				
Ba	0.312**	0.295**	0.143**	0.198**	0.328**	-0.252**	-0.443**	0.551**	0.277**	-0.553**	0.078	-0.465**	1			
Na	-0.168**	-0.170**	-0.081	-0.136*	-0.125*	-0.093	0.151**	0.099	-0.229**	-0.095	-0.140**	0.323**	-0.025	1		
W	0.067	0.374**	0.051	0.340**	0.196**	-0.378**	-0.403**	0.541**	-0.069	-0.351**	-0.010	-0.350**	0.433**	0.074	1	
Sn	0.186**	-0.012	0.041	0.231**	0.482**	-0.553**	-0.177**	0.358**	-0.200**	-0.533**	0.167**	-0.013	0.390**	0.033	0.339**	1

Table 3. Pearson correlation coefficient of elements based on the results of CLR transformation.

observed, the correlation coefficient of the index elements increased after the additive log-ratio transformation in the porphyry mineralization system. However, the correlation of the index elements decreased using the CLR method. To investigate the changes caused by ALR and CLR transformations, cluster analysis was performed on both the raw and transformed data. The results of the cluster analysis of the raw data (Fig. 4a) show that except for Mn, Ba, and As, the remaining elements are clustered with gold. In the cluster analysis of the ALR-transformed data (Fig. 4b), the clustering is more distinct, and in the gold cluster, copper is grouped with lead and molybdenum. Bismuth and tungsten are also clustered with molybdenum and gold in the dendrogram of the transformed data. In the CLR method, in the gold cluster, copper is grouped with the metallic elements arsenic, antimony, silver, and manganese. Zinc, tin, lead, iron, bismuth, and tungsten are also observed with molybdenum in the dendrogram of the transformed data (Fig. 4c).

The CLR transformation is used in compositional data analysis due to its ability to map data into an unconstrained Euclidean space. However, it introduces certain limitations, particularly the irreversible nature of its covariance matrix, which can lead to geometric distortions in correlation and clustering analyses. The singularity of the CLR covariance matrix arises because the transformed components are not independent, as they sum to zero. This property can impact statistical interpretations when using conventional multivariate methods that assume full-rank covariance structures. Despite these limitations, CLR remains a valuable tool in compositional data analysis, particularly when the focus is on relative relationships between components rather than absolute values. In this study, CLR transformation was applied for clustering and correlation analyses, in comparison to ALR transform and closed data. The results indicated, the the relationships between the

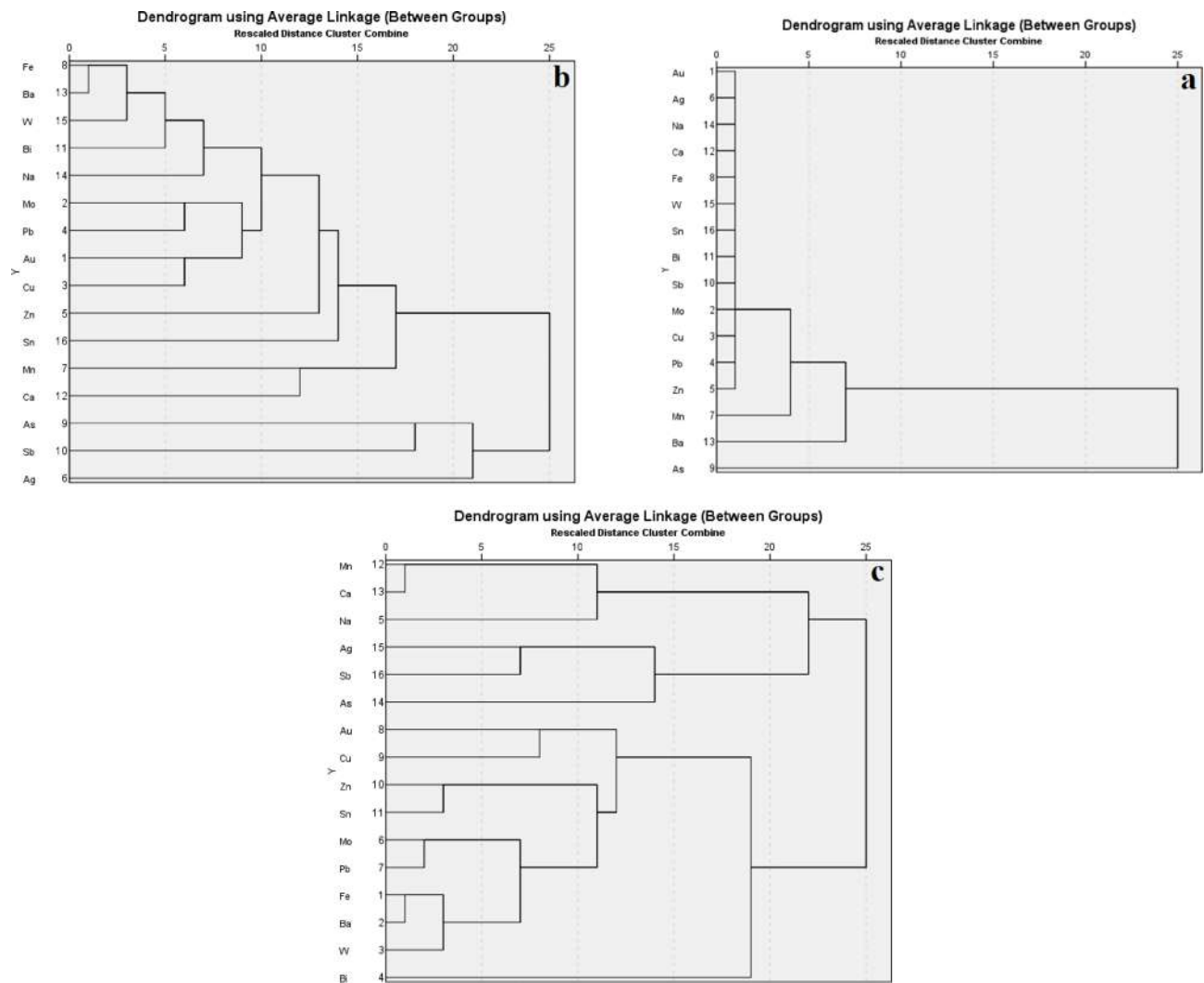


Fig. 4. Dendrogram resulting from cluster analysis on raw data (a), ALR transformed data (b), and CLR transformed data (c).

paragenesis elements are better distinguished in the CLR-transformed data compared to the untransformed data.

U-modeling of ALR and CLR data

The ALR and CLR values have been considered for the final data for U-modeling, and geochemical anomalies mapping has been performed on the U-values. This was accomplished by first calculating and interpolating the U-values of the ALR and CLR values using the U-statistics algorithm, coded in the MATLAB software. In this program, a moving window or search radius of 0 to 5000 m (r_{max}) was considered. Considering the number of samples in the neighborhood of each point within each window, the U value of that point was calculated, and its maximum value was assigned to the point as the U-value of ALR and CLR. The approach is not highly sensitive to this parameter, as seen by the results being comparatively the same when the radii of 1, 3, and 5 km were examined. To guarantee the involvement of every sample in the U-value calculation, the distance between the radii of the two circles was taken into account in two successive steps of 10 m. The computation for every sampling point begins with a circle of radius zero and extends to a radius of 5000 m. Each sampling point had 500 circles taken into consideration, and the U value was computed for each of them based on the distance of 10 m between two successive radii. The point's maximum U-values in absolute terms were then calculated and recorded as the U-value of ALR and CLR. Histograms of the absolute CLR and ALR values and U-values for the Au, Cu, and Mo elements are shown in Figs. 5 and 6. The histograms of absolute ALR and CLR values in these figures represent the normalized values of the raw data. A standard distribution curve has been fitted to the histograms. This plot, along with the standard deviation values of the log-ratio transformed data, indicates that the data transformed using the ALR method exhibits a higher degree of dispersion than the CLR method. To further investigate, a Bi-plot was constructed for both raw and log-transformed data using ALR and CLR methods (Fig. 7). This plot indicates that the correlation between gold and copper with molybdenum is negative and inverse for the raw data, making it difficult to identify the underlying realities present in the

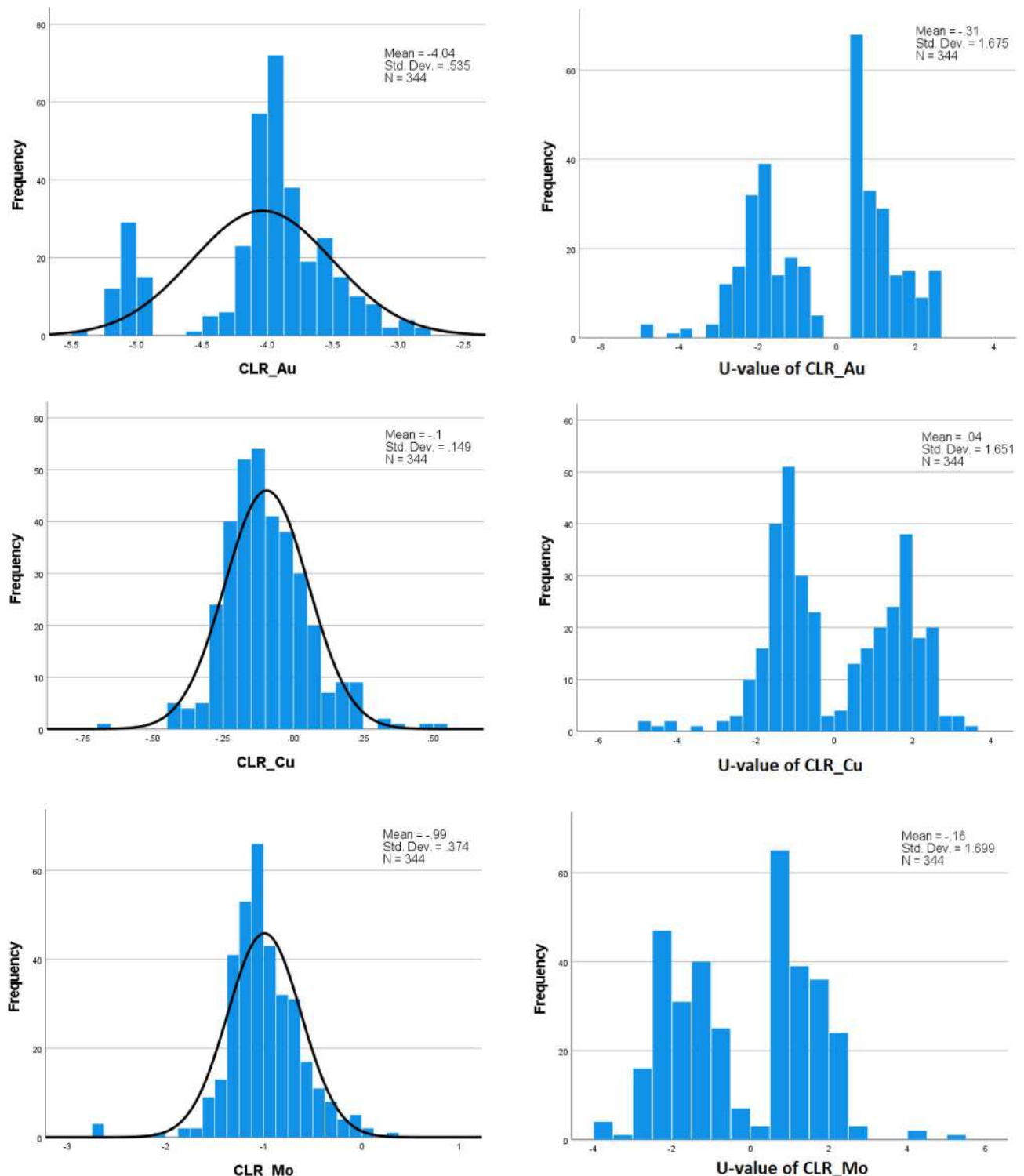


Fig. 5. Histograms of the absolute CLR values (left) and U-values of CLR (right) of the elements.

data. In the CLR model, the angles between the lines of copper, gold, and molybdenum have shown significant improvement compared to the raw data, allowing for a more accurate identification of the geochemical behavior of these three elements as paragenetic elements. In the ALR Bi-plot, the results improved significantly, revealing that these three elements exhibited very similar behavior. Notably, the angle between the lines for copper and gold approached zero, indicating that this type of transformation can be effectively utilized for data analysis. The Mardia method was employed to assess multivariate normality, and the p-values obtained from this test were quite low, suggesting a lack of multivariate normality. This is primarily due to some geochemical samples being located within mineralization zones, which are influenced by higher-concentration mineralization. One

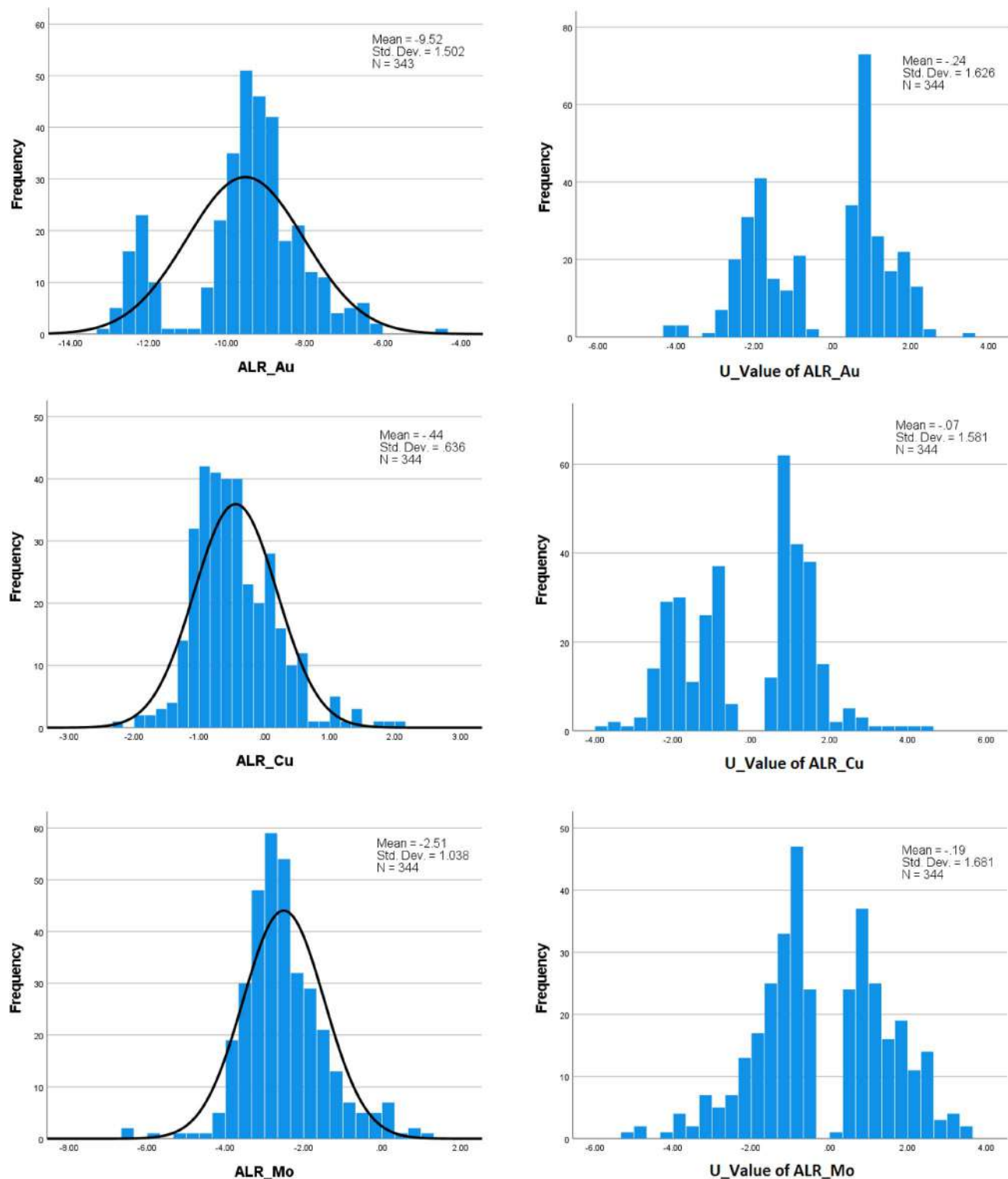


Fig. 6. Histograms of the Absolute ALR values (left) and U-values of ALR (right) of the elements.

advantage of the U-spatial statistics method is that it does not require multivariate normality in geochemical data analysis. In this algorithm, large differences in geochemical concentrations are disaggregated.

The zero point, which is the boundary that touches the anomalous data and extends past the background boundary, is also where the frequency distribution of the U-value of the ALR and CLR data displays a minimum. In other words, it is the approximate boundary of the anomalous population of the geochemical background. A maximum is seen in the frequency distribution of the U-value of the ALR and CLR data before and after this limit, indicating that the data is in two modes. Given that there is a mode in the frequency distribution of the

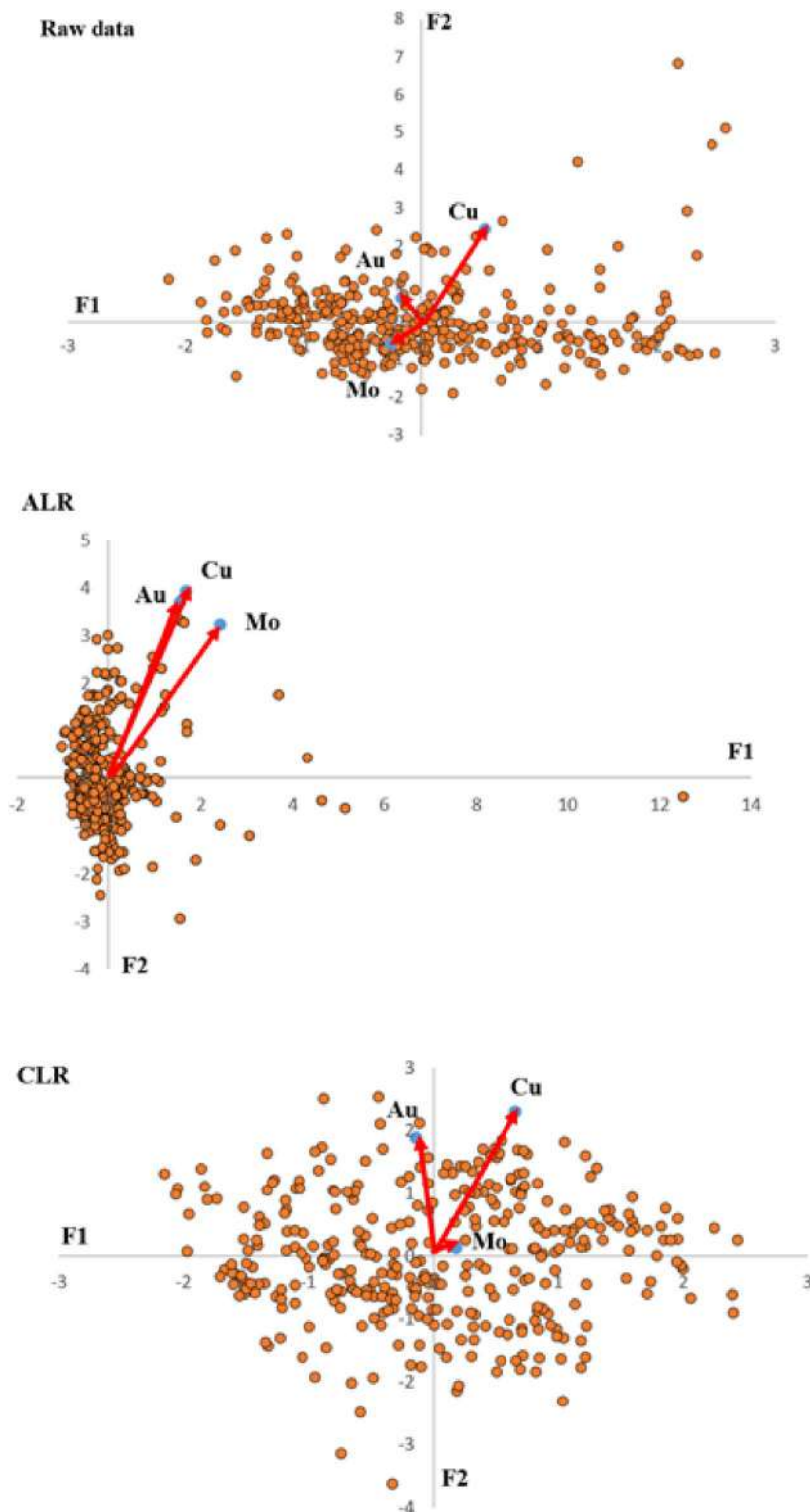


Fig. 7. Bi-plot of raw and log-transformed data by the ALR and CLR methods within XY coordinate system.

absolute ALR and CLR data, it is clear that the data set using this method has a high level of resolution. Two or more pollution populations are indicated by this distribution; the background population is the first, and the components associated with the anomalies are the subsequent populations. Two populations were found in Figs. 5 and 6 based on the distribution of ALR and CLR data. These two populations were separated using $\bar{U} + nSD$ values. The multivariate U-value of the ALR and CLR data was then shown on a geochemical map which, \bar{U} and SD are the average value and standard deviation of the estimated or modeled values, respectively.

The separation limit between geochemical anomalies from the background or geochemical threshold values was also determined by $n = 1, 2$, and 3 . This boundary was determined with $n = 1$, and the final map was produced.

Discussion

To investigate the spatial variations of the modeled data of ALR and CLR transformations by U-statistics, spatial distribution maps of geochemical elements, including delineated anomaly zones, were created. A classical statistical method was employed to separate anomalies from the background. Consequently, the threshold value was calculated using the $\bar{U} + nSD$ criterion and anomalous zones were identified on geochemical maps based on this value. Threshold values for Au, Cu, and Mo with modeled values for $n = 1$ and 2 are presented in Table 4. The $\bar{U} + SD$ criterion was utilized to determine anomalous zones. The reason why criterion $\bar{U} + SD$ was chosen to obtain anomalous areas is that the results obtained from criterion $\bar{U} + 2SD$ introduced a very small anomaly that was not consistent with the field realities. The area of the anomaly in this criterion was either zero or very small, which could not justify the secondary geological processes in this type of deposit. The reason may be due to the low mobility of gold, copper and molybdenum elements in this type of mineralization. Figure 8 illustrates the spatial distribution maps of Au, Cu, and Mo with U-values of ALR and CLR, along with the delineated anomaly zones based on the threshold values. In this distribution, the Au anomaly is observed in the region's center, which is inclined to the west with an approximate northeast-southwest trend. This mineralization trend is more evident in the U-modeling of ALR. For Cu, the mineralization trend is also northeast-southwest, confirming the close association of this element with Au in porphyry gold-copper-molybdenum systems. The variation trend of molybdenum in the U-modeling of ALR is in good agreement with gold and copper. However, in the U-modeling of CLR, the anomalous zones of this element extend to the south of the region and exhibit a more significant extent. The results of decomposing geochemical data of elements using ALR and CLR transformations and modeling these data with the U-statistic indicate the presence of copper-gold-molybdenum mineralization in the region. To further validate the methods and select the optimal method, the results of local and regional exploration conducted by Sarzamin-e-Jolgeha-e-Asemani Mining Company were utilized. These results identified two gold mineralized zones for subsurface drilling. The conducted drilling confirmed gold mineralization at depth in these zones. To validate the results of the models, the locations of these areas were considered. Following this, the anomalous limits were mapped according to the threshold values obtained from the U-modeling of the output data of the ALR and CLR algorithms (Table 4).

Figure 9 illustrates the map of anomalous limits produced by the ALR and CLR models. In the next step, the locations of the exploration drilling wells, or the mineral deposit areas, were highlighted with a pale blue hatching on these maps. To compare and validate the results, both the location of the mineral deposits and their surface extent, as well as the mineralization trends, were considered. Both models yielded similar results regarding identifying gold anomaly zones and showed excellent agreement with field data. The extent of halos detected by the U-statistics modeling of the CLR data is more extensive, which can increase the cost of continued exploration. The U-statistics modeling of ALR data is closer to field realities and more clearly indicates the mineralization trend. Therefore, these new models are proposed to evaluate the spatial distribution of elements and determine the threshold values (Fig. 8; Table 2).

Amirihanza et al. (2018) demonstrated that mineralization may be located along specific fault trends in mineralized areas, and spatial pattern analysis of structures using various methods can be effective in determining mineralized zones⁴⁴. Accordingly, in the Doostbiglou area, there are two types of faults with NW-SE and NE-SW trends. The NW-SE trending faults are associated with mineralization, while the NE-SW trending faults formed after the mineralization occurred. The NW-SE faults are also consistent with stockwork mineralization in the fractures. Derakhshani and Abdolzadeh (2009) demonstrated that porphyry copper systems may exhibit specific patterns of geochemical zonation characterized by the enrichment of certain elements associated with hydrothermal alteration zones such as potassic, phyllic, and argillic⁴⁵. The geochemical distribution of elements in the Doostbiglou area is also related to various alteration zones, supporting the hypothesis of a porphyry-style hydrothermal system with structurally controlled fluid pathways.

Geological field investigations in the study area indicate the presence of alterations associated with this type of mineralization. Figure 10a shows the phyllic alteration zone and its location. From the point of metallogeny in porphyry mineralization system, phyllic zone is one of the main alteration events that it observable in surficial and underground parts of the area with rich pyrite and sericite contents. The main host rock for this zone is the subvolcanic quartz diorite stock. It is characterized by intensive pyrite and rare chalcopyrite existence. The main

Model Results	U-modeling of ALR (Au)	U-modeling of ALR (Cu)	U-modeling of ALR (Mo)	U-modeling of CLR (Au)	U-modeling of CLR (Cu)	U-modeling of CLR (Mo)
\bar{U}	-0.23797	-0.06907	-0.18602	-0.30594	-0.1592	0.04139
S	1.625733	1.580782	1.680671	1.67507	1.699413	1.650509
$\bar{U} + S$	1.387763	1.511715	1.49465	1.369127	1.54021	1.691899
$\bar{U} + 2S$	3.013496	3.092497	3.175321	3.044197	3.239622	3.342409

Table 4. Threshold values of the modeled data by the U-modeling of the log-ratio transformed data method

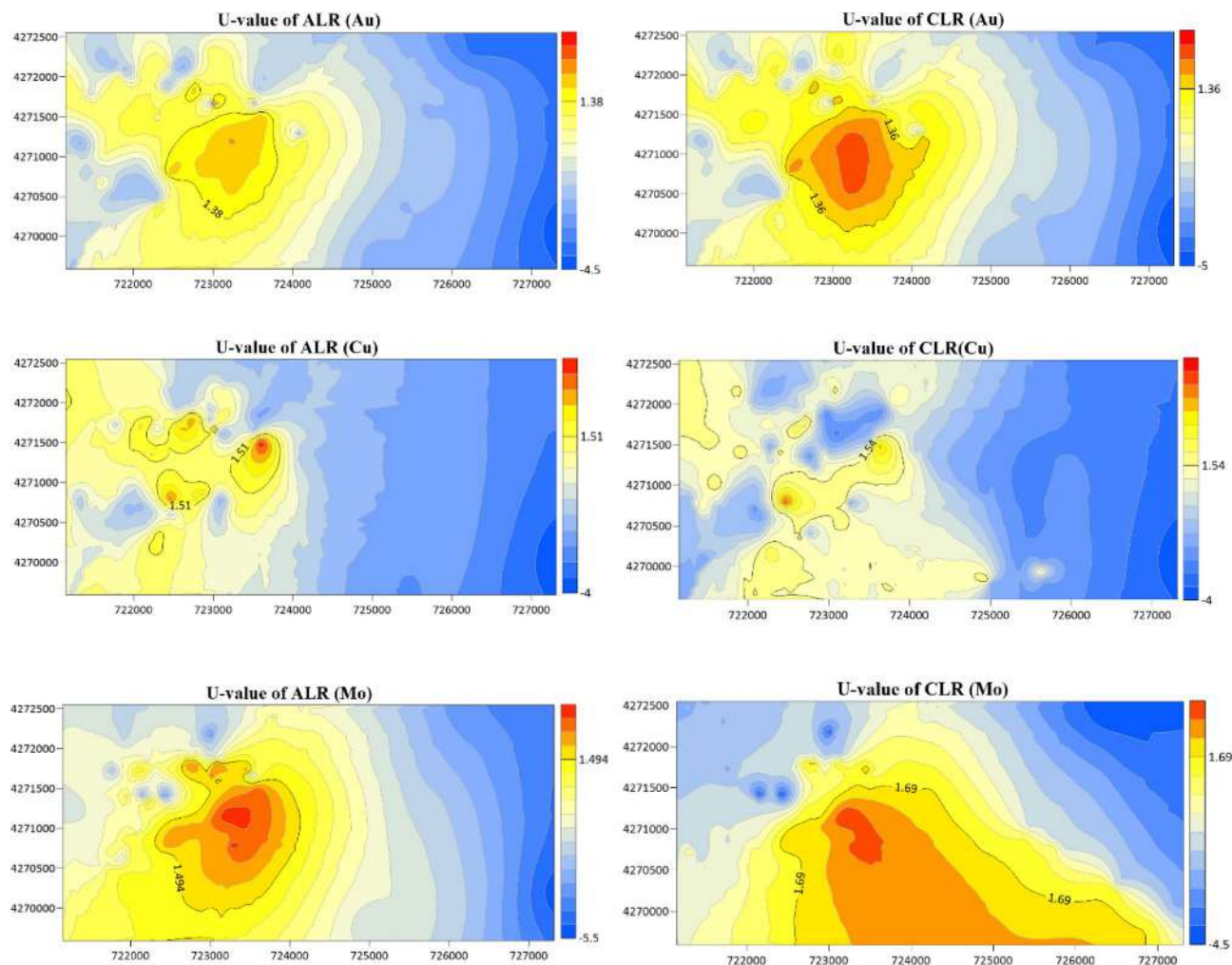


Fig. 8. Geochemical distribution maps of Au, Cu, and Mo with U-values of ALR and CLR, along with the delineated anomaly zones based on the threshold values.

alteration zone event in the area is the argillic zone with clay minerals accumulation, yellow color and developed stockworks characterization. This zone covered the phyllic zone and is located under the ferrihydrite and silica zones (Fig. 10b). Generally, the existence of gypsum in this zone indicates former high sulfide mineralization in the porphyry system. The subvolcanic quartz diorite stock is the main host rock for this alteration event.

Conclusion

In this research, a new integrated method was introduced to find geochemical exploration targets in the gold-molybdenum and copper mineralization type. This work was done by modeling the output data transformed to ALR and CLR methods with the structural method of U-spatial statistics. Analysis of the correlation coefficient in open and closed numerical systems showed that the correlation coefficient of the index elements increased after the additive log-ratio transformation in this mineralization system. However, the correlation of the index elements decreased using the CLR method.

In the cluster analysis of the ALR and CLR transformed data, the clustering is more distinct, and in the gold cluster, copper is grouped with lead and molybdenum. The spatial distribution maps of Au, Cu, and Mo with U-values of ALR and CLR showed that the results of decomposing geochemical data of elements using ALR and CLR transformations and modeling these data with the U-statistic support the presence of copper-gold-molybdenum mineralization in the region. To compare and validate the results, both the location of the mineral deposits and their surface extent, as well as the mineralization trends, were considered. Both models yielded similar results regarding the identification of gold anomaly zones and showed excellent agreement with field data. The extent of halos detected by the U-statistics modeling of the CLR data is more extensive, which can increase the cost of continued exploration. The U-statistics modeling of ALR data is closer to field realities and more clearly indicates the mineralization trend. Therefore, these new models are proposed to evaluate the spatial distribution of elements and determine the threshold values.

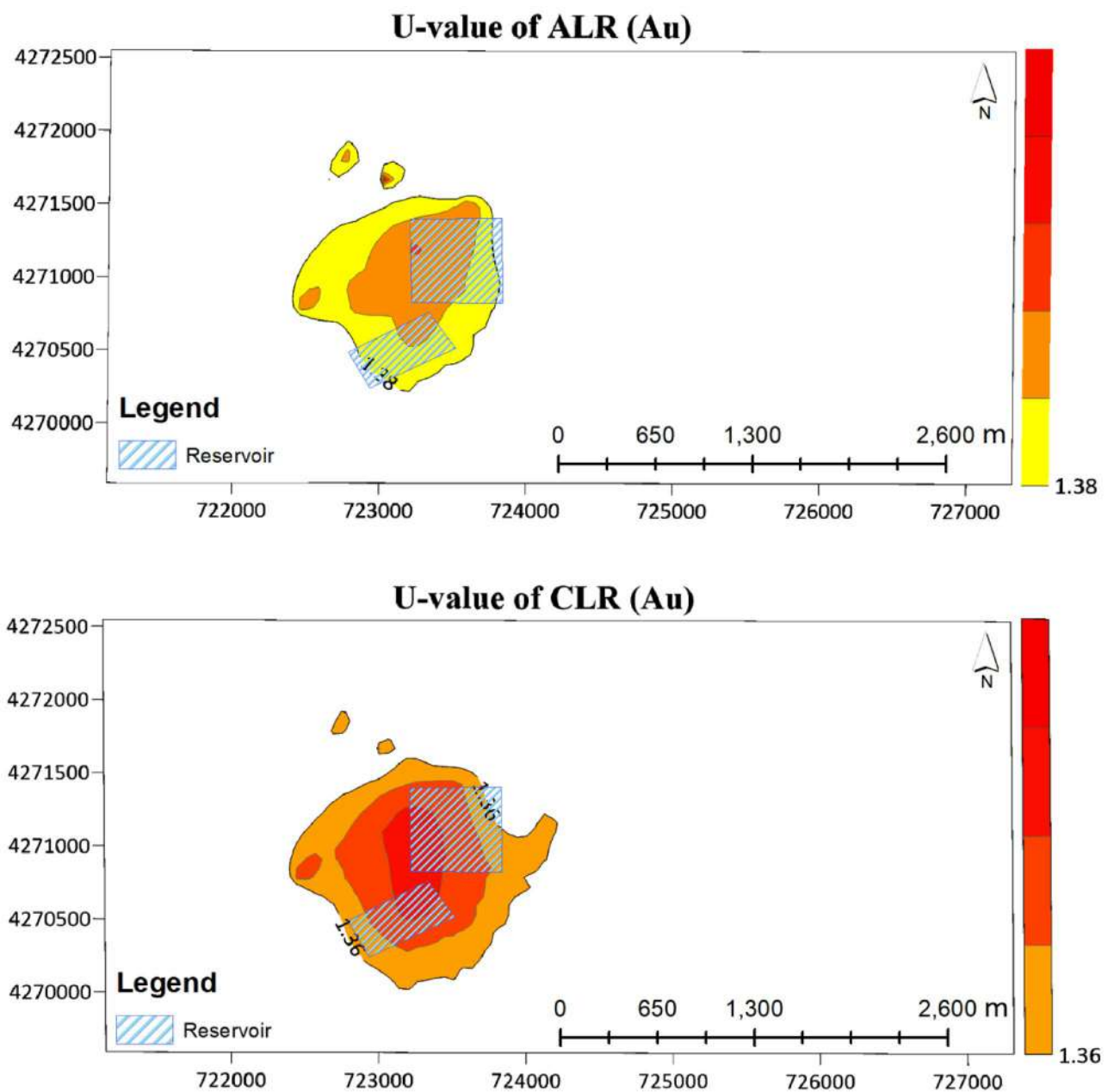


Fig. 9. Map of anomalous limits produced by U-modeling of the ALR and CLR transformations.

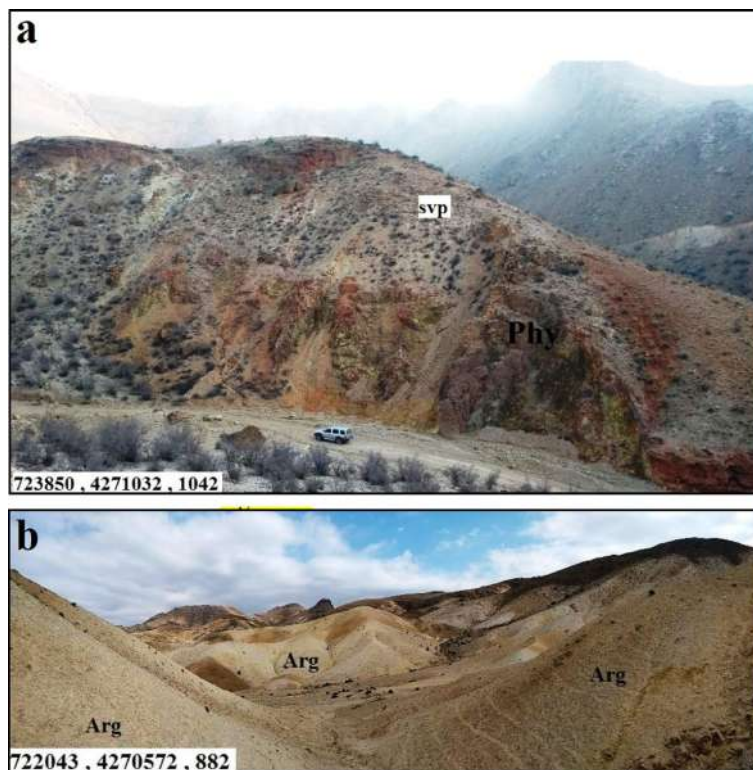


Fig. 10. Phyllic alteration zone (a) and argillization processes (b) in central part and north of the doostbiglou area³².

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 15 January 2025; Accepted: 5 June 2025

Published online: 01 July 2025

References

- Martínez, J., Llamas, J., De Miguel, E., Rey, J. & Hidalgo, M. C. Determination of the geochemical background in a metal mining site: example of the mining district of Linares (South Spain). *J. Geochem. Explor.* **94** (1–3), 19–29 (2007).
- Carranza, E. J. M. *Geochemical Anomaly and Mineral Prospectivity Mapping in GIS* (Elsevier, 2008).
- Leung, R., Balamurali, M. & Melkumyan, A. Sample Truncation strategies for outlier removal in geochemical data: the MCD robust distance approach versus t-SNE ensemble clustering. *Math. Geosci.* **53** (1), 105–130 (2021).
- Hong, J. et al. National-scale geochemical survey: distribution of chemical elements in stream sediment of South and central Asia. *J. Geochem. Explor.* **262**, 107452 (2024).
- Alasadi, S. A. & Bhaya, W. S. Review of data preprocessing techniques in data mining. *J. Eng. Appl. Sci.* **12** (16), 4102–4107 (2017).
- Garrett, R. G., Reimann, C., Hron, K., Kynčlová, P. & Filzmoser, P. Finally, a correlation coefficient that tells the geochemical truth. *Explore* **17**, 1–10 (2017).
- Balamurali, M. & Melkumyan, A. Detection of outliers in geochemical data using ensembles of subsets of variables. *Math. Geosci.* **50** (4), 369–380 (2018).
- Martín-Méndez, I., Llamas-Borrajó, J., Llamas Lois, A. & Locutura, J. Factor analysis in residual soils of the Iberian pyrite belt (Spain): comparison between Raw data, log-transformation data and compositional data. *Geochem.: Explor. Environ., Anal.* **24** (2), geochem2024–geochem2005 (2024).
- Aitchison, J. The statistical analysis of geochemical compositions. *J. Int. Assoc. Math. Geol.* **16**, 531–564 (1984).
- Filzmoser, P. & Hron, K. Outlier detection for compositional data using robust methods. *Math. Geosci.* **40**, 233–248 (2008).
- Carranza, E. J. M. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *J. Geochem. Explor.* **110** (2), 167–185 (2011).
- Khammar, F., Yousefi, S. & Joonaghani, S. A. Analysis of lithogeochemical data using log-ratio transformations and CA fractal to separate geochemical anomalies in Tak-Talar, Iran. *Arab. J. Geosci.* **14** (8), 1–15 (2021).
- Aitchison, J. The statistical analysis of compositional data. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **44** (2), 139–160 (1982).
- Pawlowsky-Glahn, V. & Buccianti, A. *Compositional Data Analysis* (Wiley, 2011).
- Graffelman, J., Pawlowsky-Glahn, V., Egozcúe, J. J. & Buccianti, A. Exploration of geochemical data with compositional canonical biplots. *J. Geochem. Explor.* **194**, 120–133 (2018).
- Pawlowsky-Glahn, V. & Egozcúe, J. J. Compositional data in geostatistics: A log-ratio based framework to analyze regionalized compositions. *Math. Geosci.* **52** (8), 1067–1084 (2020).
- Owen, D. D. R., Pawlowsky-Glahn, V., Egozcúe, J. J., Buccianti, A. & Bradd, J. M. Compositional data analysis as a robust tool to delineate hydrochemical facies within and between gas-bearing aquifers. *Water Resour. Res.* **52** (8), 5771–5793 (2016).
- Mahdianfar, H. & Salimi, A. Fractal modeling of geochemical mineralization prospectivity index based on centered Log-Ratio transformed data for geochemical targeting: a case study of Cu porphyry mineralization. *J. Min. Environ.* **13** (3), 821–838 (2022).

19. Liu, Y., Xia, Q. & Cheng, Q. Sequential Gaussian co-simulation of tectono-geochemical anomaly for concealed ore deposit prediction. *Appl. Geochem.* **157**, 105768 (2023).
20. Shahrestani, S. & Sanislav, I. Delineation of geochemical anomalies through empirical cumulative distribution function for mineral exploration. *J. Geochem. Explor.* **270**, 107662 (2025).
21. Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barcelo-Vidal, C. Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35** (3), 279–300 (2003).
22. Reimann, C., Filzmoser, P., Garrett, R. & Dutter, R. *Statistical Data Analysis Explained: Applied Environmental Statistics with R* (Wiley, 2011).
23. Liu, X., Wang, W., Pei, Y. & Yu, P. A knowledge-driven way to interpret the isometric log-ratio transformation and mixture distributions of geochemical data. *J. Geochem. Explor.* **210**, 106417 (2020).
24. Puchhammer, P. et al. A performance study of local outlier detection methods for mineral exploration with geochemical compositional data. *J. Geochem. Explor.* **258**, 107392 (2024).
25. Cheng, Q., Agterberg, F. P. & Bonham-Carter, G. F. A Spatial analysis method for geochemical anomaly separation. *J. Geochem. Explor.* **56** (3), 183–195 (1996).
26. Cheng, Q., Agterberg, F. P. & Ballantyne, S. B. The separation of geochemical anomalies from background by fractal methods. *J. Geochem. Explor.* **51** (2), 109–130 (1994).
27. Cheng, Q., Xu, Y. & Grunsky, E. Integrated Spatial and spectrum method for geochemical anomaly separation. *Nat. Resour. Res.* **9** (1), 43–52 (2000).
28. Li, C., Ma, T. & Shi, J. Application of a fractal method relating concentrations and distances for separation of geochemical anomalies from background. *J. Geochem. Explor.* **77** (2–3), 167–175 (2003).
29. Ghavami-Riabi, R., Seyedrahimi-Niaq, M. M., Khalokakaie, R. & Hazareh, M. R. U-spatial statistic data modeled on a probability diagram for investigation of mineralization phases and exploration of shear zone gold deposits. *J. Geochem. Explor.* **104** (1–2), 27–33 (2010).
30. Ghezelbash, R. & Maghsoudi, A. 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. *C.R. Geosci.* **350** (4), 180–191 (2018).
31. Seyedrahimi-Niaq, M., Shokri, N. & Lotfibakhsh, A. Improving the method of U-spatial statistics by modeling the enrichment index of stream sediments for the purpose of introducing geochemical anomalous areas of epithermal gold type mineralization. *J. Min. Eng.* **18** (59), 15–30 (2023).
32. Samani, B., Vusuq, B., Karbalaie, A. A. & Block, N. E. Geological Map of Saheb Divan Exploration Area (1:1000), Area III-II, Meshgin Shahr, Ardabil District, *Sarzamine Jolgehay Asemani Co., Ltd.* 70. (2020).
33. Nabavi, M. & H. *Tectonic Map of Iran* (Geol. Surv, 1976).
34. Teknik, V. & Ghods, A. Depth of magnetic basement in Iran based on fractal spectral analysis of aeromagnetic data. *Geophys. J. Int.* **209** (3), 1878–1891 (2017).
35. Darvishzadeh, A. *Geology of Iran* 901 (pp, In Persian, 1991).
36. Moine-Vaziri, H., Khalili-Marandi, S. & Brousse, R. Importance d'un volcanisme potassique, Au miocène supérieur, En Azerbaïdjan (Iran). *Comptes Rendus de l'Académie des sciences. Série 2, mécanique, physique, chimie, sciences de l'univers. Sci. Terre.* **313** (13), 1603–1610 (1991).
37. Moine-Vaziri, H. An introduction of geology of magmatism in Iran. *Tehran Univ. Teacher Educ. Tehran*, 440 (1996).
38. Pawlowsky-Glahn, V. & Egozcue, J. J. Compositional data and their analysis: an introduction. *Geol. Soc. Lond. Special Publications.* **264** (1), 1–10 (2006).
39. Carranza, J. *Analysis and Mapping of Stream Sediment Geochemical Anomalies: Should We Logratio Transform the Data* (University of Twente, 2006).
40. Darabi-Golestan, F., Ghavami-Riabi, R., Khalokakaie, R., Asadi-Haroni, H. & Seyedrahimi-Nyarah, M. Interpretation of litho-geochemical and geophysical data to identify the buried mineralized area in Cu-Au porphyry of Dalli-Northern hill. *Arab. J. Geosci.* **6** (11), 4499–4509 (2013).
41. Seyedrahimi-Niaq, M. & Hekmatnejad, A. The efficiency and accuracy of probability diagram, Spatial statistic and fractal methods in the identification of shear zone gold mineralization: a case study of the Saqqez gold ore district, NW Iran. *Acta Geochim.* **40**, 78–88 (2021).
42. Cheng, Q. Spatial and scaling modelling for geochemical anomaly separation. *J. Geochem. Explor.* **65** (3), 175–194 (1999).
43. Seyedrahimi-Niaq, M., Mahdianfar, H. & Mokhtari, A. R. Integrating principal component analysis and U-statistics for mapping polluted areas in mining districts. *J. Geochem. Explor.* **234**, 106924 (2022).
44. Amirhanza, H., Shafieibaf, S., Derakhshani, R. & Khojastehfar, S. Controls on Cu mineralization in central part of the Kerman porphyry copper belt, SE Iran: constraints from structural and spatial pattern analysis. *J. Struct. Geol.* **116**, 159–177 (2018).
45. Derakhshani, R. & Abdolzhadeh, M. Geochemistry, mineralization and alteration zones of Darrehzar porphyry copper deposit, Kerman, Iran. *J. Appl. Sci.* **9** (9), 1628–1646 (2009).

Acknowledgements

The authors are grateful to Sarzamin-e-Jolgeha-e-Asemani Mining Company in Iran for placing the stream sediments data of the Doostbiglou exploration limit.

Author contributions

H.M. modeled the raw data with Algorithms ALR and CLR. M.M. performed U-statistics modeling of the transformed data. He developed the research results and also developed and finalized the manuscript text.

Declarations

Competing interests

The authors declare no competing interests.

Publication statement

The authors have obtained the necessary permission to use the material of Figs. 2, 3 and 10.

Additional information

Correspondence and requests for materials should be addressed to M.S.-N.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025