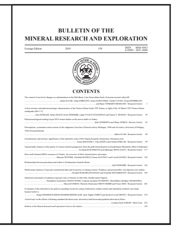
Bull. Min. Res. Exp. (2019) 159: 185-200

Bulletin of the Mineral Research and Exploration

http://bulletin.mta.gov.tr



Multivariate analysis of log-ratio transformed data and its priority in mining science: Porphyry and polymetallic vein deposits case studies

Farshad DARABİ-GOLESTAN[a*] and Ardeshir HEZARKHANI[b]

[a] *Department of Mining and Metallurgical Engineering, Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran. Orcid.org 0000-0002-4648-152X*

[b] *Department of Mining and Metallurgical Engineering, Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran orcid.org/0000-0002-1149-3440*

Research Article

Keywords:
Isometric log-ratio transformation, Subcompositional coherence, Correspondence analysis, Principal component analysis, Geochemical compositional data, Polymetallic and porphyry deposits.

ABSTRACT

Each mineralization style is characterized by typical signature associations between elements due to elemental interactions, therefore the coherence and closure effects problem must be overcome in geochemical processing. The coherence indicates the ratios between two components (rows or columns) remains the same whether they are considered in a subcomposition or in the full composition. The log-ratio transformation (LRT) has recognized as a standard procedure to support subcompositional coherence. The log-transformed data is applicable for geochemical data to unveil such associations, prior to applying the multivariate analysis like correspondence analysis (CA) and principal component analysis (PCA). At the present study, subcompositional coherence is overcome by inverse iso-metric log-ratio transformation for geochemical compositional data at two polymetallic and porphyry deposits. Based on Ilr-transformed data, Ag, Au, As, Pb, Te, Mo and rather S, W, Cu are enriched as polymetallic elements at Glojeh, while Au-Cu-(Mo) compositions indicate a porphyry deposit occurred in Dalli deposit. The ability to handle zero values in the data matrix and determining an elemental eccentricity from the center of each axis based on Euclidean distances are the advantages of CA method, with compression to LRT. Whereas, loading factors which spread in every direction and providing subcompositional coherence are the competitive advantages of PCA based on LRT, for both case studies. Results with these techniques show significant ability to draw an inference in such geochemical data, and in improving the performance of multivariate techniques using LRT.

Received Date: 20.11.2017
Accepted Date: 01.07.2018

1. Introduction

Applied geochemical data such as most mining data are often defined as compositional (Reimann et al., 2011). The compositional data are indicated as parts per million (ppm), percentages, proportions, and other quantities which is concerned with relative and sums up to a constant value like 100 percent or 1 (Aitchison and Greenacre, 2002). Hence, the individual or compositional variables are not independent and variables can not be freed to vary from the others. Accordingly, the problem of data closure effects needs to be considered in geochemical data evaluation (Reimann et al., 2012; Reimann et al., 2011; Zuo et al., 2013). Aitchison (1982, 1983, 1986) proposed that the log-ratio transformation (LRT) is an appropriate method to evaluate the total variability of the compositional data. The subcompositional coherence means that the correlation coefficient or the distance between two components in a subcomposition (subset) is the same as that for the same two components in the full composition (Greenacre, 2011). Using LRT

Citation Info: Darabi-Golestan, F., Hezarkhani, A. 2019. Multivariate analysis of log-ratio transformed data and its priority in mining science, porphyry and polymetallic vein deposits case studies. Bulletin of Mineral Research and Exploration, 159, 185-200.
http://dx.doi.org/10.19111/bulletinofmre.456958

* *Corresponding author: F. DARABİ GOLESTAN, pooyan@aut.ac.ir*

this problem is posed and consistent results were obtained because the values of a selected element represent its proportions in the complete sample, whether one works with the subcomposition or in the full composition (Filzmoser et al., 2009*c*; Pawlowsky-Glahn and Egozcue, 2006). The closure effect problem of elements and samples is commonly apparent in the simply log-transformed scores of a geochemical compositional data. In order to overcome this problem, this data can be transformed by using the LRT, e.g. centered log-ratio (clr), the iso-metric log-ratio (ilr), and the additive log-ratio (alr), prior to using any of the multivariate methods (Carranza, 2011; Filzmoser and Hron, 2008; Filzmoser et al., 2009*a*; Silverman et al., 2016; Templ et al., 2008).

Correspondence analysis (CA) is a comprehensive and exploratory method which can widely be applied to recognize various genetic associations and elemental distribution in geochemical exploration. This technique is conducted according to the two main R-mode (factor scores related to elements) and Q-mode (factor scores related to samples) methods (David et al., 1974; Março and Scarminio, 2007). These methods are used to reduce high-dimensional data into low-dimensional subspaces that explain the main variances and qualitative and/or quantitative variation among the whole data (Akbarpour et al., 2013; Carranza, 2009). Principal component analysis (PCA) is another method that is mainly used to compress large data matrices and extract the common factors (De Winter and Dodou, 2016; Pommer et al., 2004). PCA method is a data-driven multivariate technique that can reveal similarities based on the correlation or covariance matrix (Croux and Haesbroeck, 2000; Olofsson et al., 2009). In addition, the PCA based upon LRT of the geochemical compositional data is applied to reveal underlying patterns in the data, in a more vivid manner. In the geochemical exploration and to have full rank with spreading in a full circle distribution, Ilr transformation is better than Clr transformed data (Carranza, 2011; Filzmoser et al., 2009*a,* 2010). It was indicated using inverse Ilr transformed data that each variable of the biplot is related to one of the components. Greenacre (2010) pointed out that in compositional data using that PCA of Clr-transformed data (or totally LRT) and CA through the Box–Cox power transformation the same results were obtained. He indicates as long as the weighted form of LRT is used, so in this respect the sub-compositionally coherent is much better. Totally, LRT is considered as a limited case of CA (Greenacre, 2010). The obvious benefit of CA is the ability to deal with zero values in the data which can be analyzed for nonzero power parameter of the Box-Cox transformation while it is a problem in LRT (Greenacre, 2010; Martín-Fernández et al., 2003).

The objective of this paper is to apply various CA and PCA opened by log-transformed and LRT data for two different types of Au mineralization, the Glojeh polymetallic and Dalli porphyry deposit. Our case studies demonstrate that the closure effect is an inherent problem for geochemical and environmental compositional data which cannot be overcome by simply log-transformation. The present study aims to investigate: (1) A better insight to sense and explore the essential information on mineralization, and relationship between elements and samples; (2) a comparison between the log-transformed and Ilr-transformed data form two genetic types of polymetallic vein and porphyry Au deposits; and (3) a comparison between CA and PCA of compositional geochemical data.

2. Material and Method

2.1. Case studies

Two different types of polymetallic vein and porphyry mineralization were investigated to evaluate the relationship, interaction, and closure effect between elements. These specifications could ease the interpretation of geochemical dispersion pattern of elements. The first case study is the Glojeh polymetallic vein deposit. It is located in NW Iran, central part of the Tarom- Hashtjin Metallogenic Province (THMP). It is recognized as the main metallogenic province in the western Alborz magmatic arc (Darabi-Golestan and Hezarkhani, 2016; figure 1). There are two main polymetallic Au-Ag-Cu-Pb-Zn veins and several veinlets that were displaced by east-west striking in Glojeh. The offset veins (and parallel in some places) have a length of about 1,5 km and a width range from 0,1 to 4 m (averages 2,5 m). The veins were intersected by BH2N1 (borehole) at 26,2-33,9 m and 79,68-81,51 m, respectively. Different exploration works have been carried out in the area (eleven boreholes and twelve trenches), but just 153 samples (consist of duplicate and replicate samples) of BH2N1 have been analyzed by inductively coupled plasma mass spectrometry (ICP-MS) for 44 elements at Earth Sciences Development Company Lab of Iran (Darabi-Golestan and Hezarkhani, 2016). The second

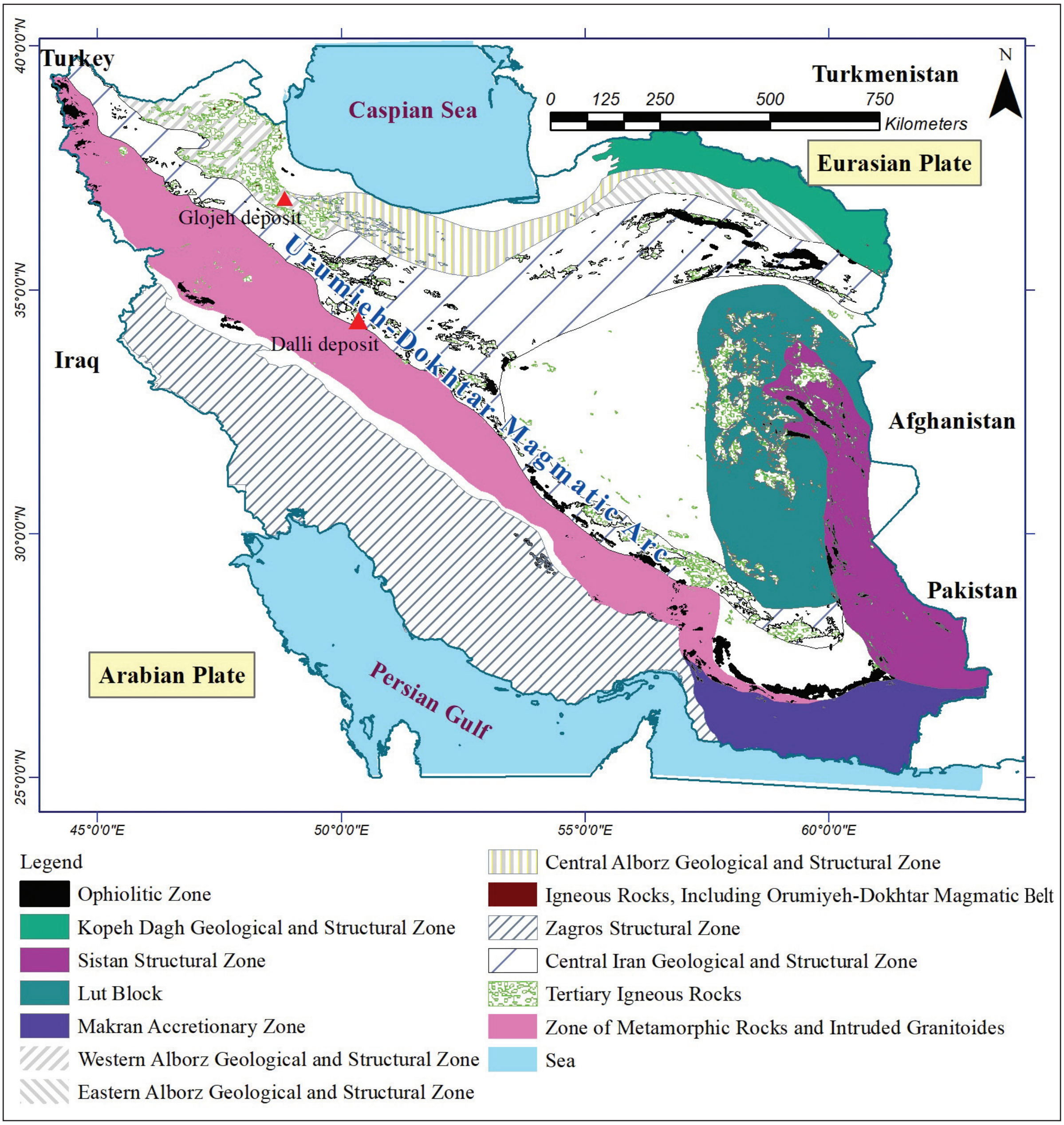


Figure 1- The location map of the deposits, major Iran's structural zones and geological units, including western Alborz zone that consist of Glojeh deposit and Urumieh-Dokhtar Magmatic Arc including Dalli deposit. The western, central, and eastern Alborz, and Sanandaj-Sirjan zone of metamorphic rocks and intruded granitoids parallel to Urumieh-Dokhtar Magmatic Arc ranges are highlighted.

case study is Northern Dalli porphyry deposit which located in the central province of Iran, 70 Km away from Arak (Darabi-Golestan et al., 2013*a*; figure 1). In total, 149 (plus 16 replicate sample) soil samples were taken from the area, with a grid net of 50 m×50 m. They were analyzed for 45 elements by using ICP-MS method in the ANDL Lab Australia (Darabi-Golestan et al., 2013*b*). Both these data sets have compositional specification.

2.2. Quality Control

The total number of 20 duplicate and 20 replicate samples were collected from original samples of Glojeh deposit. They were prepared and analyzed by ICP-MS for 44 elements, individually. The quality of analytical data were evaluated using the Thompson-Howarth graphical method (Stanley, 2006; Thompson and Howarth, 1976). The duplicate (or/and replicate) samples were applied to assessment of precision or accuracy (geochemical QA/QC) in analysis (Darabi-Golestan and Hezarkhani, 2018). The mean of the main-duplicate (or replicate) pairs are plotted along the X-axis, while the absolute difference of values is shown along the Y-axis for Au analysis at Glojeh. The two default lines corresponding to the $d_{99\%}$ and $d_{50\%}$ (or 1% and 50% error, relative standard deviation) were used onto the Thompson and Howarth (1976) scatterplot (Stanley, 2006). The lines $d_{99\%}$ and $d_{50\%}$ the 99th and 50th percentiles of the absolute difference between duplicate and replicate samples, which is represented as a function of concentration are respectively. Also, if all the samples are located under the $d_{99\%}$, it will be assumed that there is a normal distribution of error. Repeatability of measurements in the Glojeh deposit was assessed by the Thompson-Howarth error model. It indicates that the repeatability precision is under the control line of 1:1 (or y=x) for duplicate and replicate samples. The highest accuracy and lowest precision are related to the D10 and R10 samples which is located near the control line 99% for Au analyses, compared to the main sample (Figure 2).

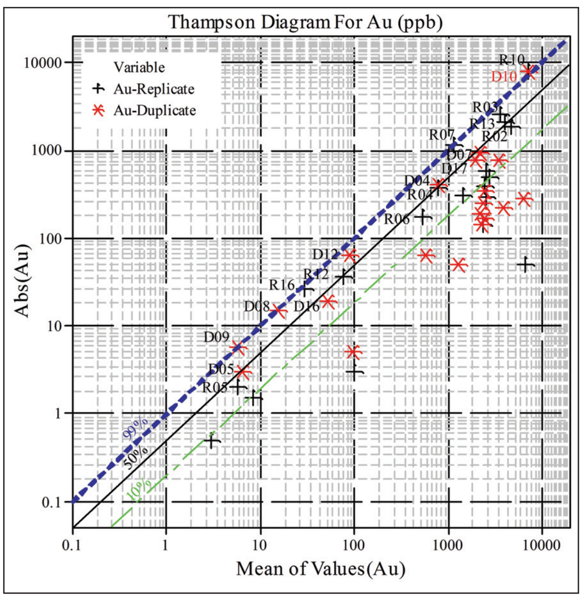


Figure 2- Estimation of precision of the Au analyses using diagram of Thompson and Howarth (1978). The mean of the replicate pairs is plotted along the X-axis, the absolute difference of the two results along the Y-axis.

In addition, just three replicate samples were taken for the porphyry deposit case study. These samples were analyzed for 44 elements that was determined by ICP-MS. Due to the small number of duplicate measurements, the accuracy of these samples were evaluated for all the elements. Quality control of main vs. replicate pair samples at Dalli deposit were given at figure 3, where all the main vs. replicate samples overlay on y=x line.

2.3. Log-ratio Transformation

The LRT has become a frequently used standard procedure to support subcompositional coherence within data (Aitchison, 1990; Greenacre, 2011). The ratios between the two components remain same, whether they are considered in a subcomposition or in the full composition (Greenacre, 2007). A family of log-ratio transformation consist of the clr, ilr, and alr space (Aitchison, 1986; Carranza, 2017; Egozcue et al., 2003; Thió-Henestrosa and Martín-Fernández, 2005). LRT may be done weighted (not for samples, but compositional elements can be weighted by the average level of each component) or unweighted (Greenacre, 2011). The frequently used unweighted version is applied for this paper based on the method that show sincoherence in the present data set using the Euclidean distance. The unweighted LRT has been proposed in detail by Kazmierczak (1985), afterward Aitchison and Greenacre (2002) suggests a biplot to a better representation of results.

In this study, the Ilr-transformed values were used for PCA analysis, while they were obtained by the CoDaPack software v 2.01 (available at http://www.compositionaldata.com/). The clr transformation (Aitchison, 1986) could be applied on composition of the data set (X), so can be written as:

$$\mathbf{Y} = clr(\mathbf{X}) = \{y_i\}_{i=1,....,n} = \left\{\log\frac{x_i}{g(x)}\right\}_{i=1,...,n} \quad \text{Eqs. (1)}$$

where all components have been divided by the g(X). The g(X) is the geometric mean of X_i components from the X symmetrically (Egozcue et al., 2003; Filzmoser et al., 2009*a*). It is calculated as follows:

$$g(\mathbf{X}) = \sqrt[n]{x_1.x_2...x_n} = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}} \quad \text{Eqs. (2)}$$

The resulted data (Y) based on this transformation show collinear relationship as $\sum_{i=1}^{n} y_i = 0$, i.e. the sum of each clr transformed data for a variable is zero (Aitchison, 1986), which rely on full rank data matrices, like standard robust covariance estimators

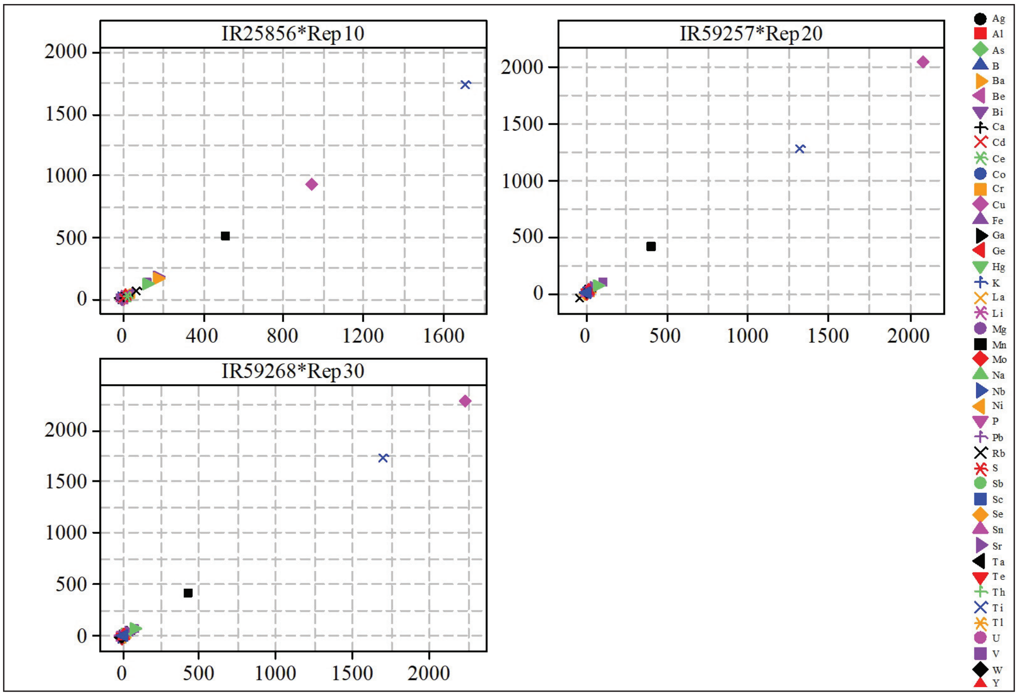


Figure 3- Quality control of main vs. replicated samples for ICP-MS analyzed data at Dalli deposit. All the main vs. replicated samples overlay on y=x line.

(Maronna et al., 2006). The ilr transformation could preserve all the advantages of clr transformation. In addition it can overcome on the disadvantageous treatment of data colinearity or the singularity of clr in Euclidean space (Egozcue et al., 2003; Pawlowsky-Glahn and Buccianti, 2011). For a composition X, the ilr transformation can be expressed as follows:

$$\mathbf{Y} = ilr(\mathbf{X}) = \{y_i\}_{i=1,\ldots,n-1} = (y_1, y_2, \ldots, y_{n-1}) \qquad \text{Eqs. (3)}$$

$$y_i = \frac{1}{\sqrt{i(i+1)}} \ln\left(\frac{\prod_{j=1}^{i} x_j}{(x_{i+1})^i}\right) \quad i = 1, \ldots, n-1 \qquad \text{Eqs. (4)}$$

It is clear based on the fundamental of Ilr transformation, only n-1 variables can be obtained for n components data (Thió-Henestrosa and Martín-Fernández, 2005). Therefore, the interpretation of new ilr transformed variables is not straightforward and there is no direct connection to the original variables (Filzmoser et al., 2009*b*). Whatever, the Ilr transformation does not show any clear simplest or canonical basis, but it has significant conceptual advantages (Pawlowsky-Glahn and Buccianti, 2011). The results obtained from ilr variables can be back-transformed to clr coefficients aided by an orthonormal basis (Egozcue et al., 2003; Liu et al., 2016). By using LRT, the subcompositional coherence is aided since the ratio between two data values (elements or samples) remains the same whether or not rows or columns are excluded from the table (Greenacre, 2007).

2.4. Correspondence Analysis (CA)

With application CA on the indicator matrix will provide two sets of factor scores for the rows (samples) and columns (elements). These factor scores are scaled such that their variance is equal to their corresponding eigenvalues (Abdi and Valentin, 2007; Greenacre and Blasius, 2006). At the CA method, it is possible to calculate the squared (χ^2) distance (Euclidean distances) between the observations or variables to look for the quantitative or qualitative objects of a given point or a set of points (Tekaia, 2016). It is supposed thatthe data matrix is X, and it has I ×J table of compositional data array. Each of these vectors was calculated from the following equations:

$$r_i = x_{i1} + x_{i2} + \ldots + x_{im} \qquad \text{Eqs. (5)}$$

$$c_j = x_{j1} + x_{j2} + \ldots + x_{jn} \qquad \text{Eqs. (6)}$$

Two diagonal matrices of $R_{n\times n}$ and $C_{m\times m}$ are defined as (Diday and Noirhomme-Fraiture, 2008; Greenacre, 1984; Ji et al., 2007; Ji et al., 1995):

$$R = \mathrm{diag}(r_1, r_2, \ldots, r_n) \qquad \text{Eqs. (7)}$$

$$C = \mathrm{diag}(c_1, c_2, \ldots, c_m) \qquad \text{Eqs. (8)}$$

The W and H matrices are defined as [(Eqs. (9) and (10)]:

$$W = R^{-1/2} X C^{-1/2} \quad \text{Eqs. (9)}$$

$$H = W^T W \quad \text{Eqs. (10)}$$

The H matrix eigenvalues and eigenvectors will be calculated. The information resulted in *p* (where $p \leq m-1$) dimension. Thus, the number of P eigenvalues were obtained as $0<\lambda p\leq...\leq\lambda_2\leq\lambda_1<1$ (Diday and Noirhomme-Fraiture, 2008; Gu et al., 2015). Afterward, eigenvectors [a_p] corresponding to each eigenvalue (λ_p) were calculated. Then the two matrices of Λ and A were generated based on eigenvalues and eigenvectors:

$$\Lambda_{(p\times p)} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p) \quad \text{Eqs. (11)}$$

$$A_{(m\times p)} = (a_1, a_2, ..., a_p) \quad \text{Eqs. (12)}$$

$U_{m\times p}$ and $V_{n\times p}$ matrices could introduce the relationship between variables and samples, respectively (Golestan et al., 2013; Ji et al., 2007). They were calculated as follows:

$$U = C^{-1/2} A \Lambda^{1/2} \quad \text{Eqs. (13)}$$

$$V = R^{-1/2} W A \quad \text{Eqs. (14)}$$

$$F = \begin{bmatrix} \mathbf{V} \\ \mathbf{U} \end{bmatrix} \quad \text{Eqs. (15)}$$

The $F_{(m+n)\times p}$ matrix (Eqs. 15) is a combination of the two important results of the CA, namely the U-matrix (R-mode; represented as a column-to-column or element-to-element criterion) and the V-matrix (Q-mode; represented as a row-to-row or sample-to-sample criterion) as a two important results of the CA (Darabi-Golestan and Hezarkhani, 2018; Ji et al., 2007). Therefore, the CA calculates the association or similarity between each variable (R-mode) and samples (Q-mode) within factor scores.

In another definition of CA, X is divided by its grand total x_{++} to obtain the so-called correspondence matrix $P=(1/x_{++})X$ (Greenacre, 2010). Let r and c vectors be the row and column of P, respectively. Therefore, the Eqs. (9) can be rewritten as Eqs. (16):

$$W = R^{-1/2} (P-rc^T) C^{-1/2} \quad \text{Eqs. (16)}$$

By performing a power (Box–Cox) transformation of the X matrix as $x_{ij}(\alpha) = x_{ij}^{\alpha}$, the CA is performed on the new correspondence matrix as X(α). The convergence and similar results of CA with LRT is a direct result of the Box-Cox transformation represented as $f(x) = (1/a)(X^{\alpha}-1)$, $\alpha>0$ or $f(x) = g(x)$ when $\alpha=0$ which f(x) tends to log(x) as tends to 0 (Greenacre, 2010, 2011).

2.5. Principal Component Analysis (PCA)

The PCA is a multivariate technique that applies frequently to illustrate the multidimensional data into lower dimensional factors without losing important information in the data (Collins and Ovalles, 1988; Fávaro et al., 2007). It could be done with both normalized and non-normalized data (Pawlowsky-Glahn et al., 2007). In this method, original data are transformed into a new set of data which led to a better results to explore the essential information (García-Izquierdo and Ríos-Rísquez, 2012; Ramasamy et al., 2013). The data matrix **X** consists of n samples (objects) which is analyzed for p elements (variables), in environmental, mining and geosciences studies. PCA decomposes the initial matrix **X** into two main produced matrices, introduced as a score matrix (**T**) and loading matrix (**P**). It can be expressed as following equation:

$$\mathbf{X} = \mathrm{T}_{n\times q} \mathbf{P}^{T}_{q\times p} + \mathrm{E}_{n\times p} \quad \text{Eqs. (17)}$$

where E is a matrix of residuals and $\mathbf{P}^T$ is the transpose of **P** (Março and Scarminio, 2007). Therefore, the major patterns of the data variance and correlations among measurements (according to their similarities) were showed in q vectors that were known as the principal components (PC; Bitner-Mathé and Klaczko, 1999; Darabi-Golestan et al., 2017; Karamanis et al., 2009; Março and Scarminio, 2007). These reduced factors indicate the largest amount of variability and associations between variables (Abdi et al., 2013; Golestan et al., 2013; Tokatli et al., 2014). Applying the visualization technique to show the factor scores and loading factors make this method comprehensible and could improve the prediction performance by visualizing large amounts of data (Hayton et al., 2004; Jeong et al., 2009). Reducing dimensionality of data at the PCA has a similar process to that of CA (Greenacre, 2007).

3. Results

3.1. Correspondence Analysis (CA)

The R-mode and Q-mode on the first and second dimension of U and V-matrices are represented in the biplot for the Glojeh (Figure 4) and Dalli (Figure 5)

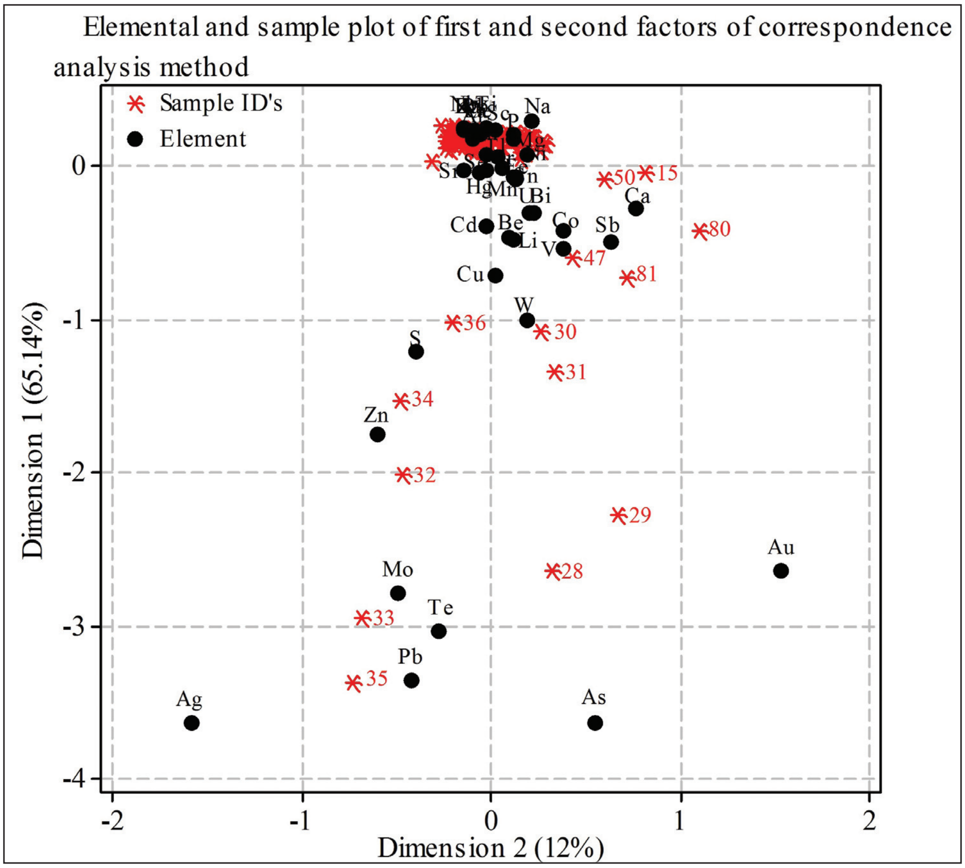


Figure 4- The two-dimensional visual representation of the two main factors; first and second factor explained 77.1% of the total variance at Glojeh using correspondence analysis. Both the row and column were displayed as principal coordinates in simultaneous plot.

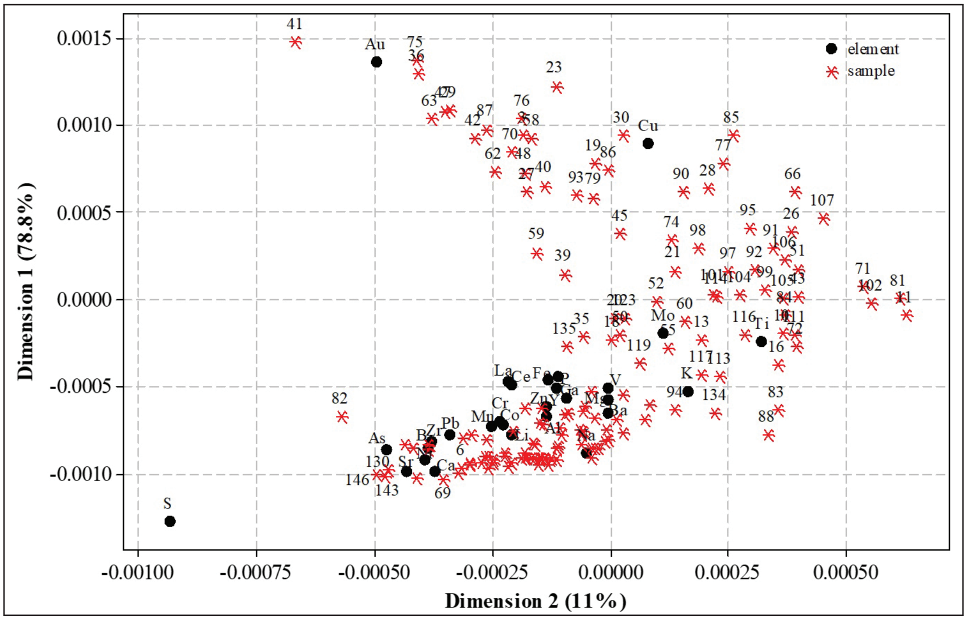


Figure 5- The two-dimensional visual representation of the two main factors; first and second factor explained 89.8% of the total variance at Dalli using correspondence analysis. Both the row and column were displayed as principal coordinates in symmetric plot.

deposits simultaneously. The Dimension 1 versus Dimension 2 (D1 versus D2) of Q-mode and R-mode combination analysis showed that two factors explained 77,14% of the total variance in Glojeh deposit. This percentage is sufficiently high to indicate a visual representation of the relationship between elements and samples. The first dimension indicates the strongest phase of mineralization that explains 65,14% of the total inertia and variance (Figure 4). On the other hand, the second dimension of variations (D2) indicates that Au and Ag show two different trends. The positive Au mineralization accompanied by Ca, As, and Sb association, while negative trend of Ag occurred with Zn, Mo, Pb, S, and Te elements.

The Dalli deposit is the second case study, which has been evaluated by 149 soil samples. The first dimension of CA indicates a strong mineralization of Cu-Au accompanied by high anomalous data for Mo, Ti, K, V and Fe, due to the related distances from other elements on this axis (Figure 5). On the other hand, the second phase of variations (Dimension 2) indicates that S, As, Au, and Sr show anomalous trends. The sample ID's of 36, 41, 75, 82, 130 and 146 show enrichment for them. The two-dimensional visual representation of indicated first and second dimension of CA explains 89.8% of the total variance, while the first dimension explains 78.8% of the total variance in district (Figure 5). The intuitive graphical figure 5 represents that D1 and D2 combination can display Cu-Au (Mo) mineralization, based on the highest distance and stretching outward from the center cluster, at D1 axis. As we discussed the first dimension was about seven times stronger than the second dimension.

3.2. Principal Component Analysis (PCA)

The geochemical compositional data were opened within the CoDaPack software v 2.01 and were transformed using the Isometric log-ratio (Ilr) transformations, prior to using any of the multivariate methods. The PCA for Ilr transformed data can be used to enhance the results of the CA for simultaneous study of elements in polymetallic and porphyry deposits when the data are compositional and suffering from closure problem.

Polymetallic vein deposit: At first, the PCA was done by log-transformation of geochemical data at Glojeh polymetallic vein deposit. The scree plot of eigenvalues indicates that the first (PC1) and second (PC2) significant factors of PCA explaining 56.3% of the total variance (Figure 6a). The PC1 explains the largest variance equal to 34.9% of the total variance. The PC1 indicates that the red group consists of Cu, Ag, Mo, Pb, Zn, Te, As, Au, Be, W, Se, and Cd elements are gained, while the blue group which is composed of Nb, K, La, Zr, Ba, Ce, Rb, Al, and Y elements is depleted at the Glojeh deposit. The factor loadings for these components are shown in figure 6a, while they are occurring between the –1 to +1 values and imply how the factors characterize the variables. Therefore, the association of Au, Ag, Cu, Pb, and Zn are more considerable for mineralization with As, Be, W, Te, and Mo in the Glojeh deposit.

On the other hand, based on the compositional properties of the data, all the data were transformed to Ilr space. The PCA method was applied to all above (43 elements) described elements within Ilr space. The first and second PC loadings of Ilr-transformed geochemical data explains 51,9% of the total variance in district (Figure 7a). The PC1 which includes the mineralization process explains 40,2% of total variation at the area (Figure 7a). Accordingly, it is comparable with the results of log-transformed data, by the higher value of variance (40,2% vs. 34,9%) and spread loading vectors and scores (Figure 7b).

Porphyry deposit: As a comparative case study, the Northern Dalli deposit has been investigated by log-transformed and LRT, based on 149 soil samples. The PC1 indicates 27,8% of total variation according to strong loading of Au and Cu against Li, Al, Mn, Ni, B, Co, Sr, and Cr elements. The PC2 shows 19% of total variance, indicated by V, Fe, To, Ga, Mg, K versus lower intense value of B, As and certainly La and Ce in this axis (Figure 8a). It is confirmed by score of samples such as 74 and 114 which are the strongest mineralized and depleted samples for Au and Cu respectively (Figure 8b). On the other hand, sample ID's of 11, 42 and 106 are the mineralized ones for V, Fe and Ti and 53, 142, and 131 are the loosest sample for them.

The biplot of the corresponding PC1 (29,5% of total variance) and PC2 (19,6% of total variance) of the Ilr transformed data, shows a clear separation between loadings and score of samples (Figure 9b). Therefore, the closure problem is overcome by Ilr-transformation, and curved shape of scores at figure 8b is depicted according to Ilr-transformation at figure 9b become spread apart and according to the spatial distribution of the observations become further apart.

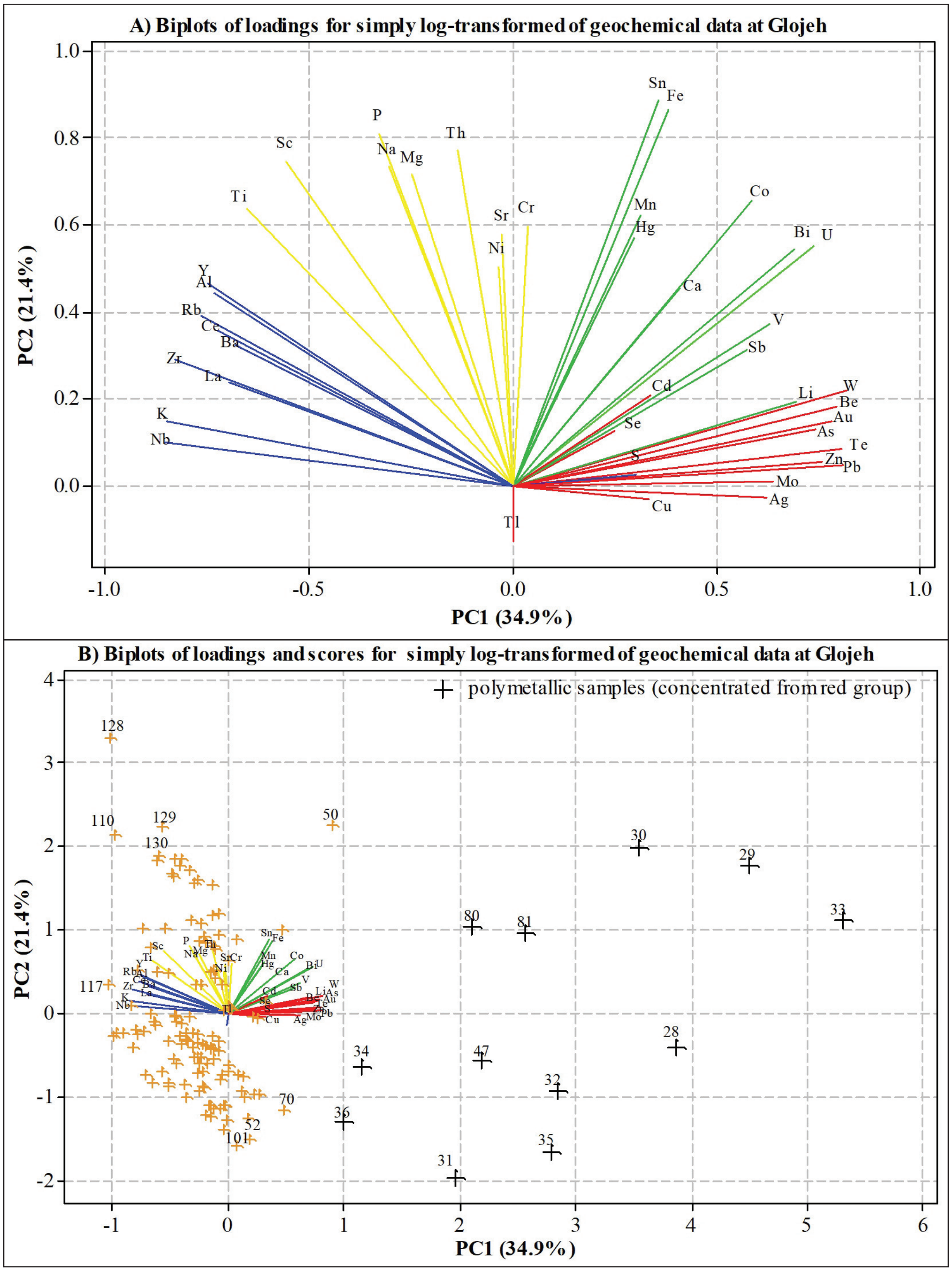


Figure 6- The loading plot (A), and symetric scores and loadings plot (B) on the PC1 versus PC2 of log-transformed data at Glojeh polymetallic deposit.

4. Discussion

4.1. CA

Polymetallic deposit: Zhu et al. (2011) proposed that the associations of Au-As or Au-Sb is common in different gold deposit, certainly in polymetallic veins that is consistent with the D2 here. The D1– D2 biplot (Figure 4) demonstrates the multi-element associations of Au-As-Ag-Pb-Te-Mo-Zn accompanied by S-Cu-W and Sb describing a polymetallic mineralization at Glojeh deposits directly. They have very high eccentricity from the center of the axis (0, 0) and sample ID's of 28, 29, 30, 31, 32, 33, 34, 35, 36 and 81 confirming this mineralization as indicator samples. The results show the potential for polymetallic

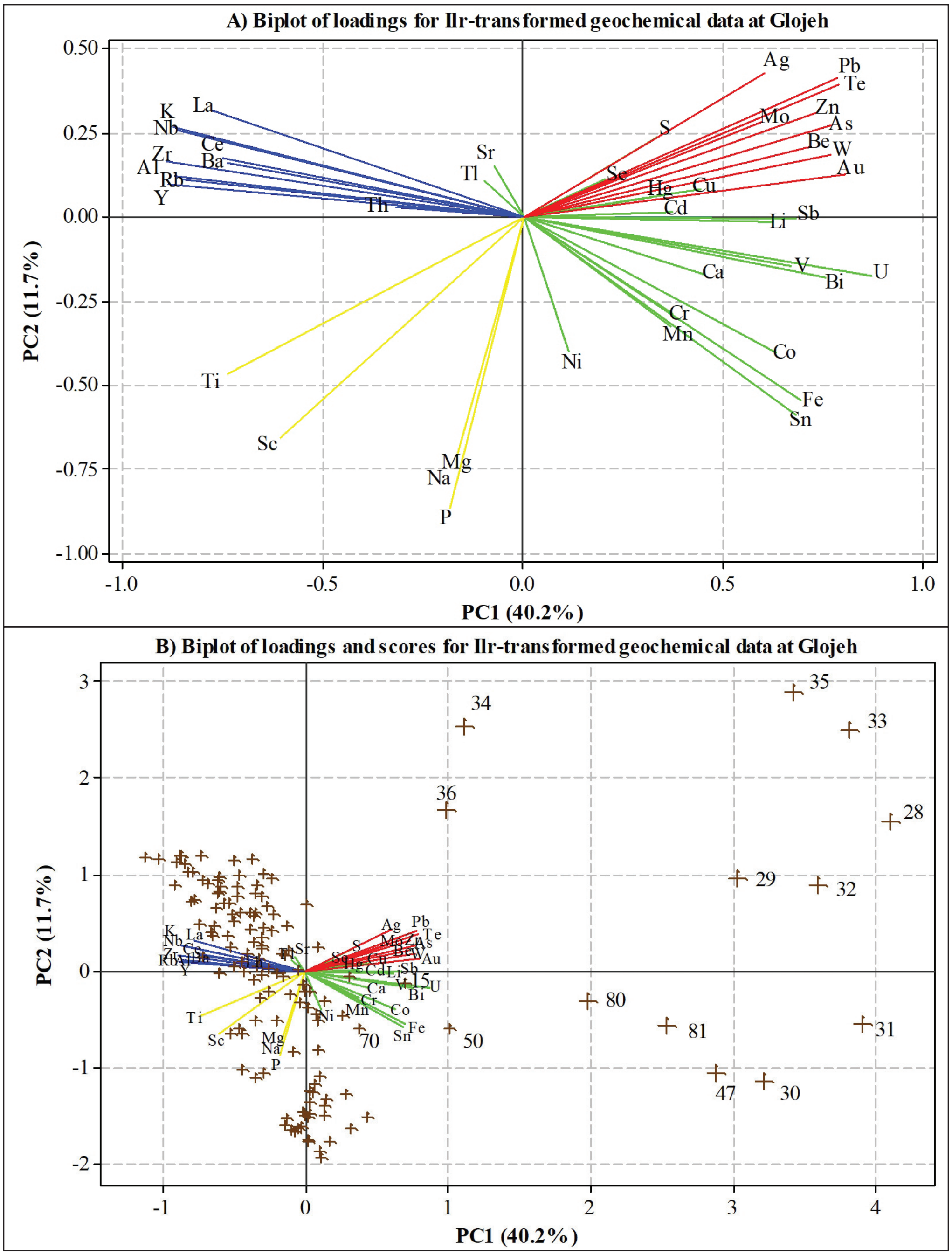


Figure 7- The loading plot (A), and symetric scores and loadings plot (B) on the PC1 versus PC2 of log-ratio transformed (Ilr space) data at Glojeh polymetallic deposit.

(highly anomalous elements are Ag, As, Pb and Au) veins mineralization. Cu and S association may be indicated that the polymetallic Au, Ag, Cu, Pb and Zn mineralization in the Glojeh deposit can be linked to a porphyry deposit in depth. Therefore, As, Te, Mo and minor S, W are more considerable for mineralization with Au, Ag, Cu, Pb and Zn in the Glojeh deposit (Darabi-Golestan and Hezarkhani, 2017).

Porphyry deposit: On the other hand, a porphyry mineralization at the Dalli deposit is verified by sample ID's of 23, 29, 35, 41, 42, 47, 63, 75, 76, 87 and too many samples. Figure 5 emphasizes on the S, As, Sr anomalous data according to the sample ID's of 82, 130 and 146. The Q-mode (score of samples) analysis approved the R-mode (score of elements) results, elemental associations, and mineralization in

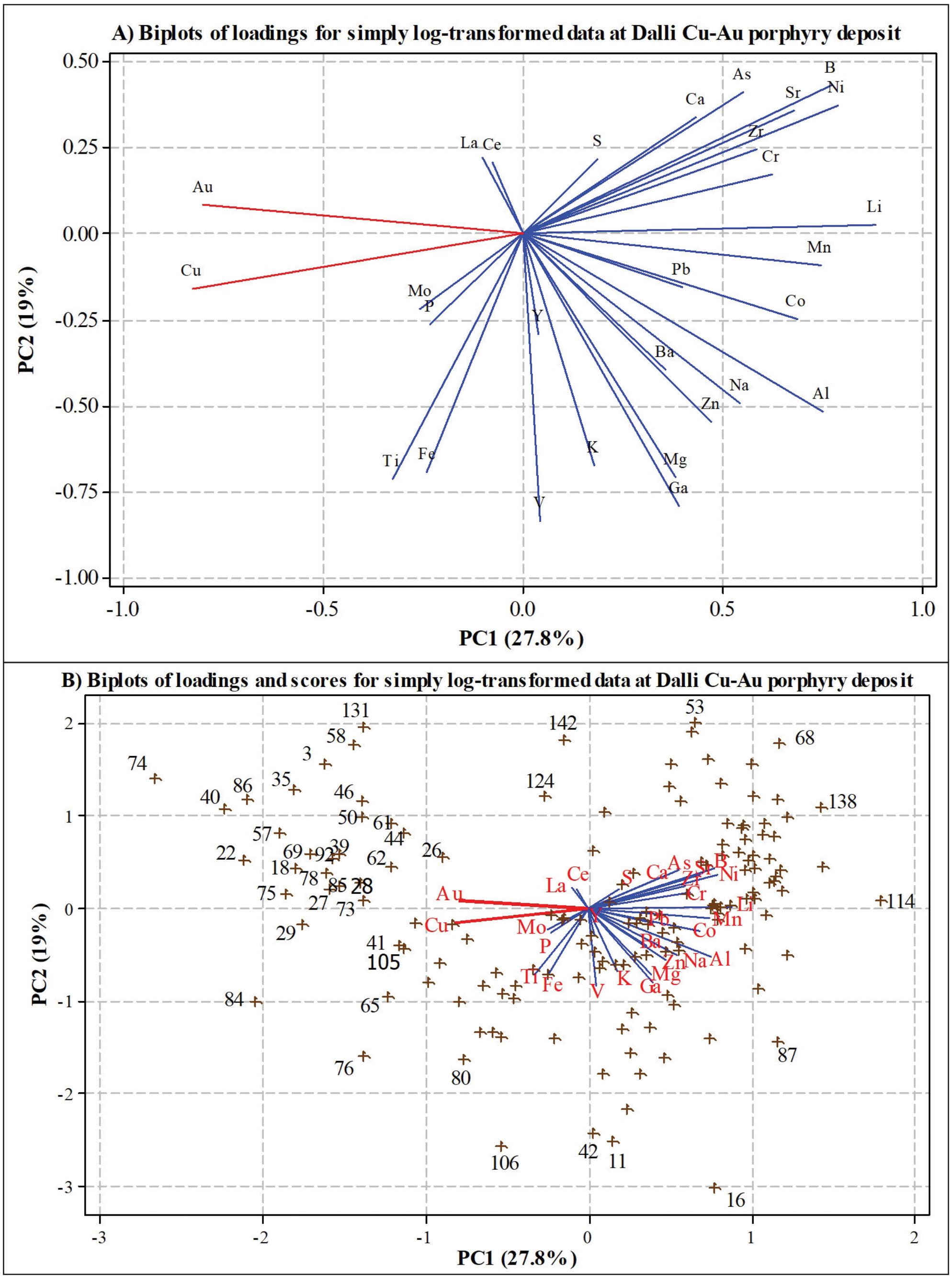


Figure 8- The loading plot (A), and symmetric scores and loadings plot (B) on the PC1 versus PC2 of log-transformed data at Dalli porphyry deposit.

order to avoid the creation of false anomalies due to different dimensions between or among the different elements. , the curved-shape configuration of the sample and element scores is a typical distribution for data closure problems in this coordinate.

4.2. PCA for log and Ilr Transformed Data

Polymetallic deposit: The biplot of the first two PC loadings for log-transformed data in a polymetallic deposit shows clearly the closure problem between

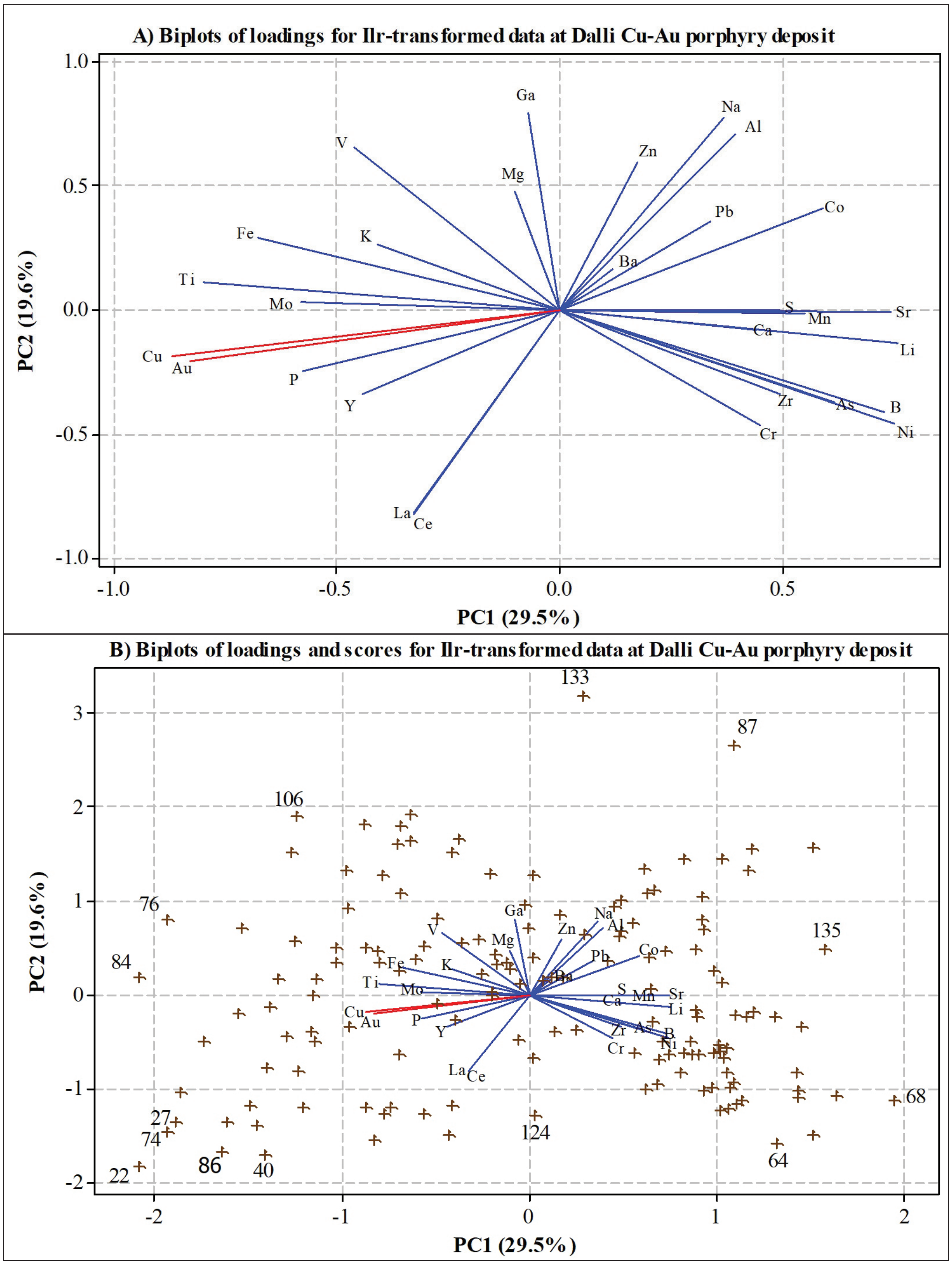


Figure 9- The loading plot (A), and symetric scores and loadings plot (B) on the PC1 versus PC2 of log-ratio transformed data at Dalli porphyry deposit.

loadings, as they are constrained in a semicircle (Figure 6a). The biplot of loadings and scores for log-transformed data (Figure 6b) shows that the red group is concentrated in sample ID's of 28, 29, 30, 31, 32, 33, 34, 35, 36, 47, 50 and 81. These samples indicate vein, veinlet and brecciated zones, while they are depleted from blue group. By the way, it is resulted that the samples 128, 110,129 and 130 are concentrated for the yellow group. The yellow and green groups are more relevant by the PC2. On the other hand, the visualized plot of multi-elemental association from PC1 vs. PC2 indicates that the closure problem in compositional data is posed and the loadings are spread in a circular, using Ilr-transformed data. The combination of loadings and scores at figure 7b indicates a better relationship and interaction between elemental loadings and sample scores, whereas the closure effect problem is minimized. Figure 7b could ease the interpretation of geochemical dispersion pattern of elements and samples.

Both the PCA opened using log-transformed and Ilr-transformed geochemical data indicate sample ID's of 28, 29, 30, 31, 32, 33, 34, 35, 36, 47, 80 and 81 as anomalous samples that taken from polymetallic (Au, Ag, Cu, Pb, and Zn) veins. But Ilr-transformed data show better inherent in compositional and overcome on the closure effect of loadings and scores.

Porphyry deposit: The loadings (Figure 8a) and scores (Figure 8b) of the two first PC's (from log-transformed data) show the closure problem, give that all the loadings except Au, Cu, La, and Ce are restricted to a semicircle (180°) and score of samples distributed closely. Because of data closure effect problem there is compositional data at Dalli deposit, the score of samples shows a strong curve shape as shown in figure 8b.

Using appropriate Ilr transformations prior to PCA analysis, the loadings of compositional data plot along all directions on the biplot (Figure 9a), while the loadings of the log-transformed data are mostly constrained in a semi-circle (Figure 8a). Figure 9a suggests that Mo-Ti-Fe-K-V and Y-La-Ce associations that represented in PC1 vs. PC2 biplot of Ilr space, show more realistic signature in Cu-Au porphyry deposit, compared to their associations represented in simply log-transformed biplot at figure 8a. This interpretation is supported by the different spatial distribution of scores at figure 8b and 9b. Figure 8b demonstrates that the known Cu-Au-(Mo) deposit in the study area has stronger positive correlation with the Mo-Ti-Fe-K-V and Y-La-Ce associations, which is concordant with Cu-Au-(Mo) porphyry deposit. With a comparison to log-transformed, loadings of the ilr-transformed show a strong Cu-Au mineralization followed by lower Mo concentrations, that are supported by too many scores of samples. Previous studies at this deposit, revealed a clear trend of sequential enrichment of Mo→Cu→Au from depth to surface (Darabi-Golestan et al., 2013*a*), that is covered by the results of ilr-transformed data from soil samples at this study. The intensely mineralized Au-Cu-(Mo) occur at sample ID's of 22, 28, 35, 40, 46, 74, 84, 86 and many others, and these are in connection with Cu-Au porphyry mineralization at the deposit (Figure 8b and 9b). According to lithological soil samples using CA (Figure 5), PCA analysis of log-transformed (Figure 8) and Ilr-transformed data (Figure 9), this trend is confirmed by higher intensity of (Au, Cu)>Mo. Darabi-Golestan et al. (2013*a*) proposed that the higher concentrations of Au and Cu are associated with mineralized zone within quartz diorite porphyry rocks that were enriched partly from Fe and Ti that were dominantly covered by potassic (K) alterations, which are more consistent with the CA and PCA of LRT data of graphical plot at the present study.

5. Conclusion

Most of mineralization has occurred because of the elemental interaction effect and physico-chemical exchanges within or/and between fluids and host rock. Therefore, the closure effects problem needs to be considered to understand the geochemical dispersion pattern of elements. Subcompositional coherence is overcome by LRT (certainly, Ilr transformation) for two genetic types of polymetallic vein and porphyry Au deposits at this study, while simply used log-transformed values are not good enough. Accordingly, by application Ilr-transformed data compared to log-transformed data the PC1 are improved from 34,9% to 40,2% at Glojeh and 27,8% to 29,5% at Dalli deposit. The ability to handle zero values in the data matrix and determining an elemental eccentricity from the center of axis based on Euclidean distances are the advantages of CA method. On the other hand, a clear picture of loading factors which spread in a full circle providing subcompositional coherence are the competitive advantages of PCA based on Log-Ratio transformation. Accordingly, Ag, Au, As, Pb, Te, Mo and rather S, W, Cu are enriched as polymetallic elements at Glojeh, while Au-Cu (Mo) indicates a porphyry deposit occurred in Dalli. Both anomalies are accompanied by a lot of samples that show enrichment for them. However, the closure problem between samples and elements is solved by LRT and these techniques show significant ability to draw an inference in such deposits.

Acknowledgements

The authors are grateful to the Iranian Mines and Mining Industries Development and Renovation Organization (IMIDRO) for their permission to have access to Glojeh deposit dataset. The authors are also thankful to Mr. Fattahi and Mr. Hemmati for their encouragement and valuable help hugely.

References

Abdi, H., Valentin, D. 2007. Multiple correspondence analysis. Encyclopedia of measurement and statistics, 651-657.

Abdi, H., Williams, L. J., Valentin, D. 2013. Multiple factor analysis: Principal component analysis for multi-table and multi-block data sets. Computational Statistics 5, 149-179.

Aitchison, J. 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society. Series B (Methodological), 139-177.

Aitchison, J. 1983. Principal component analysis of compositional data. Biometrika, 57-65.

Aitchison, J. 1986. The statistical analysis of compositional data. Monographs on Statistics and Applied Probability, 416 p.

Aitchison, J. 1990. Relative variation diagrams for describing patterns of compositional variability. Mathematical Geology 22, 487-511.

Aitchison, J., Greenacre, M. 2002. Biplots of compositional data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 51, 375-392.

Akbarpour, A., Gholami, N., Azizi, H., Torab, F. M. 2013. Cluster and R-mode factor analyses on soil geochemical data of Masjed-Daghi exploration area, northwestern Iran. Arabian Journal of Geosciences 6, 3397-3408.

Bitner-Mathé, B. C., Klaczko, L. B. 1999. Heritability, phenotypic and genetic correlations of size and shape of Drosophila mediopunctata wings. Heredity 83, 688-696.

Carranza, E. J. M. 2009. Controls on mineral deposit occurrence inferred from analysis of their spatial pattern and spatial association with geological features. Ore Geology Reviews 35, 383-400.

Carranza, E. J. M. 2011. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. Journal of Geochemical Exploration 110, 167-185.

Carranza, E. J. M. 2017. Geochemical Mineral Exploration: Should We Use Enrichment Factors or Log-Ratios? Natural Resources Research 26, 411-428.

Collins, M., Ovalles, F. 1988. Variability of northwest Florida soils by principal component analysis. Soil Science Society of America Journal 52, 1430-1435.

Croux, C., Haesbroeck, G. 2000. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. Biometrika 87, 603-618.

Darabi-Golestan, F., Ghavami-Riabi, R., Asadi-Harooni, H. 2013*a*. Alteration, zoning model, and mineralogical structure considering lithogeochemical investigation in Northern Dalli Cu–Au porphyry. Arabian Journal of Geosciences 6, 4821-4831.

Darabi-Golestan, F., Ghavami-Riabi, R., Khalokakaie, R., Asadi-Haroni, H., Seyedrahimi-Nyaragh, M. 2013*b*. Interpretation of lithogeochemical and geophysical data to identify the buried mineralized area in Cu-Au porphyry of Dalli-Northern Hill. Arabian Journal of Geosciences 6, 4499-4509.

Darabi-Golestan, F., Hezarkhani, A. 2016. High precision analysis modeling by backward elimination with attitude on interaction effects on Au (Ag)-polymetallic mineralization of Glojeh, Iran. Journal of African Earth Sciences 124, 505-516.

Darabi-Golestan, F., Hezarkhani, A. 2017. R- and Q-mode multivariate analysis to sense spatial mineralization rather than uni-elemental fractal modeling in polymetallic vein deposits. Geosystem Engineering, 1-10.

Darabi-Golestan, F., Hezarkhani, A. 2018. Evaluation of elemental mineralization rank using fractal and multivariate techniques and improving the performance by log-ratio transformation. Journal of Geochemical Exploration 189, 11-24.

Darabi-Golestan, F., Hezarkhani, A., Zare, M. 2017. Assessment of 226 Ra, 238 U, 232 Th, 137 Cs and 40 K activities from the northern coastline of Oman Sea (water and sediments). Marine Pollution Bulletin 118, 197-205.

David, M., Campiglio, C., Darling, R. 1974. Progresses in R-and Q-mode analysis: correspondence analysis and its application to the study of geological processes. Canadian Journal of Earth Sciences 11, 131-146.

Diday, E., Noirhomme-Fraiture, M. 2008. Symbolic data analysis and the SODAS software, Wiley Online Library, Namur, Belgium.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcelo-Vidal, C. 2003. Isometric logratio transformations for compositional data analysis. Mathematical Geology 35, 279-300.

Fávaro, D., Damatto, S., Moreira, E., Mazzilli, B., Campagnoli, F. 2007. Chemical characterization and recent sedimentation rates in sediment cores from Rio Grande reservoir, SP, Brazil. Journal of Radioanalytical and Nuclear Chemistry 273, 451-463.

Filzmoser, P., Hron, K. 2008. Outlier detection for compositional data using robust methods. Mathematical Geosciences 40, 233-248.

Filzmoser, P., Hron, K., Reimann, C. 2009*a*. Principal component analysis for compositional data with outliers. Environmetrics 20, 621-632.

Filzmoser, P., Hron, K., Reimann, C. 2009*b*. Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Science of the Total Environment 407, 6100-6108.

Filzmoser, P., Hron, K., Reimann, C., Garrett, R. 2009*c*. Robust factor analysis for compositional data. Computers ve Geosciences 35, 1854-1861.

Filzmoser, P., Hron, K., Reimann, C. 2010. The bivariate statistical analysis of environmental (compositional) data. Science of the Total Environment 408, 4230-4238.

García-Izquierdo, M., Ríos-Rísquez, M. I. 2012. The relationship between psychosocial job stress and burnout in emergency departments: an exploratory study. Nursing outlook 60, 322-329.

Golestan, F. D., Hezarkhani, A., Zare, M. 2013. Interpretation of the Sources of Radioactive Elements and Relationship between them by Using Multivariate Analyses in Anzali Wetland Area. Geoinformatics & Geostatistics: An Overview 1, 1-10.

Greenacre, M. 2007. “Correspondence analysis in practice,” CRC press.

Greenacre, M. 2010. Log-ratio analysis is a limiting case of correspondence analysis. Mathematical Geosciences 42, 129-134.

Greenacre, M. 2011. Measuring subcompositional incoherence. Mathematical Geosciences 43, 681-693.

Greenacre, M., Blasius, J. 2006. “Multiple correspondence analysis and related methods,” CRC press.

Greenacre, M. J. 1984. “Theory and applications of correspondence analysis.”

Gu, X., Liu, C., Wang, S., Zhao, C. 2015. Feature extraction using adaptive slow feature discriminant analysis. Neurocomputing 154, 139-148.

Hayton, J. C., Allen, D. G., Scarpello, V. 2004. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational research methods 7, 191-205.

Jeong, D. H., Ziemkiewicz, C., Fisher, B., Ribarsky, W., Chang, R. 2009. iPCA: An Interactive System for PCA-based Visual Analytics. In “Computer Graphics Forum”, Vol. 28, pp. 767-774. Wiley Online Library.

Ji, H., Zhu, Y., Wu, X. 1995. Correspondence cluster analysis and its application in exploration geochemistry. Journal of geochemical Exploration 55, 137-144.

Ji, H., Zeng, D., Shi, Y., Wu, Y., Wu, X. 2007. Semi-hierarchical correspondence cluster analysis and regional geochemical pattern recognition. Journal of Geochemical Exploration 93, 109-119.

Karamanis, D., Ioannides, K., Stamoulis, K. 2009. Environmental assessment of natural radionuclides and heavy metals in waters discharged from a lignite-fired power plant. Fuel 88, 2046-2052.

Kazmierczak, J. 1985. Analyse logarithmique: deux exemples d’application. Revue de statistique appliquée 33, 13-24.

Liu, Y., Cheng, Q., Zhou, K., Xia, Q., Wang, X. 2016. Multivariate analysis for geochemical process identification using stream sediment geochemical data: A perspective from compositional data. Geochemical Journal 50, 293-314.

Março, P. H., Scarminio, I. S. 2007. Q-mode curve resolution of UV–vis spectra for structural transformation studies of anthocyanins in acidic solutions. Analytica chimica acta 583, 138-146.

Maronna, R., Martin, R. D., Yohai, V. 2006. “Robust statistics: Theory and Methods,” John Wiley & Sons, Chichester. ISBN.

Martín-Fernández, J. A., Barceló-Vidal, C., Pawlowsky-Glahn, V. 2003. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. Mathematical Geology 35, 253-278.

Olofsson, T., Andersson, S., Sjögren, J.-U. 2009. Building energy parameter investigations based on multivariate analysis. Energy and Buildings 41, 71-80.

Pawlowsky-Glahn, V., Egozcue, J. J. 2006. Compositional data and their analysis: an introduction. Geological Society, London, Special Publications 264, 1-10.

Pawlowsky-Glahn, V., Buccianti, A. 2011. “Compositional data analysis: Theory and applications,” John Wiley & Sons.

Pawlowsky-Glahn, V., Egozcue, J. J., Tolosana Delgado, R. 2007. Lecture notes on compositional data analysis.

Pommer, L., Fick, J., Sundell, J., Nilsson, C., Sjöström, M., Stenberg, B., Andersson, B. 2004. Class separation of buildings with high and low prevalence of SBS by principal component analysis. Indoor Air 14, 16-23.

Ramasamy, V., Sundarrajan, M., Paramasivam, K., Meenakshisundaram, V., Suresh, G. 2013. Assessment of spatial distribution and radiological hazardous nature of radionuclides in high background radiation area, Kerala, India. Applied Radiation and Isotopes 73, 21-31.

Reimann, C., Filzmoser, P., Garrett, R., Dutter, R. 2011. “Statistical data analysis explained: applied environmental statistics with R,” John Wiley & Sons.

Reimann, C., Filzmoser, P., Fabian, K., Hron, K., Birke, M., Demetriades, A., Dinelli, E., Ladenberger, A., Team, T. G. P. 2012. The concept of compositional data analysis in practice—total major element concentrations in agricultural and grazing land soils of Europe. Science of the total environment 426, 196-210.

Silverman, J. D., Washburne, A., Mukherjee, S., David, L. A. 2016. A phylogenetic transform enhances analysis of compositional microbiota data. bioRxiv, 072413.

Stanley, C. R. 2006. On the special application of Thompson–Howarth error analysis to geochemical variables exhibiting a nugget effect. Geochemistry: Exploration, Environment, Analysis 6, 357-368.

Templ, M., Filzmoser, P., Reimann, C. 2008. Cluster analysis applied to regional geochemical data: problems and possibilities. Applied Geochemistry 23, 2198-2213.

Thió-Henestrosa, S., Martín-Fernández, J. 2005. Dealing with compositional data: the freeware CoDaPack. Mathematical Geology 37, 773-793.

Thompson, M., Howarth, R. J. 1976. Duplicate analysis in geochemical practice. Part I. Theoretical approach and estimation of analytical reproducibility. Analyst 101, 690-698.

Tokatlı, C., Köse, E., Çiçek, A. 2014. Assessment of the effects of large borate deposits on surface water quality by multi statistical approaches: A case study of Seydisuyu Stream (Turkey). Polish Journal of Environmental Studies 23, 1741-1751.

Zhu, Y., An, F., Tan, J. 2011. Geochemistry of hydrothermal gold deposits: a review. Geoscience Frontiers 2, 367-374.

Zuo, R., Xia, Q., Wang, H. 2013. Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. Applied geochemistry 28, 202-211.