Comparison of multivariate statistical analysis and traditional techniques as anomaly detection methods in Qare Char, northwest of Iran

Mehdi Badel[1] & Siamak Abolhassani[2]

[1] School of Mining Eng., Faculty of Eng., University of Tehran.

[2] Faculty of Eng., Islamic Azad University (Southern Tehran Branch).

ABSTRACT

Detecting the mineralization component is one of the main subjects in geochemical exploration. Various techniques have been introduced to explore favorite zones and separate the anomaly from background, such as bivariate or multivariate statistical analysis. The main objectives is to identify the spatial variability and main sources of some special elements in Qare Char, in northwest of Iran, by conducting multivariate statistical analyses, including principle component analysis assisted with GIS tools. The analytical results of some rare elements in 1099 lithogeochemical samples taken from the study area have been used to identify the main components of lithological components and separate them from mineralization components. Finally, some favorite zones have been detected and then mapped.

Keywords: *Mineralization Component; Spatial Variability; Multivariate Statistical Analysis; Principle Component Analysis.*

1. Introduction

As a traditional method of data interpretation in geochemical prospecting, in univariate analysis the computer is often performing functions that could be undertaken manually, although once the data have been encoded, the human is generally slower and less accurate. Given a large amount of samples analyzed for a number of chemical elements, the essence of interpretative strategy is to make some assumptions about the geochemical process causing the observed variations in the data in order to make the interpretation procedure manageable. The more theoretical or statistical assumptions one is willing to impose on the data, the more the complexity of the analysis can be reduced [1].

The traditional way of anomaly detections is by first choosing a univariate threshold value for each element. All concentrations above this threshold are then anomalous. Several methods exist for defining the threshold, such as: two standard deviations above the mean value or two times a chosen background value. However, as Bölviken (1971) and Tennant and White (1959) showed, there is generally not a single threshold concentration, but rather a distribution of background values and a distribution of anomalous values. If a given value has a higher probability of belonging to the anomalous distribution than the background distribution, the corresponding sample site is termed anomalous for that element. These traditional ways can be misleading if there is several subareas with different background and anomalous populations in the survey area [1].

In addition to the problem of having several background and anomalous populations within the same survey area, univariate methods only consider single elements and may thus fail to

detect anomalies if the interaction between several elements is the feature which characterizes the anomalous locality.

2. Prospecting in Qare Char

The Qare Char area with 8.6 km^2 extent locates at south west of Saqez in Kurdistan province of Iran. Geology of the area is not very complicated. It has been covered widely with granite and gneiss and the gn^1 and gn^2 units in Fig. 1 represent that gneiss is more. The gr unit has been injected in southeast of the area. For more detail about the geology of the area has been presented in Fig. 1.

The study area has been sampled in a regularly spaced in a sampling network plan which was 100m×100m and 50m×50m and all samples were analyzed for 10 elements Ag, As, Au, Cu, Mo, Pb, Sn, Sb, Bi and Zn by atomic absorption method. The outliers have been replaced and the statistical distribution patterns of the measured elements were checked by calculating the skewness and kurtosis coefficients. Results (not shown) indicate that all variables are positively skewed and their kurtosis coefficients are significantly larger than zero. Since log transformation could significantly reduce both the skewness and kurtosis coefficients for all variables, all concentrations were transformed with the 'ln' operator before the analysis to stabilize their variations. Bi and Sb removed from further process because these elements have a large number of censored data.

3. Univariate analysis

After dedicating each sample to its own rock assemblages (in this study the rocks have been divided into 4 populations based on lithology), the background was determined for each population, on the basis of median for each element in lithological unit, and finally the enrichment index was calculated for each element. This is a direct technique to remove the lithological component for a better understanding of mineralization. The enrichment index for Au has been mapped and showed in Fig. 2. As can be seen in this figure, there are some main zones which can be offered as favorite zones.

4. Multivariate Statistical Analysis

Methods of multivariate statistical analysis can simultaneously provide multiple variables analysis. The analysis of one, two or even three variables could be imagined, displayed as graphics and analyzed, but sometimes in some geochemical exploration problems with 10 or even 20 variables, it has been very difficult to review the relationship between variables. In such cases is necessary to use multivariate statistical methods to reduce dimensions in space so that the results with a number of new dimensions of the case less than before, can describe a large part of variability of data. The number of samples in the populations under investigation is an important point in the multivariate statistics should pay attention to it. Multivariate methods usually require a large number of samples [1].

4.1. Correlation

Correlation and dependence are any of a broad class of statistical relationships between two or more random variables or observed data values. In general, correlation can be thought of as a normalized measure of covariance. Rank correlation coefficients, such as Spearman's

rank correlation coefficient measure the extent to which, as one variable increases, the other variable tends to increase, without requiring that increase to be represented by a linear relationship. If, as the one variable increase, the other decreases, the rank correlation coefficients will be negative. It is common to regard these rank correlation coefficients as alternatives to Pearson's coefficient, used either to reduce the amount of calculation or to make the coefficient less sensitive to non-normality in distributions.

The Spearman correlation coefficient is often thought of as being the Pearson correlation coefficient between the ranked variables. The Spearman correlation coefficients among the elements, as shown in Table 1, reveal that most of the elements are poorly correlated with one another. It must say that the accuracy and precision of the analytical method, which the elements have been analyzed, plays an important role to reveal the real relationship among the variables. In this case, as Table 1 shows, the relationship between Cu, Zn and Pb and also Au and As is very weak and this is very odd. These elements usually have a highly relationship with each other because some of them are pathfinders of another.

4.2. Principle component analysis and factor analysis

So-called "factor" analysis was one of the first multivariate techniques to become widely used among geochemists. The intention underlying the use of principal components or factor analysis in exploration geochemistry has generally been to separate the element associations inherent in the structure of the correlation matrix into a number of groups of elements that together account for the greater part of the observed variability of the original data. The aim being to represent the large number of elements in the original data by a smaller number of "factors", each of which is a linear function (transformation) of the element concentrations, thus giving a greater efficiency in terms of information compression over the original data, and hopefully also gaining something in interpretability [1].

The relation between a set of elements can be described by that proportion of the total variance which is common to them, since it is this which produces their common variance.

Although the techniques of principal components analysis (PCA) and factor analysis (FA) have much in common and are often referred to in the literature as one and the same thing there are important mathematical and conceptual differences between them. Each component in PCA can be said to maximize the common variance. PCA is thus variance-oriented whereas FA is correlation-oriented. For geochemical data this could suggest that PCA is favorable in situations in which the range of variation of the elements is characteristic of the geochemical environment, whereas FA is favorable in situations in which element associations characterize the geochemical environment.

In this study, the FA was performed on the homogenous data (Ln transformation of the enrichment indices) and, as it is presented in Table 2, 4 factors have been extracted by principle component method. These factors fulfill about 68 percent of the total variability after varimax rotation.

As shown in Table 2, component one explains only about 20% of the total variation and about 70% of the total variation could be accounted for by the four components. The 1st component could be largely attributed to Zn and Ag. The 2nd component mainly represents Au, As and Cu (Table 2).

The relationships among the principal components and mineralization were then explored. As it can be seen in Fig.3, the multiplication of Au, As and Cu has been represented and some main favorable areas have been discovered. Fig. 4 represents the scores of the 2nd component which attributed to these three elements (Au, As and Cu). It rejects some areas offered from Fig. 3 and shows other areas which can be highly focused in further and detailed studies because results from FA have more credit due to data processing through large number of variables and detect their probable relationship.

5. Conclusion

The results coming from the enrichment plots, which have been produced by a univariate analysis, can not be trusted since in geochemical surveying of trace elements, cross-correlations of them must be surveyed to reveal the main process of rock formation (lithological components) and mineralization and make them understandable.

Regarding to interpretation of rock and mineral formations in different regions, one of the significant issues in the multivariate analysis is the chemical analysis of different elements (trace and major) and its accuracy and precision is also important. In this study, just by having eight trace elements analysis in a large number of samples, it is hard to interpret the area and represent the main process of both rock and mineral formations. By analyzing the main rock forming elements such as Al, Si, Fe, S, Na, Ca, K, and so on it can be predicted that in which chemical phases these trace elements have been formed and whether more detailed studies are cost-effective or not.

As can be seen from the results presented on the maps in this paper, it can be concluded that a map provided by the univariate analysis can not be trusted and it is necessary to pay special attention to components of exploring and detecting elements and their corresponding cross-correlation.

References

[1] Govett, G.J.S. (1983), Handbook of Exploration Geochemistry, Vol. 2 (Statistical Data Analysis in Geochemical Prospecting), Amesterdam, Elsevier, P 437.

[2] Hassanipak A.A., Sharafodin M., (2005), Exploration Data Analysis, Tehran University press, P.987.

[3] Novin Ekteshaf Consulting Eng. Co. (2007), Report of Geochemical prospecting in Qare Char, Geological survey of Iran.

[4] Wackernagel, H., 1995. "Multivariate geostatistics – An introduction with applications", Springer (Berlin), 256p.

Table.1. Correlation matrix of eight elements in Qare Char (Spearman correlation coefficient)

	Au	Cu	Pb	Zn	Ag	Sn	Mo	As
Au	**1**							
Cu	**0.213****	**1**						
Pb	**0.007**	**0.029**	**1**					
Zn	**0.052**	**0.218****	**0.086****	**1**				
Ag	**0.078****	**0.172****	**0.159****	**0.523****	**1**			
Sn	**0.112****	**-0.130****	**0.213****	**0.094****	**0.221****	**1**		
Mo	**0.094****	**-0.082****	**0.188****	**-0.217****	**0.004**	**0.251****	**1**	
As	**0.190****	**0.419****	**-0.005**	**0.359****	**0.294****	**-0.064***	**-0.118****	**1**

Table.2. Total variance explained and component matrixes for eight elements in Qare Char (Rotation method: Varimax with Kaiser Normalization)

Component	**Rotation Sums of Squared Loadings**		
	Total	% of Variance	Cumulative %
1	1.615	20.189	20.189
2	1.499	18.743	38.932
3	1.394	17.431	56.363
4	1.006	12.577	**68.940**

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Communalities
Au	-0.019	**0.682**	0.398	-0.202	0.664
Cu	0.046	**0.757**	-0.171	0.239	0.661
Pb	0.111	0.053	0.171	**0.932**	0.914
Zn	**0.819**	0.129	-0.175	0.086	0.725
Ag	**0.770**	0.087	0.202	0.057	0.645
Sn	0.364	-0.196	**0.690**	0.126	0.663
Mo	-0.185	0.077	**0.786**	0.110	0.671
As	0.412	**0.625**	-0.110	-0.018	0.573

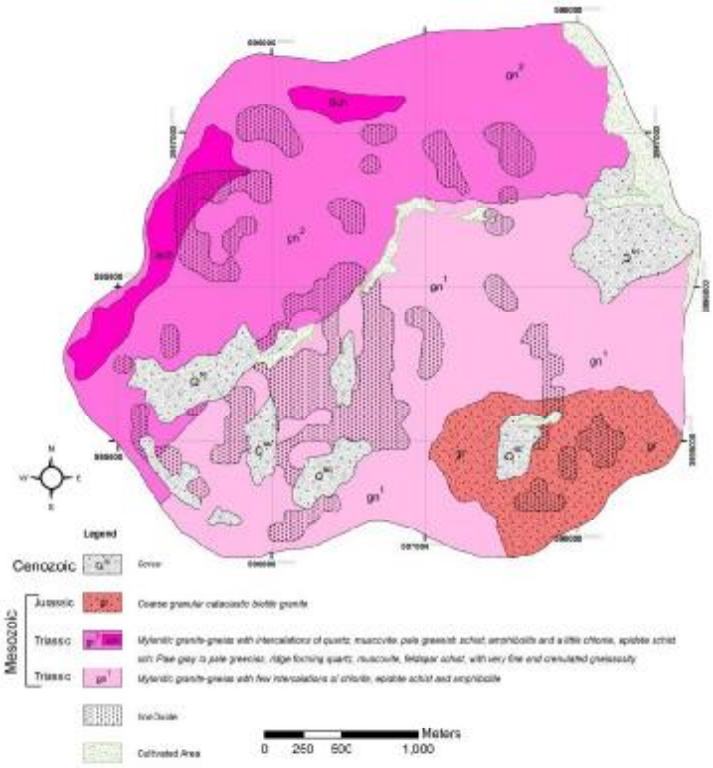

Figure.1. Geological map of the study area [3].

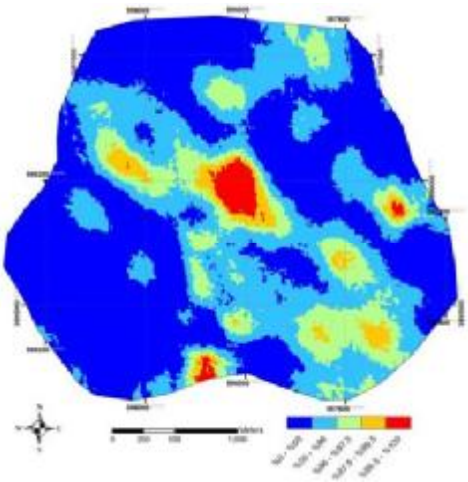

Figure.2. Kriged map of Enrichment index of Au.

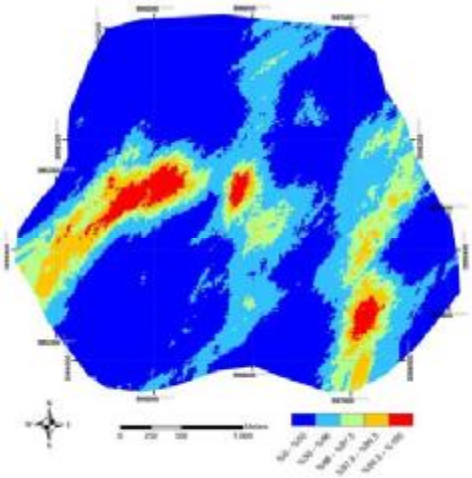

Figure.3. Kriged map of Au×As×Cu (Multiplication of Enrichment indices).

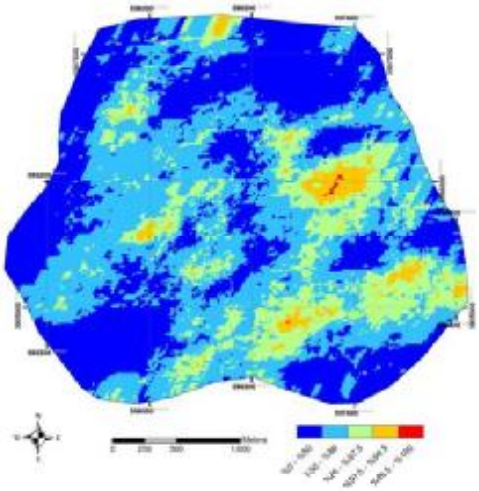

Figure.4. Kriged map of Factor II scores.