Journal of African Earth Sciences 134 (2017) 557–563

Contents lists available at ScienceDirect

Journal of African Earth Sciences

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

A comparison study on detection of key geochemical variables and factors through three different types of factor analysis

Zohre Hoseinzade, Ahmad Reza Mokhtari*

Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran

ARTICLE INFO

Article history:
Received 11 February 2017
Received in revised form
18 July 2017
Accepted 24 July 2017
Available online 24 July 2017

Keywords:
Sequential factor analysis
Staged factor analysis
MVT
Geochemistry
Iran

ABSTRACT

Large numbers of variables have been measured to explain different phenomena. Factor analysis has widely been used in order to reduce the dimension of datasets. Additionally, the technique has been employed to highlight underlying factors hidden in a complex system. As geochemical studies benefit from multivariate assays, application of this method is widespread in geochemistry. However, the conventional protocols in implementing factor analysis have some drawbacks in spite of their advantages. In the present study, a geochemical dataset including 804 soil samples collected from a mining area in central Iran in order to search for MVT type Pb-Zn deposits was considered to outline geochemical analysis through various fractal methods. Routine factor analysis, sequential factor analysis, and staged factor analysis were applied to the dataset after opening the data with (additive logratio) alr-transformation to extract mineralization factor in the dataset. A comparison between these methods indicated that sequential factor analysis has more clearly revealed MVT paragenesis elements in surface samples with nearly 50% variation in F1. In addition, staged factor analysis has given acceptable results while it is easy to practice. It could detect mineralization related elements while larger factor loadings are given to these elements resulting in better pronunciation of mineralization.

1. Introduction

For many years, multivariate analysis techniques have been employed in geochemistry to see through a complex dataset the naturally processing scenarios occurring in the earth. This has been more scrutinized in geochemical explorations to separate different phenomena required to delineate what is known as anomalism, as a fingerprint of economic deposits (Kaiser, 1960; Dalton and Upchurch, 1978; Briz-Kishore and Murali, 1992; Reiman et al.,. 2002; Filzmoser et al., 2005, 2009b; Zuo, 2014; Afzal et al., 2016). One of the multivariate analyses to lay out the information in a large data set, is the factor analysis. It was firstly introduced by Karel Pearson (1901) through measuring the intelligence. The technique is employed to determine the most effective factors in a set comprised a large number of variables whose relationships are unknown. The method tries to locate the factors so that variation decreases from the first to the last factor (Davis and Sampson, 1986).

Although in order to implement FA, it is necessary to meet several assumptions including the normal distribution of the test, this has been rarely seen in geochemistry data. The distortion causes outliers to demonstrate skewed distribution that might be due to existence of FA results (Filzmoser and Riemann, 2000; Van Helvoort, 2003; Reimann et al., 2002). A solution to this issue is using the different robust types of factor analyses, though their calculations become complicated. Minimum Covariance Determinant (MCD) is one of the robust statistical methods wildly used, as it consists more appropriate statistical properties and it is more intuitive and understandable (Rousseeuw, 1999; Filzmoser, 1999; Rantitsch, 2007; Filzmoser et al., 2009a, b).

In geochemical explorations, the idea behind using FA is to find key factors and elements related to mineralization. However, this is likely to be undermined by the presence of other variables. They can be some unnecessary variables amongst assayed elements through which the FA results fail to delineate mineralization discernably. There is a high probability that variance of elements irrelevant to mineralization, dominate the paragenetic elements of mineralization thereby masking key elements' role in dataset. Hence, FA on all variables could be inappropriate, as it causes variability to be allocated to all elements. While allocating variation

* Corresponding author.
E-mail address: ar.mokhtari@cc.iut.ac.ir (A.R. Mokhtari).

http://dx.doi.org/10.1016/j.jafrearsci.2017.07.025
1464-343X/

just to more important elements causes their importance to be more noticeable and to yield significant results. Having small variability can make delineation of anomalous areas which are difficult to find via other ways. To overcome this problem, different studies have been conducted based on different methods including sequential factor analysis and staged factor analysis (Cattell, 1966; Cattell and Vogelmann, 1977; Johnson and Wichern, 1998; Filzmoser and Reimann, 2002; Reimann et al., 2002; Van Helvoort et al., 2005; Yousefi and Carranza, 2015; Yousefi, 2017; Yousefi and Carranza, 2016; Yousefi et al., 2012, 2014a, b; Afzal et al., 2016a, b; Zhao et al., 2017). The present research has employed three different types of FA known as routine FA, sequential FA, and Staged FA on soil geochemical data collected from an area in central Iran having different MVT Pb-Zn deposits.

2. Materials and methodology

2.1. Types of factor analysis

The following sections describe three different methodologies employed in this study, known as a) Factor analysis, b) Sequential factor analysis, and c) staged factor analysis. The MCD algorithm was used in all mentioned procedures. Additionally, data were opened prior to analysis by alr-transformation (Rousseeuw, 1999; Filzmoser, 1999).

A) Routine factor analysis

A summary of symbols used in the current study are defined as follow: (see Table 1, Fig. 1)

B) Sequential factor analysis

Sequential Factor analysis was suggested by Van Helvoort et al. (2005) as an improved method dampening the drawbacks of routine factor analysis. It has been created as a series of sequential decision-making processes based on statistical and quantitative criteria so that it optimizes rotation of the factors and the factors supported by less variables are extracted. This technique easily extracts such underlying geochemical properties unrecognized through other methods and makes interpretation of factors easier via recognizing key factors.

In general, this method involves four steps as follows (van Helvoort, 2003; van Helvoort et al., 2005):

1. At the first step, outlier data are removed from the data set. Then the number of factors are determined through a cutoff value. The procedure is iterative and starts by choosing a predetermined loading value known as cutoff value. First, two factors are extracted and if, in rows and columns, at least one loading is found greater than cutoff, new factor analysis will be carried out through three factors. This procedure continues until none of the loadings in the last factor displays the value greater than the threshold value. At this point, iteration ceases but the model is still a complete factor model with all (k) measured variables (van Helvoort, 2003; van Helvoort et al., 2005).
2. The second step is the main step in which key variables are to be recognized. Similar to the previous step, it is first required to select a cutoff value. Therefore, the difference among loadings of the first variable in all factors and all other variables in corresponding factors are calculated. If any of these differences becomes less than the cutoff, an element will be left that has a larger communality and will be known as the key variable. If none of the differences among the first variable's loadings and other variables' loadings is less than the cutoff, it will indicate these two variables are highly correlated and will be omitted from the variables set. This procedure goes on until all variables and a new matrix of factors with lower number of variables known as key variables (m

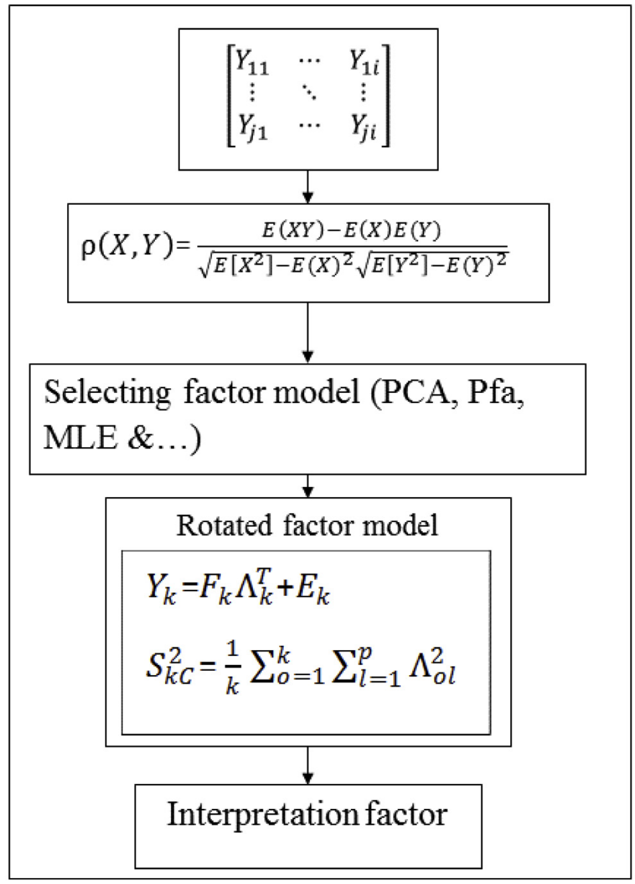

Fig. 1. Flowchart of routine factor analysis (Child, 2006).

Table 1
The summary of symbols used in the present study.

Variables	Description	Variables	Description	Variables	Description
Y	Standardized data matrix	i	I^{th} key variables	S	Stripped
ρ	Correlation coefficient	j	Jth stripped variables	C	Complete
E	Error term	o	O th measured variables	E	expanded
F	Factor scores matrix	l	L th factors	r	Remain
Λ	Loadings matrix	α	constant	a,b	Variable a and b
S^2	variance	s	Key variables	d	D^{th} row in matrix
K	Measured variables	n	Highly correlated variables	Φ_s	Experimental correlation coefficients