Origin of skewed frequency distribution of regional geochemical data from stream sediments and a data processing method

Xinyun Zhao, Libo Hao, Jilong Lu, Yuyan Zhao, Chengyou Ma, Qiaoqiao Wei

PII: S0375-6742(18)30245-0
DOI: doi:10.1016/j.gexplo.2018.07.007
Reference: GEXPLO 6168

To appear in: *Journal of Geochemical Exploration*

Received date: 16 April 2018
Revised date: 6 June 2018
Accepted date: 12 July 2018

Please cite this article as: Xinyun Zhao, Libo Hao, Jilong Lu, Yuyan Zhao, Chengyou Ma, Qiaoqiao Wei , Origin of skewed frequency distribution of regional geochemical data from stream sediments and a data processing method. Gexplo (2018), doi:10.1016/j.gexplo.2018.07.007

Origin of skewed frequency distribution of regional geochemical data from stream sediments and a data processing method

Xinyun Zhao, Libo Hao[a], Jilong Lu, Yuyan Zhao, Chengyou Ma, Qiaoqiao Wei

Department of Geochemistry, Jilin University, Changchun 130021, China

Abstract

Data should approach a normal distribution before application of a variety of methods for separating geochemical anomalies from background. In fact, regional geochemical data from stream sediments generally follow a skewed distribution. It is demonstrated that the skewed frequency distributions are mainly derived from mixed populations, which originate mainly from several geological sources or processes. Two kinds of regional stream sediment samples can be defined, one is a mixture of stream sediments from several lithology backgrounds, and the other is stream sediment from a single lithology background. The combination of these sample data tends to cause skewed frequency distribution. If these mixed data are processed as a whole, in other words, the multiple populations are not recognized and separated in frequency distribution, then serious errors may be made in statistical interpretation. In this study, a method for separating multiple populations is proposed from the point of view of geology, geochemistry and mathematics in an attempt to solve the problem of skewed frequency distribution. Based on the understanding that multiple populations originates mainly from differences in regional lithology background, elements that can reflect lithologies were chosen as classification indicators, then the EM algorithm was employed to separate samples from different populations, and the optimal number of populations can be reasonably determined according to geological map and

[a] Corresponding author. Tel.: +86 431 88502441.

E-mail address: haolb@jlu.edu.cn.