Journal of Geochemical Exploration 190 (2018) 381–389

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Recognition of geochemical anomalies based on geographically weighted regression: A case study across the boundary areas of China and Mongolia

Mi Tian[a,b], Xueqiu Wang[a,b,*], Lanshi Nie[a,b], Chaosheng Zhang[c]

[a] Key Laboratory of Geochemical Exploration, Institute of Geophysical and Geochemical Exploration (IGGE), Chinese Academy of Geological Sciences, Langfang 065000, China
[b] UNESCO International Centre on Global-scale Geochemistry (ICGG), Langfang 065000, China
[c] International Network for Environment and Health, School of Geography and Archaeology, National University of Ireland, Galway, Ireland

ARTICLE INFO

Keywords:
Stream sediment
Geochemical exploration
Geographically weighted regression
Boundary areas of China and Mongolia

ABSTRACT

The identification of stream sediment geochemical anomalies related to mineralization from background is critically needed for mineral exploration using data processing method in the diverse lithological background and complex regolith terrains. In this research, a method based on geographically weighted regression (GWR) was presented for the identification of stream sediment geochemical anomalies. The concentrations of rock-forming oxides, lithophile elements, organic carbon and total carbon were taken as proxies for parent lithology and regolith type to adjust for variations in background of trace element geochemical patterns. Robust principal component analysis (RPCA) was conducted, and then the principal components were taken as spatially independent variables. The metallogenic elements were taken as dependent variables in GWR model, and the geochemical residuals were used to indicate local anomalies. The 1:1,000,000 stream sediment geochemical data across the boundary areas of China and Mongolia were analyzed, and the result of GWR was compared with that of a traditional method. It is found that the efficiency of GWR was highly improved compared with that of the traditional method, indicating that the proposed method can model and eliminate the background differences of elements due to lithological settings and landscapes. Anomalies identified by GWR had stronger spatial association with the known deposits, and thus can be used as guides to new exploration targets.

1. Introduction

Wide-spaced geochemical sampling of fine-grained sediments was successfully used for delineation of mineralization prospects in diverse terrains, particularly in desert windblown sand covered areas (Wang et al., 2007, 2011, 2016). However, the identification of stream sediment geochemical anomalies from background is a hard task for mineral exploration (Deng et al., 2010; Pazand et al., 2011; Rezaei et al., 2015), because the background values of elements in stream sediments are influenced by lithological background and geographic landscapes etc. The paper uses the data from the China and Mongolia Cooperation Geochemical Mapping Project across the boundary areas at a scale of 1:1,000,000 as a part of *Global-scale Geochemical Mapping Program* (Wang et al., 2016) as an example for accurate identifying geochemical anomalies in the diverse regolith-covered terrains. Various proxies for parent lithology and regolith type were applied to adjust for variation in background trace element geochemical patterns and spatial components based on the geographically weighted regression (GWR) model.

A geochemical anomaly can be defined as a concentration of an element that is greater than a threshold value (i.e., the upper limit of background population). Various statistical methods have been used to process geochemical data for the determination of threshold values. Traditionally, a global single uni-element threshold has been used to separate anomalies from background, including the probability plots method, 85% cumulative frequency method, median + 2mad method, mean + *n*SDEV, QQ graph method and fractal method (Hawkes and Webb, 1962; Tukey, 1977; Cheng, 2007; Zhang et al., 2008; Zuo et al., 2009), etc. However, the above-mentioned methods treat the whole area as an "average". If the lithological units and geographic landscapes of the study area were complex, the anomalies in areas with low background value were usually ignored and the concentrations of elements in area with high background value were often identified as anomalies mistakenly, thus undermining the utility of geochemical exploration to define new targets.

Some researchers divided the large area with complex lithological units and geographic landscapes into smaller areas and then identified

* Corresponding author at: Key Laboratory of Geochemical Exploration, Institute of Geophysical and Geochemical Exploration (IGGE), Chinese Academy of Geological Sciences, Langfang 065000, China.
E-mail address: wangxueqiu@igge.cn (X. Wang).

https://doi.org/10.1016/j.gexplo.2018.04.003
Received 6 September 2017; Received in revised form 3 January 2018; Accepted 15 April 2018
Available online 19 April 2018
0375-6742/

the anomalies in each sub-area (Wang et al., 2012; Zuo, 2014). However, the data in the subarea were still taken as an average, therefore leading to the smoothness of data and the suppression of local information inevitably with the loss of useful information. Considering the multivariate nature of most geochemical datasets, some multivariate statistical methods were used to identify anomalies, including cluster analysis, principal component analysis and factor analysis, etc. (Chandrajith et al., 2001; Johnson and Wichern, 2002; Yaylalı-Abanuz, 2013). However, these methods did not take into account the spatial variation of geochemical data.

Regarding stream sediment geochemical data, it is necessary to consider both the characteristics of "stream sediment" and "geochemical data". Firstly, stream sediments are natural composite materials derived from the erosion and migration of rocks and their weathering products upstream of a sampling site. Therefore, uni-element contents of stream sediments are derived from multiple (usually background and rarely anomalous) sources. In most cases, a major proportion of variation in uni-element background contents in stream sediment is caused by the lithological units underlying the areas upstream (Carranza, 2010). Additionally, geographic landscape is another important influencing factor of the background variation of geochemical data in stream sediments. The supergenesis of different landscapes is variable, and thus the dispersion and migration mechanism of elements differ. Therefore, the factors influencing element contents in stream sediments should be considered to determine the geochemical anomalies.

As for geochemical data, two distinct characteristics must be paid attention. Firstly, a stream sediment geochemical data set is a closed number system because it contains compositional variables that are parts of a whole (Filzmoser et al., 2009a; Carranza, 2011), and the sum of concentrations of elements is usually 100%. Because of the closure effect, elevation of concentrations of an element may certainly leads to the reduction of concentrations of another element, thus the false correlations of elements may be artificially introduced (Van den Boogaart and Tolosana-Delgado, 2013; Pawlowsky-Glahn et al., 2015). Moreover, the geochemical data are spatial data, which are spatially heterogeneous. However, most geochemical anomaly identification methods are "global" methods based on the whole dataset, while the local features of the relationships between variables may be hidden. The relationships between variables of geochemical data from any study area may be different owing to the complex lithological background, landscapes and other factors. Therefore, the above-mentioned methods are of limited use in situations where there are extensive overlaps between background and anomalous values, or where weak anomalous values are hidden within the strong variance of background (Lalor and Zhang, 2001).

The observed value, that is not high in the global scope, but is different from its neighborhoods, may contain important information (Lalor and Zhang, 2001). Compared with the traditional global outliers, the outliers can be identified by comparing with its adjacent samples, which are called "spatial outlier" or "local outlier" (Zhang et al., 2009). Given the above, the geographically weighted regression (GWR) was proposed to process geochemical data and extract local geochemical anomalies related to mineralization avoiding the influence of background variation caused by complex lithological background and geographic landscapes. GWR was proposed by Fotheringham in 1996. It focuses on the quantitative relations between two or more spatial variables by regression, carrying out the local rather than global parameter estimation. The GWR considers the non-stationarity relations of geochemical data to estimate the parameters of all points in the dataset, and shows corresponding regression parameters to each point (Fotheringham et al., 2002). GWR can consider not only multiple influencing factors but also the spatially local variation, so it is applicable to geochemical data processing and anomaly identification in stream sediments from complex settings.

A reasonable way to recognize geochemical anomalies in complex lithological units or in regolith-covered areas was to determine the underlying relationships among geochemical anomalies and plausible causative geological processes (Cohen et al., 2012; Reimann and Garrett, 2005; Sadeghi et al., 2015). Since there are often clear relationships between rock types and concentrations of major oxides and some trace elements in rocks and the inheritance of stream sediments from rocks, using concentrations of major oxides and some trace elements in stream sediments to represent lithological features is a common logical tool to recognize geochemical anomalies in complex lithological units or in regolith-covered areas (Hao et al., 2014). In addition, some geochemical indicators, such as total carbon (TC) and organic carbon (OrgC), are usually associated with supergene geochemical features such as climate, landforms and regolith-type, so they can be used to indicate the difference of landscapes.

In this research, a method based on GWR was established to identify geochemical anomalies related to mineralization for different lithological background and geographic landscapes. Major elements, some trace elements, total carbon and organic carbon, which are the appropriate proxies for parent lithology and regolith type to adjust for variation in trace element background were used as spatial components to the modeling. Robust principal component analysis (RPCA) was taken to reduce the dimensions and number of independent variables to avoid the influence of outliers. The principal components and measured metallogenic elements were taken as independent and dependent variables respectively in GWR. The ***geochemical residuals,*** subtracting the predicted values from measured values, can be used to indicate abnormal areas, the greater the residual, the stronger the anomaly (Zhang et al., 2009). In order to show the efficiency of the proposed method, the results from this proposed methodology were compared with that from using traditional method.

2. The study area and sampling

2.1. The location and geological setting of the study area

The study area is located across the boundary areas of China and Mongolia, covering the Altai Mountains, south of the Mongolia Plateau and west of the Great Khingan, extending about 50–100 km from each side of the border (Figs. 1, 2). The length of borderline between China and Mongolia is up to 4673 km. Geographical coordinates of the study area are between East longitude 86°–120° and North latitude 41°–50°, with a total area of about 1.3 million km^2.

The area is located in central Asia-Mongolia Giant Orogenic Belt among Kazakhstan, Tarim, North China Craton and Siberia Platform, belonging to the middle east of Junggar-South Mongolia-the Great Khingan Paleozoic Orogenic Belt, mainly composed of a series of south convex arc tectonic-magmatic belt. The western area belongs to the Altai Orogenic Belt, Junggar Orogenic Belt, north Tianshan Orogenic Belt and Beishan Orogenic Belt, and the mid-east belongs to the middle Mongolian-Erguna Orogenic Belt, south Mongolia-Xingan Orogenic Belt and Inner Mongolia Orogenic Belt.

2.2. Sampling and analysis

The samples were collected by China and Mongolia cooperation team and were analyzed in the same laboratory. Therefore, data in this paper from the China and Mongolia Cooperation Geochemical Mapping Project are consistent.

2.2.1. Sampling density and distribution

The sampling density was one sample per 100 km^2. GPS units were used to record the coordinates. A total of 10,189 sampling sites were evenly distributed in the region. Composite samples were collected in a range of 50 m (generally 3–4 sites). Duplicate samples were distributed at the same site but different locations (at least 2 m apart), and their quantity was 5% of the total samples.

The landscapes of the area can be divided into four main types from