



Bulletin of the Mineral Research and Exploration

<http://bulletin.mta.gov.tr>



Use of Moran's I and robust statistics to separate geochemical anomalies in Jiurui area (Southeast China)

Tien Thanh NGUYEN^{a*}

^aHa Noi University of Natural Resources and Environment, Ha Noi, Viet Nam. orcid.org/0000-0002-7600-5858.

Research Article

Keywords:

Anomaly distribution,
Terrestrial Moran
Statistic, Robust
Statistics, Jiurui Copper
Field.

ABSTRACT

Separation of geochemical anomalies from background plays an important role in the study of exploration geochemistry. The limitations of commonly used methods are not taken into account spatial correlation, variability and the unsatisfactory of the statistical assumption of the normality of geochemical data. For solving these limitations, an indirect method for the separation of geochemical anomalies is proposed based on anomaly separation of local Moran's I_i values using robust statistics in this study. The experiment was carried out using 1481 samples collected from Jiurui copper prospect (southeast China). The steps for the anomaly separation are (i) spatial association and variability were first analyzed by means of Moran scatterplots at six spatial scales (2, 4, 6, 8, 10 and 12 km) using both raw data and Box-Cox transformed data; (ii) local Moran's I_i was used to measure spatial autocorrelation at these six local scales; (iii) anomalous separation was finally performed using the $MEDIAN \pm 1.5 * IQR$ (IQR: interquartile range) rule on local Moran's I_i values. The results show that geochemical anomalies are mostly concentrated around known ore-deposits, according to the objective reality and a strong correlation with known ore-deposits in Jiurui copper prospect.

Received Date: 15.03.2017

Accepted Date: 04.11.2017

1. Introduction

Geochemical anomaly separation is a fundamental issue in the operation of geochemical exploration, which is defined as a region where the concentration of an element is greater than a certain threshold value by statistical parameters, such as mean, median, mode, and standard deviation (Bolviken, 1992; Cheng, 1996; Davis, 2002; Li, 2003; Liu et al., 2013). In the 1960s, several procedures were recommended to select a threshold from background such as top 2½% selecting, the $MEAN \pm 2SDEV$ (SDEV: standard deviation) rule (Hawkes and Webb, 1962), frequency analysis including histogram, probability, Q-Q plots, and box-plot (Sinclair, 1974, 1976, 1991; Stanley and Sinclair, 1989; Garrett, 1989). The data distribution is important because these classical methods assume normality or lognormality. They do not, however, consider the shape, extent and magnitude of anomalous areas and

disregard its spatial variability (Afzal et al., 2010; 2013; Hashemi et al., 2010; Momeni et al., 2016; Nazarpour et al., 2015, 2016; Rajabzadeh et al., 2015) and that must be known when applying these classical statistical methods. But this basic requirement has been still widely neglected in many studies (Philip and Watson, 1987; Rock, 1988). These statistical methods usually assumed that the concentration of ore elements in the crust follow a normal or log-normal distribution (Ahrens, 1953, 1954a, b, 1957) due to the existence of extreme values and outliers (Reiman et al., 2005), even for Box-Cox or log-transformed data (Liu et al., 2017; McGrath and Loveland, 1992; Nguyen et al., 2014). The widely-used, classical statistical methods are likely to fail for strongly skewed data (Reimann and Filzmoser, 2000) for not being warranted for assumptions of normality, independence, and identical distribution. Therefore, many methods for outlier detection have been proposed based on robust

*Corresponding author: Tien Thanh NGUYEN, tdgis_nthanh@163.com
<http://dx.doi.org/10.19111/bulletinofmre.351376>

statistics (Barnett and Lewis, 1994; Dutter et al., 2003; Huber, 1981; Hampel et al., 1986; Rousseeuw and Leroy, 1987). Recently, Reimann (2005) used robust statistics (Huber, 1981) and exploratory data analysis (EDA) techniques (Tukey, 1977) including the boxplot and the $\text{MEDIAN} \pm 2 * \text{MAD}$ (MAD: median absolute deviation) rule for the estimation of threshold values and the range of background data, in which the $\text{MEDIAN} \pm 2 * \text{MAD}$ does not adequately build on statistical assumptions. Another approach for the identification of extreme values is the $\text{MEDIAN} \pm 1.5 * \text{IQR}$ (IQR: interquartile range; Tukey, 1977). The estimators of MEDIAN, MAD and IQR are widely available statistical software packages (Reiman, 2005).

Geochemistry properties and processes are defined in different scales. In many cases, spatial variation is not random, but tends to follow a pattern in which variability decreases as distance decreases between points in space as Tobler (1979). Therefore, it is important to take into account spatial correlation and variability of geochemical data when separating geochemical anomaly. These non-spatial statistical methods (classical and robust statistics) ignore the spatial information or autocorrelation structure of the geochemical data. In geography, spatial information or autocorrelation structure can be identified by association statistics such as Getis's G index (Getis and Ord, 1992), Geary's C (Geary, 1954), spatial scan statistics (Ishioka et al., 2007), Tango's C index (Tango, 1995; Zhang and Lin, 2006), and local Moran's I. These statistics have been widely used for many fields of study such as biological sciences (Legendre and Gauthier, 2014) environmental planning (Brody et al., 2006), environmental sciences (Zhang and McGrath, 2004; Ruiz-Rivera and Paul, 2016), criminal study (Wang et al., 2016), and mortality rates (Martins-Melo, 2016, 2017), especially in geochemical exploration (Nguyen et al., 2014, 2016). The local Moran's I_i has been used to identify Cu anomalies in Jiurui copper mining area, but spatial scales were chosen randomly without scientific basis (Nguyen et al., 2014). Recently, Nguyen et al. (2016) proposed a method for the identification of spatial clusters and outlier of an association of Cu, Au, Mo, Ag, Pb, Zn, As, and Sb elements based on a combination of robust Mahalanobis distance and local Moran's I_i . In this study, we focus on the performance of a robust statistic (the $\text{MEDIAN} \pm 1.5 * \text{IQR}$) and Moran's I to separate Cu anomalies from 1341 stream sediment samples.

2. Geological Setting and Dataset

Jiurui district is an important copper mining area in China's Jiangxi province and rich in copper reserves. It is located in the middle and lower Yangtze River metallogenic belt with copper-iron mineralization and North West Gate - Fengshan mountain cave. Around 154,968 million tons of copper, gold and silver reserves were discovered in Jiurui area. A total of 16 copper mineralized zones, three molybdenite orebodies and 120 small orebodies were found in the mining area with an average copper grade of 0.69%. According to the Jiangxi provincial government, the region produced 651,000 million tons of refined copper in the first half of 2014, up 11% from the previous year. Iron and copper deposits are one of the main ore deposits in the downstream in the area. The formation of rocks and ores control the distribution of large ore deposit. The main tectonic belt and its vicinity are located in shallow ore prospecting target area, Wushan Chengmenshan mining area, the edge and deep prospecting potential. Iron copper deposits are divided into two metallogenic series: (i) submarine exhalative related activities found that metallogenic series are any hydrothermal deposits from the injection to the bottom of the sea environment, (ii) intermediate acid hypabyssal intrusive activities-related ore deposits; refers to the formation of intrusive rocks of carboniferous sand Triassic strata in contact zone and rock deposits. Skarn, iron-copper, porphyry copper, key, vein copper, gold are the main types of deposits in the study area.

A total of 1482 multi-element sediment geochemical samples (1 composite sample per 4 km²) representing 5364 km² were collected from regional stream sediment surveys (Figure 3). Twenty-three indexes were found in a composite sample, including Ag, As, Au, Be, Cd, Cu, Hg, Li, Mn, Mo, Nb, Pb, Sb, Sn, Th, V, W, Y, Zn, Al₂O₃, CaO, K₂O, Na₂O in which silver, gold, copper are three ore-forming elements. Three geochemical associations of elements causing anomalous area in the study area include Cu, Au, Mo, Ag, Pb, Zn, As, Sb; Mo, Sb, Zn, Ag, V and Sn, W, Sb, Mo, Zn, Ag, As. Two metallogenic with a total of 13 ore deposits were found at the time of data collection.

Raw data was first checked including two aspects: data checking and processing of missing semi-quantitative and unreliable data. The initial phase for raw data checking consists of simple and obvious procedures carried out in the first stage of acceptance

the results from the laboratory. The integrity and validity of dataset were checked. During the initial data checking phase missing, non-numeric, coded an unreliable results were dealt with. Obvious misfit such as results that are way out of expected ranges and control samples that clearly do not conform to expected values were identified by looking at the data. Unreliable data and coordinates were also corrected.

A total of 51 duplicate samples in 1482 collected sediment samples were removed.

3. Methodology

The process of separation of Cu anomalies using Moran's and robust statistics can be expressed in figure 1.

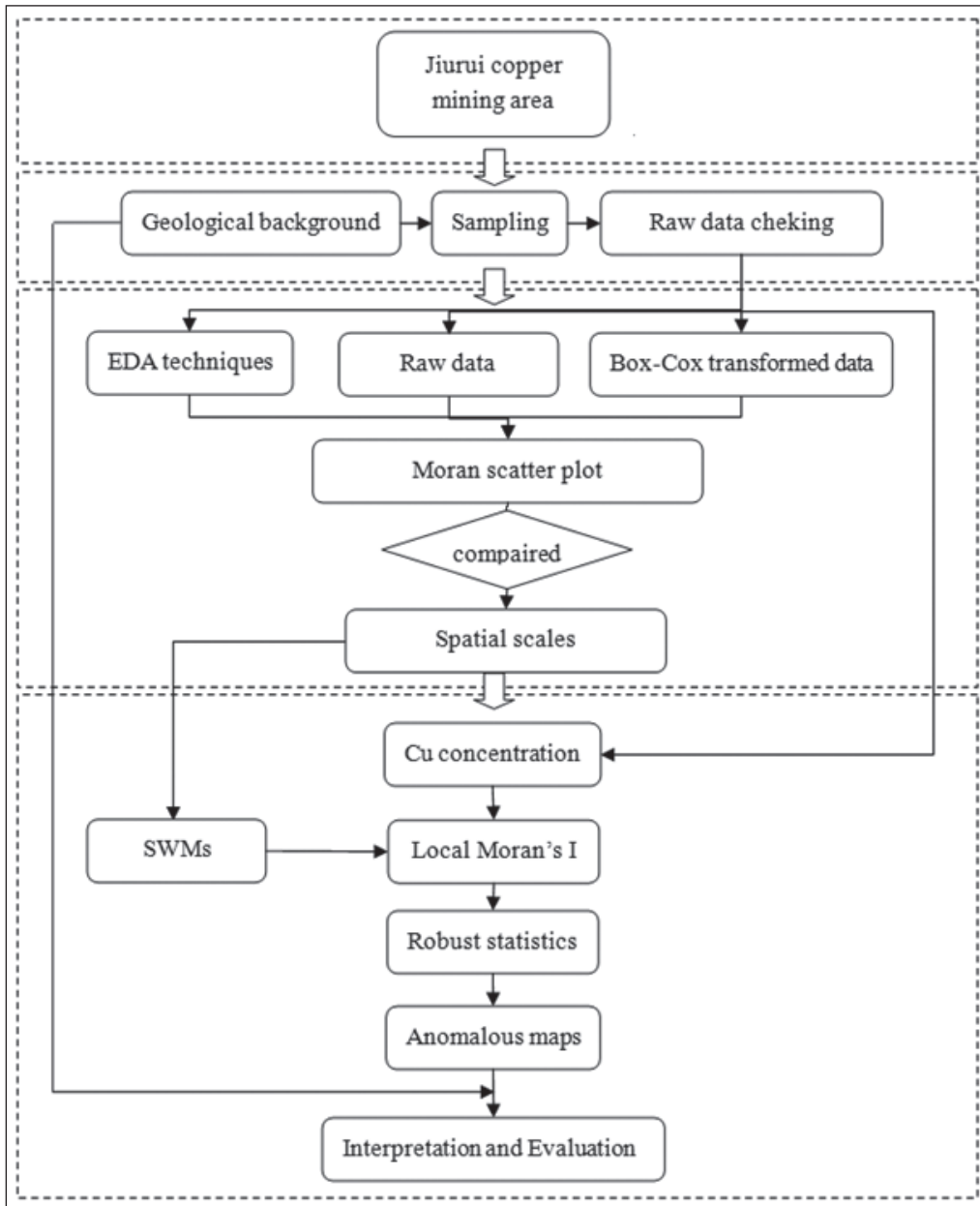


Figure 1- Flow chat for the separation of geochemical anomalies using Moran's I and robust statistics.

3.1. Moran Scatterplot

Global Moran's I statistic is given by the expression (Cliff and Ord, 1981):

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

where: x_i and x_j are the values of observed variable at sites i and j ; d is the distance class considered in the calculation; S_0 is the sum of the weights W_{ij} , given by: $S_0 = \sum_{i=1}^N \sum_{j=1}^N W_{ij}$; \bar{x} is the sample mean of the observed variable, given by: $\bar{x} = \sum_{i=1}^N \frac{x_i}{N}$; N stands for the number of observations with

$$W_{ij} = \begin{cases} 1 & \text{if } d_{ij} < d \\ 0 & \text{otherwise} \end{cases}$$

W_{ij} is spatial weight matrix for a given distance d and d_{ij} is the distance between the observations i and j . The spatial weight matrix computed from a given distance is row-standardized so that the sums of the weights for a trap are equal to one (Cocu et al., 2005):

$$W_{ij}^s = W_{ij} * \left(\sum_j W_{ij} \right)^{-1} \text{ such that } \sum_j W_{ij}^s = 1$$

where each row sums to 1 leading to the normalizing factor S_0 equals to N . In this case, this is the preferred way to implement this test (Anselin, 1992). This allows a comparison of the spatial autocorrelation statistics between the various neighboring distances because, for each of them, identical importance is attached to all the neighbors independently of their number. The numerator of Equ. (1) corresponds to the covariance between contiguous samples. This covariance is zero in the absence of spatial autocorrelation, positive in the presence of positive spatial autocorrelation, and negative in the presence of negative spatial autocorrelation. The covariance is standardized by the denominator, which is a measure of the variance of the samples. In the calculation of the statistic, the mean of the samples is a reference (Cocu, 2005): (i) when $I = 0$: the covariance between contiguous samples is zero, the neighborhood does not play any part; there is no spatial autocorrelation. The observed spatial pattern of values is equally as likely as any other spatial pattern; (ii) when $I > 0$: the places are more similar. If they are contiguous, the spatial autocorrelation is positive. Similar values, either high or low values are more spatially clustered than could be caused purely by chance; and (iii) when $I < 0$: the places are more alike if they are far apart; the spatial autocorrelation is negative, the spatial pattern is regular.

Values for global Moran's I statistics need not be constrained to the interval $[-1, 1]$. Usually, $|I| < 1$, unless regions with extreme values of $x_i - \bar{x}$ are heavily weighted. The theoretical upper bound is (Cliff and Ord, 1981; Haining, 1990; Bailey and Gatrell, 1995):

$$|I| \leq \frac{N}{\sum_{i \neq j} \sum_{j=1}^N W_{ij}} \left\{ \frac{\sum_{i \neq j} \left[\sum_{j=1}^N W_{ij} (x_i - \bar{x}) \right]^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right\}^{1/2} \quad (2)$$

The significance of any observed value of Moran I can be judged by comparing it to its expected value, but it must also account for the expected variability in the I statistic under the appropriate null hypothesis. This is where the distributional assumptions about the data become very important. With the randomization assumptions, the data values are reassigned among the N fixed locations, providing a randomization distribution against which we can judge our observed value. If our observed value of I lies in the tails of this distribution, we reject the assumption of independence among the observations and conclude that there is significant spatial autocorrelation in the data (Waller and Gotway, 2004). If relying on the normality assumptions, the z-score $z = [I - E(I)] / \sqrt{\text{Var}(I)}$ is compared to a standard normal distribution where $E(I)$ and $\text{Var}(I)$ are given in Equ. (3) and (4) (Cliff and Ord, 1981):

$$E(I) = \frac{-1}{N - 1} \quad (3)$$

$$\text{Var}(I) = E(I^2) - [E(I)]^2 \quad (4)$$

3.2. The Local Moran Statistic

Anselin (1992, 1995) developed measures of spatial autocorrelation at a local scale. A local Moran statistic for statistic for an observation i may be defined as (Anselin, 1995):

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_{j \in J_i} W_{ij} (x_j - \bar{x}) \quad (5)$$

where the summation over j is such that only neighboring values $j \in J_i$ are included. For ease of interpretation, the weights may be in row-standardized form, though this is not necessary, and by convention, $W_{ij} = 0$;

$$m_2 = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})^2,$$

m_2 as the second moment. For a row-standardized spatial weight matrix $S_0 = N$, and for standardized variables (i.e., with the mean subtracted and divided by the standard deviation), $m_{2=1}$. When: (i) $I_i = 0$: there

is no spatial autocorrelation; (ii) $I_i > 0$: the spatial autocorrelation is positive. It conveys the presence of an association of values similar to the place i where the statistic is measured (hot spots); and (iii) $I_i < 0$: the spatial autocorrelation is negative, corresponding to an association of values that are opposed to the place i where the statistic is measured (cold spots).

The moments for I_i under the null hypothesis of no spatial association can be derived using the principles outlined in Cliff and Ord (1981). An advantage of this technique is that the significance of the null hypothesis can be tested. It is possible to quantify the probability that a spatial pattern as extreme as that observed could have appeared by chance. For a randomization hypothesis, the expected value $E(I_i)$ turns out to be (Anselin, 1995):

$$E(I_i) = \frac{-W_i}{N-1}, \tag{6}$$

with W_i as the sum of the row elements, $\sum_{j=1}^N W_{ij}$, and the variance $\text{Var}(I_i)$ is found as:

$$\text{Var}(I_i) = \frac{W_{i(z)}(N-b_2)}{N-1} + \frac{2W_{i(kh)}(2b_2-N)}{(N-1)(N-2)} - \frac{W_i^2}{(N-1)^2}, \tag{7}$$

where:

$$b_2 = \frac{m_4}{m_2^2}; m_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4; W_{i(z)} = \sum_{j=1}^N W_{ij}^2; \text{ and } 2W_{i(kh)} = \sum_{k=1}^N \sum_{h=1}^N W_{ik}W_{ih}$$

A test for significant local Moran I_i may be based on these moments, although the exact distribution of local Moran I_i is still unknown (Anselin, 1995). Moran's I_i can also be standardized so that its significance level can be tested based on an assumption of a normal distribution (Anselin, 1995; Levine, 2004). However, since the probability distribution of Moran's I_i may not necessarily be normal, especially when the raw data are heavily skewed, a method called "conditional permutation" (Anselin, 1995) is preferred as it makes no assumption about the data. Under a conditional permutation, when the value on a location is being assessed, its value is fixed and all the other values are randomly shuffled on all the other locations. Each time when the other values are shuffled, the local Moran's I_i statistic is calculated to form a reference distribution. The significance level can be estimated by comparing the observed index with these simulated values, and it is called "pseudo significance" (Anselin, 2005). The pseudo significance is computed as $(M+1)/(R+1)$ where R is the number of permutations and M is the number of instances where a statistic computed from the permutations is equal to or greater than the

observed value (for positive local Moran's I_i) or less or equal to the observed value (for negative local Moran's I_i) (Spatial Analysis Laboratory, 2007).

3.3. Robust Statistics

The local Moran's I_i statistics well suited to identifying the existence of pockets of spatial association, to assess assumptions of stationarity and to identify distances beyond which no discernible association remains (Getis and Ord, 1996). If the underlying process is stable throughout the data, then one would expect the local indications to show little variation around their average. In other words, local values that are very different from the mean (or median) would indicate locations that contribute more than their expected share to the global statistic. These may be outliers or high leverage points (Anselin, 1995), local Moran's I_i values are not normally distributed. Therefore, robust statistics were used to identify outliers in a set of local Moran's I_i values by means of the $\text{MEDIAN} \pm 1.5 \cdot \text{IQR}$ rule expressed in Equ. (8):

$$\text{IQR} = Q_3 - Q_1 \tag{8}$$

where: quartile 1 (Q_1) and 3 (Q_3) are the 25th and 75th percentile of ordered local Moran's I_i values or median of lower half and upper half of local Moran's I_i values.

3.4. Data Processing

The global, local Moran's I_i statistic and Moran's scatter plots were created in spatial statistics software - *GeoDA* (version 095i) in which both the global and local Moran's I_i statistics were tested using 999 permutations, and the significance level (p-value) was set to 0.05 (5%). The descriptive statistical parameters (mean, median, IQR), EDA techniques (histograms, density trace, 1-D scatter plots, boxplots, QQ plots and empirical cumulative distribution function plots) and Box-Cox transformation of geochemical data were performed by using the *StatDA* and *MASS* packages of Statistical Modeling and Computing - R Language (version i386 2.15.0). All of maps were visualized using *ArcGIS 9.3* software.

4. Results and Discussions

4.1. Cu Concentration in the Study Area

The spatial distribution of Cu in the study area shows the influence of mineral deposits. There is a high

value pattern of Cu in the north and the east. There are also some relatively high values scattered in the north-west where no deposits have so far been detected showing the complexity or spatial heterogeneity of Cu concentration (Figure 2). It is likely the high values (or extreme values) of Cu concentration were caused by mineralization.

The minimum, median, mean, and maximum values of Cu concentration are 9.2, 27.1, 37.3 and 1926 µg/g, respectively. It can be realized that the mean value of Cu concentration is 10.2 µg/g larger than the median value of 27.1 µg/g, therefore the distribution

of Cu concentration has a positive skew. The skewness for Cu is 15.8 (much larger than zero), indicating that the size of the right-handed tail is larger than the left-handed tail. Cu has an asymmetrical distribution with a long tail to the right and a positive skew. The kurtosis of 288.5 (much larger than zero) indicates the distribution has heavier tails and a sharper peak than the normal distribution. Essentially it can be concluded that the distribution of Cu concentration is far from symmetrical. Many high values occur mostly in the north and in the east, where Cu and multi-metal deposits were found in the east-northern part (Figure 2). They may be influenced by extraneous and exotic

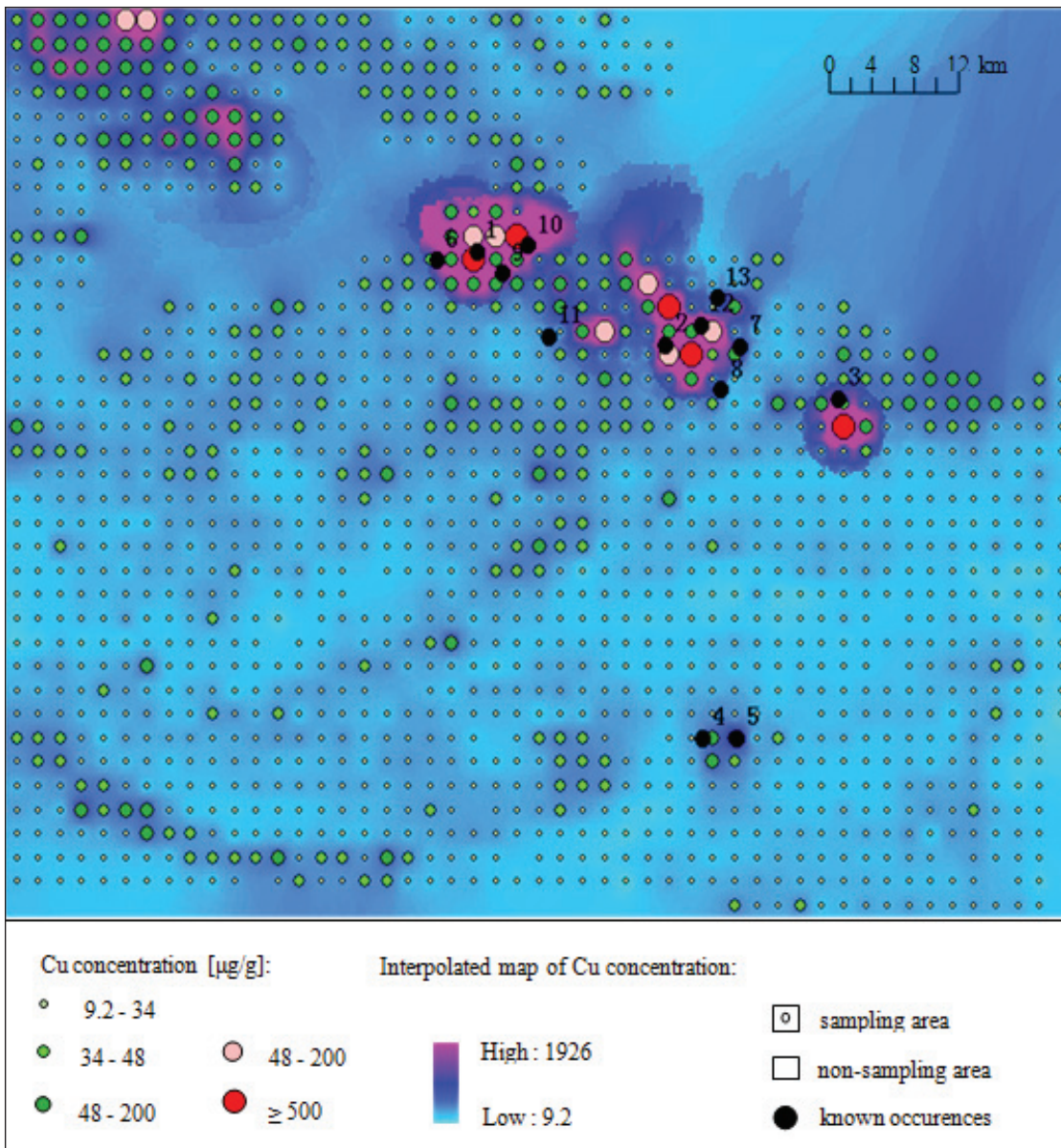


Figure 2- Spatial distribution maps of Cu concentration and of 13 known occurrences.

processes such as those related to rare rock types and mineral deposit formation processes, thus, care must be taken in interpreting the high values

Figure 3 shows the distribution of the variable Cu. It can be seen from the histogram, the distribution of raw data is skewed strongly to the right caused by many very high values (Figure 3 left). The distances of these high values are far from the main body of the data. The distribution is obviously not normal but extremely skewed to the right. Due to typical asymmetrical, a sigmoidal S-shape is not present in the empirical cumulative distribution function plot (Figure 3 right).

A curved pattern is found in the Q-Q plot (quantile-by-quantile plot) and it is in the shape of a bow (Figure 4 left). This indicates skewing because data is consistently below the line representing the quantiles of the standard normal distribution. Most of the points lie quite close to the line, and the ends are below it, the skewing is thus to the right. Data points do not follow a straight line. The plot shows the upper tail of the distribution strongly deviates from normality. Cu concentrations are strongly skewed right and the rightmost data points are outliers.

The distribution of Cu concentration is right skewed, in order to reduce the influences of extreme

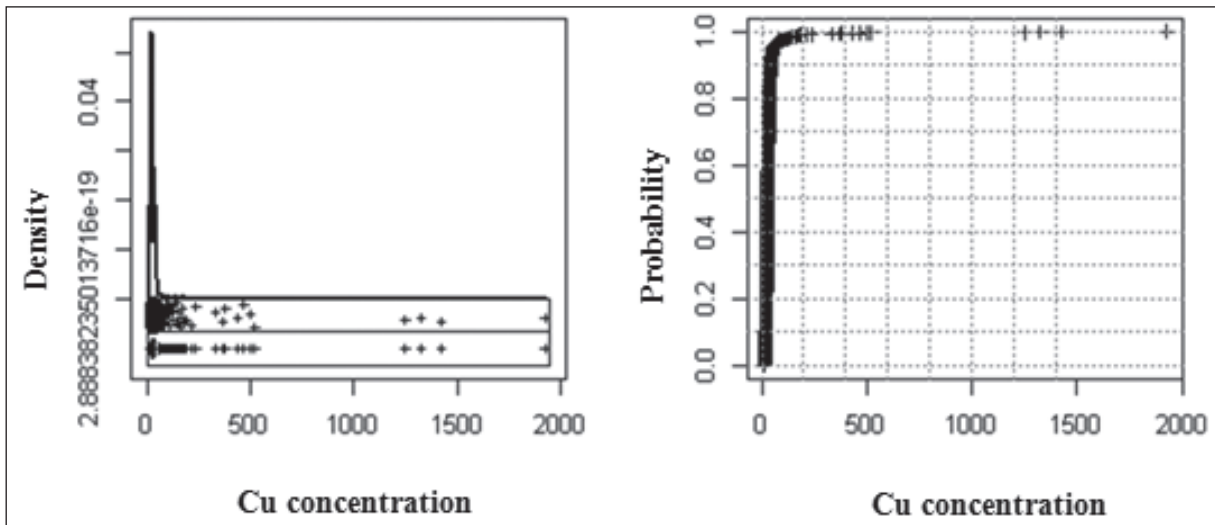


Figure 3- Histogram, density trace, 1-D scatter plot, boxplot (left) and empirical cumulative distribution function plot (right) for raw data..

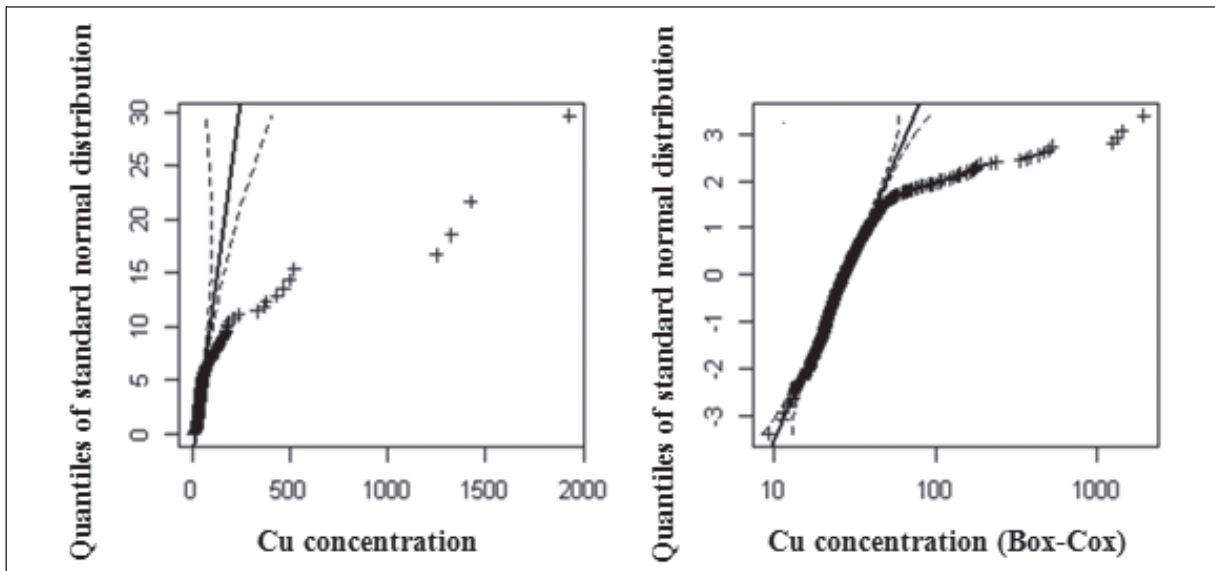


Figure 4- QQ-plots for raw data (left) and Box-Cox transformed data (right).

values, outliers and to obtain vital information of values about the shape of the distribution. Data is rescaled by using Box-Cox transformation to make the data normal. For Box-Cox transformed data, the minimum, median, mean and maximum values are 0.91, 0.99, 0.99 and 1.03 $\mu\text{g/g}$, respectively. The mean is equivalent to the median, which shows that transformed data is evenly divided around the mean. The kurtosis of -0.156 is close to zero (greater than -1.0 and less than 1.0), indicating that the distribution of the transformed data is close to symmetrical. The kurtosis of 2.034 is much smaller for the Box-Cox data, indicating the peak and tails of the distribution almost does not differ from the normal distribution.

The Box-Cox transformation results in an almost symmetrical distribution of the strongly right skewed data. The histogram and density trace still show a slight right skew (Figure 5 left). The empirical cumulative distribution function plot begins to display an S-shape (Figure 5 right), however the right skew is still clearly reflected. The QQ-plot from Box-Cox transformed data show the upper tail of the distribution less deviates from normality (Figure 4 right). More data points follow the straight line. Less extreme values dominate the Q-Q plot.

4.2. Spatial Scales

It has been proven in Hoang et al. (2017) that the variance of the Cu variable levels off when at distance of 12.4km, beyond this distance, the sampling units are not spatially correlated. Therefore, in order to

investigate the effects of different spatial scales on the results, six spatial scales were considered in this study where $d = 2, 4, 6, 8, 10$ and 12 km in this study. In this case, the weights for neighboring locations were assigned as 1 if the distances were within the band, otherwise the weights were 0. It can be seen from Moran scatter plots (Figure 6) that the values of all of global Moran's I statistic at these spatial scales (below 12 km) are positive, indicating neighboring samples tend to have similar values. The spatial pattern of Cu concentration is clustered. The values of the global Moran's I statistic decreased when bands increased.

For the raw data, the Moran scatter plots are dominated by high-high clusters (above the mean) in the upper right quadrants (Figure 6-1a, b, c, d, e, and f) representing positive spatial association. A few data points occurred in the upper left quadrants representing low values surrounded by high values and at the lower right quadrants representing high values surrounded by low values. Almost no data points were found in the lower right quadrant, which indicates no spatial association between low values. The global Moran's I value of 0.1591 indicates the strongest spatial association at spatial scale 2km. Moran's I values reduce to 0.1186 and 0.0993 at spatial scales of 4 and 6km respectively, indicating weaker spatial association. The values of Moran's I are of 0.0712, 0.0526 and 0.0437 and close to zero when spatial scales reach 8, 10 and 12 km. Number of neighboring samples N increased at these bands, $E(I)$ approaches zero, resulting in no spatial correlation between neighboring samples.

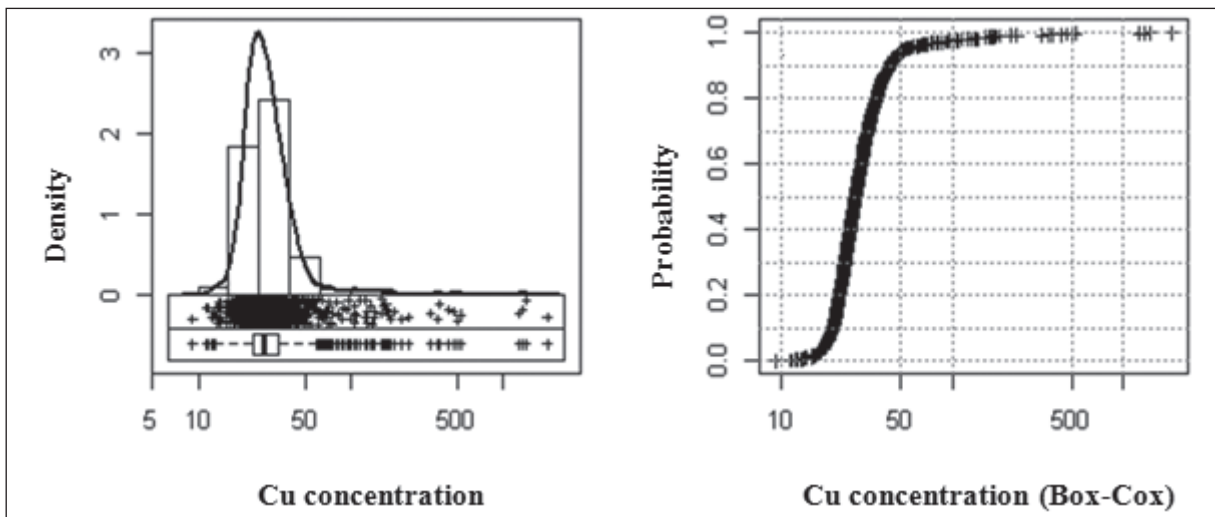


Figure 5- Histogram, density trace, 1-D scatter plot, boxplot (left) and empirical cumulative distribution function plot (right) for Box-Cox transformed data.

For Box-Cox transformed data, global Moran's I values are much higher at all six spatial scales when compared with the results for raw data, therefore, we can see spatial patterns more clearly (Figure 6-2a, b, c, d, e, and f). The values of Moran's I for the first three bands are 0.5642, 0.4755 and 0.3947, indicating a stronger positive spatial association in Box-Cox transformed data. Data points are dominated in the upper right and lower left quadrants in these plots. More points were found in the upper right quadrants where high values are surrounded by high values.

Moran's values reduce to 0.3460, 0.3075 and 0.2874 and gradually approach to zero at bands of 8, 10 and 12km. However, spatial pattern of high-high clusters are still dominated by these plots.

4.3. Anomaly Separation and Evaluation

The data distribution of local Moran's I_i is strongly skewed-right due to high values and outliers as mentioned above. The $MEDIAN \pm 1.5 * IQR$ formula was applied to separate anomalies from Box-Cox transformed local Moran's I_i values. The results of

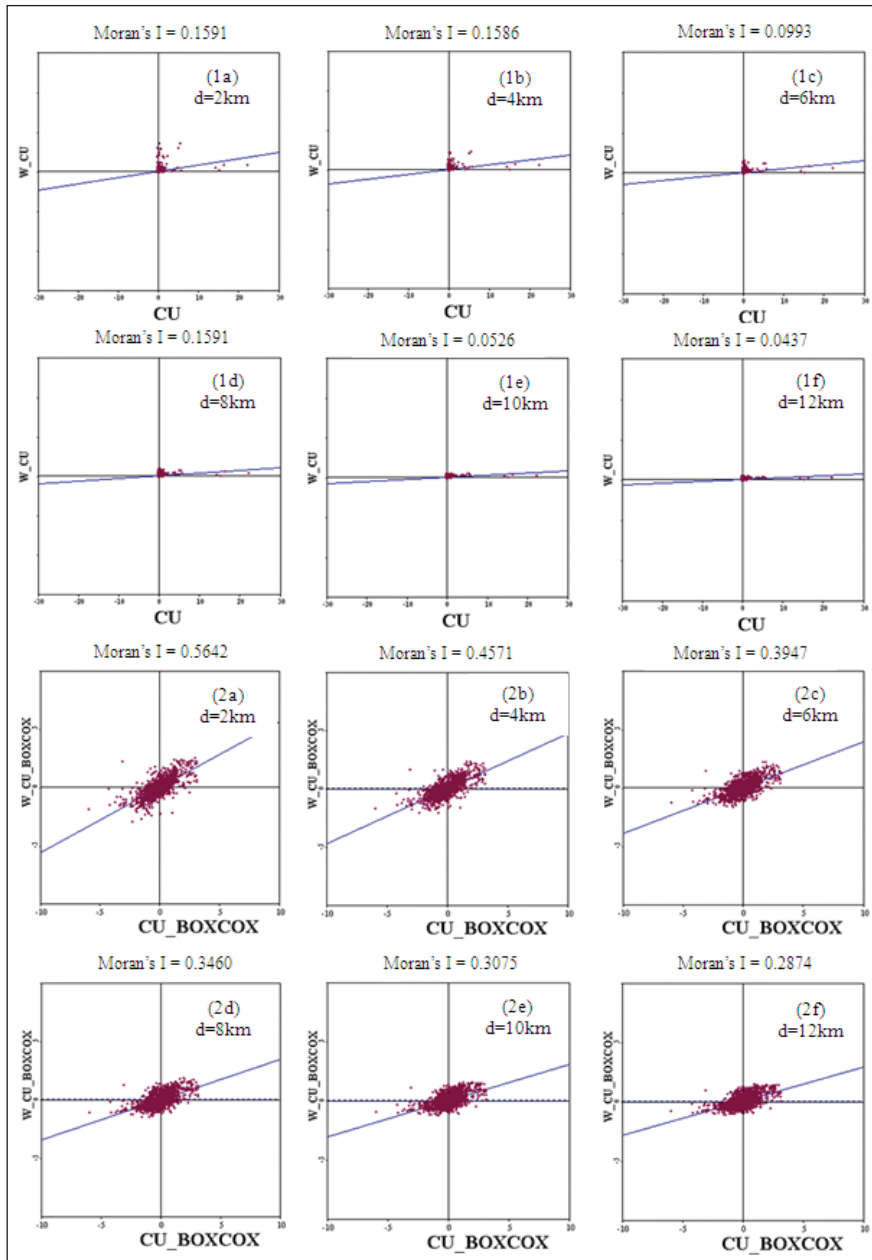


Figure 6- Moran scatterplots for Cu: raw data (1a,b,c,d,e,f), Box-Cox transformed data (2a,b,c,d,e,f).

Table 1- Summary table of statistical parameters of local Moran's I_i

Spatial scales (km)	1 st Qu.	Median	2 nd Qu.	IQR	(Med+1.5*IQR)	Anomalous area (km ²)	Background area (km ²)
d=2	0.0039	0.0146	0.0278	0.0240	0.05061	372	5.552
d=4	0.0031	0.0141	0.0261	0.0240	0.04858	392	5.332
d=6	0.0033	0.0135	0.0248	0.0216	0.04587	388	5.336
d=8	0.0026	0.0134	0.0243	0.0216	0.04586	364	5.360
d=10	0.0015	0.0133	0.0234	0.0219	0.04612	352	5.372
d=12	0.0009	0.0127	0.0234	0.0225	0.04655	332	5.392

(1st Qu. = 1st quartile; Median = median value; 3rd Qu. = 3rd quartile; IQR = interquartile range)

anomaly separation are summarized in table 1 and visualized in figure 7.

It can be seen from figure 7, generally, anomalous areas are found in the north and in the east of the study area at all of six spatial scales, where a metallogenic belt was found including Cu and multi-metal deposits marked by 6, 1, 9, 10, 2, 13, 7 and 8 in the east-northern part. Total anomalous area is 372 km² at the first spatial scale and increased slightly to 392 km² and 388 km² at spatial scales of 4 to 6 km, while background areas are 5552, 5332 and 5336 km² respectively (Figure 7-a, b, c). However, the anomalous areas are significantly

reduced to 364, 352 and 332 km² as the spatial scales increase to 8, 10 and 12 km, while background areas are increased by 5,360, 5,372 and 5.392 km² (Figure 7-d, e, f). Anomalies are mostly concentrated around known ore-deposits marked by 6, 1, 9, 10, 2, 13, 7 and 8 in the east-northern part (except ore-deposits No. 4 and 5) due to the influences of extraneous and exotic process such as mineral mineralization. A part of anomalies occurred in the northwest of the study area at all six spatial scales, where no known ore deposits were found in this area. No anomalies were found in the southern part where ore deposit 4 and 5 were located (Figure 6) due to weak anomalies in this area.

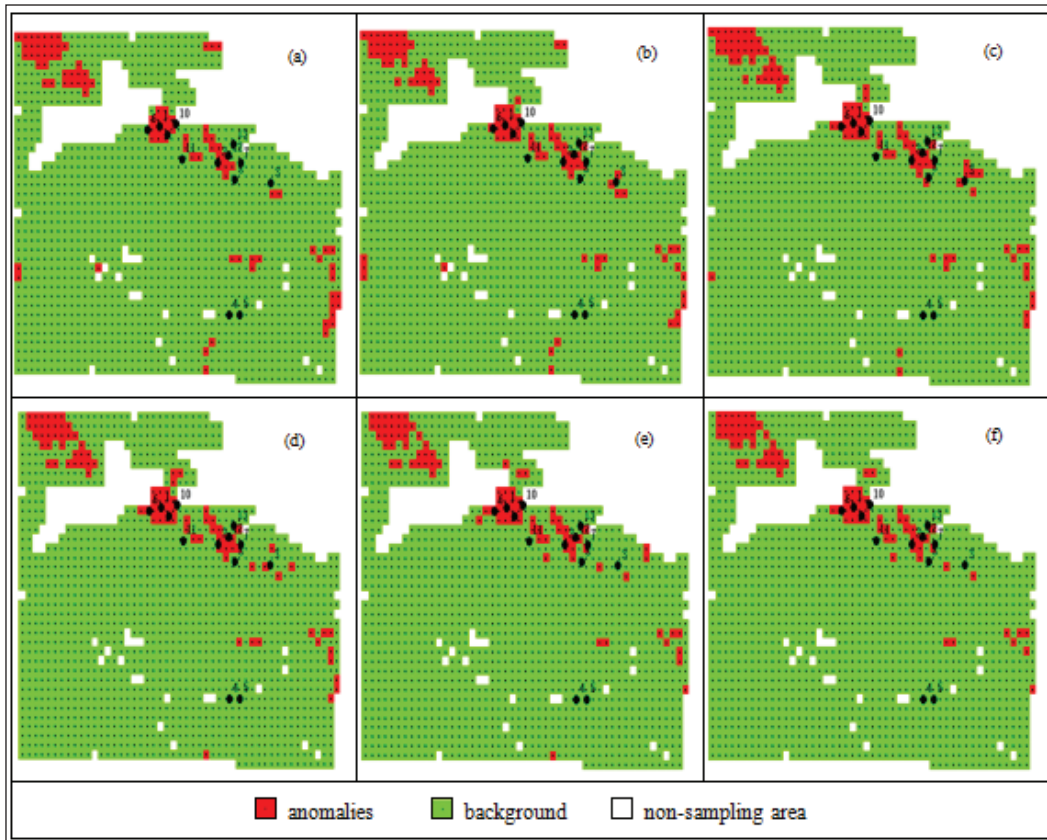


Figure 7- Spatial distribution map of Cu anomalies using local Moran's I_i at six different spatial scales.

5. Conclusions

In this study, a new method was used to separate geochemical anomalies based on local Moran's I_i and robust statistics. The method takes into account spatial correlation and variability, which demonstrates a strong case for geochemical data where the assumption of normality is usually violated. Cu anomalies were separated from 1481 composite samples in Jiurui copper mining area. The process of the anomalous separation includes (1) the analysis of spatial association and variability using Moran scatterplot at different spatial scales for both raw data and Box-Cox transformed data, (2) the measure of spatial autocorrelation using local Moran's I_i and (3) the application of the $MEDIAN \pm 1.5 * IQR$ formula on local Moran's I_i values to separate anomalies. Two important conclusions can be drawn when applying this method: (i) the anomalous separation is strongly affected by the existence of extreme and high values (outliers), therefore, data transformation is a necessary step to reduce their influences; (ii) spatial weight matrices (spatial scales) do affect the results, therefore, it is important to measure spatial autocorrelation of geochemical data at several scales. It was shown that, anomalies were mostly concentrated around known ore-deposits, according to the objective reality and have strong conformity with known ore-deposits in Jiurui copper mining area. Weak anomalies were still not found in areas around ore-deposits No 4 and 5, therefore, the method needs further improvement to separate weak anomalies.

Acknowledgments

I would like to thank to reviewers for insightful comments on this manuscript and Dr. Peng Gong from China University of Geosciences for the data collection.

References

- Afzal, P., Khakzad, A., Moarefvand, P., Rashidnejad Omran, N., Esfandiari, B., Fadakar Alghalandis, Y. 2010. Geochemical anomaly separation by multifractal modeling in Kahang (Gor Gor) porphyry system, Central Iran, *Journal of Geochemical Exploration*, 104, 34-46.
- Afzal, P., Harati, H., Fadakar Alghalandis, Y., Yasrebi, A.B. 2013. Application of spectrum-area fractal model to identify of geochemical anomalies based on soil data in Kahang porphyry-type Cu deposit, Iran. *Chemie der Erde/Geochemistry*, 73, 533-543.
- Ahrens, L. 1953. A fundamental law of geochemistry. *Nature*, 172, 1148.
- Ahrens, L. 1954a. The lognormal distribution of the elements (a fundamental law of geochemistry and its subsidiary). *Geochim Cosmochi Acta*, 5, 49-74.
- Ahrens, L. 1954b. The lognormal distribution of the elements II. *Geochim Cosmochi Acta*, 6, 121-132.
- Ahrens, L. 1957. Lognormal-type distribution III. *Geochim Cosmochi Acta*, 11, 205-213.
- Anselin L. 1992. *Spacestat Tutorial: A workbook for using SpaceStat in the analysis of spatial data*. Urbana: University of Illinois.
- Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis*, 27, 2, 93-115.
- Anselin, L. 2005. *Exploring spatial data with GeoDa™: A workbook*. Urbana: Spatial Analysis Laboratory and Center for Spatially Integrated Social Science, Department of Agriculture and Consumer Economics, University of Illinois, Urbana-Champaign.
- Bailey T. C., Gatrell A. C. 1995. *Interactive Spatial Data Analysis*. Harlow: Addison Wesley Longman, 413p.
- Barnett, V., Lewis T. 1994. *Outliers in statistical data*. New York: Wiley and Sons, 582p.
- Bolviken, B., Stokke, P.R., Feder, J., Jossang, T. 1992. *Journal of Geochemical Exploration*, 43, 91-109.
- Brody, S.D., Highfield, W.E., Thornton, S. 2006. Planning at the urban fringe: an examination of the factors influencing nonconforming development patterns in southern Florida. *Environment and Planning B Planning and Design*, 2006, 33, 1, 75-96.
- Cheng, Q., Agterberg, F.P. 1996. *Mathematical Geology*, 28, 1-16.
- Cocu N., Harrington R., Hulle M., Rounsevell, M.D.A. 2005. Spatial autocorrelation as a tool for identifying the geographical patterns of aphid annual abundance. *Agricultural and Forest Entomology*, 2005, 7, 1, 31-43.
- Cliff, A.D., Ord, J.K. 1981. *Spatial Processes, Models and Applications*. London – Pion, 266p.
- Davis, J.C. 2002. *Statistics and data analysis in Geology* (3th ed.). John Wiley and Sons Inc., New York, 656p.
- Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P. 2003. *Developments in Robust Statistics*. International

- Conference on Robust Statistics, Vorau, Austria, 23-27 July, 2001. Physica Verlag, Heidelberg.
- Garrett, R. 1989. A cry from the heart. *Explore*, 66, 18-20
- Geary, R.C. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5, 3, 115-145.
- Getis, A., Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 3, 189-206.
- Getis, A., Ord, J.K. 1996. Local spatial statistics: an overview. In Longley P, Batty M (eds) *Spatial analysis: modelling in a GIS environment*. Cambridge (UK), GeoInformation International, 269–285.
- Haining R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press, 432p.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. 1986. *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.
- Hashemi, M., Afzal, P., Ras, I. Abedini, M.V. 2010. Geochemical anomaly separation by Concentration-Area fractal model in Bardaskan area, NE Iran. *Journal of Mining and Metallurgy*, 46A, 1-10.
- Hawkes, H.E., Webb, J.S. 1962. *Geochemistry in mineral exploration*. New York: Harper.
- Hoang, A.H., Vu, D.T., Nguyen, T.T. 2017. Spatial Variability Analysis of Cu Content: A Case Study in Jirui Copper Mining Area. *International Journal of Applied Geospatial Research*, 8,1, 81- 93.
- Huber, P.J. 1981. *Robust statistics*. New York: Wiley and Sons, 1981.
- Ishioka, F., Kurihara, K., Suito, H., Horikawa, Y., Ono, Y. 2007. Detection of hotspots for three-dimensional spatial data and its application to environmental pollution data. *Journal of Environmental Science for Sustainable Society*, 1, 15-24.
- Legendre, P., Gauthier, O. 2014. Statistical methods for temporal and space –time analysis of community composition data. *Proc. R. Soc. B*, 281, 20132728. <http://dx.doi.org/10.1098/rspb.2013.2728>.
- Levine, N. 2004. *CrimeStat III: A spatial statistics program for the analysis of crime incident locations*. Ned Levine and Associates, Houston, TX, and the National Institute of Justice, Washington, DC.
- Li, Ch., Ma, T., Shi, J. 2003. *Journal of Geochemical Exploration*, 77, 167-175.
- Liu, Y., Cheng, Q., Xia, Q., Wang, X. 2013. Application of singularity analysis for mineral potential identification using geochemical data - A case study: Nanling W–Sn–Mo polymetallic metallogenic belt, South China. *Journal of Geochemical Exploration*, 134, 61-72.
- Liu, Y., Zhou, K., Cheng, Q. 2017. A new method for geochemical anomaly separation based on the distribution patterns of singularity indices. *Computers and Geosciences*, 105, 139-147.
- Martins-Melo, F.R., Ramos, A.N., Alencar, C.H., Heukelbach, J. 2016. Trends and spatial patterns of mortality related to neglected tropical diseases in Brazil. *Parasite Epidemiology and Control*, 1, 2, 56-65.
- Martins-Melo, F.R., Ramos, A.N., Alencar, C.H., Lima, M.S. 2017. Epidemiology of soil-transmitted helminthiasis-related mortality in Brazil. *Parasitology*, 144, 5, 669-679.
- McGrath, S.P., Loveland, P.J. 1992. *The Soil Geochemical Atlas of England and Wales*. Blackie Academic and Professional, Glasgow, 101p.
- Momeni, S., Shahrokhi, S.V., Afzal, P., Sadeghi, B., Farhadinejad, T., Nikzad, M.R. 2016. Delineation of the Cr mineralization based on the stream sediment data utilizing fractal modeling and factor analysis in the Khoy 1:100,000 sheet, NW Iran. *Bulletin of the Mineral Research and Exploration*, 152, 1-17.
- Nazarpour, A., Omran, N.R., Paydar, G.R., Sadeghi, B., Matroudi, F., Nejad, A.M. 2015. Application of classical statistics, logratio transformation and multifractal approaches to delineate geochemical anomalies in the Zarshuran gold district, NW Iran. *Chemie der Erde-Geochemistry*, 75,1, 117-132.
- Nazarpour, A., Paydar, G.R., Carranza, E.J.M. 2016. Stepwise regression for recognition of geochemical anomalies: Case study in Takab area, NW Iran. *Journal of Geochemical Exploration*, 168, 150-162.
- Nguyen, T. T., Liu, X. G., Ren, Z. 2014. A Study Of Geochemical Exploration Spatial Cluster Identification Based On Local Spatial Autocorrelation. *Geophysical and Geochemical Exploration*, 38, 2, 370-376.
- Nguyen, T.T., Vu, D.T., Trinh, L.H., Nguyen, T.L.H. 2016. Spatial Cluster and Outlier Identification

- of Geochemical Association of Elements: A Case Study in Juirui Copper Mining Area. *Bull. Min. Res. Exp.*, 153, 159-167.
- Philip, G.M., Watson, D.F. 1987. Probabilism in geological data analysis. *Geological Magazine*, 124, 6, 577-583.
- Rajabzadeh, M.A., Yazadni, S., Nazarpour, A., Ahmadi, A. 2015. Application Of Fractal Concentration-Area Method In Identification Of Geochemical Anomalies In Stream Sediments From Mazayejan Area, Suorian 1: 100.000 Sheet, Fars Province. *Geochemistry*, 4, 1, 15-20.
- Reimann, C., Filzmoser, P. 2000. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental geology*, 39, 9, 1001-1014.
- Reimann, C., Filzmoser P., Garrett, R.G.2005. Background and threshold: critical comparison of methods of determination. *Science of the Total Environment*, 346, 1-3,1-16.
- Rock, N., 1988. *Numerical geology*. New York Berlin Heidelberg: Springer Verlag, 427p.
- Rousseeuw, P. J., Leroy, A. M. 1987. *Robust regression and outlier detection*. John Wiley and Sons - New York, 329p.
- Ruiz-Rivera, N., Paul, V.L. 2016. Urban segregation in Latin America. *Habitat International*, 54, 1-2.
- Sinclair, A.J. 1974. Selection of threshold in geochemical data using probability graphs. *Journal of Geochemical Exploration*, 3,129-149.
- Sinclair, A.J. 1976. Application of probability graphs in mineral exploration. *Association of Applied Geochemists*, 4, 95.
- Sinclair, A.J. 1991. A fundamental approach to threshold estimation in exploration geochemistry. *Probability plots revisited*. *Journal of Geochemical Exploration*, 41,1-22.
- Spatial Analysis Laboratory. 2007. *GeoDa: an introduction to spatial data analyses*. Spatial Analysis Laboratory Department of Geography. Urbana: University of Illinois.
- Stanley, C.R., Sinclair, A.J. 1989. Comparison of probability plots and gap statistics in the selection of threshold for exploration geochemistry data. *Journal of Geochemical Exploration*, 32, 355-357.
- Tango, T. 1995. A class of test for detecting ‘general’ and ‘focused’ clustering of rare diseases. *Statistics in Medicine*, 14, 21-22, 2323-2334.
- Tobler, W.R. 1979. Smooth Pchnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, 74, 367, 519-530.
- Tukey, J.W. 1977. *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977.
- Waller, L., Gotway, C.A. 2004. *Applied Spatial Statistics for Public Health Data*. John Wiley and Sons - New Jersey, 520p.
- Wang, F., Hu, Y., Wang, S., Li, X. 2016. Local indicator of colocation quotient with a statistical significance test: examining spatial association of crime and facilities. *The Professional Geographer*, 69, 22-31.
- Zhang, C.S., McGrath, D. 2004. Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods. *Geoderma*, 119, 3-4, 261-275.
- Zhang, T.L., Lin G. 2006. A supplemental indicator of high-value or low-value spatial clustering. *Geographical Analysis*, 38, 2, 209-225.

