Bull. Min. Res. Exp. (2016) 153:159-167

Bulletin of the Mineral Research and Exploration

http://bulletin.mta.gov.tr

SPATIAL CLUSTER AND OUTLIER IDENTIFICATION OF GEOCHEMICAL ASSOCIATION OF ELEMENTS: A CASE STUDY IN JUIRUI COPPER MINING AREA

Tien Thanh NGUYEN[a*], Danh Tuyen VU[a], Le Hung TRINH[b] and Thi Le Hang NGUYEN[a]

[a] Ha Noi University of Natural Resources and Environment, Ha Noi, Viet Nam
Postal Address: No 41A Phu Dien Road, North-Tu Liem District, Ha Noi, Viet Nam
[b] Le Quy Don Technical University, Ha Noi, Viet Nam
Postal Address: 236 Hoang Quoc Viet Road, Cau Giay District, Ha Noi, Viet Nam

Research Article

Keywords:
Spatial Outliers, Spatial Clusters, Spatial Variability, Local Moran Statistic, Geochemistry.

Received: 03.12.2015
Accepted: 15.03.2016

ABSTRACT

Spatial clusters and spatial outliers play an important role in the study of the spatial distribution patterns of geochemical data. They characterize the fundamental properties of mineralization processes, the spatial distribution of mineral deposits, and ore element concentrations in mineral districts. In this study, a new method for the study of spatial distribution patterns of multivariate data is proposed based on a combination of robust Mahalanobis distance and local Moran's I_i. In order to construct the spatial matrix, the Moran's I spatial correlogram was first used to determine the range. The robust Mahalanobis distances were then computed for an association of elements. Finally, local Moran's I_i statistics was used to measure the degree of spatial association and discover the spatial distribution patterns of associations of Cu, Au, Mo, Ag, Pb, Zn, As, and Sb elements including spatial clusters and spatial outliers. Spatial patterns were analyzed at six different spatial scales (2km, 4 km, 6 km, 8 km, 10 km and 12 km) for both the raw data and Box-Cox transformed data. The results show that identified spatial cluster and spatial outlier areas using local Moran's I_i and the robust Mahalanobis accord the objective reality and have a good conformity with known deposits in the study area.

1. Introduction

The properties of living beings are distributed neither uniformly nor at random. They are aggregated in patches or formed gradients or other kinds of spatial patterns (Legendre and Fortin, 1989). Spatial patterns have been widely used in different research fields such as biometrics (Fuentes et al., 2006), landscape ecology (Bagchi et al., 2011; Irl et al., 2015), regional economics (Monastiriotis, 2009) and medicine (Waller and Gotway, 2004; McLaughlin and Boscoe, 2007; Goovaerts and Jacquez, 2004) etc. Extraction of spatial information for mineral exploration can be achieved through the study of spatial distribution patterns of regional geochemical elements such as spatial structures, spatial variability and spatial association patterns (spatial clustering and spatial outliers). Various methods and techniques have been proposed for spatial cluster and outlier identification such as Getis's G index (Getis and Ord, 1992), Geary's C (Geary, 1954), spatial scan statistics (Ishioka et al., 2007) and Tango's C index (Tango, 1995). The local Moran's I is by far the most commonly used test statistic (Anselin, 1995; Getis and Ord, 1996). The local Moran's I statistic has been successfully applied to the spatial cluster identification of diseases (Ruiz et al., 2004; Goovaerts and Jacquez, 2004), mortality rates (James et al., 2004; McLaughlin and Boscoe, 2007), environmental planning (Brody et al., 2006), and environmental sciences (McGrath and Zhang, 2003; Zhang and McGrath, 2004).

Many traditional methods for detecting outliers have been used for many years (Hawkins, 1980). However, the determination of outliers is still being extremely difficult (Rose et al., 1979; Reimann et al., 2005). A number of study (Rocke and Woodruff, 1996; Rousseeuw and Leroy, 1987 and Tyler, 1991) have shown that there have been many methods for the detection of multivariate outliers. Recently, Filzmoser et al. (2005) followed an idea of Gervini (2003) for increasing the efficiency of the robust estimation of multivariate location and scatter. However, this method can be seen as an automation of the method

* *Corresponding author: Tien Thanh NGUYEN, email: tdgis_ntthanh@163.com*
http://dx.doi.org/10.19111/bmre.01695

proposed by Garrett (1989). Filzmoser and Hron (2013) used the Mahananobis as a reliable distance measure for the analysis of multivariate data. For identifying outliers, it is crucial to estimate the mean and covariance from the data. However, geochemical data is a spatial data, all of the aforementioned methods have a common drawback that is they do not take the spatial distribution of data into account. Nguyen et al. (2014) identified spatial patterns for Cu geochemical univariable using local Moran's I, however, distance bands were chosen randomly without scientific basis and the effectiveness of spatial pattern identification for geochemical multiviables using local Moran's I was not investigated. Therefore, in this study, focuses will be made on the identification of spatial patterns of geochemical multivariables by means of a combination of robust Mahalanobis distance and local Moran's I statistic using 1341 stream sediment samples collected at scale of 1:200.000 in Jiurui copper mining area.

2. General Geological Setting of Study Area and Data Used

Jiurui area is a mining district in Jiangxi province (southeast China). Jiurui is rich in copper reserves. The Cu deposits in the Jiurui district are an important component of the Middle–Lower Yangtze River metallogenic belt, extending from Daye in Hubei Province in the west to Zhenjiang in Jiangsu Province in the east. Jiurui area are located at the edge of the active zone on the para platform. The exposed strata are Ordovician limestone, shale, sandstone in upper Silurian system, Huang long formation conglomeratic sandstone in upper carboniferous system, sandstone, dolomite, limestone, limestone in lower Daye, Changxing group's limestone, lower Triassic system Daye group's limestone, middle Jialing river group's limestone, dolomite limestone. Iron copper deposits in the area are one of the main ore deposits in the downstream. They are divided into two metallogenetic series: (i) submarine exhalative activities-related metallogenetic series are any hydrothermal deposits from the injection to the bottom of the sea environment, (ii) intermediate acid hypabyssal intrusive activities-activities-related ore deposits; refers to the formation of intrusive rocks of Carboniferous sand Triassic strata in contact zone and rock deposits. The main types are of skarn type iron and copper deposits, porphyry copper deposit, key, vein copper, gold deposits. Porphy, skarn, copper deposits in the study area belong to this series.

According to requirements of 1:200.000 regional stream sediment survey, a multi-element sediment geochemical survey of streams was carried out in Jiurui area. A total of 1482 composite samples representing about 5364 km^2 were collected. Some sampling areas at the upper part of the study area were not able to access. The sampling density was 1 composite sample per 4 km^2. There are more than 20 indexes in a composite sample, including Ag, As, Au, Be, Cd, Cu, Hg, Li, Mn, Mo, Nb, Pb, Sb, Sn, Th, V, W, Y, Zn, Al_2O_3, CaO, K_2O, Na_2O. Silver, gold, copper are three ore-forming elements. One of three geochemical associations of elements caused anomalous area in the study area is Cu, Au, Mo, Ag, Pb, Zn, As and Sb elements. There are two metallogenic series in the study area. A total of 13 ore deposits were found marked by numeric characters from 1 to 13 (Figure 1).

3. Methodology

3.1. Spatial Correlogram

Spatial correlation of a single variable can be measured by using Moran's I (1950) or Geary's c (1954) spatial correlation statistics (Cliff and Ord, 1981). The Moran's I seems to be the most commonly used statistic (Anselin, 1995; Getis and Ord, 1996), which is given by:

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

where: x_i and x_j are the values of the observed variables at sites i and j, $\bar{x}$ is the average of an observed variable, d is the distance class considered in the calculation, S_o is the sum of the weights W_{ij} and N stands for the number of observations.

Spatial correlation coefficients I(d) are tested for statistical significance by computing the expected values (E) and the variance of the I.

3.2. Robust Distance-based Local Spatial Outlier and Cluster Identification

Standard methods for multivariate outlier detection are based on the Mahalanobis distance (Filzmoser, 2005) which is defined as

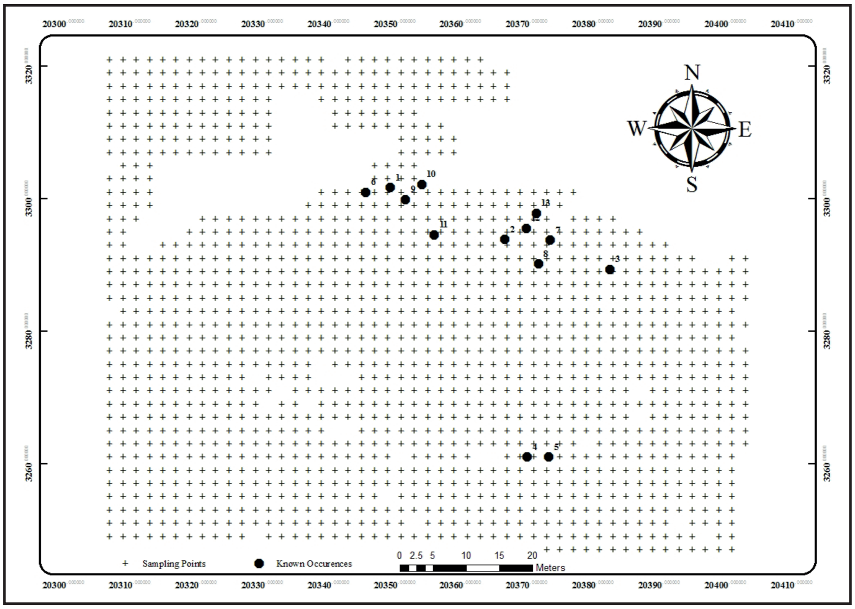

Figure 1- Location map of 1482 stream sediment samples and 13 ore deposits at scale 1:200.000.

$$\mathrm{MD_i} = ((\mathrm{x_i} - \mathrm{t})^{\mathrm{T}} \mathrm{C}^{-1} (\mathrm{x_i} - \mathrm{t}))^{\frac{1}{2}} \qquad (2)$$

for a p-dimensional observation x_i and i = 1, 2,…, n. t is the multivariate arithmetic mean, the centroid, and C is the sample covariance matrix. The Mahalanobis distance is sensitive to the presence of outliers (Rousseeuw and Van Zomeren, 1990)

Using robust estimators of location and scatter in the formula for the Mahalanobis distance equation (2) leads to the so-called robust distances (RDs). Rousseeuw and Van Zomeren (1990) used these RDs for multivariate outlier detection. A global outlier is a measured sample point that has a very high or a very low value relative to all of the values in a dataset. If the squared RD for an observation is larger than $\chi^2_{2;0.98}$, it can be declared a global outlier. In the study, local outliers were considered. The robust Mahanalobis distance for the geochemical association of elements was first used as an investigated variable. Local spatial clusters and outliers were then identified by local Moran statistic using RD. Moran I statistic (Moran, 1948; Cliff and Ord, 1973, 1981) gives a formal indication of the degree of linear association between a vector of observed values y and a weighted average of the neighboring values, or spatial lag, Wy. When the spatial weight matrix is row-standardized such that the elements in each row sum to 1, Moran I is defined as (Anselin, 1995):

$$\mathrm{I} = \frac{\mathrm{y'Wy}}{\mathrm{y'y}} \qquad (3)$$

where, the y is in deviations from their mean, $(x_i - \bar{x})$, and $\mathrm{S_0} = \mathrm{N}$. I is formally equivalent to the regression coefficient in a regression of *Wy* on *y*. A high positive local Moran's I value implies that the location under study has similarly high or low values as its neighbors, thus the locations are spatial clusters (Zhang et al., 2008). If $p(\mathrm{I_i})<\alpha$, $\mathrm{I_i}>0$ and $x_i - \bar{x}>0$, then x_i and $x_{j\in J_i}$ belong to association between high values (high-high clusters) (Figure 2-a). If $p(I_i)<\alpha$, $\mathrm{I_i}>0$ and $x_i - \bar{x}<0$, then x_i and $x_{j\in J_i}$ belong to association between low values (low-low clusters) (Figure 2-d). Spatial outliers are those values that are significantly different from the values of their surrounding locations (Lalor and Zhang, 2001). If $p(\mathrm{I_i})<\alpha$, $\mathrm{I_i}<0$ and $x_i - \bar{x}>0$, then high value, x_i is surrounded by low values, $x_{j\in J_i}$ (high-low outliers) (Figure 2-c). If $p(\mathrm{I_i})<\alpha$, $\mathrm{I_i}<0$ and $x_i - \bar{x}>0$, low value, x_i is surrounded by high values, $x_{j\in J_i}$ (low-high outliers) (Figure 2-b).

3.3. Data Processing

The robust Mahalanobis distances, the descriptive statistical parameters and exploratory data analysis plots were performed and conducted using the StatDA and MASS packages of Statistical Modeling and Computing - R Language (version i386 2.15.0). Moran's I spatial correlograms were carried out using

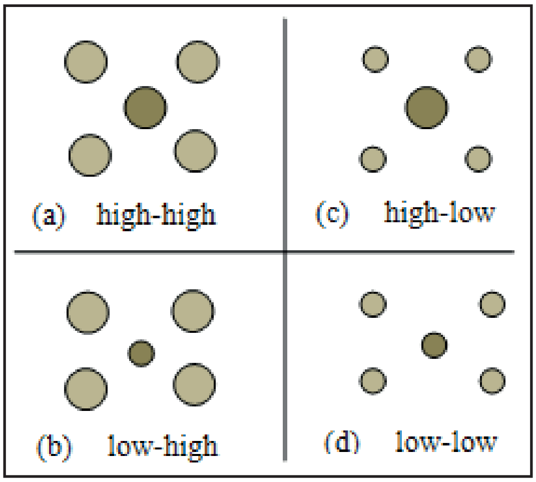

Figure 2- Spatial cluster and outlier schematic diagram (Zhang et al., 2008).

the *ncf* packages in R. The increment of the distance classes were 2 km. Testing for the significance of correlograms with each distance class was performed by a randomization test using 500 permutations. Spatial weight matrix and local Moran's I statistic were carried out with spatial statistics software - GeoDA (version 095i). The local Moran's I statistic was tested using 999 permutations, and the significance level (p-value) was set to 0.05 (5%). The results of spatial clusters and outliers were visualized using ArcGIS 9.3.

4. Results and Discussions

4.1. Distribution of Robust Mahalanobis Distances

Robust Mahalanobis distances (RMD) for original data were calculated using equation (2) are approximately chi-square distributed with a degree of freedom of 8. The distributions of RMDs can be seen in figure 3. The distribution of RMD is obviously strongly right-skewed, as a result typical asymmetrical (Figure 3a). The data points do not follow a straight line of normal Q-Q plots (Figure 3b). Box-Cox transformation was applied to make the distribution of RMD close to normality. The transformed RMD results in a symmetrical distribution and the histogram and density trace show symmetry as shown in figure 3c. The transformed data mostly follow straight lines of Q-Q plots (Figure 3d).

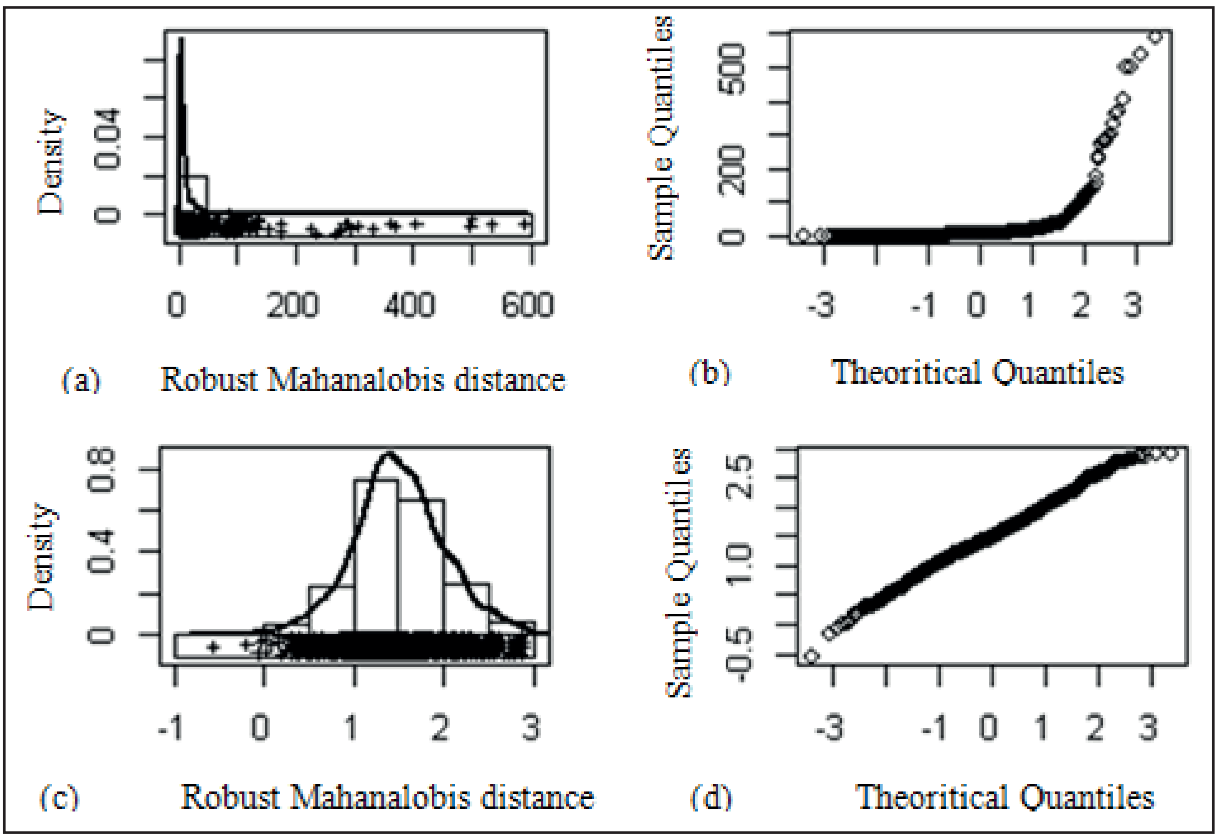

Figure 3- Histogram, density trace, 1-D scatter and Q-Q plot for the RMD of association of Cu, Au, Mo, Ag, Pb, Zn, As, Sb elements: raw data (upper), transformed data (lower).

4.2. Spatial Variability Analysis of RMDs

Both Moran correlograms for RMD and transformed RMD show a phenomenon that Moran correlation coefficients gradually decrease as distances get longer (Figure 4). Moran spatial correlogram found the strongest, positive and significant correlation at distance band ranging from 0 to 2 km for both non-transformed and transformed data. Spatial correlation decreased as distances increased. No significant spatial correlation was found at a distance above 11.1 km for non-transformed data (Figure 4a) and above 14.8 km for Box-Cox transformed data (Figure 4b). It can be concluded that spatial correlation for non-transformed RMDs and transformed RMDs are available when the distance is below 11.1 km. A distance of 12 km was thus applied to construct a spatial weight matrix.

4.3. Spatial Cluster and Outlier Analysis

Six different distance bands were used to construct spatial weights matrix including d = 2, 4, 6, 8, 10 and 12 km. The weights for neighboring locations were assigned 1 if the distances were within the band d, otherwise the weights were 0. To reduce the influence of high values (extreme values and outliers), Box-Cox transformation was applied for the identification of spatial clusters and outliers.

Table 1 shows the results of the identification of spatial clusters and spatial outliers for the association of elements. It can be seen that the number of significant spatial clusters (high-high, low-low) and spatial outliers (low-high, high-low) increased as distance bands increased for both the raw data and Box-Cox transformed data.

Table 1- Summary table of spatial clusters and outliers for association of Cu, Au, Mo, Ag, Pb, Zn, As and Sb elements using the raw RMD and Box-Cox transformed data.

Distance bands (km)	not significant	high-high	low-low	low-high	high-low
d=2	1191	65	145	26	4
d=4	1003	78	301	41	8
d=6	785	101	439	80	26
d=8	676	106	501	110	38
d=10	591	109	544	144	43
d=12	585	94	534	170	48
d*=2	1118	160	112	18	23
d*=4	908	231	199	38	55
d*=6	762	267	255	57	90
d*=8	643	274	295	82	137
d*=10	511	282	356	106	176
d*=12	464	284	384	107	192

For the raw data, the majority of samples were not significant, indicating no presence of spatial correlation or spatial dependence as distance bands were short, such as at 2 km with 1191 insignificant samples, at 4 km with 1003 insignificant samples and at 6 km with 785 insignificant samples in the southern part of the study area (Table 1 and Figure 5-1a, 1b, 1c). There were only 65 high-high samples clustered and 26 low-high, 4 high-low spatial outliers were detected at distance band of 2km (Figure 5-1a). When the distance band increased to 4 km, the number of clusters and outliers increased to 78 high-high samples clustered, 41 low-high and 8 high-low outliers (Figure 5-1a). Following this pattern, the number of clusters and outliers increased as distance bands increased to 6, 8, 10 and 12 km. Most of high-high spatial clusters were found in the north and in the east together with

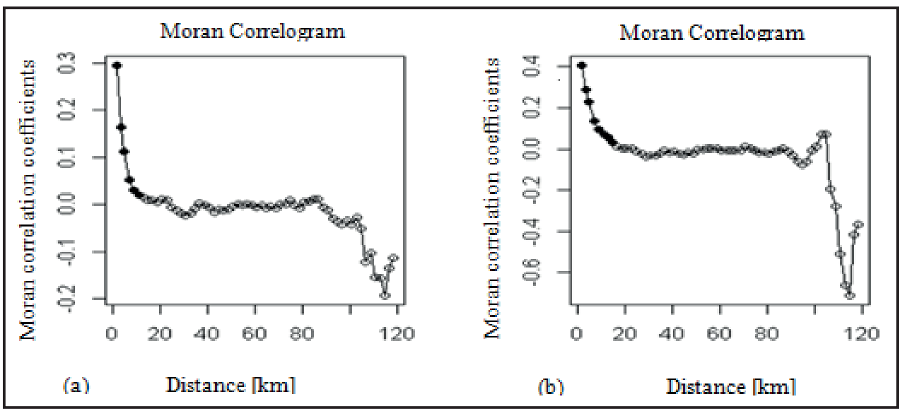

Figure 4- Moran correlograms for original Mahalanobis distance (a) and Box-Cox transformed one (b) of associations of Cu, Au, Mo, Ag, Pb, Zn, As and Sb elements.

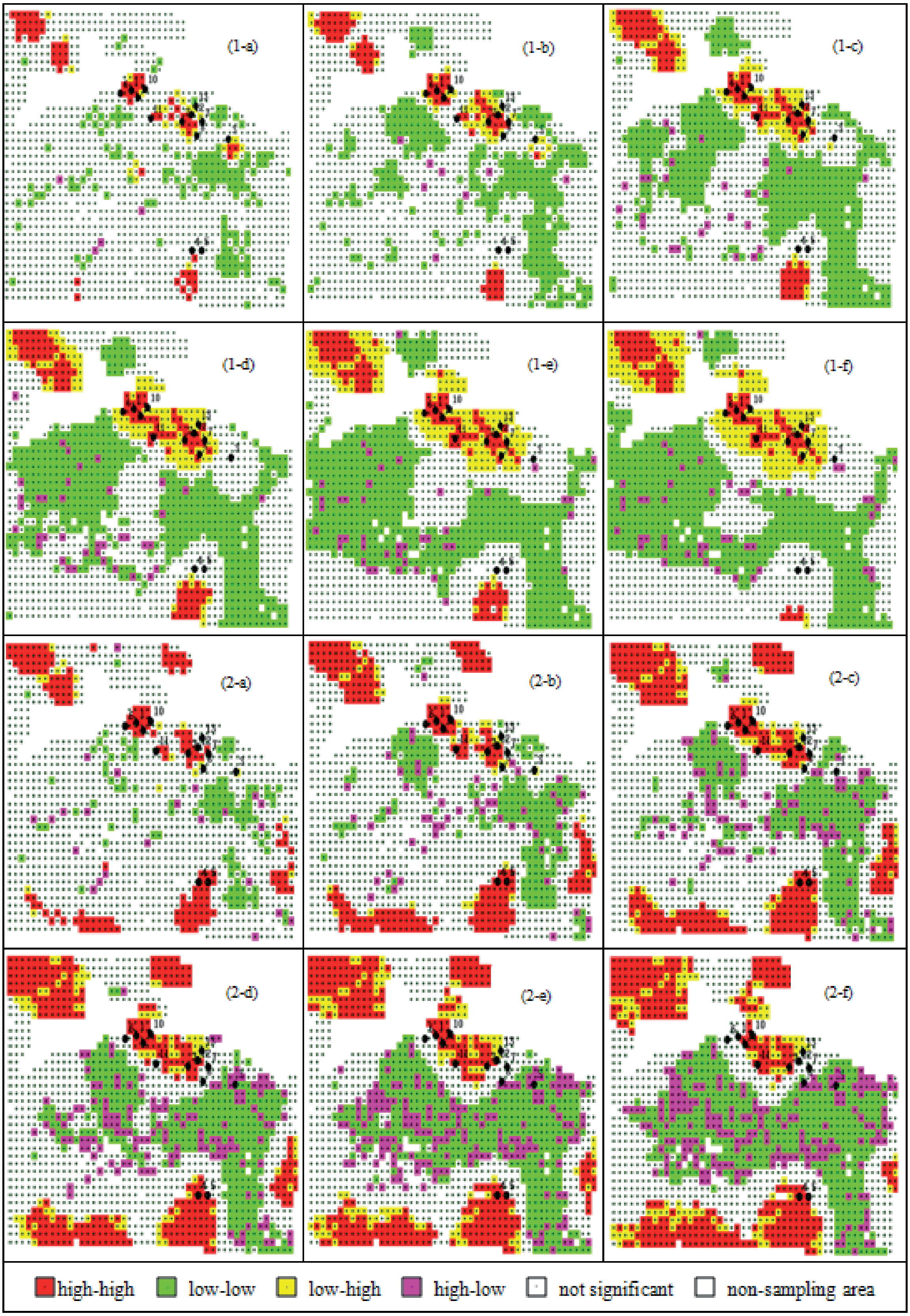

Figure 5- Spatial distribution maps of clusters and outliers of association of Cu, Au, Mo, Ag, Pb, Zn, As, Sb elements for Mahalanobis distances using raw data (1-a,b,c,d,e,f) and Box-Cox transformed data (2-a,b,c,d,e,f) at six different distance bands d = 2, 4, 6, 8, 10 and 12 km.

many low-high spatial outliers, 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 (Figure 5-1a, b, c, e, and f) except at a distance 8 km (Figure 5-1d). Moran's I_i also detected a significant high-high spatial cluster and low-high spatial outlier area in the north-west. Several significant high-low outliers were detected in the west, the south-west and the east, especially as distance bands increased. Significant low-low spatial clusters indicate that local stability occurred in these areas where no known ore deposits were found. The area of high-low outliers detected does not have conformity with known occurrences. It is therefore concluded that high-low spatial outliers played no role in the identification of ore deposits. No significant high-high clusters were found in the southern part where ore deposits 4 and 5 were located (Figure 5-1a, b, c, d, e, and f).

For Box-Cox transformed data, compared with the results of using the raw data, the number of spatial clusters and spatial outliers detected by local Moran's I_i are much more than that of non-transformed data. For example, at a distance of 2 km, there were 160 high-high spatial clusters, 18 low-high and 23 high-low spatial outliers for Box-Cox transformed data. There were 65 high-high spatial clusters, 26 low-high and 4 high-low spatial outliers for the raw data. In this case, high-low spatial outliers came into play. Moran's I_i did not detect spatial outliers or spatial clusters at distance bands of 2, 4 and 6 km where ore deposit 3 was found before (Figure 5-2a, b, and c), but it detected significant high-low spatial clusters at distances of 2 km, 4 km and 6 km (Figure 5-2d, e, and f). Also contrary to the case of non-transformed data, some significant high-high clusters were found in the southern part for Box-Cox transformed data where ore deposits 4 and 5 were located (Figure 5-2a, b, c, d, e, and f).

5. Conclusions

In this study, a new method to study spatial distribution patterns of multivariate data in an association of elements was proposed using robust Mahalanobis distance and local Moran's I_i. Four important issues can be concluded: (1) The results of the identification of spatial distribution patterns of clusters and outliers were strongly affected by the existence of high values (extreme values and outliers), as well as data transformation. It is suggested that data should be transformed first or high values (especially outliers) should be removed before calculation to reduce their influences on the results; (2) The results of the identification of spatial distribution patterns were also influenced by the construction of a spatial weights matrix. Generally, the number of spatial clusters and spatial outliers increased as distance bands increased. Therefore, it is suggested to study spatial distribution patterns at different bands; (3) High-high spatial clusters and low-high spatial outliers played an important role in the identification of local spatial instability and spatial heterogeneity in geochemical data. They may be influenced by extraneous and exotic processes such as those related to rare rock types and mineral deposit formation processes. High-high spatial clusters and low-high spatial outliers provided significant ore-finding information, which can help geochemists to have a better understanding of the potential for mineralization of element associations in the study area; (4) Low-high spatial outliers are a kind of multivariate outliers indicating the existence of local spatial instability and spatial heterogeneity, but they did not come into play in providing ore-finding information.

Acknowledgments

The authors thank Dr. Dieu Tien Bui (Associate Professor, Telemark University College (HiT), Norway) for assistance with suggested improvements to this manuscript and Dr. Peng Gong (formerly of China University of Geosciences) for the data collection.

References

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis,* 27, 2, 93-115.

Bagchi, R., Henrys, P.A., Brown, P.E., Burslem, D.F.R.P., Diggle, P.J., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N., Kassim, A.R., Law, R., Noor, S., Valencia, R.L. 2011. Spatial patterns reveal negative density dependence and habitat associations in tropical trees. *Ecology*, 92, 9, 1723-1729.

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*, 33, 1, 75-96.

Cliff, A.D., Ord, J.K. 1973. Spatial Autocorrelation. London - Pion.

Cliff, A.D., Ord, J.K. 1981. Spatial Processes, *Models and Applications*. London - Pion.

Filzmoser, P., Garrett, R.G., Reimann, C. 2005. Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences*, 31,5, 579-587.

Filzmoser, P., Hron, K. 2013. Robustness for compositional data. In C. Becker, R. Fried and S. Kuhnt, editors, Robustness and Complex Data Structures, Festschrift in Honour of Ursula Gather, *Springer Verlag,* Heidelberg, 117-131.

Fuentes, M., Song, H.R., Ghosh, S.K., Holland, D.M., Davis, J.M. 2006. Spatial Association Between Speciated Fine Particles and Mortality. *Biometrics*, 62, 855-863.

Garrett, R.G. 1989. The chi-square plot: A tool for multivariate outlier recognition. Journal of *Geochemical Exploration*, 32, 1-3, 319-341.

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, editors. Spatial Analysis: Modelling in a GIS Environment. Cambridge: *GeoInformation International*, 239-251.

Gervini, D. 2003. A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *Journal of Multivariate Analysis*, 84,1, 116-144.

Goovaerts, P., Jacquez, G.M. 2004. Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. *International Journal of Health Geographics,* 3, 14-46.

Hawkins, D.M. 1980. Identification of Outliers. *Chapman and Hall* - London.

Irl, S., Harter, D., Steinbauer, M., Puyol, D., Fernández-Palacios, J., Jentsch, A., Beierkuhnlein, K. 2015. Climate vs. topography - spatial patterns of plant species diversity and endemism on a high-elevation island. *Journal of Ecology*, 103, 1621-1633.

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.

James, W.L., Cossman, R.E., Cossman, J.S., Campbell, C., Blanchard, T. 2004. A brief visual primer for the mapping of mortality trend data. *International Journal of Health Geographics,* 3, 1, 1-7.

Legendre, P., Fortin, M.J. 1989. Spatial pattern and ecological analysis. *Plant Ecology*, 80, 2, 107-138.

Lalor, G. C., Zhang, C. S. 2001. Multivariate outlier detection and remediation in geochemical databases. Sci. *Total Environ*, 281, 99-109.

McGrath, D., Zhang, C.S. 2003. Spatial distribution of soil organic carbon concentrations in grassland of Ireland. *Applied Geochemistry*, 18, 10, 1629-1639.

McLaughlin, C.C., Boscoe, F.P. 2007. Effects of randomization methods on statistical inference in disease cluster detection. *Health Place,* 13, 1, 152-63.

Monastiriotis, V. 2009. Examining the consistency of spatial association patterns across socio-economic indicators: an application to the Greek regions. *Empirical Economics*, 37, 1, 25-49.

Moran, P. A. P. 1948. *Biometrika*, 35, 255-60.

Moran, P. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37, 17-23.

Nguyen, T., Liu, X., Ren, Z. 2014. A Study Of Geochemical Exploration Spatial Cluster Identificaton Based On Local Spatial Autocorrelation. *Geophysical and Geochemical Exploration*, 38, 2, 370-376.

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

Rose, A.W., Hawkes, H.E., Webb, J.S. 1979. Geochemistry in Mineral Exploration. *Academic Press* - London.

Rocke, D.M., Woodruff, D.L. 1996. Identification of Outliers in Multivariate Data. *Journal of the American Statistical Association,* 91, 435, 1047-1061.

Rousseeuw, P. J., Leroy, A. M. 1987. Robust regression and outlier detection. *John Wiley and Sons* - New York.

Rousseeuw, P.J., Van Zomeren, B.C. 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85, 411, 633-651.

Ruiz, M.O., Tedesco, C., McTighe, T.J., Austin, C., Kitron, U. 2004. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. *International Journal of Health Geographics*, 3, 8, 2-11.

Tango, T. 1995. A class of test for detecting ‘general’ and ‘focused’ clustering of rare diseases. *Statistics in Medicine*, 14, 21-22, 2323-2334.

Tyler, D.E. 1991. Some issues in the robust estimation of multivariate location and scatter [C]. In W. Stahel and S. Weisberg, editors, Directions in Robust Statistics and Diagnostics 2. *Springer* - New York, 327-336.

Waller, L., Gotway, C.A. 2004. Applied Spatial Statistics for Public Health Data. *John Wiley and Sons* - New Jersey.

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, C.S., Luo, L., Xu, W.L., Ledwith, V. 2008. Use of local Moran’s I and GIS to identify pollution hotspots of Pb in urban soils of Galway, *Ireland. Science of the total environment,* 398, 1-3, 212-221.