

局部空间自相关及其
在勘探地球化学中的空间聚集研究

Nguyen Tien Thanh^{1,2}, 刘修国¹, 任泽¹

(1.中国地质大学 信息工程学院,湖北 武汉 430074;2. 河内矿业与地质大学 测绘学院,越南 河内 130510)

摘 要:运用局部空间自相关 Moran 的 I_i 和 G_i^* 指数,研究 8 130 km² 区域内采集的 1 : 20 万水系沉积物 1 482 个组合样点的空间聚集情况,与基于稳健统计学的地球化学异常下限确定方法所圈定的异常范围对空间聚集范围进行对比。结果表明,局部 I_i 指数能够很好地探测出程度较弱和范围较小的空间聚集,局部 G_i^* 指数适合于程度较强和范围大的高高聚集探测,所探测出的空间聚集范围较大。使用该方法所发现的地球化学空间聚集范围和程度符合客观实际,与已知矿床(点)吻合很好。

关键词:局部 Moran 的 I_i 指数;局部 G_i^* 指数;空间聚集;勘探地球化学

中图分类号: P632 **文献标识码:** A **文章编号:** 1000-8918(2014)02-0370-07

地学变量区别于数学上随机变量的显著特征,在于其在空间分布上的相关性,它们既具随机性,又有规律性。用传统的数理统计方法对其进行研究,由于其不完全满足随机条件而可能产生偏差,因此产生了空间统计学,对空间分布变量及其分布结构进行定量研究^[1]。空间自相关是空间统计学中可用于空间结构、空间聚集研究的有效方法。空间自相关是指同一变量在不同空间位置上的相关性,是空间单元属性值聚集程度的一种度量^[2-4]。空间自相关性使用全局和局部两种指标来度量,全局指标用于探测整个研究区域的空间模式,局部指标计算每一个空间单元与邻近单元就某一属性的相关程度。众多学者把局部指标成功地应用于环境科学、环境规划,以及鉴定疾病和死亡率热点等多学科领域^[5]。McGrath 研究爱尔兰浓度草原的土壤有机碳的空间分布^[6],Zhang^[7-8]使用了 Moran 的 I_i 和 GIS 识别铅污染热点地区并确定异常,Brody^[9]将 Moran 的 I_i 用于环境规划,McLaughlin^[10]和 Goovaerts^[11]将 I_i 用于疾病和死亡率鉴定研究,张朝生^[1]利用了空间自相关与分形方法研究长江水系沉积物重金属含量空间分布特征。

本研究中,首次引入局部空间自相关(局部 Moran 的 I_i 和 G_i^* 指数)研究地球化学找矿中的地球化学指标的空间聚集情况,与基于稳健统计学的地

球化学异常下限确定方法所圈的异常范围对空间聚集范围进行对比,并且以一个具体的区域为例。该区总面积为 8 130 km²,取样面积为 5 718 km²,采集 1 : 20 万水系沉积物的 1 482 个组合样点。

1 研究方法

1.1 局部空间自相关 Moran 指数

空间位置 i 上,其局部 Moran 的 I_i 指数的定义^[12]为

$$I_i = (z_i - \bar{z}) \sum_{j \in J_i}^n w_{ij} (z_j - \bar{z}),$$

式中, z_i 、 z_j 分别是第 i 、 j 个空间对象的观测值(空间属性), $\bar{z}$ 是均值; J_i 是第 i 个空间对象的相邻对象集合; w_{ij} 是表达 n 个空间对象的空间相邻关系。空间权重矩阵可根据相邻(公共边界)标准或距离标准来度量^[13-14],如果第 i 个空间对象和第 j 个空间对象属于相邻关系,则 $w_{ij} = 1$,否则 $w_{ij} = 0$ 。通常约定,一个空间对象与其自身不属于相邻关系,即 $w_{ij} = 0$ 。如果局部 Moran 的 I_i 指数呈现正态分布,则可以通过理论期望 $E(I_i)$ 和理论方差 $V(I_i)$:

$$Z(I_i) = [I_i - E(I_i)] / \sqrt{V(I_i)}$$

来确定 $Z(I_i)$ 得分。如 $Z(I_i)$ 是一个较高的正值且 $z_j - \bar{z} > 0$,则表明位置 i 和周围相邻的观测值都相对较高(高高集聚);如 $Z(I_i)$ 是一个较高的正值且 $z_j -$

$\bar{z}<0$,则表明位置 i 和周围相邻的观测值都相对比较低(低低集聚);如 $Z(I_i)$ 是一个较高的负值且 $z_j-\bar{z}>0$,则表明周围相邻观测值远低于位置 i 上的观测值(高低集聚);如 $Z(I_i)$ 是一个较高的正值且 $z_j-\bar{z}<0$,则表明周围相邻观测值远高于位置 i 上的观测值(高高集聚)(图 1)。

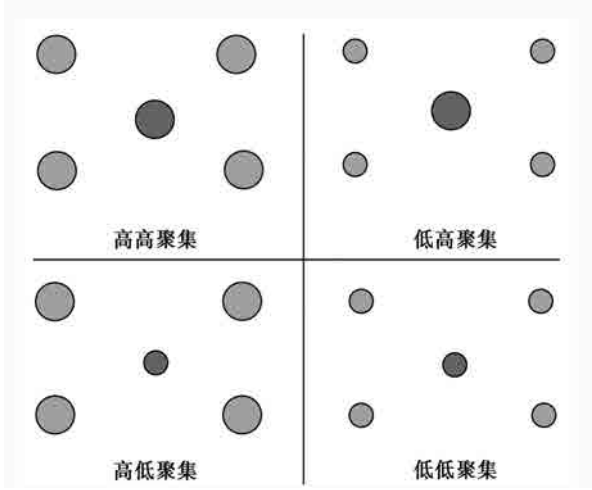


图 1 空间聚集示意

1.2 局部空间自相关 G_i^* 指数

Getis 和 Ord^[2] 提出了度量每个空间位置上的观测值与周围相邻观测值之间是否存在局部空间关联的 G_i^* 指数,其的表达式为

$$G_i^* = \frac{\sum_{j \in J_i}^m w_{ij}(d) z_j}{\sum_j^n z_j}。$$

式中: $w_{ij}(d)$ 是根据距离规则定义的空间权重矩阵, m 是临界距离(d)范围内的空间对象。

局部 G_i^* 指数可以识别高值或低值在空间上发生聚集的位置(热点)。通过查看第 i 空间对象邻近环境中的每一个对象,高值往往容易引起注意,但可能不是具有显著统计学意义的热点。要成为具有显著统计学意义的热点,对象 i 需具有高值,且被其他同样具有高值的对象所包围。某个对象及其相邻对象的局部总和将与所有对象的总和进行比较。当局部总和与所预期的局部总和有很大差异,以致于无法成为随机产生的结果时,会产生一个具有显著统计学意义的 $Z(G_i^*)$ 。 $Z(G_i^*)$ 为正,且非常显著,则表明对象 i 的相邻值属于高值空间集聚, $Z(G_i^*)$ 为负,且非常显著,则表明对象 i 的相邻值属于低值空间集聚(冷点)。 $Z(G_i^*)$ 得分可以通过

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{V(G_i^*)}$$

来确定。

1.3 Z 得分和 p 值的确定

数据和空间自相关性质的假设不同,理论表达式 $E(I_i)$ 、 $V(I_i)$ 、 $E(G_i^*)$ 和 $V(G_i^*)$ 会有所不同,则 $Z(I_i)$ 和 $Z(G_i^*)$ 得分会有所不同^[14-15]。最常见的方法是假设讨论中的变量服从正态分布,基于渐近考虑(即假设的样本可能变得无限大) Z 得分,采用适当措施确定平均值和标准偏差,服从标准正态分布。统计显著性水平可以根据 Z 得分和它在标准正态分布表的概率的比较作出判断。在观测值服从正态的前提下, I_i 和 G_i^* 的分布才是正态的,但很多与地球化学相关的著作都表明地球化学数据往往不服从正态分布^[16],所以该方法不适合应用于空间自相关检验。第二种常用的方法是假设每一个观察值同样可能发生在所有地方。换言之,观测值的位置与它们的空间安排无关,这被称为随机化假设。基于这样的假设,得到不同的理论方差,产生不同的 Z 得分。同样,这些 Z 得分服从标准正态分布(渐进),可使用标准正态分布表查显著性水平。最后一种方法是类似于随机化假设,其原理是固定观测值 i ,对剩下的 $n-1$ 观测值进行排列。每个观测值都有同样的可能在某个位置上出现,根据所产生的参考经验分布计算期望和方差。在实践中,可以在所有位置随机地改组观测值(也就是排列这些观测值),并重新计算每个新样品的 I_i 和 G_i^* 指数,这被称为随机排列法。 I_i 和 G_i^* 的期望和方差通过在整个空间位置上随机排列参考分布中的期望和方差来计算。

模式分析的零假设是完全随机性的,通过 $Z(I_i)$ 和 $Z(G_i^*)$ 得分和 $p(I_i)$ 和 $p(G_i^*)$ 值来判断是否拒绝零假设。我们希望的是 $Z(I_i)$ 和 $Z(G_i^*)$ 得分和 $p(I_i)$ 、 $p(G_i^*)$ 值可以拒绝零假设,这就意味着观测值应表现出统计意义上的显著性集聚或离散模式,而不是随机模式。 $p(I_i)$ 和 $p(G_i^*)$ 值表示所观测到的空间模式是由某一随机过程创建而成的概率,当 $p(I_i)$ 和 $p(G_i^*)$ 很小时,意味着所观测到的空间模式不太可能产生于随机过程,因此可以拒绝零假设。要拒绝零假设,必须对所愿承担的可能作出错误选择的风险程度做出主观判断。

因此,选择一个置信度,然后再执行空间统计。典型的置信度为 90%、95% 或 99%。 $Z(I_i)$ 和 $Z(G_i^*)$ 得分和 $p(I_i)$ 和 $p(G_i^*)$ 值都与标准正态分布相关联^[17](图 2)。

1.4 实验数据和所使用的软件

按照 1 : 20 万水系沉积物测量要求,笔者在一

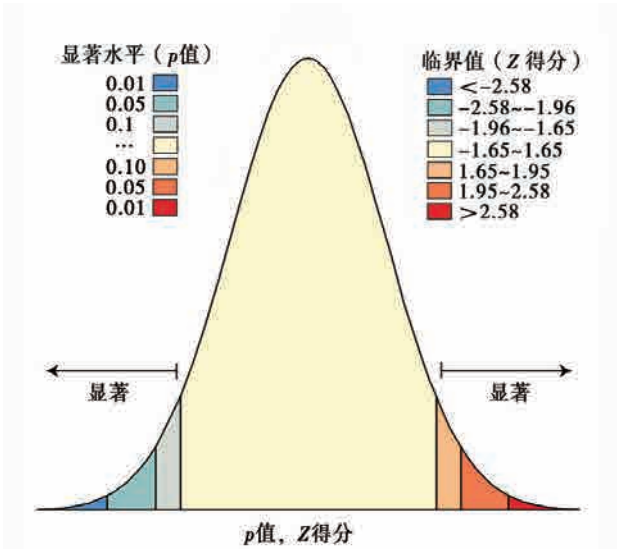


图2 标准正态分布

个具体的地区进行了取样工作。取样面积为 5 730 km², 1 482 个组合样点(图 3)。平均采样密度是 1 个组合样/4 km²。每件样品测试项目包括 Ag、As、Au、Be、Cd、Cu、Hg、Li、Mn、Mo、Nb、Pb、Sb、Sn、Th、V、W、Y、Zn、Al₂O₃、CaO、K₂O、Na₂O 等指标。

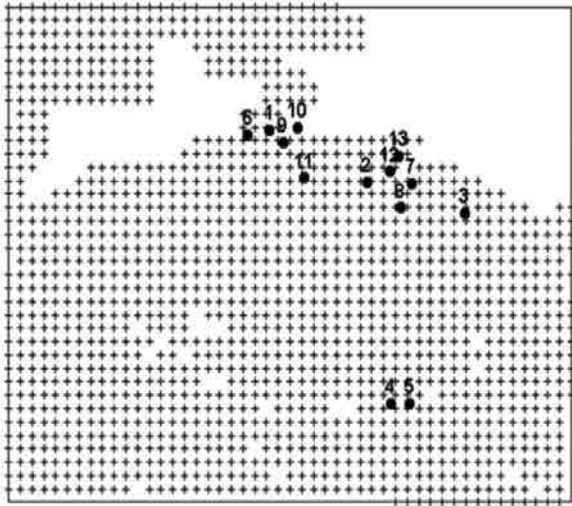


图3 研究区采样点位

本区域有两组铜成矿带,发现了 13 个矿点,主要成矿元素为 Cu、Ag、Au。为了探讨空间聚类在地球化学找矿中的应用,本次只对 Cu 指标进行研究。

利用 ArcGIS 9.3 和 GeoDA 095i^[18] 地理信息系统软件提供的空间分析方法进行空间自相关分析,用 StaDA 程序包确定异常下限和做统计图。为了检验 I_i 和 G_i^* 统计量是否显著,在 GeoDA 中采用 Monte Carlo 模拟的方法,用 999 次随机排列构造参考分布。利用 ArcGIS 计算 G_i^* 指数,然后通过 ArcGIS 可视化技术软件直观地揭示空间数据特性,探测

数据空间联系的格局,统计空间聚类位置及面积。为了全面地研究阈值距离对空间聚类确定的影响,用不同距离(2、4、6、8、10、12 km)构建空间权重矩阵,分别计算 I_i 和 G_i^* 指数,并进行比较,比较的侧重点是两种指标得出的聚集范围和已知矿点的吻合程度。

2 结果和讨论

从图 4 和图 5 可以看出,Cu 指标含量是正偏态分布(偏度为 15.2),最小值为 9.2×10^{-6} ,最大值为 $1\,926 \times 10^{-6}$,均值为 38.2×10^{-6} ,中位数为 27.1×10^{-6} 。基于传统方法(均值加两倍标准差)确定异常下限的结果受异常点的影响。为了进行对比,笔者应用了基于稳健统计学的异常下限确定方法,由中位数(median)和偏离绝对值(MAD)公式^[19]来确定的异常下限为 48,在累计概率图上明显发现从该异常点逐渐偏离直线(图 5),所圈的异常范围见图 6,基于 I_i 和 G_i^* 的空间聚集面积确定如表 1 所示。

由于地球化学找矿只关心高含量元素的空间分布,所以这里只对高高,低高和高低聚集进行讨论。所选择的置信度为 95%,意味着有随机过程创建的

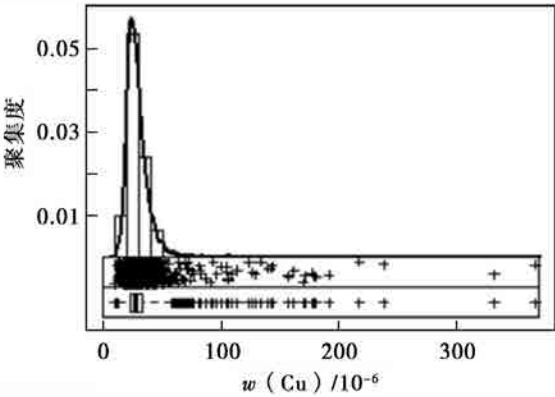


图4 研究区 Cu 含量分布

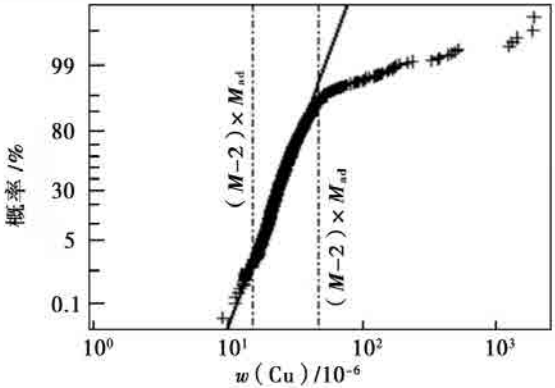
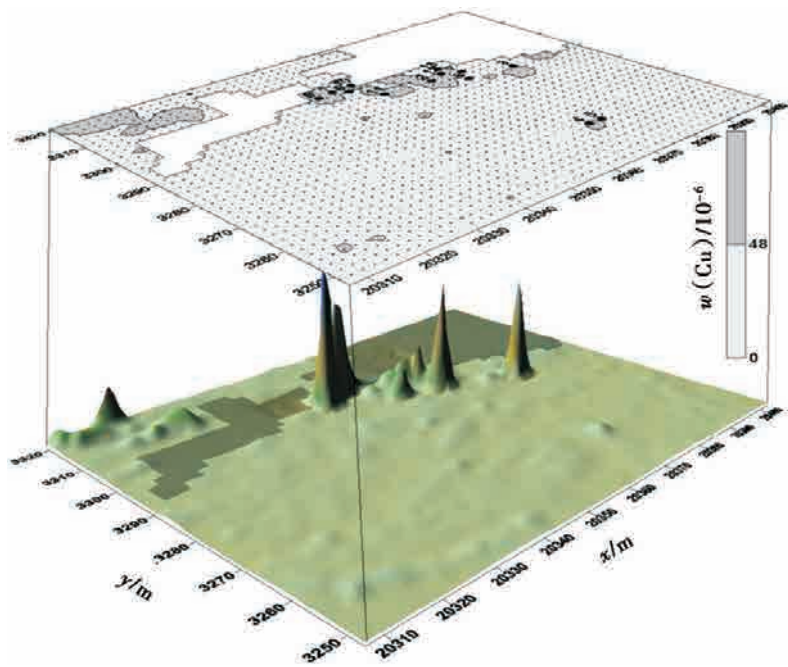


图5 研究区 Cu 含量累积概率



上图为基于稳健统计学的异常,下图为原始数据的三维图

图 6 研究区 Cu 异常

表 1 空间聚类面积 km²

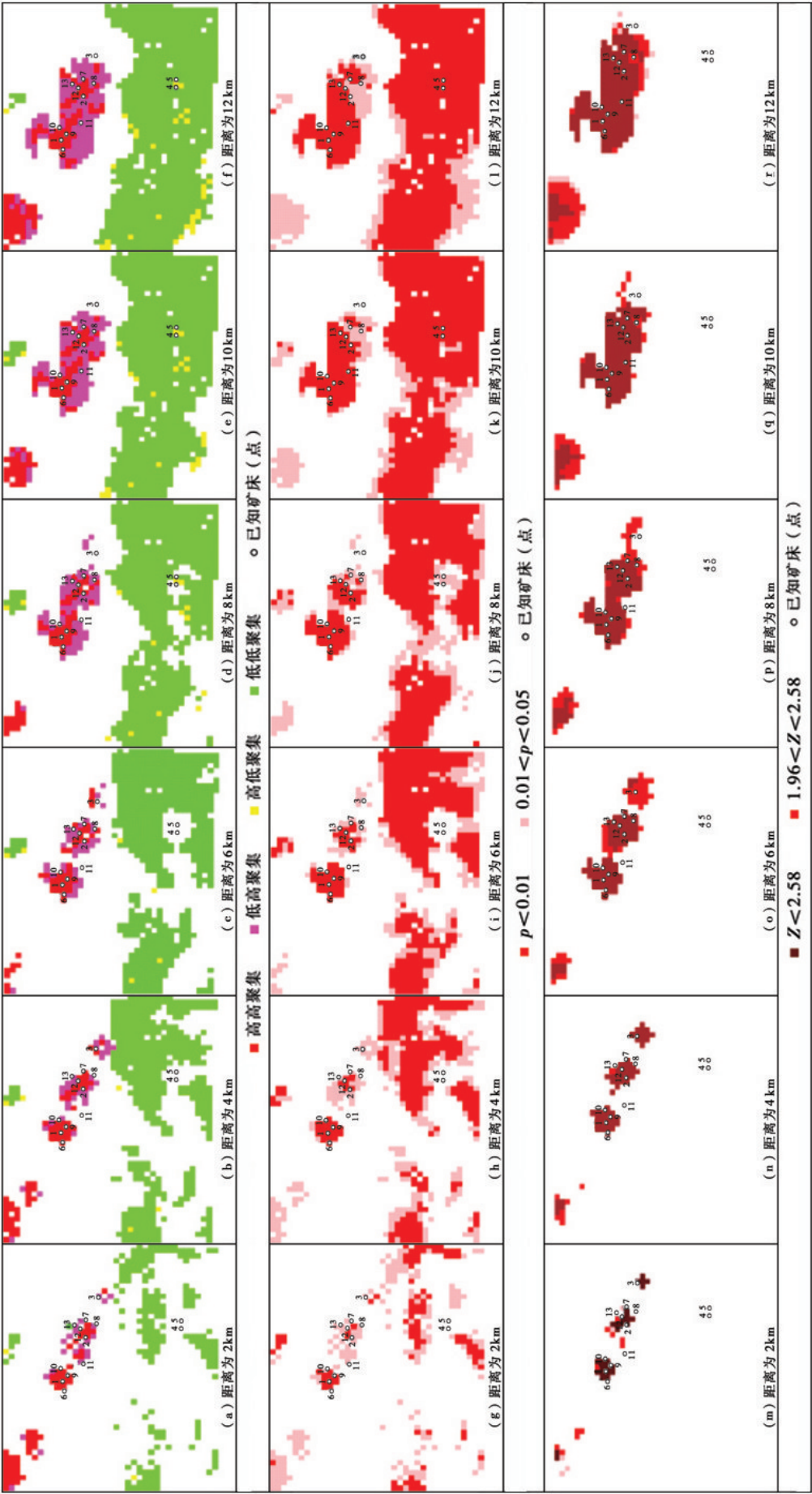
距离 km	I _i 指数			G _i [*] 指数
	高高	低高	高低	
2	228	88	8	156
4	248	128	20	276
6	188	148	24	484
8	228	248	60	576
10	328	376	100	760
12	396	492	132	916

概率 p 值小于 5%(具有比较强的空间相关性)。除此之外还提供 99% 的置信度, p 值小于 1%(具有很强的空间相关性)。

总体来看, I_i 指数所确定的空间聚集面积随阈值距离增大而增大, 在给定不同距离范围内, I_i 都能较好地探测出聚集区域的中心, 在不同程度上, I_i 所探测出的聚集区域与稳健统计学方法一致, 并与实际已发现的矿床(点)较吻合。给定的距离不同, 在北向和东北向 I_i 都发现非常显著的高高聚集(p 值 <0.01), 这些聚集位置与已知矿床 6、1、9、10、2、13、7、8 的位置很吻合(图 7a~m), 并且在高高聚集范围的附近发现较显著的低高聚集, 它合理地解释了地质体逐渐从低到高改变的趋势。虽然在范围距离为 8、10、12 km 的情况之下 I_i 没能发现在东向矿床 3 位置上的空间聚集(图 7d~f, j~m), 但是距离为 2、4、5 km 它却发现具有较显著的高高聚集和低高聚集(图 7a~c, g~i)。同样, 在距离为 2、4、6 km 和显著水平为 0.05 情况之下, I_i 没能探测出西向(矿床 4 和 5)和北向的矿床 11 具有显著的空间聚

集(在 0.1 显著水平之下在矿床 4 的位置上 I_i 有空间聚集的出现)。距离为 8 km 的情况下, I_i 能较好地探测出在矿床 5 的位置上非常显著的高高聚集和在矿床 4 位置上显著的高低聚集(图 7d、h)。距离为 10、12 km 的情况下, I_i 准确地探测出在矿床 4 和 5 的位置上出现非常强的高低聚集(图 7e、f, k、m)。与稳健统计学方法相似, I_i 也发现西北向出现较显著的高高聚集和低高聚集面积较大, 西向、西南向、东向有一些地方有较显著的高低聚集, 尤其是距离为 8、10、12 km 的情况下(图 7d~f, j~m), 但实际情况是该地没有矿点, 这些位置的空间聚集由地质体引起。

局部 G_i^* 指数所确定的空间聚集面积也随着距离增大而增大。在不同距离范围内, 除了矿床 4 和 5 之外, G_i^* 所发现的高高聚集范围与北向、东北向、东向的已知矿床基本吻合。与局部 I_i 指数相比, 在同距离的情况下, G_i^* 能更好地探测出高高聚集位置, G_i^* 所发现的高高聚集面积更大并且具有非常强的空间自相关性。与 I_i 相似, 虽然在距离为 2、4、6 km 和显著性水平为 0.05 的情况下 G_i^* 没能探测出在矿点 11 的位置上有高高聚集的出现(图 7m~o), 但是在 8、10、12 km 情况下都能准确地探测出非常显著的高高聚集。在距离为 8、10、12 km 的情况下 G_i^* 和 I_i 都没能发现在矿点 3 位置上的高高聚集(图 7p~r), 但在距离为 2、4、6 km 处都很准确地探测出非常显著的高高聚集, 与实际情况吻合(图 7l~



a~f、g~l 分别为在不同距离内基于 Moran 指数 I_i 在显著性水平为 0.01 和 0.05 下的空间聚集程度；m~r 为 G_i^* 在显著性水平为 0.01 和 0.05 下的分布

图 7 Cu 含量的空间聚集和显著性关系

o)。局部 G_i^* 指数在东向不同距离的情况下没能发现高高聚集(实际情况是有两个矿点 4 和 5),原因是在该区域 Cu 指标有三个观测值大于均值,其含量依次为 $(50.5, 52, 105) \times 10^{-6}$ (实际上是一个较弱的高低聚集,见图 6),范围很小,在不同的空间尺度下,它们周围的观测值数量比较大,含量都小于均值,导致局部 G_i^* 指数没能准确地探测出该位置的空间聚集情况。在西北向 G_i^* 也发现了面积较大、较显著的高高聚集,但这都是由地质体引起的空间聚集。

3 结论

将局部空间自相关 Moran 的 I_i 和 G_i^* 指数引入到地球化学找矿中研究空间聚集。该方法的优越性是在计算的过程中考虑样品之间的空间关系,同时通过统计检验使得结论更加可靠,并且与基于稳健统计学的地球化学异常下限确定方法所圈的异常范围对空间聚集范围进行对比。研究结果表明,局部 I_i 和 G_i^* 指数在不同的距离范围内都能比较好地探测空间聚集,特别是对高含量地球化学元素在空间上的高高聚集确定(类似于异常)具有重要的意义,符合客观实际,与已知矿床(点)吻合很好,与基于稳健统计学的异常范围确定一致。 I_i 的优点是在给定距离范围内能很好地探测出聚集区域的中心,特别适合于探测较弱和范围小的空间聚集(如上述所讨论矿点 4 和 5 的位置上空间聚集的出现), I_i 还能探测出高低聚集、低高聚集。这两者对认识地球化学元素含量改变的趋势和较弱的空间自相关(若异常)具有重要的意义。 G_i^* 适用于元素含量较高、分布范围较大的区域,在这种情况下 G_i^* 比 I_i 能更好地探测出高高聚集位置。实际工作中可以综合应用 I_i 和 G_i^* 来研究地球化学元素的空间分布。

参考文献:

[1] 张朝生,章申,何建邦. 长江水系沉积物重金属含量空间分布特征研究——空间自相关与分形方法[J]. 地理学报, 1998, 53(1): 88-95.

[2] Getis A and Ord J K. The analysis of spatial association by use of distance statistics[J]. Geographical Analysis, 1992, 24: 189-206.

[3] Bivand R. A review of spatial statistical techniques for location studies[R]. CEPR symposium on new issues in trade and location,

Lund, Sweden, 1998: 1-21.

[4] Goodchild M F. Spatial autocorrelation, concepts and techniques in modern geography[M]. 47 Geo Books, Norwich, 1986.

[5] 李勇,周永章,张澄博,等. 基于局部 Moran's I 和 GIS 的珠江三角洲肝癌高发区蔬菜土壤中 Ni、Cr 的空间热点分析[J]. 环境科学, 2010, 31(6): 1617-1623.

[6] McGrath D, Zhang C S. Spatial distribution of soil organic carbon concentrations in grassland of Ireland [J]. Applied Geochemistry, 2003, 18(10): 1629-1639.

[7] Zhang C S, Luo L, Xu W L, et al. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland [J]. The Science of the total environment, 2008, 398(1-3): 212-221.

[8] Zhang C S, Tang Y, Luo L, et al. Outlier identification and visualization for Pb concentration in urban soils and its implications for identification of potential contaminated land [J]. Environmental Pollution, 2009, 157(11): 3083-3090.

[9] Brody S D, Highfield W E, Thornton S. Planning at the urban fringe: an examination of the factors influencing nonconforming development patterns in southern Florida[J]. Environment and Planning, 2006, 33: 75-96.

[10] McLaughlin C C, Boscoe F P. Effects of randomization methods on statistical inference in disease cluster detection[J]. Health Place, 2007, 13(1): 152-163.

[11] Goovaerts P, Jacquez G M. 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[J]. International Journal of Health Geographics, 2004(3): 14.

[12] Anselin L. Local indicators of spatial association -LISA[J]. Geographical Analysis, 1995, 27(2): 93-115

[13] Anselin L. Spatial Econometrics: Methods and Models [M]. Boston: Kluwer Academic Publishers, 1988.

[14] Cliff A D, Ord J K. Spatial Processes, Models and Applications [M]. Pion, London, 1981.

[15] Cliff A D, Ord J K. Spatial Autocorrelation[M]. London: Pion, 1973.

[16] Ahrens L H. Lognormal-type distribution III [J]. Geochim Cosmochim Acta, 1957, 11: 205-213.

[17] Mitchell, Andy. The ESRI Guide to GIS Analysis Volume 2 [M]. ESRI Press, 2005.

[18] Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. In Fischer MM, Getis A (eds) Handbook of applied spatial analysis [M]. Springer, Berlin, Heidelberg and New York, 2009: 73-89.

[19] Emmanuel J M. Carranza. Geochemical Anomaly and Mineral Prospectivity Mapping in GIS[M]. The Netherlands, 2008: 57-58.

A STUDY OF GEOCHEMICAL EXPLORATION SPATIAL CLUSTER IDENTIFICATON BASED ON LOCAL SPATIAL AUTOCORRELATION

Nguyen Tien-thanh^{1,2}, LIU Xiu-guo¹, REN Ze¹

(1. Faculty of International Education, China University of Geosciences, Wuhan 430074, China; 2. Faculty of Surveying and Mapping Ha Noi University of Mining and Geology, Ha Noi 130510, Viet Nam)

Abstract: Using local Moran’s and indexes, the authors studied spatial cluster of 1482 stream sediment samples in a specific area of 8130 km² on the scale of 1:200000. The comparison of the anomaly range computed by robust statistics with that computed by local spatial autocorrelation shows that local Moran’s index can well detect relatively weak spatial cluster and small areas while local index is suitable for spatial cluster detection in a strong degree of spatial cluster and large areas. The degree and area of geochemical spatial cluster by using these approaches accord with the objective reality and are in good conformity with known deposits (ore spots).

Key words: local Moran’s index; local index; spatial cluster; geochemical exploration

作者简介: Nguyen Tien Thanh (1984–), 越南人, 博士研究生, 从事 GIS 与成矿预测工作。
通讯作者: 刘修国 (1969–), 博士生导师, 研究方向: 地理信息系统。

上接 362 页

A COMPARATIVE STUDY OF THE DATA INTEGRATION METHODS FOR GEOCHEMICAL AND CULTIVATED LAND QUALITY INVESTIGATION

SANG Ling-ling^{1,2}, GAO Ge³, YUN Wen-ju^{2,4,5}, ZHU De-hai^{4,5}

(1. College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China; 2. Land Consolidation and Rehabilitation Center, Ministry of Land and Resources, Beijing 100035, China; 3. Shaanxi Institute of Surveying and Mapping of Land, Xi’an 710054, China; 4. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; 5. Key Laboratory of Land Quality, Ministry of Land and Resources, Beijing 100035, China)

Abstract: Land quality geochemical assessment and cultivated land quality evaluation constitute the basic work of the land resources quality investigation and evaluation. Integration research on the two results is an inevitable trend between the current data sharing and cross-disciplinary data integration. In this paper, from data independence, results practicability and scientificity, the two results were combined by simple superposition method, comprehensive method or factor method. With Qianguo County in Jinlin Province as a case study area, the authors investigated the advantages and disadvantage of the integration results. The results show that different integration methods had different advantages and disadvantages: the simple superposition method based on data of two results could maintain their respective independence; the comprehensive method based on the results is easy to express, but with too much emphasis placed on experience; the factor method based on two survey data can clearly express the physical meaning, but with insufficient theoretical support.

Key words: geochemistry, land resource evaluation; cultivated land quality survey; data integration; Qianguo County in Jinlin Province

作者简介: 桑玲玲 (1983–), 博士后, 主要研究方向为土地利用变化与土地评价。E-mail: sangling@foxmail.com。
通讯作者: 郎文聚, 教授, 博士生导师, 主要从事土地整治和土地评价研究。E-mail: yunwenju@vip.sina.com。