

空间自相关分析方法在流行病学中的应用

姜庆五 赵飞

【关键词】 空间自相关; 地理信息系统

Application of spatial autocorrelation method in epidemiology JIANG Qing-wu, ZHAO Fei. Key Laboratory of Public Health Safety, Ministry of Education, School of Public Health, Fudan University, Shanghai 200032, China

Corresponding author: JIANG Qing-wu, Email: jiangqw@fudan.edu.cn

【Key words】 Spatial autocorrelation; Geographical information system

一、前言

在多元统计分析中,如果开展线性回归分析,则至少满足两个基本条件:一是解释变量之间相互正交,二是样本要素(样品或者样点)之间彼此独立。当只有一个解释变量的时候,样品独立性就是最基本的前提条件。假如分析对象是空间取样结果,则样点之间要求不存在显著的空间关系,否则回归模型参数的可靠性不能保证。因此,开展空间回归分析的预备工作是空间样点之间的相关性分析。如果空间样点之间相关性不显著,则可以建立空间回归模型;否则,常规的回归模型失效,这时可以通过样本要素之间的空间相关分析揭示某些统计规律或特征。于是,空间自相关理论和方法应运而生^[1]。

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]。一种现象的观测值如果在空间分布上呈现出“高的地方周围也高,低的地方周围也低”,称为空间正相关,表明这种现象具有空间扩散的特性;如果呈现出“高的地方周围低,低的地方周围高”,则称为空间负相关,表明这种现象具有空间极化的特性;如果观测值在空间分布上呈现出随机性,表明空间相关性不明显,是一种随机分布的现象^[5]。

二、空间自相关分析方法

空间自相关分析可分以下几个步骤:首先建立空间权重矩阵,以明确研究对象在空间位置上的相互关系;然后进行空间自相关分析,包括全局空间自相关和局部空间自相关分析。全局空间自相关分析是判断研究区域空间自相关现象存在性的分析;局

部空间自相关分析是找出空间自相关现象存在的局部区域的分析^[5]。

1. 空间权重矩阵的建立:根据 Tobler 地理学第一定律,所有地理事物均存在关系,但距离较近的事物比距离较远的事物更有关系^[6]。所以,任何空间单元的属性均与其周围邻近的其他空间单元有关,其互相影响的程度与它们之间的距离相关。空间权重矩阵就是用来确定并描述空间关系、表达空间相邻关系的重要方式^[7]。

近年来,随着空间统计技术的广泛使用,国内外许多学者对空间权重矩阵展开了较为深入的研究。国内主要集中在“如何构建”的技术层面,如刘旭华和王劲峰^[8]利用 ArcInfo 生成的弧段属性表建立了要素之间的二进制邻接关系;魏晓峰和吴健平^[5]开发的自相关分析模块包括了 0、1 空间权重矩阵的构建;潘海燕等^[9]利用 ArcObject 实现了 Rooks 情形下 0、1 权重矩阵的构建。上述研究基本都只涉及一两种简单的权重矩阵的建立,可供选择性太少,不能满足更多复杂的空间关系建模。徐彬^[10]比较详细地总结了国外对空间权重矩阵的各种不同定义,不仅介绍了相应的构建方法,而且还定量分析了各种方法对 Moran's *I* 统计量的影响,但文中没有涉及如何约束要素的最少邻居数、如何处理空间均质性等一些复杂情况。国外具有代表性的研究成果主要有 Anselin^[11]开发的 GeoDa 软件和 ESRI 公司^[12]开发的 ArcMap 软件中的 Spatial Statistics Tools 模块。这两个软件提供了多种权重矩阵建立方法,可以满足大多数空间关系的模拟;其中 ArcMap 还允许用户将自定义的二维关系表转换为权重文件,用于模拟更为复杂的空间关系。

空间自相关的度量首先要确定和建立一个空间权重矩阵 w , 下面以 Spatial Statistics Tools 模块为例

来介绍几种权重矩阵的构建方法和它们所表示的意义。在 ArcGIS 9.3 软件中,使用 Generate Spatial Weights Matrix 工具来产生权重矩阵,其产生的权重矩阵基于以下两种策略^[13]:权重策略(Weighted Strategies)和二分类策略(Binary Strategies)。权重策略要素之间的权值取值在 0 至 1 之间,对应的方法包括反距离加权法(Inverse Distance)、反距离平方加权法(Inverse Distance Squared)和无差异带法(Zone of Indifference);二分类策略要素之间的权重定义为 0 或 1,对应的方法主要有固定距离区带法(Fixed Distance Band)和多边形相邻法(Polygon Contiguity)^[13]。

(1) 权重策略:使用权重策略构建权重矩阵,基于如下公式^[13]:

$$w_{ij} = 1/d^k \quad (i, j \text{ 之间的距离} < d) \quad (1)$$

$$w_{ij} = 0 \quad (\text{其他}) \quad (2)$$

反距离加权用于产生权重矩阵, $k=1$, 同时需要设定门槛距离 d (threshold distance) 和相邻数 N (number of neighbor)。 d 有三种赋值: Null、0 或自定义。若为自定义,则超过该距离之后,要素之间的交互权重为 0, 否则用公式(1)计算, 权重 w_{ij} 随门槛距离 d 的变化关系如图 1 所示;若为 Null, d 值将自动调整为保证每个要素至少有 1 个邻居的距离;若为 0, d 值将自动调整为全部范围内每个要素都为其他要素的邻居所对应的距离。邻居数 N 用来设定每个要素的最小邻居数, 根据图 1 可知, 当 i, j 之间的距离 $< d$ 时, 距离越小, 两要素之间的交互权重越大, 相互对方属性值产生的影响将越大, 即它们之间的相关性越大; 反之, 距离越大, 它们之间相关性就越小。当两要素之间的距离大到一定程度超过门槛距离 d 时, 交互权重为 0, 即它们之间没有相关性^[14]。

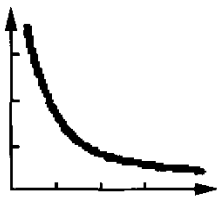


图 1 以反距离加权法计算权重随距离 (d) 变化示意图

距离的计算有两种方法可供选择, 即欧氏距离法和马氏距离法。

欧氏距离 (Euclidean distance) 的计算公式:

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3)$$

式中, (x_1, y_1) 表示 A 点的坐标, (x_2, y_2) 表示 B 点的坐标, A 到 B 的欧氏距离就等于两点之间的直线距离(图 2)。

欧氏距离是一种点到点的距离, 没有考虑各个

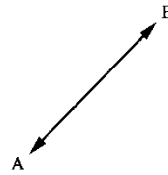


图 2 A 与 B 之间的欧氏距离

变量之间的相关性。为了解决这个问题, 1936 年, 印度著名的统计学家马哈拉诺比斯 (Mahalanobis) 提出了“基于多元统计的马氏距离 (Manhattan distance) 方法”, 它是一种可有效地计算两个未知样本集相似度的方法^[15-18], 其计算公式:

$$D = |x_1 - x_2| + |y_1 - y_2| \quad (4)$$

式中, (x_1, y_1) 表示 A 点的坐标, (x_2, y_2) 表示 B 点的坐标, A 到 B 的马氏距离等于两点之间水平距离与垂直距离之和(图 3)。假设从 A 到 B, 只能沿着东西方向或者南北方向移动, 这时候就只能计算马氏距离。所以, 当移动方向受限制或者只能沿着轴线移动的时候, 马氏距离比欧氏距离更实用。欧氏反距离法 (Inverse Euclidean Distance) 适用于连续型数据, 比如温度的变化; 当空间分析涉及到某些商店或其他城市公共设施的定位, 并且交通网络中可能遇到障碍物的时候, 使用马氏反距离法 (Inverse Manhattan Distance) 会更合适^[14]。

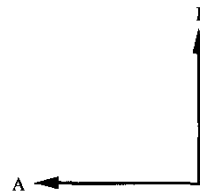


图 3 A 与 B 之间的马氏距离

反距离平方加权法的原理同反距离加权法相似, 两者的主要区别在于公式(1)中 $k=2$ 。同样地, 当 i, j 之间的距离 $< d$ 时, 两个空间单元之间距离越远, 空间权重越小, 代表其相关性越小。跟反距离加权法相比, 对于同一门槛距离 d , 反距离平方加权法对应的权重更小, 即图 1 中的曲线随着门槛距离 d 的增加斜率下降得更快更明显。所以对目标空间而言, 使用反距离平方加权法计算距离比使用反距离加权法需要更近距离的相邻空间才存在相关性^[19]。

无差异带法和上面两种方法的差异在于, 先设定门槛距离 d , 若要素之间的距离超过 d , 权重用公式(1)计算, 如图 4 中的曲线部分所示; 若要素之间的距离在门槛距离 d 范围内, 则权重为 1, 如图 4 中的直线部分所示。该方法实际上是考虑某一范围的

空间均质性:在均质范围之内,权重相同;超过该范围,则权重随距离衰减^[13]。

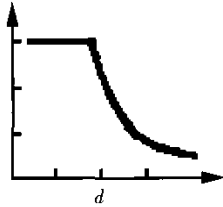


图4 以无差异带法计算权重随距离变化示意图

(2)二分类策略:二分类策略空间权重矩阵构建公式:

$$w_{ij}=1 (i, j \text{ 互为邻居}) \quad (5)$$

$$w_{ij}=0 (\text{其他}) \quad (6)$$

固定距离区带法是基于固定距离定义空间邻接的方法。以某一研究对象为中心,一定距离(门槛距离 d)为半径,将落入半径范围内的研究对象定义为相邻,权重为 1(图 5),否则定义为不相邻,权重为 0。基于距离的空间邻接定义可以选择权重字段,用以指定多边形参与计算的重心坐标。比如在分析某地人口分布状况时,可以指定其人口重心作为划分距离带的中心;若不指定权重字段,则采用多边形的几何中心。注意 d 不能为 0;若 d 设定为空值,则表示每个要素至少有 1 个邻居^[5,13]。

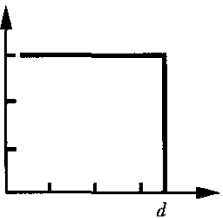


图5 固定距离区带法计算权重随距离($<d$)变化示意图

该方法认为,目标空间单元的属性值只受到与之相邻的指定距离范围(门槛距离 d)之内的邻近空间单元的影响,即它们之间有空间相关性;在此指定距离范围之外的其他邻近空间单元都不对目标空间单元产生影响。要评价一批在特定空间范围内数据的统计学性质时,可以选择这种方法^[14]。

多边形相邻法是基于多边形数据定义空间邻接的方法,看两个空间单元是否具有公共边或公共点,有公共边或公共点的空间单元互为相邻,前者定义为边相邻,后者定义为角相邻,两者之间权重取 1;否则取 0^[5,13]。当构建权重矩阵的研究属性本身具有传染性时,如对某种传染性疾病的空间自相关分析时,可以采用这种方法^[14]。但应当注意,该方法主

要针对多边形数据构建空间权重矩阵。

在建立空间权重矩阵的时候,除了要考虑选择何种构建方法和计算距离的方法之外,还要考虑另外一个问题:是否要对属性值进行标准化。

当目标属性值在空间的分布受到取样方法或统计方法的影响而造成偏倚的时候,应该进行标准化。行标化是指每一个样本的权重都除以相邻空间单元对应权重的合计以便进行标化。行标化计算权值经常采用固定距离邻接或共同边邻接,以减少某些特异值的影响,也可以用其他距离邻接方法进行标化。在一个基于距离邻接的全局空间区域的研究中,如果没有规定邻居的门槛距离,那么任一目标空间单元都具有同等数量的邻居,也就是说,除了自身以外的其他所有空间单元都是目标空间单元的邻居。在基于距离邻接的空间关系研究中,行标化将使所有属性值的权重比例位于 0 至 1 之间,形成一个相对的而不是绝对的空间权重矩阵^[14]。

综上所述,进行空间自相关分析的第一步是建立空间权重矩阵,建立空间权重矩阵的方法有许多,计算距离的时候又有欧氏距离和馬氏距离两种计算方法,并且每一次构建还应考虑是否对属性值进行标化处理。这样对于同一个分析对象可以构建出多种不同的空间权重矩阵。因此,在建立空间权重矩阵时,应根据研究内容的特点和实际需要,结合各种方法本身的适用条件和特点选择合理的方法和指标,以获得最佳的分析效果。

2. 全局空间自相关分析:用于分析描述某现象的整体分布状况,探测该现象在整个研究区域的空间聚集模式,使用单一的值来反映整个区域的自相关^[3,4,20]。衡量空间自相关的全局指标有全局 Moran's I 指数(Global Moran's I)、全局 Getis-Ord G 系数(Getis-Ord General G)和 Geary C 系数,此外还有 Cliff-Ord statistic、Join-count statistic 等不常用的指标^[3]。

(1)全局 Moran's I 指数:由 Moran 于 1948 年提出,它反映的是空间邻接或邻近的区域单元属性值的相似程度,其计算公式^[3,21]:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{j=1}^n (x_i - \bar{x})^2}, \forall j \neq i \quad (7)$$

式中, n 表示所有空间单元的个数, x 表示空间单元的属性值, i 和 j 代表不同的空间单元, \bar{x} 为所有空间单元属性值的平均值, w_{ij} 为空间权重矩阵,反映空间单元 i 和 j 的空间关系。

全局 Moran's I 指数的期望值:

$$E[I] = \frac{1}{n-1} \tag{8}$$

式中,随着样本数 n 的增大,期望值将逐渐趋近于 0。

Moran's I 取值范围介于 $-1 \sim +1$ 之间,当 $I >$ 期望值时,表示空间正相关,表明相邻空间单元具有相似的属性值,并且越接近于 1 表示空间单元之间关系越密切,性质越相似(高值聚集或低值聚集),代表整体呈聚集性分布;当 $I <$ 期望值时,表示空间负相关,表明相邻空间单元的属性值呈此消彼长的状态,并且越接近于 -1 表示空间单元之间差异越大或分布越不集中,代表整体呈离散型分布;当 I 接近期望值时,则表明空间单元之间不存在自相关现象。Moran's I 的结果在很大程度上取决于所选择的空

间权重矩阵^[5]。对于全局 Moran's I 指数,可以用标准化统计量 z_i 来检验空间自相关的显著性水平,其计算公式:

$$z_i = \frac{I - E[I]}{\sqrt{V[I]}} \tag{9}$$

$$V[I] = E[I^2] - E[I]^2 \tag{10}$$

式中, $E[I]$ 为全局 Moran's I 指数的期望值, $V[I]$ 为其理论方差。

z 值的显著性可以通过查标准正态分布表获知,如取显著性水平 $\alpha = 0.05$,那么 > 1.96 或者 < -1.96 的 z 值都被认为是存在统计学差异的,表明整体存在空间自相关现象。

(2)全局 Getis-Ord G 系数:由于 Moran's I 指数不能判断空间数据是高值聚集还是低值聚集,Getis 和 Ord 于 1992 年提出了全局 Getis-Ord G 系数。Getis-Ord G 系数一般采用距离相关的空间权重,要求空间单元的属性值为正,其计算公式^[22,23]:

$$G(d) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, \forall j \neq i \tag{11}$$

式中, n 表示所有空间单元的个数, d 为距离, x 表示空间单元的属性值, i 和 j 代表不同的空间单元代号, $w_{ij}(d)$ 为 i 和 j 之间的距离权。类似于全局 Moran's I 指数,也可以对全局 Getis-Ord G 系数进行标准化:

$$z_G = \frac{G(d) - E[G(d)]}{\sqrt{V[G(d)]}} \tag{12}$$

$$E[G(d)] = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(d)}{n(n-1)}, \forall j \neq i \tag{13}$$

$$V[G(d)] = E[G(d)^2] - E[G(d)]^2 \tag{14}$$

式中, $z_G > 0$ 表示存在高值聚集, $z_G < 0$ 表示存在低值聚集。

为了比较上述两种全局指数在探测空间聚集模式的能力差异,张松林和张昆^[3]设计了一个属性值随机分布的 11×11 研究区域,采用 ArcGIS 的 Spatial Statistics Tools 进行模拟计算,分别用全局 Moran's I 指数和全局 Getis-Ord G 系数进行分析,得出了以下结论:全局 Getis-Ord G 系数对于高值聚集和低值聚集的探测能力并不相同,检测出高值聚集的正确率远大于低值聚集的正确率。当研究范围内同时存在高值和低值聚集时, G 系数会受到聚集区域规模的影响。而且聚集区域靠近研究区域的边界时,会受到极大地影响。当高值聚集区域与低值聚集区域规模相当时, G 往往为正数,有明显判断为高值聚集的倾向。尽管 G 系数有区分高值和低值聚集的优势,但是判断一个区域是否存在空间聚集,尤其是估计聚集区域位于研究区域的边缘时,采用 Moran's I 指数统计的结果更为可靠,且随着空间聚集范围的扩展, Moran's I 指数的准确性会增加。如要区分是高值还是低值聚集,可以通过属性值的相对大小直接判断。

3. 局部空间自相关分析:计算每一个空间单元与邻近单元就某一属性的相关程度,检验研究区域内每个空间单元相对于整体范围而言其空间自相关是否足够显著,从而指出研究属性在空间的高高、高低、低低的分布^[4,20,24]。其中最常用的局部空间自相关指标是局部 Moran's I 指数 (Anselin's Local Moran's I) 和局部 Getis-Ord G 系数 (Getis-Ord G_i^*)。

(1)局部 Moran's I 指数:由于全局指标有时会掩盖局部状态的不稳定性^[4],因此在很多场合需要采用局部指标来探测空间自相关。如地震学家往往需要对地震数据进行局部空间自相关分析,基于地震发生是否存在空间格局的有关信息来研究地震的区域分布特性,以辅助地震预报^[25]。常用的局部 Moran's I 指数是由 Anselin^[11] 提出的,其计算公式^[26]:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{ij} (x_j - \bar{X}) \tag{15}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \tag{16}$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n-1} - \bar{X}^2 \tag{17}$$

式中, n 表示所有空间单元的个数, x_i 表示空间单元 i 的属性值, w 为空间矩阵, w_{ij} 代表空间单元 i 和 j 之间的影响程度。

同全局 Moran's I 指数一样, I_i 也可以用 z 值来进行标准化:

$$z(I_i) = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \tag{18}$$

$$E[I_i] = -\frac{\sum_{j=1, j \neq i}^n}{n-1} \quad (19)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2 \quad (20)$$

式中, $E[I_i]$ 和 $V[I_i]$ 为理论期望值和理论方差。

局部 Moran's I 指数反映了某区域周围相似属性值的空间聚集程度。 $I_i > 0$ 表示空间单元 i 与邻近单元的属性值相似(高值聚集或低值聚集), 即存在空间正相关现象; $I_i < 0$ 表示空间单元 i 与邻近单元的属性值不相似, 即存在空间负相关现象。局部 Moran's I 可以看作是全局 Moran's I 的各区域分量, 它反映了各区域的空间自相关现象对区域整体空间自相关的影响程度^[5]。

(2) 局部 Getis-Ord G 系数: 是一种基于距离权重矩阵的局部空间自相关指标, 能探测高值聚集和低值聚集^[27], 其计算公式^[28]:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{S \sqrt{\frac{[n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2]}{n-1}}} \quad (21)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (22)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (23)$$

式中, n 表示所有空间单元的个数, x_j 代表空间单元 j 的属性值, w_{ij} 代表空间权重。局部 Getis-Ord G 系数也可以用 z 值进行标准化:

$$z(G_i^*) = \frac{G_i^* - E[G_i^*]}{\sqrt{V[G_i^*]}} \quad (24)$$

式中, $G_i^* > 0$ 表示空间单元 i 的邻居的属性值高, $G_i^* < 0$ 表示空间单元 i 的邻居的属性值低^[4]。为了研究局部 Moran's I 指数和局部 Getis-Ord G 系数在空间自相关分析中的探测效果差异, 张松林和张昆^[4]设计了一个属性值随机分布的 11×11 的规则网格模拟地震带, 采用 ArcGIS 的 Spatial Statistics Tools 进行局部空间自相关模拟计算, 分别采用局部 Moran's I 指数和局部 Getis-Ord G 系数进行计算, 结果发现: 局部 Moran's I 指数和局部 Getis-Ord G 系数探测空间聚集的能力有显著地差异。局部 Getis-Ord G 系数能较准确地探测出聚集区域; 而局部 Moran's I 指数一般能大致探测出聚集区域的中心, 但对聚集范围的识别偏差较大, 探测出的范围 $<$ 实际范围。在同样的条件下, 局部 Moran's I 指数探测高值聚集的能

力逊色于对低值聚集的探测, 但其探测位于边、角聚集区域的能力相对强于其他位置。

三、空间自相关分析技术在医学及相关领域中的应用

空间自相关分析在医学及相关领域的应用可以深化研究者对疾病相关因素和人群疾病发生率空间分布规律的认识, 分析疾病发生的原因并对其进行动态监测, 同时合理配置医疗资源, 为人群疾病的预防和防治策略的制定提供依据。

1. 在医学领域的应用: 早在 1979 年, Glick^[29] 就将空间自相关分析引入到美国宾夕法尼亚州肿瘤地理分布变异的研究中。Sokal 和 Thomson^[23] 分别以 Moran's I 指数、局部 Getis-Ord G 系数和 Geary C 系数对南美洲某部落 43 个村庄村民们的基因多态性进行了局部空间自相关分析, 研究结果表明, 一个村庄相对于其相邻村庄的位置会影响局部空间自相关分析的结果和遗传基因的多态性。Schuurman 等^[30] 研究了肥胖聚集性和参加一般体育锻炼活动的空间聚集性以及它们与居民居住密度之间的关系。分别以全局 Moran's I 指数和局部 Moran's I 指数进行全局及局部空间自相关分析。研究结果提示, 没有足够证据证明肥胖的聚集性与一般体育锻炼活动的聚集性有关。Demirel 和 Erdogan^[31] 利用空间自相关分析技术中的全局和局部空间自相关, 对 1988—2006 年土耳其的皮肤利什曼病在各省区分布情况进行分析, 以期了解疾病的发展趋势、聚集状况以及病情特殊的省份。结果显示, 皮肤利什曼病在各个省份之间不是随机分布的, 具有显著的空间聚集性, 高发地区集中在东南地区, 并且发病有明显朝东南地区发展的趋势。除了疾病以外, 空间自相关分析还被应用到了公共卫生的相关研究中。Araujo 等^[32] 运用全局和局部 Moran's I 指数评价 1998—2003 年间巴西沙尔瓦多不同种族的谋杀发生率之间的空间分布, 并提示谋杀发生率存在空间聚集性, 究其原因主要是由于长期的社会不平等所致。

除了外国学者利用空间自相关技术在医学和公共卫生领域从事研究以外^[33-46]; 许多国内学者也对空间自相关技术进行了很好的理论与应用研究^[47-53]。张志杰等^[54] 采用 Moran's I 指数研究了湖沼地区血吸虫病传播惟一中间宿主——湖北钉螺的空间分布。赵飞等^[55] 采用全局空间自相关 Moran's I 指数、局部空间自相关 Getis-Ord G 系数和 SaTScan 统计软件, 系统深入地探测了中国湖区 5 省血吸虫病聚集区域。结果提示, 中国湖区 5 省沿江地带仍然是血

吸虫病空间聚集的主要区域。5 个主要聚集中心分布于长江流经湖北、湖南、江西、安徽 4 省的区域,聚集范围从上游至下游呈现逐渐缩小的趋势,尤以湖北、湖南省交界地带空间聚集性最高、范围最大。杨国静等^[56]收集了长江下游江苏、安徽、江西 3 省 20 世纪 80—90 年代代表年份的血吸虫病流行病学数据,应用空间自相关分析方法分析了血吸虫病疫情的空间分布规律,结果显示安徽省和江西省的血吸虫病患者总数和钉螺总面积以及江苏省的钉螺总面积均呈现出一定的空间聚集性。以上研究为评价滩地钉螺螺情、研究钉螺生态学和血吸虫病流行规律提供了理论基础,为血吸虫病的预防控制提供指导。蒋敏等^[57]采用全局空间自相关 Moran's *I* 指数对四川省 HIV/AIDS 的空间分布模式进行了研究,采用局部空间自相关 Getis-Ord *G* 系数统计量对聚集位置与属性进行探测,结果提示四川省 HIV/AIDS 以凉山彝族自治州为中心,向整个四川省西南部扩张并呈现聚集分布的特点。陆青云和殷勇^[58]利用地理信息系统(GIS)技术和空间统计分析方法对 2007 年江苏省南通地区食管癌死亡的空间分布规律进行了研究,研究最终探测出了南通地区食管癌的“热点区域”,为进一步探索食管癌的病因和危险因素、制定防治规划、评价干预措施等提供了线索和科学依据。

2. 在医学相关领域的应用:空间自相关分析在医学相关领域也有很好的应用^[59-62]。Trisalyn 等^[63]基于遥感影像采用全局 Moran's *I* 指数对成树和幼树进行了精确的分类。吕安民等^[64]采用全局 Moran's *I* 指数分析了 1982—1990 年和 1990—1998 年两个时期中国省级人口增长率的空間自相关性。吴拥政^[65]运用我国省级、地级经济增长数据和基于 Moran's *I* 指数的空间自相关方法对我国经济活动的空间关联性进行了实证分析。

四、应用前景和展望

数字地球战略的实施和空间信息基础设施的建设,使我们能够快速获取和掌握自然以及社会经济等大量数据信息,从而客观上促进了空间分析技术的发展及其在各个领域的应用^[66]。其中,空间自相关技术在医学领域已经得到了初步应用,并展示出良好的发展态势和应用前景。

然而从整体上看,空间自相关技术在医学领域的应用仍然受到一些条件的限制,有待进一步改进和完善。首先,空间自相关技术的有效平台——GIS 是在 20 世纪 60 年代出现并发展起来的,一些模型、方法和技术本身尚存在许多不完善之处,例如空

间权重的选择等。尽管空间自相关分析在医学领域中的应用已经初步显示出巨大的潜力,但 GIS 和空间自相关技术本身仍需要拓展其在各个领域的应用范围,并结合各个领域自身特点进一步完善其功能^[66,67]。其次,空间自相关分析方法在医学领域应用受到数据源的限制。由于我国医疗卫生保健体系的不完善,缺乏相应的空间理论、方法和技术手段,因此,长期以来积累的数据缺乏空间属性,限制了空间自相关技术在医学领域的展开和发展^[66]。

但是,GIS 可以处理海量数据,和空间数据分析结合,决定了其在分析疾病地理分布模式和社会、自然环境条件关系中的核心作用,作为疾病预防和干预决策支持系统的核心,其潜力是不可替代的^[68]。因此,随着空间自相关理论、技术和方法体系的不断完善,信息技术和 GIS 空间分析技术的应用和发展,空间自相关分析方法将会在未来的医学领域中作出更大贡献。

参 考 文 献

- [1] Barbujani G. Geographic patterns: how to identify them and why. *Hum Biol*, 2000, 72(1): 133-153.
- [2] Getis A, Ord K. The analysis of spatial association by use of distance statistics. *Geogr Analysis*, 1992, 24(3): 189-206.
- [3] Zhang SL, Zhang K. Comparison between general Moran's *I* index and Getis-Ord general *G* of spatial autocorrelation. *J Sun Yatsen University (Nat Sci Ed.)*, 2007, 46(4): 93-97. (in Chinese)
张松林,张昆. 全局空间自相关 Moran 指数和 *G* 系数对比研究. *中山大学学报(自然科学版)*, 2007, 46(4): 93-97.
- [4] Zhang SL, Zhang K. Contrast study on Moran and Getis-Ord indexes of local spatial autocorrelation. *J Geodesy Geodynamics*, 2007, 27(3): 31-34. (in Chinese)
张松林,张昆. 空间自相关局部指标 Moran 指数和 *G* 系数研究. *大地测量与地球动力学*, 2007, 27(3): 31-34.
- [5] Wei XF, Wu JP. The development and application of spatial autocorrelation based on ArcGIS. *Geomatics Spatial Information Technol*, 2005, 28(6): 77-80. (in Chinese)
魏晓峰,吴健平. 基于 ArcGIS 的空间自相关分析模块的开发与应用. *测绘与空间地理信息*, 2005, 28(6): 77-80.
- [6] Chen YG. Reconstructing the mathematical process of spatial autocorrelation based on Moran's statistics. *Geogr Res*, 2009, 28(6): 1449-1463. (in Chinese)
陈彦光. 基于 Moran 统计量的空间自相关理论发展和方法改进. *地理研究*, 2009, 28(6): 1449-1463.
- [7] Chen YS. Investigation of local indexes of spatial autocorrelation application. East China Normal University, 2009. (in Chinese)
陈雅淑. 局部空间自相关指标的适用性研究. 华东师范大学硕士学位论文, 2009.
- [8] Liu XH, Wang JF. Analysis and application on the specification methods of the spatial weight matrix. *Geo-information Sci*, 2002, 4(2): 38-44. (in Chinese)
刘旭华,王劲峰. 空间权重矩阵的生成方法分析与实验. *地球信息科学*, 2002, 4(2): 38-44.
- [9] Pan HY, Cheng PG, Xiao GR, et al. The constitution and realization of spatial weight matrix based on ArcObjects. *Sci*

- Survey Map, 2007, 32(6):130-131. (in Chinese)
潘海燕, 程明根, 肖根如, 等. 基于 ArcObjects 的空间权重矩阵的建立与实现. 测绘科学, 2007, 32(6):130-131.
- [10] Xu B. Simulation analysis of the influence of spatial weight matrix on Moran's I . Nanjing Normal University, 2007. (in Chinese)
徐彬. 空间权重矩阵对 Moran's I 指数影响的模拟分析. 南京师范大学硕士论文, 2007.
- [11] Anselin L. Exploring spatial data with GeoDaTM [EB/OL]. (2004-05-05) [2010-05-30]. <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>.
- [12] ESRI. ArcGIS desktop help [EB/OL]. (2009-04-24) [2010-05-30]. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2214&pid=2213&topicname=An_overview_of_the_spatial_statistics_toolbox.
- [13] Wang HL, Hu WP, Wu C. Analyzing the effect of spatial weighted matrix on spatial autocorrelation-taking hunan's income gap between urban and rural areas as a case. J South Chin Normal University (Nat Sci Ed.), 2010(1):110-115. (in Chinese)
王红亮, 胡伟平, 吴驰. 空间权重矩阵对空间自相关的影响分析——以湖南省城乡收入差距为例. 华南师范大学学报(自然科学版), 2010(1):110-115.
- [14] ESRI. ArcGIS desktop help [EB/OL]. (2009-04-24) [2010-05-20]. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Modeling_spatial_relationships.
- [15] Grubestic TH. Efficiency in broadband service provision: a spatial analysis. Telecommunications Policy, 2010, 34(3):117-131.
- [16] Tempel M, Filzmoser P, Reimann CC. Cluster analysis applied to regional geochemical data: problems and possibilities. Applied Geochemistry, 2008, 23(8):2198-2213.
- [17] Theodoridis S, Koutroumbas K. Clustering: basic concepts. Academic Press, 2006.
- [18] Deza E, Deza M. Distances and similarities in data analysis. Elsevier, 2006.
- [19] Wartenberg D, Uchirin C, Coogan P. Estimating exposure using kriging: a simulation study. Environ Health Perspect, 1991, 94:75-82.
- [20] Qi XP, Zhou MG, Hu YS, et al. Spatial hotspot exploration on digestive tract cancer mortality with geographic information system. Geogr Res, 2010, 29(1):181-187. (in Chinese)
戚晓鹏, 周耕群, 胡以松, 等. 应用地理信息系统探测消化道癌症死亡率空间聚集性. 地理研究, 2010, 29(1):181-187.
- [21] ESRI. ArcGIS desktop help [EB/OL]. (2009-01-28) [2010-05-20]. [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_spatial_autocorrelation;Moran's_I\(Spatial_Statistics\)_works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_spatial_autocorrelation;Moran's_I(Spatial_Statistics)_works).
- [22] ESRI. ArcGIS Desktop Help [EB/OL]. (2009-01-28) [2010-05-20]. [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_High/Low_Clustering;Getis-Ord_General_G\(Spatial_Statistics\)_works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_High/Low_Clustering;Getis-Ord_General_G(Spatial_Statistics)_works).
- [23] Sokal RR, Thomson BA. Population structure inferred by local spatial autocorrelation: an example from an Amerindian tribal population. Am J Phys Anthropol, 2006, 129(1):121-131.
- [24] Flahaut B, Mouchart M, San ME, et al. The local spatial autocorrelation and the kernel method for identifying black zones. A comparative approach. Accid Anal Prev, 2003, 35(6):991-1004.
- [25] Wang JF, Sun YJ, Han WG, et al. Introduction of spatial analysis. Geomatics World, 2004, 2(5):6-10. (in Chinese)
王劲峰, 孙英君, 韩卫国, 等. 空间分析引论. 地理信息世界, 2004, 2(5):6-10.
- [26] ESRI. ArcGIS desktop help [EB/OL]. (2009-01-28) [2010-05-20]. [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_Cluster_and_outlier_analysis;Anselin_local_Moran's_I\(spatial_statistics\)_works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_Cluster_and_outlier_analysis;Anselin_local_Moran's_I(spatial_statistics)_works).
- [27] Oed JK, Getis A. Testing for local spatial autocorrelation in the presence of global autocorrelation. J Regional Sci, 2001, 41(3):411-432.
- [28] ESRI. ArcGIS desktop help [EB/OL]. (2009-01-28) [2010-05-20]. [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_hot_spot_analysis;Getis-Ord_G\(Spatial_Statistics\)_works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How_hot_spot_analysis;Getis-Ord_G(Spatial_Statistics)_works).
- [29] Glick B. The spatial autocorrelation of cancer mortality. Soc Sci Med Med Geogr, 1979, 13 D(2):123-130.
- [30] Schuurman N, Peters PA, Oliver LN. Are obesity and physical activity clustered? A spatial analysis linked to residential density. Obesity (Silver Spring), 2009, 17(12):2202-2209.
- [31] Demirel R, Erdogan S. Determination of high risk regions of Cutaneous leishmaniasis in Turkey using spatial analysis. Turkiye Parazitoloj Derg, 2009, 33(1):8-14.
- [32] Araujo EM, Costa MC, Oliveira NF, et al. Spatial distribution of mortality by homicide and social inequalities according to race/skin color in an intra-urban Brazilian space. Rev Bras Epidemiol, 2010, 13(4):549-560.
- [33] Queiroz JW, Dias GH, Nobre ML, et al. Geographic information systems and applied spatial statistics are efficient tools to study Hansen's disease (Leprosy) and to determine areas of greater risk of disease. American J Tropical Med Hygiene, 2010, 82(2):306-314.
- [34] Erdogan S. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. J Safety Res, 2009, 40(5):341-351.
- [35] Beaudreau F, Bjorkman C, Alenius S, et al. Spatial patterns of bovine corona virus and bovine respiratory syncytial virus in the Swedish beef cattle population. Acta Vet Scand, 2010, 52(1):33.
- [36] Loobuyck M, Frossling J, Lindberg A, et al. Seroprevalence and spatial distribution of Neospora caninum in a population of beef cattle. Prev Vet Med, 2009, 92(1-2):116-122.
- [37] Hinman SE, Blackburn JK, Curtis A. Spatial and temporal structure of typhoid outbreaks in Washington, D.C., 1906-1909: evaluating local clustering with the G_i^* statistic. Int J Health Geogr, 2006, 5:13.
- [38] Rainey JJ, Omenah D, Sumba PO, et al. Spatial clustering of endemic Burkitt's lymphoma in high-risk regions of Kenya. Int J Cancer, 2007, 120(1):121-127.
- [39] Chowell G, Rivas AL, Smith SD, et al. Identification of case clusters and counties with high infective connectivity in the 2001 epidemic of foot-and-mouth disease in Uruguay. Am J Vet Res, 2006, 67(1):102-113.
- [40] Alvarez G, Lara F, Harlow SD, et al. Infant mortality and urban marginalization: a spatial analysis of their relationship in a medium-sized city in northwest Mexico. Rev Panam Salud Publica, 2009, 26(1):31-38.
- [41] Neto OL, Barros MB, Martelli CM, et al. Differential patterns of neonatal and post-neonatal mortality rates in Goiania, Brazil, 1992-1996: use of spatial analysis to identify high-risk areas. Cad Saude Publica, 2001, 17(5):1241-1250.
- [42] Depina MF, Alves SM, Barbosa M, et al. Hip fractures cluster in space: an epidemiological analysis in Portugal. Osteoporos Int, 2008, 19(12):1797-1804.
- [43] Sabel CE, Wilson JG, Kingham S, et al. Spatial implications of covariate adjustment on patterns of risk: respiratory hospital admissions in Christchurch, New Zealand. Soc Sci Med, 2007, 65(1):43-59.
- [44] Portnov BA, Dubnov J, Barchana M. On ecological fallacy, assessment errors stemming from misguided variable selection,

- and the effect of aggregation on the outcome of epidemiological study. *J Expo Sci Environ Epidemiol*, 2007, 17(1): 106-121.
- [45] Chiang CT, Hwang YH, Su CC, et al. Elucidating the underlying causes of oral cancer through spatial clustering in high-risk areas of Taiwan with a distinct gender ratio of incidence. *Geospat Health*, 2010, 4(2): 230-242.
- [46] Mitra R, Buliung RN, Faulkner GE. Spatial clustering and the temporal mobility of walking school trips in the Greater Toronto Area, Canada. *Health Place*, 2010, 16(4): 646-655.
- [47] Tsai PJ, Lin ML, Chu CM, et al. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health*, 2009, 9: 464-476.
- [48] Shen ZZ, Feng ZJ, Ma JQ, et al. Pilot study and spatial clustering of pulmonary tuberculosis in sichuan province in 2004. *Modern Prev Med*, 2008, 35(8): 1412-1413. (in Chinese)
沈卓之, 冯子健, 马家奇, 等. 四川省2004年肺结核流行特征及空间聚集性分析. *现代预防医学*, 2008, 35(8): 1412-1413.
- [49] Ba YE, Yao ZY, Zhang LJ, et al. Relationship between spatial aggregation of *Spermophilus dauricus* and animal plague in hilly region. *Chin J Control Endemic Dis*, 2004(3): 134. (in Chinese)
巴雅尔, 姚忠友, 张丽娟, 等. 低山丘陵区达乌尔黄鼠的空间聚集性及与动物鼠疫的关系. *中国地方病防治杂志*, 2004(3): 134.
- [50] Pei XQ, Feng ZJ, Li XS, et al. Spatial autocorrelation analysis on brucellosis in Inner Mongolia Autonomous Region from 2004 to 2005. *Chin J Dis Control Prev*, 2009, 13(6): 654-658. (in Chinese)
裴小琴, 冯子健, 李晓松, 等. 2004-2005年内蒙古自治区布鲁氏菌病发病的空间自相关分析. *中华疾病控制杂志*, 2009, 13(6): 654-658.
- [51] Chen BW, Xu BY, Li DY, et al. Application of local indicators of spatial autocorrelation to analysis of disease cluster 48. *Chin Pub Health*, 2006, 22(9): 1146-1147. (in Chinese)
陈炳为, 许碧云, 李德云, 等. 应用区域型空间自相关系数分析疾病的聚集性. *中国公共卫生*, 2006, 22(9): 1146-1147.
- [52] Tang XY, Huang TR, Zhu XD, et al. Using spatial autocorrelation analysis to study spatial heterogeneity of liver cancer in Guangxi. *Chin J Epidemiol*, 2009, 30(2): 167-170. (in Chinese)
唐咸艳, 黄天王, 朱小东, 等. 应用空间自相关分析研究广西壮族自治区肝癌的空间异质性分布特征. *中华流行病学杂志*, 2009, 30(2): 167-170.
- [53] Guo JT, Zhou YB, Zhang ZJ, et al. Simple sequence repeat variation and small-scale spatial autocorrelation analysis on smooth-shell populations of *Oncomelania hupensis* in Sichuan province. *Chin J Epidemiol*, 2009, 30(5): 497-501. (in Chinese)
郭俊涛, 周艺彪, 张志杰, 等. 四川壳钉螺微卫星遗传变异的小尺度空间自相关分析. *中华流行病学杂志*, 2009, 30(5): 497-501.
- [54] Zhang ZJ, Peng WX, Zhou YB, et al. Spatial autocorrelation analysis of the small-scale distribution of *Oncomelania hupensis* in marshland and lake regions. *Chin J Schistosomiasis Control*, 2007(6): 418-423. (in Chinese)
张志杰, 彭文祥, 周艺彪, 等. 湖沼地区湖北钉螺小尺度分布的空间自相关分析. *中国血吸虫病防治杂志*, 2007(6): 418-423.
- [55] Zhao F, Zhu R, Zhang LJ, et al. Integrated detection and analysis on the clusters of schistosomiasis based on geographic information system. *Chin J Epidemiol*, 2010, 31(11): 1272-1275. (in Chinese)
赵飞, 朱蓉, 张丽娟, 等. 中国湖区五省血吸虫病聚集区域地理信息系统综合探测分析. *中华流行病学杂志*, 2010, 31(11): 1272-1275.
- [56] Yang GJ, Zhou XN, Wang TP, et al. Spatial autocorrelation analysis on schistosomiasis cases and *Oncomelania* snails in three provinces of the lower reach of Yangtze River. *Chin J Parasitol Parasitic Dis*, 2002, 20(1): 6-9. (in Chinese)
杨国静, 周晓农, 汪天平, 等. 安徽、江西及江苏三省血吸虫病患者与钉螺分布的空间自相关分析. *中国寄生虫学与寄生虫病杂志*, 2002, 20(1): 6-9.
- [57] Jiang M, Li XS, Feng ZJ, et al. The spatial autocorrelation of HIV/AIDS in Sichuan. *Modern Prev med*, 2008, 35(22): 4329-4331. (in Chinese)
蒋敏, 李晓松, 冯子健, 等. 四川省HIV/AIDS空间自相关分析. *现代预防医学*, 2008, 35(22): 4329-4331.
- [58] Lu QY, Yin Y. Research of GIS-based spatial distribution of death of esophageal cancer in Nantong region. *Med J Communications*, 2010, 24(1): 101-102, 104. (in Chinese)
陆青云, 殷勇. 基于GIS的南通地区食管癌死亡空间分布研究. *交通医学*, 2010, 24(1): 101-102, 104.
- [59] Yan FH, Jin YQ. Spatial auto-correlation of characteristic parameters in remotely sensing image using optimal multi-scale Getis statistic. *J Image Graphics*, 2006, 11(2): 191-196. (in Chinese)
颜锋华, 金亚秋. 尺度分布的Getis统计对遥感图像特征参量空间自相关性的研究. *中国图象图形学报*, 2006, 11(2): 191-196.
- [60] Zhao ZW, Wang WC, Zhang YQ, et al. The spatial research on average health consumption in China during fifteen period. *Chongqing Med J*, 2008, 37(17): 1950-1952. (in Chinese)
赵增炜, 王文昌, 张彦琦, 等. “十五”期间中国地区人均卫生消费的空间分析. *重庆医学*, 2008, 37(17): 1950-1952.
- [61] Xu LH, Yue WZ. Spatial patterns of population in Shanghai based on spatial-statistic. *Resources Environment Yangtze Basin*, 2009, 18(3): 222-228. (in Chinese)
徐丽华, 岳文泽. 上海市人口分布格局动态变化的空间统计研究. *长江流域资源与环境*, 2009, 18(3): 222-228.
- [62] Song TQ, Peng WX, Zeng FP, et al. A spatial heterogeneity of surface soil moisture content in dry season in Mulun National Natural Reserve in Karst area. *Chin J Applied Ecol*, 2009, 20(1): 98-104. (in Chinese)
宋同清, 彭晚霞, 曾毓平, 等. 喀斯特木论自然保护区旱季土壤水分的空间异质性. *应用生态学报*, 2009, 20(1): 98-104.
- [63] Trisalyn N, Barry B, Michael AW. Techniques for accuracy assessment of tree locations extracted from remotely sensed imagery. *J Environmental Management*, 2005, 74(3): 265-271.
- [64] Lv AM, Li CM, Lin ZJ, et al. Population growth rate and its spatial association by providence in China. *Acta Geogr Sinica*, 2002, 57(2): 143-150. (in Chinese)
吕安民, 李成名, 林宗坚, 等. 中国省级人口增长率及其空间关联分析. *地理学报*, 2002, 57(2): 143-150.
- [65] Wu YZ. An empirical analysis on longitudinal change stages and horizontal difference of spatial relatedness of regional economy growth: based on province-level and prefecture-level city data and spatial statistical Moran's I method. *Economic Res Guide*, 2010(4): 87-89. (in Chinese)
吴拥政. 区域经济增长空间关联性的纵向变化与横向差异实证分析——基于省级、地级数据和空间统计Moran's I指数方法. *经济研究导刊*, 2010(4): 87-89.
- [66] Wu JL, Wang JF, Zheng XY, et al. A review on application of spatial data analysis technology in public health. *Progress Geogr*, 2003, 22(3): 219-227. (in Chinese)
武继磊, 王劲峰, 郑晓暎, 等. 空间数据分析技术在公共卫生领域的应用. *地理科学进展*, 2003, 22(3): 219-227.
- [67] Anselin L, Getis A. Spatial statistical analysis and geographic information systems. *Annals Regional Sci*, 1992, 26(1): 19-33.
- [68] Rushton G, Elmes G, McMaster R. Considerations for improving geographic information system research in public health. *Urisa J*, 2000, 12(2): 31-49.

(收稿日期: 2010-12-02)

(本文编辑: 万玉立)