



# 一种空间权重矩阵的优化方法

翟卫欣<sup>1</sup> 程承旗<sup>2</sup>

1 北京大学地球与空间科学学院,北京,100871

2 北京大学工学院,北京,100871

**摘要:**在地理学空间自相关的分析中,权重矩阵对整个分析结果有着较大影响。常见的权重矩阵,例如车矩阵、皇后矩阵、距离权重矩阵和  $k$ -邻近矩阵,都有各自的优势和缺点。提出了一种基于长度面积比例的空间权重矩阵(ratio of length and area, RLA),并以近年来危害最大的几种传染病之一——病毒性肝炎在中国大陆各省份的发病率为例进行了实验分析。实验结果表明,RLA 矩阵能够很好地实现空间权重矩阵的基本功能,是常见的车矩阵的一种更为广义的定义,并且可以更加自由地实现空间自相关的分析。利用本空间权重矩阵能够更好地模拟不同地理单元之间的邻接关系,为流行病的预防提供支持。

**关键词:**空间自相关;空间权重矩阵;病毒性肝炎

中图法分类号:P208

文献标志码:A

空间统计学作为一种处理和分析空间数据的计算方法,已经被广泛地应用到与空间信息操作相关的诸多领域,例如地质、防疫、防灾、土壤、农业、环境、经济、公共卫生等<sup>[1-5]</sup>。与传染病相关的研究在国内<sup>[6-8]</sup>和国外<sup>[9-12]</sup>已成为地理学研究的热点之一。空间统计的方法能有效处理与疾病相关的信息数据,分析不同区域传染性疾病分布和变化的联系,探索传染性疾病病因及各种影响因素<sup>[13-14]</sup>,被广泛应用。由于传染性疾病的发生和流行与空间环境因素密切相关,疫区分布具有明显的地理特征,因此,找出传染性疾病与空间环境因素之间的内在联系,是预测流行趋势并进行有效控制的关键<sup>[15]</sup>。常用的空间自相关分析的方法均是基于空间权重矩阵的,一个符合真实地理情况的权重矩阵能够更好地帮助实现空间统计,得到事物空间分布的内在规律<sup>[16]</sup>。空间权重矩阵的判断一般为二值型,即认为两个空间单元相邻或者不相邻。对于具体的空间实体,将其本身的空间属性(例如面积)信息融入空间权重的判断方法可以有效增强对空间单元之间邻近关系判断的适应和表达能力<sup>[17-18]</sup>。

传统的空间权重矩阵的构造有3类方法:①

按照面域之间是否有邻接关系的邻接法;②基于面域之间距离的距离权重法;③ $k$ -近邻方法。相应地,空间权重矩阵也分为三类。每一类方法均可以以一个  $N \times N$  的空间权重矩阵来表示,在该权重矩阵中每一个元素的值  $\omega_{ij}$  为 0 或者 1,0 代表第  $i$  个和第  $j$  个空间单元不邻接,1 代表这两个空间单元邻接<sup>[19]</sup>。

第一种方法的空间权重矩阵第  $i$  行第  $j$  列的元素值表示为:

$$\omega_{ij} = \begin{cases} 1, & A_i \text{ 与 } A_j \text{ 相邻} \\ 0, & \text{其他情况} \end{cases} \quad (1)$$

式中,  $A_i$  与  $A_j$  表示两个面域(下同)。这种权重矩阵的邻接性的度量可分为车矩阵和皇后矩阵两种,定义是:如果  $A_i$  与  $A_j$  有重合的边(对于皇后矩阵,除了边之外,重合的点也考虑),则认为两者相邻接,矩阵中  $\omega_{ij}$  元素和  $\omega_{ji}$  元素的值均为 1。

第二种方法的空间权重矩阵第  $i$  行第  $i$  列的元素值表示为:

$$\omega_{ii} = \begin{cases} 1, & d(A_i, A_i) < \text{距离阈值} \\ 0, & \text{其他情况} \end{cases} \quad (2)$$

式中,  $d(A_i, A_i)$  代表  $A_i$  与  $A_i$  与之间的距离。这种权重矩阵对距离函数的定义和距离阈值的选择

有一定的要求,常见的距离函数计算方法是中心距离法,即选取面状单元的空间中心来计算距离。

第三种方法的空间权重矩阵第  $i$  行第  $j$  列的元素值表示为:

$$\omega_{ij} = \begin{cases} 1, (A_i, A_j \text{ 属于 } k\text{-邻近}) < \text{距离阈值} \\ 0, \text{ 其他情况} \end{cases} \quad (3)$$

$k$ -邻近方法是数据挖掘中一种常见方法,在空间统计学方面主要是代表选取每一个单元最近的  $k$  个单元作为其邻接单元,相应的空间权重矩阵中的元素为 1,而其他的相应元素则为 0。

本文提出了一种新的计算空间权重矩阵(ratio of length and area, RLA)的度量方法。这种方法相比传统度量方法能够更精确地刻画不同空间实体之间的邻接关系。基本思想是,选取空间单元的面积作为其自身综合影响力的标度,选取任意两个空间单元的相重合的边长作为其邻接的绝对标度。如果邻接的绝对标度与空间单元自身影响力的标度之比值大于某一阈值,说明邻接单元对该空间单元的影响不容忽视,认为两者相连。这种对于空间相邻区域影响的度量与边长正相关、与面积负相关,能够刻画出相邻的两个空间单元的接壤部分对于各自空间单元的影响。本文以我国近年病毒性肝炎发病情况为例,进行全局和局部的病毒性肝炎疾病空间自相关对比分析。

本文选取 GeoDa 作为跨平台开源软件,可以用于空间统计的各个方面<sup>[20]</sup>。GeoDa 软件采用随机模拟方法对 Moran's I 值进行显著性检验<sup>[21]</sup>。判断显著性的参数  $p$  值大小反映了观测数据为随机分布的可能性。一般认为,当  $p \leq 0.05$  时,观测数据为随机分布的可能性不到 5%,即区域单元的观测值存在显著的空间自相关性。

## 1 空间自相关分析

### 1.1 RLA 空间权重矩阵构造方法

本文提出的 RLA 空间权重矩阵构造方法是一种要求更为严格的有一定区别的度量方法,其空间权重矩阵元素值表示为:

$$\omega_{ij} = \begin{cases} 1, L_{ij}/S_i > \text{距离阈值} \\ 0, \text{ 其他情况} \end{cases} \quad (4)$$

式中,  $L_{ij}$  分别表示  $A_i$  与  $A_j$  邻接部分线段的长度;  $S_i$  表示  $A_i$  的面积。相对于传统的权重矩阵,避免了某两个空间单元虽然有一定的重合边,但是这个重合边的长度仅仅是整个空间单元的一小

部分而导致两者实际上的联系很小甚至可以忽略的“伪邻接”。有效去除“伪邻接”能够突出重要的空间邻接性质。

### 1.2 Moran's I 统计

#### 1.2.1 全局 Moran's I 统计

全局空间相关性分析从宏观层面出发,通过比较区域汇总的均值与每个对象的值,得出整个研究区域的相关性数值即 Moran's I 值,用于判断该传染病分布在整个研究区域中有无聚集性。设研究区域中有  $n$  个面域单元,第  $i$  个单元上的观测值为  $x_i$ ,观测变量在  $n$  个单元中的均值记为  $\bar{x}$ ,  $\omega_{ji}$  为空间权重矩阵,则 Moran's I 定义为<sup>[22]</sup>:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (x_i - \bar{x})(x_j - \bar{x}) \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

式中,空间权重  $\omega_{ji}$  和  $(x_i - \bar{x})(x_j - \bar{x})$  的乘积限定了  $(x_i - \bar{x})(x_j - \bar{x})$  只对相邻的单元有效。因此,  $I$  值的大小取决于  $i$  和  $j$  单元的变量值对于均值的偏离符号。不同的矩阵可定义不同的邻域单元,因而会产生不同的相关性。Moran's I 的取值一般在  $[-1, 1]$  之间,小于 0 表示负相关,等于 0 表示不相关,大于 0 表示正相关。

#### 1.2.2 局部 Moran's I 统计

局部自相关指标(local indicators of spatial autocorrelation, LISA)从微观出发,对研究区域内每个对象的取值进行分析,得出每个对象在领域空间的分布状态。局部 Moran's I 值定义为<sup>[21]</sup>:

$$I_i = \frac{n(x_i - \bar{x}) \sum_{j=1}^n \omega_{ij} (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \quad (6)$$

式中,  $n$  是参与分析的空间单元数;  $x_i$  和  $x_j$  分别表示某现象(或某属性特征)  $x$  在空间单元  $i$  和  $j$  上的观测值;  $\omega_{ij}$  是空间权重。

## 2 RLA 空间权重矩阵与空间邻接矩阵对比实验

### 2.1 实验数据

本文以全国 1:100 万省级行政区划数字地图作为底图,将病毒性肝炎发病数据进行关联。主要数据有:①2003~2011 年全国 31 个省区病毒性肝炎发病率、死亡率统计数据资料;②基础地图数据。由于要进行邻近性分析,所以未考虑香港、澳门、台湾、海南区域。

分析过程中运用了多种不同权重矩阵构造方式,当选择 RLA 矩阵时,特别注意阈值大小、相邻多边形数量与空间自相关统计量 Moran's I 值之间的关系。

## 2.2 实验结果

针对 2011 年中国大陆大部分省区病毒性肝炎发病率数据,分别使用空间邻接矩阵和 RLA 矩阵计算全局和局部 Moran's I 值。

1) 空间邻接矩阵。实验共选取了一阶车矩阵、二阶车矩阵两种最常见的邻接矩阵来求得相应的 Moran's I 值和  $p$  值。由于中国的行政区划不存在某两个省份只有一个点相重合的情况,因此皇后矩阵与车矩阵是相同的。一阶车矩阵的 Moran's I 值大于二阶车矩阵的 Moran's I 值,而且这两种情况下的  $p$  值均是小于 0.01 的,因而这两个结果均表明发病率存在显著的空间自相关性。

2) RLA 矩阵。基于长度面积比的权重矩阵中,阈值  $t$  直接决定了对象的邻域范围,不同阈值定义了不同的权重矩阵。

当  $1.2 \times 10^{-3} < t < 1.5 \times 10^{-3}$  时, Moran's I 值快速升高,  $p$  值快速降低; 当  $1.5 \times 10^{-3} < t < 2.0 \times 10^{-3}$  时, Moran's I 值在保持在较高水平的情况下缓慢下降,  $p$  值在一直处于较低的水平中,之后两者分别在下降和上升,如图 1 所示。由上述分析可知,当  $1.5 \times 10^{-3} < t < 2.0 \times 10^{-3}$  时,所建立的矩阵能较好地反映 2011 年病毒性肝炎发病率的空间分布自相关情况。

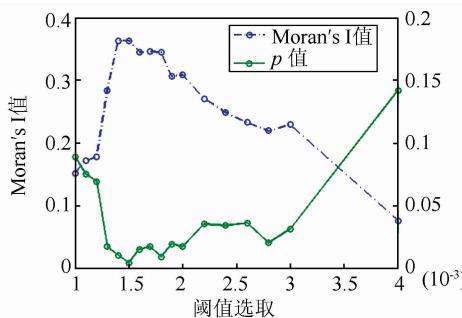


图 1 RLA 的阈值选取

Fig. 1 Determination of RLA Threshold

3) RLA 矩阵与 Moran's I 值的关系。不同权重矩阵本质上定义了每个区域的相邻区域的个数,从而决定了 Moran's I 值的大小。为了揭示空间自相关系数与相邻区域个数的关系,在不同距离下,实验统计与每个区域重心距离小于距离阈值  $t$  的区域个数,如图 2 所示。根据定义可知阈值  $t$  选取越大,对于邻接条件的要求越高,每个

单元所对应的相邻区域的个数越小。

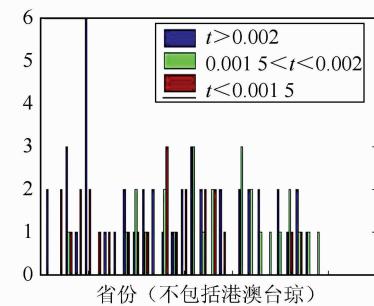


图 2 相邻区域的个数统计图

Fig. 2 Number of Adjacent Areas

经过统计,在阈值大于  $2.0 \times 10^{-3}$  时,符合条件的邻接总数是 44 个,权重矩阵所定义的邻域条件内,平均一个区域有约 1.5 个相邻区域,整体数据没有显著的空间自相关( $p=0.1980$ )。当选取阈值为  $2.0 \times 10^{-3}$  时,整体上又增加了 27 个相邻接的区域,此时整体数据呈现了显著的空间自相关( $p=0.0040$ );如果进一步放低对于阈值选取的要求,会增加新的 25 个空间邻接关系,但是空间自相关特征就如之前的显著( $p=0.0890$ )。可以看出,虽然同样是增加了邻接的数量,但并非使结果的显著性向同一方向变化。

中国大陆的行政区划界限不是按照规则的地理排列方式,且各省区形状不规则,有可能出现某两个不同的行政区划只有较短的边界相重合,邻接关系并不明显。实验数据表明,在运用 RLA 矩阵时,许多空间上虽然接壤但是实际关联并不强的关联不计入空间权重矩阵。例如,山东和安徽两个省区的界限只有很短的重合长度,这个长度占据各自省界的较小部分,所以尽管两省区有重合边界,但是实际上关联并不大,将这种关联等同于其他的相邻关联(例如山东和河南)显然不合理。

本文选取常见的车矩阵和本文提出的 RLA 空间权重矩阵(取  $t=1.5 \times 10^{-3}$ )进行对比实验。对于全局 Moran's I 值,结果显示,用常规的车矩阵计算结果为 0.17,用 RLA 空间权重矩阵的计算结果为 0.23,两种计算结果略有不同,均表现出了流行病传播与空间聚集度的正向相关性。

分析 30 个区域的局部 Moran's I 值可以发现,运用两种不同的方法得到的分布和显著程度同样均有所差异。本文提出的 RLA 空间权重矩阵是一种以邻接关系为基础、要求相对更为严格的矩阵,所以其每个空间单元对应的邻接数量一定小于或等于车矩阵。RLA 矩阵保留了更为有

意义的邻接关系而去除了相对无效的“伪邻接”，使得权重矩阵能够更加真实地反应空间的邻接关系，而且车矩阵只是 RLA 空间权重矩阵的一种特例(取  $t=0$ )。西部的新疆、西藏、青海、陕西等省区的局部 Moran's I 值表明其聚集方式发生了显著的变化，而其他地区变化不大。同时，两类图像的各个省区的置信程度( $p$  值)也有所变化。发生这种变化的原因主要在于权重的变化导致原来的  $I$  值和新的  $I$  值有一定的偏差，相对的“高”和“低”的程度也有了变化，因而会有了“高-高”、

“低-低”、“高-低”、“低-高”模式的转化。

两种方法分析的结果如图 3 和图 4 所示，其中图 3(a)、图 4(a)均为车矩阵的结果，图 3(b)、图 4(b)均为 RLA 空间权重矩阵的结果。RLA 空间权重矩阵的结果表明我国新疆、西藏、宁夏、陕西省区均是低发病区而被高发病区所围绕，而江苏则是自身发病率不高周围发病率也不高的“低-低”模式。新疆、西藏、宁夏和浙江省区都在  $p=0.01$  的条件下置信，陕西是在  $p=0.05$  的条件下置信。

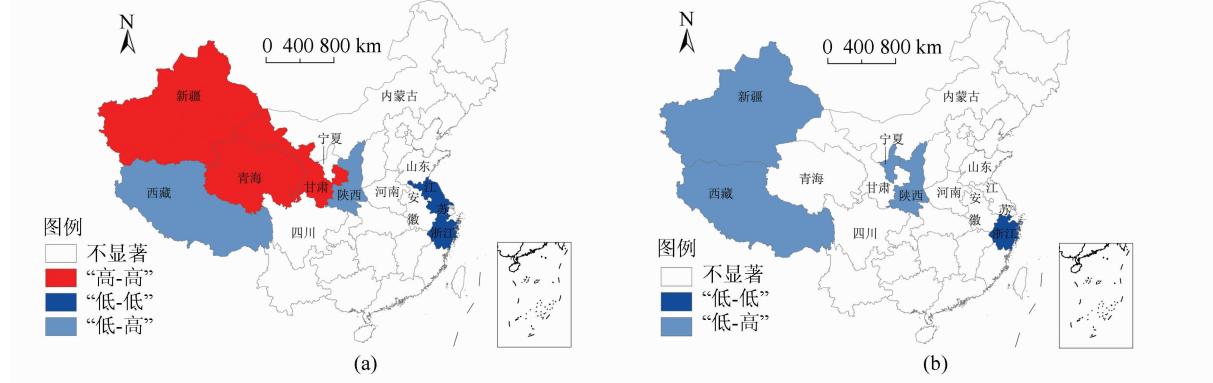


图 3 两种方法的聚类分析图

Fig. 3 Cluster Map of Two Methods

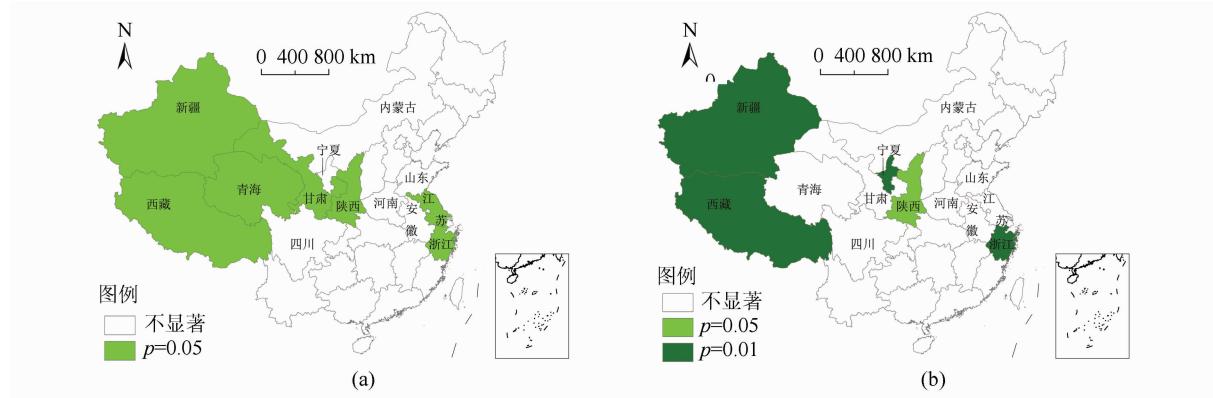


图 4 显著性分析图

Fig. 4 Significant Map of Two Methods

本文所使用的权重矩阵与常见的权重矩阵的形式无差异，所以进行空间分析的速度相同；但是由于其含义有变化，因而分析结果有不同。对于宁夏，如果用传统的车矩阵进行空间分析，宁夏是属于没有明显聚集特征的单元，但是如果用 RLA 矩阵，则可以看出宁夏属于“低-低”模式，并且在  $p=0.01$  的条件下置信。造成这一差异的主要原因是，在传统的矩阵中，宁夏邻接的陕西、甘肃、内蒙古几个省对宁夏的影响是相同的，而这显然不符合事实，陕西与宁夏接壤的省界长度很短，空间联系实际上不高。在本文的模型中减少了陕西省对宁夏的影响，由此得到陕西省的聚集模式是“低-

高”模式。对于甘肃省，如果用传统的车矩阵来进行空间分析，邻接省区新疆、青海、四川、陕西、宁夏、内蒙古对其影响相同，新疆、青海、甘肃三省形成了“高-高”聚集模式，但是这种模式事实上并不精确，从空间邻接性的角度来看，甘肃受到新疆的影响相对于其他各个省区的影响较小(排第四)，甘肃周围有相对发病率更高的新疆、青海、内蒙古省份，也有更低的四川、陕西、宁夏省区。其中，内蒙古、青海和陕西的影响力更大，新疆与甘肃的邻接性不强，其形成的“高-高”模式并不是稳健的。对于浙江省，两种方法均得到“低-低”的聚集模式，新的矩阵表明置信程度更高，说明浙江与其周

围省区的病毒性肝炎发病率均表现偏低。对于江苏等地,RLA 矩阵相对于传统的车矩阵同样得到了一定的改进。

### 3 结 语

权重矩阵不同,对相同数据采用相同的分析方法会得出不同的相关性结论。原因在于权重矩阵不同导致邻域范围不同,同样两个区域在一种矩阵定义下相邻,在另一种矩阵定义下则未必相邻。在不同的邻域范围作用下,可能会产生不同的结果,与全局的 Moran's I 值和  $p$  值没有正向或者负向关系。本文提出的 RLA 空间权重矩阵实际上是车矩阵的一种广义矩阵,这种矩阵能够恰当地处理好空间自相关分析中的大量无意义的“伪邻接”关系。

本文以 2011 年病毒性肝炎的中国大陆各省发病情况为统计数据,进行 RLA 矩阵的阈值选取,证明在  $1.5 \times 10^{-3} < t < 2.0 \times 10^{-3}$  时,所建立的矩阵能较好地反映 2011 年病毒性肝炎发病率的空间分布特征。将传统的车矩阵和 RLA 矩阵进行对比,从运算时间上来看没有差异;从效果上看,RLA 矩阵在进行了一些相关性的优化之后,得到了更为符合实际的聚类分析图和显著性分析图。

### 参 考 文 献

- [1] Chen J, Hill A A, Urbano L D. A GIS-based Model for Urban Flood Inundation[J]. *Journal of Hydrology*, 2009, 373(1): 184-192
- [2] Yilma J M, Malone J B. A Geographic Information System Forecast Model for Strategic Control of Fasciolosis in Ethiopia[J]. *Veterinary Parasitology*, 1998, 78(2): 103-127
- [3] Kristensen T K, Malone J B, McCarroll J C. Use of Satellite Remote Sensing and Geographic Information Systems to Model the Distribution and Abundance of Snail Intermediate Hosts in Africa: A Preliminary Model for Biomphalaria Pfeifferi in Ethiopia[J]. *Acta Tropica*, 2001, 79(1): 73-78
- [4] Chen Fei, Du Daosheng. Application of the Integration of Spatial Statistical Analysis with GIS to the Analysis of Regional Economy[J]. *Geomatics and Information Science of Wuhan University*, 2002, 27(4): 391-396(陈斐, 杜道生. 空间统计分析与 GIS 在区域经济分析中的应用[J]. 武汉大学学报 · 信息科学版, 2002, 27(4): 391-396)
- [5] Feng Xin, Du Shihong, Shu Hong. Influence of Spatial Weight Matrices on Spatial Autocorrelation: A Case Study of HFRS in China[J]. *Geomatics and Information Science of Wuhan University*, 2011, 36(12): 1 410-1 413(冯昕, 杜世宏, 舒红. 空间权重矩阵对空间自相关的影响分析——以我国肾综合征出血热疾病为例[J]. 武汉大学学报 · 信息科学版, 2011, 36(12): 1 410-1 413)
- [6] Lai Shengjie, Liao Yilan, Zhang Honglong, et al. Comparing the Performance of Temporal Model and Temporal-spatial Model for Outbreak Detection in China Infectious Diseases Automated-Alert and Response System, 2011-2013, China [J]. *Chinese Journal of Preventive Medicine*, 2014, 48(4): 259-264(赖圣杰, 廖一兰, 张洪龙, 等. 2011-2013 年国家传染病自动预警系统中时间模型和时空模型应用效果比较[J]. 中华预防医学杂志, 2014, 48(4): 259-264)
- [7] Song Yongze, Ge Yong, Peng Junhuan, et al. Application of Genetic Programming on Predicting and-Mapping Malaria in Anhui Province[J]. *Journal of Geo-information Science*, 2015, 17(8): 954-962(宋泳泽, 葛咏, 彭军还, 等. 疟疾预测的遗传规划方法与应用——以安徽省县(市)疟疾发病率为例[J]. 地球信息科学学报, 2015, 17(8): 954-962)
- [8] Yang Xunfeng, Li Lianfa, Wang Jinfeng, et al. The Use of Generalized Additive Model in Association of Temperature and Cardiovascular Mortality in Anhui[J]. *Journal of Geo-information Science*, 2015, 17(11): 1 388-1 394(杨勋凤, 李连发, 王劲峰, 等. 安徽省温度与心脑血管疾病死亡关系的广义叠加模型分析[J]. 地球信息科学学报, 2015, 17(11): 1 388-1 394)
- [9] Buckeridge D L, Okhmatovskaia A, Tu S, et al. Understanding Detection Performance in Public Health Surveillance: Modeling Aberrancy-Detection Algorithms[J]. *Journal of the American Medical Informatics Association*, 2008, 15(6): 760-769
- [10] Brownstein J S, Freifeld C C, Madoff L C. Digital Disease Detection-Harnessing the Web for Public Health Surveillance[J]. *New England Journal of Medicine*, 2009, 360(21): 2 153-2 157
- [11] Noufaily A, Enki D G, Farrington P, et al. An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems[J]. *Statistics in Medicine*, 2013, 32(7): 1 206-1 222
- [12] Amaral P V, Anselin L. Finite Sample Properties of Moran's I Test for Spatial Autocorrelation in Tobit Models[J]. *Papers in Regional Science*, 2014, 93(4): 773-781
- [13] Gao Fenghua, Feng Qianjin, Jiang Lifeng, et al. Analysis on Epidemic of Legal Infectious Diseases of

- Mailand China[J]. *Modern Preventive Medicine*, 2013, 4(4): 756-759(高风华, 冯谦谨, 江丽凤, 等. 中国大陆 2002-2010 年法定传染病疫情分析[J]. 现代预防医学, 2013, 4(4): 756-759)
- [14] Du Juan, Guan Zequn. A Review of the Applications of Spatial Analysis in Epidemiology[J]. *Modern Preventive Medicine*, 2007, 34(9): 3 691-3 698(杜娟, 关泽群. GIS 在流行病学研究中的应用[J]. 现代预防医学, 2007, 34(9): 3 691-3 698)
- [15] Fang Liqun. Study on Spatiotemporal Distribution and Environmental Risk Factors of Hemorrhagic Fever with Renal Syndrome[D]. Beijing: Academy of Military Medical Sciences, 2009(方立群. 肾综合征出血热时空分布及环境危险因素研究[D]. 北京: 中国人民解放军军事医学科学院, 2009)
- [16] Ma Hongchao, Li Dren. Geographic Window Fixing and Its Application Based on Spatial Statistics[J]. *Geomatics and Information Science of Wuhan University*, 2001, 26(1): 18-23(马洪超, 李德仁. 基于空间统计学的空间数据窗口大小的确定[J]. 武汉大学学报·信息科学版, 2001, 26(1): 18-23)
- [17] Guan X, Jin X, Pan Q, et al. Analysis on Spatial Difference of Newly Increased Farmland by Land Consolidation in China at County Level[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(20): 226-233
- [18] She Bing, Zhu Xinyan, Su Kehua, et al. Test Methods for Space-Time Interaction of Events Under Road Network Constraints[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(3): 353-356(余冰, 朱欣焰, 苏科华, 等. 道路网约束下的事件时空交互检验方法研究[J]. 武汉大学学报·信息科学版, 2015, 40(3): 353-356)
- [19] Anselin L, Rey S. Properties of Tests for Spatial Dependence in Linear Regression Models[J]. *Geographical Analysis*, 1991, 23(2): 112-131
- [20] Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to Spatial Data Analysis[J]. *Geographical Analysis*, 2006, 38(1): 5-22
- [21] Anselin L. Local Indicators of Spatial Association-LISA[J]. *Geographical Analysis*, 1995, 27(2): 93-115
- [22] Moran P A P. Notes on Continuous Stochastic Phenomena[J]. *Biometrika*, 1950: 17-23

## An Improved Spatial Weights Matrix Construction Strategy

ZHAI Weixin<sup>1</sup> CHENG Chengqi<sup>2</sup>

1 School of Earth and Space Science, Peking University, Beijing 100871, China

2 College of Engineering, Peking University, Beijing 100871, China

**Abstract:** In the study of geographical spatial autocorrelation, the spatial weights matrix is regarded as a fundamental and essential field area of research. In this paper, we firstly analyze the advantages and disadvantages of several common weights matrixes like ROOK matrix, QUEEN matrix and K-Nearest matrix. Considering the improvements for in the traditional spatial weights matrix, we put forward RLA (Ratio of Length and Area) which is that can be applied with a stricter measurement standards and represents the relative relationship of a spatial unit to other adjacent ones units instead of a simple true or false discrimination to improve the accuracy. For verification, we carried out experiments on the Mainland China viral hepatitis statistical data from 2004 to 2012 of Mainland China based on the provincial administrative divisions. The results indicate that the proposed weights matrix not only achieves the fundamental functions of a spatial weights matrix, but also is treated as a general definition of ROOK matrix, freely applicable to when implementing spatial autocorrelation analysis. The adoption of this RLA spatial weights matrix will further reveal the spatial relationships among different geographical units, also providing support for the prevention of epidemics.

**Key words:** spatial autocorrelation; spatial weights matrix; viralhepatitis

**First author:** ZHAI Weixin, PhD, specializes in spatial modelling. E-mail: pkuzhaiweixin@gmail.com

**Corresponding author:** CHENG Chengqi, professor. E-mail: ccq@pku.edu.cn

**Foundation support:** The National Major Projects on High-Resolution Earth Observation System, Nos. 03-Y30B06-9001-13/15, 30-Y30B13-9003-14/16; the Research Projects on Public Services of Mapping and Geography Information, No. 201512020.