

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

王红亮, 胡伟平*, 吴 驰

(华南师范大学地理科学学院, 广东广州 510631)

摘要:空间权重矩阵定量表达了地理要素之间的空间关系,是空间统计分析的重要工具.该文详细分析了 Arcgis 9.3用于空间自相关分析的空间权重矩阵的生成方法,并以湖南省各县的城乡收入差距为例进行了实证分析.分析结果表明,随着各要素的平均影响范围增大,全局 Moran'I呈曲线下下降趋势,局部自相关的结果也出现明显差异.

关键词:空间权重矩阵;空间自相关; Moran'I

中图分类号: K903 **文献标志码:** A

空间自相关是空间统计分析的前提条件,是认识空间分布特征的一种常用方法^[1].要进行空间自相关度量,首先需要通过空间权重矩阵定量表达地理要素之间的空间关系.近几年,随着空间统计技术的广泛应用,国内外一些学者对空间权重矩阵展开了较为深入研究.国内主要集中在如何构建的技术层面,如刘旭华、王劲峰^[2]利用 ArcInfo生成的弧段属性表来建立要素之间的二进制邻接关系;魏晓峰、吴健平^[3]开发的空間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^[4]利用 Arcobject实现了 Rooks情形下的 0、1权重矩阵.上述这些研究中,基本只涉及到了一两种简单的权重矩阵建立,可选择性太少,不能满足复杂的空间关系建模;徐彬^[5]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,而且定量分析了各种方法对于 Moran'I的影响,但文中没有涉及到如何约束要素的最少邻居数、如何处理岛状多边形及如何处理空间均质性等一些复杂情况,也没有找出权重矩阵对于 Moran'I的影响规律.国外具有代表性的研究成果有 ANSELIN^[6]开发的 GeoDa软件以及 Esri公司开发的 ArcMap软件中的 Spatial Statistics Tools模块.2款软件提供了多种权重矩阵建立方法,可满足大多数的空间关系的模拟.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.本文详细分析了 Spatial Statistics Tools模块中各种权重矩阵的基本构建方法,并探讨了权重矩阵对于空间自相关指数的影响规律,以期空间自相关分析合理构建权重作为参考.

1 空间自相关分析及权重矩阵的构建

1.1 空间自相关分析

空间自相关分析是用于检验某一要素的属性值是否显著地与其相邻空间点上的属性值相关联的重要指标,可分为正相关和负相关 2类,正相关表明某单元的属性值变化与其相邻空间单元具有相同变化趋势,负相关则正好相反^[7].

1.1.1 **全局空间自相关指数** 最为常用的全局空间自相关指数为 Moran'I统计量(本文涉及到的空间自相关统计量均为 Moran'I),具体计算为^[8-9]:

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

式中 n 表示观测单元数目, w_{ij} 表示互为邻居单元之间交互权重值, x_i 表示单元观测值, \bar{x} 表示所有观测

收稿日期: 2009-09-23

基金项目: 国家自然科学基金资助项目(40871086)

作者简介: 王红亮(1981-),男,湖南祁阳人,华南师范大学2007级硕士研究生,Email: 277653923@qq.com;胡伟平(1969-),男,湖南衡阳人,华南师范大学副教授,主要研究方向:遥感与地理信息系统应用,Email: huwup@scnu.edu.cn

* 通讯作者

值的均值. $Moran'$ I 取值范围在 $(-1, 1)$, 如果空间过程不相关, 则 I 的期望接近 0. 当 I 取负值时, 一般表示负的自相关, 取正值表示正的自相关. 计算出 I 指数后, 还需要对其进行检验, 一般采用 Z 检验:

$$Z_I = \frac{I - E[I]}{\sqrt{Var[I]}}$$

(2)

其中 $E[I]$ 为空间自相关指数的期望值, $Var[I]$ 为标准差.

1.1.2 局部空间自相关 $Moran'$ I 统计量通过适当修正可用于探测局部尺度上的空间自相关性, 也就是空间异质性. 计算方法如下:

$$I_i = \frac{x_i - \bar{x}}{\sum_{j=1, j \neq i}^n w_{ij} [x_j - \bar{x}]} \cdot \frac{\sum_{j=1, j \neq i}^n w_{ij} [x_j - \bar{x}]}{n - 1} \quad (3)$$

局部统计量 I_i 的检验方法类似全局指数的检验^[8-9].

1.2 ArcGIS 9.3 的空间统计工具箱中几种不同的空间关系建模方法

从式 (1)、(3)可以看出, 需要获取 w_{ij} 才能计算出空间自相关指数. ArcGIS 9.3 用 Generate Spatial Weights Matrix工具来产生这个权重矩阵. 其产生的权重矩阵基于 2种策略(表 1).

表 1 ArcGIS 9.3 中各种构建权重矩阵的方法
Tab 1 The methods applied to generate spatial weight matrix in ArcGIS 9.3

策略	方法
Binary strategies	FIXED_DISTANCE
	K_NEAREST_NEIGHBORS
	CONTIGUITY_EDGES_ONLY
	CONTIGUITY_EDGES_CORNERS
	DELAUNAY_TRIANGULATION
Weighted strategies	INVERSE_DISTANCE
	ZONE_OF_DIFFERENCE

对于 Binary strategies 要素之间的交互权重非 0 即 1; 对于 Weighted strategies 要素之间的权重取值在 0 至 1 区间.

1.2.1 Weighted strategies 使用 Weighted strategies 构建权重矩阵, 基于如下公式^[10]:

$$w_{ij} = \frac{1}{d^k} \quad (\text{当 } i, j \text{ 之间的距离小于 } d); \quad (4)$$

$$w_{ij} = 0 \quad (\text{其他}). \quad (5)$$

INVERSE_DISTANCE(以下简称 ID): 用该方法产生权重矩阵, k 可以取 1 或者 2. 对该方法而言, 需设置门槛距离 d (Threshold distance)、邻居数 N (Num-

ber of neighbor). d 有 3 种赋值: Null 0 或自定义. 若自定义, 则超过该距离之后, 要素之间的交互权重为 0. 否则用公式 (4) 计算; 若为 Null d 值将自动调整为保证每个要素至少有 1 个邻居 (Neighbor) 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.

ZONE_OF_DIFFERENCE(以下简称 ZD): 该方法和上面的差异在于, 先设定门槛距离 $d > 0$. 若要素之间的距离超过 d 权重用公式 (4) 计算. 要素之间的距离在设定的 d 范围内, 权重为 1. 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^[10].

1.2.2 Binary strategies 若采用 Binary strategies 构建权重矩阵, 则采用公式 $w_{ij} = 1$ (当 i, j 互为邻居) 和 $w_{ij} = 0$ (其他) 计算.

FIXED_DISTANCE(以下简称 FD): 该方法需设置门槛距离 d 、邻居数 N . 其中 d 不能为 0. 若要素之间距离在 d 范围之内, 权重为 1. 否则取 0; 若 d 设定为空值, 则表示每个要素至少有 1 个邻居.

K_NEAREST_NEIGHBORS(以下简称 KNN): 该方法一般只需设置邻居数 K . K 值用于保证每个要素拥有 K 个最近的邻居, 每个要素与这最近的 K 个邻居之间权重为 1. 与剩下其他要素之间权重取 0.

CONTIGUITY_EDGES_ONLY(以下简称 CE): 该方法主要针对多边形要素类, 有公共边的要素互为邻居^[11], 两者之间权重取 1. 无公共边的 2 个要素之间权重取 0. 当存在岛状多边形的时候, 需要设置邻居数 N . N 仅仅用来设置岛状多边形的邻居个数, 对其他非岛状多边形, 该数值无效.

CONTIGUITY_EDGES_CORNERS(以下简称 CEC): 该方法和 CE 基本一样, 差异在于, 有公共边或公共点的要素之间权重取 1. 否则取 0.

DELAUNAY_TRIANGULATION(以下简称 DT): 该方法基于要素质心产生狄罗尼三角形, 享有公共边的 2 个顶点所在的要素互为邻居, 二者之间权重取 1. 否则取 0.

上述所有方法中, 对多边形要素而言, 取质心计算两者之间距离, 其中距离的计算多采用欧氏距离公式.

2 案例研究——湖南省县域居民城乡收入差距的空间自相关实证分析

2.1 研究区域与数据处理

本文选取湖南省 2007 年 88 个县域单元的城乡

收入差距作为研究,为了可比性,本文将市辖区排除在外.城乡收入差距用城镇居民人均可支配收入减去农村人均纯收入来衡量.统计数据来源《2008年湖南省统计年鉴》及各市州的统计公报.2007年县域城乡收入差距数据按自然断点法符号化进行分级(图1).从图1可以看出,湖南省2007年县域城乡收入差距空间差异明显,按照空间位置,大体呈现北部、中部、南部3个差异较为显著的地带.其中差距最大是长沙县,8833元,最小为洪江市,2902元.城乡收入差距在7106~8833元的有13个县域单元,6107~7106元的有25个县域单元,介于5244~6107元的有25个县域单元,介于4224~5244元的有19个县域单元,介于2902~4224元的有6个单元.

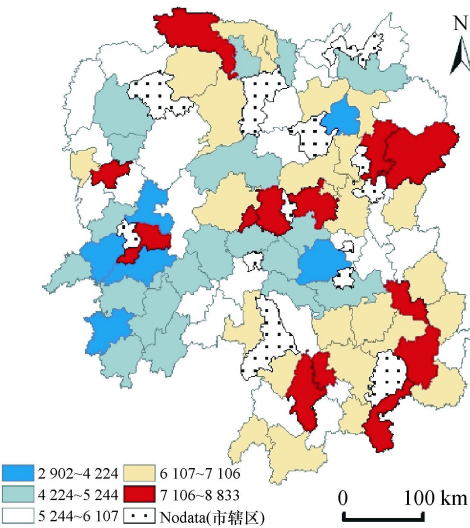


图1 收入差距的空间分布

Fig.1 The spatial distribution of income gap

2.2 不同权重矩阵构建方法对每个县域单元拥有邻居数的差异

不同方法构建权重矩阵会使得每个县域单元拥有的邻居数有所不同,从而造成权重矩阵有所差异.表2是各种方法在不同参数设置下得到的湖南省县域单元最大最小邻居数,可以看出:

对ID方法而言:若门槛距离 $d=0$ m,则每个县域单元的邻居数都为87,此时参数邻居数 N 的设置无意义;

若设门槛距离 d 为空值,则程序会保证每个县域单元至少有1个邻居,由此产生1个新的门槛距离 d_1 ,在 d_1 约束下,某些县域单元最多拥有7个邻居.此时可用邻居数 N 来设定每个县域单元的最小邻居数.若 $N>7$,则 d_1 再次更新为另一距离 d_2 ,用来保证每个县域至少拥有 N 个邻居,这种情况会导致每个县域拥有的最大邻居数和最小邻居数一样都

为 N ;若设 $N<7$,则在 d_1 条件下,某些县域最小邻居数为 N ,一些县域会拥有最大邻居数7;

若要设定1个大于零的门槛距离,可以采用Calculate Distance Band from Neighbor Count工具加以配合.这里用该工具计算出了每个单元至少拥有10个邻居的距离,结果为:最小距离74413 m,最大距离153344 m,平均距离101432 m;

若设门槛距离 $d=153344$ m,在该距离下,每个单元最少会有10个邻居,一些单元可拥有最大32个邻居数.若设 $N<11$,不会改变县域单元拥有的最小和最大邻居数;若设 $10<N<32$,则最小邻居数改为 N ,但最大邻居数仍为32;若 $N>32$,原有的最小邻居数和最大邻居数都将改为 N .

对FD方法而言,若设门槛距离 $d=74413$ m,在该值约束下,每个县域最多拥有10个邻居.若设 $N>10$,程序会调整每个县域拥有的最大邻居数和最小邻居数都为 N ;若设 $N<11$,不会改变某些县域所拥有的最大邻居数10,但最小邻居数等于设定的 N 值.

对KNN方法而言,本文设 $N=10$,该值保证每个县域的邻居数都为10.

用CE/CEC及DT构建权重矩阵,无需设置参数,每个县域拥有的邻居数见表2.

2.3 湖南省县域居民城乡收入差距的分布与特征空间分析

为进一步揭示不同方法构建的权重矩阵差异,下面对湖南省城乡收入差距进行空间自相关分析.

2.3.1 不同方法构建的权重矩阵对全局空间自相关指数的影响分析 本文根据不同的方法构建了8个权重矩阵文件(表3).所有结果表明,全局存在较弱空间自相关,相关程度因Moran'I而异.

进一步观察得知,对于ID方法而言,若设置门槛距离 $d=0$ m,计算出来的Moran'I与其他两者相差很远.对于FD方法,表中二者设置的最小邻居数都为10,但由于二者的最大邻居数不一样,结果计算出来的Moran'I相差近1倍.这很有可能是由于每个县域单元考虑影响范围过大导致Moran'I发生了变化.

为进一步证实该猜测,对这2种方法,本文将门槛距离设置为Null, N 取一系列值,然后计算相应的平均邻居数及Moran'I值.

结果表明,二者的平均邻居数小于某一阈值水平时(本文为3.95),Moran'I相接近,超过该阈值后,随着县域单元考虑的平均影响范围越来越大,2种方法构建的权重矩阵所对应的全局Moran'I值都呈曲线下降趋势;在平均邻居数一样的情况下,ID方法对应的Moran'I要大于FD方法(图2).

表 2 不同参数设置下的邻居数
Tab 2 The neighbors by different preferences

方法	参数设置		结果		
	门槛距离 d/m	邻居数 N>0(整数) /个	最小邻居数 /个	最大邻居数 /个	平均邻居数 /个
ID (k=1, 下文同)	0	N<88	87	87	87
	Null	N<8	N	7	未定
	Null	N>7	N	N	N
	153 344	N<11	10	32	21. 14
	153 344	10<N<32	N	32	未定
	153 344	N>32	N	N	N
FD	74 413	N<11	N	10	未定
	74 413	N>10	N	N	N
KNN	Null	10	10	10	10
CE	Null	Null	1	9	4. 68
CEC	Null	Null	1	9	4. 68
DT	Null	Null	3	9	5. 64

表 3 权重矩阵对全局空间自相关的影响 ($\alpha=0. 01$)
Tab 3 The effect of spatial weightmatrix on General Moran' I
($\alpha=0. 01$)

方法	参数 d	参数 N	全局 Moran' I	Zscore
ID	0	Null	0. 048 9	4. 811
	Null	10	0. 217 4	4. 902
	153 344	10	0. 150 0	4. 778
FD	Null	10	0. 209 3	5. 148
	153 344	10	0. 111 6	4. 137
KNN	Null	10	0. 209 3	5. 148
CE /CEC	Null	Null	0. 271 4	4. 002
DT	Null	Null	0. 236 3	3. 996

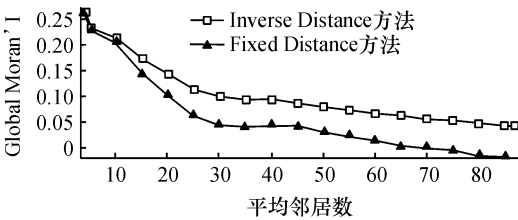


图 2 平均邻居数与全局空间自相关指数
Fig 2 The average neighbors and Global Moran' I

对 KNN 方法和 FD 方法,在取 $N>7$ 、门槛距离为空值时的权重结果一样,计算的指数没有差异.对于 DT 与 CE /CEC 方法,结果不受参数影响,虽然二者的最大邻居数一样,但 DT 最小的邻居数要大于 CE /CEC 对应的单元考虑的影响范围要大于 CE /CEC (二者的平均邻居数量参见表 2),因此计算的 Moran' I 要小于 CE /CEC 方法计算的 Moran' I 值.

2. 3. 2 不同方法构建的权重矩阵对局部空间自相关指数的影响分析 本文采用 4 种不同的空间权重

矩阵进行局部空间自相关分析,结果见表 4 和图 3.

表 4 权重矩阵对局部空间自相关的影响 ($\alpha=0. 01$)
Tab 4 The effect of spatial weightmatrix on Local Moran' I
($\alpha=0. 01$)

编号	方法	d	N	HH	HL	LH	LL
图 3(a)	CE /CEC	Null	Null	3	1	1	8
图 3(b)	ID	Null	80	6	2	1	10
图 3(c)	FD	Null	80	1	6	0	0
图 3(d)	FD	Null	87	0	1	4	0

表 4 中 HH 表示城乡居民人均收入差距呈现高值聚集分布 (正相关), HL 表示呈现高值被低值包围分布 (负相关), LH 表示低值被高值所包围分布 (负相关), LL 表示低值聚集分布 (正相关), 剩余地方为随机分布 (图 3).

图 3(a)至图 3(d)为表 4 对应的聚类分布图. 图 3(a)与图 3(b)都只存在 1 个 LH, 但分布位置不一样,前者为临澧县,后者为湘阴县;前者比后者少了 1 个 HL 为中方县;前者衡阳县为 LL 相对后者的随机分布更为合理 (对比图 1 可看出);二者在探测 HH 分布时差别显著,前者只有 3 个 HH 分布,后者则相对合理 (对比图 1 可看出),有 6 个. 图 3(c)相对于图 3(a)与图 3(b)差别更为明显,前两者的 LL 分布在图 3(c)全为随机分布. 而图 3(d)则彻底改变了前三者的分布模式,HH 和 LL 分布全部消失.

为进一步分析单元的平均邻居数对局部自相关的影响,类似 2. 3. 1 中采用的方法,对于 FD 和 ID 方法,将门槛距离设置为 Null N 取一系列值,然后

计算对应的各类集聚数目.结果见图 4.

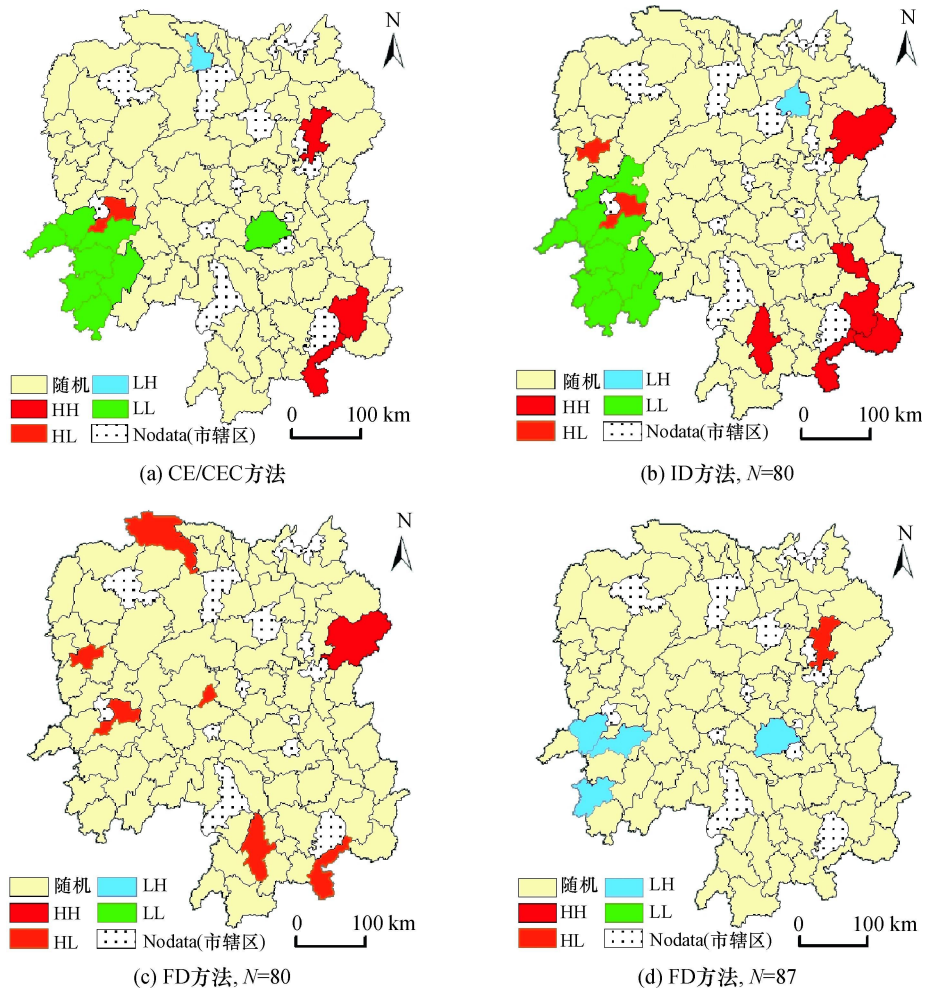


图 3 局部自相关结果
Fig 3 The results of Local Moran 'I

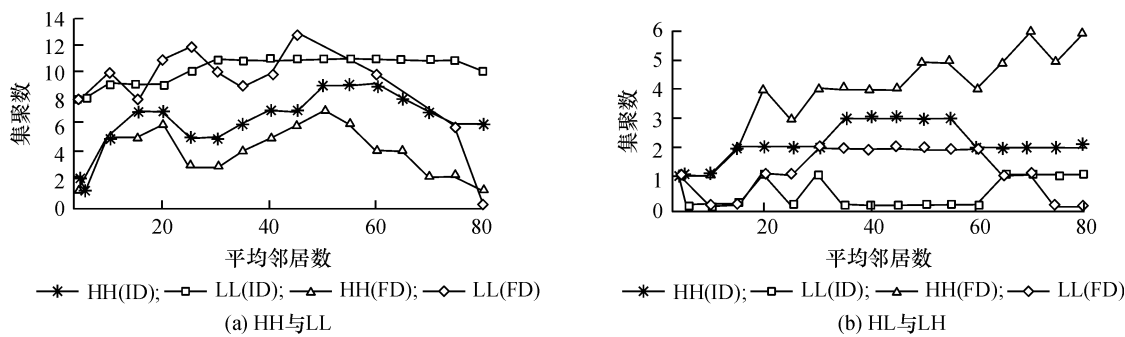


图 4 集聚数与平均邻居数
Fig 4 The number of clusters and average neighbors

从图 4(a)可以看出,随着平均邻居数的增加,2 种方法探测的 HH 和 LL 数目都存在明显的波动,其中 ID 方法对应的 LL 数目波动相对 FD 方法较为缓和;在平均邻居数一样的情况, ID 方法对应的 HH 数要高于 FD 方法.

从图 4(b)可以看出,随着平均邻居数的增加,2

种方法探测的 HL 和 LH 数目也存在明显的波动;在平均邻居数一样的情况下, FD 方法探测的 HL 与 LL 数目均要大于 ID 方法.

CEC /CE 及 DT 方法对应的结果为一定值, KNN 方法和 FD 方法在门槛距离为空值时产生的权重结果基本一样,故未进行讨论.

3 结论

CONTIGUITY_EDGES_CORNERS方法所产生权重矩阵反映的空间关系较为简单,因其构建简便而成为大多数研究者在进行空间自相关分析时的主要选择.但对同一研究区域而言,通常随着要素考虑的影响范围增大,全局 Moran' I呈曲线下降趋势,探测局部自相关的 4类集聚空间分布和数量也会出现明显差异甚至完全不一样的结果.因此,研究者应该充分考虑空间相互作用的复杂性,尝试多种方法的空间关系建模,选择最符合实际情况的权重矩阵进行空间自相关分析.本文只分析了权重矩阵对于 Moran' I的影响,而对于其他空间自相关指数,如 Geary' s C统计量及 G统计量的影响,有待进一步深入探讨.

参考文献:

[1] 何宗贵,韩世民,崔道.空间自相关分析的统计量探讨[J].中国血吸虫病防治杂志,2008,20(4):317.
HE Zonggui HAN Shimin CUIDao Discussion on statistic of spatial autocorrelation analysis[J]. China Schistose Control 2008,20(4):317.
[2] 刘旭华,王劲峰.空间权重矩阵的生成方法分析与实验[J].地球信息科学,2002(2):38-44.
LIU Xuhua WANG Jinfeng Analysis and application on the specification methods of the spatial weight matrix[J]. Geo-Information Science 2002(2):38-44.
[3] 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[J].测绘与空间地理信息,2005,28

(6):78-79.
WEIXiaofeng WU Jianping The Development and application of spatial autocorrelation based on ArcGIS[J]. Geomatics & spatial information technology 2005,28(6):78-79.
[4] 潘海燕,程朋根,肖根如.基于 ArcObjects的空间权重矩阵的建立与实现[J].测绘科学,2007,32(6):130-132.
PAN Haiyan CHENG Penggen XIAO Genru The constitution and realization of spatial weight matrix based on ArcObjects[J]. Science of Surveying and Mapping 2007,32(6):130-132.
[5] 徐彬.空间权重矩阵对 Moran' I指数影响的模拟分析[D].南京:南京师范大学,2007:2-19.
[6] ANSELIN L Exploring Spatial Data with GeoDaTM[EB/OL]. (2004-05-05)[2009-07-20]. <http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf>
[7] 孟斌,张景秋,王劲峰.空间分析方法在房地产市场研究中的应用——以北京市为例[J].地理研究,2005,24(6):956.
MENG Bin ZHANG Jingqiu WANG Jinfeng Application of spatial analysis to the research of real estate, taking Beijing as a case[J]. Geography research 2005,24(6):956.
[8] 王远飞,何洪林.空间数据分析方法[M].北京:科学出版社,2007:110-119.
[9] 徐建华.计量地理学[M].北京:高等教育出版社,2006:120-126.
[10] ESRI ArcGIS Desktop Help[EB/OL]. (2009-04-24)[2009-07-20]. <http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=welcome>
[11] LEE J WONG W S Statistical analysis with ArcView GIS[M]. New York: John Wiley & Sons Incorporated 2000:132-133.

ANALYZING THE EFFECT OF SPATIAL WEIGHTED MATRIX ON SPATIAL AUTOCORRELATION
—— TAKING HUNAN'S INCOME GAP BETWEEN URBAN AND RURAL AREAS AS A CASE

WANG Hongliang HU Weiping*, WU Chi
(School of Geography, South China Normal University, Guangzhou 510631, China)

Abstract: Spatial weight matrix quantificationally expresses the spatial relationships between the geographic features, which is an important tool for spatial statistical analysis. The different methods in generating spatial weight matrix, which is used for spatial autocorrelation analysis in ArcGIS 9.3, are analyzed by Hunan's county income gap between urban and rural as a case for empirical analysis. The analysis indicates that with the expansion of average influenced scope of each feature, the global Moran' I shows a curve drop and the result of local spatial autocorrelation presents significant differences.
Key words: spatial autocorrelation; spatial weight matrix; Moran' I

【责任编辑 成文】