

洞庭湖区钉螺与血吸虫人畜感染的空间回归分析

秦建新¹, 仇应山

摘要:目的 分析钉螺面积与人、畜感染数量的空间回归关系,了解洞庭湖区各县、区钉螺对血吸虫病人、畜感染的影响程度及影响范围,指导湖区血吸虫病的防治工作。方法 利用ArcGIS10空间关系模块中的普通最小二乘法回归模型、地理加权回归模型对人、畜感染进行建模。结果 钉螺面积对人、畜感染影响程度较大,每增加一单位的钉螺面积会相应增加0.153487个感染病人和0.003838头病牛;钉螺面积对人、畜感染存在空间异质性的正相关。结论 空间回归分析可以较好地用于洞庭湖区钉螺面积与血吸虫人畜感染的关系研究,且地理加权回归模型对这种异质性的模拟较普通最小二乘法回归模型更加优越。

关键词:血吸虫病;钉螺;OSL;GWR;洞庭湖

中图分类号:R188.1,R532.21 文献标识码:A 文章编号:1009-9727(2013)9-1071-03

Spatial regression analysis of schistosoma infection among inhabitants and livestock and oncomelania in Dongting Lake area. QIN Jian-xin, QIU Ying-shan. (College of Resource and Environment Science, Hunan Normal University, Changsha, 410081, Hunan, P. R. China)

Abstracts: Objective To analyze spatial regression relationship between the schistosoma infection in inhabitants and livestock and oncomelania in Dongting Lake area. Methods The ordinary least squares regression model and the geographically weighted regression model in ArcGIS10 were used to establish model of the schistosoma infection in inhabitants and livestock. Results The relationship of the snail's area with human and livestock infection was relatively close, each unit increase in the snail area led to the increase of 0.153 487 infected patients and 0.003838 infected livestock. There exists positive correlation of spatial heterogeneity between snail area and the infection of human and livestock. Conclusion Spatial regression analysis can be well used in analysis of the relationship between snail area and human/livestock infection in Dongting Lake area, and the simulation of spatial heterogeneity with geographically weighted regression model is better than the ordinary least squares regression model.

Key words: Schistosomiasis; Snail; OSL; GWR; Dongting Lake

洞庭湖流域是日本血吸虫病的流行区域,钉螺是日本血吸虫的唯一中间宿主,毛蚴寄生在钉螺内发育成尾蚴,尾蚴感染终宿主是人、畜。因此,钉螺的分布和数量与人、畜血吸虫病的感染存在着内在的联系。空间回归分析,是将传统线性回归分析结合空间权重矩阵,从而引入空间依赖的一种分析方法^[1]。通过空间回归分析,可以对空间关系进行建模、检查和探究,有助于解释所观察到的空间模式背后的诸多因素。我们利用ArcGIS10空间关系模块中的普通最小二乘法回归模型、地理加权回归模型对人、畜感染血吸虫病进行建模,分析钉螺面积与人、畜感染数量的空间回归关系,得出其回归系数的空间分布图,从而了解各县、区钉螺对血吸虫病人、畜感染的影响及影响程度,指导血吸虫病的防治。

1 资料与方法

1.1 资料 血吸虫统计资料来自湖南省血防办2000年至2010年统计资料,并选取感染病人、病牛头数和钉螺面积为分析指标;空间矢量数据来自国家基础地

理信息数据,选用洞庭湖疫区17个县、区的行政区划shapefile面矢量数据,并对对应17个行政区输入上述属性数据。17县、区分别为:岳阳市区、岳阳县、华容县、湘阴县、汨罗市、临湘市、常德市区、安乡县、汉寿县、澧县、临澧县、桃源县、石门县、津市市、益阳市区、南县、沅江市,如表1所示。

1.2 方法 分别采用普通最小二乘法和地理加权回归分析普通最小二乘法(Ordinary least squares linear regression,下简称OLS)假设在研究区域内的所有空间单元的因变量与自变量间的空间关系为相同的线性关系,即回归参数在空间上不变,适合所有变不随地理位置变化而变化的观测对象,OLS是进行地理加权回归分析之前的必要步骤,是空间回归分析的起点。其表达式为 $y = x\beta + \epsilon$ (1)

$$\text{其中 } y = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{Bmatrix}, x = \begin{Bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{12} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{Bmatrix}, \beta = \begin{Bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{Bmatrix}, \epsilon = \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{Bmatrix}$$

基金项目:国家自然科学基金项目(No.40971038);科技部公益性行业科研专项(No.201009021)

作者单位:湖南师范大学资源与环境科学学院,湖南,长沙410081

作者简介:秦建新(1969~),男,汉族,湖南省永州人,博士,教授,主要从事“3S”技术在地方病传播中的应用。

式中, y 为因变量; x 为自变量的 $n \times (p+1)$ 阶回归矩阵; β 为参数向量; ϵ 为随机误差向量。并且, 自变量之间线性无关, 是满秩矩阵, 样本容量 n 大于自变量个数 p , 随机误差项 ϵ 满足正态分布且独立等方差^[2]。

表1 17个县血吸虫指标

Tab 1 Statistics indicators of schistosomiasis of the 17 counties			
行政区	感染病人	病牛头数	钉螺面积(千 m ²)
County(Distiret)	No.infected	No.infected cow	Snail area(1 000m ²)
岳阳市区 Yueyang	162 117	4 109	917 032.13
岳阳县 Yueyang	106 225	1 644	482 017.13
华容县 Huanrong	53 968	1 640	405 966.38
湘阴县 Xiangyin	92 123	1 272	432 088.25
汨罗市 Guluo	44 081	1 398	80 019.79
临湘市 Linxiang	30 373	1 134	55 395.80
常德市区 Changde	64 806	1 847	260 429.44
安乡县 Anxiang	47 767	1 252	434 390.63
汉寿县 hanshou	108 544	3 226	728 253.00
澧县 Lixian	78 644	2 649	539 861.88
临澧县 Linli	9 563	168	21 741.94
桃源县 Taoyuan	1 130	761	32 813.14
石门县 Shimen	912	0	13 540.67
津市市 Jinshi	13 792	1 036	81 561.12
益阳市区 Yiyang	32 705	1 665	126 573.78
南县 nanxian	125 623	1 836	122 445.43
沅江市 yuanjiang	167 871	5 066	648 975.00

:各指标为 2000~2010 年累计值, 其中钉螺面积单位为(千 m²)

地理加权回归 (Geographically weighted regression, 下文称 GWR) 模型将数据的地理位置引入到回归参数中。GWR 模型是 OLS 模型的空间扩展, 允许随回归参数的地理位置变化而变化^[3]。模型的数学表达式为 $W_i Y = W_i X \beta_i + \epsilon_i$ (2)

式中, W_i 为 $n \times n$ 阶对角矩阵, 是观测点 i 到邻近观测点结论的函数; Y 为 n 个点上采集的因变量的 $n \times 1$ 阶观测向量; X 为 $n \times k$ 阶自变量矩阵; β_i 为观测点 i 对应的 $n \times 1$ 阶参数向量; ϵ_i 为 $n \times 1$ 阶误差向量, 服从方差为常数的正态分布。

2 结果

2.1 感染病人、病牛头数和钉螺面积的 OLS 分析
OLS 回归模型解释了钉螺面积在空间全局下对人、畜血吸虫病感染的影响(表 2)。用于评估模型性能的“校正 R^2 ”结果分别达到 0.639937 和 0.656270, 表明 OLS 模型能够解释感染病人和病牛头数的 64% 和 66% 的变化; 由“系数”项知, 每增加一单位的钉螺面积会相应增加 0.153487 个感染病人和 0.003838 头病牛; “概率 p ”结果和 T 检验结果表明以上解释变量具有统计学意义。同时对两次 OLS₁ 和 OLS₂ 模型的标准化残差进行一阶邻接多边形的全局空间自相关分析得出 Moran I 指数分别为 -0.005483 和 -0.172145, 标准化 Z 值分别为 0.366648 和 -0.668399, 标准化残差不存在聚类现象, 模型是可靠的。

2.2 感染病人、病牛头数和钉螺面积的 GWR 分析
通过不同参数设置表明, 采用“自适应”核函数和 AICc 带宽的 GWR 分析可以得到最佳结果, 比 OLS 模型有显著提高(表 3, OLS₁ 表示以感染病人和钉螺面积为变量的 OLS 模型, 其他类似; 保留小数点后 3 位)。GWR₁ 的决定系数(校正 r^2)较 OLS₁ 增加 0.244, 虽然 GWR₂ 的决定系数较之 OLS₂ 几乎不增加, 但总体上 GWR₂ 拟合效果优于 OLS₂。

对 GWR₁ 和 GWR₂ 结果的残差进行空间自相关分析得到 Moran I 指数分别为 -0.214553 和 -0.256551, 表明 GWR 模型的残差更具有空间离散性。

2.3 钉螺面积与感染病人的空间关系
由于 GWR 模型在空间位置变化条件下对因变量和自变量的关系进行建模, 因此根据每个县、区的回归系数可以得出钉螺面积对人、畜感染的影响程度及这种程度在空间上的分布。

由图 1 可知, 大多数县、区钉螺面积与感染病人的回归系数介于 0.7 与 1.0 之间, 华容县、南县和沅江市的回归系数处在 0.3 与 0.7 之间, 安乡县最低, 呈现出高相关度包围较低相关度的 U 形包围圈。出现这种格局的原因可能是 U 形圈内的县、区布置有大面积的水域, 水域周长与面积比值相对其他县、区偏小, 这种情况会减少人可接触的钉螺面积, 从而钉螺面积对感染人数的影响程度出现下降。

表2 人、畜感染和钉螺面积的 OLS 计算结果统计表

Tab 2 OLS regression results of Oncomelania area & infected Human and farm cattle

模型 Model	变量 Variable	系数 Coefficient	标准差 SD	t 统计量 Counted	概率 p Probability	校正 r^2 Corrected r^2
OLS ₁	截距 Intercept	18 471.095	11 830.710	1.561284	0.139307	-
	钉螺面积 Snail area	0.153487	0.028290	5.425561	0.000071*	0.639937
OLS ₂	截距 Intercept	590.673	285.777	2.066901	0.056454	-
	钉螺面积 Snail area	0.003838	0.000683	5.616779	0.000049*	0.656270

表3 OLS与GWR拟合效果对比
Tab 3 Comparison of OLS & GWR results

回归模型 Regression model	AIC 值	r^2	校正 r^2 Corrected r^2
OLS ₁	406.515	0.662	0.640
GWR ₁	-343.676	0.997	0.884
OLS ₂	279.925	0.678	0.656
GWR ₂	-450.945	0.990	0.656

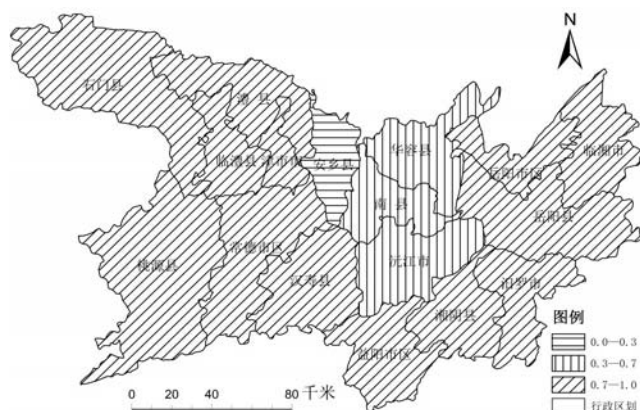


图1 钉螺面积与感染病人回归系数空间分布

Fig 1 Coefficient distribution of infected Human

2.4 钉螺面积与病牛头数的空间关系 当病牛头数作为因变量时,出现了高回归系数和低回归系数的两级分化(图2),津市市、安乡县、华容县、汨罗市、湘阴县和益阳市区的钉螺面积和病牛数量相关系数小于0.3,钉螺面积的大小对病牛数量的影响程度较小。其余11县、区的相关系数大于0.7,南县和沅江市的相关系数由以感染人数为因变量时的低值转变为以病牛头数为因变量的高值,津市市、汨罗市、湘阴县和益阳市区则由高值变为低值。

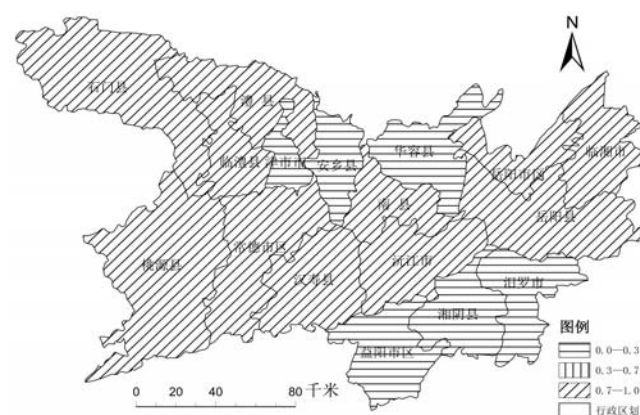


图2 病牛头数回归系数空间分布图

Fig 2 Coefficient distribution of infected farm cattle

3 讨论

本文在钉螺面积影响人、畜感染的前提下通过GWR模型分析得出其影响的空间差异。一方面反映

各县、区钉螺传播血吸虫病的危害程度;一方面为血吸病的防治和卫生资源的配置提供指导信息。GWR模型将地理位置引入回归拟合,效果优于OLS模型,故GWR模型可以用于空间异质性探索。以钉螺面积作为单一的解释变量与人、畜感染数量进行GWR模型拟合能够取得较好的效果。影响钉螺分布区域与人、畜接触的因素还包括水域分布、地形和土地利用类型等,因此将这些因素加入到模型的解释变量将会起到跟好的拟合效果。

参考文献:

- [1] SUN N, PU YX.. Spatial Regression Analysis with Python'S Functional Programming [J]. Geomatics World, 2009(3):66 - 68. (In Chinese) (孙宁,蒲英霞. 使用Python函数式编程进行空间回归分析[J]. 地理信息世界, 2009(3):66 - 68).
- [2] Huang XY, Xie BC. Statistics: methods and Applications [M]. Beijing, China Renmin University Press, 2009. (In Chinese) (黄向阳,谢邦昌. 统计学:方法与应用[M]. 北京:中国人民大学出版社, 2009.)
- [3] Yang Y, Yang JY.. Spatial regression analysis on influence factors of maize lodging stress[J]. Transact CSAE, 2011, 27(6): 244 - 249. (In Chinese) (杨杨,杨建宇. 玉米倒伏胁迫影像因子的空间回归分析[J]. 农业工程学报, 2011(6):244-249.)
- [4] Tang GA, Yang X.. ArcGIS Spatial analysis of geographic information systems experiment tutorial[M]. Beijing, Science Press, 2006(In Chinese) (汤国安,杨昕. ArcGIS地理信息系统空间分析实验教程[M]. 北京:科学出版社, 2006)
- [5] Brunson C., Fotheringham AS, Brunson C, et al... Geographically Weighted Regression: The Analysis of Spatially Varying Relationships [M]. Chichester, Wiley, 2002.
- [6] Ying LG, Ning YM.. Spatial Data: ITS Nature, Effects and Analysis[J]. Advance Earth Sci, 2005(1):49 - 56. (In Chinese) (应龙根,宁越敏. 空间数据:性质、影响和分析方法[J]. 地球科学进展, 2005,(1):49- 56.)
- [7] ESRI virtual campus: http://training.esri.com/Courses/ts_RegAnalysis93/viewPresentation1.cfm.
- [8] Qin JX, Tan ZF, Zhang C.Environment factors and spatial characters of distribution of Oncomelania Snails in marshland beach of Dongting Lake area [J]. J Nature Disas, 2008, 17(4): 19 - 27(In Chinese) (秦建新,谭子芳,张春. 洞庭湖区江滩钉螺分布的环境因子及空间特征-基于RS/GIS的研究[J]. 自然灾害学报, 2008, 17(4):19-27)
- [9] Wang TP, Zhou XN, Malone. JB. Prediction of Schistosomiasis Using GIS in Jiangsu, Anhui and Jiangxi Provinces.[J]. Chin J Schistosom Control, 2004, 16,(2): 86-90. (In Chinese) (汪天平,周晓农,JB. Malone,等. 地理信息系统(GIS)用于江苏、安徽和江西省血吸虫病流行预测的研究[J]. 中国血吸虫病防治杂志, 2004, 16(2):86-90)
- [10] Barbour AD. Modeling the transmission of schistosomiasis: an introductory view[J]. Am J Trop Med Hyg, 1996, 55(Suppl): 135-143.

收稿日期:2013-05-09 编辑:崔宜庆