

地理探测器:原理与展望

王劲峰^{1,2}, 徐成东¹

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101;
2. 中国科学院大学, 北京 100049)

摘要: 空间分异是自然和社会经济过程的空间表现,也是自亚里士多德以来人类认识自然的重要途径。地理探测器是探测空间分异性,以及揭示其背后驱动因子的一种新的统计学方法,此方法无线性假设,具有优雅的形式和明确的物理含义。基本思想是:假设研究区分为若干子区域,如果子区域的方差之和小于区域总方差,则存在空间分异性;如果两变量的空间分布趋于一致,则两者存在统计关联性。地理探测器 q 统计量,可用以度量空间分异性、探测解释因子、分析变量之间交互关系,已经在自然和社会科学多领域应用。本文阐述地理探测器的原理,并对其特点及应用进行了归纳总结,以利于读者方便灵活地使用地理探测器来认识、挖掘和利用空间分异性。

关键词: 空间分异性;地理探测器; q 统计;类型量

DOI: 10.11821/dlxb201701010

1 引言

空间统计学于20世纪60年代兴起,其理论基础为空间自相关,突破了经典统计学的独立同分布假设前提^[1]。基于空间自相关,一系列方法工具被研制出来,主要包括空间自相关检验 Moran's $I^{[2]}$ 、半变异函数^[3]、Ripley $K^{[4]}$; 热点探测 Gi^[5]、LISA^[6]、SatScan^[7]; 空间插值方法 Kriging^[3]、贝叶斯最大熵 BME^[8]; 空间回归 SAR/MAR/CAR^[9-11]、GWR^[12]、空间贝叶斯层次模型 BHM^[13]等。这些方法工具大大促进了人们对于具有空间自相关性数据的分析和利用。

随着定位及观测技术的发展和普及,无论是更精细或者更大范围的研究,还是空间大数据,空间分层异质性 (spatial stratified heterogeneity) 问题凸显。空间分层异质性,简称空间分异性或区异性,是指层内方差小于层间方差的地理现象,例如地理分区、气候带、土地利用图、地貌图、生物区系、区际经济差异、城乡差异以及主体功能区等等^[14-19],是空间数据的另一大特性。“层 (strata)”是统计学概念,大体对应地理上的类 (classes) 或子区域 (sub regions)。

虽然已有上百种分类算法 K-means、SOM 等可用于分类或分区,但是针对空间分异

收稿日期: 2016-10-15; 修订日期: 2011-01-05

基金项目: 国家自然科学基金项目(41531179, 41601419, 41421001); 科技部科技基础性工作专项(2014FY121100); 公益性行业科研专项(GYHY20140616); LREIS 自主创新项目(088RA200YA) [Foundation: National Natural Science Foundation of China, No.41531179, No.41601419, No.41421001; Ministry of Science and Technology of China, No.2014FY121100, Special Scientific Research Fund of Public Welfare Profession of China, No. GYHY20140616; Science and Technology Innovation Project of LREIS, No.088RA200YA]

作者简介: 王劲峰(1965-), 男, 研究员, 中国地理学会会员(S110009913M), 主要从事地理信息科学的理论创新和实践。
E-mail: wangjf@igsnrr.ac.cn

性的统计学方法尚十分有限。目前, 主要方法包括空间分异性测度和因子分析的地理探测器 q 统计^[20-21]; 在不同样本条件下的区域总量估算、空间插值和空间抽样的系列统计量: 当有分层样本时的 MSN ^[22-23]、当样本有偏时的 $Bshade$ ^[24-26]、当只有单样本点时的 SPA 模型^[27]、小样本多单元报告以及多边形图层之间的互相转换的 $Sandwich$ 模型^[28]。

地理探测器是探测空间分异性, 以及揭示其背后驱动力的一组统计学方法。其核心思想是基于这样的假设: 如果某个自变量对某个因变量有重要影响, 那么自变量和因变量的空间分布应该具有相似性^[29-30]。地理分异既可以用分类算法来表达, 例如环境遥感分类; 也可以根据经验确定, 例如胡焕庸线。地理探测器擅长分析类型量, 而对于顺序量、比值量或间隔量, 只要进行适当的离散化^[31], 也可以利用地理探测器对其进行统计分析。因此, 地理探测器既可以探测数值型数据, 也可以探测定性数据, 这正是地理探测器的一大优势。地理探测器的另一个独特优势是探测两因子交互作用于因变量。交互作用一般的识别方法是在回归模型中增加两因子的乘积项, 检验其统计显著性。然而, 两因子交互作用不一定就是相乘关系。地理探测器通过分别计算和比较各单因子 q 值及两因子叠加后的 q 值, 可以判断两因子是否存在交互作用, 以及交互作用的强弱、方向、线性还是非线性等。两因子叠加既包括相乘关系, 也包括其他关系, 只要有关系, 就能检验出来。

本文主要阐述了地理探测器的基本原理, 并介绍地理探测器软件; 通过对比分析地理探测器在自然科学、社会科学、环境科学和人类健康方面的几个典型应用案例, 讨论了地理探测器的适用条件和优势, 提出了可能遇到问题的对策。

2 地理探测器原理

空间分异性是地理现象的基本特点之一。地理探测器是探测和利用空间分异性的工具。地理探测器包括4个探测器。

分异及因子探测: 探测 Y 的空间分异性; 以及探测某因子 X 多大程度上解释了属性 Y 的空间分异(图1)。用 q 值度量^[29], 表达式为:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \quad (1)$$

$$SSW = \sum_{h=1}^L N_h \sigma_h^2, \quad SST = N \sigma^2$$

式中: $h = 1, \dots, L$ 为变量 Y 或因子 X 的分层(Strata), 即分类或分区; N_h 和 N 分别为层 h 和全区的单元数; σ_h^2 和 σ^2 分别是层 h 和全区的 Y 值的方差。 SSW 和 SST 分别为层内方差之和(Within Sum of Squares)和全区总方差(Total Sum of Squares)。 q 的值域为 $[0, 1]$,

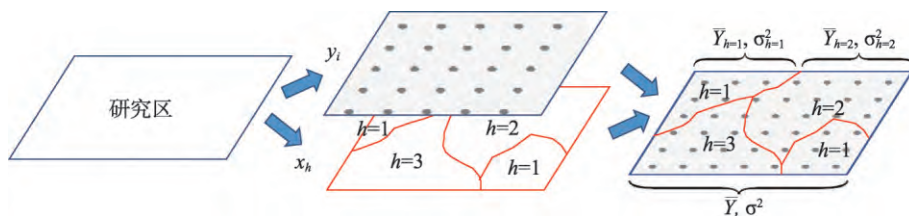


图1 地理探测器原理

Fig. 1 The principle of geographical detector

值越大说明 Y 的空间分异性越明显；如果分层是由自变量 X 生成的，则 q 值越大表示自变量 X 对属性 Y 的解释力越强，反之则越弱。极端情况下， q 值为 1 表明因子 X 完全控制了 Y 的空间分布， q 值为 0 则表明因子 X 与 Y 没有任何关系， q 值表示 X 解释了 $100 \times q\%$ 的 Y 。

q 值的一个简单变换满足非中心 F 分布^[21]：

$$F = \frac{N-L}{L-1} \frac{q}{1-q} \sim F(L-1, N-L; \lambda)$$

(2)

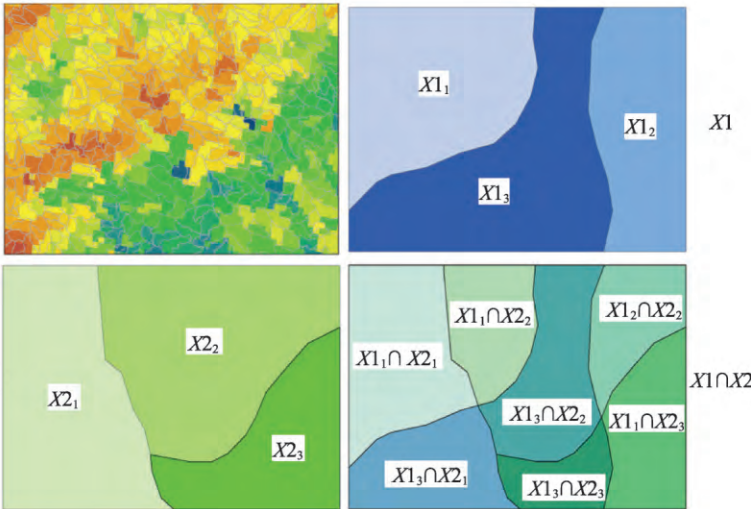
$$\lambda = \frac{1}{\sigma^2} \left[\sum_{h=1}^L \bar{Y}_h^2 - \frac{1}{N} \left(\sum_{h=1}^L \sqrt{N_h} \bar{Y}_h \right)^2 \right]$$

(3)

式中： λ 为非中心参数； $\bar{Y}_h$ 为层 h 的均值。根据式 (3)，可以查表或者使用地理探测器软件^[32] 来检验 q 值是否显著。

交互作用探测：识别不同风险因子 X s 之间的交互作用，即评估因子 $X1$ 和 $X2$ 共同作用时是否会增加或减弱

对因变量 Y 的解释力，或这些因子对 Y 的影响是相互独立的。评估的方法是首先分别计算两种因子 $X1$ 和 $X2$ 对 Y 的 q 值： $q(X1)$ 和 $q(X2)$ ，并且计算它们交互（叠加变量 $X1$ 和 $X2$ 两个图层相切所形成的新的多边形分布，图 2) 时的 q 值： $q(X1 \cap X2)$ ，并对 $q(X1)$ 、 $q(X2)$ 与 $q(X1 \cap X2)$ 进行比较。两个因子之间的关系可分为以下几类（图 3）：



注：分别计算出 $q(X1)$ 和 $q(X2)$ ；将 $X1$ 和 $X2$ 两个图层叠加得到新图层 $X1 \cap X2$ ，计算 $q(X1 \cap X2)$ ；按照图 3 判断两因子交互的类型。

图 2 交互作用探测

Fig. 2 Detection of interaction

风险区探测：用于判断两个子区域间的属

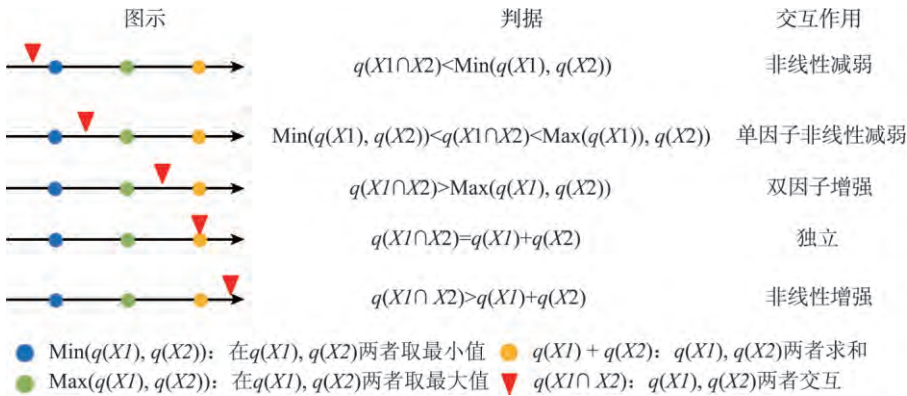


图 3 两个自变量对因变量交互作用的类型

Fig. 3 Types of interaction between two covariates

性均值是否有显著的差别,用 t 统计量来检验:

$$t_{\bar{y}_{h=1}-\bar{y}_{h=2}} = \frac{\bar{Y}_{h=1} - \bar{Y}_{h=2}}{\left[\frac{\text{Var}(\bar{Y}_{h=1})}{n_{h=1}} + \frac{\text{Var}(\bar{Y}_{h=2})}{n_{h=2}} \right]^{1/2}} \quad (4)$$

式中: $\bar{Y}_h$ 表示子区域 h 内的属性均值,如发病率或流行率; n_h 为子区域 h 内样本数量, Var 表示方差。统计量 t 近似地服从Student's t 分布,其中自由度的计算方法为:

$$df = \frac{\frac{\text{Var}(\bar{Y}_{h=1})}{n_{h=1}} + \frac{\text{Var}(\bar{Y}_{h=2})}{n_{h=2}}}{\frac{1}{n_{h=1}-1} \left[\frac{\text{Var}(\bar{Y}_{h=1})}{n_{h=1}} \right]^2 + \frac{1}{n_{h=2}-1} \left[\frac{\text{Var}(\bar{Y}_{h=2})}{n_{h=2}} \right]^2} \quad (5)$$

零假设 $H_0: \bar{Y}_{h=1} = \bar{Y}_{h=2}$,如果在置信水平 α 下拒绝 H_0 ,则认为两个子区域间的属性均值存在着明显的差异。

生态探测:用于比较两因子 X_1 和 X_2 对属性 Y 的空间分布的影响是否有显著的差异,以 F 统计量来衡量:

$$F = \frac{N_{x1}(N_{x2}-1)SSW_{x1}}{N_{x2}(N_{x1}-1)SSW_{x2}} \quad (6)$$

$$SSW_{x1} = \sum_{h=1}^{L1} N_h \sigma_h^2, \quad SSW_{x2} = \sum_{h=1}^{L2} N_h \sigma_h^2$$

式中: N_{x1} 及 N_{x2} 分别表示两个因子 X_1 和 X_2 的样本量; SSW_{x1} 和 SSW_{x2} 分别表示由 X_1 和 X_2 形成的分层的层内方差之和; $L1$ 和 $L2$ 分别表示变量 X_1 和 X_2 分层数目。其中零假设 $H_0: SSW_{x1} = SSW_{x2}$ 。如果在 α 的显著性水平上拒绝 H_0 ,这表明两因子 X_1 和 X_2 对属性 Y 的空间分布的影响存在着显著的差异。

3 地理探测器软件

GeoDetector是根据上述原理,用Excel编制的地理探测器软件,可从以下网址免费下载: <http://www.geodetector.org/>。

地理探测器使用步骤包括:

(1) 数据的收集与整理:这些数据包括因变量 Y 和自变量数据 X 。自变量应为类型量;如果自变量为数值量,则需要离散化处理。离散可以基于专家知识,也可以直接等分或使用分类算法如K-means等。

(2) 将样本 (Y, X) 读入地理探测器软件,然后运行软件,结果主要包括4个部分:比较俩区域因变量均值是否有显著差异;自变量 X 对因变量的解释力;不同自变量对因变量的影响是否有显著的差异,以及这些自变量对因变量影响的交互作用。

地理探测器探测两变量 Y 和 X 的关系时,对于面数据(多边形数据)和点数据,有不同的处理方式。

对于面数据,两变量 Y 和 X 的空间粒度经常是不同的。例如,因变量 Y 为疾病数据,一般以行政单元记录;环境自变量或其代理变量 X 的空间格局往往是循自然或经济社会因素而形成的,如不同水文流域、地形分区、城乡分区等。因此,为了在空间上匹配这两个变量,首先将 Y 均匀空间离散化,再将其与 X 分布叠加,从而提取每个离散点上的因变量和自变量值 (Y, X) 。格点密度可以根据研究的目标而提前指定。如果格点密度

大，计算结果的精度会较高，但是计算量也会较大。因此在实际操作时须要考虑精度与效率的平衡。GeoDetector的输入数据如表1所示，其中每条记录都是一个格点，各列分别存储因变量 Y 和自变量 X 。

对于点数据：如果观测数据是通过随机抽样或系统抽样而得到，并且样本量足够大，可以代表总体，则可以直接利用此数据在地理探测器软件中进行计算。如果样本有偏，不能代表总体，则需要用一些纠偏的方法对数据进一步的处理之后再在地理探测器软件中进行计算。图4示意了地理探测器软件的数据准备、软件界面和输出。

表 1 GeoDetector数据输入格式

Tab. 1 Input table for the GeoDetector

	Y	X_1	X_2	$\cdots$	X_n
格点1	y_1	x_{11}	x_{12}	$\cdots$	x_{1n}
格点2	y_2	x_{21}	x_{22}	$\cdots$	x_{2n}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
格点 n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nm}

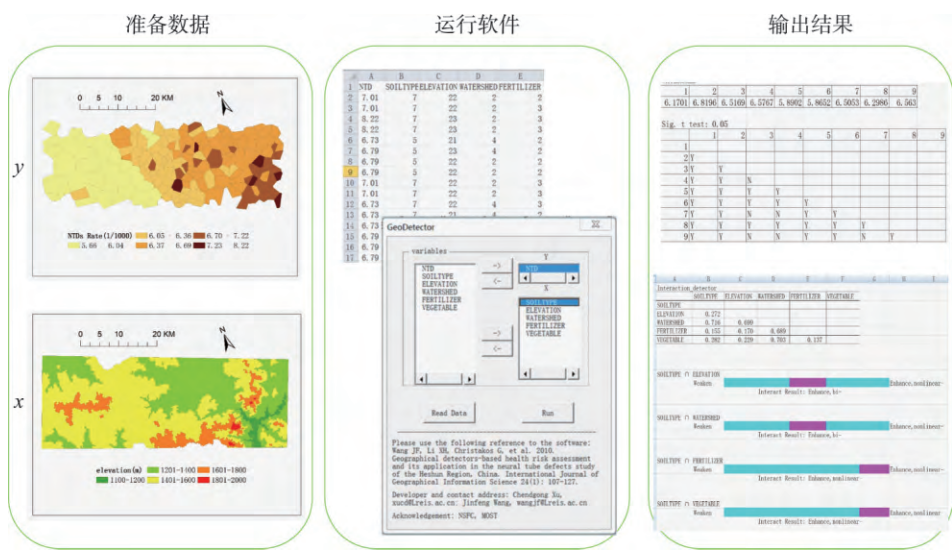


图4 地理探测器软件用户界面
Fig. 4 Users interface of the GeoDetector

4 应用案例

首先，结合新生儿神经管畸形空间变异的环境因子识别详尽演示GeoDetector的使用步骤；随后，分别介绍了地理探测器在自然科学（美国地表切割因子分析）、社会科学（中国城市化驱动力识别）和环境污染（土壤抗生素空间差异的控制因子分析）等不同领域的典型应用；最后，对地理探测器的应用案例进行分析比较，并归纳总结。

4.1 新生儿神经管畸形空间变异的环境因子识别^[20]

GeoDetector对某县1998-2006年的神经管畸形出生缺陷（NTDs）的发生（ Y ）进行了环境风险因子分析。环境风险因子或其代理变量（ X ）包括：高程、土壤类型、流域分区以及蔬菜产量和化肥使用量等社会经济变量^[20]（图5）。图6显示了本研究中所用的网格点，格点间距为1 km，共1145个，用于将 Y （数值量）和 X （类型量）匹配起来。

输入数据以及软件运行输出结果如图7所示。图8显示了风险因子探测的结果。图8a显示了对于单个风险因子（高程等级）而言的风险区探测的结果，其中图8a的表格第一行的数字“21”、“22”、“23”等是此环境因子各分区的编号，为类型量；第二行是在每

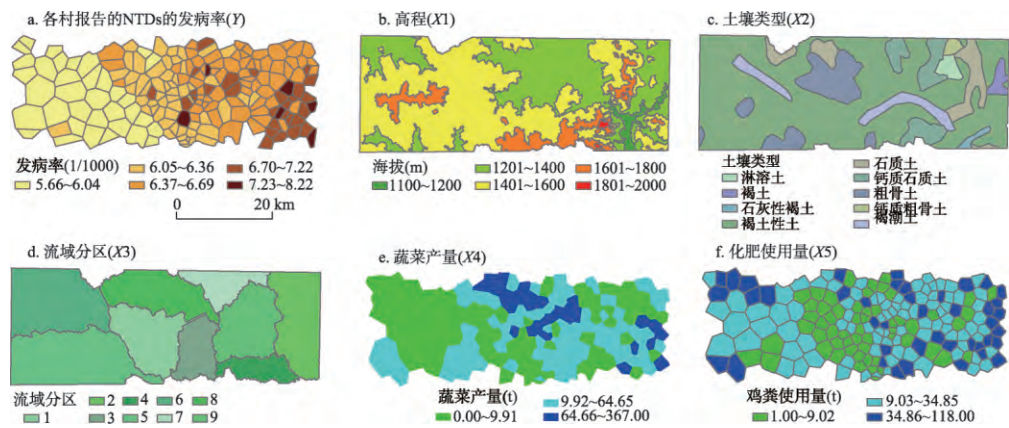


图5 某县新生儿神经管畸形的环境因子分析的数据^[20]
Fig. 5 The prevalence of neural tube defects and its covariates or proxies

个类型区内的NTDs的平均发病率，为数值量。接下来的第6~10行是各类型分区的NTDs发病率之间是否存在统计差异的判断，采用显著性水平为0.05的 t 检验，“Y”表示存在显著性差异，“N”表示不存在显著性差异。图8b展示的是所有风险因子 q 值的计算结果，结果表明水文流域变量（X3）具有最高的 q 值，说明这些变量中河流是决定NTDs空间格局的最主要的环境因子。图8c是生态探测的输出，结果显示就对NTD空间分布的作用而言，水文流域与其他变量存在着显著差异。图8d是交互探测的结果，其中第3~5行是两两变量交互作用后的 q 值，结果表明，任何两种变量对NTDs空间分布的交互作用都要大于第一种变量的独自作用。

4.2 美国陆表切割度空间变异的主导因素探测^[34]

地表切割度控制了水土流失、土地利用和生态功能，是历史和现实多因子作用的综合结果。美国的地表切割密度及其格局呈现出空间分层异质性（spatial stratified heterogeneity）。Luo等^[34]将地理探测器分别运用于美国8大地形区探测其形成的主导因子。发现最大 q 值所对应的因子随着不同特征和地质历史的地形区而变化。例如，岩性主导了山区地形，地形曲率（curvature）控制了平原区地形，冰期控制中央低地（图9）。地理探测器提供了反演地表过程因子的客观构架。

4.3 中国县域城市化空间变异的驱动力^[35]

城市化是中国过去和未来各20年最大的社会经济空间运动。认识县域城镇化的形成机制

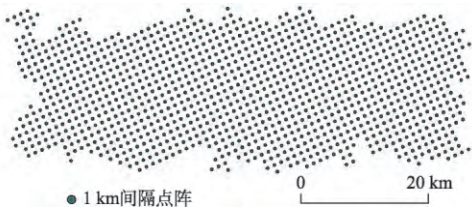


图6 某县新生儿神经管畸形的环境因子分析中用到的网格点^[20]
Fig. 6 Grid points to match neural tube defects and its covariates

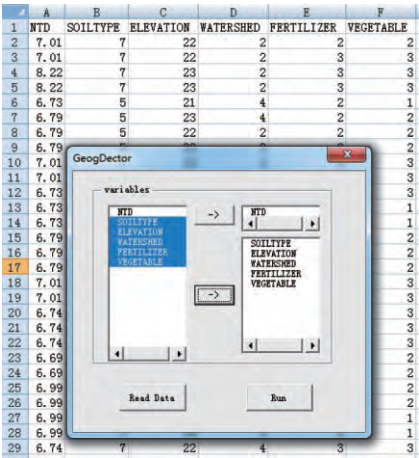


图7 数据输入的格式及运行界面
Fig. 7 Table of the input data and the users' interface of GeoDetector software

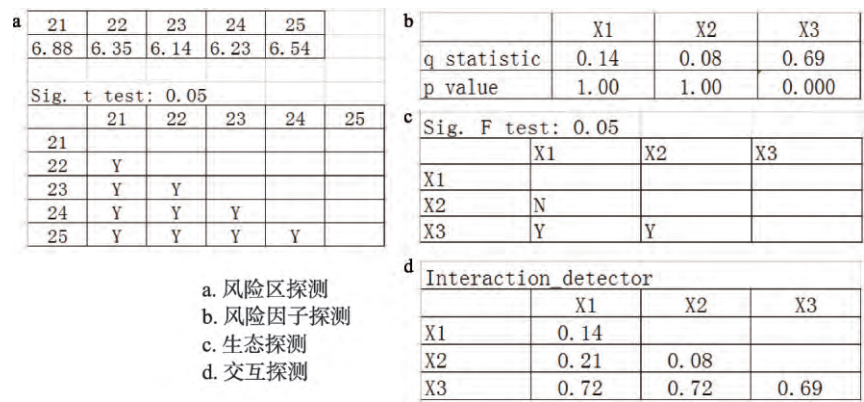
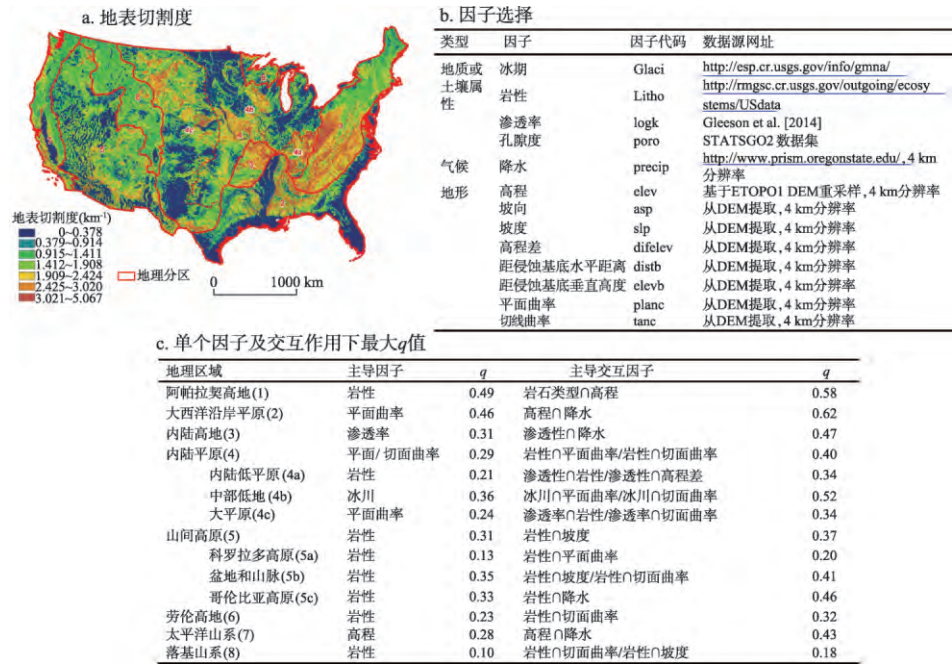


图8 风险区探测、风险因子探测、生态探测以及交互探测的结果

Fig. 8 Risk area detector (a), Risk factor detector (b), Ecological detector (c) and Interaction detector (d)



注: Y为地表切割度; X为候选因子 Y-X为主导因子识别。

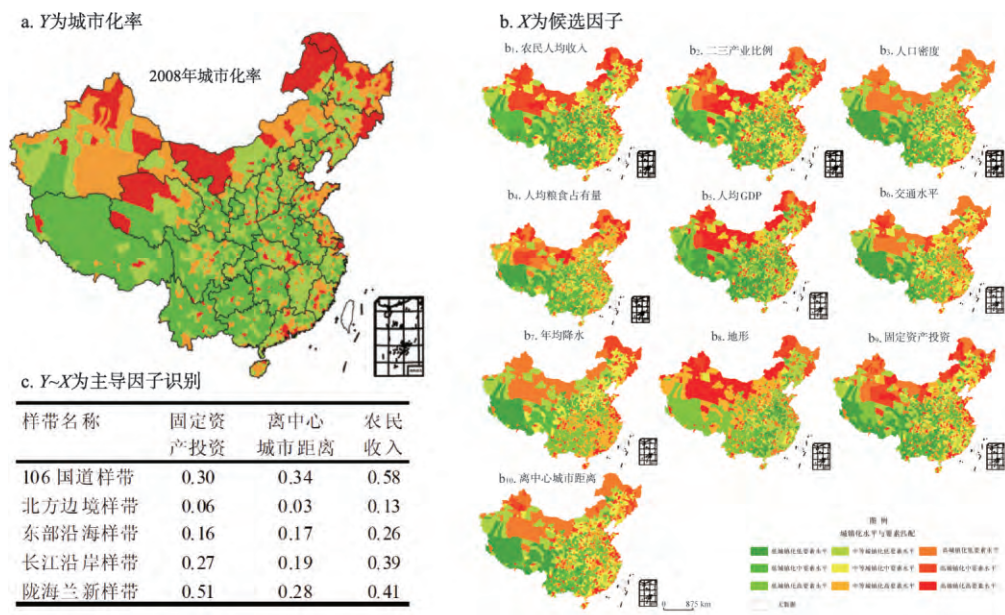
图9 美国地表切割度及其控制因子探测^[34]

Fig. 9 Land dissection in US and the detection of its controlling factors

为确立县域城镇化的发展战略,因地制宜地引导城镇化的地域模式提供参考依据。收集全国各县城市化率数据(图10a)以及候选解释变量数据(图10b),将地理探测器分别运用于106国道样带、北方边境样带、东部沿海样带、长江沿岸样带和陇海兰新铁路沿线样带,发现城市化驱动力在不同样带不同(图10c)。

4.4 土壤抗生素残留空间变异的因子分析^[36]

土壤和蔬菜里的抗生素残留通过食物链传递影响人类健康。山东省寿光县是中国最大的有机蔬菜种植基地,施用鸡粪;而鸡的饲料和饮水含有抗生素。研究区是寿光县的一个乡,面积约160 km²。寻找FQs类抗生素(Y)与候选影响因子(X)(包括蔬菜种植模式、肥料类型和数量、种植年限、大棚面积、地形起伏)之间的关系(图11)。地理探

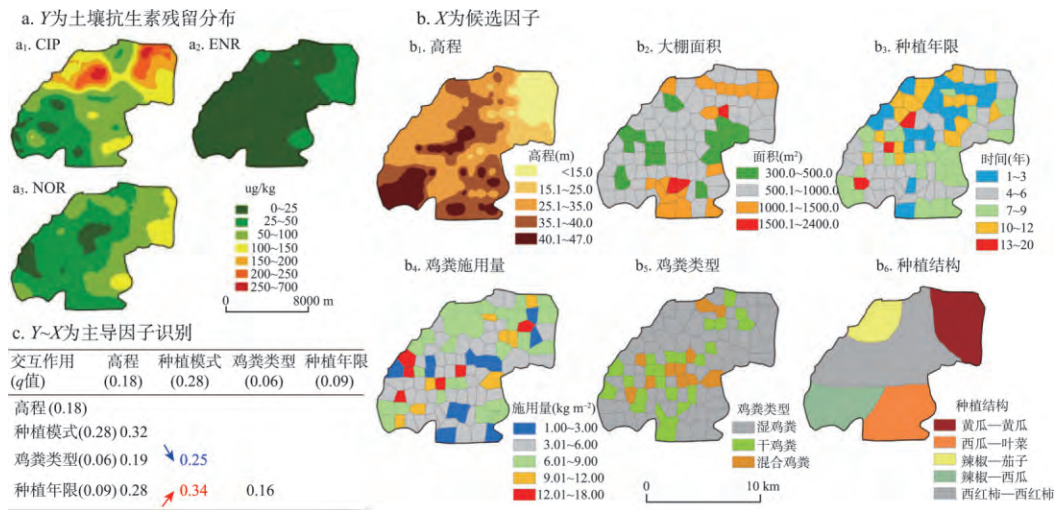


注： Y 为城市化率； X 为候选因子； $Y \sim X$ 为主导因子识别。

图 10 中国城市化率及其驱动因子探测^[35]

Fig. 10 Urbanization rate in China and the detection of its driving forces

测器 q 值如下：蔬菜种植模式（0.28）>使用鸡粪肥料数量（0.20）>高程（0.18）>种植年限（0.09）>鸡粪类型（0.060）>大棚面积（0.02）。可见，蔬菜种植模式是FQs抗生素空间分异的最重要的控制因素，例如，“黄瓜—黄瓜”模式（先种黄瓜，收获后再种黄瓜）对于FQs空间分异的决定力，是“辣椒—西瓜”模式（先种辣椒，收获后接着种西瓜）的3倍；排在第二位的鸡粪肥料使用量对FQs污染也有相当的影响。基于以上发现，政策建议是，为了达到既减少FQs土壤污染，又不对当地蔬菜生产产生过多影响的



注： Y 为土壤抗生素残留分布； X 为候选因子； $Y \sim X$ 为主导因子识别。

图 11 土壤抗生素残留及其影响因素探测^[36]

Fig. 11 Antibiotic residual in soil and the detection of its determinaints

目的, 一个有效而又可行的方案是调节种植模式和鸡粪使用模式 ($< 6 \text{ kg/m}^2$ 每年; 干鸡粪多于湿鸡粪)。

4.5 地理探测器应用案例分析比较

本文归纳了至2016年底, 地理探测器在各方面的运用, 对案例问题、研究区域、因变量 Y 、自变量或其代理变量 X 、研究发现和结论等方面进行归纳(附录), 分析比较可以发现地理探测器主要运用领域包括: 土地利用^[37-45]、公共健康^[46-54]、区域经济^[35, 44, 55-61]、区域规划^[45, 62-63]、旅游^[64-66]、考古^[67]、地质^[34]、气象^[68-70]、植物^[71]、生态^[72-75]、环境^[76-78]、污染^[34, 79-81]、遥感^[82]和计算机网络^[83]等。

地理探测器已被运用于从自然到社会十分广泛的领域; 其研究区域大到国家尺度, 小到一个乡镇尺度。在这些应用中, 地理探测器主要被用来分析各种现象的驱动力和影响因子以及多因子交互作用。这主要是因为地理探测器 q 值具有明确的物理含义, 没有线性假设, 客观地探测出自变量解释了 $100 \times q\%$ 的因变量。地理探测器在运用于因子和驱动力分析时, 因变量是数值量, 而自变量是类型量; 如果自变量是数值量, 需要将其离散化为类型量。

地理探测器作为驱动力和因子分析的有力工具^[20]已经在以上案例以及关于精准扶贫的研究中得到验证。实际上, 地理探测器还可以作为空间分异性分析的有力工具^[21], 这在关于胡焕庸线、中国热带北界的研究中可得到充分证明。

5 结论和讨论

5.1 空间分异性检验是空间数据分析的必要环节

空间自相关性和空间分异性是空间数据的两大特性, 也是空间数据的两大信息资源, 可以挖掘利用, 以认识其背后过程机理。现代空间统计学是围绕空间自相关展开的。地理数据中的类型量广泛存在, 它体现了地理对象的空间分异性, 表现为层内方差小于层间方差。地理探测器是度量、挖掘和利用空间异质性的新工具, 其理论核心是通过空间异质性来探测因变量与自变量之间空间分布格局的一致性, 据此度量自变量对因变量的解释度, 即 q 值。地理探测器比一般统计量有更强的势(Power), 更加确信, 强烈提示因果关系, 因为两个变量在二维空间分布一致比两个变量的一维曲线的一致要难得多。

地理探测器是空间数据探索性分析的有力工具。产生空间分异性的原因是多样的: 可能由于各层(类)的机理不同, 也可能是由于各层(类)的因子不同、或者各层(类)的主导因子不同。这些不同都会导致空间分异性。用全局模型分析具有异质性的对象将掩盖对象的异质性, 被混杂效应所干扰, 甚至导致错误的结论。因此, 在数据分析开始时, 就应当首先探测是否存在空间异质性, 据此确定是使用全局模型、还是选取局域模型; 是使用全域变量还是选用局域变量; 是使用全局参数还是局域参数? 表2为不同对象性质所对应的建模策略。

5.2 地理探测器的适用条件

表3比较了空间自相关检验、地理探测器和线性回归的研究对象, 变量类型, 统计量, 模型原理, 以及统计推论的差异。表4比较了地理探测器与方差分析的异同, 可见地理探测器包含方差分析, 比方差分析适用面更加广泛, 并且具有明确的物理含义。

归纳起来, 地理探测器可以为3方面使用: ① 度量给定数据的空间分异性; ② 寻找

表2 空间分异性和建模策略

Tab. 2 Spatial stratified heterogeneity and modelling strategies

空间静态		空间分异			
		机理分异	因素分异	因素影响力分异	原因不清
A全局模型		B不同模型	C不同变量	D不同参数	E数据分组
模型 f	一个	多个模型	一个模型	一个模型	尝试B
变量 x	一个(组)	一个(组)/每个模型	多组变量	一组变量	尝试C
参数 b	一个(组)	一个(组)/每个模型	多组参数	多组参数	尝试D
举例	一个模型 $y=f(x, b)$ CLR; Krige	不同层不同模型 $y_h=f_h(x_h, b_h)$	不同层不同变量 $y_h=f(x_h, b_h)$	不同层不同参数 $y_h=f(x, b_h)$ GWR; MLM	洪水预报：春季 融雪/秋季降雨

注：“层”指Strata，即不同类或不同区；CLR表示经典线性回归；GWR：地理加权回归；MLM：多层模拟。

变量最大的空间分异；③寻找因变量的解释变量。

地理探测器的适用条件：①擅长自变量 X 为类型量（如土地利用图），因变量 Y 为数值量（碳排放）的分析；②当因变量 Y 和自变量 X 均为数值量，对 X 离散化转换为类型量后，运用地理探测器建立的 Y 和 X 之间的关系将比经典回归更加可靠，尤其当样本量 <30 时。因为统计学一般要求样本单元数 >30 ；而地理探测器的 X 为类型量，同类相似，因此样本单元的代表性增加了，或者说地理探测器可以用 <30 的样本量达到更大样本量其他模型才能达到的统计精度；③对变量无线性假设，属于方差分析（ANOVA）范畴，物理含义明确的，其大小反映了 X （分层或分类）对 Y 解释的百分比 $100\times q\%$ 。表4展示了地理探测器与方差分析的差异。④地理探测器探测两变量真正的交互作用，而不同于计量经济学预先指定的乘性交互。⑤地理探测器原理保证了其对多自变量共线性免疫。⑥在分层中，要求每层至少有2个样本单元。样本越多，估计方差越小。

表3 空间热点、空间分异性和空间线性回归

Tab. 3 Comparison between spatial hotspot, spatial stratified heterogeneity and spatial linear regression

	空间热点检验	空间分异性检验	空间线性回归
空间过程	空间局域异质性	空间分层异质性	平稳线性过程
变量	数值量 y	数值量 y ~类型量 x	数值量 y ~数值量 x
统计量	Getis Gi; LISA; SatScan	q 统计	β
原理	统计量的观测值与随机期望值之差	两变量空间分布的一致性； x 分层（分类/分区）使 q 值最大	两变量回归误差； x 最优系数使 R^2 最小
结论	y 的空间热点区	y 的解释变量 x ， x 解释了 $100\times q\%$ $q = SSB_y/SS\hat{T}y$	y 对 x 的弹性， b $b = dy/dx$

注： SSB 为层间方差，即式(1)中 $SST-SSW$ 。

表4 地理探测器 q 统计与方差分析的区别

Tab. 4 Comparison between the GeoDetector q -statistic and ANOVA

均值；方差					
方差分析	$F = \infty$	$F = 9894$	N/A	N/A	$F = 0$
q -统计	$q = 1$	$q = 0.8$	$q = 0.7$	$q = 0.6$	$q = 0$

注：第一行表示两层（strata）的均值和方差

5.3 灵活使用地理探测器

地理探测器可以灵活使用，①既可以探测全局驱动力（最大 q 值所对应的自变量），如某县神经管畸形发生的环境因子^[20]；也可以探测不同地区的局域驱动力（不同地区最

大 q 值对应的自变量), 如在不同样带城市化主导驱动力不同^[35]和美国不同区域地表切割度主导因子^[34]; 还可以探测驱动力的时间变化(不同时间段的 q 值), 如干预前后神经管畸形主导因子的变化^[74]和遥感图像滤波前后地物可分性变化^[82]; 也可以探测不同尺度的驱动力, 例如中国住宅价格在全国、省会、地级市、县不同层次各因子解释力不同^[58]。

② 数值量可以离散化为类型量, 例如将GDP分级, 从而数值量也可以使用地理探测器。

③ Y 和 X 均有空间分布, 但并非必须在地理空间上分层, 也可以对属性分层, 即 Y 或 X 的分层可以是地理空间、时间、或者属性。以上抛砖引玉, 读者可以思考各种有创意的使用方法。

④ 基于Excel的地理探测器软件GeoDetector只有一个用户界面, “一键式”操作很容易掌握; 不需要安装任何GIS组件, 不需要用户具有GIS知识和操作技能, 所有的空间信息都存储于格点中, 小巧、免费; 在软件使用时须要注意, 对于变量采用不同的离散化粒度会对模型结果有影响, 一般选择 q 值最大的离散化方案^[31]。

5.4 地理探测器使用中常见问题和解决方案

地理探测器使用中经常遇到的问题和应对方法。

① 地理探测器中的自变量必须为类型量, 如果数据中此变量为连续型变量, 软件运行时则会报错。因变量可以为连续型变量或是二值变量。

② 数据准备阶段, 一般来说有两种方式生成类型数据(即形成空间分区或分层): 一是通过专业知识, 比如气温或降水等数据, 可以通过气候带的定义等方式来确定空间上的分区; 二是通过数据的分类算法来形成空间分区, 目前的分类算法包括多种, 如等间距法、分位数法、K-means等, 也包括多种基于专业领域知识的分类算法^[18], 分类的效果可通过地理探测器的 q 统计量来评价, q 值越大分区效果越好。

③ 类型数据不要求必需空间连续。如土地利用类型数据中, 同一类型的数据可能分布在不同的区域, 空间上不一定连续。

④ 空间面数据及点数据需要转换为软件输入数据: 对于不规则的面状数据(如行政区域), 首先通过格点化的方法生成空间上系统分布(如间距固定的规则格点)的样本点, 然后把每个格点所在位置的自变量及因变量的信息提取出来, 最后在把这些提取出的数据作为输入数据在软件中运行; 对于规则的面状数据, 如网格化数据(遥感数据), 可直接利用网格提取出点数据(如中心点), 然后与“不规则的面状数据”处理流程相同, 利用点数据把相关变量信息提取出来, 作为输入数据在软件中运行; 对于随机抽样或系统抽样获取的点状数据, 可直接利用此数据, 并输入数据在软件中运行。对于采样有偏的点状数据, 可以通过多种纠偏的方法处理后再计算。例如, 样本量较少, 但可以覆盖所有类型的情况, 可通过Sandwich方法(www.sssampling.org/sandwich)面插值后, 再利用前面所述“不规则的面状数据”的方法进一步处理; 样本量较少, 且不能覆盖所有类型区域, 则可通过Bshade方法(www.sssampling.org/bshade)进行插值后再做进一步的处理。

⑤ 每个分层中要求至少有两个样本点, 软件就可以计算。实际计算时, 样本量越大, 计算结果越准确。对于样本量的具体要求, 与研究对象的性质有关。如果研究对象在每个层内都是均质的, 则每个层内两个样本点就足够; 如果研究对象在每个层内存在较大的空间异质性, 则需要较大的样本量才能反映对象的真实性质, 文献中常把30个样本作为大样本。

⑥ q 值多大时才有意义? 就像其它模型(如线性回归中的 R^2)一样, 没有明确的定义, 不同专业有不同的要求。 q 值的物理含义是自变量 x 解释了 $100 \times q\%$ 的因变量 y 。

⑦ 如何选择影响因子? 与其它模型一样, 选择影响因子(自变量)需要根据专业知识或借助数据探索性分析。

⑧ 变量是否有正态分布的假设? 如果不对 q 值进行统计显著性检验, 此时 q 值仍有明确的物理意义, 但不必正态分布的假设。

⑨ 因变量是否要求为正值? 是否需要做规一化处理? 因变量不要求必需为正值, 也不需要做规一化。

⑩ 软件最大可容纳多少行的数据? 是否可以在WPS中运

行? 是否可以在 MAC 的 Excel 中运行? 软件最大可容纳数据量与 Excel 中 VBA 程序的限制有关, 最大行是 32767。软件目前版本不可以在 WPS 中运行, 也不可以在 MAC 的 Excel 中运行。^⑪ 软件是否需要借助 GIS 软件的处理? 对于数据准备过程中的格点化以及空间叠置分析中需借助于 GIS 软件。^⑫ 对于自变量的共线性, 如何处理? 地理探测器对于共线性免疫, 可不用处理。

参考文献(References)

- [1] Fischer M, Getis A. Handbook of Spatial Analysis. Springer, 2010.
- [2] Moran PAP. Notes on continuous stochastic phenomena. Biometrika, 1950, 37: 17-23.
- [3] Matheron G. Principles of geostatistics. Economic Geology, 1963, 58: 1246-1266.
- [4] Ripley B D. Modelling spatial patterns. Journal of the Royal Statistical Society, 1977, 39(2): 172-192.
- [5] Getis A, Ord J K. The analysis of spatial association by distance statistics. Geographical Analysis, 1992, 24(3): 189-206.
- [6] Anselin L. Local indicators of spatial association-LISA. Geographical Analysis, 1995, 27(2): 93-115.
- [7] Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods, 1997. 26: 1481-1496.
- [8] Christakos G. Random Field Models in Earth Sciences. CA, San Diego: Academic Press, 1992.
- [9] Cliff A D, Ord J K. Spatial Processes: Models and Applications. London: Pion, 1981.
- [10] Anselin L. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer, 1988.
- [11] Haining R. Spatial Data Analysis in Environmental and Social Sciences. Cambridge University Press, 1990.
- [12] Fotheringham A S, Brunsdon C, Charlton M E. Quantitative Geography: Perspectives on Spatial Data Analysis. London: SAGE Publications, 2000.
- [13] Haining R. Spatial Data Analysis: Theory and Practice. Cambridge University Press, 2003.
- [14] Cai Yunlong. Geographical sciences under the current sciences and society. Nature Magazine, 2013, 35(1): 30-39 [蔡运龙. 当代科学和社会视角下的地理学. 自然杂志, 2013, 35(1): 30-39.]
- [15] Cai Yunlong. Key concepts and studies in contemporary geographical science. Courses-Teaching Materials-Education, 2015, 35(11): 108-112. [蔡运龙. 当代地理学的关键概念和研究核心. 课程·教材·教法, 2015, 35(11): 108-112.]
- [16] Fan Jie. Draft of major function oriented zoning of China. Acta Geographica Sinica, 2015, 70(2): 186-201. [樊杰. 中国主体功能区划方案. 地理学报, 2015, 70(2): 186-201.]
- [17] Wang Zheng, Sun Yi. A simulation on regional coordination and industrial structure evolution oriented to the main functional areas of China. Scientia Geographica Sinica, 2013, 33(6): 641-648. [王铮, 孙艺. 中国主体功能区协调发展与产业结构演化. 地理科学, 2013, 33(6): 641-648.]
- [18] Li Xia, Yeh Anthony Gar-on, Liao Qifang. Case-based reasoning (CBR) for land use classification using radar images. Journal of Remote Sensing, 2004, 8(3): 246-253. [黎夏, 叶嘉安, 廖启芳. 利用案例推理(CBR)方法对雷达图像进行土地利用分类. 遥感学报, 2004, 8(3): 246-253.]
- [19] Li X, Yeh A. Zoning land for agricultural protection by the integration of remote sensing. GIS and cellular automata. Photogrammetric Engineering & Remote Sensing, 2001, 67(4): 471-477.
- [20] Wang J F, Li X H, Christakos G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. International Journal of Geographical Information Science, 2010, 24(1): 107-127.
- [21] Wang J F, Zhang T L, Fu B J. A measure of spatial stratified heterogeneity. Ecological Indicators, 2016, 67(2016): 250-256.
- [22] Wang J F, Christakos G, Hu M G. Modeling spatial means of surfaces with stratified non-homogeneity. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(12): 4167-4174.
- [23] Hu M G, Wang J F. A meteorological network optimization package using MSN theory. Environmental Modelling & Software, 2011, 26: 546-548.
- [24] Wang J F, Reis B Y, Hu M G, et al. Area disease estimation based on sentinel hospital records. PLoS ONE, 2011, 6(8): e23428.
- [25] Hu M G, Wang J F, Zhao Y. A B-SHADE based best linear unbiased estimation tool for biased samples. Environmental Modelling & Software, 2013, 48: 93-97.
- [26] Xu C D, Wang J F, Hu M G, et al. Interpolation of missing temperature data at meteorological stations using P-

附表 地理探测器更多典型应用

问题	区域	因变量Y	自变量X	研究发现和结论Y~X	文献
2008年汶川地震儿童死亡的影响因素	四川汶川	儿童地震死亡率	地震烈度、高程、坡度、距断裂距离、地貌、GDP、人口密度、房屋类型	地震烈度、房屋倒塌和坡度是5岁以下儿童地震死亡的主要因子。高程、断裂、地貌、人口密度、GDP与地震烈度和房屋倒塌交互作用增加了死亡率。坡度引起滑坡。	[23]
土壤抗生素空间分异的影响因素	山东寿光	土壤抗生素浓度	高程、温室面积、种植年、鸡粪数量、鸡粪类型、种植模式	① 种植模式是土壤抗生素空间差异的主导因子,例如黄瓜—黄瓜模式(种植黄瓜收获后再接着种黄瓜)的 q 值是辣椒—西瓜模式的3倍;② 种植模式与鸡粪量交互大大增加了土壤抗生素污染。政策建议:调整种植结构以缓解土壤抗生素污染。	[36]
伤寒副伤寒的影响因素	云南玉溪红塔区	伤寒副伤寒发病率	人均GDP、医疗机构数量、农民比例、人口密度、NDVI、地貌	社会因子对发病率影响最大。农民比例和人口密度分别解释了发病率的 $q = 78.3\%$ 和 $q = 73.8\%$;两者交互解释了 $q = 84.7\%$ 的发病率。	[50]
手足口病的影响因素	中国	0~9岁儿童手足口病发病率	月均温、月降水、相对湿度、人口密度、小学生密度、中学生密度、GDP、产业结构	0~9岁以下儿童密度主导了手足口病的空间变异,随后的控制因子是第三产业和GDP。第三产业和儿童密度交互解释了 $q = 42\%$ 的手足口病空间差异,GDP与第三产业交互 $q = 34\%$,它们均大于单因子作用。第三产业具有人口流动性强的特征。	[47]
城市景观的控制因素	福建厦门	森林景观连通度	天气、FMPI、高程、人口	人口密度与各种生态因子交互作用,线性或非线性,帮助增强了森林景观的连通性。	[38]
建设用地演变机理	南通市	建设用地变化速率	反映自然基底条件、经济发展水平、社会生活状况、政策制度环境方面15个指标	城乡建设用地变化形成机理差异显著,城镇用地增长受控于多种复杂因素,影响力最高的为城镇居民社会生活状况,而农村居民点用地变化则更多依赖于区位因素条件的影响,表现出其用地变化的先天依赖性与变化单一性。	[39]
分析2010年国家级经开区经济增长率的空间分异,其核心影响因素	全国	开发区GDP增长率	共15个因素。内部:优惠政策、土地和劳动力成本等;外部:母城的社会经济基础,交通可达性、产业结构、投资水平、资源束缚和城市环境质量	影响国家级经开区经济增长率的核心因素构成在三大地区间存在差异:中部为劳动力成本变化、开发区外贸水平变化、开发区优惠政策调整、交通可达性和城市产业支撑能力;西部为开发区优惠政策调整、城市产业支撑能力、开发区外贸水平变化、开发区产业相对规模和城市投资水平;东部为开发区优惠政策调整、劳动力成本变化、城市投资水平、城市经济发展水平和城市经济活力。	[56]
县域人均粮食占有量的形成机理	江西省	人均粮食占有量	人均粮食、种植结构、复种指数、灌溉指数、人均耕地面积、人均GDP、地均化肥、农机总动力、农民人均收入、农林牧渔劳动力	江西省县域人均粮食的空间格局主要受人均粮食占有量与人均耕地面积的影响;其次为农机总动力投入、农村居民人均纯收入、种植结构、复种指数;较低的经济水平使得全省人均粮食对自然地理条件表现出较大的依赖;4条样带县域人均粮食主导因子有区别。	[57]
血吸虫螺情地理因素探测	湖北省	活螺密度	DEM、水系、土壤类型和土壤质地、植被类型和植被覆盖度、土地利用	筛选出的地理环境指示因子及其适宜类型或范围,可应用于钉螺孳生地的遥感监测,减少钉螺调查采样的成本。	[75]
县域新增耕地影响因素?	湖北省	新增耕地数量	耕地面积、地形地貌、投资规模、建设规模、人均GDP、未利用地面积、经济密度、城镇化率、产业结构、人口密度等	2010年县域新增耕地主要影响因子为现有耕地面积、地貌类型、项目投资、建设规模和人均GDP。自然条件和土地整治政策为主要影响因素。	[43]
甘肃省县域城镇化地域分异成因	甘肃省	县域城镇化率	共17个指标:反映空间集聚、经济发展、社会发展、基础设施	各地区城镇化进程差异实质机理并不相同,除陇东地区受到基础设施建设的影响最为突出外,其他地区都显示为受经济增长及社会发展影响最大。	[44]

续附表

问题	区域	因变量Y	自变量X	研究发现和结论 Y~X	文献
人类活动和生态因素对城市地表温影响	厦门	城市林地温度	人口密度、林地调查、DEM、遥感影像等	树种和高程是城市地表温度的最主要的因子;人类活动与特定的生态因子交互作用线性或非线性地增加了地表温度;高程和主导树种交互在高温和低温区交互显著。该发现帮助改进城市冷岛空间格局的模拟预测。	[68]
水电站污染的来源分析	多瑙河	潜在生态风险(PERI)	沿河河流及沉积物采样点集;重金属	沉积物中生态风险指标危险度PERI因子排序: $Hg > Cd > As \sim Cu > Pb \gg Zn$ 。	[79]
工业CO ₂ 排放的因子分析	内蒙古	CO ₂ 排放量	GDP, 产业结构、城市化率、经济增长率人口和公路密度	决定力排序: 人口 > 城市化 > 人均GDP > 产业水平 > 能源强度; 存在交互作用, 最大的交互作用是GDP和公路密度解释CO ₂ 排放的71%。	[80]
乡村居民点的空间分布特征与优化	中国乡镇	乡村居民点人口密度	中国行政和交通图、县级经济社会统计	乡村居民点分布受传统和经济影响, 经济发展影响变得更加显著; 其他因素包括据公路距离、农业机械化、人均粮食产量、人均耕地面积、人口密度、高程、降水等。	[63]
蚂蚱种群动力学	内蒙古	蚂蚱密度	生物量等多生态因子	主导因子和交互作用, 为分异空间上蚂蚱疫病控制提供科学依据。	[73]
耕地占用与补偿的驱动力及其时变	江苏省	2005-2010年3个时段耕地占用/补偿	自然基底、经济发展水平、社会状况、政策制度共15个指标	区域经济发展水平与耕地占用和补充数量的空间分布一致性最高, 主要驱动因素随时间变化, 2005-2010年非经济因素驱动作用增强, 经济因素是主要驱动力。	[40]
贵州煤矿区植被指数变化影响因素	贵州煤矿区	2000-2014年NDVI	高程、坡度、坡向、土壤、多年平均降水量、多年平均气温和煤炭年产量	煤炭产量对矿区植被变化的解释力在7个因子中处于第3位, 在遵义、毕节和六盘水市, 其解释力分别排第4、第5和第5位, 对植被变化解释力最大的是自然环境因子中的降水、土壤和坡度。	[71]
大熊猫适宜区建模	四川雅安	大熊猫出现点	海拔、坡度、坡向、地形指数、距水源距离、植被类型、主食竹及距公路距离	各因子解释力(PD值)依次为: 主食竹(0.6936)、海拔(0.2745)、植被类型(0.2705)、坡度(0.1887)、地形指数(0.0582)、距道路距离(0.0315)、距水源距离(0.0126)、坡向(0.0026)。具有好的生态学可解释性。	[74]
中国老龄化空间差异的影响因子	中国县域	2000-2010年人口普查2283县人口老龄化	自然因素为11类生态区划; 人口因素2000年年龄; 经济增长、教育、卫生等	各因子解释力(PD值)依次为: 2000年老龄化程度($q = 0.63$) > 10年间步入老年人口比重($q = 0.51$) > 平均存活子女数($q = 0.01$); 自然区划($q = 0.37$), 人均受教育年限($q = 0.14$)。	[42]
乡村聚落格局分异因子	宁夏盐池风沙区	2005-2012年乡村聚落点	自然地形、区位条件及社会经济发展水平因素	乡村聚落空间布局分异主要受区位条件和自然地形的影响, 其中区位条件影响作用力最显著, 社会经济发展水平对其影响力不显著。	[53]
县域住宅价格空间差异的影响因子	中国县城	2014年中国2873县住宅平均单价	租房户比重、城镇流动人口数量、城镇在岗职工平均工资、房地产从业人口比重、土地平均出让价格	租房户比例、流动人口规模、住房支付能力、住房市场活跃度、土地成本是中国住宅价格差异的5个核心影响因素, 不同行政等级市场的影响因素作用强度各异。全国、省会、地级市、县不同层次各因子解释力不同。	[58]
村庄空间分布影响因素	中国	340万个自然村庄点数据来自2012年1:25万《中国电子地图》	高程、降水、城镇化率、交通线、固定资产投资、离中心城市距离、农村经济、人口密度等13个指标	村庄分布受到传统因素和经济发展双重因子的影响, 传统影响因素依然在发挥作用, 但经济发展的影响愈加明显。不同区域地形、水资源条件对村庄分布影响显著。交通条件、产业非农化、经济发展、农业现代化发展对乡村生活、生产空间的空间形态和分布模式产生剧烈影响。	[62]
PM _{2.5} 污染事件的影响因素	京津冀	PM _{2.5} 污染事件数	污染企业数、降水、地形、人口数量、地表覆盖类型、年均相对湿度、年均风速	污染企业总数(影响力为0.97, 下同)、降水量(0.93)、地形坡度(0.89)对PM _{2.5} 污染的影响显著高于其他风险因子。	[81]

续附表

问题	区域	因变量Y	自变量X	研究发现和结论 Y~X	文献
中国产业结构升级的控制因子	全国	产业结构层次系数	对外直接投资存量、投资流量、能源、资本、劳动力、消费水平、技术进步、国际贸易、经济水平	中国产业结构升级水平还受多种因素的影响,不同时期不同要素对产业结构升级水平的影响力存在差异。	[60]
蚂蚱疫病的适生区分析	内蒙古镶黄旗	蚂蚱密度	高程、坡向、坡度、植被类型、土地覆被	两因子交互对蚂蚱密度作用明显:高程与土壤沙化等;坡向与土壤沙化等。主导因子识别帮助改进模型预报和生态控制管理。	[72]
量化史前文化与河流的关系	河南省	遗址的密度\裴李岗时期\仰韶前期\仰韶后期\龙山时期	离河距离	克服前人研究史前聚落遗址分布与水系相关性研究方法的主观性。裴李岗文化时期、仰韶文化前期、仰韶文化后期、龙山文化时期,离水距离因子对遗址密度分布的决定力分别为39.5%、70.8%、73.0%和59.8%。	[67]
产业结构升级因子时空差异	中国	产业结构层次系数	GDP、电力消耗、资本、劳动力、消费水平、技术进步、贸易	2003-2008年和2010-2012年两个阶段,对外投资流量和存量对产业结构升级水平有很高的影响力。产业结构升级受到能源、资本、劳动力、消费、技术、贸易影响。	[61]
农村用地地理要素识别	环渤海	农村居民点用地1985、1995、2000	1985、1995、2000年土地利用、地理、经济社会共17个指标	农村居民点用地主要驱动力是农村经济发展水平、劳动力非农化程度、和距中心城区距离。	[45]
气温变化影响因素	东北	1960-2011各点平均气温数据时间线性倾向率	气候区划图、地貌区划图、植被区划图、土壤区划图、DEM、社会经济数据等	因子贡献量由大到小为植被类型、地貌类型、湿地率、GDP、土壤类型、人口密度、农田率、气候类型、海拔高度、森林率;社会因子增强了自然因子的贡献量,共同导致气温变化。	[70]
居民宜居满意度的影响机理	北京	7000个居民个人宜居综合评分	7000个居民问卷34个宜居要素及个体属性。	北京市居民宜居满意度受噪音、生活设施、人文环境、出行便利、自然环境、安全环境和休闲等7个因子控制。个体与家庭对宜居满意度影响存在交互增强作用。	[78]
植被物候影响因素	陕西	滤波前后NDVI	各类林地和农田共14类	q 值越大,说明图像各个地物类型间的区分性越大,图像的噪声信息越少。Whittaker滤波能够有效减少NDVI原始影像的噪声,提高影像辨识度。	[82]
城市休闲商务区RBD的成因	北京	RBD面积	交通、居民密度、游客密度、景区位置、地价	大型购物中心型RBD规模受旅客密度、本地居民密度影响。休闲型商业街类型RBD受游客密度控制。城市休闲区类型RBD受与景区距离、游客密度影响较强。	[65]
沙漠化驱动力	陕西	植被覆盖率	3类9个变量:自然条件、气候、人类因子	土壤类型、降水和风速是沙漠化的主导因子。自然要素而不是人为活动对沙漠化起更大的作用	[76]
土地扩张驱动力	北京	建成区扩张率	地形、距河距离、人口、社会经济、政策	多因子相互作用较单因子对于建成区扩张有更大的作用。	[37]
神经管畸形影响因子	山西	神经管畸形发病率	高程、土壤类型、土地利用、距河远近、距断裂线距离	在和顺 q 值:土壤(6%)>据煤区距离(4.7%)>距河流距离(4.2%);在原平 q 值:土壤(16.8%)>高程(16.3%)>断裂带(14%)。 Q 值随县不同,干预前后不同。	[48]
美国陆表切割度的影响因素识别	美国	陆表切割度	气候、坡度、地形、岩石、土壤、植被	岩性控制了山区的切割度;曲率控制了平原区的地表切割度;冰川控制了前冰川地区的陆表切割度。地理探测器提供了一个客观的框架用来揭示控制地球表层过程的影响因素。	[33]
沙区景观的控制因子分析	毛乌素沙区	NDVI;沙覆盖比例等	地形、水系、气候潜在植物生产力、风力、岩性、人文因子	气候潜在生产力、局地地形和水系是景观的主导因子;东部沙漠化以人类活动成因为主;西两沙漠化以自然因子为主。	[77]
残疾人就业的影响因素分析	中国	残疾人就业率	城市化、产业结构、人均GDP、旅客周转量、工资、社保率、医生数、医疗保险、失业险、人口密度	决定力 q 排序:医生数>旅客周转量>城市化率>产业结构>医疗保险>失业险>社保率。帮助政府制定因地制宜的残疾人就业政策。	[49]

- BSHADE. *Journal of Climate*, 2013, 26(19): 7452-7463.
- [27] Wang J F, Hu M G, Xu C D, et al. Estimation of citywide air pollution in Beijing. *PLoS ONE*, 2013, 8(1): e53400.
- [28] Wang J F, Haining R, Liu T J, et al. Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. *Environment and Planning A*, 2013, 45(10):2515-2534.
- [29] Wang J F, Li X H, Christakos G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *International Journal of Geographical Information Science*, 2010, 24(1): 107-127.
- [30] Wang J F, Hu Y. Environmental health risk detection with GeogDetector. *Environmental Modelling & Software*, 2012, 33: 114-115.
- [31] Cao F, Ge Y, Wang J F. Optimal discretization for geographical detectors-based risk assessment. *GIScience & Remote Sensing*, 2013, 50(1): 78-92.
- [32] <http://www.geodetector.org>
- [33] Luo W, Jasiewicz J, Stepinski T, et al. Spatial association between dissection density and environmental factors over the entire conterminous United States. *Geophysical Research Letters*, 2015, 43(2): 692-700.
- [34] Lou C R, Liu H Y, Li Y F, et al. Socioeconomic drivers of PM2.5 in the accumulation phase of air pollution episodes in the Yangtze River Delta of China. *International Journal of Environmental Research and Public Health*, 2016, 13: 928.
- [35] Liu Yansui, Yang Ren. Spatial characteristics and mechanisms of county level urbanization in China. *Acta Geographica Sinica*, 2012, 67(8): 1011-1020. [刘彦随, 杨忍. 中国县城城镇化的空间特征与形成机理. *地理学报*, 2012, 67(8): 1011-1020.]
- [36] Li X W, Xie Y F, Wang J F, et al. Influence of planting patterns on Fluoroquinolone residues in the soil of an intensive vegetable cultivation area in north China. *Science of the Total Environment*, 2013, 458-460: 63-69.
- [37] Ju H R, Zhang Z X, Zuo L J, et al. Driving forces and their interactions of built-up land expansion based on the geographical detector: A case study of Beijing, China. *International Journal of Geographical Information Science*, 2016 (11): 1-20. <http://dx.doi.org/10.1080/13658816.2016.1165228>.
- [38] Ren Y, Deng L Y, Zuo S D, et al. Geographical modeling of spatial interaction between human activity and forest connectivity in an urban landscape of southeast China. *Landscape Ecology*, 2014, 29(10): 1741-1758. DOI 10.1007/s10980-014-0094-z.
- [39] Cai Fangfang, Pu Lijie. Spatial-temporal characteristics and formation mechanism of urban-rural construction land in Nantong city. *Resources Science*, 2014, 36(4): 731-740. [蔡芳芳, 濮励杰. 南通市城乡建设用地演变时空特征与形成机理. *资源科学*, 2014, 36(4): 731-740.]
- [40] Chen Changling, Zhang Quanqing, Lu Xiao, et al. Analysis on spatial-temporal characteristics and driving mechanisms of cropland occupation and supplement in Jiangsu province. *Economic Geography*, 2016, 36(4): 155-163. [陈昌玲, 张全景, 吕晓, 等. 江苏省耕地占补过程的时空特征及驱动机理. *经济地理*, 2016, 36(4): 155-163.]
- [41] Li Tao, Liao Heping, Zu Yuanheng, et al. Spatial disequilibrium and its formation mechanism of farmland conversion in Chongqing. *Journal of Nature Resources*, 2016, 31(11): 1844-1857. [李涛, 廖和平, 褚远恒, 等. 重庆市农地非农化空间非均衡及形成机理. *自然资源学报*, 2016, 31(11): 1844-1857.]
- [42] Wang Lucang, Wu Rongwei, Liu Haimeng, et al. Spatial patterns and regional differences of population ageing in China based on the county scale. *Progress in Physical Geography*, 2016, 35(8): 921-931. [王录仓, 武荣伟, 刘海猛, 等. 县域尺度下中国人口老龄化的空间格局与区域差异. *地理科学进展*, 2016, 35(8): 921-931.]
- [43] Wei Fengjuan, Li Jiangfeng, Liu Yanzhong. Spatial-temporal characteristics and impact factors of newly increased farmland by land consolidation in Hubei province at county level. *Transactions of the Chinese Society of Agricultural Engineering*, 2014, 30(14): 267-275. [魏凤娟, 李江风, 刘艳中. 湖北县域土地整治新增耕地的时空特征及其影响因素分析. *农业工程学报*, 2014, 30(14): 267-275.]
- [44] Yang Bo, Shi Peiji. Geographical features and formation mechanisms of county level urbanization in Gansu Province. *Arid Land Geography*, 2014, 37(4): 838-845. [杨勃, 石培基. 甘肃省县城城镇化地域差异及形成机理. *干旱区地理*, 2014, 37(4): 838-845.]
- [45] Yang Ren, Liu Yansui, Long Hualou, et al. Spatial-temporal characteristics of rural residential land use change and spatial directivity identification based on grid in the Bohai Rim in China. *Geographical Research*, 2015, 34(6): 1077-1087. [杨忍, 刘彦随, 龙花楼, 等. 基于格网的农村居民点用地时空特征及空间指向性的地理要素识别: 以环渤海地区为例. *地理研究*, 2015, 34(6): 1077-1087.]
- [46] Hu Y, Wang J F, Ren D, et al. Geographical-detector-based risk assessment of the under-five mortality in the 2008

- Wenchuan Earthquake, China. PLoS ONE, 2011, 6(6): e21427.
- [47] Huang J X, Wang J F, Bo Y C, et al. Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. International Journal of Environmental Research and Public Health, 2014, 11: 3407-3423.
- [48] Liao Y L, Zhang Y, He L, Wang J F, et al. Temporal and spatial analysis of neural tube defects and detection of geographical factors in Shanxi Province, China. PLoS ONE, 2016, 11(4): e0150332. doi:10.1371/journal.pone.0150332.
- [49] Liao Y L, Wang J, Du W, et al. Using spatial analysis to understand the spatial heterogeneity of disability employment in China. Transactions in GIS, 2016. doi: 10.1111/tgis.12217.
- [50] Wang J F, Wang Y, Zhang J, et al. Spatiotemporal transmission and determinants of typhoid and paratyphoid fever in Hongta District, China. PLoS Neglected Tropical Diseases, 2013, 7(3): e2112.
- [51] Chen Yebin, Li Weihong, Huang Yuxing, et al. Spatio-temporal spreading features and the influence factors of Dengue fever in downtown Guangzhou. Tropical Geography 2016, 36(5): 767-775. [陈业滨, 李卫红, 黄玉兴, 等. 广州市登革热时空传播特征及影响因素. 热带地理, 2016, 36(5): 767-775.]
- [52] Tao Haiyan, Pan Zhongzhe, Pang Maolin, et al. Mixing spatial-temporal transmission patterns of metropolis dengue fever: A case study of Guangzhou, China. Acta Geographic Sinica, 2016, 71(9): 1653-1662. [陶海燕, 潘中哲, 潘茂林, 等. 2016. 广州大都市登革热时空传播混合模式. 地理学报, 2016, 71(9): 1653-1662.]
- [53] Wang Manman, Wu Xiuqin, Wu Bin, et al. Evolution analysis of spatial pattern of rural settlements in sandy area of northern Yanchi. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(8): 260-271. [王曼曼, 吴秀芹, 吴斌, 等. 盐池北部风沙区乡村聚落空间格局演变分析. 农业工程学报, 2016, 32(8): 260-271.]
- [54] Ni Shuhua. Spatial statistics and its application to the field of public health. Editorial Department of Shantou University Journal, 2014, 29(4): 61-67. [倪书华. 空间统计学及其在公共卫生领域中的应用. 汕头大学学报(自然科学版), 2014, 29(4): 61-67.]
- [55] Tan J T, Zhang P Y, Lo K V, et al. The urban transition performance of resource-based cities in Northeast China. Sustainability, 2016, 8: 1022. doi:10.3390/su8101022.
- [56] Ding Yue, Cai Jianming, Ren Zhoupeng, et al. Spatial disparities of economic growth rate of China's National-level ETDZs and their determinants based on geographical detector analysis. Progress in Geography, 2014, 33(5): 657-666. [丁悦, 蔡建明, 任周鹏, 等. 基于地理探测器的国家级经济技术开发区经济增长率空间分异及影响因素. 地理科学进展, 2014, 33(5): 657-666.]
- [57] Hu Dan, Shu Xiaobo, Yao Bo, et al. The evolvement of spatio-temporal pattern of per capita grain possession in counties of Jiangxi province. Areal Research and Development, 2014, 33(4): 157-162. [胡丹, 舒晓波, 尧波, 等. 江西省县域人均粮食占有量的时空格局演变. 地域研究与开发, 2014, 33(4): 157-162.]
- [58] Wang Shaojian, Wang Yang, Lin Xueqin, et al. Spatial differentiation patterns and influencing mechanism of housing prices in China: Based on data of 2872 counties. Acta Geographica Sinica, 2016, 71(8): 1329-1342. [王少剑, 王洋, 蔺雪芹, 等. 中国县域住宅价格的空间差异特征与影响机制. 地理学报, 2016, 71(8): 1329-1342.]
- [59] Xu Qirong, Zheng Xinqi. Analysis of influencing mechanism of urban growth using geographical detector. Acta Geodaetica et Cartographica Sinica, 2015, 44(S0): 96-101. [徐秋蓉, 郑新奇. 一种基于地理探测器的城镇扩展影响机理分析法. 测绘学报, 2015, 44(S0): 96-101.]
- [60] Yu Jiageng, Ye Shikang. Direct investment to foreign countries and industrial structure update in China. Commercial Economic Studies, 2014, 34: 127-128. [俞佳根, 叶世康. 空间视角下中国对外直接投资与产业结构升级水平研究. 商业经济研究, 2014, 34: 127-128.]
- [61] Cui Rimin, Yu Jiageng. Spatial perspective of the direct investment and industrial structure update in China. Fujian Forum (The Humanities & Social Sciences Monthly), 2015(2): 26-33. [崔日明, 俞佳根. 基于空间视角的中国对外直接投资与产业结构升级水平研究. 福建论坛(人文社会科学版), 2015(2): 26-33.]
- [62] Yang Ren, Liu Yansui, Long Hualou, et al. Spatial distribution characteristics and optimized reconstructing analysis of rural settlement in China. Scientia Geographica Sinica, 2016, 36(2): 170-179. [杨忍, 刘彦随, 龙花楼, 等. 中国村庄空间分布特征及空间优化重组解析. 地理科学, 2016, 36(2): 170-179.]
- [63] Yang R, Xu Q, Long H L. Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization. Journal of Rural Studies, 2016. <http://dx.doi.org/10.1016/j.jrurstud.2016.05.013>.
- [64] Wang X G, Xi J C, Yang D Y, et al. Spatial differentiation of rural touristization and its determinants in China: A geo-detector-based case study of Yesanpo scenic area. Journal of Resources and Ecology, 2016, 7(6): 464-471.

- [65] Zhu He, Liu Jiamin, Tao Hui, et al. Temporal-spatial pattern and contributing factors of urban RBDs in Beijing. *Acta Geographica Sinica*, 2015, 70(8): 1215-1228. [朱鹤, 刘家明, 陶慧, 等. 北京城市休闲商务区的时空分布特征与成因. *地理学报*, 2015, 70(8): 1215-1228.]
- [66] Zhu H, Liu J M, Chen C, et al. A spatial-temporal analysis of urban recreational business districts: A case study in Beijing, China. *Journal of Geographical Sciences*, 2016, 25(12): 1521-1536.
- [67] Bi Shuoben, Ji Han, Chen Changchun, et al. Application of geographical detector in human-environment relationship study of prehistoric settlements. *Progress in Geography*, 2015, 34(1): 118-127. [毕硕本, 计晗, 陈昌春, 等. 地理探测器在史前聚落人地关系研究中的应用与分析. *地理科学进展* 34(1): 118-127.]
- [68] Ren Y, Deng L Y, Zuo S D, et al. Quantifying the influences of various ecological factors on land surface temperature of urban forests. *Environmental Pollution*, 2016, 216: 519-529. <http://dx.doi.org/10.1016/j.envpol.2016.06.004>.
- [69] Li J, Zhu Z W, Dong W J. A new mean-extreme vector for the trends of temperature and precipitation over China during 1960-2013. *Meteorology & Atmospheric Physics*, 2016: 1-10. doi: 10.1007/s00703-016-0464-y.
- [70] Yu Jia, Liu Jiping. Quantitative analysis with geographical detector on the influence factor of temperature variation in Northeast China. *Hubei Agriculture Sciences*, 2015, 54(19): 4682-4687. [于佳, 刘吉平. 基于地理探测器的东北地区气温变化影响因素定量分析. *湖北农业科学*, 2015, 54(19): 4682-4687.]
- [71] Li Jungang, Yan Qingwu, Xiong Jibin, et al. Variation of vegetation index in coal mining areas in Guizhou province and its affecting factor. *Journal of Ecology and Rural Environment*, 2016, 32(3): 374-378. [李俊刚, 闫庆武, 熊集兵, 等. 贵州省煤矿区植被指数变化及其影响因子分析. *生态与农村环境学报*, 2016, 32(3): 374-378.]
- [72] Shen J, Zhang N, Gexi G D R, et al. Construction of a GeogDetector-based model system to indicate the potential occurrence of grasshoppers in Inner Mongolia steppe habitats. *Bulletin of Entomological Research*, 2015, 105: 335-346.
- [73] Zhang N, Jiang Y C, Liu C Y, et al. A cellular automaton model for grasshopper population dynamics in Inner Mongolia steppe habitats. *Ecological Modelling*, 2016, 329: 5-17.
- [74] Liao Ying, Wang Xinyuan, Zhou Junming. Suitability assessment and validation of giant panda habitat based on geographical detector. *Journal of Geoinformation Science*, 2016, 18(6): 767-778. [廖颖, 王心源, 周俊明. 基于地理探测器的大熊猫生境适宜度评价模型及验证. *地球信息科学学报*, 2016, 18(6): 767-778.]
- [75] Tong Laga, Xu Xinliang, Fu Ying, et al. Impact of environmental factors on snail distribution using geographical detector model. *Progress in Geography*, 2014, 33(5): 625-635. [通拉嘎, 徐新良, 付颖, 等. 地理环境因子对螺情影响的探测分析. *地理科学进展*, 2014, 33(5): 625-635.]
- [76] Du Z, Xu X, Zhang H, et al. Geographical detector-based identification of the impact of major determinants on aeolian desertification risk. *PLoS ONE*, 2016, 11(3): e0151331. doi: 10.1371/journal.pone.0151331.
- [77] Liang P, Yang X P. Landscape spatial patterns in the Maowusu (Mu Us) sandy land, northern China and their impact factors. *Catena*, 2016, 145: 321-333.
- [78] Zhan Donsheng, Zhang Wenzhong, Yu Jianhui, et al. Analysis of influencing mechanism of residents' livability satisfaction in Beijing using geographical detector. *Progress in Geography*, 2015, 34(8): 966-975. [湛东升, 张文忠, 余建辉, 等. 基于地理探测器的北京市居民宜居满意度影响机理. *地理科学进展*, 2015, 34(8): 966-975.]
- [79] Todorova Y, Lincheva S, Yotinov I, et al. Contamination and ecological risk assessment of long-term polluted sediments with heavy metals in small hydropower cascade. *Water Resources Management*, 2016, 30: 4171-4184.
- [80] Wu R N, Zhang J Q, Bao Y H, et al. Geographical detector model for influencing factors of industrial sector carbon dioxide emissions in Inner Mongolia, China. *Sustainability*, 2016, 8(2): 149.
- [81] Zhou Lei, Wu Jianjun, Jia Ruijing, et al. Investigation of temporal-spatial characteristics and underlying risk factors of PM_{2.5} pollution in Beijing-Tianjin-Hebei Area. *Research of Environmental Sciences* 2016, 29(4): 483-493. [周磊, 武建军, 贾瑞静, 等. 京津冀PM_{2.5}时空分布特征及其污染风险因素. *环境科学研究*, 2016, 29(4): 483-493.]
- [82] Zhang Han, Ren Zhiyuan. Phenological characteristics of vegetable in Shaanxi province detected using Whittaker filter. *Journal of Desert Research*, 2015, 45(4): 901-906. [张晗, 任志远. 基于Whittaker滤波的陕西省植被物候特征. *中国沙漠*, 2015, 45(4): 901-906.]
- [83] Xie Shuai, Liu Shibin, Duan Jianbo, et al. Spatial distribution characteristics of OSDS registered users and its influencing factors. *Journal of Geo-information Science*, 2016, 18(10): 1332-1340. [谢帅, 刘士彬, 段建波, 等. 2016. OSDS注册用户空间分布特征及影响因素分析. *地球信息科学学报*, 2016, 18(10): 1332-1340.]

Geodetector: Principle and prospective

WANG Jinfeng^{1,2}, XU Chengdong¹

(1. State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Spatial stratified heterogeneity is the spatial expression of natural and socio-economic process, which is an important approach for human to recognize nature since Aristotle. Geodetector is a new statistical method to detect spatial stratified heterogeneity and reveal the driving factors behind it. This method with no linear hypothesis has elegant form and definite physical meaning. Here is the basic idea behind Geodetector: assuming that the study area is divided into several subareas. The study area is characterized by spatial stratified heterogeneity if the sum of the variance of subareas is less than the regional total variance; and if the spatial distribution of the two variables tends to be consistent, there is statistical correlation between them. Q-statistic in Geodetector has already been applied in many fields of natural and social sciences which can be used to measure spatial stratified heterogeneity, detect explanatory factors and analyze the interactive relationship between variables. In this paper, the authors will illustrate the principle of Geodetector and summarize the characteristics and applications in order to facilitate the using of Geodetector and help readers to recognize, mine and utilize spatial stratified heterogeneity.

Keywords: spatial stratified heterogeneity; Geodetector; q -statistic; category variable