

文章编号:1673-6338(2017)01-0079-06

朝鲜人口统计数据空间化模拟及影响因子分析

史婷婷^{1,2}, 杨晓梅^{1,3}, 蓝荣钦⁴

(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.中国中医科学院中药资源中心,北京 100700; 3.江苏省地理信息资源开发与利用协同创新中心,江苏南京 210023;
4.信息工程大学,河南 郑州 450001)

摘要:朝鲜人口统计数据空间化是解决朝鲜统计数据与自然要素数据融合分析的重要途径。基于朝鲜市郡级人口普查数据,将GIS空间分析技术与统计学方法相结合,分析了朝鲜人口密度与空间因子的关系,采用多元回归的方法建立了朝鲜人口密度空间化模型,在GIS平台中实现了朝鲜人口密度的空间格网化模拟,并利用地理探测器对影响朝鲜人口密度空间分布因素的决定力进行了有效探测。结果表明,多元回归模型拟合精度达到0.769,生成的栅格人口密度数据与朝鲜三级行政区人口统计数据保持一致。同时,影响该地区人口密度的因子依次为道路网密度、居民点密度、居民地指数、海拔、坡度和耕地指数。

关键词:人口密度;空间模拟;格网化;地理探测器;朝鲜

中图分类号:P208;K901.3

文献标识码:A

DOI 编码:10.3969/j.issn.1673-6338.2017.01.016

Spatial Simulation and Influence Factors Detection of Population Density in North Korea

SHI Tingting^{1,2}, YANG Xiaomei^{1,3}, LAN Rongqin⁴

(1.State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;
2.National Resource Center for Chinese Material Medical, China Academy of Chinese Medical Science, Beijing 100700, China;
3.Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 4.Information Engineering University, Zhengzhou 450001, China)

Abstract: Spatialization of North Korean demographic data is an important way to solve the fusion analysis of North Korea's statistical data and natural elements. Based on the population census data of city-county level in the North Korea, the relationship between population density and spatial factors was firstly analyzed, and a spatial model of population density was established by means of multiple variables regression analysis and GIS technique, and then a rasterized population density map for North Korea was calculated. The geographical detector tool was used to analyze the determinant of population density. The result shows that the correlation coefficient between population density and spatial factors is 0.769. The resulting grid population density data is consistent with North Korea demographic data and the correlation is satisfied. Meantime, the decision forces of potential determinants are road network density, residential density, residential index, elevation, slope and farmland index.

Key words: population density; spatial simulation; grid; geographical detector; North Korea

人口是研究区域可持续发展的一个重要因素,而人口的分布是人口数据在空间上的表现形式^[1]。通常人口数据是以行政区划为单元,通过人口统计、普查以及逐级汇总得到的。这种统计

型的人口数据只能获得行政区划单元内的人口总量,不仅不能准确地反映行政区划内人口分布的实际状况,而且不利于进行各种空间分析。因此,研究人口空间分布及其影响因素,不仅能外在把

收稿日期:2016-06-07;修回日期:2016-09-22。

基金项目:国家863计划项目(2012AA121201;2013AA122901);国家科技支撑计划项目(2014BAL01B01)。

作者简介:史婷婷(1986-),女,河南南阳人,助理研究员,博士,主要从事地理空间分析模型研究。E-mail:shitt@lreis.ac.cn

通讯作者:杨晓梅,女,研究员。E-mail:yangxm@lreis.ac.cn

握地域人口分布特征,还能内在挖掘导致地域人口变化的驱动力。

已有不少学者对人口密度空间化问题做了大量研究^[2-13],但现有研究侧重于国内,而对我国具有很大战略意义的周边国家的研究尚属空白,且对影响人口分布内在机理的研究较少。因此,研究朝鲜地区人口密度分布的空间格局及其影响因素,对揭示朝鲜人口变化规律,辨识朝鲜人口变化的关键驱动因子,进而建立朝鲜人口综合评价指标和制订我国对外发展政策具有重要意义。

基于上述考虑,以朝鲜作为研究区域,根据研究区的特点和数据获取情况,选取自然因素中的地形和土地利用两个指标,分别用朝鲜数字高程(DEM)和土地利用指数来度量;经济因素选取交通指标,利用道路网密度来度量;人文因素选取居民点指标,利用居民点密度进行量度。通过定量分析这些因素与人口分布的相关性,继而建立人口密度空间化模型,以实现人口数据的空间化。同时,利用地理探测器方法对朝鲜人口密度空间分布的内在机理进行定量分析,以期对相关领域研究和工作提供有益的参考和依据。

1 研究区和数据来源

1.1 研究区概况

朝鲜位于东北亚朝鲜半岛北部,北纬 $37^{\circ}\sim 43^{\circ}$,东经 $124^{\circ}\sim 131^{\circ}$,面积为 $120\,540\text{ km}^2$ 。朝鲜地势北高南低、东高西低,平均海拔 440 m ,山地多平原少,山地约占国土面积的 80% ,平原大部分在西部和南部地区,中东部是山地丘陵区,高原与山地间盆地较多,西、南海岸是丘陵低地区。朝鲜的耕地大多位于朝鲜南部,耕地面积有限,只占总面积的 14.4% ^[14]。截止2016年7月,朝鲜总人口达 $25\,115\,311$ 人^[15]。

1.2 数据来源与预处理

数据包括:1) DEM数据,来源于空间分辨率为 30 m 的ASTER GEDM数据;2)土地利用数据,由Landsat TM遥感影像解译获得,依据中国科学院资源环境信息数据库的土地利用/覆被分类标准,将朝鲜土地利用类型分为建设用地、耕地、林地、草地、水域和未利用地;3)市郡级行政区划数据,来源于DIVA-GIS国家数据^[16];4)道路和居民点数据,来源于Geofabrik网站^[17];5)人口数据,来源于朝鲜2008年人口普查数据。

数据预处理主要在ArcGIS中完成,包括:1)对来源不同的行政区划、道路和居民点矢量数

据进行预处理,通过扫描矢量化、格式转换、坐标转换、几何校正及标准化处理等步骤,形成格式、坐标和属性均统一的朝鲜基础地理数据;2)将各种类型的土地利用矢量数据转化为 $1\text{ km}\times 1\text{ km}$ 分辨率的栅格土地利用数据,每个类型用1个栅格数据层表示;3)对DEM进行投影转换,并重采样至 $1\text{ km}\times 1\text{ km}$ 的空间分辨率,同时生成坡度图和坡向图;4)统计各市郡的平均高程、平均坡度和平均坡向;5)计算各市郡的不同土地利用类型指数,即某市郡某土地利用类型的面积占该市郡总面积的百分比;6)计算各市郡级行政区内道路总长度,求出道路网密度;7)计算各市郡的居民点密度;8)利用朝鲜人口普查统计的各三级行政区总人口数,得出各行政区的平均人口密度。

2 研究方法

2.1 人口密度空间模拟基本思路

根据朝鲜整个区域的特点,选取3类影响该区域人口密度分布的空间因子:1)地形因子,包括海拔高度、坡度和坡向;2)交通因子,即道路网密度;3)土地利用因子,包括6种土地利用指数。选取能够体现研究区特点的若干空间因子,在GIS平台中求得其在每个市郡中的平均值;采用相关分析方法考查每个空间因子与人口密度的相关关系,找出影响人口密度分布的主导因子;分别将这些主导因子与人口密度进行回归分析,具体考查每个因子对人口分布密度的影响;根据市郡尺度上的数据,采用多元回归方法建立人口密度与空间因子的回归方程;根据这一方程,基于 $1\text{ km}\times 1\text{ km}$ 栅格数据,得到每个栅格的人口数量,即栅格尺度上的人口密度,从而实现人口密度的栅格化模拟,并利用GIS平台中的图层叠加技术实现人口密度数据的空间化。

2.2 地理探测器分析法

地理探测器分析方法最初应用于地方性疾病风险评价及相关地理影响因素的研究^[18]。该方法的主要思想是:地理现象的发生总是存在于特定的空间位置上,影响其发展变化的环境因素往往表现出地理空间变异的特征。如果某个环境因素和地理现象的变化在空间上具有统计学的显著意义,则说明这种环境因素对地理现象的发生有决定作用^[18]。

地理探测器模型基于空间分异理论,可以有效获得因子变量和结果变量的相关性。该方法的最大优势在于没有过多的假设条件,可以有效克

服传统统计分析方法处理类别变量的局限性,被逐渐应用到地理学科研究中的各个领域^[19-21]。

探测朝鲜人口密度影响因子的计算模型为

$$F_D = 1 - \frac{1}{n\sigma^2} \sum_{i=1}^m (n_{D,i} \times \sigma_{D,i}^2). \quad (1)$$

式中: F_D 为人口密度影响因子 D 的探测力指标; $n_{D,i}$ 为在影响因子 D 的子区域 i 内的样本数; n 为整个区域样本数; m 为子区域个数; σ^2 为整个区域人口密度的离散方差; $\sigma_{D,i}^2$ 为子区域人口密度的方差。 $F_D \in [0, 1]$, F_D 值越大,说明影响因子 D 对人口密度空间分布的控制力越大。选取高程、坡度、坡向、土地利用、居民点密度及道路网密度等作为地理探测器分析的探测因子,分别探测各个因子对朝鲜人口密度空间格局的影响作用。

3 结果与分析

3.1 人口密度与环境因素之间的关系

朝鲜人口分布特征是在一定的自然和社会经济条件下形成的。其中自然因素如地形是朝鲜人口分布的基础条件;社会经济和生成布局的变化对朝鲜的人口分布也起着一定的积极作用。基于以上考虑,选取海拔、坡度、坡向等地形因素,林地指数、草地指数、耕地指数、水域指数、居民地指数和建筑用地指数等土地利用类型因素,以及道路网密度等交通因素,采用相关分析中 Pearson 相关系数来分析各因素与朝鲜人口密度之间的关系,分析结果如表 1 所示。

表 1 2008 年朝鲜人口密度与环境影响因素的相关性分析结果

环境影响因素	市郡个数	Pearson 相关性	显著性(双侧)
平均坡向	172	-0.096	0.209
平均坡度	172	-0.459**	0.000
平均海拔	172	-0.407**	0.000
居民点密度	172	0.629**	0.000
道路网密度	172	0.663**	0.000
居民地指数	172	0.645**	0.000
耕地指数	172	0.403**	0.000
林地指数	172	-0.211**	0.005
草地指数	172	-0.049	0.525
水域指数	172	0.203	0.008
建筑用地指数	172	0.115	0.132

注:**表示在 0.01 水平(双侧)上显著相关。

从表 1 可以看出,显著相关的因素有坡度、海拔、居民点密度、道路网密度、居民地指数、耕地指数和林地指数等 7 个因子。从影响的程度上看,居民点密度、道路网密度、耕地指数和居民地指数对朝鲜人口密度的影响较大;林地指数相对较小;其他一些因素的影响不显著。从影响的方向上

看,坡度、海拔和林地对朝鲜人口密度分布呈负相关;而居民点密度、道路网密度、居民地指数和耕地指数与朝鲜人口密度分布呈正相关。

通过绘制各影响因子和人口密度的散点图,来分析人口密度和各影响因子的关系。用 R^2 表示人口密度与各影响因子之间的相关系数的平方。

地形因素中,从平均坡度和人口密度的关系来看(图 1(a)),朝鲜各市郡的人口密度集中在平均坡度 25° 以下的地区,说明人口主要分布在地形起伏程度较低的地区;而地形起伏程度高的地区人口分布稀少。从平均海拔和人口密度的关系看(图 1(b)),朝鲜各市郡人口密度随着海拔的增高而降低,并且主要集中在海拔 1 km 以下;海拔超过 1 km 的地区人口分布稀少。坡度和海拔的共同影响,使得地势低且平坦的地区利于工农业的发展,易于形成人口稠密区。

交通因素中,从道路网密度与人口密度的关系看(图 2),道路网密度越大的区域,人口密度越大,即经济较发达地区的人口分布较为稠密。

土地利用因素中(图 3),随着居民地指数和耕地指数的增加,人口密度增大,说明土地利用信息能很好地反应人口的微观分布。

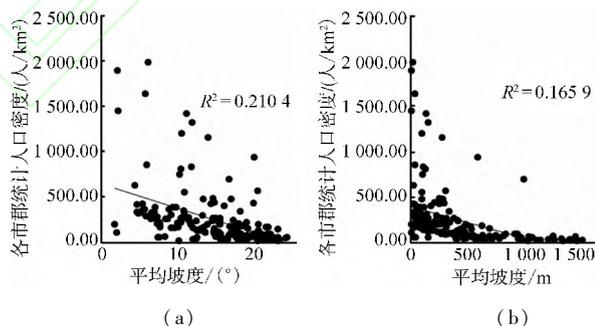


图 1 地形因子与各市郡统计人口密度的关系

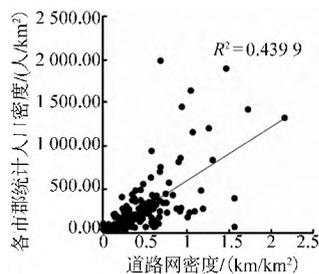


图 2 各市郡道路网密度与人口密度的关系

3.2 人口密度空间化模型建立及验证

从相关性分析可以发现,朝鲜人口密度受多种空间因素的影响,仅依赖单一因素无法很好地模拟朝鲜人口密度的空间分布。因此选择相关性强且具有代表性的多个影响因素(坡度、海拔、居民点密度、道路网密度、居民地指数和耕地指数)

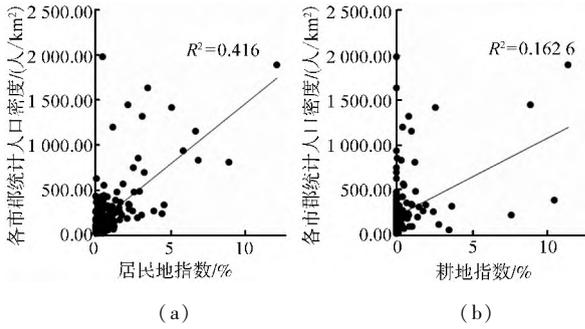


图3 土地利用因子与人口密度的关系

作为建模参数,以朝鲜各市郡的平均人口密度为因变量,选择不同组合的建模参数作为自变量,进行多元回归分析。采用逐步回归方法对朝鲜人口密度空间化进行建模,得到不同的人口数据模型,建模的结果如表2所示。

表2 回归模型验证结果

模型	相关性 R	相关平方 R ²	调整相关平方 R ²	标准估计误差
1	0.663	0.440	0.437	249.199 289 0
2	0.737	0.542	0.537	225.898 537 9
3	0.752	0.566	0.558	220.749 996 0
4	0.769	0.591	0.581	214.935 815 1

模型1中预测变量的多元回归方程由常量和因变量道路网密度构成;模型2中预测变量的多元回归方程由常量和因变量道路网密度及居民地指数构成;模型3中预测变量的多元回归方程由常量和因变量道路网密度、居民地指数及居民点密度构成;模型4中预测变量的多元回归方程由常量和因变量道路网密度、居民地指数、居民点密度及坡度构成。

对4个模型的各市郡预测人口密度与实际统计人口密度进行相关性分析发现,模型4的相关性最高,达到0.769。比较4个模型的标准估计误差发现,模型4的标准估计误差最小,说明模型4的精度最高。模型4中通过显著性检验的因子为道路网密度、居民地指数、居民点密度和坡度,说明这4个因子对朝鲜人口密度的影响最大。由模型4得到多元回归方程为

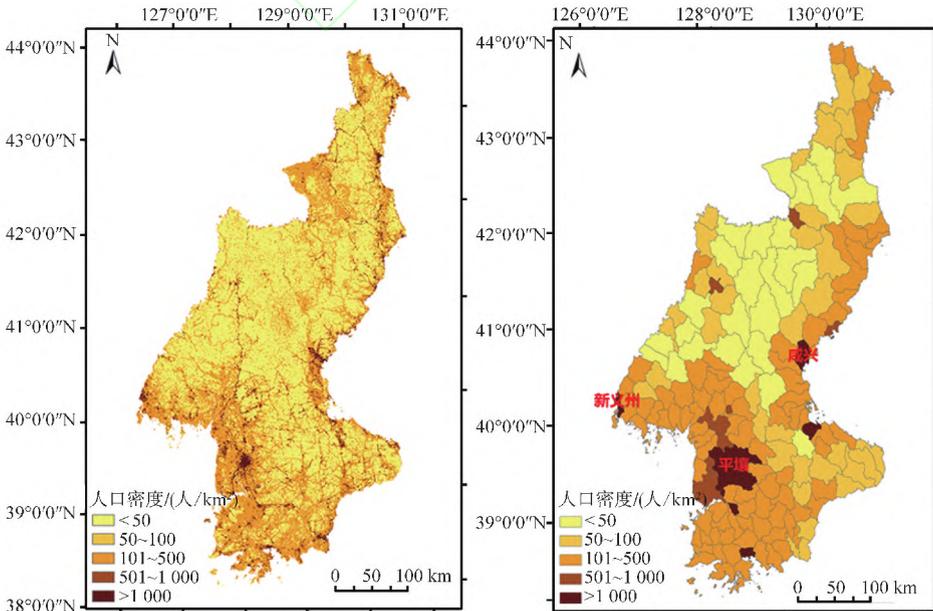
$$P = 213.064 + 183.840X_1 + 72.119X_2 + 1939.621X_3 - 11.091X_4 \quad (1)$$

式中:P为人口密度;X₁、X₂、X₃和X₄分别为道路网密度、居民地指数、居民点密度和坡度。

3.3 朝鲜人口密度的空间模拟

将道路网密度、居民点密度、居民地指数和坡度格网化到1 km×1 km网格上。根据基于市郡级行政单元建立的人口密度空间化模型(模型4),利用ArcGIS平台的图层叠加工具,计算得到朝鲜区域每个1 km×1 km栅格上的人口数量,生成基于栅格的朝鲜人口密度图(图4(a))。根据朝鲜人口统计数据得到朝鲜每个三级行政区的总人口,将每个三级行政区的总人口除以其面积,得到每个行政区内的平均人口,如图4(b)所示。

由图4可知,从地理位置上看,朝鲜人口集中分布在西南沿海平原地区,少量分布在东部沿海区及北部与中国接壤地区,中、北部高海拔地区几乎无人居住。从行政区划上看,朝鲜人口最稠密的地方集中在首都平壤,其次是西部的新义州及东海岸的咸兴。



(a) 基于栅格的朝鲜人口密度图

(b) 基于市郡级行政区划的人口密度图

图4 朝鲜人口密度图

3.4 朝鲜人口密度空间特征成因分析

朝鲜人口密度与地形、土地利用和交通因素有着不同程度的相关关系,但这种关系很难定量地揭示人口密度空间格局的形成机理。利用地理

探测器方法对内在机理进行探测分析。首先,对各影响因子根据自然断点法进行离散化,离散化结果如图 5 所示;其次,根据式(1)计算出各因子对人口密度空间分布的决定力 F_d 值,见表 3。

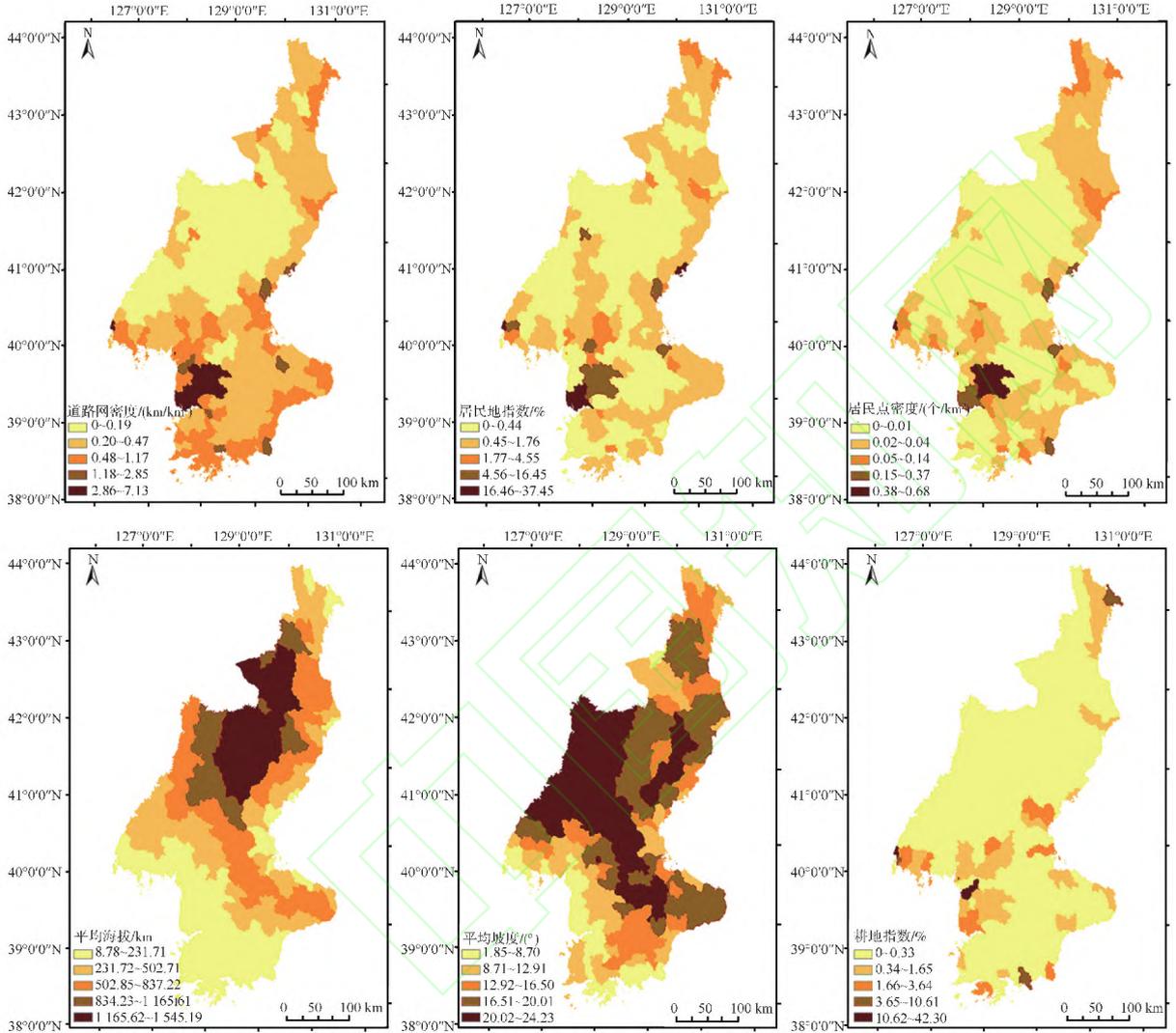


图 5 离散化地理探测因子的空间分布

表 3 朝鲜人口密度空间分布影响因子探测结果

探测指标	F_d 值	决定力排序
海拔	0.182 529	4
坡度	0.181 185	5
道路网密度	0.455 115	1
居民点密度	0.414 069	2
耕地指数	0.180 58	6
居民地指数	0.388 679	3

海岸和北部与中国接壤区,中北部山区基本无人定居。与朝鲜人口密度相关的主要因素有坡度、海拔、道路网密度、居民地指数、居民点密度和耕地指数,其中,海拔和坡度与朝鲜人口密度成负相关。对朝鲜人口密度空间分布有较大决定力的有道路网密度、居民点密度和居民地指数,说明朝鲜人口密度受经济条件的制约比较明显。

4 结束语

选取地形、土地利用和居民点等因素,利用 GIS 和统计分析法对朝鲜人口密度进行空间化模拟,并利用地理探测器方法探测影响朝鲜人口分布的决定力因子。研究表明,朝鲜人口主要集中在西南沿海平原地区,少量分布在东部

模拟人口与实际统计人口相关性达到0.769,模拟结果较为理想。但人口密度的分布是各种自然和社会经济因素共同作用的结果,影响因子众多。受数据获取难度的限制,算法选取的因子不够丰富,使得模拟朝鲜人口密度的精度比较有限,无法全面揭示朝鲜人口密度分布的内在机理,在今后的研究中需要进一步完善。

参考文献:

- [1] 胡焕庸.论中国人口之分布[M].北京:科学出版社,1983:52-92.
HU H Y. On the distribution of Chinese population[M]. Beijing: Science Press, 1983:52-92.
- [2] 廖顺宝,孙九林.基于GIS的青藏高原人口统计数据空间化[J].地理学报,2003,58(1):25-33.
LIAO S B, SUN J L. GIS based spatialization of population census data in Qinghai-Tibet Plateau[J]. Acta Geographica Sinica, 2003,58(1):25-33.
- [3] 曹彦波,李永强,李敏,等.基于多源数据的云南人口空间分布模拟研究[J].中国地震,2014,30(1):118-131.
CAO Y B, LI Y Q, LI M, et al. Simulation of spatial distribution of the population of Yunnan based on the integration of multi-resource data[J]. Earthquake Research in China, 2014, 30(1):118-131.
- [4] 曾展,涂平.基于信息关联的城市人口空间分布建模方法研究[J].福州大学学报:自然科学版,2014,42(6):853-858.
ZENG Z, TU P. The research on methods of modeling spatial distribution of urban population and their application[J]. Journal of Fuzhou University: Natural Science Edition, 2014, 42(6):853-858.
- [5] LU A M, LI C M, LIN Z J, et al. Spatial continuous surface model of population density[J]. Acta Geodaetica et Cartographica Sinica, 2003,32(4):344-348.
- [6] 刘纪远,岳天祥,王英安,等.中国人口密度数字模拟[J].地理学报,2003,58(1):17-24.
LIU J Y, YUE T X, WANG Y A, et al. Digital simulation of population density in China[J]. Acta Geographica Sinica, 2003, 58(1):17-24.
- [7] MARTIN D. An assessment of surface and zonal models of population[J]. International Journal of Geographical Information Systems, 1996,10(8):973-989.
- [8] 王珂靖,蔡红艳,杨小唤,等.基于城镇居民用地再分类的人口数据空间化方法研究——以长江中游4省为例[J].遥感技术与应用,2015,30(5):987-995.
WANG K J, CAI H Y, YANG X H, et al. Spatialization method for census data based on reclassifying residential land use in urban areas—a case study in the middle reaches of the Yangtze River Watershed[J]. Remote Sensing Technology and Application, 2015,30(5):987-995.
- [9] 于欣鑫,马安青,邢容容,等.青岛地区人口密度空间化模拟[J].中国海洋大学学报:自然科学版,2015(8):97-102.
YU X X, MA A Q, XING R R, et al. Simulation of population spatial distribution in Qingdao region[J]. Periodical of Ocean University of China, 2015(8):97-102.
- [10] 王磊,蔡运龙.人口密度的空间降尺度分析与模拟——以贵州猫跳河流域为例[J].地理科学进展,2011,30(5):635-640.
WANG L, CAI Y L. Spatial down-scaling analysis and simulation of population density in Maotiaohe Basin, Guizhou Province[J]. Progress in Geography, 2011,30(5):635-640.
- [11] 叶靖,杨小唤,江东.乡镇级人口统计数据空间化的格网尺度效应分析——以义乌市为例[J].地球信息科学,2010,12(1):40-47.
YE J, YANG X H, JIANG D. The grid scale effect analysis on town leveled population statistical data spatialization[J]. Journal of Geo-Information Science, 2010,12(1):40-47.
- [12] 卓莉,陈晋,史培军,等.基于夜间灯光数据的中国人口密度模拟[J].地理学报,2005,60(2):266-276.
ZHUO L, CHEN J, SHI P J, et al. Modeling population density of China in 1998 based on DMSP/OLS nighttime light image[J]. Acta Geographica Sinica, 2005,60(2):266-276.
- [13] 蓝荣钦,王家耀.智慧城市空间信息基础设施支撑力评价体系研究[J].测绘科学技术学报,2015,32(1):78-81.
LAN R Q, WANG J Y. Research on the support capacity evaluation system of spatial information infrastructure for smart city[J]. Journal of Geomatics Science and Technology, 2015,32(1):78-81.
- [14] Wikipedia. North Korea[EB/OL]. [2016-08-03]. https://en.wikipedia.org/wiki/North_Korea.
- [15] Central Intelligence Agency. The world factbook[EB/OL]. [2016-08-03]. <https://www.cia.gov/library/publications/the-world-factbook/geos/kn.html>.
- [16] DIVA-GIS. DATA[EB/OL]. [2015-04-15]. <http://www.diva-gis.org/gdata>.
- [17] GEOFABRIK. ASIA[EB/OL]. [2015-04-15]. <http://download.geofabrik.de/asia.html>.
- [18] 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[J]. International Journal of Geographical Information Science, 2010, 24(1):107-127.
- [19] 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[J].地球信息科学学报,2016,18(6):767-778.
LIAO Y, WANG X Y, ZHOU J M. Suitability assessment and validation of giant panda habitat based on geographical detector[J]. Journal of Geo-Information Science, 2016,18(6):767-778.
- [20] 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[J].地理科学进展,2015,34(8):966-975.
ZHAN D S, ZHANG W Z, YU J H, et al. Analysis of influencing mechanism of residents' livability satisfaction in Beijing using geographical detector[J]. Progress in Geography, 2015, 34(8):966-975.
- [21] 毕硕本,计哈,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[J].地理科学进展,2015,34(1):118-127.
BI S B, JI H, CHEN C C, et al. Application of geographical detector in human-environment relationship study of prehistoric settlements[J]. Progress in Geography, 2015,34(1):118-127.