

基于多源数据的多元综合法实现泉州市人口网格化

黄耀裔, 陈文成

(泉州师范学院 资源与环境科学学院, 福建 泉州 362000)

摘要:以泉州市(不包括金门县)为例,应用人口关联因素相关分析探索泉州市的植被指数 EVI、夜间灯光(NPP-VIIRS、LJ1-01)、土地利用数据、DEM 数据、交通与水系矢量数据等多源异构数据的关联性,筛选关联因子后再利用地理探测器的“分异及因子探测”法进行指标赋权,实现基于多元综合法的泉州市人口网格化。实验结果显示:利用 LJ1-01 夜间灯光的人口网格化结果优于 NPP-VIIRS,说明多源数据的精度越高,效果越好。人口网格化结果可为其他数据模型的集成研究提供数据来源。

关键词:NPP-VIIRS;LJ1-01;多元综合法;人口网格化;泉州市

中图分类号:P315 **文献标志码:**A **文章编号:**1671-1807(2020)08-0135-05

人口数据是表征人类活动最直接的指标之一,在生态环境保护、灾害风险评估救援、疾病防疫防控、商业决策、区域规划与开发等领域具有广泛的运用。当前按行政单元的人口统计则存在有可变面元、MAUP 分区、形状效应、尺度/面积效应等诸多问题,而人口网格化具有比行政单元下的平均人口密度更能够反映行政区划内部差异,更客观反映人口实际空间状态,目前许多学者如柏中强^[1],王明明^[2],卓莉^[3]等基于 DMSP/OLS 夜间灯光、植被指数、土地利用等多源数据实现人口网格化进行了诸多研究,各国也开发如 UNEP/GRID^[4]、LandScan^[5]、中国公里格网人口分布数据集^[6]等人口网格数据集,然而主要集中在全球、国家或者省域尺度,而对于乡镇(街道)等中小尺度的人口网格化研究则很少。局部区域的人口网格化与大区域尺度同样重要。由于人口的空间分布与夜间灯光、土地利用等多因素存在着密切关联,因此可借助相关分析、回归分析等方法筛选其关联指标后进一步综合。本文以泉州市(不包括金门县)为例,应用人口关联因素相关分析探索泉州市的植被指数 EVI、夜间灯光(NPP-VIIRS、LJ1-01)、土地利用数据、DEM 数据、交通与水系矢量数据等多源异构数据的关联性,利用地理探测器的“分异及因子探测”法进行指标赋权,实现基于多元综合法的 500 m×500 m 尺度大小的泉州市人口网格化。

1 数据来源

1.1 研究区概况

泉州市(117°25′—119°05′E, 24°30′—25°56′N)隶属于福建省的一个地级市,位于福建省东南沿海,北承福州市,南接厦门市,东邻台湾海峡。现管辖鲤城等 4 个县级区、晋江市等 3 个县级市、惠安县等 5 个县及泉州经济技术开发区、泉州台商投资区,下划为 170 个乡镇(街道)。土地面积为 11 015 km²(含金门县),2019 年末常住人口 874 万(不含金门县)。2019 年泉州市实现地区生产总值达 9 946.66 亿元,连续 20 年位居福建省首位。

1.2 多源数据来源与预处理

1.2.1 行政区划及其网格数据

泉州市行政区划数据来源于泉州市栅格地图矢量化,共采集了泉州市市界、县级、乡镇三个级行政单元的面状矢量数据(*.shp 格式,UTM 投影;WGS84 地理坐标系),网格数据(*.shp 格式)利用 ArcGIS 的“网格”工具实现 500 m×500 m 尺度的面状网格,再与泉州市市界行政区划相交得区划内网格作为基准网格,以利于其他多源数据以属性值赋值于网格中,可消除不同遥感数据的空间位置和像元大小不一致等问题,实现多源数据空间位置的同一。

1.2.2 人口普查数据

人口普查数据来源于由村、乡镇(街道)、县级普查后逐级往上汇总单元的 2015 年泉州市 1%常住人

收稿日期:2020-04-21

基金项目:福建省自然科学基金项目(2016J01195);泉州市科技项目(2018Z024;2018Z027)。

作者简介:黄耀裔(1983—),男,福建泉州人,泉州师范学院,高级实验师,计算机技术硕士,研究方向:环境信息系统应用管理与实验教学。

口数抽样调查数据,将其作为属性赋值给前述县级、乡镇行政区划矢量数据当中。

1.2.3 EVI 植被指数数据

相关研究发现植被丰富度与不透水地表呈显著负相关,与人类活动有密切相关,能在一定程度上反映人口分布与人类活动强度,利用植被指数融合夜间灯光缓解过饱和与溢出效应,强化夜间灯光区域的内部强度差异^[7]。本文 EVI 植被数据指数数据来源于中国科学院计算机网络信息中心国际科学数据镜像网(<http://www.gscloud.cn>)。本文获取 2015 年的 MODIS 传感器影像,经公式(1)处理获得 EVI。

$$EVI = 2.5 \frac{NIR - R}{NIR + 6R - 7.5B + 1} \quad (1)$$

式(1)中 NIR、R 同式(1),B 为蓝波段即 MODIS 的第 3 波段。

依次输入数据进行月合成植被指数后,再计算全年 12 个月月合成植被指数的算术平均值即为本文采用的 EVI 年均数据。通过 QGIS 软件的栅格分析工具中的统计分析工具统计网格尺度的 EVI 指数像元平均值作为 EVI 指数网格属性值,作为变量 X_1 。

1.2.4 NPP-VIIRS 夜间灯光数据

NPP-VIIRS 夜间灯光数据来源于 NOAA/NGDC 所属的地球观测小组利用 Suomi 国家极地合作卫星 Suomi-NPP 上搭载的可见红外辐射成像计 VIIRS 的 Day/Night 波段(DNB)获取的夜间灯光数据产品。为了与 2015 年人口普查数据保持一致,使用 2015 年年均夜间灯光光合成数据(SVDNB_npp_20150101-20151231_75N060E_vcm-ntl_v10_avg_rade9.tif),泉州市区域像元值范围为 0~149。按上述预处理方法赋值于网格属性值中,作为变量 X_2 。

1.2.5 珞珈一号(LJ1-01)夜间灯光数据

珞珈一号(LJ1-01)是由我国武汉大学研发的全球首颗专业夜光遥感卫星,空间分辨率为 130 m,泉州市区域像元值范围为 1~850 220。本文利用其年均合成数据产品,按上述预处理方法赋值于网格属性值中,作为变量 X_3 。

1.2.6 土地利用数据

土地利用数据通过 2015 年泉州市范围的 Landsat8 陆地卫星遥感影像数据,空间分辨率为 30 m×30 m,基于 SVM 分类法按建筑用地、植被、裸地、水体土地类型进行监督分类后,分别提取每一种土地类型并统计其在 500 m×500 m 网格中的占比分别作为变量 $X_4 - X_7$ 。

1.2.7 DEM 数据

DEM 数据采用全球数字高程模型数据 ASTER

GDEM,其空间分辨率为 30 m×30 m,按上述预处理方法赋值于网格属性值中,作为变量 X_8 。坡度数据则根据原始 DEM 数据生成,同理按上述预处理方法赋值于网格属性值中,作为变量 X_9 。

1.2.8 交通、水系矢量数据

交通数据来源于 OpenStreetMap 数据,水系数据则来源于福建省 1:50 000 的水系数据。均将其投影为网格一致的投影坐标系,分别按欧氏距离、最小 DEM 成本距离、最小坡度成本距离作为变量 $X_{10} - X_{15}$ 。

2 方法

2.1 人口关联因素探索性相关分析

由于人口分布受诸多因素的影响,包括直接或间接、稳定或变化、静态或动态的作用与反馈的自然、人文社会经济等因素、现象及其所表现出的一些间接特征(如像元的波谱反射率、夜间灯光辐射)等。人口关联因素探索性相关分析则可以揭示要素之间相互关系的密切程度,相互关系密切程度的测定主要通过相关系数的计算与显著性检验来完成,计算模型如下:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

式(2)中 $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$, $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$; y_i, x_i 分别为人口

(因变量)与潜在关联因素(自变量)的样本值即时间序列或截面数据。在给定显著水平 α 若通过检验则为人口分布关联因素,不能通过则不是关联因素, $R > 0$ 为正关联、 $R < 0$ 为负关联,而且当 α 越小关联性越大,并按 0.05、0.01、0.001 划分为显著、高度显著、极端高度显著 3 级。

2.2 人口网格化指标因子体系

据上述数据来源等实际情况选取以下指标因子: EVI 植被指数(X_1), NPP-VIIRS 夜间灯光强度(X_2), LJ1-01 珞珈一号夜间灯光强度(X_3), 土地利用—建设用地面积占比(X_4), 土地利用—植被面积占比(X_5), 土地利用—裸地面积占比(X_6), 土地利用—水体面积占比(X_7), DEM(X_8), 坡度(X_9), 水系欧氏距离(X_{10}), 道路欧氏距离(X_{11}), 水系最小高程成本距离(X_{12}), 水系最小坡度成本距离(X_{13}), 道路最小高程成本距离(X_{14}), 道路最小坡度成本距离(X_{15})。

2.3 人口网格化指标规范化变换

根据人口分布关联因素分析,上述 15 个指标因

子与人口网格化存在正关联或者负关联关系。由于上述指标单位(量纲)不一,指标之间数值大小悬殊,因此首先必须规范化变换,模型如下:

$$\begin{cases} x_{j,k} = \frac{X_{j,k} - \min(X_{j,k})}{\max(X_{j,k}) - \min(X_{j,k})} (\text{正关联}), a \\ x_{j,k} = \frac{\max(X_{j,k}) - X_{j,k}}{\max(X_{j,k}) - \min(X_{j,k})} (\text{负关联}), b \end{cases} (\forall j, k) \quad (3)$$

式(3)中 $x_{j,k}$ 为 j 网格 k 指标规范变换前与变换后的值, $\max(X_{j,k})$ 、 $\min(X_{j,k})$ 分别为泉州市所有网格中 k 指标的最大值与最小值。指标体系中指标的相关系数若大于 0 则应用式(3)-a 变换,相关系数若小于 0 则采用式(3)-b 变换。

2.4 人口网格化指标赋权

行政区划单元类型是多指标综合的结果,与人口空间分布具一定关系。因此基于地理探测器法(Geodetector)^[8]对人口分布区按类划分(区域分类),地理探测器作为一种用来探析空间分异性和验证地理现象与其潜在的驱动因子之间关系的新的一组统计学方法,在地理探测器 q 变量值下利用“分异及因子探测”挖掘人口分布的关联因素及两两间的交互作用。以潜在关联因素为自变量 x 、人口密度为因变量 y ,则地理探测器法计算如下:

$$q = 1 - \frac{\sum_{k=1}^K N_k \sigma_k^2}{N \sigma^2} \quad (4)$$

式(4)中: $k=1, \dots, K$ 为潜在关联因子即自变量 x 的分类/分区数,从而人口密度即因变量 y 也被 x 划分为相同的 k 类(或分区), N 、 N_k 分别为总的(或全区)和 k 类(或分区)的人口密度 y 的样本容量(空间单元数也即乡镇域或网格数), σ^2 、 σ_k^2 分别为 y 的总及 k 类方差:

$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - \bar{y})^2 \quad (5)$$

$$\sigma_k^2 = \frac{1}{N_k} \cdot \sum_{i=1}^{N_k} (y_i - \bar{y}_k)^2 \quad (6)$$

式(5)和式(6)中, y_i 空间 i 的人口密度、 $\bar{y}$ 为全区总的平均人口密度、 $\bar{y}_k$ 为 k 类分区的平均人口密度。则 q 为测度因变量 y 即人口密度空间分异程度的指标,也是测度自变量 x 解释(控制) y 空间分异能力(程度、贡献)的指标, $0 \leq q \leq 1$, q 值越大 y 的空间分异越明显, x 对 y 空间分异解释的贡献 $q\%$ 越大。

由于指标体系中不同指标对人口空间分布的贡献、作用或重要性大小也不同,因此需要对指标进行赋权,本文采用基于“分异及因子探测”的方法进行赋权,如下:

$$r_{h,k} = \frac{q_j}{\sum_{j=1}^n q_j} \quad (h = 1, 2, \dots, i; k = 1, 2, \dots, n) \quad (7)$$

式(7)中 q_j 为分异及因子探测的 j 指标的 q 值, j 表示指标体系中 n 个指标的序号, h 表示 i 组指标体系组合的序号, $r_{h,k}$ 表示第 h 种指标组合中第 k 指标的权重。

2.5 多源加权融合的人口综合建模与网格化

经指标规范变换、赋权后,则多源加权融合的模式如下:

$$\zeta_{i,j,h} = \sum_{k=1}^n (r_{h,k} \cdot x_{i,j,k}) \quad (\forall h, i, j) \quad (8)$$

式(8)中 $x_{i,j,k}$ 表示 i 乡镇域中的 j 网格 k 指标规范变换后的值, $r_{h,k}$ 见式,则 $\zeta_{i,j,h}$ 为 h 组指标组合的 i 乡镇域中的 j 网格的人口分布的综合关联度,则多元综合人口网格化模型为:

$$p_{i,j,h} = \frac{Q_i}{\sum_{j=1}^{M_i} \zeta_{i,j,h}} \cdot \zeta_{i,j,h} \quad (\forall i, j, h) \quad (9)$$

式(9)中 Q_i 为 i 乡镇的 2015 年人口 1% 抽查的乡镇域常住人口, M_i 为 i 乡镇域中的 500 m×500 m 尺度网格的总个数, i 为乡镇 ID 码、 j 为网格 ID 码, $\zeta_{i,j,h}$ 见式, $p_{i,j,h}$ 为 h 组指标组合的 i 乡镇域中 j 网格的网格人口。

2.6 技术路线

以泉州市乡镇行政单元,基于单元内的人口数据与 EVI、NPP-VIIRS、LJ01-1、土地利用等多源数据提取的变量因子进行相关性关联分析,地理探测器的“分异及因子探测”法进行赋权,实现基于多元综合法的 500 m×500 m 尺度大小的泉州市人口网格化,技术路线见图 1。

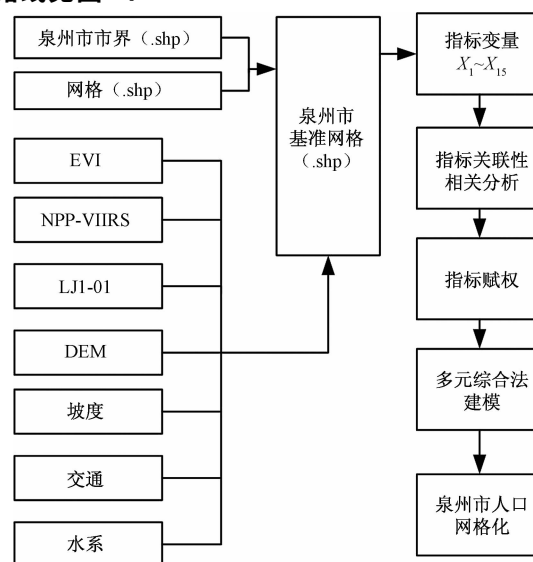


图 1 人口网格化技术路线图

3 泉州市人口网格化

根据人口分布关联因素挖掘结果(见表 1):①据相关分析结果,仅水体面积占比(X_7)未通过显著性检验,因此不是人口分布的关联因素;②根据地理探测器法,交通可达性仅道路欧氏距离(X_{11})、水系可达性仅水系欧氏距离(X_{10})是人口分布关联因素,而水

系最小高程成本距离(X_{12})、水系最小坡度成本距离(X_{13})、道路最小高程成本距离(X_{14})、道路最小坡度成本距离(X_{15})的 P 值未通过检验;③灯光强度 NPP-VIIRS(X_2)、LJ1-01(X_3)分别是人口分布关联因素,分别作为不同变量实现不同多元综合模型进行人口网格化。

表 1 泉州市人口分布因素挖掘及其对人口网格化中的作用、贡献(权重)

因素(指标)		相关分析			地理探测器		基于 q 值的人口网格化权重	
名称	代号	Pearson	. sig	Pearson*	q	p	r_1	r_2
EVI	X_1	0.321	0.000	0.500	0.128 5	0.003	0.081 2	0.080 5
NPP-VIIRS 灯光强度	X_2	0.512	0.000	0.684	0.421 0	0.000	0.266 0	0
LJ1-01 灯光强度	X_3	0.515	0.000	0.667	0.433 8	0.000	0	0.271 9
建设用地面积占比	X_4	0.404	0.000	0.526	0.370 1	0.000	0.233 8	0.232 0
植被面积占比	X_5	-0.350	0.000	-0.498	0.191 6	0.000	0.121 1	0.120 1
裸地面积占比	X_6	-0.213	0.006	-0.026	0.126 8	0.003	0.080 1	0.079 5
水体面积占比	X_7	0.044	0.581	0.186	0.009 1	0.976	0	0
DEM	X_8	-0.250	0.001	-0.404	0.072 3	0.071	0.045 7	0.045 3
坡度	X_9	-0.277	0.000	-0.481	0.121 3	0.004	0.076 6	0.076 0
水系欧氏距离	X_{10}	-0.206	0.008	-0.181	0.078 3	0.052	0.049 5	0.049 1
道路欧氏距离	X_{11}	-0.214	0.006	-0.262	0.072 8	0.070	0.046 0	0.045 6
水系最小高程成本距离	X_{12}	-0.179	0.022	-0.259	0.039 7	0.386 0	0	0
水系最小坡度成本距离	X_{13}	-0.224	0.004	-0.274	0.061 5	0.130 5	0	0
道路最小高程成本距离	X_{14}	-0.171	0.029	-0.274	0.025 9	0.658 6	0	0
道路最小坡度成本距离	X_{15}	-0.205	0.009	-0.295	0.037 3	0.427 7	0	0

注: * 表示加权个案的相关系数,其显著性水平(双侧)均为 0.000。

根据①—③,指标体系为 X_1 、 X_4 、 X_5 、 X_6 、 X_8 、 X_9 、 X_{10} 、 X_{11} 及夜间灯光数据(X_2 或 X_3)择一构建不同灯光的 2 种组合的指标体系,如表 1 所示,即分别由基于 q 值的人口网格化权重 r_1 、 r_2 、每列中不为 0 的指标组

成了略有差异的两组指标体系进而构建人口网格化。

综上述变量与公式,计算得到 2015 年泉州市多源多元加权综合融合的 500 m×500 m 人口网格化的结果如图 2 所示,两组指标组合的结果基本一致,可以互证。

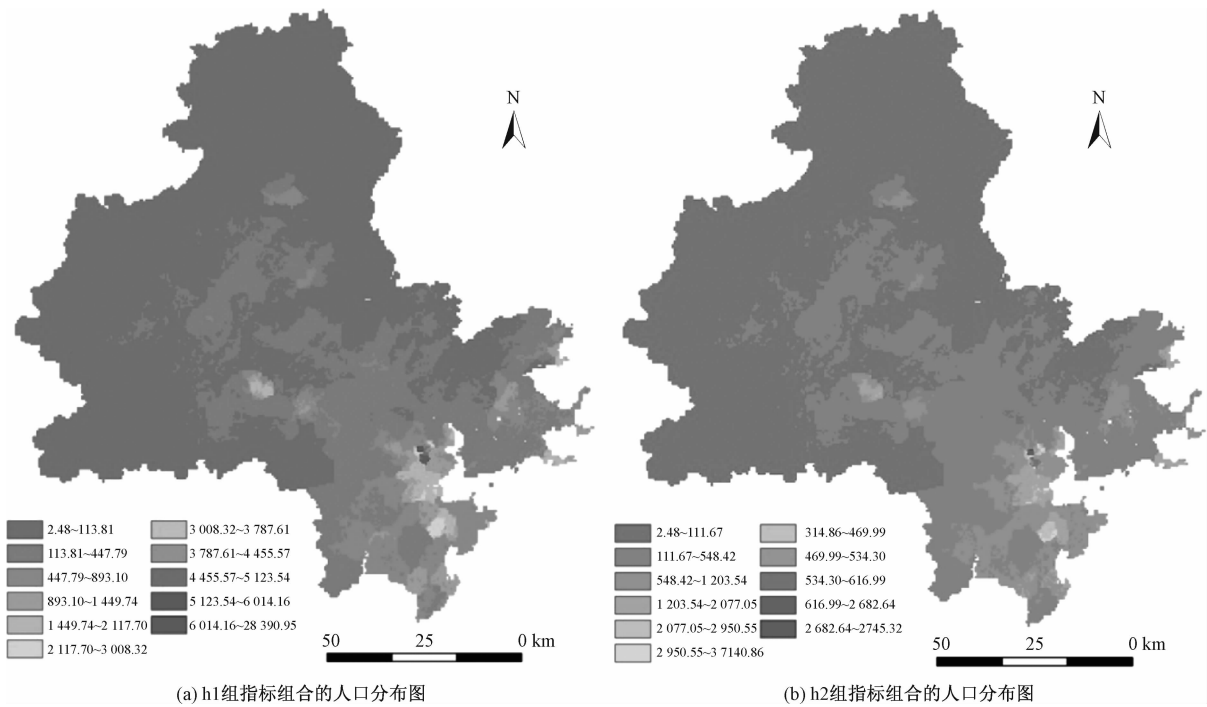


图 2 泉州市(不包括金门县)0.25 km² 尺度的人口网格化结果

进一步为了验证应用精度,通过 h_1 和 h_2 得到的人口网格化结果,根据相对误差公式 $\delta = (X - X')/X'$ 计算其相对误差部分结果表 2,从所有乡镇累积相对误差来看,则基于 LJ1-01 夜间灯光模型比基于 NPP-VIIRS 夜间灯光的相对误差低,累计相对误差也更小,说明 LJ-01 数据能够更加有效地反映人类活动特征,其在人类空间活动领域有更大的应用价值与应用潜力,可为人类持续开展作出更长远的贡献。

表 2 泉州市部分乡镇人口网格化相对误差

泉州市乡镇	δ_{h1}	δ_{h2}
岵山镇	0.03	0.01
湖头镇	0.445	0.205
横口乡	0.538	-0.263
丰泽街道	0.061	-0.027
东湖街道	-0.400	-0.250
前黄镇	-0.280	-0.086
累积相对误差	25.725	8.982

4 结语

本文在 GIS 技术和多源遥感数据的支持下,应用人口关联因素相关分析探索泉州市的植被指数 EVI、夜间灯光(NPP、LJ1-01)、土地利用数据、高程数据及交通与水系矢量数据等多源异构数据的关联性,利用地理探测器的“分异及因子探测”法进行赋权,实现基于多元综合法的泉州市人口网格化。两组指标组合的结果基本一致,可以互证,但从相对误差

比较,则 LJ1-01 数据在人口网格化的精度由于 NPP-VIIRS 数据,表明说明 LJ-01 数据具有更大潜力,人口网格化结果图可为相关部门在灾害风险评估救援、疾病防疫防控、商业决策、区域规划与开发等领域提供辅助决策。

参考文献

- [1] 柏中强,王卷乐,姜浩,等. 基于多源信息的人口分布格网化方法研究[J]. 地球信息科学学报,2015,17(6):653—660.
- [2] 王明明,王卷乐. 基于夜间灯光与土地利用数据的山东省乡镇级人口数据空间化[J]. 地球信息科学学报,2019,21(5):699—709.
- [3] 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.
- [4] LO C P. Modelling the population of China using DMSP operational linescan system nighttime data[J]. Photogrammetric Engineering and Remote Sensing,2001,67(9):1037—1047.
- [5] LANDSCAN. Release of landsacn global[EB/OL]. [2020—02—15]. <https://landscan.ornl.gov/>.
- [6] 江东. 中国公里网格人口数据集[EB/OL]. (2016—11—02). <http://www.geodata.cn/data/datadetails.html?dataguid=161949057751763&docId=19773>.
- [7] CAO X, CHEN J, IMUTA H, et al. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data [J]. Remote Sensing of Environment, 2009, 113: 2205—2209.
- [8] 王劲峰,徐成东. 地理探测器:原理与展望[J]. 地理学报,2017,72(1):116—134.

The Realization of Population Gridding in Quanzhou City Based on multi-source Data

HUANG Yao-yi, CHEN Wen-cheng

(School of Resources and Environmental Science, Quanzhou Normal University, Quanzhou Fujian 362000, China)

Abstract: This paper taking Quanzhou City (excluding Jinmen county) as an example, the correlation analysis of population correlation factors was used to explore the correlation of EVI vegetation index, NPP-VIIRS and LJ1-01 night lighting data, land use data, DEM data, traffic and water vector data and so on other multi-source heterogeneous data. After selecting the correlation factors, the "differentiation and factor detection" method of geographical detector was used to carry out index empowerment to realize the population grid of Quanzhou City based on the multiple comprehensive method. The experimental results show that LJ 1-01 night lighting is better than NPP-VIIRS, which indicates that the higher the accuracy of multi-source data. The results of population gridding can provide data sources for the integration of other data models.

Key words: NPP-VIIRS; LJ1-01; multiple comprehensive method; population gridding; Quanzhou city