

基于 DMSP/OLS 修正的福建省人口数据网格化

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:基于福建省人口普查的县域常住人口、DMSP/OLS、NDVI 等数据,据研究的植被指数能有效降低夜间灯光数据过饱和问题,结合植被指数改进优化 HSI 指数模型,在改进的 HSI 指数下,基于地理探测器分区后分别直接以常住人口的网格个数为权进行加权个案与加权最小二乘法进行人口网格化建模,再进行人口零误差调整得到 1 km×1 km 尺度的福建省人口统计数据网格化.结果表明:结合地理探测器的改进 HSI 指数模型能优化夜间灯光过饱和,有效回避可变面元效应等问题,提高模型拟合精度,对人口网格化效果更好,更符合实际的人口空间分布.

关键词:DMSP/OLS;NDVI;改进 HSI;人口网格化;福建省

中图分类号:K910

文献标识码:A

文章编号:1009-8224(2020)06-0086-08

人口分布具有空间信息和可移动性的特点,是研究区域经济发展、抗震救灾、自然资源规划等的重要依据^[1].现行统计制度下的人口统计主要依行政区划单元,而行政单元存在形状不规则、边界不稳定、同级规模悬殊、内部差异、“可变面元”、与研究单元不一致等问题^[2].人口网格化具有比行政单元的平均人口密度更能够反映行政区划内部人口差异,更接近人口实际分布,有利于分析不同尺度行政区域中人口内部的空间信息,在实现人口与社会经济、自然资源、环境等的有效融合具有重要意义.

夜间灯光数据作为能反映人类经济活动的数据集,备受关注^[3].美国普查局自 1973 年开始对 DMSP/OLS(defense meteorological satellite program/operational linescan system)与人口分布研究显示两者呈显著相关性^[4].然而 DMSP 数据受限于传感器的缺陷,其 DN 值域区间只介于 0~63,导致在灯光强度较高的城市中心会出现灯光过饱和现象,掩盖了饱和区内部实际存在的夜间灯光强度差异,进而影响人口、电力、GDP 等的估算精度和差异程度^[5].如何有效降低过饱和问题成为近年来夜间灯光数据研究的热点,许多学者致力于这方面研究^[6-8].如 Letu^[9]应用三次回归模型校正行政单位,但此法难以应用于像元尺度.Ziskin 等^[10]借助遥感增益校正思想,在非饱和区利用“低增益”校正,在城市饱和密集区利用“高增益”校正,但此法当前应用于 DMSP/OLS 会非常耗费人力和成本.

植被作为地球表面的重要地物,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^[11],因此可考虑利用各种植被指数(如 NDVI、EVI 等)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^[12-16],衍生出 VANUI^[17]、EANTLI^[18]等相关校正指数缓解夜间灯光过饱和与溢出效应.

本文通过研究了前人的各种降低过饱和的方法,发现还有待进一步提高的空间.由于 HSI 指数背后的假设是基于 NTL 和 NDVI 遵循幂规律.当 NTL 具有最大值时,随着 NDVI 接近于零,将呈指数增长,存在着过度修正城市周边地区的饱和和核心等不足加以修正,试图得到更优的人口网格化模型.因此根据人口普查数据(福建省县域常住人口^[19]、DMSP/OLS、NDVI),改进 HSI 指数模型后基于 GIS 空间分析,借助地理探测器分区优化调节,直接面向网格加权个案及加权最小二乘法建模后再进行零误差调整网格化,最后进行 1 km×1 km 尺度的福建省人口数据网格化.该法克服多种数据来源综合的复杂性,降

收稿日期:2020-06-03

作者简介:黄耀裔(1983—),男,福建泉州人,高级实验师,硕士,主要从事环境信息系统应用、开发评价研究.

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

低数据计算的复杂度,能够应用于不同地区的人口空间网格化.

1 研究区及数据来源

1.1 研究区概况

福建省位于中国东部沿海 $23^{\circ}33'N\sim 28^{\circ}20'N$ 、 $115^{\circ}50'E\sim 120^{\circ}40'E$,陆地面积 $12.4\times 10^4\text{ km}^2$,下辖福州市、厦门市、泉州市等 9 个地级市,共 84 个县级行政区划单元(其中区 28 个、县 43 个、县级市 13 个,不包含金门县和马祖县),下辖乡、镇、街道办共 1 184 个.由于境内地形地貌复杂多变,丘陵连绵,植被覆盖程度高,社会经济发展差异显著造成境内人口分布极度不均匀,区域间人口密度差异大.

1.2 数据来源

1.2.1 行政区划及其网格数据、人口数据 (1)行政区划数据.福建省行政区划数据来源于国家地理信息公共服务平台(<http://www.tianditu.gov.cn>)的福建省底图扫描矢量化,采集设区市、县、乡镇级行政单元等不同尺度的面状矢量数据(*.shp 格式,比例尺 1:50 000;投影 UTM;地理坐标系 WGS84).约定县级行政单元统称为县域,记 ID 码为 j .

(2)网格数据.由于福建省为(省域)中观尺度^[21],而辅助数据 DMSP/OLS 数据的像元尺度约 1 km^2 ,因此确定人口网格化的尺度为 1 km^2 .利用 ArcGIS 软件的 ArcToolbox 的“网格”工具创建生成面状网格,经由面转点为点状网格后与福建省行政区划相交得基本网格(点状),基本网格点与网格相交得基本网格(面状),对福建省省域进行规则划分,并继承了县域的名称、ID 码等相关属性.其中,点状网格与面状网格为“等价”关系,若无特别要求则不予区分,并设 ID 码统一记为 i .

(3)人口数据.人口数据来源第六次人口普查的乡镇级、县级统计逐级往上汇总单元的福建省常住人口数,作为属性赋值给前述行政区划矢量数据.其中,设 y 为县域总常住人口, j 县域总常住人口为 y_j .

1.2.2 DMSP/OLS 遥感夜间灯光数据 数据为来源于美国海洋暨大气总署 NOAA(网站 <http://ngdc.noaa.gov/eog/download.html>)发射的国防气象卫星计划的 DMSP Block5D21SatelliteF21 卫星搭载的 OLS 传感器获取稳定夜间灯光数据 F182010.v4b_web.stable_lights.avg_vis.tif^[22],能够探测到地球表面的城市、小型居民地等夜间灯光及车流等可见光-近红外波段的发射辐射.

1.2.3 NDVI 植被指数 NDVI 植被指数是根据 NASA 发射的 EOS 卫星的所携带的中分辨率成像光谱仪 MODIS 所拍摄的卫星数据反演,MODIS 数据作为中、低分辨率遥感影像的典型代表,契合了上述的中观尺度网格.其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>),影像数据时间尺度与人口普查时间保持一致.

$$NDVI=NIR/R. \tag{1}$$

其中:NIR、R 分别为近红外波段、红波段,即 MODIS 的第 2、第 1 波段.根据式(1)计算得 NDVI 后,取月内每天最大值的平均值为月合成数据,年平均 NDVI 为 12 个月的月 NDVI 合成数据,仍表示为 NDVI,作为 HSI 指数灯光修正的参变量.

1.3 网格属性

首先分别将 DMSP/OLS、NDVI 数据(LAT/LON 坐标系)统一投影为 UTM,使之与行政区划数据、基本网格数据具相同的空间参照系统,再利用福建省行政区进行裁剪,并统一采用与人口网格化尺度一致即 $1\text{ km}\times 1\text{ km}$ 网格利用最邻近法进行重采样,得 NTL、NDVI,后经栅格转为点后,再通过 GIS 叠置分析给基本网格赋属性,因此后述灯光修正、网格人口估计/预测等都是通过网格计算.

2 研究方法

2.1 HSI 指数改进

2.1.1 夜间灯光数据的归一化变化 DMSP/OLS 传感器的辐射分辨率较低,灰度值/像元 DN 值介于 0~63,容易出现过饱和现象,及存在溢出效应,应用之前须加以修正,先作规范化变换:

$$NTL'=NTL/NTL_{\max}. \tag{2}$$

其中:NTL 为夜间灯光强度,为 NTL 的最大值归一化值, $0 \leq \text{NTL}' \leq 1$, $\text{NTL}'=0$ 表示无灯光, $\text{NTL}'=1$ 表示灯光饱和.采用最大值归一化是为了确保 $\text{NTL}' \neq 0$ 时变换值也不为 0(即 $\neq 0$),为简便下述 NTL' 仍记为 NTL.

2.1.2 HSI 指数改进 NDVI 植被指数不仅分辨率及精度比 DMSP/OLS 高,也比 DMSP/OLS 具有更宽广的光谱范围,在反映人口分布与人类活动强度等方面具互补性,二者数据融合可以有效缓解夜间灯光过饱和与溢出效应,强化夜间灯光区域的内部强度差异.Liu 等^[22]、杨等^[14]及 Lu 等^[16]基于 NDVI 提出灯光降饱和指数,但是应用于人口数据等网格化/空间化则存在前述的欠妥当之处:当 $\text{NDVI}=1$ 时,仅将 NTL 减少一半.这表明城市活动同时发生在植被茂密的地区.然而,在世界上许多地区,城市扩张主要在在植被减少或低的地区展开.本研究试图利用 NTL 与 NDVI 之间的反比关系,建立一个更稳健的城市化地区光谱指数.本文进一步提出对算法模型的改正,即在 $0.1 \leq \text{NDVI} \leq 0.9$ 的适应居住区中,若 DMSP/OLS 的 DN 值即 $\text{NTL}=0$,则说明缺少夜间人类活动,为适应居住区中的暂无人居住区,因此令 $\text{HSI}_{i,j,k}=0$,这对于在人口网格化/空间化、电力消耗量估算与空间化、城市扩张等方面的应用非常重要.改正后的指数如下:

$$\text{HSI}_{i,j,k} = \begin{cases} 0 & \text{NDVI} < 0.1 \text{ or } \text{NDVI} > 0.9, \\ \frac{1 - \text{NDVI}_{i,j,k} + \text{NTL}_{i,j,k}}{(1 - \text{NTL}_{i,j,k}) + (1 + \text{NTL}_{i,j,k}) \times \text{NDVI}_{i,j,k}} & 0.1 \leq \text{NDVI} \leq 0.9, \\ 0 & \text{NTL} = 0. \end{cases} \quad (3)$$

其中: i,j 分别为网格、县域的 ID 码, k 表示人口分区的类型, $\text{NTL}_{i,j,k}$ 为规范化后的 i 网格的灯光强度, $\text{NDVI}_{i,j,k}$ 为 i 网格的年均 NDVI(见 1.2.3),则 $x_{i,j,k}$ 中的 i,j,k 为经 NDVI 植被指数修正后的 k 类人口分区的 j 县域中 i 网格灯光指数(HSI 指数),其中 $\text{NDVI} < 0.1$ 及 $\text{NDVI} > 0.9$ 的区域为不适应居住区、无常住人口区,DN 值=0 或因灯光溢出而 $\text{ND} \neq 0$ 都定义 $\text{HSI}=0$.

2.2 地理探测器法分类

我国行政区划是综合考虑历史沿革历史传统因素、政治与社会经济发展、自然地理条件等的结果,行政区划单元(地方政府)的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^[24]对人口分布区按类划分(区域分类),以县域类型即县级地方政府类型为自变量(变量值 $k=1,2,3$ 分别表示区、市、县),乡镇级行政单元的人口密度为因变量,利用地理探测器的“分异及因子探测”探测人口的空间分异性及其县域类型 k 多大程度上来解释人口的空间分异,通过“风险区探测”探测市、区、县的人口密度是否存在显著差异(即是否存在空间异质性)对县域分类.通过相关分析方法等筛选与人口空间分布相关的社会经济发展指标与自然地理环境因素,再采用“二次分区法”等进行综合的人口分布区划分,按不同区域构建人口网格化模型来提高模型精度.

2.3 县域人口网格化

2.3.1 直接面向网格建模的数据准备 HSI 改进后数值存为基本网格的新属性 $x_{i,j,k}$,通过 GIS 空间分析后分别按县域统计:(1)统计县域中非零(即 $x_{i,j,k} \neq 0$ 、有常住人口)网格数 $n_{j,k}$;(2)计算非零网格的灯光指数均值 $x_{j,k} = \sum x_{i,j,k} / n_{j,k}$ 并保存为县域属性,其中 i 为网格的 ID 码, j 为县域的 ID 码, k 表示县域的人口分区类型($k=1,2,3$ 依次表示区、市、县),则 $x_{j,k}$ 是 j 的网格的代表值 j 的以网格为单位的平均密度;(3)计算非零网格即有常住人口的平均人口 $y_{j,k} = \sum y_{i,j,k} / n_{j,k} = y_i / n_{j,k}$,其中 $y_{i,j,k}$ 为 k 类区 j 县域 i 网格的人口数/以网格为单位的人口密度(未知的人口网格化就是对 $y_{i,j,k}$ 的估计), $n_{j,k}$ 为 $x_{i,j,k} \neq 0$ 即 $y_{i,j,k} \neq 0$ 的网格数, y_j 为 j 县域的常住人口,则 $y_{j,k}$ 为与改进后 HSI 指数对应的 j 的网格的人口代表值、也即以网格为单位 j 的平均人口密度.然后,导出属性 $n_{j,k}$ 、 $y_{j,k}$ 、 $x_{j,k}$ 等,利用 SPSS 的相关分析、回归分析等分别按县域进行建模.

2.3.2 人口加权个案及加权最小二乘估计建模 因为灯光数据本身已包含城市、居民地、道路等信息,进行了比与土地利用融合更详细的归一化植被指数 NDVI 指数数据融合,因此仅选择改正后的 HSI 指数(即 $x_{j,k}$)为自变量.先通过 SPSS 软件分别进行无加权个案与以 $n_{j,k}$ 为权加权个案的相关分析,探索

$y_{j,k}$ 与 $x_{j,k}(k)$ 相关性作为选择人口网格化模型函数形式的依据,再利用 SPSS 的回归分析进行建模:

$$\hat{y}_{j,k} = f_k(x_{j,k}). \tag{4}$$

其中: k 表示人口分区的类型; j 表示归属 k 类的县域,将根据地理探测器探索; $f_k()$ 为 k 类分区人口网格化模型的函数, $x_{j,k}$ 为 $f_k(x_{j,k})$ 的自变量,为 $f_k(x_{j,k})$ 的因变量的回归值、估计值、预测值.因为上述的相关分析、回归分析分别都是根据以县域网格代表值 $x_{j,k}$ 、 $y_{j,k}(k)$ 为自变量、因变量的样本,是直接面向网格的,回避了从不规则县域到规则网格的尺度效应,但是由于“网格代表 $x_{j,k}$ 、 $y_{j,k}$ ”所代表的网格数量 $n_{j,k}$ 即县域规模相差悬殊,因此本文提出以 $n_{j,k}$ 为权重,即式(4)模型建模的.

2.3.3 县域总人口零误差调整 因为式(4)的人口网格化模型 $f_k(x_{j,k})$ 的回归系数,是 k 类县域的公共的平均回归系数,未顾及类内县域之间必然存在着人口分布的“次级”差异性.同时也是为了最充分利用已知的人口数据以及作为削弱可变面元效应影响的一种手段/方法,因此根据“维持县域总人口不变”的县域控制原则、方法,对式(4)初估的网格人口进行优化调整:

$$\xi_{j,k} = \sum_i y_{i,j,k} / \sum_i \hat{y}_{i,j,k} = y_j / \sum_i \hat{y}_{i,j,k} \quad (\forall k). \tag{5}$$

$$P_{i,j,k} = \hat{y}_{i,j,k} \xi_{j,k} = \hat{y}_{i,j,k} (y_j / \sum_i y'_{i,j,k}) \quad (\forall k). \tag{6}$$

其中: y_j 为 k 类区 j 县域的总人口; $y_{i,j,k}$ 为其 i 网格的人口数,为根据 $x_{i,j,k}$ 、式(4)初估 i 网格的人口数,则为 k 类区 j 县域的网格人口初估数的优化调节系数,作为 j 的属性,通过叠置传递给相关网格.所以,式(6)中的 $p_{i,j,k}$ 为最终预测即人口网格化的 k 类区 j 县域 i 网格的人口数(人口密度网格化).

3 结果与分析

3.1 地理探测器分区

由于人口分布区域差异,因此借助地理探测器分区建模来提高网格化精度.从地理探测器分区分析结果(表 1):在显著水平 $P=0.000$ 下经“分异及因子探测”, q 通过了显著性检验,由此可得县域类型 $k(k=1、2、3$ 即 $k=$ 区、市、县)是人口密度空间分异的驱动力、控制因素.经“风险区探测”,区市县两两之间的人口密度差异有统计学意义.根据 q 显著说明 k 分类内人口密度离差平方和的和小于总离差平方和是显著的,因此依地理探测器分区结果将福建省人口密度按区市县划分为 3 类.

表 1 地理探测器分区

Tab.1 Results of geodetector

分异及因子探测		风险区探测*			
q 值	P 值	k	市	县	均值
0.160 9	0.000	区	Y	Y	6 398.4
		市		Y	1 364.1
		县			516.0

注:显著水平 $\alpha=0.05$,Y 表示差异有统计学意义.

3.2 地理探测器分区下的加权个案及加权最小二乘法模型分析

分别检索区(28 个区)、市(13 个市)、县(43 个县)的 $n_{j,k}$ 、 $y_{j,k}$ 、 $x_{j,k}$ 等数据,利用 SPSS 进行无个案加权及以 $n_{j,k}$ 为权进行个案加权的皮尔逊相关分析(表 2).从表 2 可知:(1) 当无个案加权时,区类的网格代表值 $x_{j,1}$ 与 $y_{j,1}$ 呈高度显著正线性相关(显著水平 $P=0.01$),而市类、县类的 $x_{j,2}$ 与 $y_{j,2}$ 、 $x_{j,3}$ 与 $y_{j,3}$ 均呈极端高度显著正线性相关(显著水平 $P=0.001$). (2) 加权个案后,区、市、县的网格代表值 $x_{j,k}$ 与 $y_{j,k}$ 均呈极端高度显著正线性相关(显著水平 $P=0.001$). (3) 区、市、县的 HSI(改进前的 HSI、EANTLI、NTL 分别与相应的非零网格人口代表值 $y_{j,k}$ (因变量)之间均呈线性关系($\alpha=0.01$ 高度显著或 $\alpha=0.001$ 极端高度显著).

因 $x_{j,k}$ 、 $y_{j,k}(k)$ 呈线性关系,以及无人/无灯光即 $x_{i,j,k}=0$ 则无人,因此式(4)为无常数项的一元线性回归模型,如表 3 所示:(1) 对于自变量(改进后的 HSI),其模型的 F 检验与系数 t 检验的 Sig. 均为 0.000,通过显著性检验, R^2 值介于 0.750 至 0.977,对人口变化的解释达 75% 及以上,模型拟合精度高. (2) 其他自变量为 HSI 改进前、EANTLI、NTL 的模型、系数的 Sig. 也全为 0.000,都分别通过 F 检验、 t 检验的显著性检验. $R^2 \geq 0.663$,值也较大,模型的拟合精度也都较高. (3) 相应县域类型的模型,除区的 EANTLI 的 $R^2=0.759$ 略大于改进后的 HSI 的 $R^2=0.750$ 外,其他都是改进后 HSI 模型的 R^2 较大,尤

其明显比 HSI 模型大,因此改正后 HSI 指数对人口变化的解释性更好,模型的拟合精度更高.因此,表 3 所示的模型都可用于人口网格化,即将网格的自变量值如 $x_{i,j,k}$ 代入模型中则得网格的人口.通过式(5)和式(6)实现人口统计数据的 $1\text{ km}\times 1\text{ km}$ 尺度的网格化.

表 2 与人口密度的相关分析结果

Tab.2 Correlation analysis results of population density

k	自变量	系数	p 值	加权个案		
				系数	P 值	$\sum n_{j,k}$
1 区	HSI 改进后	0.548	0.003	0.566	0.000	10 931
2 市		0.951	0.000	0.952	0.000	11 222
3 县		0.788	0.000	0.728	0.000	25 586
1 区	HSI 改进前	0.545	0.003	0.604	0.000	18 834
2 市		0.960	0.000	0.973	0.000	24 424
3 县		0.951	0.000	0.908	0.000	75 432
1 区	EANTLI	0.551	0.002	0.610	0.000	10 972
2 市		0.936	0.000	0.924	0.000	11 265
3 县		0.398	0.008	0.372	0.000	25 757
1 区	NTL	0.489	0.008	0.551	0.000	11 215
2 市		0.928	0.000	0.921	0.000	11 631
3 县		0.597	0.000	0.586	0.000	26 310

表 3 SPSS 的加权个案及加权最小二乘法回归分析的主要成果

Tab.3 Main results of weighted case and weighted least square regression analysis of SPSS

k	自变量	式(4): $y'_{j,k}=f_k(x_{j,k})=a_k\cdot x_{j,k}$	系数检验		模型拟合与检验				
			t 值	Sig.	R	R^2	调整 R^2	F 值	Sig.
1 区	HSI	$y_{j,1}=2\,396.267\cdot\text{HSI 改进后},j,1$	181.000	0.000	0.866	0.750	0.750	32\,745	0.000
2 市	改进	$y_{j,2}=2\,147.511\cdot\text{HSI 改进后},j,2$	689.100	0.000	0.988	0.977	0.977	474\,849	0.000
3 县	后	$y_{j,3}=2\,169.138\cdot\text{HSI 改进后},j,3$	568.800	0.000	0.963	0.927	0.927	323\,542	0.000
1 区	HSI	$f(\text{HSI}_{j,1})=1\,871.724\cdot\text{HSI 改进前},j,1$	192.400	0.000	0.814	0.663	0.663	37\,031	0.000
2 市	改进	$f(\text{HSI}_{j,2})=1\,764.754\cdot\text{HSI 改进前},j,2$	430.600	0.000	0.940	0.884	0.884	185\,405	0.000
3 县	前	$f(\text{HSI}_{j,3})=1\,257.651\cdot\text{HSI 改进前},j,3$	569.300	0.000	0.901	0.811	0.811	324\,106	0.000
1 区	EANTLI	$f(\text{EANTLI}_{j,1})=3\,227.767\cdot\text{EANTLI},j,1$	186.100	0.000	0.871	0.759	0.759	34\,617	0.000
2 市		$f(\text{EANTLI}_{j,2})=3\,286.031\cdot\text{EANTLI},j,2$	482.100	0.000	0.977	0.954	0.954	232\,390	0.000
3 县		$f(\text{EANTLI}_{j,3})=3\,592.265\cdot\text{EANTLI},j,3$	345.700	0.000	0.907	0.823	0.823	119\,538	0.000
1 区	NTL	$f(\text{NTL}_{j,1})=3\,236.964\cdot\text{NTL},j,1$	170.500	0.000	0.850	0.722	0.722	29\,078	0.000
2 市		$f(\text{NTL}_{j,2})=2\,838.422\cdot\text{NTL},j,2$	506.800	0.000	0.978	0.957	0.957	256\,836	0.000
3 县		$f(\text{NTL}_{j,3})=2\,952.622\cdot\text{NTL},j,3$	496.000	0.000	0.950	0.903	0.903	246\,030	0.000

3.3 HSI 改进的人口网格化结果分析

由于县级统计单元的人口数据为已知数,因此统计单元人口与其估计值计算的误差如均方根误差、相对误差等值大小与网格人口估计精度不存在必然的关系.此外,利用 SPSS 回归分析建模已经对模型进行了各种检验,县域控制已使县域内的统计人口与其估计值的相关系数为 1、各种误差值也均为 0.因此,为了说明对 HSI 改进的效果,采用相同方法分别利用 HSI 改进前、HSI 改进后、EANTLI、NTL 进行人口网格化,彼此间的网格人口的二元相关系数矩阵可见表 4,人口网格化结果见图 1(a~c).

据表 4,分别以 HSI 改进前后有常住人口即 $p_{i,j,k}\geqslant 1$ 网格不同辅助数据的网格人口为样本,两两之间的相关系数分别为 $R^2\geqslant 0.833$ 和 $R^2\geqslant 0.821$,接近于 1,呈非常极端高度显著的正相关关系,且 HSI 改进后大于 HSI 改正前,说明 HSI 改进后效果更好.根据图 1(a~c)可视化结果,在实现 $1\text{ km}\times 1\text{ km}$ 尺度的福建省人口数据网格化专题图和福州市、三明市局局部放大专题图(图 1(d~k))显示,原本灯光溢出效应区域也存在明显内部差异,改善了行政区划单元内部人口的差异度,人口网格化更符合实际状况.

EANTLI 与 HSI 改进后(即图 1(a)和 2(b))的一致性比 EANTLI 与 HSI 改进前(即图 1(b)和 1(c))的一致性大,尤其西部、西北部更明显.从局域放大图看,在地理探测器分区建模下,对比灯光过饱和典型区(福州市局部)与灯光不足典型区(三明市局部),同样的人口分级人口分布,HSI 改进后对人口高密度区(福州图 1(d~f))的刻画比 HSI 改进前更为细腻,HSI 改进前对低值区(三明市图 1(h),图 1(i),图 1(k))的刻画虽更细致但与 EANTLI 却相差很大,而 HSI 改进后与 EANTLI 的非常接近(图 1(i)、图 1(j)).

表 4 网格人口的二元相关系数矩阵

Tab.4 Matrix of binary correlation coefficient of grid population

方法	HSI(HSI 改进前)>0(N=118 690)			HSI(HSI 改进后)>0(N=47 739)		
	HSI 改进前	HSI 改进后	EANTLI	HSI 改进前	HSI 改进后	EANTLI
HSI 改进后	0.985	—	—	0.993	—	—
EANTLI	0.833	0.852	—	0.821	0.832	—
NTL	0.938	0.957	0.924	0.940	0.950	0.914

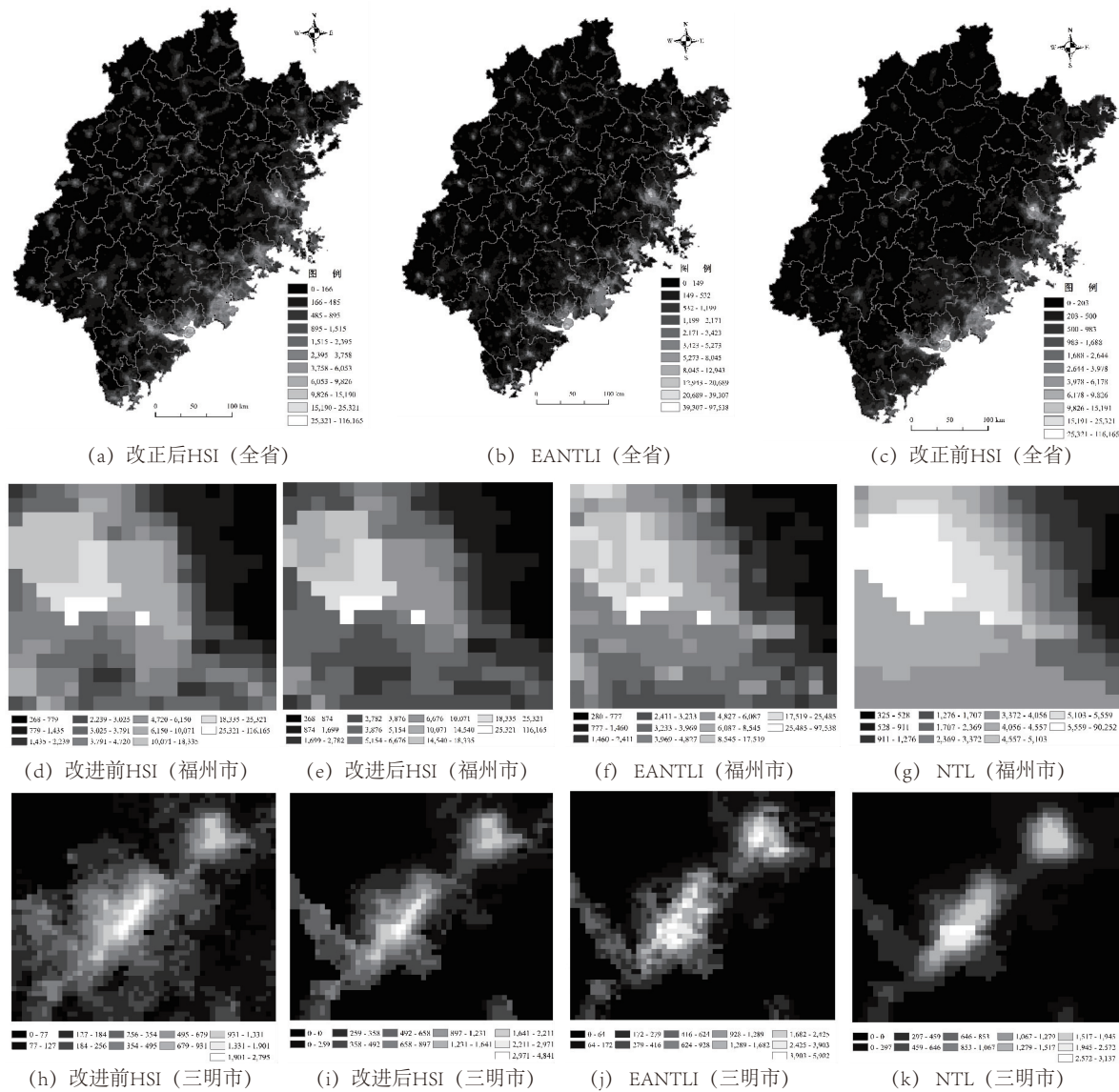


图 1 加权个案及加权最小二乘法模型人口网格化结果

Fig.1 Population gridding results of weighted case and weighted least square model

为了进一步分析网格化后县域人口分布是否更接近符合实际分布,汇总县域网格人口数见表 5.根据表中结果,HSI 改进前的常住人口空间分布范围即 $\sum n_{j,k} = 118\ 690$ 太大,偏离实际情况,而 HSI 改

进后的 $\sum n_{j,k}=47\ 739$ 与 EANTLI 的 $\sum n_{j,k}=47\ 994$ 基本一致.按县域统计 HSI 改进前的均值及最大值的最小值很小,与 HSI 改进后及 EANTLI 的相差很大.虽然按县域统计的 HSI 改进前的最小值与 EANTLI 相近并介于 EANTLI 与 HSI 改进后的之间,但最小值的平均值却是三者中最小的.综上所述,HSI 改正后在地理探测器分区下的辅助人口网格化的效果更好,更符合实际.

表 5 按县域的网格人口统计的汇总

Tab.5 Summary of grid demographics by county										人/(1 km×1 km)		
参数	HSI 改进后			HSI 改进前			EANTLI			NTL		
	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均
均值	89 378	309	2 375	89 378	58	2 277	89 378	308	2 507	89 378	285	2 475
最大值	116 165	1 092	5 494	116 165	290	4 815	97 538	2 635	9 910	90 252	1 118	4 150
最小值	89 378	125	2 507	74 502	26	1 350	81 329	16	1 422	87 340	104	1 870
极差	74 502	125	1 495	41 663	263	3 465	35 166	2 591	8 488	6 072	1 009	2 281
标准差	41 663	868	3 999	17 121	25	749	6 428	264	1 167	1 903	171	515
变异系数	1.173	0.192	0.564	1.260	0.192	0.620	2.215	0.067	1.214	1.058	0.013	0.670
总网格 $\sum n_{j,k}$	47 739			118 690			47 994			49 156		

HSI 改进后的人口网格化的成果分析福建省的人口分布:依图 1(a)呈现,1 km×1 km 网格的福建省人口密度空间分异显著,总体上呈 NE 向的带状、串珠状分布,其次为呈 NW 向的带状、串珠状分布,二者组合则呈现为不甚规则的网状,交叉处为局域人口分布中心,在灯光过饱和与典型区与灯光不足典型区均更符合实际的人口分布,整体分布于闽东南沿海的福州市区、厦门市、泉州市区等地区.

4 结束语

人口空间分布会直接影响其经济、社会、环境、资源等,本研究正视了夜间灯光像元饱和与溢出效应对网格化的不利影响借助植被指数与人口空间分布的相关性,而植被指数较夜间灯光具有更宽广的光谱范围,将其融合在 DMSP/OLS 夜间灯光中,拓展其光谱(像元)范围,具有能力降低夜间灯光数据过饱和区过饱和问题.基于此思想侧重方法的改进,对 HSI 指数改进,将适合人类居住类网格($0.1\leq NDVI\leq 0.9$)进一步划分为有人居住和尚未人口居住,引入地理探测器法分区下直接面向网格的加权个案与加权最小二乘法建模,克服可变面元问题的影响,考虑了县域规模的差异,更能反映城市间夜间光度变化特征.此方法简单、直观且具有一定的普适性.但由于人口变化是一个长期、连续的动态变化,而我国进行人口普查的周期也较长,时效性存在不足,同时多源遥感数据受卫星运行周期影响而存在一定的周期性,难于长期连续性的获取.DMSP/OLS 夜间灯光数据于 2013 年被具有更高分辨率的 NPP-VIIRS 数据所替代,因此后续年份的人口网格化则需从新构建植被指数来校正新的夜间灯光数据 NPP-VIIRS 的模型,从而达到连续性.本研究成果可以为其他数据模型的集成研究提供数据来源,进一步从自然、人文、经济等方面考虑其驱动力,这也是本研究后续开展的工作方向.

参考文献:

[1] 张国俊,黄婉玲,周春山,等.城市群视角下中国人口分布演变特征[J].地理学报,2018,73(8):1513-1525.

[2] 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.

[3] 李小敏,郑新奇,袁涛.DMSP/OLS 夜间灯光数据研究成果知识图谱分析[J].地球信息科学学报,2018,20(3):351-359.

[4] HE C Y,MA Q,LI T,Spatiotemporal dynamics of electric power consumption in Chinese Mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data[J].Journal of Geographical,2012,22(1):125-136.

[5] SHI K,YU B L HUANG Y X,et al.Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: a comparison with DMSP-OLS data [J].Remote Sensing,2014(6):1705-1724.

[6] 郑子豪,陈颖彪,吴志峰,等.单元路网长度的 DMSP/OLS 夜间灯光数据去饱和方法[J].遥感学报,2018,22(1):161-173.

- [7] 张梦琪,何宗宜,樊勇.DMSP/OLS 稳定夜间灯光影像的校正方法[J].测绘通报,2017(12):58-62.
- [8] 王鹤饶,郑新奇,袁涛.DMSP/OLS 数据应用研究综述[J].地理科学进展,2012,31(1):11-19.
- [9] LETU H, HARA M, YAGI H, et al. Estimating energy consumption from night-time DMPS/OLS imagery after correcting for saturation effects[J]. International Journal of Remote Sensing, 2010(31):4443-4458.
- [10] ZISJIN D, BAUGH K, HSU F C, et al.Methods used for the 2006 radiance lights[J].Proceedings of the Asia Pacific Advanced Network,2010:131-142.
- [11] POZZI F,SMALL C. Analysis of urban land cover and population density in the United States[J]. Photogrammetric Engineering and Remote Sensing ,2005(71):719-726.
- [12] 毛卫华,胡德勇,曹冉,等.利用 MODIS 产品和 DMSP/OLS 夜间灯光数据监测城市扩张[J].地理研究,2013,32(7):1325-1335.
- [13] ROUSE J,HAAS R,SCHELL J,et al. Monitoring vegetation systems in the Great Plains with ERTS[J].Third ERTS Symposium,1973:309-317.
- [14] 杨续超,高大伟,丁明军,等.基于多源遥感数据及 DEM 的人口统计数据空间化:以浙江省为例[J].长江流域资源与环境,2013,22(6):729-734.
- [15] 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.
- [16] LU D,WENG Q,LI G Residential population estimation using a remote sensing derived impervious surface approach[J].International Journal of Remote Sensing,2006,27(16):3553-3570 .
- [17] ZHANG Q L, SCHAAF C, SETO K C. The vegetation adjusted NTL urban index;a new approach to reducesaturation and increase variation in nighttime luminosity[J]. Remote Sensing of Environment,2013,129(2):32-41.
- [18] 卓莉,张晓帆,邓璟,等.基于 EVI 指数的 DMSP/OLS 夜间灯光数据去饱和方法[J].地理学报,2015,70(8):1339-1350.
- [19] 福建省统计局.福建省统计年鉴[EB/OL].[2020-06-07].www.stats-fj.gov.cn.
- [20] 胡冲冲,毕如田.省域尺度下土壤养分的最佳采样尺度确定[J].山西农业大学学报(自然科学版),2016,36(3):209-218.
- [21] Nation Geophysical Data Center. Operational linescan system-data description[EB/OL].[2020-06-08].http ://www.ngdc.noaa.
- [22] LIU H Q,HUETE A.A feedback based modification of the NDVI to minimize canopy background and atmospheric noise[J].Transactions on Geoscience and Remote Sensing,1995,33(2):457-465.
- [23] ZHANG Y,ZHANG K C,AN Z S,et al. Quantification of driving factors on NDVI in oasis-desert ecotone using geographical detector method[J]. Journal of Mountain Science,2019,16(11):2615-2624.
- [24] 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.

Data Griding of Population in Fujian Province Based on Modified DMSP/OLS

HUANG Yaoyi,CHEN Wencheng

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

Abstract: According to the data of resident population, DMSP/OLS, NDVI in Fujian Province, the vegetation index studied can effectively reduce the oversaturation of the night light data, and improve the HSI index model with the vegetation index, under the improved HSI index, based on the number of grids of the resident population, the weighted cases and the weighted least squares were modeled by the grid number of the resident population, and then the population zero-error adjustment was carried out to get the grid of the population statistics data of Fujian Province at the scale of $1\text{ km}\times 1\text{ km}$. The results showed that the improved HSI index model combined with the geographical detector can optimize the night light oversaturation, effectively avoid the variable panel effect, improve the fitting accuracy of the model, to obtain better effect of population gridding, and to be more in line with the actual spatial distribution of population.

Keywords: DMSP/OLS; NDVI; model modification; population gridding; Fujian Province

(责任编辑 杨珠)