

文章编号:1671-4229(2023)04-0020-09

4种典型人口空间化产品精度的比较评价

徐勇^{1,2}, 郑梓聪¹, 曹泳茵¹, 郑从威¹, 朱竑^{1,2*}

(1. 广州大学 地球科学与遥感学院, 广东 广州 510006;

2. 南方海洋科学与工程广东省实验室(珠海), 广东 珠海 519080)

摘要: 人口作为社会经济的重要指标,对城市韧性及可持续发展研究具有重要意义。国内外研究团队利用多源地理遥感数据已然研制了不同的全球人口空间化的数据产品。然而,受限于真实人口统计调查数据的长周期性和人口分布空间的复杂性,现有产品对于模拟国家级人口空间分布的实际准确度并未得到充分验证。本研究利用不同年份的人口普查数据与对应年份的现有4种人口空间数据集(GPW、Landsat、Worldpop、Cnpop)进行验证及评价。结果显示,4种不同人口产品对于模拟我国2010年和2015年区县级人口的总体精度均达到75%以上。其中,Worldpop和Cnpop数据产品对于模拟我国2010年的人口空间分布更精确,而Worldpop和Landsat产品对于模拟我国2015年的人口空间分布效果更优。同时,研究还发现,Worldpop数据产品对于模拟我国高密度人口地区,特别是城乡过渡带总体精度较高,即更加适宜于城市区域的相关研究。本研究为学界合理选择和使用不同人口空间化产品,以提高研究成果的精确度和针对性提供了重要参考。

关键词: 人口数据集; 空间分布; 多源遥感; 精度评价

中图分类号: K 909 **文献标志码:** A

Comparative assessment of the accuracy of four typical gridded population datasets in China

XU Yong^{1,2}, ZHENG Zi-chong¹, CAO Yong-yin¹, ZHENG Cong-wei¹, ZHU Hong^{1,2*}

(1. School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China;

2. Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, China)

Abstract: As one of the most important socioeconomic indicators, population data is vital for urban resilience and urban sustainability studies. Based on multi-sourced geographical data, various gridded population datasets have been developed by several international teams. However, due to the long-term periodicity of demographic surveying data as well as the uneven distributions of population in regions of China, the overall accuracy of existing population datasets in simulating the real population distributions at a national level in China has not been fully assessed and verified. The purpose of this study is thus to comprehensively validate and evaluate the accuracy of existing gridded population datasets based on the census data at different times. Among them, the verified products include four mainstreaming gridded population datasets, including GPW, Landsat, Worldpop and Cnpop developed by the Chinese Academy of Sciences. The results show that the overall accuracies of the four different population products in simulating the distribution of population in China at years 2010 and 2015

收稿日期: 2022-10-25; 修回日期: 2023-06-26

基金项目: 国家自然科学基金资助项目(41971184, 42071394)

作者简介: 徐勇(1982—), 男, 副教授, 博士. E-mail: xu1129@gzhu.edu.cn

* 通信作者. E-mail: zhuhong@gzhu.edu.cn

引文格式: 徐勇, 郑梓聪, 曹泳茵, 等. 4种典型人口空间化产品精度的比较评价[J]. 广州大学学报(自然科学版), 2023, 22(4): 20-28.

are above 75%. Among them, the Worldpop and Cnpop data products are more accurate in simulating the distribution of population in China at year 2010, while Worldpop and Landscan products perform better in acquiring the distribution of population at year 2015. Meanwhile, the study also found that the Worldpop product has higher overall accuracy in simulating the population of high-density urban areas in China, like urban-rural transition zones, indicating that the Worldpop might be suitable for the studies in large cities with more high-density urban areas. The findings and results could provide an important guide for the community to reasonably select and utilize different gridded population products, and thus avoid unnecessary errors and improve the accuracy and pertinence of related research.

Key words: population dataset; spatial distribution; multi-sourced remote sensing; accuracy assessment

人口指标反映了特定时间和空间范围内人类的真实聚集状况。基础人口数据对于城市环境、城市规划、公共健康、资源管理及社会治理等研究至关重要^[1]。通常人口统计数据是基于一定行政边界,进行实地人口统计调查而得到的。例如我国的人口普查工作主要基于每10年一次的全国人口普查和每5年一次的1%人口抽样调查而得到。其余年份则一般基于人口普查数据进行插值和平滑等方法进行估算而得到。由于人口统计数据获取的时间周期长,难以及时反映城市地区实时人口规模及动态迁移的状态,如春运期间的人口流动及其交通拥堵风险评估等,这些人口信息的不确定性为城市的规划和有序管理带了诸多不便。面对大数据带来的各种新的可能性,近来也不断出现利用百度迁徙数据来探讨城市间人口的流动^[2]、利用手机信令数据对城际居民出行的关注^[3],以及利用腾讯社交用户数据估算全国人口空间分布^[4]等的研究,并取得了一定进展。与此同时,由于新兴社交媒体及手机用户的不穷尽性,这些相关研究反映的人口特征亦具有明显的局限性和数据缺陷。

随着对地观测技术的发展,无接触式的遥感观测数据为快速估算和模拟城市大范围人口分布及动态信息带来了新的可能^[5-8]。目前,有两类主要遥感数据被广泛用于城市人口的估算研究,即夜间灯光遥感数据和传统光学遥感数据。其中,夜间灯光数据主要通过人类的夜晚社会经济活动丰富程度来反映人口规模,如美国的DMSP/OLS、NPP/VIIRS夜光遥感数据及我国自主研发的珞珈一号和仰望一号夜光遥感数据。早期夜光遥感数据受制于其有限的空间和辐射分辨率,主要

被用于探讨国家或区域等大尺度行政单元人口规模,如Elvidge等^[9]尝试利用DMSP/OLS夜光遥感数据探测城市内部的人口规模,发现DMSP/OLS数据在反映城市内部人口分布空间差异方面存在亮度值过饱和的先天缺陷。现有NPP-VIIRS夜光遥感数据,无论是在空间分辨率上,还是在辐射数值上,较原来的DMSP-OLS夜光数据有了较大的提升^[10-11]。利用NPP-VIIRS夜光数据,Wang等^[7]采用非线性模型,探讨了夜光数据同我国城市人口规模之间的关系,发现夜光总强度因子要优于其他夜光遥感因子(如夜光平均强度、夜光区域)。Li等^[12]利用夜光遥感对北京市各区的人口进行定量模拟,并结合真实人口调查数据验证了利用夜光遥感数据对城市人口估算的精度约为62%,反映出与真实数据尚有较大的误差。类似地,Zhang等^[13]探讨了我国自主研发的珞珈一号夜光遥感数据对于省市级人口调查的应用潜力,也发现夜光数据可以较好地模拟省市人口总数,回归R²达到0.6以上。尽管以上工作目前还都难以完全真实地反映人口的实际情况,但这在一定程度上说明,可以利用精确度更高的夜光指数来研究中宏观尺度城市区域人口的分布及变化。

除了夜光遥感数据,中高分辨率可见光遥感数据也常常被用于人口活动状况的监测,并成为估算人口规模的有效方法之一。有别于夜光遥感数据,可见光遥感数据主要依据人的物理活动空间(如工业区、居民点及建筑物等)来估算城市的人口规模。Stevens等^[14]就选择以遥感数据为主,开源地理信息数据为辅的方式,利用随机森林方法估算得到世界范围内人口分布信息(被称为Worldpop人口计划)。类似地,哥伦比亚大学发起

的 Gridded Population of the World (GPW) 人口计划,探讨结合遥感及开放地理信息数据以生成全球范围的人口分布信息。此外,欧洲学者 Freire 等^[15]利用中低分辨率陆地遥感数据,率先提取了全球人居环境分布图(Global Human Settlement Layer, GHSL),再在人居环境分布图的基础上,改善已有人口数据产品在部分国家存在数据丢失或不一致引发的问题,该方案已在圭亚那、黎巴嫩等国家取得一定效果。我国学者柏中强等^[16]尝试利用土地利用及夜光等数据,开展我国人口数据的空间化研究,并利用陕西省延安市 24 个乡镇人口进行验证,结果显示,有 19 个乡镇的估计人口数与普查统计值误差绝对值小于 10%。

综合国内外已有的研究成果, Bai 等^[17]对比分析了不同人口数据产品的精度,包括 WorldPop 及 GPW 人口数据产品,发现中国科学院资源环境科学与数据中心(Cnpop, 后简称中科院)及美国 WorldPop 人口产品在同类结果中,对我国人口分布的模拟效果表现得最好。Xu 等^[18]利用 GPW、GHS、Landsat 以及 Worldpop 4 种数据集对 2015 年中国滇黔桂乡镇一级地区进行人口估算,与人口统计数据进行对比,结果表明在乡镇一级尺度下,Worldpop 数据集往往能更好地描绘滇黔桂地区的人口空间分布。

虽然,已有一些研究对不同人口产品进行了精度的分析。然而,现有分析结果均表明,当利用我国不同地区或不同时期人口调查数据去评估不同人口产品精度时,存在不一致的研究结论,导致用户难以作出合理的数据选择。如 Bai 等^[17]对不同人口数据的验证结果表明,各类人口数据的模拟准确度约为 60% 左右,而谭敏等^[19]对 2010 年广州市 166 个街道的人口模拟结果显示不同人口产品的精度可以高达 80%。同时,不同人口数据产品的性能表现各异,有学者发现 Worldpop 结果相对其他产品对于模拟我国乡镇人口,特别是我国东南地区人口,总体表现较优^[17]。亦有学者发现 GPW 数据对于模拟广州街道的人口情况,总体要优于 Worldpop 等人口数据^[19]。还有学者发现 Landsat 产品对于模拟一些东南亚国家(如老挝、泰国等)省级人口时,总体要优于 Worldpop 及 GPW 等人口产品^[20]。这些不太一致的研究结论,给学者选择和合理使用不同人口数据带来

了困惑。

综上所述,为了让用户能够合理地选择合适的的人口数据产品,加强研究的针对性。本研究将利用我国区县尺度多年份的人口统计调查数据,拟系统评价现有国内外几种典型人口空间数据产品的精度和适用性。同时,研究将充分结合我国人口分布差异大、流动性强等特点,对不同时人口数据产品进行综合对比和分析,以期提高数据选择的精准性和有效性。

1 研究区域与数据

1.1 研究区概况

根据第七次人口普查,我国拥有超过 14 亿的人口。由于我国自然资源分布及区域发展不平衡,导致我国人口分布极不均匀,且人口流动性大。综合考虑研究范围大及统计数据的可获取性,本研究拟基于我国 2 800 多个区县级行政单元作为研究对象,分析和评价不同人口空间化数据产品的人口模拟性能和精度。图 1 为我国 2010 年人口普查数据所得的人口密度分布,红色表示人口密度高,而绿色表示人口密度小。从图 1 可以看出,我国东部人口相对密集,特别是东南沿海人口最为集中。人口统计数据亦表明,人口数量排名前五的省份依次是广东、山东、四川、河南和江苏,其人口均超过 7 800 万。本研究将充分对比和评价不同人口产品对于我国不同地区的适用度及精度,为合理选择和使用人口数据提供参考。

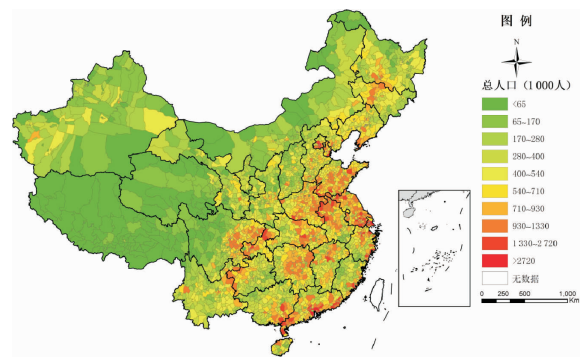


图 1 研究区域及相关人口统计数据

Fig. 1 Study area and relevant demographic data

注:基于自然资源部标准地图服务网站 GS(2019)1822 号标准地图制作,底图边界无修改。

1.2 主要数据

本研究涉及验证的是4套人口数据集,包括GPW人口数据集、Landscan人口数据集、Worldpop人口数据集和中科院制作的Cnpop人口数据集,

提供了年份以及空间分辨率等相关信息,见表1。为了更好地验证数据精度,研究采用我国2010年和2015年所有县级单元的人口数据进行精度验证。4类数据产品的情况如下:

表1 数据集相关信息
Table 1 Information about the datasets

数据集	数据来源	提供年份/年	提供空间分辨率
GPW	哥伦比亚大学国际地球科学信息网络中心和NASA的社会经济数据 https://sedac.ciesin.columbia.edu/	2000、2005、2010、2015、2020	1 km、5 km、30 km、55 km
Landscan	美国橡树岭实验室(ORNL) https://landscan.ornl.gov/	2000-2021	1 km
Worldpop	佛罗里达大学地理系和新型病原研究所 https://www.worldpop.org/	2000-2020	100 m、1 km
Cnpop	中国科学院地理科学与资源研究所 https://www.resdc.cn/	1995、2000、2005、2010、2015、2019	1 km

(1)GPW是由哥伦比亚大学国际地球科学信息网络中心和NASA的社会经济数据和应用中心合作完成。至今,GPW以及在SEDAC网站发布了4个版本的数据集。数据空间分辨率为30弧秒,精读相当于1 km。GPW的生产方法是通过面积加权法将人口普查数据分配到网格中,同时利用一些辅助数据,如湖泊、河流等。

(2)Landscan是由美国橡树岭实验室(ORNL)制作开发的一个世界范围内的30弧秒分辨率(即1 km分辨率)的人口数据集。该数据结合不同类型的土地利用数据、遥感数据以及城市基础设施数据等,利用多元分区密度模型以权重形式分解行政单元内的人口统计值到不同地理层,并依据数据层估算各区人口分布。

(3)Worldpop项目系由佛罗里达大学地理系和新型病原研究所一起开发,并设立在南安普敦大学地理与环境科学学院,旨在为中南美洲、亚洲和非洲地区提供人口数据集,为支持发展、灾害和卫生应用等研究。Worldpop数据集在实现人口空间化的过程中结合了人口普查数据和包括土地利用、夜间灯光及各类不同土地利用和地形等空间数据集为辅助,通过随机森林模型生成人口空间分布的预测结果。Worldpop数据产品精度有100 m和1 000 m两种分辨率版本。为了同其他产品保持一致,本文采用其分辨率为1 km数据开展精度评估研究。

(4)Cnpop数据集是由中科院提供的,其空间分辨率为1 km,利用全国分县人口统计数据和多

因子权重分配法,以实现人口空间化的人口空间分布数据集^[21]。

2 研究方法

2.1 定量评价

以真实的人口数据作为标准参考值,对不同人口空间的数据集进行精度评价。评价指标包括平均绝对误差(Mean Absolute Error, MAE),平均相对误差(Mean Relative Error, MRE)、均方根误差(Root Mean Square Error, RMSE)及相关系数(Correlation Coefficient, CC)。利用上述4种指标,分别对比分析不同人口数据集的精度,并制作误差空间分布图。

各评价指标公式如下:

$$MAE = \frac{1}{N} \sum |f_i - r_i|, \quad (1)$$

$$MRE = \frac{1}{N} \sum \frac{|f_i - r_i|}{r_i} \times 100\%, \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum (f_i - r_i)^2}, \quad (3)$$

$$CC = \frac{cov(f,r)}{\sigma_f \sigma_r}, \quad (4)$$

其中, f_i 表示第*i*个区县估算所得人口数值; r_i 表示第*i*个区县的真实人口统计数据; N 表示区县的个数; $cov(f,r)$ 表示估算人口数与人口普查数的协方差; σ_f 表示估算人口数的标准差; σ_r 表示人口普查人口数的标准差。

2.2 定性评价

定性分析采用两种方式:①基于同一分级方

式对不同人口分布产品进行划分,对比分析不同人口数据产品对于模拟我国人口分布的情况;②基于区县尺度单元,利用人口产品同真实人口统计调查数据制作散点图和开展回归分析,拟合优度可用于判读不同人口数据产品同真实人口的偏差及一致性情况。

3 结果与分析

3.1 人口数据产品的定性评价

图 2 显示了不同人口数据集所得的我国区县级人口模拟结果,其中,图 2(a)~图 2(d)分别表

示 GPW、Landscan、Worldpop 和 Cnpop 数据集对我国 2010 年人口分布的模拟结果。不同数据的模拟结果,均显示出东部人口相对密集,特别是东南沿海地区。亦可以发现,人口较多的省份包括广东、山东、河南、四川、江苏、安徽、湖南等地。相对于东部,西部地区人口较少,东西部人口差异明显。不同人口数据的 2010 年模拟结果同我国区县普查数据(图 1)所得人口分布基本一致,除 GPW 人口数据发现西部部分地区颜色稍深,表示部分地区人口具有一定高估的现象之外,其余 3 种人口数据均同真实人口普查数据分布高度一致。

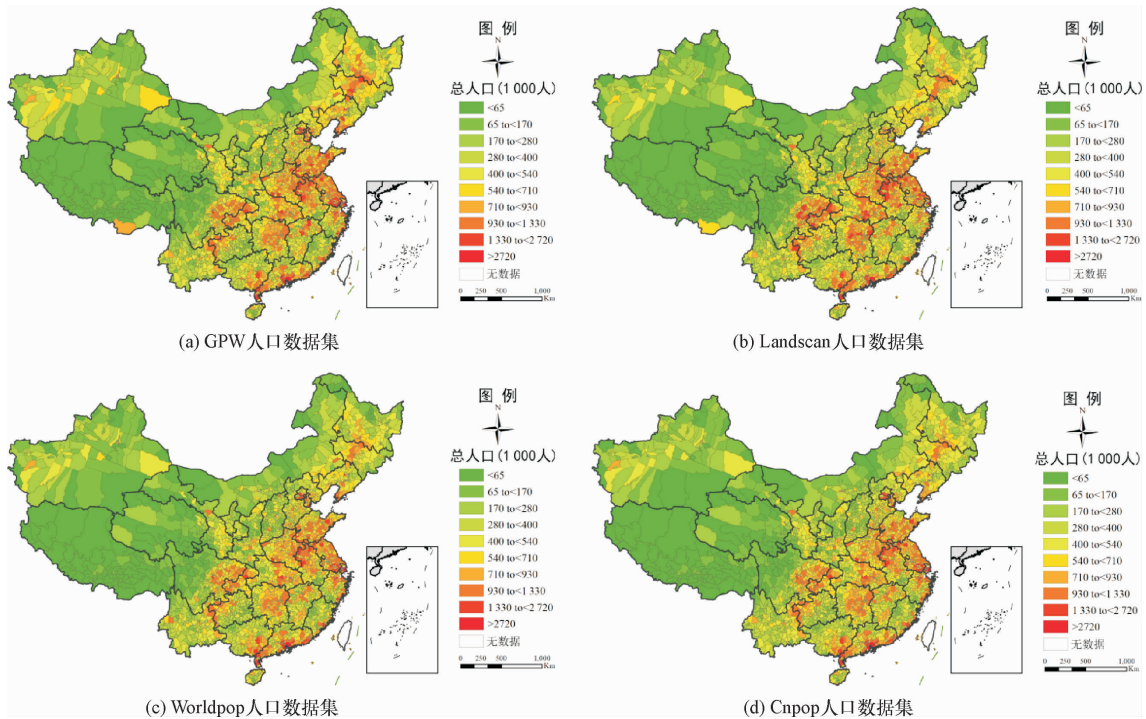


图 2 2010 年不同人口数据分布图

Fig. 2 Population distributions of different datasets for year 2010

注:基于自然资源部标准地图服务网站 GS(2019)1822 号地图制作,底图边界无修改。

图 3 所示为 2010 年区县范围内不同人口空间分布数据集所得结果与人口普查数据所做出的散点图,其中 X 轴为真实人口普查数据, Y 轴为不同人口数据集所得结果,散点越集中于斜率为 1 的直线,表示人口估算结果同人口普查数据越一致,即精度也越高。同时,利用线性函数对估算人口和人口普查数据进行直线拟合,拟合精度高,表示效果佳。结果显示,4 种不同人口空间化数据均具有较高的拟合优度, R^2 均达到 0.84 以上,能较

好地拟合我国人口的空间分布状态。其中,GPW、Landscan、Worldpop 和 Cnpop 4 种人口数据产品的拟合优度分别为 0.92,0.84,0.95 和 0.94。此外,结果亦发现,GPW 人口产品对我国区县人口存在一定程度人口高估现象,拟合系数为 1.09。而 Landscan 人口数据则存在着一定程度的人口低估,拟合系数为 0.85。与前面两种数据相比,2010 年 Worldpop 和 Cnpop 人口产品的总体精度要略优于 GPW 和 Landscan 产品。

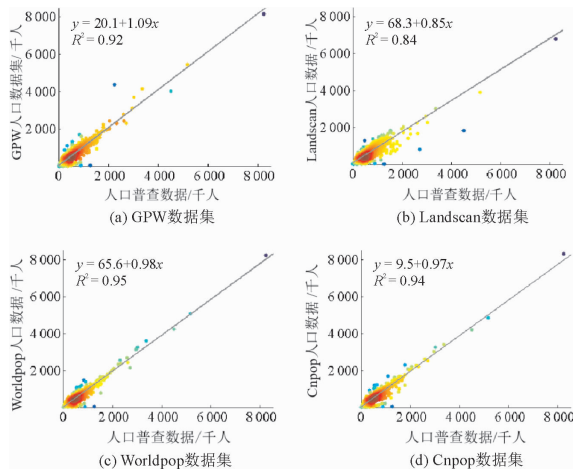


图3 4种人口数据集与人口普查数据的散点图

Fig.3 Scatter plots of four population data sets and census data

表2 基于2000、2010和2020年人口普查数据对不同人口数据产品相关系数

Table 2 Correlation coefficient of different population data products based on 2010,2010,2020 census data

数据集	皮尔逊相关系数			R ² 决定系数		
	2000年人口	2010年人口	2020年人口	2000年人口	2010年人口	2020年人口
GPW	0.928 **	0.954 **	0.943 **	0.862	0.924	0.890
Landscan	0.869 **	0.916 **	0.945 **	0.755	0.836	0.893
Worldpop	0.923 **	0.968 **	0.945 **	0.872	0.945	0.893
Cnpop	0.913 **	0.959 **	/	0.834	0.935	/

注:**. 在0.01级别(双尾),相关性显著。

表3 基于2010年人口普查数据对不同人口数据产品精度验证

Table 3 Precision verification of different population data products based on 2010 census data

数据集	MAE/万人	RMSE/万人	MRE	CC
GPW	8.9	12.3	0.25	0.95
Landscan	8.4	14.5	0.21	0.92
Worldpop	4.0	9.1	0.12	0.97
Cnpop	4.7	9.8	0.13	0.96

表4 基于2015年人口普查数据对不同人口数据产品精度验证

Table 4 Precision verification of different population data products based on 2015 census data

数据集	MAE/万人	RMSE/万人	MRE	CC
GPW	8.2	13.0	0.23	0.96
Landscan	7.7	12.3	0.21	0.96
Worldpop	6.8	11.7	0.19	0.96
Cnpop	9.0	15.1	0.25	0.94

3.2 人口数据产品的定量评价

基于2000年、2010年、2015年和2020年的我国区县人口普查数据,采用平均绝对误差(Mean Absolute Error, MAE)、平均相对误差(Mean Relative Error, MRE)、均方根误差(Root Mean Square Error, RMSE)和相关系数(Correlation Coefficient, CC)4种指标,分别对两个年份的不同人口数据产品进行精度验证,验证结果见表2~表4,其中表2表示基于2000年、2010年和2020年人口普查数据对不同人口数据产品相关系数,表3表示2010年不同人口产品的精度验证结果,表4表示2015年不同人口产品的精度验证结果。

基于不同精度评价指标的统计结果显示,2010年人口产品中,Worldpop和Cnpop两种数据整体精度较好,如Worldpop和Cnpop同人口普查数据的相关系数分别为0.97和0.96,优于GPW和Landscan人口数据的0.95和0.92,且其他指标具有一致的表现。这一发现也同图3所示散点图的结果一致,如Worldpop和Cnpop同人口普查数据的拟合优度(0.95和0.94)要高于GPW和Landscan数据(0.92和0.84),均表明Worldpop和Cnpop具有更好的精度。

相比于2010年精度验证结果,2015年的结果则有一定差异。利用不同指标,表4所得2015年不同产品的验证结果显示,Landscan和Worldpop具有较好的精度,优于GPW和Cnpop人口数据的结果,如Landscan和Worldpop的平均相对误差为0.21和0.19,小于GPW和Cnpop产品的0.23和0.25。这也表明,同一人口产品对我国不同时期的人口模拟结果也存在差异化表现。总体而言,

Worldpop 在不同时期均具有较好的表现,无论是 2010 年,还是 2015 年,各个指标均表现稳定,具有较高的精度。除了 Worldpop 人口数据,其他 3 种人口数据亦有不错的验证精度,在不同年份所有产品的总体误差均在 25% 之内。其中,2010 年人口产品中,Worldpop 和 Cnpop 数据要优于 GPW 和 Landscan 人口数据。而 2015 年,Worldpop 和 Landscan 数据要略优于 Cnpop 和 GPW 人口数据。这一结果也表明,Worldpop 产品对于模拟我国人口具有较好的稳定性,在不同年份均具有较优的结果。不同产品在不同年份的表现呈现一定的差异,主要因为使用的数据存在时间差异,同时,缺乏现势性人口普查数据作为有效训练数据也会导致结果发生变化。

3.3 人口数据产品的视觉评价

除了定量评价,视觉对比则可直观地显示不同人口产品的模拟效果。为了使样本具有一定的代表性,使用 2010 年 4 种不同人口数据 (GPW、Landscan、Worldpop、Cnpop) 与 2010 年人口普查数据作出误差空间分布图(图 4),可以看出各区域的误差,Worldpop 数据总体精度最好,在新疆和西藏等区域,Worldpop 和 Cnpop 数据要优于 GPW 和 Landscan 人口数据。在中国中部和沿海,4 种人口数据表现得比较平均。同时,在甘肃和陕西中部地区,4 种人口数据得误差都比较大。

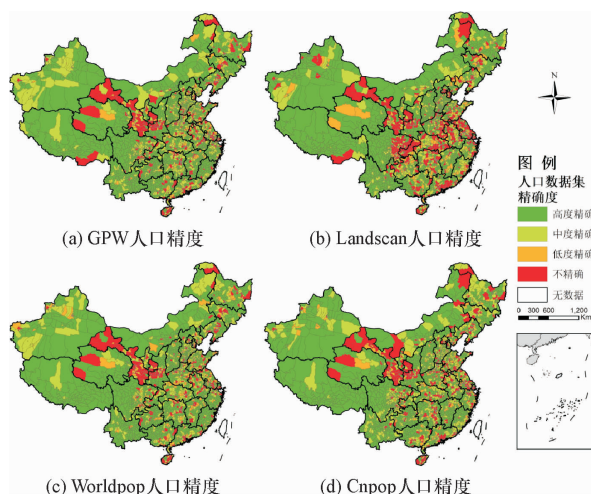


图 4 2010 年不同人口数据与真实值的误差空间分布图
Fig. 4 Distribution of population estimation error of different products for year 2010 when compared with actual census data

注:基于自然资源部标准服务网站 GS(2019)1822 号地图制作,底图边界无修改。

基于以上分析,研究选取我国中部城市武汉以及东部城市盐城进行对照分析,分别代表我国人口高密度及中低密度地区情况。如图 5 和图 6 分别表示 2010 年武汉和盐城 4 种不同数据集所对应人口空间分布及该区域的高分辨率遥感影像。从高分辨率遥感影像可清晰辨别不同土地利用类型,包括居民区、道路、植被及水体等地物类别。因此,可作为参考数据从侧面反映人口的真实空间分布。

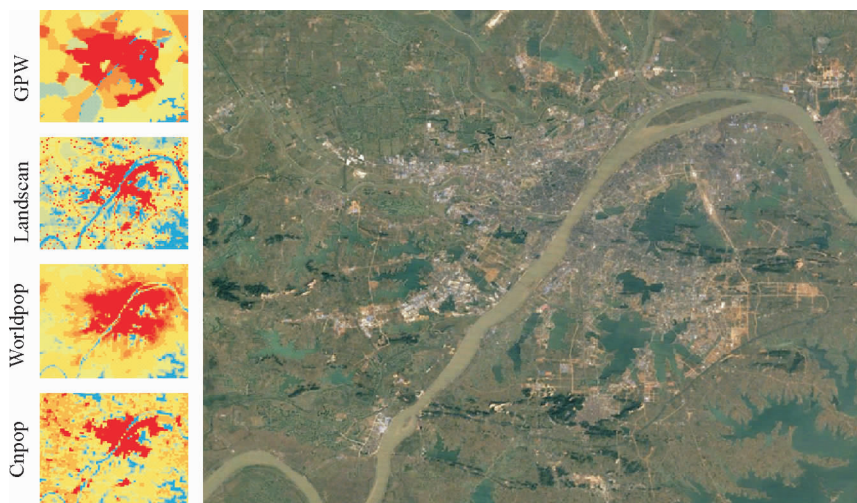


图 5 利用高分辨率遥感数据对 4 种人口数据集进行视觉对比:以武汉市为例

Fig. 5 Visual comparison of four population data sets using high-resolution remote sensing data; Wuhan city as an example

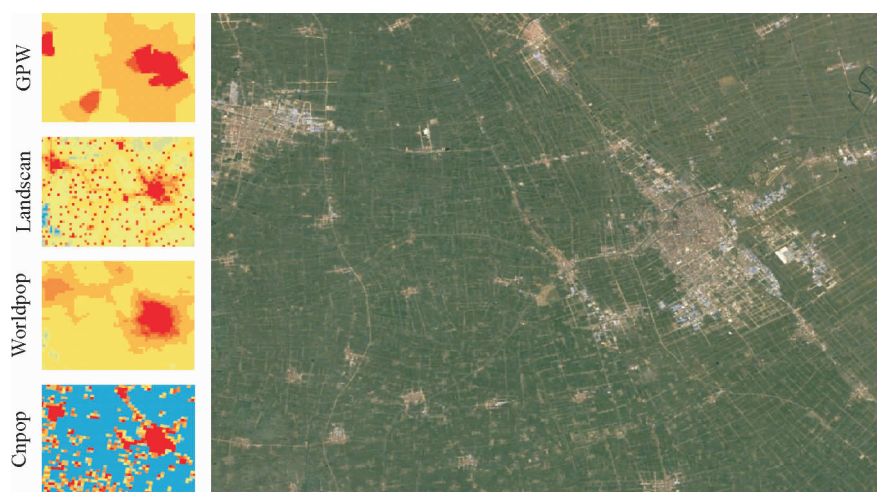


图6 利用高分辨率遥感数据对4种人口数据集进行视觉对比:以盐城市为例

Fig. 6 Visual comparison of four population data sets using high-resolution remote sensing data; Yancheng city as an example

图5和图6所示为不同数据集对我国中部城市武汉和东部城市盐城的人口模拟效果对比。从空间细节上看,由好到差依次为Landscan、Cnpop、Worldpop以及GPW人口数据。其中,GPW数据集由于采用了基于行政单元的面积加权法得到不同行政单元的平均人口密度,无法有效反映行政单元内部的人口空间差异,空间细节表现不佳。通过对照同一地区的高分辨率遥感数据,可以发现不同的人口数据集,均能较好地反映城乡人口差异,如深红色表示城市人口,浅黄到蓝色表示乡村人口。

从城市范围来看,GPW和Worldpop能较好地模拟城市核心区域。然而,相比于GPW和Worldpop数据,Landscan和Cnpop所得的城市核心区域有一定程度的缩小,无论是武汉,还是盐城均出现类似情况,偏小的区域主要发生在城乡过渡带。从数值来看,4种人口数据均能较好模拟城市核心地区人口,如深红色结果表示人口高密度地区。而对于城乡过渡带地区,Worldpop模拟效果最好,如Worldpop结果中出现深红到橙色的过渡带人口分布特征,而Landscan和Cnpop的模拟结果则在城乡过渡带中出现一定程度的人口低估现象。综合空间细节和城乡差异化模拟效果来看,Worldpop略优于其他数据产品。

4 结论与讨论

本研究利用我国2010年、2015年和2020年

人口普查数据,基于2800多个区县行政单元,对4种典型人口数据集,包括国际主流GPW、Landscan、Worldpop产品以及我国中科院Cnpop人口数据集,利用定性及定量相结合方法进行性能评价,为学者合理选择人口数据集提供参考依据。

基于不同精度指标,研究发现,4种典型人口数据,对于我国2010年和2015年区县级人口分布模拟总体误差均在25%以下,这意味着采用不同人口数据产品,均可较好地模拟我国区县人口分布,总体精度超过75%。其中,对于2010年人口模拟结果,Worldpop和Cnpop人口数据精度要优于GPW和Landscan人口数据,精度高出约8%~13%。而对于2015年人口模拟结果,Worldpop和Landscan人口数据产品要略优于GPW和Cnpop人口数据,精度高约2%~6%。相比其他3套数据集,Worldpop数据在2010年和2015年均具有较高精度,且各项指标均表现稳定。进一步研究亦发现,Worldpop数据集能够较好地模拟城乡过渡带的人口分布,而其他同类人口产品在该区域会出现一定程度的人口低估。

受限于测试数据的一致性及验证数据的可获得性,研究仅验证4种典型人口数据集,未有验证其他人口数据集,也未尝试利用更多时序人口普查数据进行验证。同时,考虑到新型社会感知数据的广泛使用,如腾讯位置服务、手机信令定位数据、健康码数据及天网监测大数据等,未来可结合多源地理遥感数据,采用我国最新的人口普查数

据,开展更广泛的人口空间化精细模拟和验证工作。同时,结合我国城市人口流动强度大、实时性强等特点,发展基于多源天空地数据的人口空间化自主技术,开展流动人口实时监测及预测预报研究,以期为未来城市发展及前瞻性空间规划研究提供快速、丰富和全面的人口基础调查数据。

参考文献:

- [1] 朱宇,丁金宏,王桂新,等. 近 40 年来的中国人口地理学——一个跨学科研究领域的进展[J]. 地理科学进展, 2017, 36(4): 466-482.
- [2] 叶强,张丽璇,彭鹏,等. 基于百度迁徙数据的长江中游城市群网络特征研究[J]. 经济地理, 2017, 37(8): 53-59.
- [3] 王梅红,侯笑宇,司连法,等. 地理空间数据结合手机信令等多源数据刻画城市居民出行特征[J]. 测绘通报, 2022(5): 162-165.
- [4] Xu Y, Song Y M, Cai J X, et al. Population mapping in China with Tencent social user and remote sensing data[J]. Applied Geography, 2021, 130: 102450.
- [5] 曹丽琴,李平湘,张良培. 基于 DMSP/OLS 夜间灯光数据的城市人口估算——以湖北省各州市为例[J]. 遥感信息, 2009, 24(1): 83-87.
- [6] 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.
- [7] Wang L T, Wang S X, Zhou Y, et al. Mapping population density in China between 1990 and 2010 using remote sensing [J]. Remote Sensing of Environment, 2018, 210: 269-281.
- [8] Chen J D, Fan W, Li K, et al. Fitting Chinese cities' population distributions using remote sensing satellite data[J]. Ecological Indicators, 2019, 98: 327-333.
- [9] Elvidge C D, Baugh K E, Dietz J B, et al. Radiance calibration of DMSP-OLS low-light imaging data of human settlements [J]. Remote Sensing of Environment, 1999, 68(1): 77-88.
- [10] Elvidge C D, Zhizhin M, Hsu F C, et al. VIIRS nightfire: Satellite pyrometry at night[J]. Remote Sensing, 2013, 5(9): 4423-4449.
- [11] Chen X, Nordhaus W. A test of the new VIIRS lights data set: Population and economic output in Africa[J]. Remote Sensing, 2015, 7(4): 4937-4947.
- [12] Li X M, Zhou W Q. Dasymetric mapping of urban population in China based on radiance corrected DMSP-OLS nighttime light and land cover data[J]. Science of the Total Environment, 2018, 643: 1248-1256.
- [13] Zhang G, Guo X Y, Li D R, et al. Evaluating the potential of LJ1-01 nighttime light data for modeling socio-economic parameters[J]. Sensors, 2019, 19(6): 1465.
- [14] Stevens F R, Gaughan A E, Linard C, et al. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data[J]. PLoS One, 2015, 10(2): e0107042.
- [15] Freire S, Schiavina M, Florczyk A J, et al. Enhanced data and methods for improving open and free global population grids: Putting 'leaving no one behind' into practice[J]. International Journal of Digital Earth, 2020, 13(1): 61-77.
- [16] 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.
- [17] Bai Z Q, Wang J L, Wang M M, et al. Accuracy assessment of multi-source gridded population distribution datasets in China[J]. Sustainability, 2018, 10(5): 1363.
- [18] Xu Y M, Ho H C, Knudby A, et al. Comparative assessment of gridded population data sets for complex topography: A study of Southwest China[J]. Population and Environment, 2021, 42(3): 360-378.
- [19] 谭敏,刘凯,柳林,等. 基于随机森林模型的珠江三角洲 30 m 格网人口空间化[J]. 地理科学进展, 2017, 36(10): 1304-1312.
- [20] Yin X, Li P, Feng Z M, et al. Which gridded population data product is better? Evidences from mainland Southeast Asia (MSEA)[J]. ISPRS International Journal of Geo-Information, 2021, 10(10): 681.
- [21] 付晶莹,江东,黄耀欢. 中国公里网格人口分布数据集[J]. 地理学报, 2014, 69(S1): 41-44.

【责任编辑:孙向荣】