

基于居住空间属性的人口数据空间化方法研究

董 南^{1,2}, 杨小唤^{1,2*}, 蔡红艳¹

(1. 中国科学院地理科学与资源研究所, 资源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:精细尺度的人口分布是当前人口地理学研究的热点和难点,在灾害评估、资源配置、智慧城市建设等方面应用广泛。居住建筑物尺度作为精细尺度的重要内容,其人口数据空间化日益引起学术界的关注。本文以居住建筑斑块面积、斑块内建筑面积比重、建筑物层数、公摊率等居住空间属性为人口分布数量的指示因子,以居住建筑的轮廓斑块为人口分布位置的指示因子,利用街道界线和街道常住人口数据为控制单元,建立线性模型,获得了2015年宣城市宣州区6个街道的居住建筑物尺度的人口分布矢量数据,刻画了城市市区人口空间分布的细节信息。结果表明:①以居住空间属性作为人口空间分布的指示因子,获取的人口空间数据精度高,结果可信。29个社区(村)估算人数的相对误差绝对值的平均值低于7%,其中25个社区(村)的相对误差绝对值低于10%。在1102个居住建筑斑块中,估算人数在合理区内的斑块个数占比高于74%,轻微低估区(-10%, 0)和轻微高估区(0, 10%)的斑块总数占比高于9%;②由斑块面积和建筑物层数共同表征的建筑物体积,是建筑物尺度上影响人口空间分布的关键因素;斑块内建筑面积比重属性能进一步提高模型精度;公摊率属性具有“降高升低”作用,但将估算人数调节到合理区的“能力”较弱。

关 键 词:人口;空间化;居住空间;居住建筑;斑块;宣城市宣州区

1 引言

人口的空间分布是指一定的时间点,人口在地域上的分布状况,是人口过程在空间上的表现形式,是人口地理学研究的核心问题(胡焕庸, 1983)。人口数据空间化是获取人口空间分布数据的有效途径,在精确刻画人口分布、多源数据融合研究等方面具有重要科学意义(符海月等, 2006; 林丽洁等, 2010; 柏中强等, 2013)。精细尺度的人口空间数据在灾害评估、资源配置、智慧城市建设等方面应用广泛。建筑物尺度是精细尺度的重要方面,该尺度上人口空间数据的获取,正引起学术界的重视(卓莉等, 2014; Jia et al, 2016)。目前,建筑物尺度上人口数据空间化研究较少,成果精度有待提高,因

此值得进行深入研究。

经过近30年的发展,人口数据空间化研究水平逐渐成熟。常用的空间化方法包括分区密度方法(Mennis, 2003; Gallego, 2010; Dmowska et al, 2014)、多元回归方法(卓莉等, 2005; 杨小唤等, 2006; 王磊等, 2011; 柏中强等, 2015)、多因素融合方法(廖顺宝等, 2003; Bhaduri et al, 2007; Zhang et al, 2015)、随机森林方法(Stevens et al, 2015; Gaughan et al, 2016)等。但是,其人口空间化数据成果大多以栅格呈现,格网大小多为20~1000 m,这种方式不可避免地会面临尺度效应带来的格网尺度适宜性问题(杜国明, 张树文, 张有全, 2007; 叶靖等, 2010; 王培震等, 2012; 李月娇等, 2014)。本文以矢量形式呈现人口空间数据,可以有效避免格网

收稿日期:2016-05;修订日期:2016-09。

基金项目:国家自然科学基金项目(41271173);国家科技支撑计划课题(2012BAI32B06) [Foundation: National Natural Science Foundation of China, No.41271173; National Science and Technology Support Program, No.2012BAI32B06]。

作者简介:董南(1984-),男,河北唐山人,博士研究生,主要从事人口地理、遥感与GIS应用研究,E-mail: dongnan67@126.com。

通讯作者:杨小唤(1965-),男,安徽潜山人,博士,研究员,主要从事遥感和GIS应用、资源环境科学空间数据库等方面的研究,E-mail: yangxh@igsnrr.ac.cn。

引用格式:董南, 杨小唤, 蔡红艳. 2016. 基于居住空间属性的人口数据空间化方法研究[J]. 地理科学进展, 35(11): 1317-1328. [Dong N, Yang X H, Cai H Y. 2016. A method for demographic data spatialization based on residential space attributes[J]. Progress in Geography, 35(11): 1317-1328.]. DOI: 10.18306/dlkxjz.2016.11.002

尺度适宜性研究问题。

以摄影测量与遥感为代表的现代测绘技术,为获取高分影像、大比例尺正射影像、地表三维立体提供技术支持,进而获取的大比例尺房屋矢量数据(谷国梁等, 2016)包括建筑物类型、轮廓、体积、层高(Lwin et al, 2009; Ural et al, 2011; Lung et al, 2013)等精细地理数据,其应用能将人口空间数据的分辨率提高到10 m级或建筑物尺度。因此,本文选取居住建筑斑块面积、斑块内建筑面积比重、建筑物层数、公摊率等居住空间属性作为人口分布数量的指示因子,以居住建筑斑块作为人口分布位置的指示因子,以街道界线、街道常住人口作为控制单元,建立线性回归模型。以安徽省宣城市宣州区6个街道为研究区,提出基于居住空间属性的人口数据空间化方案,探讨不同居住空间属性组合对模型精度的影响,获得了研究区2015年居住建筑尺度上人口分布矢量数据,以期丰富精细尺度人口空间化方法及数据成果。

2 研究区域及数据来源

2.1 研究区域概况

研究区宣城市宣州区位于安徽省东南部,包括

双桥、鳌峰、西林、澄江、济川、敬亭山6个街道,含29个社区(村),总面积约90 km²,东西跨度约10 km,南北跨度约16 km(图1)。截至2015年,研究区常住人口超过23万人,是宣州区乃至宣城市的人口集中区域。研究区内居住建筑以多层楼房(4~6层)为主,多分布在市区中部;其次以平房居多,多分布在敬亭山街道、双桥街道、澄江街道北部、济川街道东部、鳌峰街道南部(图2)。

2.2 常住人口及行政界线数据

通过对各街道办事处进行实地调研,获取各街道、社区(村)2015年的常住人口数据;从宣州区民政局获取各街道、社区(村)的行政界线图片资料,并以Quick Bird遥感影像为底图,进行矢量化生成行政界线矢量数据。

2.3 居住建筑斑块提取

居住建筑是供人们生活起居用的建筑物,包括普通住宅、公寓、别墅、宿舍等。居住建筑斑块,是指对遥感影像进行目视解译,通过矢量化方法提取居住建筑物的轮廓多边形。本文选用Quick Bird遥感影像(拍摄时间为2015年7月),根据居住建筑在遥感影像上反映的颜色、形状、相关布局、地域分布等,并参考其他学者关于不同类型住宅用地的解译标志(杜国明, 张树文, 2007; 陈振拓等, 2012),建立

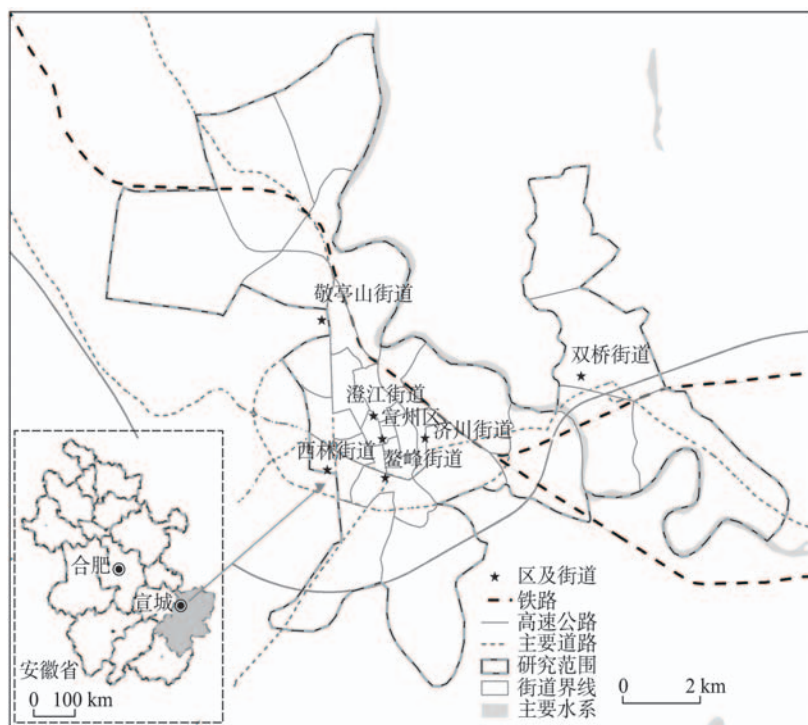


图1 研究区地理位置示意图

Fig.1 Geographical location of the study area

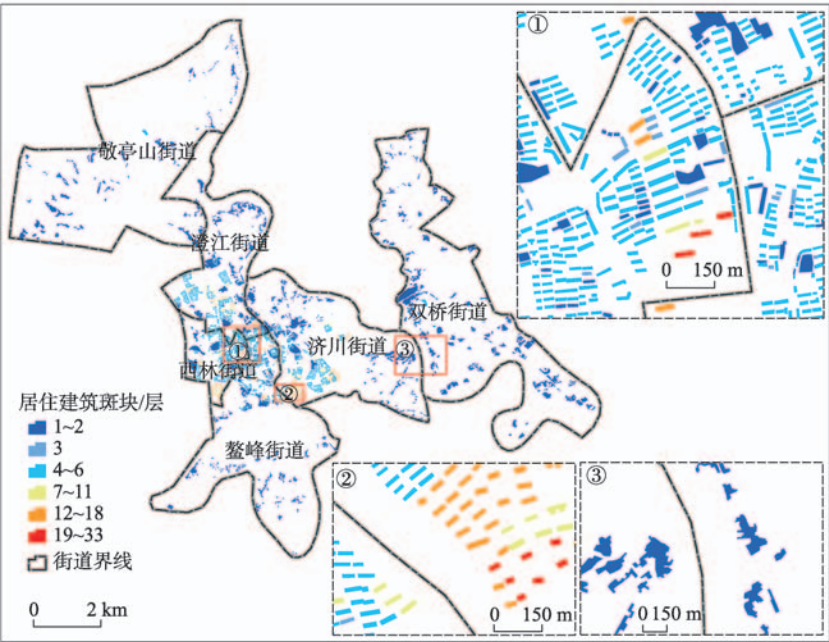


图2 居住建筑斑块空间分布图

Fig.2 Spatial distribution of residential building patches in Xuanzhou District, Anhui Province

居住建筑斑块解译标志(表 1)。在此解译标志基础上,借助百度街景地图、搜狗建筑物三维景观图等辅助数据,对遥感影像进行目视解译,去除公共建筑、工业建筑及农业建筑,提取居住建筑的外轮廓图斑。此种做法可最大限度地消除非居住建筑用地对人口空间分布的影响。

2.3.1 提取规则

(1) 1~2层居住建筑,以住宅用地斑块为最小提取单元。在该斑块内,以居住建筑物为主要部分,同时含有满足生活需求的附属设施用地,如建筑物之间的植被、绿化带、空地、道路等区域(如果相邻建筑物间的最大间距大于 15 m,则分别生成不同斑块,旨在降低斑块内附属设施用地比重过大引起的影响)。

该类居住建筑多位于城中村、乡村地区,乡村

的居住建筑遵循“一户一宅”的原则,即便是盖的 2 层小楼,依旧是一户人家;同时,乡村的居住建筑具有排列无序、大小不一及人口密度低等特点,因此,将居住建筑与附属设施用地作为整体,来提取住宅用地斑块,用该斑块而非单个的居住建筑斑块作为提取单元是适宜的。

(2) 3 层及以上居住建筑,以单个居住建筑物(独栋居民楼)为最小提取单元。楼房是研究区内居住建筑的主体,更是人口集中区的载体。在人口集中区,以单个居住建筑物作为最小提取单元,能有效地避免居住建筑之外的用地对人口分布的影响。

2.3.2 斑块提取及居住空间属性信息获取

依据以上原则,共提取 2862 个居住建筑斑块(图 2),其中 4~6 层居住建筑斑块比重最大,为 58.0%;1~2 层居住建筑斑块比重次之,为 30.1%。

表 1 居住建筑斑块解译标志

Tab.1 Interpretation keys for residential building patches

类型	解译标志描述			
	颜色	形状	相关布局	地域分布
1~2层居住建筑	屋顶以暗红、灰色为主;阴影不明显	房屋朝向不一致;房屋大小不一,几何形状规则	房屋排列无序,房屋间距不等,周围为绿色植被、耕地等	城乡结合处,城中村
3层及以上居住建筑	屋顶以暗红、暗灰、灰白色为主;阴影显著,并随楼房的变高而变长	方向特征显著,并且一致;几何形状规则且明显	房屋排列有序,与其他居住建筑呈近似等间距关系,周围绿化带及道路明显	城市市区

斑块的居住空间属性包括:

(1) 斑块面积(*Area*)

斑块面积(*Area*)是衡量居住空间大小的关键因素之一,依据GIS的Calculate Geometry工具自动计算Krasovsky 1940 Albers投影坐标系统下的各居住建筑斑块面积。

(2) 斑块内建筑面积比重(*PHA*)

斑块内建筑面积比重(*PHA*)即斑块内居住建筑物投影面积之和与该斑块总面积的比例。对1~2层居住建筑斑块,设置此空间属性,能够避免生活附属设施用地面积对人口分布数量的影响。*PHA*的确定:1~2层斑块,该属性值等于斑块内居住建筑的面积极量测值/斑块总面积,计算结果精确到0.05;3层及以上层数斑块,该属性值都为1。

(3) 斑块内建筑物层数(*BFN*)

城市建设为了集约利用土地资源以及解决大量人口的居住问题,多层及高层居住建筑的比重很大,因此层数属性必不可少。*BFN*的确定:1~2层斑块,依据居住建筑在遥感影像上的地物特征,能够准确获取层数;3层及以上层数斑块,依据“宣城房产网”提供的楼房层数信息、百度街景地图及搜狗的建筑物三维景观图,获取层数属性。

(4) 公摊率(*k*)

公摊率(*k*)为公摊面积与建筑面积的比值。对于楼房建筑来说,公摊率也是影响居住空间的因素。目前,国家对公摊率没有明确规定数值范围,依据房地产行业关于公摊率或公摊系数的参考值范围(有哈相伴, 2009; 王立新, 2014; 胡长菊, 2015),以相对平均水平标准确定方案 k^1 ;通常情况下,高层住宅因有电梯、消防、设备管线等公共设施,公摊面积相对较大,而多层住宅因无电梯等公共设施,公摊面积相对较小(王立新, 2014; 胡长菊, 2015),因此将7层以上住宅的公摊率在平均水平的基础上上浮,将6层以下住宅的公摊率在平均水平的基础上下浮,确定方案 k^2 (表2)。

(5) 居住建筑房屋套数、建筑面积属性

套数、建筑面积数据主要用于模型结果精度的

验证。数据来源于“宣城房产网”提供的楼房房屋套数、建筑面积等信息;此外,利用搜狗的建筑物三维景观图对套数信息进行补充。

3 基于居住空间属性的人口数据空间化建模

本文以午夜时刻为时间节点,研究常住人口在住宅用地上的静态空间分布,侧重揭示人口在住宅用地内部的分布差异。

3.1 模型提出

本文以居住建筑斑块面积(*Area*)、斑块内建筑面积比重(*PHA*)、斑块内建筑物层数(*BFN*)、公摊率(*k*)来描述居住空间,建立常住人口与居住建筑的空间属性的关系,实现建筑物尺度上人口数据空间化。

依据基于土地利用数据的人口数据空间化模型(江东等, 2002; 杨小唤等, 2002)可知,人口数量与各土地利用类型面积存在线性关系。其中住宅用地是指用于人们生活居住的房基地及其附属设施的土地,是土地利用类型中最直接反映人口分布的因子(刘焕金, 2012)。其他土地利用类型,如耕地、园地、林地、草地等,虽会存在一定的人口,但主要是少量的、临时性的人口,在规模和存在时间上都无法与住宅用地相比。因此,鉴于住宅用地的突出及独特地位,在进行人口数据空间化时,侧重研究人口在住宅用地内部的分布差异,认为其他土地利用类型均无人口分布,即将人口展布在居住建筑上。

首先,构建人口数量与居住建筑斑块面积的线性回归模型,数学表达式见表3中的类型1。其次,对于1~2层居住建筑斑块,斑块内建筑面积比重(*PHA*)会直接影响该斑块内人口数量,引入*PHA*消除斑块内生活附属设施用地面积对人口数量的影响。另外,居住空间是一个三维问题,引入斑块内建筑物层数(*BFN*)能更好地体现居住空间的大小。同时,为探讨*PHA*和*BFN*对人口空间分布的综合影响,引入这2个属性信息建立模型。数学表达式见表3中的类型2、类型3和类型4。对于楼房建筑来说,公摊率也是居住空间的固有属性。研究区内,层数大于4层的楼房,其建筑面积占总研究区建筑面积的72.6%,故引入公摊率(*k*),探讨其对人口空间分布的影响。数学表达式见表3中的类型5和

表2 居住建筑物公摊率方案

Tab.2 Public area rate for residential buildings

公摊率(<i>k</i>)	层数			
	1~3	4~6	7~18	19~
方案 k^1	0.0	0.10	0.15	0.20
方案 k^2	0.0	0.08	0.18	0.22

表3 基于居住空间属性的人口数据空间化模型

Tab.3 Models for demographic data spatialization based on residential space attributes

模型名称	回归模型
类型 1	$\hat{P}=a\times Area+b$
类型 2	$\hat{P}=a\times Area\times PHA+b$
类型 3	$\hat{P}=a\times Area\times BFN+b$
类型 4	$\hat{P}=a\times Area\times PHA\times BFN+b$
类型 5	$\hat{P}=a\times Area\times PHA\times BFN\times(1-k^1)+b$
类型 6	$\hat{P}=a\times Area\times PHA\times BFN\times(1-k^2)+b$

注： \hat{P} 为单个居住建筑斑块的估算人口； a 、 b 为模型系数。

类型 6。

3.2 各类型的比较

分别汇总每个社区(村)的所有斑块的 6 个参数值： $Area$ 、 $Area\times PHA$ 、 $Area\times BFN$ 、 $Area\times PHA\times BFN$ 、 $Area\times PHA\times BFN\times(1-k^1)$ 、 $Area\times PHA\times BFN\times(1-k^2)$ ，将 29 个社区(村)的人口数量与对应社区(村)的参数值分别进行线性回归分析，根据计算结果，用决定系数指标来判定线性回归模型对样本数据的拟合程度。

通过比较发现(表 4)，类型 1 和类型 2 的决定系数较低，估计值误差的标准差较高，表明采用 $Area$ 、 PHA 不能准确推算人数。引入 BFN 属性的类型 3，决定系数达到 0.914，与类型 2 相比，提高幅度很大，估计值误差标准差下降很多，表明由 $Area$ 和 BFN 共同表征的建筑物体积是准确估算人数的关键因素，这与建筑物体积是精细尺度人口分布模拟的最优因子的结论(Dong et al, 2010; Silván-Cárdenas et al, 2010; 卓莉等, 2014)一致。与类型 3 相比，类型 4 的决定系数虽然只提高了 0.06，但是估计值误差标准差下降了 40.5%，同时类型 4 各项拟合度参数也是 6 个类型中的最优值，表明在 $Area$ 、 BFN 属性的基础上，引入 PHA 能更好地估算斑块人数。与类型 4 相

表4 回归模型拟合度参数对比

Tab.4 Comparison of fitting parameters of regression models

模型名称	相关系数	决定系数	校正决定系数	估计值误差标准差
类型 1	0.173	0.030	-0.006	8097.17
类型 2	0.438	0.192	0.162	7388.19
类型 3	0.956	0.914	0.911	2405.26
类型 4	0.985	0.970	0.969	1430.66
类型 5	0.982	0.965	0.963	1548.80
类型 6	0.983	0.965	0.964	1530.76

比，类型 5 和类型 6 的决定系数有所降低，估计值误差标准差有所升高，但幅度不大。结合实际情况，公摊率也可能具有一定的使用价值，选取类型 4、类型 5 和类型 6 进行下一步讨论。

3.3 模型的显著性检验

在 3.2 小节中，人口数量与居住空间属性之间的关系仅被假设为线性关系，这种假设是否恰当，能否确保所使用的回归模型具有统计学意义，必须通过对模型的假设检验来说明。依据方差分析显示的假设检验结果可知(表 5)，显著性水平均小于 0.05，说明类型 4、类型 5 和类型 6 均具有统计学意义，能用于表达人口数量与相关居住空间属性变量之间的线性关系。

3.4 人口数据空间化模型建立

分乡镇控制是获取精细尺度人口空间数据的前提条件之一。胡焕庸先生(1936)在“句容县之人口分布”中阐述：“制作精密之人口地图，必须备有各县分乡人口统计，及分乡区划地图。”可见乡镇(街道)级别的人口统计数据 and 行政区划图是绘制高精度人口密度图的 2 个必要条件。

本文以街道为建模控制单元，旨在减少模型空间尺度转换的跨度，保证精细尺度人口数据空间化

表5 类型4、类型5与类型6的显著性检验分析表

Tab.5 Significance test of type 4, type 5, and type 6 models

模型名称	变差来源	平方和	自由度	均方	F统计量	显著性水平(Sig.)
类型 4	回归	1.769E+09	1	1.769E+09	864.431	0.000
	残差	55263490	27	2046795.9		
	总离差	1.825E+09	28	—	—	—
类型 5	回归	1.760E+09	1	1.760E+09	733.626	0.000
	残差	64767166	27	2398783.9		
	总离差	1.825E+09	28	—	—	—
类型 6	回归	1.761E+09	1	1.761E+09	751.662	0.000
	残差	63266949	27	2343220.3		
	总离差	1.825E+09	28	—	—	—

的精度。同一街道内,人均居住空间相似,即街道内各个居住建筑斑块的人均居住空间系数 a 相同,同时根据“无土地则无人口”的原则,即无居住建筑则无人口,常数项 $b=0$ 。从而确定类型4、类型5、类型6的公式:

$$\begin{aligned} \text{类型4: } P_i &= a_i \times \sum_{j=1}^n (Area_{ij} \times PHA_{ij} \times BFN_{ij}) \\ P_i &= a_i \times \\ \text{类型5: } \sum_{j=1}^n (Area_{ij} \times PHA_{ij} \times BFN_{ij} \times (1 - k_{ij}^1)) & \quad (1) \\ P_i &= a_i \times \\ \text{类型6: } \sum_{j=1}^n (Area_{ij} \times PHA_{ij} \times BFN_{ij} \times (1 - k_{ij}^2)) \end{aligned}$$

式中: P_i 为街道 i 的常住人口; a_i 为街道 i 人均居住空间系数; $Area_{ij}$ 、 PHA_{ij} 、 BFN_{ij} 为街道 i 斑块 j 的斑块面积、斑块内建筑面积比重、建筑物层数; k_{ij}^1 、 k_{ij}^2 分别为方案1、方案2确定的公摊率。

图3为依据类型4获取的宣州区6个街道的人口空间分布矢量数据。该数据展示了居住建筑物尺度上的人口数量,刻画了城市市区人口空间分布的细节信息。即人口主要集中在市区中部,敬亭山街道、双桥街道北部及鳌峰街道南部的人口分布较少。

4 精度验证及误差分析

4.1 精度验证

精度验证拟从2个方面进行:一是社区(村)尺度验证。以街道尺度进行人口数据空间化,对其结果进行精度验证,最具有说服力的方法就是用低一级行政区套合人口空间数据,同对应行政区的人口统计数据进行比较,本文用相对误差及误差范围内社区(村)个数指标进行精度评价;二是居住建筑尺度验证。用居住建筑斑块的估算人数,同该斑块内人数参考真值进行比较,该方案更能从精细尺度上体现人口数据空间化模型的精度。人数参考真值等于每套房屋的平均人数乘以斑块内房屋套数。

(1) 社区(村)尺度验证

对人口空间化结果进行基于29个社区的相对误差统计分析(图4)。结果显示:类型4、类型5和类型6的相对误差绝对值的平均值低于7%,25个社区(村)的相对误差绝对值低于10%。表明类型4、类型5和类型6的精度较高,类型4具有微弱的精度优势。

(2) 居住建筑尺度验证

从2862个居住建筑斑块中,随机选取分布在29个社区的1102个居住建筑斑块,对估算人数进行验证。本文以每套房屋平均居住2~3人,确定每

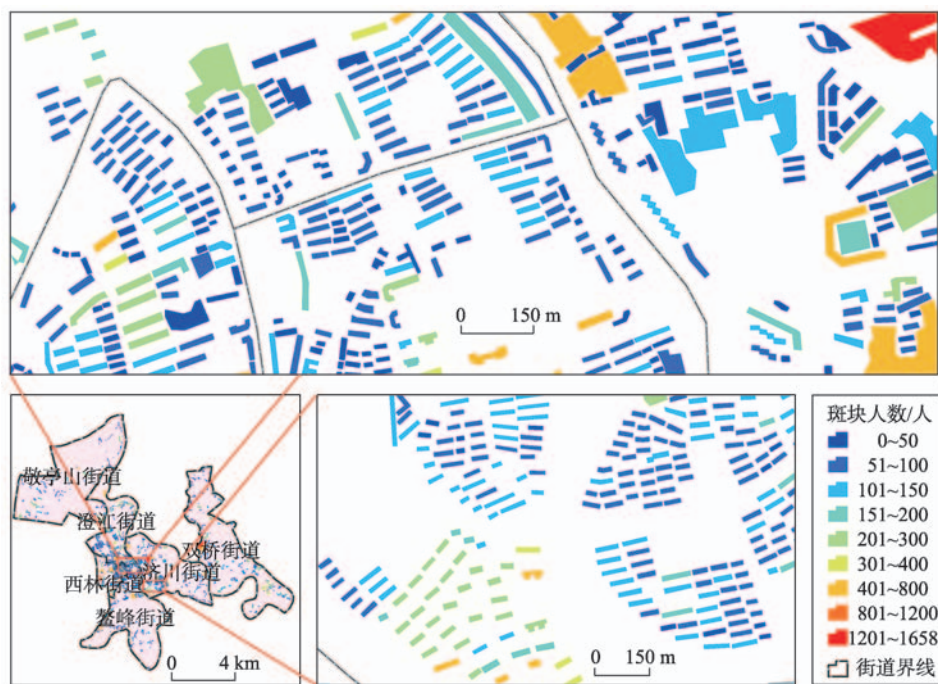


图3 2015年宣州区6街道居住建筑尺度人口空间分布图

Fig.3 Population distribution at the residential building scale for six residential committees in Xuanzhou District in 2015

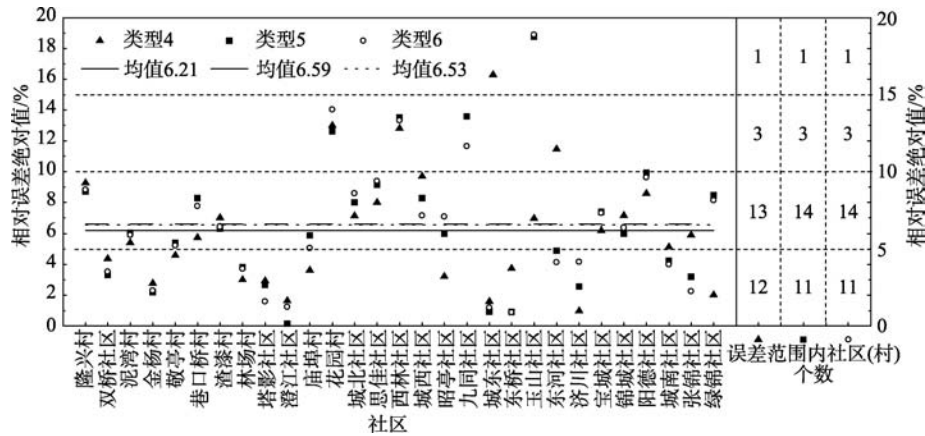


图4 社区(村)尺度的估算人数相对误差统计图

Fig.4 Relative error statistics of estimated number of people at the communities (villages) scale

个斑块应该居住的人数,作为人数参考真值。人数低估区,以每套房屋2人乘以套数作为人数下限参考真值;人数高估区,以每套房屋3人乘以套数作为人数上限参考真值。依据式(2)对类型4、类型5和类型6估算的人数与参考真值进行对比分析。

$$E_{ij} = \begin{cases} P'_{ij}/2 \times T_{ij} - 1, & P'_{ij} < 2 \times T_{ij} & \text{低估区} \\ 0, & 2 \times T_{ij} \leq P'_{ij} \leq 3 \times T_{ij} & \text{合理区} \\ P'_{ij}/3 \times T_{ij} - 1, & P'_{ij} > 3 \times T_{ij} & \text{高估区} \end{cases} \quad (2)$$

式中: E_{ij} 、 P'_{ij} 、 T_{ij} 分别为街道 i 斑块 j 的相对误差、估算人数和房屋套数; $2 \times T_{ij}$ 和 $3 \times T_{ij}$ 分别为街道 i 斑块 j 内应该居住的人数下限和人数上限值。

从图5可以看出:类型4、类型5和类型6的居住建筑斑块估算人数在合理区内的斑块个数分别为824、816、823,所占比例分别高达74.77%、74.04%、74.68%;轻微低估区(-10%, 0)和轻微高估区(0, 10%)的斑块个数比例之和分别为9.98%、9.43%、9.16%;其他误差区间内的斑块个数占比很低,并呈现随着误差变大斑块个数减少的趋势。表明类型4、类型5和类型6的精度较高,类型4具有微弱的精度优势。

4.2 误差分析与讨论

不考虑由模型产生的随机误差,分析造成人数估算误差的原因,发现估算误差主要来源于以下4个方面:一是由模型求得的人均居住面积偏大(偏小)引起的误差;二是居住建筑矢量化误差引起的误差;三是由斑块内建筑面积比重偏高(偏低)引起的误差;四是由以上3个因素两两或综合作用引起

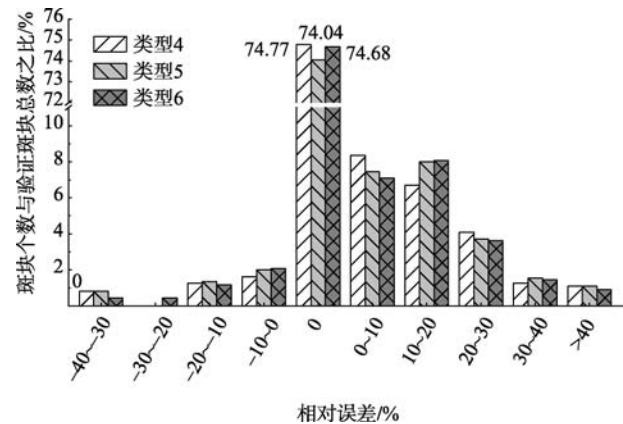


图5 居住建筑尺度的估算人数相对误差统计图

Fig.5 Relative error statistics of estimated number of people at the residential building scale

的误差。

4.2.1 总体分析

鉴于类型4、类型5和类型6的精度相似,以类型4结果为例进行误差分析。从高估区和低估区的278个斑块中,随机抽取100个居住建筑斑块,对其人均居住面积的参考真值(建筑面积/人数参考真值)与计算值、建筑面积的真实值与计算值进行对比分析(表6)。

表6表明,人均居住面积计算值偏大、矢量化提取的建筑面积偏大的斑块个数占有所有斑块的3/4左右,并且各自的相对误差平均值较高,这2类原因会导致人数高估;人数低估的斑块个数相对较少,占1/4左右。图5表明整个研究区内大部分斑块的人数估算偏高,只有少数斑块的人数估算偏低,这与表6的数据分析结果一致。

4.2.2 分项原因说明

(1) 人均居住面积计算值引起的误差

以图6①中A栋楼为例。由式(1)可得其人均居住面积为32.4 m²/人,较参考真值(37.7 m²/人)偏小5.3 m²/人,导致人数高估19.7人。矢量化提取的居住建筑斑块,其建筑面积为4405.4 m²,较真实值(4522.2 m²)偏小116.8 m²,导致人数低估3.6人。综上,人均居住面积偏小和矢量化错误共同作用导致人数高估16人,人均居住面积偏小是主因。人均居住面积计算值偏大则会导致人数低估(图6⑤中E栋楼)。

(2) 居住建筑矢量化误差引起的误差

以图6②中B栋楼为例。矢量化提取的居住建筑斑块,其建筑面积为1716.1 m²,较真实值(1500.1 m²)大216 m²,导致人数高估5人。由式(1)可得其人均居住面积为41.7 m²/人,与参考真值(41.7 m²/人)相同,可知人均居住面积没有引起误差。即矢量化误差是主因,导致人数高估5人。矢量化提取的建筑面积偏小,则会导致人数低估(图6⑤中F栋楼)。

(3) 斑块内建筑面积比重偏高引起的误差

以图6③中C斑块为例。矢量化提取的居住建筑斑块,其斑块面积为8383.7 m²,因斑块内建筑面

表6 人均居住面积、建筑面积数值对比

Tab.6 Numerical comparison of per capita living space and building area

类别	计算值减真实值	斑块个数比重/%	相对误差平均值/%	描述
人均居住面积	<0	25.9	-5.8	人均居住面积计算值偏小
	>0	74.1	15.5	人均居住面积计算值偏大
建筑面积	<0	22.2	-8.2	矢量化提取的建筑面积偏小
	>0	77.8	13.1	矢量化提取的建筑面积偏大

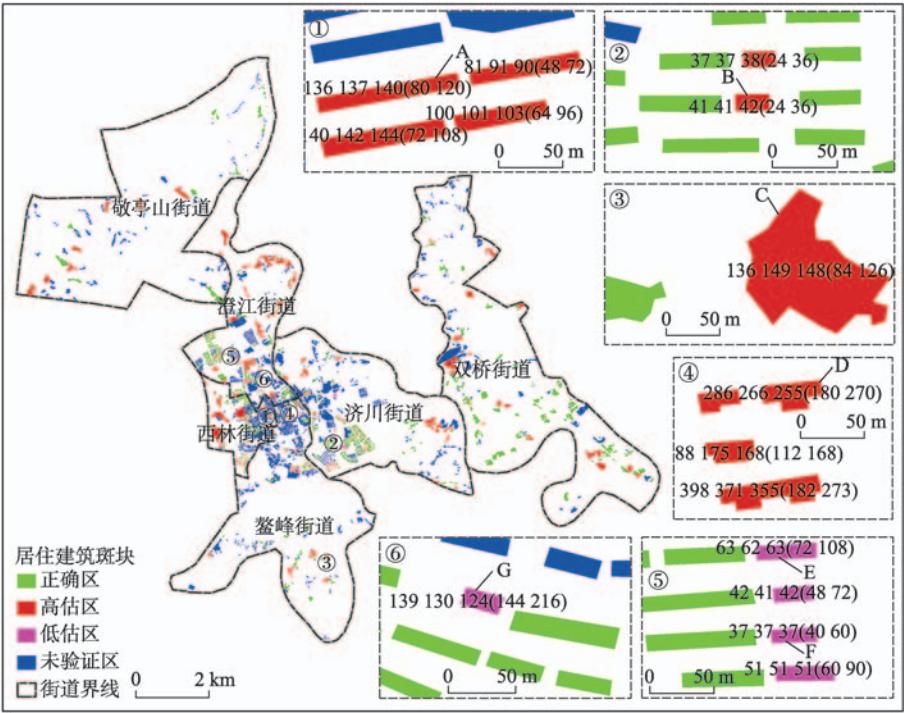


图6 估算人数相对误差空间分布图^①

Fig.6 Spatial distribution of relative error of estimated number of people

①图6数值序列含义为:类型4估算人数、类型5估算人数、类型6估算人数(人数下限参考真值、人数上限参考真值)。例如,图6⑥中G栋楼的数值序列为:139、130、124(144、216),即类型4估算人数为139人、类型5估算人数为130人、类型6估算人数为124人(人数下限参考真值为144人、人数上限参考真值为216人)。

积比重赋值为0.55,所以斑块内居住建筑面积为4611 m²,较真实值(2978.5 m²)偏大1632.5 m²,导致人数高估。

(4) 人均居住面积计算值偏大与矢量化误差共同作用引起的误差

以图6④中D栋楼为例。由式(1)可得其人均居住面积为33.9 m²/人,较参考真值(31.2 m²/人)偏大2.7 m²/人,导致人数低估21.6人。矢量化提取的居住建筑斑块,其建筑面积为9685.0 m²,较真实值(8421.3 m²)增加1263.7 m²,导致人数高估37.2人。综上,人均居住面积偏大和矢量化误差共同作用导致人数高估16人。两者共同作用也会导致人数低估(图6⑥中G栋楼)。

4.2.3 类型对比分析

图4-5表明,类型4、类型5、类型6的精度均较高,类型4具有微弱的精度优势。与类型4相比,类型5、6引入了公摊率,该属性对估算人数具有“降高”的调节作用,但是调节到合理区的“能力”

较弱。

表7展示了类型5、类型6与类型4估算人数差值信息。对于1~3层斑块,类型5、类型6的估算人数比类型4偏高的斑块个数比重分别为18.7%、18.1%,平均高出6.8、6.5人,表明公摊率属性具有“升降”作用,如图6③中C斑块所示;对于4~6层斑块,估算人数彼此差别不大,如图6②中B斑块所示;对于7~18层斑块,类型5、类型6的估算人数比类型4偏低的斑块个数比重分别为11.6%、11.6%,平均降低13.1、21.9人;对于19~33层,这一差距更大,表明公摊率属性具有“降高”作用,可以适当避免高层居住建筑估算人数过于偏高,如图6④中D斑块所示。

表8汇总了各类型估算人数在合理区内的斑块个数比重。对于1~3层斑块,与类型4相比,类型5、类型6反而将估算人数“调出”合理区,合理区内斑块个数比重分别降低2.0%;对于7~18层斑块,与类型4相比,类型5、类型6将估算人数“调入”合理区

表7 类型5、类型6与类型4估算人数差值信息汇总
Tab.7 Summary of the differences of estimated number of people between type 5, type 6, and type 4 models

层数/层	类型5与类型4			类型6与类型4		
	估算人数之差	人数之差平均值	斑块个数比重/%	估算人数之差	人数之差平均值	斑块个数比重/%
1~3	>0	6.8	18.7	>0	6.5	18.1
	=0	0.0	6.4	=0	0.0	7.2
	<0	-1.0	0.1	<0	—	0.0
4~6	>0	1.1	8.2	>0	1.8	40.9
	=0	0.0	32.3	=0	0.0	20.5
	<0	-1.1	21.2	<0	-1.0	0.3
7~18	>0	—	0.0	>0	—	0.0
	=0	—	0.0	=0	—	0.0
	<0	-13.1	11.6	<0	-21.9	11.6
19~33	>0	—	0.0	>0	—	0.0
	=0	—	0.0	=0	—	0.0
	<0	-44.1	1.5	<0	-54.0	1.5

表8 合理区内斑块个数比重信息汇总
Tab.8 Summary of the proportion of patches in reasonable range

层数/层	合理区内斑块个数比重/%		
	类型4	类型5	类型6
1~3	18.9	16.9	16.9
4~6	49.5	49.3	49.4
7~18	6.3	6.8	7.3
19~33	0.2	1.1	1.2

的斑块个数比重分别为0.5%、1.0%;对于19~33层斑块,类型5、类型6将估算人数“调入”合理区的斑块个数比重分别为0.9%、1.0%。综合来看,公摊率属性的“降高”作用,并没有提高合理区内斑块个数比重,即调节到合理区的“能力”较弱,这也是公摊率属性不能提高模型整体估算精度的原因。然而作为居住空间属性之一的公摊率,在避免高层居住建筑估算人数过于偏高等方面具有应用潜力,值得深入探讨。

5 结论与讨论

5.1 结论

本文选用 Quick Bird 遥感影像,依据解译标志,参照百度街景地图、搜狗三维地图,提取居住建筑的轮廓斑块。以街道为人口数据空间化转换尺度,基于居住空间属性,建立线性回归模型,将常住人口展布在居住建筑斑块上,获取了2015年宣州区居住建筑物尺度上的人口分布矢量数据。主要结论如下:

(1) 提出了一种以居住建筑斑块面积、斑块内建筑面积比重、建筑物层数、公摊率等居住空间属性来实现城区建筑物尺度人口空间化的方案。精度分析表明,利用该方案生成的人口空间数据精度较高,结果可信。29个社区(村)相对误差绝对值的平均值低于7%;在1102个居住建筑斑块中,估算人数在合理区内的斑块个数占比高于74%,轻微低估区(-10%, 0)和轻微高估区(0, 10%)的斑块总数占比高于9%。

(2) 由斑块面积和建筑物层数共同表征的建筑物体积,是建筑物尺度上影响人口空间分布的关键因素。在此基础上,引入斑块内建筑面积比重属性,能提高模型估算精度。作为居住空间属性之一的公摊率,具有“降高升低”的作用,但将估算人数调节到合理区的“能力”较弱,尽管如此,其应用潜力值得深入探讨。

5.2 讨论

本文的不足之处是,在建模时假定同一街道内各个居住建筑斑块的人均居住空间相同,导致模型求得的人均居住面积有偏差,进而影响模型精度。在后续的研究中,将重点细分同一街道内居住建筑类别,揭示不同类型建筑的人均居住面积的异同,以期进一步提高模型精度。精细尺度上,城市公共设施要素数据(何莲娜, 2011; 康停军等, 2012)、带有地理位置的 tweets 数据(Patel et al, 2016)等也是城市人口空间分布的指示因素,下一步将引入此类地理信息数据,以期提高城区建筑物尺度人口数据空间化精度。

参考文献(References)

柏中强, 王卷乐, 姜浩, 等. 2015. 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, 17(6): 653-660. [Bai Z Q, Wang J L, Jiang H, et al. 2015. The gridding approach to redistribute population based on multi-source data[J]. Journal of Geo-Information Science, 17(6): 653-660.]

柏中强, 王卷乐, 杨飞. 2013. 人口数据空间化研究综述[J]. 地理科学进展, 32(11): 1692-1702. [Bai Z Q, Wang J L, Yang F. 2013. Research progress in spatialization of population data[J]. Progress in Geography, 32(11): 1692-1702.]

陈振拓, 李志强, 丁文秀, 等. 2012. 面向防震减灾的人口数据空间化研究: 以2007年宁洱地震灾区为例[J]. 震灾防御技术, 7(3): 273-284. [Chen Z T, Li Z Q, Ding W X, et al. 2012. Study of spatial population distribution in earthquake disaster reduction: A case study of 2007 Ning'er earthquake[J]. Technology for Earthquake Disaster Prevention, 7(3): 273-284.]

杜国明, 张树文. 2007. 面向防洪救灾的人口统计数据空间化研究: 以扶余县为例[J]. 长江流域资源与环境, 16(2): 265-268. [Du G M, Zhang S W. 2007. Spatial distribution in population statistics in relation to flood prevention and rescue: A case study in Fuyu County[J]. Resources and Environment in the Yangtze Basin, 16(2): 265-268.]

杜国明, 张树文, 张有全. 2007. 城市人口密度的尺度效应分析: 以沈阳市为例[J]. 中国科学院研究生院学报, 24(2): 186-192. [Du G M, Zhang S W, Zhang Y Q. 2007. Analyzing scale effects of population density with Shenyang City as a case[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 24(2): 186-192.]

符海月, 李满春, 赵军, 等. 2006. 人口数据格网化模型研究进展综述[J]. 人文地理, 21(3): 115-119, 114. [Fu H Y, Li M C, Zhao J, et al. 2006. Summary of grid transformation models of population data[J]. Human Geography, 21(3): 115-119, 114.]

谷国梁, 王晓蕾, 李雅静, 等. 2016. 天津市面向震害快速评估的房屋和人口空间化研究[J]. 地震, 36(2): 149-158. [Gu G L, Wang X L, Li Y J, et al. 2016. Spatialization of population and housing data in Tianjin oriented to rapid earthquake loss assessment[J]. Earthquake, 36(2): 149-158.]

何莲娜. 2011. 基于城市布局要素的人口数据空间化研究[J]. 测绘科学, 36(1): 38-41. [He L N. 2011. A discussion of population spatial distribution based on urban layout parameter[J]. Science of Surveying and Mapping, 36(1): 38-41.]

胡长菊. 2015. 国家建设部规定高层住宅楼公摊系数是多少[EB/OL]. 华律网, 2015-06-07[2016-09-21]. <http://www.66law.cn/question/5226524.aspx>. [Hu C J. 2015. Guojia jianshebu guiding gaoceng zhuzhailou gongtan xishu shi duoshao[EB/OL]. Hualv Network, 2015-06-07[2016-09-21]. <http://www.66law.cn/question/5226524.aspx>.]

胡焕庸. 1936. 句容县之人口分布[J]. 地理学报, (3): 621-627, 14. [Hu H Y. 1936. The distribution of population in Chuyong District[J]. Acta Geographica Sinica, (3): 621-627, 14.]

胡焕庸. 1983. 论中国人口之分布[M]. 上海: 华东师范大学出版社. [Hu H Y. 1983. Lun Zhongguo renkou zhi fenbu[M]. Shanghai, China: East China Normal University Press.]

江东, 杨小唤, 王乃斌, 等. 2002. 基于RS、GIS的人口空间分布研究[J]. 地球科学进展, 17(5): 734-738. [Jiang D, Yang

- X H, Wang N B, et al. 2002. Study on spatial distribution of population based on remote sensing and GIS[J]. *Advances in Earth Sciences*, 17(5): 734-738.]
- 康停军, 张新长, 赵元, 等. 2012. 基于多智能体的城市人口分布模型[J]. *地理科学*, 32(7): 790-797. [Kang T J, Zhang X C, Zhao Y, et al. 2012. Agent-based urban population distribution model[J]. *Scientia Geographica Sinica*, 32(7): 790-797.]
- 李月娇, 杨小唤, 王静. 2014. 基于景观生态学的人口空间数据适宜格网尺度研究: 以山东省为例[J]. *地理与地理信息科学*, 30(1): 97-100. [Li Y J, Yang X H, Wang J. 2014. Grid size suitability of population spatial distribution in Shandong Province based on landscape ecology[J]. *Geography and Geo-Information Science*, 30(1): 97-100.]
- 廖顺宝, 孙九林. 2003. 基于GIS的青藏高原人口统计数据空间化[J]. *地理学报*, 58(1): 25-33. [Liao S B, Sun J L. 2003. GIS based spatialization of population census data in Qinghai-Tibet Plateau[J]. *Acta Geographica Sinica*, 58(1): 25-33.]
- 林丽洁, 林广发, 颜小霞, 等. 2010. 人口统计数据空间化模型综述[J]. *亚热带资源与环境学报*, 5(4): 10-16. [Lin L J, Lin G F, Yan X X, et al. 2010. Spatialization models of census data: A review[J]. *Journal of Subtropical Resources and Environment*, 5(4): 10-16.]
- 刘焕金. 2012. 基于多源数据的太湖流域人口空间化研究[D]. 南京: 南京农业大学. [Liu H J. 2012. Research on the population specialization basic on mutil- source data in Lake-Tai Basin[D]. Nanjing, China: Nanjing Agricultural University.]
- 王磊, 蔡运龙. 2011. 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*, 30(5): 635-640. [Wang L, Cai Y L. 2011. Spatial down-scaling analysis and simulation of population density in Maotiaohe Basin, Guizhou Province[J]. *Progress in Geography*, 30(5): 635-640.]
- 王立新. 2014. 专家给出银川住宅楼公摊率参考[N/OL]. *银川晚报*, 2014-08-26[2016-9-21]. <http://finance.chinanews.com/house/2014/08-26/6532614.shtml>. [Wang L X. 2014. Zhuanjia geichu Yinchuan zhuzhailou gongtanlv cankao[N/OL]. *Yinchuan Evening News*, 2014-08-26[2016-09-21]. <http://finance.chinanews.com/house/2014/08-26/6532614.shtml>.]
- 王培震, 石培基, 魏伟, 等. 2012. 基于空间自相关特征的人口密度格网尺度效应与空间化研究: 以石羊河流域为例[J]. *地球科学进展*, 27(12): 1363-1372. [Wang P Z, Shi P J, Wei W, et al. 2012. Grid scale effect and spatialization of population density based on the characteristic of spatial autocorrelation in Shiyang River Basin[J]. *Advances in Earth Science*, 27(12): 1363-1372.]
- 杨小唤, 江东, 王乃斌, 等. 2002. 人口数据空间化的处理方法[J]. *地理学报*, 57(S1): 70-75. [Yang X H, Jiang D, Wang N B, et al. 2002. Method of pixelizing population data[J]. *Acta Geographica Sinica*, 57(S1): 70-75.]
- 杨小唤, 刘业森, 江东, 等. 2006. 一种改进人口数据空间化的方法: 农村居住地重分类[J]. *地理科学进展*, 25(3): 62-69. [Yang X H, Liu Y S, Jiang D, et al. 2006. An enhanced method for spatial distributing census data: Reclassifying of rural residential[J]. *Progress in Geography*, 25(3): 62-69.]
- 叶靖, 杨小唤, 江东. 2010. 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]. *地球信息科学学报*, 12(1): 40-47. [Ye J, Yang X H, Jiang D. 2010. The grid scale effect analysis on town leveled population statistical data spatialization[J]. *Journal of Geo-Information Science*, 12(1): 40-47.]
- 有哈相伴. 2009. 套内面积售房对买房的影响[EB/OL]. 新浪博客, 2009-08-28[2016-09-21]. http://blog.sina.com.cn/s/blog_4e45ef760100e84y.html. [Youhaxiangban. 2009. Taonei mianji shoufang dui maifang de yingxiang[EB/OL]. *Sina Blog*, 2009-08-28[2016-09-21]. http://blog.sina.com.cn/s/blog_4e45ef760100e84y.html.]
- 卓莉, 陈晋, 史培军, 等. 2005. 基于夜间灯光数据的中国人口密度模拟[J]. *地理学报*, 60(2): 266-276. [Zhuo L, Chen J, Shi P J, et al. 2005. Modeling population density of China in 1998 based on DMSP/OLS nighttime light image[J]. *Acta Geographica Sinica*, 60(2): 266-276.]
- 卓莉, 黄信锐, 陶海燕, 等. 2014. 基于多智能体模型与建筑物信息的高空间分辨率人口分布模拟[J]. *地理研究*, 33(3): 520-531. [Zhou L, Huang X R, Tao H Y, et al. 2014. High spatial resolution population distribution simulation based on building information and multi-agent[J]. *Geographical Research*, 33(3): 520-531.]
- Bhaduri B, Brigh E, Coleman P, et al. 2007. LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics[J]. *GeoJournal*, 69(1-2): 103-117.
- Dmowska A, Stepinski T F. 2014. High resolution dasymetric model of U.S demographics with application to spatial distribution of racial diversity[J]. *Applied Geography*, 53: 417-426.
- Dong P L, Ramesh S, Nepalia A. 2010. Evaluation of small-area population estimation using LiDAR, Landsat TM and parcel data[J]. *International Journal of Remote Sensing*, 31(21): 5571-5586.
- Galleo F J. 2010. A population density grid of the European Union[J]. *Population and Environment*, 31(6): 460-473.
- Gaughan A E, Stevens F R, Huang Z J, et al. 2016. Spatiotemporal patterns of population in mainland China, 1990 to 2010[J]. *Scientific Data*, 3: doi: 10.1038/sdata.2016.5.
- Jia P, Gaughan A E. 2016. Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida[J]. *Applied Geography*, 66: 100-108.
- Lung T, Lübker T, Ngochoch J K, et al. 2013. Human population distribution modelling at regional level using very high resolution satellite imagery[J]. *Applied Geography*, 41: 36-45.
- Lwin K, Murayama Y. 2009. A GIS approach to estimation of

- building population for micro-spatial analysis[J]. *Transactions in GIS*, 13(4): 401-414.
- Mennis J. 2003. Generating surface models of population using dasymetric mapping[J]. *The Professional Geographer*, 55(1): 31-42.
- Patel N N, Stevens F R, Huang Z J, et al. 2016. Improving large area population mapping using geotweet densities[J]. *Transactions in GIS*, doi: 10.1111/tgis.12214.
- Silván-Cárdenas J L, Wang L, Rogerson P, et al. 2010. Assessing fine-spatial-resolution remote sensing for small-area population estimation[J]. *International Journal of Remote Sensing*, 31(21): 5605-5634.
- Stevens F R, Gaughan A E, Linard C, et al. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data[J]. *PLoS One*, 10(2): doi: 10.1371/journal.pone.0107042.
- Ural S, Hussain E, Shan J. 2011. Building population mapping with aerial imagery and GIS data[J]. *International Journal of Applied Earth Observation and Geoinformation*, 13(6): 841-852.
- Zhang Y, Dong C, Liu J P, et al. 2015. Gridded population distribution map for the Hebei Province of China[J]. *Environmental Engineering and Management Journal*, 14(3): 673-680.

A method for demographic data spatialization based on residential space attributes

DONG Nan^{1,2}, YANG Xiaohuan^{1,2*}, CAI Hongyan¹

(1. State Key Laboratory of Resources and Environmental 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: Fine spatial scale population distribution has increasingly become a research hotspot yet a difficult question in the field of population geography. It has practical values in application and scientific significance for relevant research, such as disaster risk and impact assessment, resource allocation, and construction of smart cities. Residential building scale is considered an important part of fine spatial scales for population distribution. Research on the spatialization of population data at this scale has increasingly attracted academic attention. In this study, a population distribution vector data set at the residential building scale was established for six residential committees in Xuanzhou District, Anhui Province in 2015 based on residential space attribute data. Data used in the study include residential building patch area, percentage of housing area within residential building patches, building floor number, and public area rate. The method takes residential space attributes as variables for spatializing population data and treats residential building patches as population distribution location in geographical space with town boundary and town-level demographic data as controls. The spatialization method used in this study reveals detailed information about population distribution in urban areas. The results show that: (1) The population distribution data, obtained by using residential space attributes, are proved to be of high accuracy and reliability. The mean absolute relative error for 29 communities (villages) is less than 7%. The absolute relative error of 25 out of 29 communities (villages) is less than 10%. The proportion of patches whose estimated number of people is in reasonable range is higher than 74% in a total of 1102 residential building patches. The proportion of patches whose relative error is in slightly underestimated area (-10%, 0) and overestimated area (0, 10%) is higher than 9%. (2) Building volume, defined by residential building patch area and building floor number, is a key factor to estimate accurately the number of people within a residential building. The percentage of housing area can further improve model accuracy. Public area rate plays an important role to increase estimated number of people in underestimated area and decrease that in overestimated area, but is too weak to adjust the estimated number of people to reasonable range. In conclusion, spatialization based on residential space attributes can be an important method for population spatialization research at the residential building scale.

Key words: population; spatialization; residential space; residential building; patch; Xuanzhou District of Xuancheng City