

何艳虎, 龚镇杰, 林凯荣. 基于地理大数据和多源信息融合的区域未来人口精细化空间分布模拟研究——以珠江三角洲为例 [J]. 地理科学, 2022, 42(3): 426-435. [He Yanhu, Gong Zhenjie, Lin Kairong. Simulation of fine spatial distribution of regional future population based on geographical big data and multisource fused method: A case of the Pearl River Delta. Scientia Geographica Sinica, 2022, 42(3): 426-435.] doi: 10.13249/j.cnki.sgs.2022.03.007

基于地理大数据和多源信息融合的区域未来人口精细化空间分布模拟研究 ——以珠江三角洲为例

何艳虎^{1,2}, 龚镇杰¹, 林凯荣³

(1. 广东工业大学生态环境与资源学院, 广东 广州 510006; 2. 广东省流域水环境治理与水生态修复重点实验室, 广东 广州 510006; 3. 中山大学土木与工程学院, 广东 珠海 519082)

摘要: 首先采用队列因素法和 CA-Markov 模型对区域未来人口规模和土地利用格局进行模拟预测, 并结合 POI 地理大数据, 利用多源信息融合法构建区域未来人口精细化空间分布模拟模型, 以珠江三角洲城市群 2030 年各区县精细化的人口空间分布预测进行实证分析。结果表明: ① 采用队列因素法进行珠江三角洲各区县人口规模预测的相对误差大部分在 5% 以下, 基于 CA-Markov 模型土地利用模拟的 Kappa 系数达到 0.97; ② 珠江三角洲城市群精细化的人口空间分布模拟数据与实际人口数据的拟合趋势线 R^2 达到了 0.90, 模拟效果优于 Worldpop 数据集, 体现了 POI 地理大数据与多源信息融合在精细化人口空间分布模拟上的优势; ③ 珠江三角洲未来人口呈现由中心向外围扩散和递减的空间分布格局, 空间差异显著且较为稳定, 70% 的人口集中在广州、深圳、东莞和佛山等核心城市。

关键词: 未来人口; 精细化空间分布; CA-Markov 模型; 多源信息; 珠江三角洲

中图分类号: K901.3; P208; C922 **文献标识码:** A **文章编号:** 1000-0690(2022)03-0426-10

中国作为人口大国, 人口已成为经济社会发展不可忽视的关键问题^[1]。庞大的人口数量一方面为经济社会发展提供了充足的劳动力和强劲的消费力, 另一方面也给生态环境保护、能源开发利用与区域规划发展带来了新的挑战^[2]。为解决好人口快速增长与资源可持续利用问题, 中国先后进行了 7 次全国人口普查。人口普查作为了解人口数量与人口分布的重要信息来源, 通常是以行政单元平均人口密度表征人口的空间分布信息, 但存在空间分辨率低和更新周期长等缺陷^[3]。中国人口普查每 10 a 进行一次, 较长的人口更新周期不利于探索人口空间分布规律和厘清人口、生态环境、资源开发利用与经济社会发展之间的相互作用机制^[4]。准确的人口规模预测与精细化的未来人口空间分布能为区域发展规划和生态环境

保护战略的制定提供科学依据与参考^[5]。

区域未来人口空间分布模拟包含两部分内容: 未来人口规模预测和人口空间分布模拟。当前人口规模预测方法主要分为两大类: 数学统计方法和人口学方法。其中常用的数学统计模型包括: Logistic 模型^[6]、GM(1, 1)模型^[7]、ARIMA 模型^[8]等, 其主要应用在单一人口总量规模预测上, 并没有充分考虑其他人口因素对未来人口的影响。随着人口统计工作日益受到重视, 相关统计数据如各年龄段人口数、各年龄段死亡数、各育龄段妇女人口数等^[9]日趋丰富, 为人口学方法的运用提供了数据支撑。目前, 运用较为广泛的基于人口学理论的人口预测方法是队列因素法^[10], 该方法充分利用人口统计数据中的各年龄段信息(横向信息)并结合人口变动因素(出生率、死亡率和迁移率)

收稿日期: 2020-11-20; **修订日期:** 2021-05-12

基金项目: 国家自然科学基金项目(51979043、51509127)资助。[Foundation: National Natural Science Foundation of China (51979043, 51509127).]

作者简介: 何艳虎(1985-), 男, 河南息县人, 博士, 副教授, 主要从事水资源配置与管理研究。E-mail: heyanyhu456@gdut.edu.cn

对未来人口总量、人口结构进行预测,其结果不再局限于单一人口总量的预测,更加符合实际人口预测需求。近年来,随着中国经济社会发展进入新常态,在进行“十四五”规划或更远的中长期规划中,除了掌握未来人口规模和结构,还应全面了解人口的精细化空间分布状况,为国家和地区相关规划和政策制定提供更加科学的参考。

人口数据空间化是了解人口空间分布的关键,它主要基于人口空间分布指示性因子构建人口空间分布模拟模型,进而对人口统计数据进行离散化处理,实现人口统计数据由行政单元向更小尺度的地理栅格转换^[11]。在早期人口空间分布模拟研究中,众多学者把土地利用类型作为影响人口空间分布的指示性因子,在遵循同一土地利用类型人口密度相同的假设前提下,构建人口密度与相应土地利用类型所占面积的多元回归关系,形成了土地利用类型法^[12]。该方法原理简单,已在中国 1995 年、2000 年和 2003 年公里格网人口数据集中得到应用^[11],但未能充分考虑其他因素对人口空间分布的指示性作用。实际上,除了土地利用类型,道路、坡度和夜间灯光遥感影像等因子在人口空间化过程中同样起着指示性作用^[13]。随着互联网技术的不断发展,地理大数据得到了长足发展,为实现精细化的人口空间分布预测提供了新思路,其中最具代表性的是 Point of Interest (POI)数据,包含了地理实体的名称、经纬度和地址等信息,承载着大量的人类活动信息,克服了传统数据难以反映人类活动与地理实体空间关系的问题,是当前人口数据空间化的研究热点^[14,15]。多源信息融合法^[16]通过结合土地利用类型、POI 数据和夜间灯光数据等人口空间分布指示性因子进行人口数据空间化建模,相较于土地利用类型法,更能真实地反映人口空间分布格局,而人口空间分布指示性因子权重的确定则较为复杂,在现有研究中主要分为两种方法:多元线性回归方法^[16]和机器学习^[17]。其中多元线性回归方法是将人口与人口空间分布指示性因子之间的关系假设为线性关系,但是两者间的关系并非简单的线性关系,更多体现为复杂的非线性关系^[18],而机器学习,如随机森林模型通过给出各指示性因子的权重值,能够较好地模拟人口与人口空间分布指示性因子间的复杂非线性关系,近年来受到广泛的应用^[19]。

目前区域人口空间分布模拟研究在时间维度上主要立足于过去和现在对人口分布结构及机理进行探讨^[20,21],在空间尺度上主要集中在国家和省级层面对人口进行空间精细化研究^[22,23]。针对未来人口空间分布的研究也仅以区县人口规模定量分析其空间的基本格局^[24],关于区域未来人口精细化空间分布模拟研究仍不足,尤其缺少区县尺度的未来人口精细化空间分布模拟预测研究。粤港澳大湾区是中国开放程度最高、经济活力最强的区域之一,在国家发展大局中具有重要战略地位^[25],珠江三角洲城市群是粤港澳大湾区的重要组成部分,包含了广州、深圳等湾区核心城市,开展其未来人口精细化空间分布的研究对湾区交通、能源、信息、水利等基础设施的规划布局及各类资源要素的高效配置具有重要指示作用。

基于上述分析,本文以珠江三角洲城市群为例,采用队列因素法和 CA-Markov 模型模拟预测珠江三角洲各区县 2030 年的人口规模和土地利用类型,基于 POI 大数据和随机森林模型,结合多源信息融合法构建并验证精细化的人口空间分布预测模型,最后模拟预测珠江三角洲各区县 2030 年精细化的人口空间分布格局。通过以上探索性研究,建立区域未来人口精细化空间分布预测模型,为地区公共基础设施布局、资源优化配置和生态环境保护提供科学依据。

1 研究区与数据来源

1.1 研究区概况

珠江三角洲城市群由广州、深圳、珠海、佛山、惠州、东莞、中山、江门、肇庆 9 市组成,土地面积为 $5.6 \times 10^4 \text{ km}^2$,位于广东省中南部,毗邻港澳,被称为中国的“南大门”。根据 2019 年《广东省统计年鉴》^[26],珠江三角洲城市群 GDP 总量为 8.69 万亿元,约占广东省 GDP 总量的 80%,常住人口为 6 446.89 万人,城镇化率高达 86.28%。然而,珠江三角洲城市间经济社会发展水平差异显著,人口分布存在空间异质性,对区域发展规划和资源开发利用带来了挑战。本研究以珠江三角洲城市群 9 个地级市的 47 个区县(按 2010 行政区划)为研究对象进行人口精细化空间分布研究(图 1)。

1.2 数据来源

研究数据主要包括:① 人口统计数据,2010 年、2015 年和 2018 年珠江三角洲城市群人

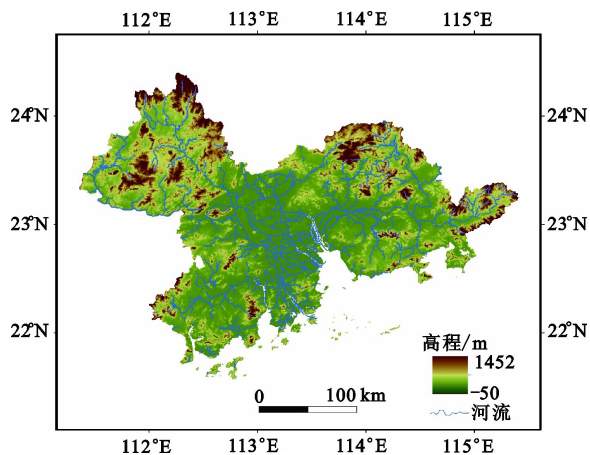


图1 研究区域位置

Fig.1 Location of the study area

口统计数据,源于《全国第六次人口普查数据》^[27]、《广东省统计年鉴》^[26]和各地市统计年鉴;② 区县级行政边界数据,1:100万广东省县级行政边界数据,源于国家地理信息公共服务平台;③ 土地利用类型数据,2005年、2010年和2015年全国分辨率为1 km×1 km土地利用类型数据,分为6大类25个小类,源于中国科学院资源环境科学数据中心(<https://www.resdc.cn>);④ DEM数据,广东省STRM 90 m分辨率DEM数据,源于中国科学院资源环境科学数据中心;⑤ 河流、道路数据,1:100万广东省河流、道路数据,源于国家地理信息公共服务平台(<https://www.tianditu.gov.cn/>);⑥ POI数据,九市各区县POI数据,分为7类,源于百度地图API(<http://map.baidu.com>)。人口分布指示因子权重和来源见表1。

表1 人口分布指示性因子

Table 1 Indicator of population distribution			
POIs名称	变量	权重	来源
超市	W_1	0.183	百度地图API
公园	W_2	0.108	百度地图API
酒店	W_3	0.046	百度地图API
警察局	W_4	0.051	百度地图API
银行	W_5	0.075	百度地图API
学校	W_6	0.054	百度地图API
医院	W_7	0.250	百度地图API
DEM	W_8	0.046	中国科学院资源环境科学数据中心
道路	W_9	0.169	国家地理信息公共服务平台
河流	W_{10}	0.018	国家地理信息公共服务平台

1.3 数据预处理

数据预处理主要包括以下:① 人口统计数据:根据2010年《全国第六次人口普查数据》的各年龄人口占比,计算珠江三角洲9市各区县各年龄人口数。② 县级行政边界数据:在1:100万广东省县级行政边界矢量图层上,提取珠江三角洲九市各区县边界,同时以2010年行政区划图进行校准。③ 土地利用类型数据:采用Transverse Mercator投影方式进行空间投影,对25个土地利用类型进行重分为6类。④ DEM数据:采用Transverse Mercator投影方式进行空间投影,栅格大小采用最邻近法重采样为1 km×1 km。⑤ 河流、道路数据:在1:100万广东省河流、道路数据矢量图层上进行裁剪得到珠江三角洲9市各区县河流、道路数据,通过欧式距离工具由矢量图层得到最近河流、道路的欧式距离栅格图层,栅格大小采用最邻近法重采样为1 km×1 km。⑥ POI数据:通过核密度分析工具,将7类POI矢量点图层转换成7个密度栅格图层,栅格大小采用最邻近法重采样为1 km×1 km。

2 研究方法

2.1 队列因素法

采用队列因素法对珠江三角洲城市群2030年的人口规模进行预测,该方法简单易懂,难点在于人口变动参数的合理设置,如各年龄段人口的存活率、育龄妇女各年龄段的生育率和人口迁移率等。其中生育率是关键参数,受国家政策影响较大。在本研究中,育龄妇女年龄范围为15~49岁^[28],考虑到100岁及以上人口数量对总人口规模影响小,在计算过程中未考虑100岁及以上人口数量。

队列因素法数学公式如下^[28]:

$$P_{t+1} = P_t - P_{d,t} + P_{b,t+1} + P_{i,t+1} \quad (1)$$

$$P_t = \sum_{y=0}^{99} P_{y,t} \quad (2)$$

$$P_{d,t} = \sum_{y=0}^{99} (P_{y,t} \times D_{y,t}) \quad (3)$$

$$P_{b,t+1} = \sum_{y=15}^{49} (F_{y,t} \times FR_{y,t}) \times (1 - D_{b,t+1}) \quad (4)$$

$$P_{t+1} = \sum_{y=0}^{99} [P_{y,t} \times (1 - D_{y,t})] + \sum_{y=15}^{49} (F_{y,t} \times FR_{y,t}) \times (1 - D_{b,t+1}) + P_{i,t+1} \quad (5)$$

式中, P_{t+1} 为 $t+1$ 年总人口; P_t 为 t 年总人口; $P_{d,t}$ 为死亡人口; $P_{b,t+1}$ 为 $t+1$ 年出生人口; $P_{i,t+1}$ 为 $t+1$ 年净迁入人口; $P_{y,t}$ 为 t 年 y 年龄人口数; $D_{y,t}$ 为 t 年 y 年龄死亡率; $F_{y,t}$ 为 t 年 y 年龄育龄妇女人口数; $FR_{y,t}$ 为 t 年 y 年龄育龄妇女生育率; $D_{b,t+1}$ 为 $t+1$ 年新生儿死亡率。

2.2 CA-Markov 模型

元胞自动机(Cellular Automata, CA)作为在时间、空间和状态上离散的动力学模型,具有模拟复杂时空演化过程的能力^[29]。土地利用类型的每个栅格称之为元胞,每个元胞对应特定土地利用类型,具有有限的状态。元胞自动机的核心是确定转换规则^[30],其规则通常可由相关政策、自然因素和经济社会因素等构成,如本研究在“耕地红线”政策上,限定耕地不变化,该动力学模型可表示为^[31]:

$$S_{t+1} = f(S_t, N) \quad (6)$$

式中, S_{t+1} 为 $t+1$ 时刻元胞的状态集合; S_t 为 t 时刻元胞的状态集合; N 为元胞的领域; f 为元胞转换规则。

Markov 模型在土地利用格局变化研究中广泛应用^[32],它是一种基于马尔科夫链过程理论描述随机过程的常用方法,该方法遵循无后效性假设条件,即随机过程 $t+1$ 时刻状态仅与 t 时刻状态有关,该过程的关键在于确定土地类型间相互转化的初始状态转移概率矩阵,从而对未来土地类型趋势做出预测,具体公式如下^[31]:

$$S_{t+1} = P_{ij} \times S_t \quad (7)$$

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (8)$$

式中, P_{ij} 为 i 土地类型转移为 j 土地类型的概率,其中 $0 \leq P_{ij} \leq 1$ 且 $\sum_{j=1}^n P_{ij} = 1 (i = 1, 2, \cdots, n)$, n 为土地类型个数。

元胞自动机与 Markov 模型各有优缺点,一方面元胞自动机具有模拟复杂系统的时空演化能力,另一方面 Markov 模型在长期预测方面具有优势,

但缺乏空间变量^[33], CA-Markov 模型通过结合两者优点,互为补短,具有模拟土地利用类型时空演化和预测土地利用类型未来趋势的能力^[34],具体操作过程参考 Aburas 等人的研究^[34]。因此,本文采用 CA-Markov 模型模拟预测珠江三角洲城市群 2030 年土地利用空间格局,为人口的精细化空间分布预测提供基础。

2.3 多源信息融合法

选择合理的人口空间分布指示性因子构建人口空间化数学模型可以较好地以行政区为载体的,实现人口的精细化空间分布^[11]。本研究采用多源信息融合法构建人口精细化空间分布模拟模型,该方法的基本流程为: ① 选择合理的人口分布指示性因子; ② 确定各人口分布指示性因子权重; ③ 依据权重值对人口分布指示性因子进行加权融合,并计算栅格单元的人口分布系数; ④ 以各区县总人口为控制条件,根据区县范围内的栅格单元人口分布系数进行人口空间化。其中人口分布指示性因子从人口分布的主要影响因素出发,综合考虑现有研究成果进行选择^[21,35],其权重的确定是多源信息融合法的关键步骤,决定了栅格单元的人口分布系数。人口与人口分布指示性因子存在复杂的非线性关系,由决策树发展而来的随机森林能较好地模拟复杂的非线性问题^[36],因此本研究采用随机森林模型进行权重值计算,具体计算原理和过程可参考 Breiman 的研究^[36]。进行权重值计算前,首先从预处理的 10 个人口分布指示性因子图层中按区县范围提取对应的栅格像元数值,为了消除各区县范围大小不一的影响,采用区县栅格的平均像元值代替区县栅格的总像元值,然后利用区县人口数与该区县 10 个人口分布指示性因子的栅格平均像元值拟合随机森林模型,最后把人口分布指示性因子输入拟合好的随机森林模型得到权重值,进而得到各栅格单元的人口分布系数。在进行人口空间化的预测过程中,依据“无居住用地则无人口”原则^[13],假定人口只分布在土地利用类型为建设用地上^[12]。具体计算公式如下:

$$P_i = P_t \times \lambda_i \quad (9)$$

$$\lambda_i = \frac{W_i}{W_t} \quad (10)$$

$$W_i = D_{i,1} \times w_1 + D_{i,2} \times w_2 + \cdots + D_{i,k} \times w_k \quad (11)$$

$$W_i = \sum_{i=1}^m W_i$$

式中, P_i 为第 i 个栅格人口数; P_i 为计算区县总人口; λ_i 为第 i 个栅格人口分布系数; W_i 为第 i 个栅格加权融合值; W_i 为计算区县内可分布人口的栅格加权融合总值; $D_{i,k}$ 为第 i 栅格 k 类人口分布因子值; w_k 为第 k 类人口分布因子权重; m 为所计算区县内可分布人口的栅格数。

3 结果分析

3.1 2030 年人口规模预测

随着“全面二胎政策”的实施,中国生育率较之前有明显的提升,同时结合广东省人口发展规划(2017—2030 年)(http://www.gd.gov.cn/gkmlpt/content/0/146/post_146685.html#7)提出的总和生育率,本研究将 2010—2015 年总和生育率设置为 1.1, 2016—2030 年总和生育率设置为 1.6, 育龄妇女年龄范围 15~49 岁(图 2a)。根据公式(1), 基于 2010 年珠江三角洲各区县人口数据进行 2030 年人口规模预测。由图 2b 可知,“全面二胎政策”实施前后,珠江三角洲各区县实际人口与预测人口的相对误差大部分在 5% 以下,相对误差较大的区县有广州的黄埔区、番禺区、南沙区和深圳的南山区和福田区,其中 2016 年比 2015 年相对误差大,主要原因可能是“全面二胎政策”对生育率

的影响存在较大的不确定性。由图 3 可看出,珠江三角洲各区县人口呈现稳定的增长趋势,与 2010 年相比,2030 年各区县常住人口增幅基本处于 13.5% 的水平,且人口集中于广州、深圳、东莞和佛山等粤港澳大湾区核心城市的格局未变。总体而言,采用队列因素法和所设人口参数来预测 2030 年珠江三角洲城市群人口规模是相对可靠的。

3.2 2010—2030 年土地利用类型空间格局变化模拟

首先利用 2005 年、2010 年两期土地利用现状图构建面积转移矩阵与适宜性图层,同时考虑到“坚守 18 亿亩耕地红线”和水域转换其他土地类型可能性小,假设耕地和水域两类土地类型在面积上保持不变,基于此,以 2010 年土地利用现状为基础模拟珠江三角洲城市群 2015 年土地利用数据,并采用珠江三角洲城市群 2015 年土地利用现状图进行检验。总体 Kappa 系数为 0.97, 大于 0.75, 两期土地利用图一致性高^[31], 表明 CA-Markov 模型模拟珠江三角洲城市群土地利用类型具有可行性。通常情况下,利用 CA-Markov 模型进行土地利用类型模拟预测时,基年与预测年的年份间隔应与建立面积转移矩阵的初始年和末期年的年份间隔相一致^[37], 即以 2015 年土地利用现状为基础模拟预测 2020 年土地利用类型数据,再以 2020 年土地利用类型数据为基础模拟预测 2025 年土地利用类型数据,依此类推得到 2030 年土地利用类型数据。由于本文假设人口只分布在土地

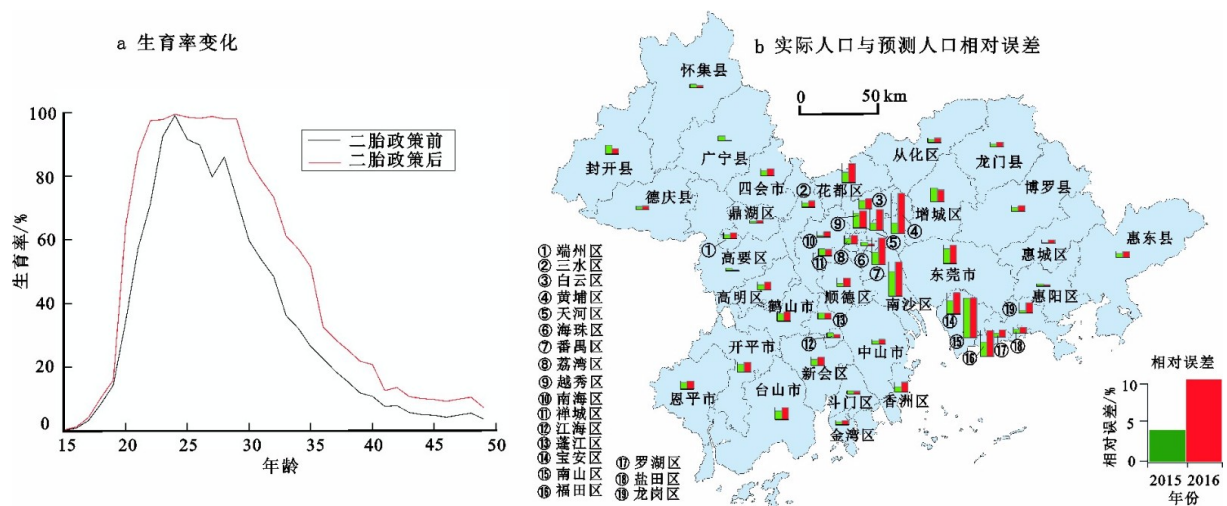


图 2 二胎政策前后珠江三角洲生育率变化及实际人口与预测人口相对误差

Fig.2 The change of fertility and the relative error between actual population and predicted population in the Pearl River Delta before and after the two-child policy

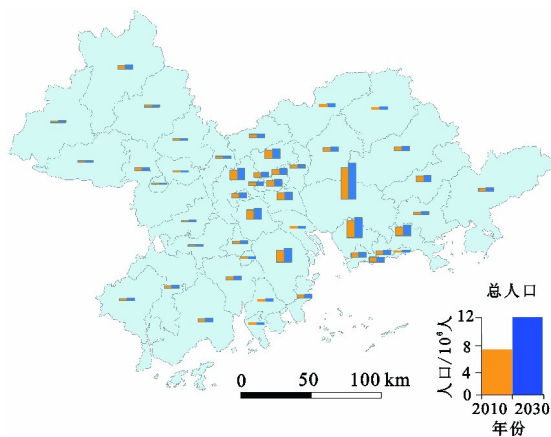


图3 珠江三角洲 2010 年实际人口与 2030 年预测人口

Fig.3 The actual population of the Pearl River Delta in 2010 and the predicted population in 2030

利用类型为建设用地上(居住区),非建设用地的土地利用类型为非居住区,通过对比 2010—2030 年居住区变化(图 4),可看出居住区面积总体呈增长趋势,主要表现为在原居住区的基础上向外围扩展,但各期居住区面积变化不大,年平均变化率为 3.6%,最大年变化率为 2010—2015 年的 6.3%,最小年变化率为 2025—2030 年的 0.3%。

3.3 2010—2030 年精细化的人口空间分布模拟

基于珠江三角洲城市群 2030 年人口数量和土地利用类型预测结果,采用多源信息融合法构建未来精细化的人口空间分布预测模型,对 2030

年珠江三角洲精细化的人口空间分布进行模拟预测。模型中人口分布系数由 10 种人口分布指示性因子确定,假设 DEM 和河流等自然因素及道路、医院、公园、银行、警务局、学校等公共基础设施未来变动幅度不大,即假定人口分布指示性因子加权融合图层不变,通过模拟预测得到的 2030 年建设用地作为掩膜,提取 2030 年建设用地相对应的人口分布指示性因子加权融合值,利用公式(9)得到 2030 年珠江三角洲精细化的人口空间分布。

3.3.1 人口空间分布模拟精度评价

采用实际人口数据和已有相关研究的成果^[38]与本研究结果进行比较,以评价珠江三角洲未来人口精细化空间分布模拟的精度。考虑到街道人口数据的可获得性,首先选取 2010 年珠江三角洲 100 个街道的实际人口数据进行模拟精度评价。由图 5 可知,模型模拟总体精度高, R^2 达到了 0.90,最小误差仅为 0.003%,平均相对误差为 17.4%,表明模拟的人口空间分布与实际的人口空间分布一致性较高。已有研究中的 Worldpop、GPWv4 和中国公里网格人口分布数据集等 3 种具有代表性的人口空间分布数据集与广东省街道/乡镇级行政区实际人口拟合的趋势线 R^2 分别为 0.85、0.84 和 0.76^[39],小于 0.9,说明本研究模拟效果更好。本文在人口分布指示性因子权重确定的方法上与 Worldpop 一致,不同点在人口分布指示性因子的

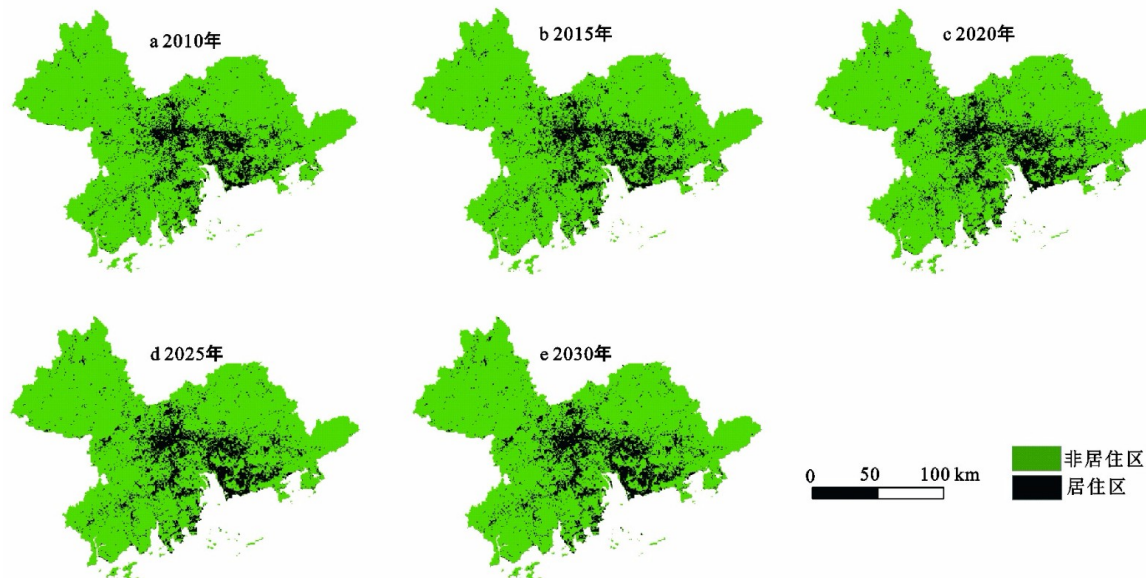


图4 2010—2030 年珠江三角洲城市群居住区模拟

Fig.4 Simulation of residential area in the Pearl River Delta from 2010 to 2030

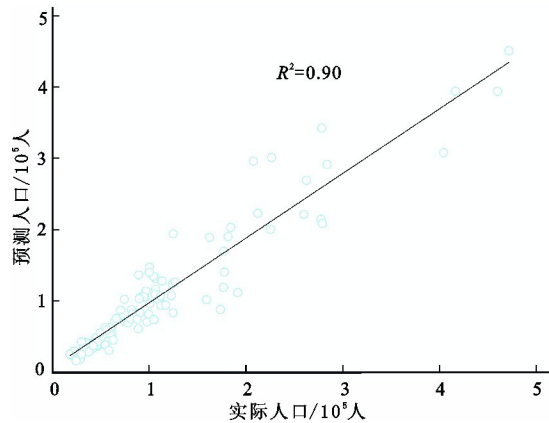


图5 2010年珠江三角洲人口空间分布精度评价

Fig.5 The accuracy of evaluation on spatial population distribution in the Pearl River Delta in 2010

选取上, Worldpop 数据集选取夜间灯光数据反映人类的活动范围, 而本研究采用 POI 地理大数据, 通过比较两者拟合的趋势线 R^2 , 可以看出, 采用 POI 地理大数据进行人口空间分布模拟所得的精度总体上高于采用夜间灯光数据所得的精度, 体现了 POI 地理大数据在区域未来人口精细化空间分布模拟上的优势。

3.3.2 2010—2030 年精细化人口空间分布模拟

由图 6 可以看出, 随着时间推移, 珠江三角洲城市群人口空间分布范围总体呈现由中心向外围

扩展趋势, 但变化幅度不大, 年平均变化率为 3.6%; 人口密度在广州、深圳、东莞和佛山显著呈现由市中心向外围递减趋势。2025—2030 年, 广州、深圳、东莞和佛山人口占珠江三角洲 9 市总人口的比例维持在 70% 左右, 说明未来一段时间珠江三角洲人口空间分布格局相对稳定。上述人口空间分布特点一方面体现了随着城镇化进程加快, 城市人口分布呈向外扩展趋势, 但在耕地面积保持不变的前提下, 扩展的速度缓慢; 另一方面珠江三角洲人口主要分布在经济社会发展水平较高的核心城市, 形成了显著的人口空间分布差异格局。

4 结论与讨论

本文基于人口普查数据和现状土地利用图, 分别采用队列因素法和 CA-Markov 模型模拟预测区域未来人口规模和土地利用空间格局, 同时利用多源信息融合方法, 结合 POI 地理大数据, 构建了精细化的未来人口空间分布模拟模型, 对珠江三角洲城市群 2030 年的精细化人口空间分布进行了实证分析。主要结论如下: ① “全面二胎政策” 实施前后 2 a 的珠江三角洲各区县实际人口与预测的人口规模相对误差大部分在 5% 以下, 预测结果相对可靠。② CA-Markov 模型模拟珠江三角洲城市群土地利用类型的总体 Kappa 系数

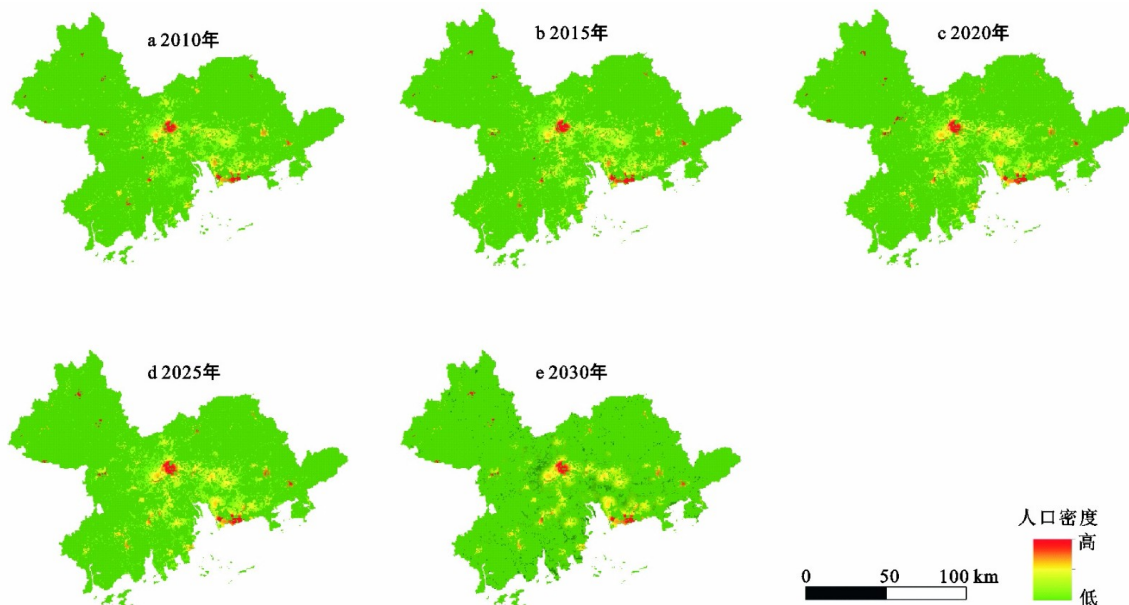


图6 2010—2030 年珠江三角洲人口空间分布

Fig.6 Spatial distribution of population in the Pearl River Delta in 2010—2030

达到 0.97, 模拟结果与土地利用现状数据一致性程度高, 具有可行性。③ 珠江三角洲城市群的精细化人口空间分布模拟数据与实际人口数据的拟合趋势线 R^2 达到了 0.90, 模拟效果优于 World-pop 数据集, 体现了 POI 地理大数据在区域未来精细化人口空间分布模拟上的优势。④ 珠江三角洲城市群人口呈现由市中心向外围递减的空间分布格局, 70% 左右的人口分布在广州、深圳、东莞和佛山等核心城市; 人口空间分布呈现由中心向外围扩展趋势, 但总体扩展幅度较小, 2010—2020 年平均变化率为 3.6%, 2025—2030 年的平均变化率仅为 0.3%, 人口空间分布格局差异显著且较为稳定。

精细化的人口空间分布预测对国家和地区制定相关发展规划至关重要, 能有效地解决现有人口统计数据更新慢和空间分辨率低的问题, 为探索人口空间分布规律和识别人口、资源与经济社会发展之间的相互作用机制提供基础。本文对区域未来精细化的人口空间分布模拟进行了初步探索, 但仍存在不足。队列因素法在未来人口规模预测中充分利用了人口统计数据的横向信息, 但人口变动参数的设置存在较大的不确定性, 如本研究中珠江三角洲各区县均采用了一致的总和生育率, 未能充分考虑地区差异性, 造成个别地区出现较高的预测偏差, 同时对“全面二胎政策”所涉及目标人群的二胎生育意愿和二胎生育分布时间, 以及人口政策的时间滞后性影响程度有所欠缺, 今后研究可有针对性地考虑总和生育率的地区差异性和人口政策的时间滞后性影响, 进一步根据目标人群的二胎生育意愿调整总和生育率, 从而提高人口数量预测的准确性; 在人口空间化过程中, 基于“无居住用地无人口”的原则, 假设人口分布于建设用地类型上, 但由于遥感土地利用影像与实际土地利用类型存在一定误差, 如遥感土地利用影像中的乡镇地区建设用地数量和面积与实际的情况相差较大, 使得部分地区人口空间化效果不佳。今后研究可进一步细化各地区的人口变动参数设置, 考虑乡镇人口的集聚性和分散性特征, 建立一定的人口分布缓冲区, 进而提高未来人口空间分布模拟精度。

参考文献(References):

[1] Li S, Yang Z W, Li H S et al. Projection of population structure

in China using least squares support vector machine in conjunction with a Leslie matrix model[J]. *Journal of Forecasting*, 2018, 37(2): 225-234.

[2] 刘劲松. 中国人口地理研究进展[J]. *地理学报*, 2014, 69(8): 1177-1189. [Liu Jinsong. A review of population geography research in China. *Acta Geographica Sinica*, 2014, 69(8): 1177-1189.]

[3] Zhao M, Cheng W M, Zhou C H et al. GDP spatialization and economic differences in South China based on NPP-VIIRS nighttime light imagery[J]. *Remote Sensing*, 2017, 9(7): 673-693.

[4] 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.

[5] 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.

[6] 阎慧臻. Logistic 模型在人口预测中的应用[J]. *大连工业大学学报*, 2008, 27(4): 333-335. [Yan Huizhen. Application of logistic model in population prediction. *Journal of Dalian Polytechnic University*, 2008, 27(4): 333-335.]

[7] Men K P, Jiang L Y, Zhu H T. Research on prediction of China's population development from 2008 to 2050[J]. *Journal of Sustainable Development*, 2008, 1(2): 55-62.

[8] 田金方, 张小斐. 干预 ARIMA 模型及其在中国人口总量预测中的实证研究[J]. *数理统计与管理*, 27(2): 263-267. [Tian Jinfang, Zhang Xiaofei. Intervention ARIMA model and its positive research in our country's population forecasting. *Application of Statistics and Management*, 27(2): 263-267.]

[9] 田飞. 人口预测方法体系研究[J]. *安徽大学学报(哲学社会科学版)*, 2011(5): 151-156. [Tian Fei. Research on population forecasting method system. *Journal of Anhui University (Philosophy and Social Sciences Edition)*, 2011(5): 151-156.]

[10] Guo A J, Ding X J, Zhong F L et al. Predicting the future Chinese population using shared socioeconomic pathways, the sixth national population census, and a PDE model[J]. *Sustainability*, 2019, 11(13): 3686.

[11] 柏中强, 王卷乐, 杨飞. 人口数据空间化研究综述[J]. *地理科学进展*, 2013, 32(11): 1693-1702. [Bai Zhongqiang, Wang Juanle, Yang Fei. Research progress in spatialization of population data. *Progress in Geography*, 2013, 32(11): 1693-1702.]

[12] 廖顺宝, 李泽辉. 基于人口分布与土地利用关系的人口数据空间化研究——以西藏自治区为例[J]. *自然资源学报*, 2003, 18(6): 599-665. [Liao Shunbao, Li Zehui. Study on spatialization of population census data based on relationship between population and land use—Taking Tibet as an example. *Journal of Natural Resources*, 2003, 18(6): 599-665.]

[13] Hu L J, He Z Y, Liu J P. Adaptive multi-scale population spatialization model constrained by multiple factors: A case study of Russia[J]. *The Cartographic Journal*, 2016, 54(3): 265-282.

[14] Yao Y, Liu X P, Li X et al. Mapping fine-scale population dis-

- tributions at the building level by integrating multisource geospatial big data[J]. *International Journal of Geographical Information Science*, 2017, 31(6): 1220-1244.
- [15] Jiang S, Alves A, Rodrigues F et al. Mining point-of-interest data from social networks for urban land use classification and disaggregation[J]. *Computers, Environment and Urban Systems*, 2015, 53: 36-46.
- [16] Tan M H, Li X B, Li S J et al. Modeling population density based on nighttime light images and land use data in China[J]. *Applied Geography*, 2018, 90: 239-247.
- [17] Zhao S, Liu Y X, Zhang R et al. China's population spatialization based on three machine learning models[J]. *Journal of Cleaner Production*, 2020, 256: 120644.
- [18] He Z, Li J, Liu K et al. Kernel low-rank multitask learning in variational mode decomposition domain for multi-hyperspectral classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(7): 4193-4208.
- [19] Gaughan A E, Stevens F R, Huang Z J et al. Spatiotemporal patterns of population in mainland China, 1990 to 2010[J]. *Sci Data*, 2016, 3: 160005.
- [20] 闫东升, 孙伟, 孙晓露. 长江三角洲人口时空格局演变及驱动因素研究[J]. *地理科学*, 2020, 40(8): 1285-1292. [Yan Dongsheng, Sun Wei, Sun Xiaolu. Spatial-temporal pattern evolution and driving factors of population in the Yangtze River Delta. *Scientia Geographica Sinica*, 2020, 40(8): 1285-1292.]
- [21] Li J G, Li J W, Yuan Y Z et al. Spatiotemporal distribution characteristics and mechanism analysis of urban population density: A case of Xi'an, Shaanxi, China[J]. *Cities*, 2019, 86: 62-70.
- [22] 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): 0107042.
- [23] Wang L Y, Fan H, Wang Y K. Improving population mapping using LuoJia 1-01 nighttime light image and location-based social media data[J]. *Science of the Total Environment*, 2020, 730: 139148.
- [24] 曾永明. 中国人口空间分布形态模拟与预测——基于“五普”和“六普”的分县尺度人口密度研究[J]. *人口与经济*, 2016, 6: 48-61. [Zeng Yongming. Simulation and projection on the pattern of Chinese population spatial distribution: Research with population density of county level based on the 5th and 6th population census of China. *Population & Economics*, 2016, 6: 48-61.]
- [25] 张日新, 谷卓桐. 粤港澳大湾区的来龙去脉与下一步[J]. *改革*, 2017(5): 64-73. [Zhang Rixin, Gu Zhuotong. The origin and next move of Guangdong-Hong Kong-Macao Greater Bay. *Reform*, 2017(5): 64-73.]
- [26] 广东省统计局. 广东省统计年鉴[M]. 北京: 中国统计出版社, 2010-2019. [Guangdong Statistical Bureau. *Guangdong statistical yearbook*. Beijing: China Statistics Press, 2010-2019.]
- [27] 国务院人口普查办公室, 国家统计局人口和就业统计司. 中国2010年人口普查分乡、镇、街道资料[M]. 北京: 中国统计出版社, 2012. [Population Census Office Under the State Council Department of Population, Employment Statistics National Bureau of Statistics. *Tabulation on the 2010 population census of the People's Republic of China By township, town, street*. Beijing: China Statistics Press, 2012.]
- [28] Chen Y D, Guo F, Wang J C et al. Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100[J]. *Sci Data*, 2020, 7(1): 83-85.
- [29] Tong X H, Feng Y J. A review of assessment methods for cellular automata models of land-use change and urban growth[J]. *International Journal of Geographical Information Science*, 2019, 34(5): 866-898.
- [30] Shu B R, Zhu S H, Qu Y et al. Modelling multi-regional urban growth with multilevel logistic cellular automata[J]. *Computers, Environment and Urban Systems*, 2020, 80: 101457.
- [31] 余健, 童秋英, 朱波. 改进CA-Markov模型的武汉市土地利用变化模拟[J]. *测绘科学*, 2020, 45(6): 165-171. [Yu Jian, Tong Qiuying, Zhu Bo. Simulation of land change in Wuhan city based on improved CA-Markov model. *Science of Surveying and Mapping*, 2020, 45(6): 165-171.]
- [32] John J S, Chithra N R, Thampi S G. Prediction of land use/cover change in the Bharathapuzha River basin, India using geospatial techniques[J]. *Environment Monitor Assessment*, 2019, 191(6): 354-369.
- [33] 张学儒, 周杰, 李梦梅. 基于土地利用格局重建的区域生境质量时空变化分析[J]. *地理学报*, 2020, 75(1): 160-178. [Zhang Xueru, Zhou Jie, Li Mengmei. Analysis on spatial and temporal changes of regional habitat quality based on the spatial pattern reconstruction of land use. *Acta Geographica Sinica*, 2020, 75(1): 160-178.]
- [34] Aburas M M, Abdullah S H, Ramli M F et al. Simulating and monitoring future land-use trends using CA-Markov and LCM models[J]. *IOP Conference Series:Earth and Environmental Science*, 2018, 169: 012050.
- [35] Yue T X, Wang Y A, Liu J Y et al. Surface modelling of human population distribution in China[J]. *Ecological Modelling*, 2005, 181(4): 461-478.
- [36] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45: 5-32.
- [37] Zhao Z L, Wu X, Zhang Y L et al. Assessment of changes in the value of ecosystem services in the Koshi River Basin, Central High Himalayas based on land cover changes and the CA-Markov Model[J]. *Journal of Resources and Ecology*, 2017, 8(1): 67-76.
- [38] 马钰琪, 朱秀芳, 刘宪锋, 等. 基于夜间灯光数据和多地理因子数据的人口空间化方法——以辽宁省为例[J]. *北京师范大学学报(自然科学版)*, 2015, 51(增刊1): 57-61. [Ma Yuqi, Zhu Xiufang, Liu Xianfeng et al. A population spatialization method based on DMSP/OLS night-time light data and weighted

multi-geographic factors: The example of Liaoning Province. Journal of Beijing Normal University(Natural Science), 2015, 51(S1): 57-61.]

[39] 林丹淳, 谭敏, 刘凯, 等. 代表性人口空间分布数据集的精度

评价——以2010年广东省为例[J]. 热带地理, 2020, 40(2): 346-356. [Lin Danchun, Tan Min, Liu Kai et al. Accuracy comparison of four gridded population datasets in Guangdong Province, China. Tropical Geography, 2020, 40(2): 346-356.]

Simulation of Fine Spatial Distribution of Regional Future Population Based on Geographical Big Data and Multisource Fused Method: A Case of the Pearl River Delta

He Yanhu^{1,2}, Gong Zhenjie¹, Lin Kairong³

(1. School of Ecology, Environment and Resources, Guangdong University of Technology, Guangzhou 510006, Guangdong, China; 2. Guangdong Provincial Key Laboratory of Water Quality Improvement and Ecological Restoration for Watersheds, Guangzhou 510006, Guangdong, China; 3. School of Civil Engineering, Sun Yat-sen University, Zhuhai 519082, Guangdong, China)

Abstract: The fine spatial distribution of regional population in future plays an important role in constructing relevant future plans for country and regions. More previous researches focused on the spatial distribution of current regional population, while little deals with the fine spatial distribution of future population. In this article, taking the fine simulation of spatial distribution of population in 2030 for counties in the Pearl River Delta (PDR) as an example, we firstly simulated future population and land use type using the methods of cohort-component and CA-Markov model, respectively, according to the census data and current land use maps. And combined with the big data of Point of Interest (POI), a model that simulates the fine spatial distribution of regional population in future was constructed by the multisource fused method. The results are as follows: Firstly, the absolute relative error is mainly less than 5% when predicting demographic data of PDR with the method of cohort-component. The Kappa coefficient of land use simulation in PDR by CA-Markov model is as high as 0.97. Secondly, the R^2 for fitting line between the simulated spatial distribution of population data and the actual population data reaches 0.90, which is better than Worldpop data set and demonstrates the advantage of big data of POI in the fine spatial distribution simulation of population. Thirdly, the spatial distribution of population in PDR decreases from the center to the periphery, performing the significant spatial difference. Meanwhile, 70% of the population concentrates in the core cities of Guangdong-Hong Kong-Macao Greater Bay, the core cities mainly include Guangzhou, Shenzhen, Dongguan and Foshan. The spatial distribution of population in PRD was significantly different and relatively stable, and showed a trend of expansion from the center to the periphery, while the overall expansion was small. This study puts forward a multisource fused method for the fine spatial distribution of regional future population, and provides scientific basis and reference for the region to make relevant development plans in the future.

Key words: future population; fine spatial distribution; CA-Markov model; multisource information; the Pearl River Delta