

文献 CSTR:

32001.14.11-6035.noda.2024.0157.zh



文献 DOI:

10.11922/11-6035.noda.2024.0157.zh

数据 DOI:

10.57760/sciencedb.j00001.00882

文献分类: 地球科学

收稿日期: 2024-09-04

录用日期: 2025-01-23

发表日期: 2025-04-28

王良^{1,2}, 张晓东^{1,3*}, 韩雪华¹, 许丹丹¹, 张敏¹, 张真继², 张兴华²

1. 北京市城市规划设计研究院, 北京 100045

2. 北京交通大学, 北京 100044

3. 清华大学, 北京 100084

摘要: 本数据集基于 2020 年北京市第七次人口普查数据, 通过将普查数据与街道和路网矢量数据进行关联, 实现了北京市地块级人口的精细空间分布。研究通过 OpenStreetMap (OSM) 道路网络数据, 形成北京市自然街区边界, 并整合高德建筑轮廓和城市建筑物屋顶矢量数据, 结合建筑层数等属性, 构建线性回归模型, 最终获得精细化的地块级人口数据。数据结果通过与 WorldPop 数据的对比校核, 模型的 R^2 达到 0.91, 皮尔逊相关系数为 0.953, 误差控制在合理范围内, 验证了数据集的高精度和适用性。本数据集覆盖北京市所有街道和社区单元, 提供了地块级的人口空间分布信息, 具有较高的空间分辨率和数据质量。数据以 Shapefile 格式存储, 数据量约为 182 MB。本数据集对城市规划、资源配置等领域的研究具有重要意义。

关键词: 人口统计学; 地块级; 空间分布; 北京; 2020 年

数据库（集）基本信息简介

数据库（集）名称	北京地块级常住人口数据集（2020 年）
数据通信作者	张晓东 (zhang-xd23@mails.tsinghua.edu.cn)
数据作者	王良, 张晓东, 韩雪华, 许丹丹, 张敏, 张真继, 张兴华
数据时间范围	2020年
地理区域	地理范围39°26′至41°03′N, 115°25′-117°30′E
数据量	182 MB
数据格式	*.shp
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.00882
基金项目	国家重点研发计划项目 (2021YFA1000300、2021YFA1000304)
数据库（集）组成	本数据集共包括 1 个压缩文件: Block_population.zip。其中, Block_population.shp 是北京市地块级人口数据的矢量图层, 包含了各地块的属性信息, 如地块所属的街道、所属的行政区划、地块面积、地块上的人口数和人口密度等。

引言

* 论文通信作者

张晓东: zhang-xd23@mails.tsinghua.

edu.cn

随着全球人口不断增长, 如何精确地描述和分析人口的空间分布, 已经成为 21 世纪地理科学的重要议题之一^[1]。人口的分布不仅影响着空间资源的配置与使

用，还直接关系到城市的可持续发展与社会治理的有效性。准确、精细的人口分布数据对于区域规划、灾害风险评估、公共服务配置等领域至关重要^[2]。

在高精度人口分布数据集的研究方面，已有不少机构和学者发布了相关的数据产品，例如在国际上，有美国橡树岭国家实验室发布的 LandScan^[3]、英国南安普顿大学的 WorldPop 数据^[4]以及哥伦比亚大学的 GPW（Gridded Population of the World）^[5]等。目前，这些数据集在全球范围内被广泛认可并使用，但在中国的应用中，它们表现出了明显的精度差异，特别是在细尺度上的人口空间表达能力，这些数据集仍存在一定的局限性^[6]。在中国，也有不少学者通过多种方法构建了不同尺度的人口分布数据集，如徐新良^[7]基于土地利用类型、夜间灯光亮度、居民点密度等人口分布影响因子，通过空间化制备了中国人口空间分布公里网格数据集 CnPop、王卷乐等^[8]利用 2010 年国家统计局发布的中国乡镇单元人口统计数据，研制了 2010 年中国乡镇（街道）级人口密度数据集等。针对这些不同数据源的精度，李紫涵等人^[6]对比了 LandScan、WorldPop、GPWv4 和 CnPop 四种数据集在陕西省乡镇尺度上的精度，发现 WorldPop 的整体精度最高，而 LandScan、GPWv4 和 CnPop 的精度偏低。然而，这些数据集大多基于较大的统计单元或者基于网格（公里网格或百米网格）的形式，因此在反映更精细的地理空间特性方面存在局限性，如道路、河流等地形特征对人口分布的影响在这些网格化数据中难以得到准确体现。目前，乡镇及以下尺度的高精度人口分布数据在中国仍然匮乏，这主要源于乡镇行政区划数据获取的困难和人口统计数据与空间数据进行精细匹配的复杂性^[9]。虽然已有研究尝试通过多种技术手段来获取更精细的人口分布数据，但依然面临着数据来源有限、技术处理复杂，以及表达精度不高等挑战^[6]。因此，现有的大多数数据源虽然在一定程度上提供了较高的空间分辨率，但在精细化地理空间表达方面仍显不足。

针对这一问题，本文以 2020 年第七次全国人口普查数据为基础，结合路网、建筑等数据，提出了一种基于大数据的地块级人口数据生成模型。通过引入居住类型兴趣面（Area of Interest, AOI）数据、建筑斑块面积和建筑物层数等关键变量，模型能够更准确地反映人口的空间分布，并最终生成了北京地块级常住人口数据集。这不仅为北京市的精细化人口研究提供了关键支持，也为其他大型城市的人口空间分布研究提供了新的方法与思路。同时，基于道路划分的自然街区人口数据研究，更精确地反映了人口的实际空间分布，为城市规划和社会治理提供了更加科学的支持。

1 数据采集和处理方法

1.1 数据采集

研究所使用的数据均为 2020 年通过官方统计、地理信息平台 and 开放数据接口等方式获取，涵盖了人口普查、行政区划、建筑轮廓、地理信息和人口分布等多个方面。各数据集的具体来源和描述如下：

1) 北京市第七次（2020 年）全国人口普查数据^[10]。该数据由国家统计局发布，包含了北京市 2020 年街道尺度的详细人口统计信息，如常住人口数量、性别、年龄和教育程度等统计信息。

2) 北京市街道行政区划数据。数据来自北京市地理信息公共服务平台（<https://beijing.tianditu.gov.cn/>），该数据集由国家基础地理信息中心提供，包含研究区域的空间范围和行政边界的详细信息。

3) OpenStreetMap（OSM）数据。OSM 是一个由全球志愿者共同维护的开源地理信息数据集，数据涵盖了全球范围内的道路、建筑物、土地利用和其他地理要素，OSM 数据被广泛地应用于城市扩展模拟、人口分布分析及灾害响应研究。

4) 城市建筑物屋顶矢量数据集（2020）^[11]。该数据集可在国家青藏高原数据中心开放获取，是一个由中国 90 个城市的建筑物屋顶区域数据组成的矢量化数据集，这些数据利用高分辨率的开放获取遥感图像，通过一个基于地理空间人工智能（GeoAI）框架的自动提取方法生成。该数据集在不同地区测试样本上的精度达到了 1 米的空间分辨率、97.95% 的总体准确率和 83.11% 的 F1 分数，目前被广泛引用。

5) AOI 数据。本研究利用高德地图 API 接口（<https://restapi.amap.com/v3/place/text?>）获取居住类 POI（Point of Interest）编号，并进一步通过“<https://ditu.amap.com/detail/get/detail?id=编号>”获取相应的 AOI 数据。

6) 建筑轮廓数据。数据来自高德地图（<https://ditu.amap.com/>），通过高德地图前端 JS 获取的 2020 年的数据集。建筑轮廓数据包含了建筑物的位置、形状和楼层数等信息。

7) WorldPop 数据^[4]。数据来源于 WorldPop 官网，是一个全球高空间分辨率的人口密度数据集，通过整合土地利用、夜间灯光强度等多源数据和回归模型生成的低至百米的人口数据源。

1.2 处理方法

1.2.1 自然街区数据处理

研究依据城市规划和土地利用的实际情况，结合 OSM 数据的特点，制定了明确的划分标准。城市地块通常指城市规划区内具有明确边界和功能的区域，而乡村地块则指城市规划区外的自然形成的居住或农业区域。地块的划分依据包括但不限于道路网络、土地利用类型、行政边界以及自然地理特征基于 OSM 的道路网络数据进行处理，生成北京市自然街区。首先，对 OSM 数据进行初步筛选，移除冗余信息，如重复的道路或非道路要素。在此过程中，还排除了一些对自然街区划分无实际意义的细小道路和步行路径。接下来，通过缓冲区操作扩展道路边界，并使用融合功能合并相邻街道的边界线。随后，通过叠加分析对街道的 shp 面数据进行裁剪，以生成更为精确的街区边界。在此基础上，对裁剪后的多边形进行拆分，确保各街区地块的合理性。进一步采用擦除操作，对剩余区域进行精细处理，移除不必要的部分。最后，利用中线提取工具（如 ArcGIS Pro 中的 Polygon To Centerline 工具）从缓冲后的地块数据中提取中心线，并通过 Polygonize 工具将中心线转换为多边形数据，最终生成北京市自然街区的空间数据集。通过系统化的处理步骤，确保了自然街区划分的准确性和适用性，为后续的空间分析提供了可靠的数据基础。

1.2.2 北京建筑数据库融合方法

在本研究中，为了构建全面的北京建筑数据库，整合了城市建筑物屋顶矢量数据集（2020）^[11]与高德的建筑轮廓数据两种数据源。首先，利用高德地图提供的建筑轮廓数据，该数据集覆盖了北京市的大部分地区，提供了建筑物的位置、形状和尺寸等详细信息。然而，由于高德建筑轮廓数据在远郊区县，尤其是建设强度较低的区域（包括两线三区中的限制建设区和非增长引导区等）的建筑识别率较低，导致这些区域的建筑数据存在明显不足。

为弥补这一不足，采用了城市建筑物屋顶矢量数据集（2020），该数据集通过高分辨率遥感图像及 GeoAI 框架的自动提取方法生成，尤其针对建设强度较低的区域进行了精确的建筑识别。通过将该数据集中的建筑信息与高德建筑轮廓数据相结合，构建了覆盖全市域的完整北京建筑数据库。此外，为了进一步完善数据库，假设建设强度较低区域的建筑高度均为 1 层，并在此基础上为整合

后的建筑数据库添加了建筑楼层数的信息，最终形成了包含建筑楼层数的完整北京建筑数据库。

1.2.3 地块级人口数据计算流程

本文基于以下假设对北京市的地块级人口进行了计算。即北京市的常住人口全部居住在居住用地上。这一假设基于居住用地通常具备完善的生活设施和住房条件，能够满足常住人口的需求。因此，将常住人口与居住用地直接关联，有助于提高模型对实际人口分布的反映精度。在处理居民楼、宾馆酒店和写字楼等建筑时，由于研究专注于常住人口分布，主要考虑其分布在居住用地上的建筑物，对于宾馆酒店和写字楼等非居住用地上的建筑，认为它们对常住人口的贡献较小，因此在计算时予以忽略。

基于这些假设，构建了一个线性回归模型，用于估算地块级人口分布。模型中，居住空间属性（如建筑面积、建筑层数）作为人口分布的指示变量，同时引入街道界线和街道常住人口数据作为控制变量。街道的总人口数基于第七次全国人口普查（2020年）数据进行校准，以确保模型结果的准确性。地块级人口的估算过程可表示为公式（1）所示：

$$P_{plot}^M = P_{street}^M \times \left(\sum_{i=1}^{n_M} \frac{A_i^M A_i}{A_{total,street}^M} \right) / A_{res,street}^M \quad (1)$$

其中， P_{plot}^M 表示M街道中地块人口数， P_{street}^M 表示第M个街道的总人口数， A_i^M 是地块内M街道中第i个建筑的总面积， n_M 是M街道中建筑的总数， $A_{total,street}^M$ 是街道所有建筑的总面积， $A_{res,street}^M$ 是M街道居住区的建筑总面积。该公式计算了地块内所有建筑的人口加权和，即根据每个建筑面积在街道所有建筑总面积中的比例，结合街道总人口数来估算每个建筑的人口分布。其数据处理流程如图1所示。

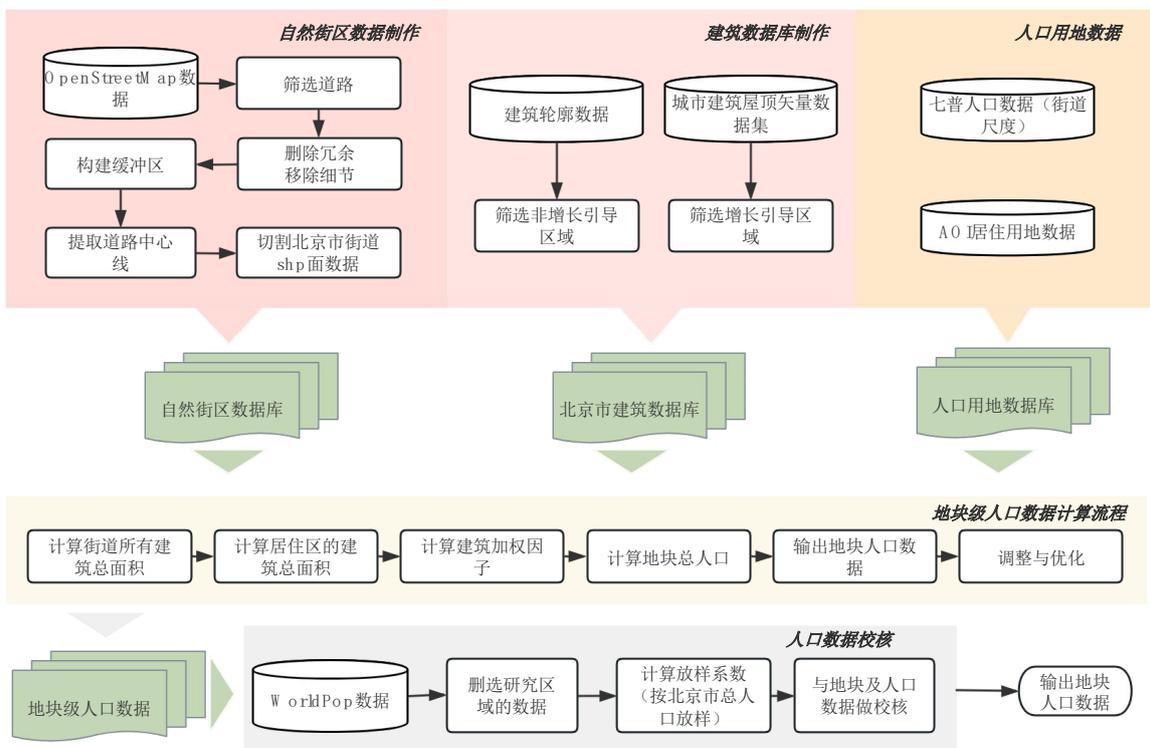


图1 数据库建设流程

Figure 1 Construction process of the dataset

2 数据样本描述

本数据集为矢量数据集，文件格式为 shp。本数据集包含 2020 年全北京市 16 个区、339 个街道乡镇的 46045 个地块的常住人口数据。数据采用 WGS 84 坐标系，投影方式为 UTM（Universal Transverse Mercator，通用横轴墨卡托）投影。北京市的常住人口主要集中在中心城区及其周边地区，呈现出中间密集、外围稀疏的分布规律。图 2 展示了 2020 年北京市分地块的人口数量示例图，图中将人口数量划分为 5 个等级，蓝色区域表示人口总数小于 1282 人的地块，红色区域表示人口总数大于或等于 31369 人的地块。

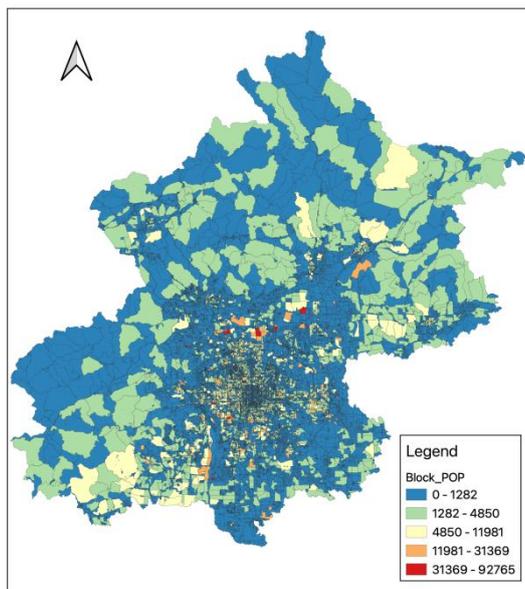


图 2 2020 年人口北京市地块级人口分布

Figure 2 Population distribution at the block-level population in Beijing in 2020

3 数据质量控制和评估

研究中主要通过 WorldPop 数据集对模型所生成的地块级人口数据进行校核，并采用多种统计分析方法和误差评估来验证模型的精度。首先，模型的预测精度通过与 WorldPop 数据在公里网格上的对比进行评估。结果如图 3 所示，模型与 WorldPop 数据之间的决定系数 R^2 为 0.91，显示出高度相关性。其皮尔逊相关系数为 0.953，回归斜率为 1.045，表明模型预测结果与 WorldPop 数据具有良好的拟合度。图 3 展示了模型预测结果与 WorldPop 数据的精度对比情况，可以看出模型与 WorldPop 数据的拟合程度较高，证明了模型在大规模数据集上的适用性和准确性。

为了进一步验证模型的稳定性和可靠性，采用了交叉验证方法。将生成的地块级人口数据集进行预处理，筛选出 blockpop 和 worldpop 字段值均大于 0 的记录，从而得到一个清洗后的数据集。该数据集被进一步划分为 5 个等大小的子集，每个子集包含原始数据集的 20%。模型在 4 个子集上进行训练，在剩余的 1 个子集上进行测试，这一过程循环进行，确保每个子集都有机会作为测试集，以评估模型在不同数据分布上的表现，交叉验证的结果表明，模型在各个子集上的表现一致，进一步增强了其泛化能力。相对误差的量化是评估模型精度的重要步骤，公式（2）所示：

$$E = \frac{p_{i1} - p_{i2}}{p_{i2}} \times 100\% \quad (2)$$

其中， E 表示相对误差， p_{i1} 表示第*i*个地块的模型预测人口值； p_{i2} 表示相应地块的实际人口统计值。图4显示了误差分布情况，可以看出模型的误差控制在合理范围内，绝大部分地块的误差较小，验证了模型的精度与可靠性。

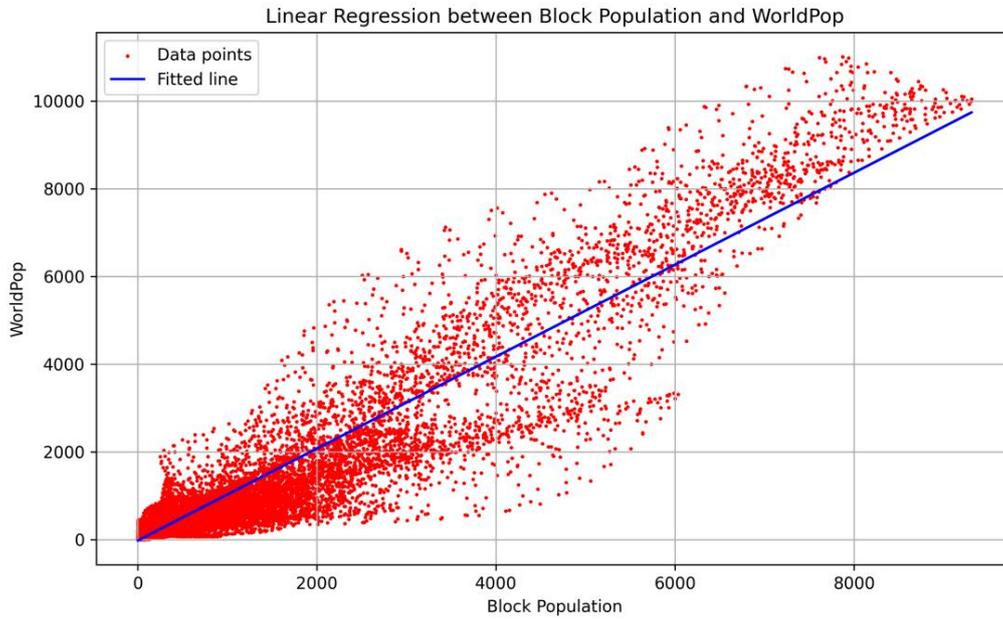


图3 WorldPop 和模型预测结果的精度对比

Figure 3 Comparison of accuracy between WorldPop and the Prediction Results

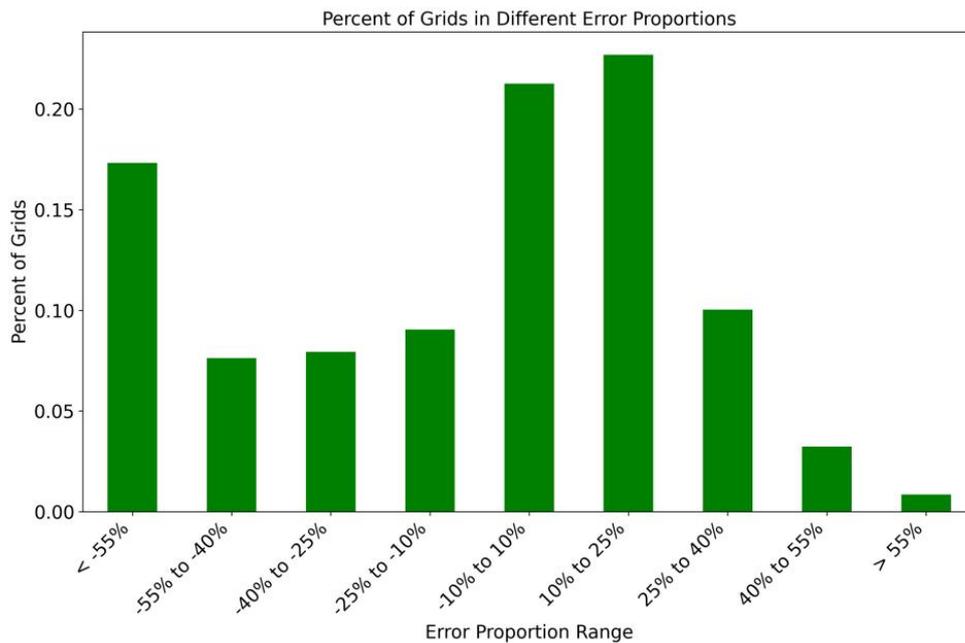


图4 误差分布统计图

Figure 4 Statistical chart of error distribution

通过精度对比图和误差分布图的结合分析，研究表明，模型能够有效生成北京市地块级人口数据，具备较高的准确性和适用性。

4 数据价值

本文基于大数据的地块级人口数据生成模型研究，以 2020 年北京市第七次人口普查数据为例，充分展示了本数据集的价值。本数据集覆盖了北京市所有街道和社区单元的人口统计数据，并通过关联数万条统计单元的矢量边界，提供了北京市精细的地块级人口分布情况。由于数据源于国家人口普查，其统计周期为 10 年一次，因此本数据集在精确反映地块级人口空间分布方面具有重要意义。

本数据集不仅在空间分辨率上优于其他国内外同类数据集，而且通过采用先进的数据处理技术和严谨的质量控制，确保了数据的可靠性和精确性。考虑到模型的可扩展性，因该模型依赖于人口普查数据、OSM 数据、建筑轮廓数据和矢量边界数据，这些数据在许多国家和地区都是可获得的，因此该方法具有较好的普适性。

通过调整模型参数和优化数据处理流程，该模型可以适应不同区域的特定需求，包括不同规模的城市或农村地区。此外，该模型的数据处理技术和质量控制措施也可以为其他区域提供参考，以确保结论的可靠性和精确性。

本数据集可以作为北京市地块级人口分布的基础数据，应用于城市规划和公共服务资源配置等多个领域，提供了重要的决策支持和科学依据。这些特性使得本数据集成为分析北京市人口分布格局、制定区域发展策略的重要工具。

尽管本数据集具有显著价值，但也存在一些局限性。主要限制包括数据时效性问题，由于依赖十年一次的人口普查，可能无法反映短期内的人口变化。此外，模型预测的准确性也依赖于输入数据的质量，不同区域数据源的可用性和质量差异可能影响模型的普适性。

5 数据使用方法和建议

北京地块级常住人口数据集（2020 年）可通过常见的地理信息系统（GIS）软件如 ArcGIS 和 QGIS 进行处理与分析。本数据集具有较高的空间分辨率，未经过复杂的统计建模，具有较强的适用性，能够直接用于研究北京市地块级人口空间分布特征、地区差异以及相关领域的深入分析。它为交通规划、灾害风险评估、环境保护、公共卫生和城市发展等多方面研究提供了重要的数据基础，特别适用于城市人口密度、土地利用和公共设施布局等领域的研究。

数据作者分工职责

王良（1990—），男，湖北省黄冈市人，硕士，高级工程师，研究方向为城市计算。主要承担工作：总体设计、技术实施与成果集成。

张晓东（1979—），男，吉林省榆树人，硕士，正高级工程师，研究方向为城市计算和规划信息化。主要承担工作：总体设计、技术实施和数据汇总。

韩雪华（1991—），女，河北省保定市人，博士，高级工程师，研究方向为城市韧性和规划大模型。主要承担工作：空间化技术支持。

许丹丹（1988—），女，山东省聊城市人，博士，高级工程师，研究方向为复杂网络和城市交通。主要承担工作：空间化技术支持。

张敏(1994—),女,河南省周口市人,博士,工程师,研究方向为城市知识图谱。主要承担工作:空间化技术支持。

张真继(1963—),男,吉林省长春市人,博士,教授,研究方向为信息经济、信息管理、物流信息管理等。主要承担工作:总体设计。

张兴华(1980—),男,吉林省吉林市人,博士,教授,研究方向为城市规划决策支持。主要承担工作:总体设计。

参考文献

- [1] 美国国家科学院国家研究理事会著. 理解正在变化的星球: 地理科学的战略方向[M]. 刘毅, 刘卫东等译. 北京: 科学出版社, 2011. [National Research Council of the National Academy of Sciences. Understanding a Changing Planet: Strategic Directions for the Geographical Sciences [M]. Translated by Liu Yi, Liu Weidong, et al. Beijing: Science Press, 2011.]
- [2] SEROK N, BLUMENFELD-LIEBERTHAL E. A simulation model for intra-urban movements[J]. PLoS One, 2015, 10(7): e0132576. DOI: 10.1371/journal.pone.0132576.
- [3] BHADURI B, BRIGHT E, COLEMAN P, et al. LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics[J]. GeoJournal, 2007, 69(1): 103 - 117. DOI: 10.1007/s10708-007-9105-9.
- [4] 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. DOI: 10.1371/journal.pone.0107042.
- [5] DOXSEY-WHITFIELD E, MACMANUS K, ADAMO S B, et al. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4[J]. Papers in Applied Geography, 2015, 1(3): 226 - 234. DOI: 10.1080/23754931.2015.1014272.
- [6] 李紫涵, 吴田军, 王洁, 等. 多源人口空间化数据集精度评析与融合: 以陕西省为例[J]. 测绘地理信息, 2024, 49(4): 76 - 82. DOI: 10.14188/j.2095-6045.2022063. [LI Z H, WU T J, WANG J, et al. The accuracy evaluation and fusion of multi-source population spatialization datasets in Shaanxi Province[J]. Journal of Geomatics, 2024, 49(4): 76 - 82. DOI: 10.14188/j.2095-6045.2022063.]
- [7] 徐新良. 中国人口空间分布公里网格数据集. 中国科学院资源环境科学数据中心数据注册与出版系统 (<http://www.resdc.cn/DOI>), 2017. DOI: 10.12078/2017121101.
- [XU X L. China Population Spatial Distribution Grid Dataset [DB]. Chinese Academy of Sciences, Resource and Environment Science Data Center, Data Registration and Publishing System, 2017. DOI: 10.12078/2017121101.]
- [8] 王卷乐, 王晓洁, 王明明, 等. 中国乡镇(街道)人口密度数据集(2010年)[J/OL]. 中国科学数据, 2020, 5(3). (2020-09-11). DOI: 10.11922/csdata.2020.0013.zh. [WANG J L, WANG X J, WANG M M, et al. A dataset of township-level population density in China(2010)[J/OL]. China Scientific Data, 2020, 5(3). (2020-09-11). DOI: 10.11922/csdata.2020.0013.zh.]

- [9] 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): 1692 - 1702. [BAI Z Q, WANG J L, YANG F. Research progress in spatialization of population data[J]. Progress in Geography, 2013, 32(11): 1692 - 1702.]
- [10] 《中国人口普查年鉴—2020》编委会. 中国人口普查年鉴—2020 [Z]. 2022: 4.10.44303/y.cnki.yzggqp.2022.000001. [Editorial Board of "China Population Census Yearbook - 2020". China Population Census Yearbook - 2020 [Z]. 2022: 4.10.44303/y.cnki.yzggqp.2022.000001.]
- [11] ZHANG Z X, QIAN Z, ZHONG T, et al. Vectorized rooftop area data for 90 cities in China[J]. Scientific Data, 2022, 9(1): 66. DOI: 10.1038/s41597-022-01168-x.

论文引用格式

王良, 张晓东, 韩雪华, 等. 北京地块级常住人口数据集(2020 年)[J/OL]. 中国科学数据, 2025, 10(2). (2025-04-28). DOI: 10.11922/11-6035.noda.2024.0157.zh.

数据引用格式

王良, 张晓东, 韩雪华, 等. 北京地块级人口数据集（2020 年）[DS/OL]. V2. Science Data Bank, 2024. (2025-02-26). DOI:10.57760/sciencedb.j00001.00882.

A dataset of block-level household living population in Beijing (2020)

WANG Liang^{1,2}, ZHANG Xiaodong^{1,3*}, HAN Xuehua¹, XU Dandan¹, ZHANG Min¹,
ZHANG Zhenji², ZHANG Xinghua²

1. Beijing Municipal Institute of City Planning and Design, Beijing 100045, P.R. China

2. Beijing Jiaotong University, Beijing 100044, P.R. China

3. Tsinghua University, Beijing 100084, P.R. China

*Email: zhang-xd23@mails.tsinghua.edu.cn

Abstract: Based on the 2020 Beijing Seventh National Census data, the study integrates the census data with street and road network vector data, and produces a refined spatial distribution of block-level population in Beijing. The study utilizes OSM road network data to delineate natural neighborhood boundaries in Beijing, incorporating Gaode building footprints and urban rooftop vector data. By combining attributes like building height, a linear regression model was developed to obtain a refined block-level population dataset. Cross-validation against WorldPop data confirms the dataset's high accuracy with an R^2 of 0.91, a Pearson correlation coefficient of 0.953, and errors maintained within reasonable limits. The dataset covers all streets and community units in Beijing, providing block-level population distribution information with high spatial resolution and data quality. The dataset is stored in Shapefile format and has a size of approximately 182 MB. The dataset is of significant importance for research on urban planning and resource allocation.

Keywords: demography; block-level; spatial distribution; Beijing; 2020

Dataset Profile

Title	A dataset of block-level household living population in Beijing (2020)
Data corresponding author	ZHANG Xiaodong (zhang-xd23@mails.tsinghua.edu.cn)
Data authors	WANG Liang, ZHANG Xiaodong, HAN Xuehua, XU Dandan, ZHANG Min, ZHANG Zhenji, ZHANG Xinghua
Time range	2020
Geographical scope	Beijing (39°26'N-41°03'N, 115°25'E-117°30'E)
Data volume	182 MB
Data format	.shp
Data service system	https://doi.org/10.57760/sciencedb.j00001.00882
Source of funding	The National Key Research and Development Program of China (2021YFA1000300, 2021YFA1000304).
Dataset composition	The dataset includes one compressed file: Block_population.zip. Within this file, Block_population.shp is a vector layer of the block-level population data for Beijing, including attribute details for each block, such as the street to which the block belongs, administrative district, block area, population, and population density.