



文献 DOI:

10.11922/csdata.2020.0013.zh

数据 DOI:

10.11922/sciencedb.964

文献分类: 地球科学

收稿日期: 2020-03-31

开放同评: 2020-04-01

录用日期: 2020-08-18

发表日期: 2020-09-17

## 中国乡镇（街道）人口密度数据集（2010 年）

王卷乐<sup>1\*</sup>, 王晓洁<sup>1,2</sup>, 王明明<sup>1,2</sup>, 姚锦一<sup>1,2</sup>, 王岚涛<sup>1,3</sup>, 邱丛丛<sup>1,3</sup>

1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101

2. 山东理工大学建筑工程学院, 山东淄博 255049

3. 武汉大学城市设计学院, 武汉 430000

**摘要:** 乡镇（街道）级人口统计数据是中国目前公开的最精细的人口统计数据。将人口统计数据和行政边界数据进行属性关联, 实现人口统计数据的空间展布, 对于地理、资源、环境、生态、灾害, 以及可持续发展等方面的研究具有重要意义。本数据集将 2010 年国家统计局发布的中国 31 个省（直辖市、自治区）的乡镇（街道）单元人口统计数据按乡镇行政区划建立关联, 解决区划间空隙和重叠等拓扑问题, 采用决策树思想对 42–122 个统计单元逐一建立空间匹配, 利用统计单元人口和面积属性计算平均人口密度, 并经矢栅转换等处理形成 2010 年乡镇（街道）级人口密度数据集。经精度评定, 数据集各省（直辖市）误差均控制在 10% 之内, 总体精度达 99%。数据集为 TIFF 栅格形式, 数据量为 108 MB。

**关键词:** 人口统计学; 人口密度; 空间分布; 乡镇级; 中国

### 数据库（集）基本信息简介

数据库（集）名称	中国乡镇（街道）人口密度数据集（2010 年）
数据作者	王卷乐、王晓洁、王明明、姚锦一、王岚涛、邱丛丛
数据通信作者	王卷乐 (wangjl@igsrr.ac.cn)
数据时间范围	2010 年
地理区域	安徽省、北京市、福建省、甘肃省、广东省、广西省、贵州省、海南省、河北省、河南省、黑龙江省、湖北省、湖南省、吉林省、江苏省、江西省、辽宁省、内蒙古自治区、宁夏回族自治区、青海省、山东省、山西省、陕西省、上海市、四川省、天津市、西藏自治区、新疆维吾尔自治区、云南省、浙江省、重庆市。
空间分辨率	1 km
数据量	108 MB
数据格式	*.tif
数据服务系统网址	<a href="http://www.sciencedb.cn/dataSet/handle/964">http://www.sciencedb.cn/dataSet/handle/964</a>
基金项目	中国科学院战略性先导科技专项（A 类）（XDA19040501）；中国科学院“十三五”信息化专项科学大数据工程项目（XXH13505-07）。

\* 论文通信作者

王卷乐: wangjl@igsrr.ac.cn

## 数据库（集）组成

本数据集共包括1个压缩文件：2010中国乡镇（街道）人口数据.zip。其中：(1) 2010中国乡镇（街道）人口数据.tif是中国31省平均人口密度的栅格图像，数据量108 MB；(2) 2010中国乡镇（街道）人口数据.tfw存储空间参考信息，数据量1 KB；(3) 2010中国乡镇（街道）人口数据.tif.ovr是金字塔文件，数据量3.1 MB；(4) 2010中国乡镇（街道）人口数据.tif.aux.xml存储其他辅助信息的XML文件，数据量1.4 KB。

## 引言

“100 亿人在地球上如何生存和分布”被认为是 21 世纪地理科学的 11 个重大战略方向之一<sup>[1]</sup>。空间分布是人口的重要属性，人口密度则是表现人口空间分布最主要的形式和衡量人口分布地区差异的主要指标<sup>[2]</sup>。中国拥有世界上最多的国家人口总数，人口分布不平衡，资源环境承载压力大<sup>[3]</sup>。准确、精细的人口密度数据能够为区域规划与开发、灾害风险评估和救援、医疗资源配置和资源环境承载力评估等研究和应用提供支撑<sup>[4]</sup>。在全球尺度，美国国家橡树岭实验室完成了 LandScan 数据集<sup>[5]</sup>，哥伦比亚大学国际地球科学信息网络中心完成了 GPW（Gridded Population of the World）数据集<sup>[6]</sup>，但这些全球尺度的数据集在中国的精度存在显著的区域差异<sup>[7-8]</sup>。在中国区域，有学者基于居民地重分类方法获取长江中游 4 省人口空间分布数据<sup>[9]</sup>，基于多元回归方法获取宣城城区人口空间数据<sup>[10]</sup>，以及基于人口统计数据 and 土地利用关联建模的中国公里网格尺度人口空间分布数据集<sup>[11-13]</sup>。然而，这些数据多数是在县级人口统计数据单元基础上的产品，缺乏更精细尺度的表达刻画。乡镇级人口统计数据是目前官方公布的最精细的人口数据，然而乡镇级的人口空间数据缺乏。究其原因在于两点，一是乡镇级行政区划矢量数据获取困难，且存在大量的行政区划和地名变更问题，难以直接利用。二是乡镇级人口统计数据的空间关联需要进行大量的涉及乡镇地名和行政区划的空间数据匹配处理工作，工作量且技术要求较高。柏中强等人<sup>[4]</sup>结合第 5 次人口普查构建 2000 年中国 27 省乡镇级平均人口密度数据集，但尚未见 2010 年的同类数据产品。本文将 2010 年国家统计局发布的中国 31 个省（直辖市、自治区）乡镇（街道）单元的人口统计数据按乡镇行政区划建立空间匹配。利用乡镇级行政区划数据的人口属性和面积属性计算乡镇级平均人口密度，经过矢栅转化形成本数据集。

## 1 数据采集和处理方法

## 1.1 数据采集

统计数据来源于中国国家统计局<sup>[14]</sup>，乡镇（街道）边界数据来源于地理国情监测云平台<sup>[15]</sup>、国家地球系统科学数据中心<sup>[16]</sup>、中国科学院资源环境科学数据中心<sup>[17]</sup>和部分地图图件。

## 1.2 处理方法

## 1.2.1 拓扑处理

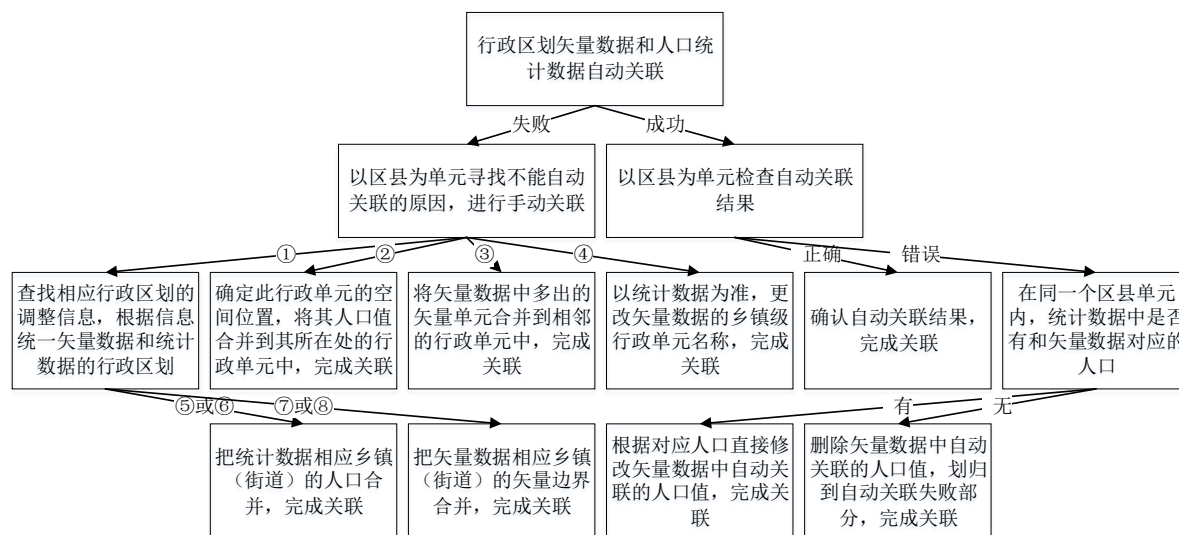
乡镇级行政区划矢量数据易存在两种拓扑错误：行政区划边界处存在重叠或空隙现象。造成这种错误有两种情况，一种是由乡镇（街道）矢量单元发生偏移造成的有规律的重叠和空隙，另一种是由乡镇（街道）矢量单元存在残缺或多余造成的不规则的重叠和空隙。其中第一种情况利用 ArcGIS 软件的空间校正（Spatial Adjustment）功能将乡镇（街道）矢量单元移回原来的位置，第二种情况利

用 ArcGIS 的建立拓扑功能中的“不能重叠（Must Not Overlap）”和“不能有缝隙（Must Not Have Gaps）”规则实现对残缺部分的补充和对多余部分的去除。

### 1.2.2 属性关联

乡镇级人口统计数据与乡镇级行政区划矢量数据的属性关联是指人口统计数据的乡镇（街道）单元的人口总值属性同行政区划矢量数据的空间属性相关联，实现乡镇级人口统计值在空间上的分布。两者属性关联可以通过两者的公用字段（键）来自动关联，但往往由于存在以下 3 种情况导致人口统计数据和行政区划矢量数据不能完全自动关联成功：①乡镇级行政区划矢量数据的表征时间和乡镇级人口统计数据的时间不一致，在这段时间差内发生过行政区划的调整；②乡镇级行政区划矢量数据存在误差，矢量数据的行政区划单元多于或少于人口统计数据的行政单元；③人口统计数据和行政区划矢量数据对乡镇（街道）的命名方式不统一。由以上原因造成自动关联失败的乡镇（街道）单元需要进行手动关联。

本文采用决策树的思想实现乡镇级人口统计数据与乡镇级行政区划矢量数据的属性关联，如图 1 所示。此决策树共包含四层。第一层根节点以乡镇（街道）的名称作为公用字段（键）实现矢量数据和统计数据的自动关联，由此会出现自动关联部分和不能自动关联部分两个有向边。第二层对这两部分做进一步处理，其中自动关联部分需要以区县为单元检验自动关联的准确性；自动关联失败的部分以区县为单元进行手动关联，寻找自动关联失败的原因，由此产生 6 个有向边。第三层包含 2 个内部节点和 4 个叶节点。左边的内部节点是由行政区划调整导致的自动关联失败，需要根据调整信息统一矢量数据和统计数据的行政区划；右边的内部节点是自动关联错误的情况下，手动寻找统计数据中对应的人口值，2 个内部节点分别产生 2 个有向边。中间的 4 个叶节点是第二层中间 4 个有向边的具体解决方案。第四层是第三层 4 个有向边的具体解决方案。具体流程如图 1 所示。



- ① 矢量数据的表征时间和统计数据不一致，且在这段时间差内进行过行政区划调整；
- ② 在矢量数据中找不到和统计数据相对应的行政单元；
- ③ 在统计数据中找不到和矢量数据相对应的行政单元的人口值；
- ④ 矢量数据和统计数据对行政单元的命名方式不一致；
- ⑤ 矢量数据表征时间早于统计数据，且在时间差内发生过行政单元的新建；
- ⑥ 矢量数据表征时间晚于统计数据，且在时间差内发生过行政单元的撤并；
- ⑦ 矢量数据表征时间晚于统计数据，且在时间差内发生过行政单元的新建；
- ⑧ 矢量数据表征时间早于统计数据，且在时间差内发生过行政单元的撤并。

图 1 乡镇级人口空间数据库建设流程<sup>[18]</sup>

### 1.2.3 平均人口密度计算

根据公式（1），计算各乡镇（街道）平均人口密度：

$$D = \frac{P}{A} \quad (1)$$

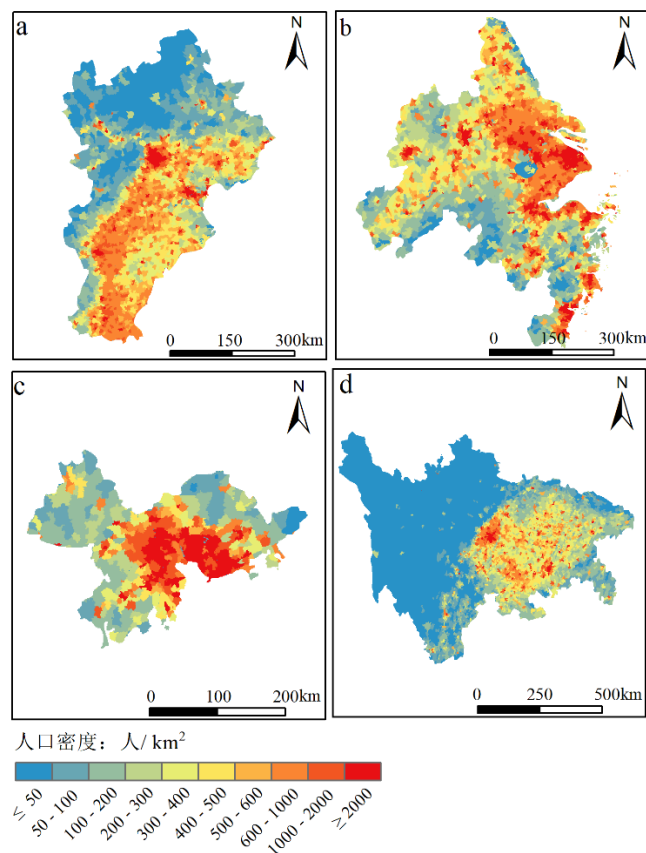
式中， $D$  为人口密度（人/ $\text{km}^2$ ）； $P$  为乡镇（街道）人口数量（人）； $A$  为乡镇（街道）面积（ $\text{km}^2$ ）。

### 1.2.4 矢量转栅格

利用 ArcGIS 的 Polygon to Raster 工具将人口密度矢量图层转化为栅格图层，空间分辨率为  $1 \text{ km} \times 1 \text{ km}$ 。

## 2 数据样本描述

本数据集为栅格数据类型，空间分辨率为  $1 \text{ km}^2$ ，后缀名为 tif。本数据集包括全国 22 个省、5 个自治区以及 4 个直辖市内 42 122 个统计单元的平均人口密度信息。数据以 Krasovsky 椭球为基准，投影方式为 Albers 投影，最后生成中国 2010 年乡镇街道人口密度。中国 31 个省（直辖市、自治区）的人口密度由东部地区至西部地区总体上呈现出密集到稀疏的分布规律，人口主要分布在中国的中东部地区，集中表现在京津冀、长三角、珠三角等经济文化较发达地区以及山东、河南等地区。图 2 是中国部分城市群 2010 年分乡镇街道人口密度示例图，图中将人口密度分为 10 级，蓝色区域代表人口平均密度小于等于  $50 \text{ 人}/\text{km}^2$ ，红色范围表示人口平均密度大于等于  $2000 \text{ 人}/\text{km}^2$ 。



a 京津冀; b 长三角; c 珠三角; d 成渝

图 2 中国部分城市群 2010 年分乡镇街道人口密度分布

### 3 数据质量控制和评估

利用省级行政区划矢量边界统计栅格数据集，得到每个省份的栅格数据集人口值，与 2010 年省级人口统计数据对比（图 3）。计算全国省级统计人口数据与栅格数据集人口数据之间的相对误差（图 4），其误差计算公式（式 2）如下所示：

$$E = \left( \frac{p_{i1} - p_{i2}}{p_{i2}} \right) \times 100, (i = 1, 2, 3, \dots, 31) \quad (2)$$

其中， $E$  表示相对误差； $p_{i1}$  表示第  $i$  个省级行政区栅格数据集的值； $p_{i2}$  表示第  $i$  个省级行政区统计人口值。

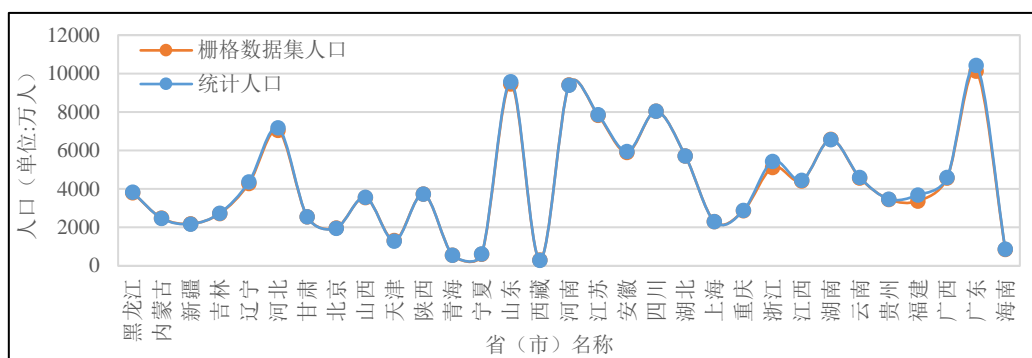


图 3 栅格数据集人口与统计人口对比

由图 3 对比可得，数据集的人口分布趋势与省市级统计人口值基本一致，总体精度高达 99%，说明本人口数据集精度较高。部分省份（例如福建省和浙江省）因地理条件复杂、乡镇（街道）行政边界破碎且多变，造成矢栅转换过程中产生误差略大，但仍然控制在 10% 以内，能够满足应用需求。

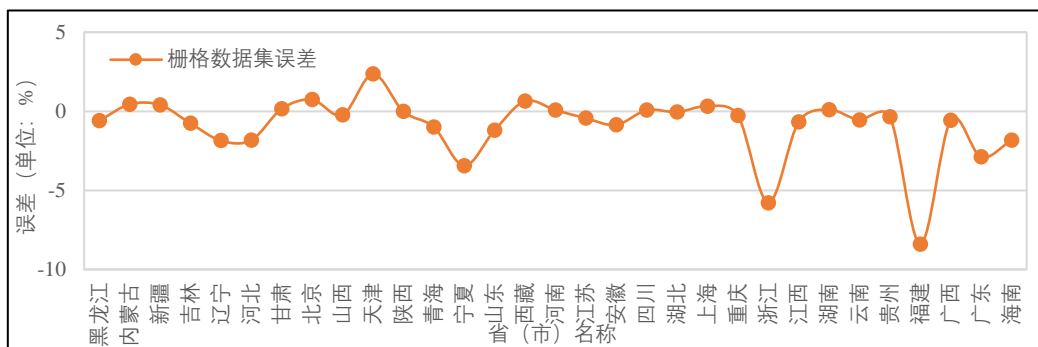


图 4 栅格数据集人口与统计人口误差

### 4 数据价值

本数据集包括 2010 年全中国 31 个省级行政区，其中 22 个省、5 个自治区和 4 个直辖市乡镇街道单元内人口统计数据。本数据集将全国范围内（不包括港、澳、台特别行政区）乡镇街道级人口统计数据与 42 122 个统计单元矢量边界进行关联，在空间上展示中国最精细的乡镇级人口分布数据。中国的全国人口普查数据每 10 年统计一次，因此本数据集可以成为乡镇（街道）级人口数据分布的本底数据集，可以在空间尺度上反映我国人口的精细分布状况，为经济发展、城市规划、医疗资源



配置、灾害评估等提供数据参考。

## 5 数据使用方法和建议

本数据集可以用能够处理栅格或矢量数据的地理信息系统相关软件打开<sup>[4]</sup>,包括开源的 QGIS 和 GRASS GIS 软件、商业化的 ArcGIS、SuperMap、ENVI 等。本数据集的空间分辨率为 1 km,可以直接用来表征中国乡镇街道级的人口空间分布,为中国人口、资源、环境、灾害、生态等科学研究提供重要基础和本底数据。同时也可以结合同一尺度的土地利用数据、居民点数据、DEM 数据、夜间灯光数据、道路交通数据、社交媒体数据等多源数据进行融合和综合分析,进一步揭示中国区域人口空间化的地理分布特征,为人类活动模拟、区域规划、资源配置、灾害评估等研究和应用提供数据产品。

## 数据作者分工职责

王卷乐（1976—），男，河南省洛阳市人，博士，研究员，研究方向为科学数据共享、地理信息系统与遥感应用研究。主要承担工作：总体设计、技术实施与成果集成。

王晓洁（1995—），女，山东省烟台人，硕士，研究方向为人口空间化。主要承担工作：乡镇人口数据空间化、数据汇总。

王明明（1993—），男，河北省沧州市人，硕士，研究方向为人口空间化。主要承担工作：提供空间化技术支持。

姚锦一（1996—），男，江苏省苏州市人，硕士，研究方向为遥感科学与技术。主要承担工作：乡镇人口数据空间化。

王岚涛（1993—），男，湖北省武汉市人，硕士，研究方向为城市设计与城市健康。主要承担工作：乡镇人口数据空间化。

邱丛丛（1994—），女，山东省临沂市人，硕士，研究方向为建筑设计。主要承担工作：乡镇人口数据空间化。

## 参考文献

- [1] National Research Council(NRC). 理解正在变化的星球: 地理科学的战略方向[M]. 刘毅, 刘卫东, 译. 北京: 科学出版社, 2011.
- [2] 祝卓. 人口地理学[M]. 北京: 中国人民大学出版社, 1991.
- [3] 葛美玲, 封志明. 基于 GIS 的中国 2000 年人口之分布格局研究——兼与胡焕庸 1935 年之研究对比[J]. 人口研究, 2008, 32(1): 51-57.
- [4] 柏中强, 王卷乐. 中国 27 省乡镇（街道）级人口密度数据集（2000 年）[J/OL]. 中国科学数据, 2016, 1(1). DOI: 10.11922/csdata.170.2015.0002.
- [5] DOBSON J E, BRIGHT E A, COLEMAN P R, et al. LandScan: A global population database for estimating populations at risk[J]. Photogrammetric Engineering and Remote Sensing, 2000, 66(7): 849-857.
- [6] BALK D L, DEICHMANN U, YETMAN G, et al. Determining global population distribution: Methods, applications and data. Advances in Parasitology[J], 2006, 62(4): 119-156.

- [7] 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): 118-128.
- [8] 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).
- [9] 王珂靖, 蔡红艳, 杨小唤, 等. 基于城镇居民用地再分类的人口数据空间化方法研究——以长江中游 4 省为例[J]. 遥感技术与应用, 2015, 30(5): 987-995.
- [10] 董南, 杨小唤, 黄栋, 等. 引入城市公共设施要素的人口数据空间化方法研究[J]. 地球信息科学学报, 2018, 20(7): 918-928.
- [11] 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.
- [12] 徐新良. 中国人口空间分布公里网格数据集[DB/OL]. 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.
- [13] 卓莉, 陈晋, 史培军, 等. 基于夜间灯光数据的中国人口密度模拟[J]. 地理学报, 2005, 60(2): 266-276.
- [14] 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.
- [15] 地理国情监测云平台. 乡镇行政区划矢量数据 [DB/OL]. [2020-08-20]. <http://www.dsac.cn/DataProduct/Index/2019>.
- [16] 国家地球系统科学数据中心. 中国 1:25 万乡镇界限数据 (2000 年) [DB/OL]. [2020-08-20]. <http://www.geodata.cn/data/datadetails.html?dataguid=23491717434844&docid=10966>.
- [17] 中国科学院资源环境科学数据中心. 中国乡镇行政边界数据 [DB/OL]. [2020-08-20]. <http://www.resdc.cn/data.aspx?DATAID=203>.
- [18] 王明明, 王卷乐. 山东省乡镇级格网化人口空间分布数据集 (2000、2010 年) [J/OL]. Science Data Bank, 2019. (2019-02-28). DOI: 10.11922/sciencedb.736.

## 论文引用格式

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.

## 数据引用格式

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）[DB/OL]. Science Data Bank, 2020. (2020-03-31). DOI: 10.11922/sciencedb.964.

## A dataset of township-level population density in China (2010)

Wang Juanle<sup>1\*</sup>, Wang Xiaojie<sup>1,2</sup>, Wang Mingming<sup>1,2</sup>, Yao Jinyi<sup>1,2</sup>,  
Wang Lantao<sup>1,3</sup>, Qiu Congcong<sup>1,3</sup>

1. State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, P.R. China

2. School of Civil and Architectural Engineering, Shandong University of Technology, Zibo, Shandong 255049, P.R. China

3. School of urban design, Wuhan University, Wuhan, Hubei 430000, P.R. China

\*Email: wangjl@igsnr.ac.cn

**Abstract:** Township is so far the smallest unit for official demographic statistics in China. When correlated with administrative boundary, demographic data can be used to analyze the spatial distribution of population, which is significant for research on geography, resources, environment, ecology, disasters, sustainable development, etc. Based on the demographic data of the 31 provinces, municipalities and autonomous regions publicized by the National Bureau of Statistics in 2010, we correlated township administrative units with the number of population in each, while attending to spatial topological problems such as pattern spot gaps and overlaps between the spatial divisions. In total, 42122 spatial units were paired with their demographic data by using the decision tree. The average population density was calculated as the number of population per spatial unit, which was then converted into vector data for generating the dataset of township-level population density in China (2010). A validation of the data shows an overall accuracy of 99%, with a less than 10% deviation for each province or municipality. The dataset is in the TIFF format, with a total volume of 108 MB.

**Keywords:** demography; population density; spatial distribution; township-level; China

### Dataset Profile

<b>Title</b>	A dataset of township-level population density in China (2010)
<b>Data corresponding author</b>	Wang Juanle (wangjl@igsnr.ac.cn)
<b>Data authors</b>	Wang Juanle, Wang Xiaojie, Wang Mingming, Yao Jinyi, Wang Lantao, Qiu Congcong
<b>Time range</b>	2010
<b>Geographical scope</b>	Anhui, Beijing, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hainan, Hebei, Henan, Heilongjiang, Hubei, Hunan, Jilin, Jiangsu, Jiangxi, Liaoning, Inner Mongolia, Ningxia, Qinghai, Shandong, Shanxi, Shaanxi, Shanghai, Sichuan, Tianjin, Xizang, Xinjiang, Yunnan, Zhejiang, Chongqing
<b>Spatial resolution</b>	1 km
<b>Data volume</b>	108 MB
<b>Data format</b>	*.tif
<b>Data service system</b>	< <a href="http://www.sciencedb.cn/dataSet/handle/964">http://www.sciencedb.cn/dataSet/handle/964</a> >
<b>Sources of funding</b>	Strategic Priority Research Program of Chinese Academy of Sciences



	(XDA19040501); Informatization Program of the 13 <sup>th</sup> Five Year Plan of Chinese Academy of Sciences (XXH13505-07).
<b>Dataset composition</b>	The dataset is named “Township-level population data in China.zip”. It is made up of the following four subsets: (1) “2010 township-level population data in China.tif”, which is a raster image mapping out the average population density of 31 provinces in China, with a data volume of 108 MB; (2) “2010 township-level population data in China.tfw”, which stores spatial reference information, with a data volume of 1KB; (3) “2010 township-level population data in China.tif.ovr”, which is a pyramid file, with a data volume of 3.1 MB; (4) “2010 township-level population data in China.tif.aux.xml”, which stores other auxiliary information in XML format, with a data volume of 1.4 KB.