



文献 DOI:

10.11922/11-6035.csd.2021.0087.zh

数据 DOI:

10.11922/sciencedb.j00001.00332

文献分类: 地球科学

收稿日期: 2021-12-06

开放同评: 2021-12-19

录用日期: 2022-01-21

发表日期: 2022-03-30

中国城市 2020 年建成区数据集

孙洁¹, 孙中昶^{2,3*}, 郭华东^{2,3}, 王建波¹, 蒋会平³, 高建⁴

1. 江苏海洋大学, 江苏连云港 222000

2. 中国科学院空天信息创新研究院, 海南省地球观测重点实验室, 海南三亚 572000

3. 中国科学院空天信息创新研究院, 数字地球重点实验室, 北京 100094

4. 南京邮电大学, 南京 210023

摘要: 如今, 全球城市化在促进区域经济发展的同时, 也带许多环境问题, 因此城市可持续发展已经成为国内外学者的研究热点。为监测和评估联合国可持续发展土地利用效率指标, 本文采用 Sentinel 数据基于 Google Earth Engine 云平台, 提出一种快速提取不透水面算法, 并参考联合国城市建成区的定义标准, 最终得到 2020 年中国 433 个人口大于 30 万城市的标准化城市建成区数据集。选取 4065 个随机验证点建立混淆矩阵进行不透水面的精度验证, 计算得到不透水面数据的平均总体精度和 kappa 系数分别为 95.57% 和 0.91 采用回归分析的方法将本数据集与官方 2020 统计年鉴进行线性分析, 计算得到拟合优度 R^2 为 0.82。另外, 本产品与国家基础地理信息中心的 GlobeLand 30 (2020) 和北京师范大学 Urban Extents (2020) 数据进行对比, 结果表明本文数据集提取精度高, 可应用于监测与评估联合国城市土地利用效率指标, 在城市扩张分析以及环境评价等应用领域亦具有重要的使用价值。

关键词: 建成区; 不透水面; 城市可持续发展目标; Google Earth Engine

数据库 (集) 基本信息简介

数据库 (集) 名称	中国城市 2020 年建成区数据集
数据作者	孙中昶、孙洁、郭华东、蒋会平、高建、王建波
数据通信作者	孙中昶 (sunzc@aircas.ac.cn)
数据时间范围	2020 年
地理区域	北纬 3°31'00"-53°33', 东经 73°29'59.79"-135°2'30''
空间分辨率	10 m
数据量	194 MB (解压后 306 MB)
数据格式	*.shp
数据服务系统网址	http://www.doi.org/10.11922/sciencedb.j00001.00332
基金项目	海南省重点研发计划 (ZDYF2020192); 国家自然科学基金面上项目 (42171291); 中国科学院战略性先导科技专项 (A 类) (XDA19090121、XDA19030104)。
数据库 (集) 组成	数据被保存在“中国城市 2020 年建成区数据集”的压缩文件内, 文件包括用于精度验证的随机点数据和 33 个以省级行政区命名的文件夹, 省级行政区文件夹下是 433 个市级城市 2020 年建成区数据。

* 论文通信作者

孙中昶: sunzc@aircas.ac.cn

引言

城市是地表覆盖的重要组成部分，与人类生活密切相关。我国地理学家早在改革开放初便提出了城市化概念。到 2001 年，城镇化首次上升为国家发展战略被写入“十五计划”^[1]。经历漫长曲折的发展，我国城镇化发展取得了巨大的成就。截至 2021 年，根据第七次全国人口普查数据显示，我国城镇常住人口已达 90199 万人，常住人口城镇化率达 63.89%。但快速、无序的城市化也伴随着诸多的问题，加快推进城市化带来了如资源短缺、生态失衡、水源污染、城市热岛效应等环境问题^[1]。因此，城市可持续发展成为提升城镇发展质量的必经之路。

为了衡量、监测和报告城市可持续发展，2015 年联合国通过了《2030 年可持续发展议程》^[2]，提出 17 项可持续发展目标（Sustainable Development Goals, SDGs）。其中，SDG11“建设包容、安全、有抵御灾害能力和可持续的城市和人类住区”^[3]对于实现所有可持续发展目标至关重要。城市扩张是关系到城市可持续发展的重要一环，对城市人口增长和城市扩张趋势的了解，有助于决策方优化城市发展，从而最大程度地减少城市化对城市环境的不利影响。可持续城市发展需要有效监测城市扩张并描绘土地消耗与人口增长之间的关系，因此联合国提出 SDG11.3.1 指标——土地消耗率与人口增长率之比^[3]。尽管有明确的评价指标，但数据缺失是度量和监测该指标面临的主要问题。因此，本研究参照联合国标准，获取全球尺度精确的城市空间分布信息，以解决该指标数据缺失问题为目标，进而开展指标扩展和监测方法模型改进以便更全面、有效衡量城市土地利用效率。

建成区是指城市行政区域范围内主要由建筑结构密集覆盖的土地表面，包括公共设施区域，这些建筑结构旨在或用于人类、动物、事物的庇护所或经济活动。中央城市区域附近的密集村镇不包括在建成区范围内。总结国内外提取城市建成区方法，主要包括：

（1）人工勾绘提取。光学遥感影像作为底图，借助地理信息处理软件，通过目视解译判读城市建成区边界。对于边界清晰的小城市，能够快速准确地勾勒；但对于边界复杂的超大城市，提取工作量巨大并且产品质量的好坏依赖于操作者经验丰富与否。

（2）基于遥感技术提取。自 20 世纪 70 年代，遥感数据就开始应用于提取城市空间分布信息研究。由于遥感具有大范围定期观测地表的能力，从遥感影像中提取的城市数据能够准确地反映城市地表信息，这一特性使城市遥感快速发展。在过去的 20 年里，许多基于遥感影像的区域和全球尺度的城市制图产品已经发布。早期城市土地产品分辨率较低，如全球城市范围 MODIS 1 km 地图（MOD1K）^[4]、2000 年全球土地覆盖（GLC00）^[5]。中低空间分辨率的遥感影像中存在大量的混合像元现象和异物同谱现象^[6]。因此，此类产品多难以区分异物同谱的土地类型，如沥青道路与干燥的土壤。近年来，随着一系列中高分辨率遥感卫星的发射并运行，以及类似 Google Earth Engine（GEE）云计算平台的兴起，研究人员也开发出更高分辨率的城市土地产品，如分辨率为 30 米的全球土地覆盖产品（GlobeLand30）^[7]、分辨率为 30 米的全球不透水地图^[8]、分辨率为 38 米的全球人类住区层（GHSL）等^[9]。全球城市地区地图产品的大部分数据源主要使用光学遥感图像或夜间灯光数据^[10]。然而目前使用的数据源依然解决不了一些问题。如低纬度地区多云雨雾天气，很难获取质量较好的光学影像，这就为城市动态遥感监测带来了很大的困难^[11]。

与光学数据相比，SAR 具有全天时、全天候、穿透性强、效率高等特点，可以解决这类问题。另外，SAR 数据在区分建筑物与其他种类地物时十分敏感，因此，SAR 数据可以提供有关城市区域的有用信息。德国航空航天中心（DLR）发布的全球城市足迹产品（GUF）是有史以来第一个使

用合成孔径雷达（SAR）图像的全球城市范围产品^[10]。但 SAR 数据由于城乡特点，尤其是在山区容易产生阴影和叠掩现象，严重影响了城市制图产品提取与精确制图。因此，研究人员提出多源数据融合的解决方案，该方案被认为是城市制图产品提取的一个有效手段。基于多源遥感数据开展高分辨率高精度全球城市制图在为联合国 SDG 城市可持续发展提供空间数据与决策支持方面具有重要应用前景。

（3）基于大数据与人工智能数据提取。近年来，随着大数据概念的提出，基于大数据的城市建成区研究兴起，包括电子地图兴趣点（POI）、出租车轨迹、道路网络数据等大数据可用于提取城市建成区。许泽宁^[12]等提出一种新的基于电子地图兴趣点的城市建成区边界识别方法，并提取 2008 年中国共计 523 个城市建成区边界。POI 数据在建成区边界附近，POI 密度显著降低，这类提取方法以此为依据，但问题也十分明显。由于城市规模、城市结构的不同，城市内部 POI 密度分布波动性较大，导致识别结果差异性较大，城乡边界，兴趣点也明显变少，边界提取精度差，因此该方法适用于 POI 数据完善的地级城市，而地级市以下的城市提取效果一般。

本研究基于 Sentinel 数据，自主生产 2020 年中国城市不透水面数据，并参考联合国城市建成区相关标准，生产了中国 433 个人口大于 30 万的城市建成区标准化数据集。本文数据是基于遥感技术提取的全球首个参考联合国建成区标准制定的中国城市数据集，并且采用高精度遥感数据源制作，数据产品精度高，满足当前城市可持续发展研究要求，在城市扩张分析以及环境评价等应用领域能发挥出更大的使用价值。

1 数据采集和处理方法

1.1 数据与预处理

表 1 为源数据详细信息，本数据集使用的主要源数据有：

表 1 研究数据详细情况

Table 1 Data Source

数据集	空间分辨率	时间覆盖	空间覆盖	数据类型
Sentinel-2 L1C 级	10 m/20 m	2015-06-23 至 2021-12-26	全球	栅格（tif）
Sentinel-1 SAR GRD	10 m	2014-10-03 至 2021-12-25	全球	栅格（tif）
哥白尼全球土地覆盖层数据，CGLS-LC100	100 m	2015-01-01 至 2019-12-31	全球	栅格（tif）
2015 年建成区	/	2015 年	中国	矢量（shp）
城市范围多边形和行政区域界线数据	/	/	中国	矢量（shp）
全球行政区域数据库	/	/	全球	矢量（shp）

（1）高分辨率遥感影像数据包括来自谷歌地球引擎（GEE）平台的线上数据 Sentinel-2 MSI L1C 产品、Sentinel-1 SAR GRD 以及哥白尼全球土地覆盖层数据（Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3, CGLS-LC100, <https://zenodo.org/record/3939050#.YdX0M2hBzb0>）^[13]；Sentinel-2A 光学影像调用时间范围为 2020 年前后各半年，合计 1 年，筛选云像素百分比（CLOUDY_PIXEL_PERCENTAGE）为 20 以下的影像。

(2) 城市基础地理信息数据包括城市范围多边形和行政区划界线数据、全球行政区域数据库 3.6 版 (GADMv3.6, <http://www.gadm.org/>) 等。

预处理分为两部分, 一是对 Sentinel-2 光学影像的去云处理, 波段 Q60 包含不透明的云 (Opaque clouds) 和卷云 (Cirrus clouds) 信息, 对应到去云处理的过程有两部分, 即云掩膜和卷云掩膜, 最终将合并的掩膜应用于光学影像中, 达到去云效果。其次, 对 Sentinel-1 原始数据进行转换, 将一级地面距离探测 (GRD) 场景处理为后向散射系数 (σ^0) [11]。

1.2 不透水面提取方法

本研究采用 Sentinel-1、Sentinel-2 数据, 精细提取不透水面, 提取流程如图 1 所示。步骤如下:

(1) Sentinel-1 数据获取城市不透水面初步提取结果。由于建筑物材质、结构的特性, 一般在图像上的后向散射回波较强, 部分区域形成白色线性亮高斑, 与其他种类地物区别明显, 因此可用来获取不透水面初步提取结果 [14]。本文选用宽幅、C 波段的 VV 极化 (垂直发送和垂直接收) 数据, 对 SAR 数据做后向散射处理获取后向散射系数, 并对多幅影像做均值处理, 减少噪声的干扰, 利用直方图双峰法 (详见朱 [11]) 确定参数文件, 最终逻辑运算得到单个图层。

(2) Sentinel-2 光学影像提取 3 种指数信息。归一化差值植被指数 (NDVI)、归一化差分水体指数 (NDWI) 和归一化差值建筑用地指数 (NDBI) 3 种指数, 将时间范围内的多幅影像图归一化制成多幅 NDBI、NDWI、NDVI, 并做均值处理, 随之得到 3 个单图层图像 NDBI_mean、NDWI_mean、NDVI_mean。

归一化差值建筑用地指数 (NDBI) 是使用短波红外波段和近红外波段来突出建筑用地的比值型指数。

$$NDBI = \frac{\rho_{SWIR,i} - \rho_{NIR,i}}{\rho_{SWIR,i} + \rho_{NIR,i}} \quad (1)$$

式 (1) 中: $\rho_{SWIR,i}$ 和 $\rho_{NIR,i}$ 分别代表短波红外波段和近红外波段的大气表面反射率 (Top of Atmosphere Reflectance, TOA)。

归一化差值植被指数 (NDVI) 旨在量化植被强度, 该指数对多光谱栅格数据集中两个波段的特征进行对比, 即红光波段中叶绿素的色素吸收率和近红外波段中植物体的高反射率。选用 NDVI 均值以提高植被识别范围, 剔除城市植被区域。

$$NDVI = \frac{\rho_{NIR,i} - \rho_{Red,i}}{\rho_{NIR,i} + \rho_{Red,i}} \quad (2)$$

式 (2) 中, $\rho_{NIR,i}$ 是 i 时间的近红外波段的 TOA, $\rho_{Red,i}$ 是 i 时间的红色波段的 TOA。

归一化水体指数 NDWI 是于监测与水体相关的含水量变化的一种指数, 本研究采用绿波段和近红外波段来制作 NDWI 以达到大面积提取影像中的水体信息的目的, 将城市内水体剔除。

$$NDWI = \frac{\rho_{GREEN,i} - \rho_{NIR,i}}{\rho_{GREEN,i} + \rho_{NIR,i}} \quad (3)$$

式 (3) 中: $\rho_{GREEN,i}$ 和 $\rho_{NIR,i}$ 分别代表绿光波段和中红外波段的 TOA。

在水体中, 桥梁、钻井平台和船舶等人造基础设施和设备的存在往往会被归一化差值建筑指数 NDBI 识别出来而成为不透水面的一部分, 但此类人造基础设施和设备并不具备成为不透水面的特点。这种与城市地区人造表面产生的混淆, 需要准确地地区分开来。如何准确地将船舶与城市陆地区域的人造地表分开是研究中的一个重要步骤。由于大多数船只都在水体 (河流、海洋等) 中, 掩蔽这些水域可以部分解决这一问题。因此引入 NDWI 的平均值, 即 NDWI_mean, 制作出一个包含大

量船舶的水体表面，以过滤掉这部分人造基础设施^[15-16]。

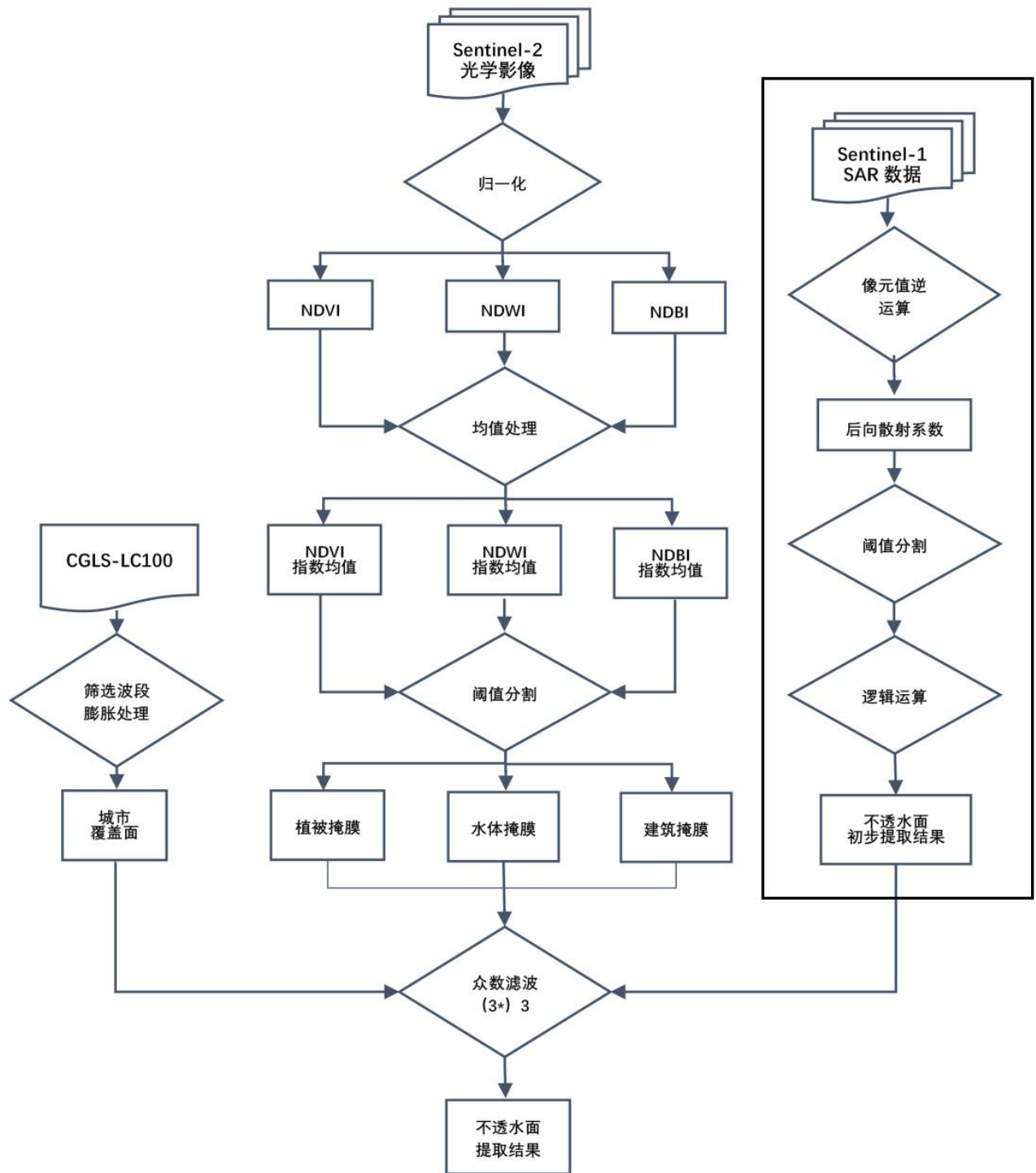


图 1 城市不透水面制作流程图

Figure 1 The flow chart for mapping urban impervious surfaces

(3) 3 种指数掩膜处理。利用直方图双峰法确定 3 种指数计算结果的阈值分割参数，通过分别对 NDBI_mean、NDWI_mean、NDVI_mean 应用阈值，获得建筑、水体、植被的掩膜。

(4) 白尼全球土地覆盖层数据 (CGLS-LC100) 中的 “urban-coverfraction” 数据，作为低分辨率的参考数据，结合 Sentinel-1 数据获取的城市不透水面初步提取结果，生成样本数据。利用建筑、水体、植被掩膜，剔除水体、植被区域，增加建筑区域，最终生成目标不透水面。

(6) 城市不透水面的边界是参考城市范围多边形和行政区域划定的。

1.3 建成区转换

目前,我国城市建成区的划分标准不一,许多城市都有自己的计算方法,影响全球城市可持续发展监测,不利于联合国《2030 年可持续发展议程》实施。因此,本文建成区严格执行联合国标准。标准要求如表 2 所示,对应各个标准,本文制定出转换方法。转换过程中,通过删除、合并、填充、增补等操作,以专家判读和人工干预辅助,对照 Google Earth 高分辨率影像,将城市不透水面转换为建成区。这些标准包括:

表 2 不透水面转换建成区

Table 2 Conversion of impervious surface into built-up area

	标准	方法
标准 1	最小城市土地面积和城市之间的距离	确保城市土地见,面积 ≥ 200 平方米,距离 ≤ 200 米
标准 2	城市土地的最小制图单位	消除细碎图斑,填补主城区内部孔洞
标准 3	功能关系密切城郊不透水区域视为建成区一部分	保留功能上密切相关的城郊区域
标准 4	城市内部公用的透水面区域被视为建成区的一部分	人工补充开放的公共透水面区域,如公园、绿地等

(1) 最小城市土地面积和城市土地之间的距离。根据联合国的建议,相距不到 200 米的地块面积在 20 公顷以上的城市土地区域,应被视为建成区的一部分。这一点,将所有符合标准的不透水面区域连接起来,形成一个连续的城市。

(2) 城市土地的最小制图单位。采用中国国家地理信息中心的标准,根据空间分辨率为 10×10 m 的遥感影像,人造地表面积应占据 8×8 像素以上。因此,城市不透水面栅格数据转成矢量数据后,消除面积小于 6400 平方米的细碎斑块,并填充了位于主要城市区域内面积小于 64600 平方米的孔洞,以保持空间格局的连续分布。

(3) 具有密切功能关系的城郊不透水区域应视为建成区的一部分。一些独立的农村居住区位于主城区周边或与主城区间隔一定距离,但它们可能通过连接良好的道路,在功能上依赖主城区提供的就业和服务。因此,在转换过程中,保留这一部分的斑块。

(4) 用作公共空间的透水区域应视为建成区的一部分。这部分区域,需要结合专家判读和人工干预,在 Google Earth 高分辨率图像的帮助下,手动补充和调整公园和绿地的斑块到主城区^[13-14]。

具体操作为:首先对不透水面栅格数据裁剪处理,裁剪数据为中国城市范围多边形和行政区域界线数据,然后将栅格数据转为矢量数据,定义投影坐标系后进行几何计算,填充连片区域的孔洞,连片区域的标准是面积大于等于 200000 平方米的区域,将该区域内的孔洞填充,重新计算图斑面积,再根据填充后的图斑面积,筛选出面积大于 6400 平方米的图斑,达到去除最小图斑的目的。接下来对之前获得的消除零碎图斑后的数据进一步操作,按照图斑间距离,将图斑分为两类,图斑间距离小于等于 200 米的部分直接提取备用,图斑间距离大于 200 米的部分,再选出面积大于 200000 平方米的图斑导出,与之前图斑间距离小于等于 200 米的部分合并,得到的数据是城市建成区初步结果,通过专家判读和人工修正,补充调整城市内部公园、绿地斑块。

2 数据样本描述

本数据主要包含 2018 版《世界城市化展望》列出的 433 个常住人口大于 30 万中国城市建成区数据。数据被保存在“中国城市 2020 年建成区数据集”的压缩文件内，文件大小总计 194 MB，解压后文件大小为 306 MB。文件包括精度验证随机点数据和 33 个以省级行政区命名的文件夹，省级行政区文件夹下是市级城市建成区数据，数据格式均为 shp，均采用 UTM 投影、WGS84 坐标系。

3 数据质量控制和评估

本文精度验证过程基于 QGIS 图像加载 Google 卫星地图，生成 4065 个随机参考点，建立混淆矩阵，根据统一的判读规则，进行目视判读，以判断表面是否为不透水面。标准精度评估措施，即总体精度（OA）、生产者精度（PA），计算用户准确度（UA）来评估提取的准确度。

表 3 为 2020 年 433 中国城市不透水面分布数据精度验证结果，显示平均 OA、PA 和 UA 分别为 95.57%、94.13%、95.45%。证明了融合光学和 SAR 数据对大面积城市土地提取方法的有效性。

表 3 2020 年 433 中国城市不透水面分布数据精度验证结果

Table 3 Accuracy verification of impervious surface data

	不透水面	透水面	总计	UA (%)
不透水面	1635	78	1713	95.45
透水面	102	2247	2349	95.66
总计	1737	2325	4062	/
PA (%)	94.13	96.65	/	/
总体精度 OA (%)	95.57			
Kappa 系数 OK	0.91			

选取我国最新发布的 2020 年城市建设统计年鉴，统计年鉴中只列出与本文数据相对应的 418 个城市（人口 30 万或以上）作为参考，因此将本数据和统计年鉴数据之间进行回归分析，建立线性回归方程，如图 2 所示。

统计年鉴数据采用逐级上报的方式进行统计，我国官方对建成区范围的定义，一般指建成区外轮廓线所能包括的地区，也就是这个城市实际建设用地所达到的范围。与本文建成区定义存在区别的是，官方定义以“成片开发建设、市政公用设施和公共设施基本具备的城市区域”为标准，类似于核心类型、一城多镇类型城市，建成区范围包括集中连片的部分以及分散的若干个已经成片建设起来的区域。本文建成区转换步骤中，与主城区间距大于 200 米的图斑，只有面积达到标准要求才能予以保留。

回归分析中，统计数据与本文数据差异较大的城市有佛山市、临沂市、淄博市，均为本数据计算面积大于统计年鉴数据。分析可知，临沂市和淄博市差异较大的主要原因是，这些城市为核心类型城市，周边村镇与中心城区有一定距离间隔，两种不同方法计算得到的结果会产生较大差异。佛山近年来依靠地域优势飞速发展，与广州市地域连接，两个城市边界模糊，因此两种数据差异较大。

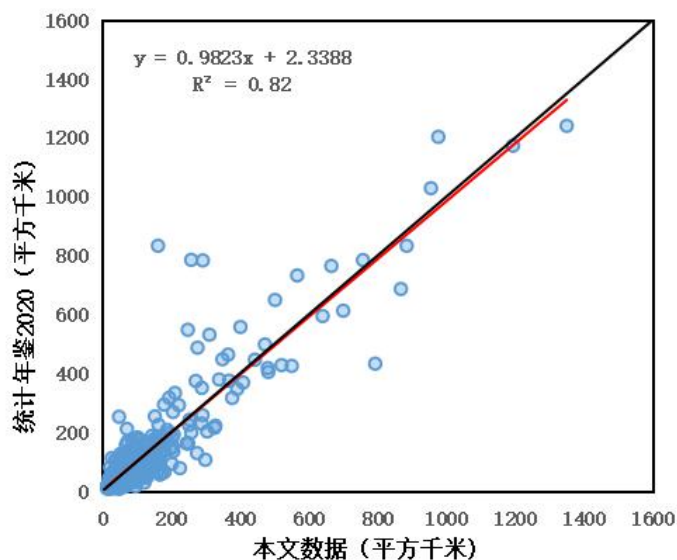
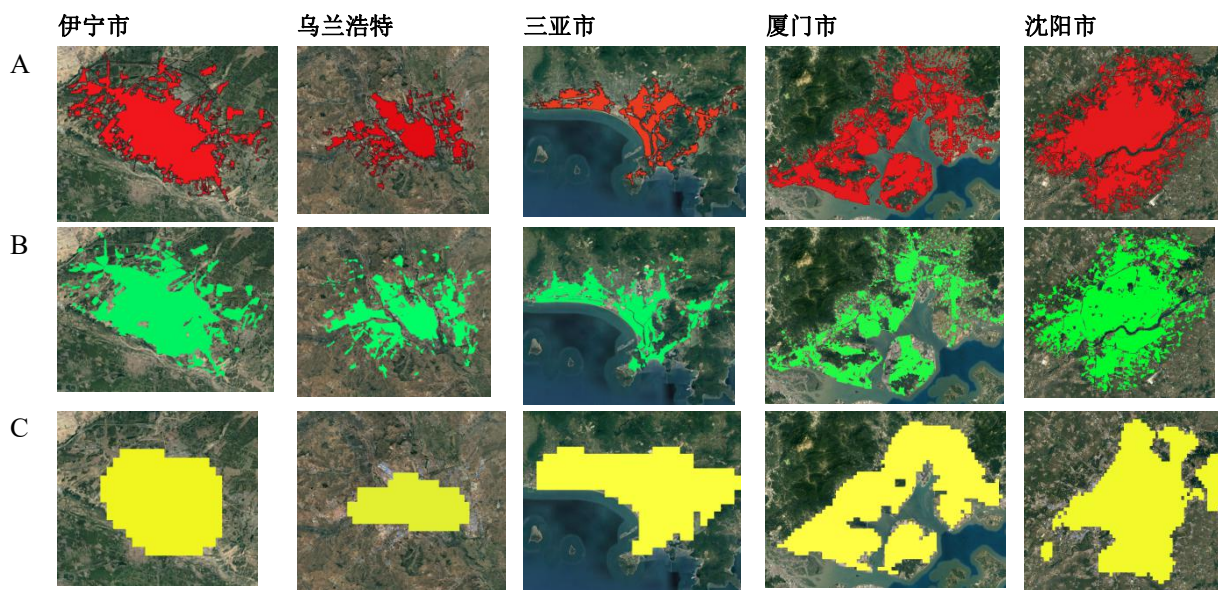


图 2 2020 中国建成区数据与 2020 年统计年鉴线性回归方程图

Figure 2 Linear regression equation between 2020 China's built-up area data and 2020 statistical yearbook



A 为本文数据；B 为 GlobeLand30 数据；C 为 Urban Extents2020 数据

图 3 提取结果与 GlobeLand30 产品细节对比图

A is the data of this paper; B is globeland30 data; C is urban extenss2020 data

Figure 3 Comparison between extraction results and product details of globeland30

本文选取中国伊宁市、乌兰浩特市、三亚市、厦门市和沈阳市，将本数据集与 GlobeLand30 2020^[17]、Urban Extents 2020^[18]进行对比。GlobeLand 30 关于人造地表的定义是：由人工建造活动形成的地表，包括城镇等各类居民地、工矿、交通设施等，但不包括建设用地内部连片绿地和水体^[19]，由图 3 中 B 可看出，GlobeLand 30 产品城市边界和内部细节提取较差，城市内部存在较多由于城市绿化、公共用地未提取而产生的孔洞，城市周边存在由于零散村庄提取产生的细碎图斑，有过估、

漏估现象，尤其是厦门市，显著可见的少了很大一部分。图 3 中 C 为北京师范大学制作的 Urban Extents2020 数据，整体显示为一块较为完整区域，由于数据源是空间分辨率较低的夜间灯光数据，因此 Urban Extents2020 分辨率也明显较低，该数据只能粗略展示城市的整体位置与形状，过估、漏估现象较为明显，无法保证细节。A 为本文数据，在较好的保留城市细节的基础上，删除掉城市外部零散的农村区域，也补充了城市内部的绿地和公园，使数据整体较为完整。图 4 为德州市、高平市、肇庆市建成区与不透水面对比图。

本研究精细提取不透水面过程中，使用 Sentinel-2 数据生成的植被、水体、建筑 3 种掩膜，以剔除掉使用 Sentinel-1 数据提取的不透水面初步提取结果中的植被、水体、建筑部分。但由于 SAR 数据对白色线性亮高斑有较高的敏感性，在提取建筑区域时有极大的优势，但同时，也能将城市中呈现为亮高斑的裸地提取出来。而在掩膜剔除的过程中，没有剔除掉这一区域。本研究对这一问题的处理方式是，在制作数据时，手动剔除目视识别城市中面积较大的白色亮高斑裸地。这一解决方法存在较大的局限性，因此，后续数据的精度提升，将以此为切入点，寻找合适的解决方法。

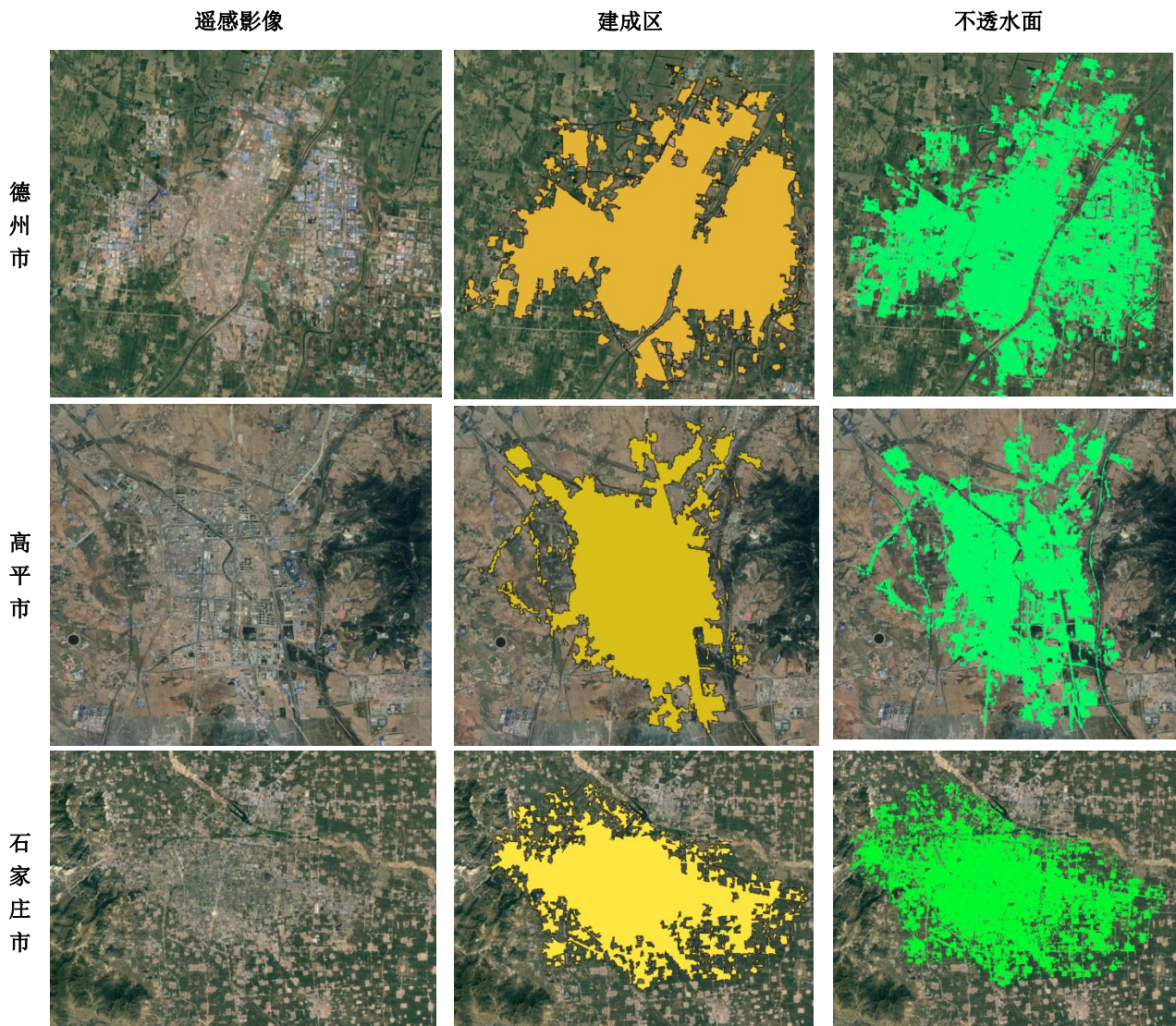


图 4 建成区与不透水面细节对比图

Figure 4 Comparison of details of built-up area and impervious surface

4 数据使用方法和建议

本数据集可用于探索城市建设用地扩张模式，验证推动城市扩张的内部和外部因素，以解决城市无节制扩张、低密度住房等问题。此外，在国家新型城镇化、区域生态保护和高质量发展的背景下，综合运用数据集中城市的时间和空间要素，在探索城市发展与生态保护，多尺度分析城市扩张对生态用地的影响、环境质量评估方面具有重大作用，如城市热环境演变研究、自然生态用地变化研究、PM 2.5 时空演变研究等。

为了监测城市可持续发展，联合国提出 SDG11.3.1 指标，本数据集解决了该指标数据缺失问题，因此，将已发布的 2015 年中国建成区数据集与本文数据集进行 SDG11.3.1 指标计算，是本数据集重要的应用之一。土地消耗率与人口增长率之比（LCRPGR）揭示了城市土地扩张与城市人口增长之间的关系^[2]。

人口增长率（PGR）：

$$PGR = \frac{\ln(Pop_{t+n}/Pop_t)}{y} \quad (4)$$

式（4）中： Pop_t 表示城市于 t 年的人口数量， Pop_{t+n} 表示城市于 $t+n$ 年的人口数量， y 表示两期年份差。

土地消耗率（LCR）：

$$LCR = \frac{\ln(Urb_{t+n}/Urb_t)}{y} \quad (5)$$

式（5）中： Urb_t 表示城市于 t 年的建成区面积， Urb_{t+n} 表示城市于 $t+n$ 年的建成区面积， y 表示两期年份差。

土地消耗率与人口增长率之间的比率（LCRPGR）：

$$LCRPGR = \frac{LCR}{PGR} \quad (6)$$

图 5 显示了各个城市 LCRPGR 的分布数情况。从图 6 可知，有 133 个城市 $1 < LCRPGR \leq 2$ ， $2 < LCRPGR \leq 4$ 的城市也占据了 132 个。其次，城市占比多的是 $0 < LCRPGR \leq 1$ ，有 90 个城市在这一阶段。整体分布较为均匀， $4 < LCRPGR$ 土地扩张率大于人口增长率很多，集中分布在一些城市群附近，如成渝城市群、京津冀城市群、长江中游城市群、珠三角城市群。仅有 4 个城市 $LCRPGR < 0$ ，处于人口流失状态。依据 JIANG^[16]发表的文章内容显示，LCRPGR 的平均值首先从 1990–1995 年的 1.34 下降到 1995–2000 年的 0.85，2000–2005 年上升至 1.82，2005–2010 年计算结果为 1.59，2010–2015 年 LCRPGR 的平均水平达到 2.15。本研究发现 2015–2020 年城市指标平均值继续上升至 2.54，增速放缓，目前我国城市整体以快速发展为主要趋势，土地扩张快于人口增长，城市化率在稳步增长。

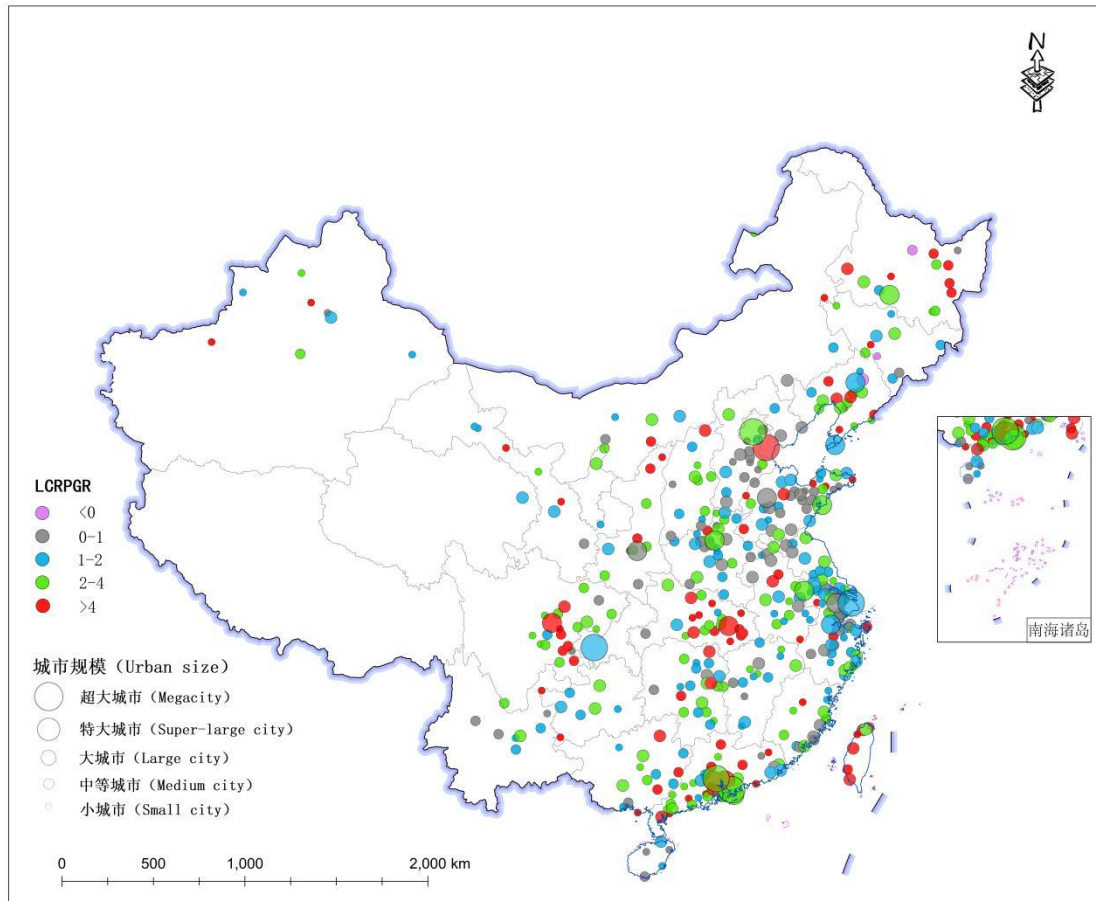


图 5 2015–2020 年中国 433 个城市市级尺度 LCRPGR 分级显示 (审图号: GS(2022)981 号)

Figure 5 LCRPGR classification of 433 cities in China from 2015 to 2020

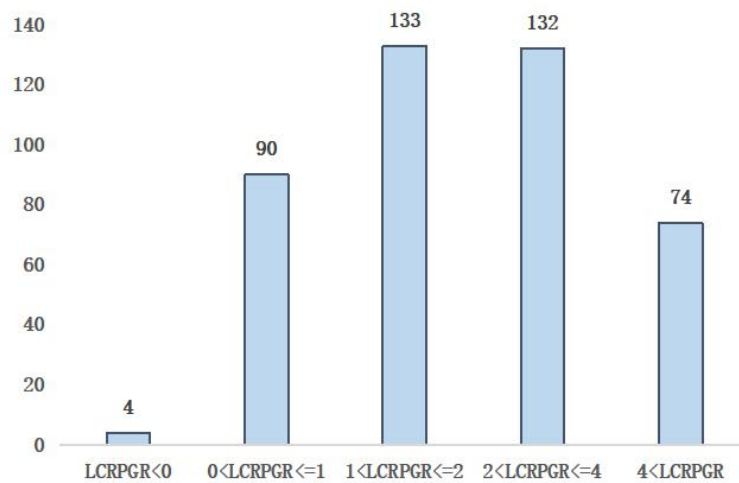


图 6 市级尺度 LCRPGR 等级分组图

Figure 6 LCRPGR level grouping diagram of municipal scale

数据作者分工职责

孙洁（1996—），女，江苏省南京市人，硕士研究生，研究方向为数字城市空间信息采集与处理。

主要承担工作：数据分析和论文撰写。

孙中昶（1983—），男，山东省济宁市人，博士，副研究员，研究方向为城市遥感、微波遥感。主

要承担工作：数据前期处理与质量控制。

郭华东（1950—），男，江苏省徐州市人，中国科学院院士，研究员，研究方向地球科学。主要承

担工作：数据质量控制。

王建波（1985—），男，山东省潍坊市人，博士，讲师，研究方向为物理大地测量研究，主要承担

工作：数据质量控制。

蒋会平（1990—），男，安徽省安庆市人，博士，助研，研究方向为城市遥感与城市可持续发展。

主要承担工作：数据处理流程的设计与质量控制。

高建（1983—），男，江苏省南京市人，博士，讲师，研究方向为物理大地测量研究，主要承担工

作：数据质量控制。

参考文献

- [1] 方创琳, 鲍超, 黄金川, 等. 中国城镇化发展的地理学贡献与责任使命[J]. 地理科学, 2018, 38(3): 321–331. DOI:10.13249/j.cnki.sgs.2018.03.001. [FANG C L, BAO C, HUANG J C, et al. Contribution, responsibility and mission of geography on China's urbanization development[J]. Scientia Geographica Sinica, 2018, 38(3): 321–331. DOI:10.13249/j.cnki.sgs.2018.03.001.]
- [2] United Nations. Transforming our world: the 2030 agenda for sustainable development[EB/OL]. (2015) [2021 - 01 - 21]. <https://sdgs.un.org/2030agenda> .
- [3] 朱秀林, 蒋会平, 孙中昶, 等. 基于 SDG11.3.1 的中国人口和土地城镇化分析与评价[J]. 国土资源情报, 2020(11): 36 - 41. DOI:10.3969/j.issn.1674-3709.2020.11.006. [ZHU X L, JIANG H P, SUN Z C, et al. Analysis and evaluation of population and land urbanization in China based on SDG 11.3.1[J]. Land and Resources Information, 2020(11): 36 - 41. DOI:10.3969/j.issn.1674-3709.2020.11.006.]
- [4] ANNEMARIE FRIEDL MARK A MCIVER DOUGLAS K WOODCOCK CURTIS E SCHNEIDER. Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data[J]. Photogrammetric Engineering & Remote Sensing, 2003, 69(12): 1377–1386. DOI: 10.14358/PERS.69.12.1377.
- [5] BARTHOLOMÉ E, BELWARD A S. GLC2000: a new approach to global land cover mapping from Earth observation data[J]. International Journal of Remote Sensing, 2005, 26(9): 1959–1977. DOI:10.1080/01431160412331291297.
- [6] FLANAGAN M, CIVCO D L. Subpixel impervious surface mapping[C]. ASPRS 2001 Annual Convention, St. Louis, MO, April 23-27, 2001.
- [7] CHEN J, CHEN J, LIAO A P, et al. Global land cover mapping at 30 m resolution: a POK-based operational approach[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 103: 7–27. DOI:10.1016/j.isprsjprs.2014.09.002.
- [8] ZHANG X, LIU L Y, WU C S, et al. Development of a global 30 m impervious surface map using

- multisource and multitemporal remote sensing datasets with the Google Earth Engine platform[J]. *Earth System Science Data*, 2020, 12(3). DOI:10.5194/essd-2019-200.
- [9] PESARESI M, EHRILCH D, FLORCZYK A, et al. GHS built-up grid, derived from Landsat, Multitemporal (1975, 1990, 2000, and 2014)[EB/OL]. (2017) [2021-1-21]. <https://publications.jrc.ec.europa.eu/repository/handle/JRC109041>.
- [10] WU F, WANG C, ZHANG H, et al. Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning[J]. *Remote Sensing of Environment*, 2021, 262: 112515. DOI: 10.1016/j.rse.2021.112515.
- [11] 朱秀林, 赵相伟, 杜文杰, 等. 基于 Sentinel-1 SAR 和 Sentinel-2A 光学影像的海南岛城市不透水面数据集[J]. *中国科学数据*, 2019, 4(2): 69–80. DOI: 10.11922/csdata.2018.0084.zh.[ZHU X L, ZHAO X W, DU W J, et al. A dataset of urban impervious surface of Hainan Island, based on Sentinel-1 SAR and Sentinel-2A optical images[J]. *China Scientific Data*, 2019, 4(2): 69–80. DOI: 10.11922/csdata.2018.0084.zh.]
- [12] 许泽宁, 高晓路. 基于电子地图兴趣点的城市建成区边界识别方法[J]. *地理学报*, 2016, 71(6): 928–939. DOI:10.11821/dlxb201606003. [XU Z N, GAO X L. A novel method for identifying the boundary of urban built-up areas with POI data[J]. *Acta Geographica Sinica*, 2016, 71(6): 928 – 939. DOI:10.11821/dlxb201606003.]
- [13] BUCHHORN M, SMETS B, BERTELS L. Copernicus Global Land Service: Land Cover 100m: Collection 3: epoch <2019>: Globe (Version V3.0.1)[EB/OL]. (2020 - 09 - 08) [2021 - 01 - 21]. <http://dx.doi.org/10.5281/zenodo.3939050>.
- [14] 芮维. 城市地区高分辨率 SAR 图像中典型地物的特征分析与解译[D]. 武汉:武汉大学, 2010. [GUO W. Feature analysis and interpretation of typical object for high resolution SAR image on urban areas[D]. Wuhan:Wuhan University, 2010.]
- [15] SUN Z C, XU R, DU W J, et al. High-resolution urban land mapping in China from sentinel 1A/2 imagery based on google earth engine[J]. *Remote Sensing*, 2019, 11(7): 752. DOI:10.3390/rs11070752.
- [16] JIANG H, SUN Z, GUO H, et al. An assessment of urbanization sustainability in China between 1990 and 2015 using land use efficiency indicators[J]. *Npj Urban Sustainability*, 2021, 34. DOI:10.1038/s42949-021-00032-y.
- [17] JUN C, JIN C, ANPING L, et al. Global land cover mapping at 30m resolution: A POK-based operational approach[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 103: DOI:7-27. 10.1016/j.isprsjprs.2014.09.002.
- [18] ZHAO M, CHENG C, ZHOU Y, et al. A global dataset of annual urban extents (1992–2020) from harmonized nighttime lights[J]. *Earth System Science Data*, 2022, 14: 517-534. DOI:10.5194/essd-14-517-2022
- [19] 陈军, 廖安平, 陈晋, 等. 全球 30m 地表覆盖遥感数据产品-GlobeLand30[J]. *地理信息世界*, 2017, 24(1): 1–8. DOI:10.3969/j.issn.1672-1586.2017.01.002. [CHEN J, LIAO A P, CHEN J, et al. 30-meter global land cover data product-GlobeLand30[J]. *Geomatics World*, 2017, 24(1): 1 – 8. DOI:10.3969/j.issn.1672-1586.2017.01.002.]

论文引用格式

孙洁, 孙中昶, 郭华东, 等. 中国城市 2020 年建成区数据集[J/OL]. *中国科学数据*, 2022, 7(1). (2022-03-

29). DOI: 10.11922/11-6035.csd.2021.0087.zh.

数据引用格式

孙洁, 孙中昶, 郭华东, 等. 中国城市 2020 年建成区数据集[DS/OL]. Science Data Bank, 2021. (2021-12-15). DOI: 10.11922/sciencedb.j00001.00332.

A dataset of built-up areas of Chinese cities in 2020

SUN Jie¹, SUN Zhongchang^{2,3*}, GUO Huadong^{2,3}, WANG Jianbo¹,
JIANG Huiping³, GAO Jian⁴

1. Jiangsu Ocean University, Lianyungang 222000, P.R. China

2. Aerospace Information Research Institute, Chinese Academy of Sciences, Hainan Laboratory of Earth Observation, Sanya 572029, P.R. China

3. Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, P.R. China

4. Nanjing University of Posts and Telecommunications, Nanjing 210023, P.R. China

*Email: sunzc@aircas.ac.cn

Abstract: Global urbanization promotes the development of regional economy; however, it also brings many environmental problems. The sustainable urban development has become a popular research topic all of the world. In order to monitor and evaluate the United Nations urban land use efficiency indicator, this study used the Sentinel data to propose an algorithm for rapidly extracting the impervious surface based on the Google Earth Engine cloud computing platform, and finally obtained a standardized urban built-up dataset of 433 cities with a population more than 300,000 in China in 2020. This study adopted the United Nations definition of urban agglomeration area. A confusion matrix of 4065 random validation points was used to conduct the accuracy validation of resultant impervious surface, and the regression analysis of this dataset with the 2020 official Statistical Yearbook showed that the average overall accuracy and kappa coefficient was 95.57% and 0.91, respectively, and the goodness of fit of R^2 was 0.82. In addition, we compared this product with the GlobeLand30 (2020) and Beijing Normal University Urban Extents (2020) data. The results showed that our dataset had high extraction accuracy and could be used for monitoring and evaluation of the United Nations urban land use efficiency indicator. It also has significant value in many applications such as urban expansion analysis and environmental evaluation.

Keywords: built-up area; impervious surface; urban sustainable development goals; Google Earth Engine

Dataset Profile

Title	A dataset of built-up areas of Chinese cities in 2020
Data corresponding author	SUN Zhongchang (sunzc@aircas.ac.cn)
Data authors	SUN Zhongchang, SUN Jie, GUO Huadong, JIANG Huiping, GAO Jian, WANG Jianbo
Time range	A specific time period (e.g., 2010; 2000 – 2013)

Geographical scope	Longitude & latitude; geographical scope (e.g., 40 ° 01'52''N – 23 ° 11'59''N, 105 ° 43'45''E – 61 ° 28'45''E; specific areas include: Tibetan Plateau, Hengduan Mountains, Himalaya Mountains, Hindu Kush Mountains and Pamirs)
Spatial resolution	10 m
Data volume	194 MB (306 MB after decompression)
Data format	*.shp
Data service system	< http://www.doi.org/10.11922/sciencedb.j00001.00332 >
Source of funding	Hainan Key R & D and Development Plan (ZDYF2020192); National Natural Science Foundation of China (42171291); Strategic Pilot Science and Technology Special Project of Chinese Academy of Sciences (Class A) (XDA19090121, XDA19030104).
Dataset composition	This dataset includes distribution map data in China in 2020, and all numbers are compressed into a file. The file contains SHP files from 433 cities.