

中国再增 2666 个传统村落空间分布数据集

余亮*, 唐铭婕, 付蒙, 刘智涛, 邱雨晨, 曹玲莉, 杨旭辉, 沈军鑫

苏州大学建筑学院, 苏州 215123

摘要: 继前四批共计 4,153 个中国传统村落公布后, 国家住房和城乡建设部、文化部、财政部于 2019 年 6 月又公布了第五批 2,666 个传统村落名录, 至此收录的中国传统村落总数有 6,819 个。本数据集对再增的 2,666 个传统村落进行定位, 空间坐标提取参考百度地图和 Google Earth 影像图, 以村落名称和村落图形的几何中心为依据, 并以古建筑屋顶影像为辅助判断材料, 提取村落的空间位置。对于地图和影像上缺少村落信息的情况, 则定位上移至其镇级行政中心。以此, 获得中国再增 2,666 个传统村落空间分布数据集。数据集存储为.shp 和.kmz 格式, 由 6 个数据文件组成, 数据量为 7.48 MB (压缩为 2 个文件, 362 KB)。

关键词: 中国; 传统村落; 名录; 空间分布; 第五批

DOI: <https://doi.org/10.3974/geodp.2022.01.03>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2022.01.03>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2020.03.22.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2020.03.22.V1>.

1 前言

传统村落遍布世界各地, 是宝贵的历史文化遗产, 呈现着人类诞生至今的聚合与居住形式, 他们因地域不同而特色各异。因长期的生活和活动之需, 传统村落 (以下简称村落) 中穿插有大量居住的传统民居, 民居和村落形成了不能分离的相互依存关系。作者十年多时间内, 实地考察调查过大江南北的几十个村落, 为村落和民居鲜活的地域特色和能工巧匠的精湛技艺叹服的同时, 也为一些村落和民居未能及时整治和修复而倒塌感到惋惜。在国家大力倡导“绿水青山就是金山银山”和乡村振兴的今天, 及时持续的保护修复工作非常必要 (图 1)。

《中国再增 2666 个传统村落空间分布数据集》是继前两次 (以后均此称), 即《中国 2555 个村落空间分布数据集》和《中国 1598 个村落空间分布数据集》^[1,2]的数据再整理和发布, 利用该数据源不仅能方便辨认解读基层村落及以上, 即镇、乡到县等不同

收稿日期: 2021-02-26; 修订日期: 2021-10-11; 出版日期: 2022-03-25

基金项目: 国家自然科学基金 (41371173)

*通讯作者: 余亮 F-8099-2018, 苏州大学建筑学院, yuliang_163cn@163.com

数据引用方式: [1] 余亮, 唐铭婕, 付蒙等. 中国再增 2666 个传统村落空间分布数据集[J]. 全球变化数据学报, 2022, 6(1): 19–24. <https://doi.org/10.3974/geodp.2022.01.03>. <https://cstr.escience.org.cn/CSTR:20146.14.2022.01.03>.

[2] 余亮, 唐铭婕, 付蒙等. 中国再增 2666 个传统村落空间分布数据集[J/DB/OL]. 全球变化科学仓储电子杂志, 2021. <https://doi.org/10.3974/geodb.2020.03.22.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2020.03.22.V1>.

层级的行政分划及隶属关系，还能比较和确认对应的村落地理位置并分析空间分布特点，以较完整地展现优秀的乡村文化，激发传统村落的精神物质活力。



图1 第五批传统村落：民居保护和修复（余亮，摄于2017（左）和2020（右）年）
（左：浙江绍兴市越城区东浦街道东浦村；右：浙江金华市义乌市佛堂镇倍磊村）

2 数据集元数据简介

《中国再增 2666 个传统村落空间分布数据集》^[3]的名称、作者、地理区域、数据年代、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

表 1 《中国再增 2666 个传统村落空间分布数据集》元数据简表

条 目	描 述
数据集名称	中国再增 2666 个传统村落空间分布数据集
数据集短名	VillagesChina2666
作者信息	余亮 F-8099-2018, 苏州大学建筑学院, yuliang_163cn@163.com 唐铭婕 O-6467-2018, 苏州大学建筑学院, 361988267@qq.com 付蒙 O-6455-2018, 苏州大学建筑学院, 821064405@qq.com 刘智涛 ABH-4639-2020, 苏州大学建筑学院, 785025073@qq.com 邱雨晨 ABH-5207-2020, 苏州大学建筑学院, 375284315@qq.com 曹玲莉 ABI-1416-2020, 苏州大学建筑学院, 1083748619@qq.com 杨旭辉 ABH-6245-2020, 苏州大学建筑学院, 654712015 @qq.com 沈军鑫 ABH-7152-2020, 同济大学浙江学院, 409209740 @qq.com
地理区域	中国, 覆盖 30 个省级行政区（港澳台无数据）
数据年代	2012–2019
数据量	4.50 MB
数据集组成	2 个文件（VillagesChina2666.kmz + VillagesChina2666.rar）
基金项目	国家自然科学基金（41371173）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《全球变化数据学报（中英文）》发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[4]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

3 村落空间点数据提取

村落名录不显示空间位置，是一行有行政隶属关系且显示地名的名称文字，村落空间点数据定位建立在文字名称信息的提取基础上，通过比较辨认村落的空间形态特征而被定位获取。空间认知上人们集聚居住的村落，主要构成要素是民居建筑，除此还有田地、河流和道路等地物，民居不仅空间上的屋顶水平投影面大，而且还不同于其他物体特征，如自然要素的山体、植物和水面，以及人工要素的道路、桥梁等，点数据定位正是利用了屋顶与其他物体有显著差异、图形规整易辨认的形体特点。

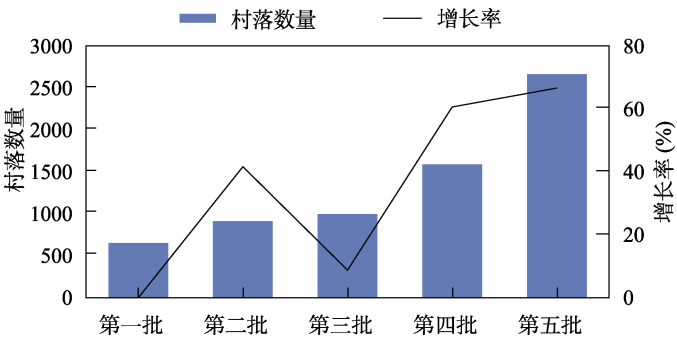


图 2 第一批至第五批村落数量变化

3.1 原始数据

村落原始数据：第五批中国传统村落名录的村落数量（最终公布）：2,666 个，与最初公示数量相比，新增村落 22 个，删去 2 个^[5]。数据集范围：覆盖中国 30 个省级行政区（不包括港澳台），加之已公布的四批全国传统村落，村落名录总体数量达到 6,819 个，递增率明显，特别是第四批至第五批的增长从 60% 达到 65%（图 4）。

3.2 村落点数据和名称

村落空间点坐标（经纬度）数据获取与前两次基本相似^[1,2]，有两个步骤，一是点坐标定位，需要分辨空间上村落与自然或人工地物的差异，聚焦空间中突起的较为规整、体量较小的个体；二是村落名称的整理，目的是为了形成规则的属性表，采用半自动结合手动的方法，半自动指的是应用 Excel 相关提高效率的工具和方法，应用 left、right、mid 函数以及高级分列等工具，在村落名称上直接辨认分段；文字的转拼音则应用 Excel 工具中的宏，在 Visual Basic 编辑器中插入新模块，输入一定代码并定义函数 getpy 和 proper 函数后对中文进行转换和首字母大小写处理，这些方法与前两次明显不同。而手动则针对非常见名称，如自治州、街道、社区、屯、寨的列段等，在半自动处理后再进行一定的手动归类，详见 3.3 节。

村落的空间点数据定位，输入至百度地图并获得该村落的点位置后，导入 Google Earth 地图，并以手动目测法在线上调整空间点位置的坐标差异，判别该点是否落在村落古建筑屋顶，且落在代表该村落空间特性的几何中心位置上。每个点数据的定位并非显而易见，手动调整需多次。包括前两次在内，定位的主要问题是判别不易，百度地图导入至 Google Earth 地图后容易改变原来的位置，如定位在屋顶上的点会移位至田野或水塘等场所，而随

着第五批数据量比第四批明显增加，调整工作量也显著增大。此外，对部分辨认困难和线索缺乏的名称，在查询相关资料或直接联系该村确认等方法的基础上，如确实难以定位的村落，则定在其上一级行政单位^[1]。

3.3 村落名称处理

村落为基层村级组织，名录的一连串文字名称是获取村落位置的直接线索，特别是处在尾部的村落标记，即村落词尾的标记称法多而丰富，并不统一，可能这与中国多民族和多气候类型等特点有关^[6]。处理方法主要是属性表上各列段行政层级和隶属关系的确定，列段从省、市、县、到镇和村，设5个行政层级，由于一般省市行政的5级表示与直辖市的4级不同，数据集采用5级公约数列段，设置1个直辖市和一般市级的重复列段，如“北京市房山区佛子庄乡黑龙关村”中的“北京市”列段就重复了第1级和第2级。

其次，观察名称从省到村的行政序列可知，5级行政隶属关系后不少村落有6级词尾标注，反映出有多个空间位置需分辨定位，这里根据屋顶影像，能确认在6级则6级，如不明显则在5级。村落名称中最多的“村”标记，但未必以“村”为词尾，第五批5级村落中有2,598个“村”词尾（表2），其他依次是“社区”的34个、“村村”的27个、“老街”的3个、“屯”的3个、“嘎查”（相当行政村）的1个，没有“居委会和村委会”标注。在5级和6级的关系上，一般是行政村下与自然村等相连，如“广西（短缩）桂林市恭城瑶族自治县莲花镇竹山村委红岩老村屯”中，“竹山村委”是5级村架构，“红岩老村屯”则为6级自然村，该定位在“红岩老村屯”上，如推测难则定在上一级“竹山村委”上。一般地一个行政村大多对应一个自然村，也有多个对应，如“广西壮族自治区来宾市金秀瑶族自治县六巷乡六巷村六巷屯、朗冲屯、上古陈屯”的3个自然村。第五批村落中的6级标注有401个（占2,666个的15.04%），后缀为“村、组、寨、屯、寨村、沟、坡、街、庄和湾”，最多的还是“村”，达262个，接着是“组”的45和“寨”的43个，“庄”和“湾”的各1个最少，整体上呈形式多样、组合字数不等外，名称上也较形象地反映了与当地自然风貌结合的特点。

表2 村5级和6级以及“古”、“老”的词尾标注(数量：个)

No	第5级	数量	第6级	数量	组合“古”、“老”的词尾	数量
1	村	2,598	村	262	古村	6
2	社区	34	组	45	古寨	3
3	村村	27	寨	43	古寨村	1
4	老街	3	寨村	10	老街	3
5	屯	3	屯	29	老街村	3
6	嘎查	1	沟	4		
7	居委或居委会	0	坡	4		
8	村委或村委会	0	街	2		
9			庄	1		
10			湾	1		
合计		2,666		401		16

此外，有与“古”和“老”组合衍生的“古村、古寨、老街”等词尾，5级和6级均

有，在此无意关注词尾“古村”和“老街”的名称何时使用？或通过与“古”“老”的组合有唤起“传统”的意图？但至少空间定位时多了引起注意的线索。第五批村落定位时未见与“古”“老”组合后的空间差异，类似“曲靖市罗平县钟山乡普理村白古村”的“古”，如不作较深的调研，确难分辨是名称还是“古”有“传统”含义。村落中有“古村”6 个、“古寨”3 个，除此还有 3 个“老街”标注，其中“老街”的“安徽省合肥市巢湖市”的“柘皋镇北闸老街”和“烔炀镇烔炀老街”，尽管是村级，经调研现场可知，空间上村和镇存在着紧密的依存关系，镇发展有村的踪影，村的壮大为镇发展提供基础，空间上往往有合二为一的特点，这些空间现象可通过遥感影像获得推测支撑并在定位时应用。一般而言，镇比村具有更多的商业和行政功能，空间上有的村被现在新镇“包围”，有的为邻，经年代更迭，不少村非昔日状态，构架和边界、机理等发生了很大变化，如第 5 批村落的“安徽省六安市金安区毛坦厂镇浸堰村”（图 3），村在镇的周边，当地有人将镇老街说成是村的老建筑，说明了两者空间关系的接近。一般的空间上如明确是村则定位在村，反之定位在镇中老街或老建筑上，类似特点也反映在安徽省合肥市巢湖市烔炀镇烔炀老街和安徽省合肥市巢湖市柘皋镇北闸老街上。村 6 级词尾（共 401 个）的空间定位点示于表 3，从中可知最多的是 6 级，其他还有 5 级和 4 级，符合本文前面介绍的如当级定位困难，则定在上一级行政单位的方法。



图 3 镇和村的发展关系（余亮，摄于 2020 年）
（第 5 批，安徽省六安市金安区毛坦厂镇浸堰村，左：浸堰村；右：毛坦厂镇老街）

表 3 村 6 级词尾的空间定位点（数量：个）

6 级词尾 定位点	村	组	寨	寨村	屯	沟	坡	街	庄	湾	合计
6 级	196	11	18	6	15	1	1	1	1	1	253
5 级	58	28	23	4	12	1	3	1	0	0	130
4 级（镇）	5	5	2	0	1	0	0	0	0	0	9
4 级（乡）	3	5	0	0	1	0	0	0	0	0	9
计	262	45	43	10	29	4	4	2	1	1	401

4 数据结果和讨论

数据集由 2 个子集构成，分别是中国再增 2,666 个传统村落空间分布点数据的 ArcGIS

的.shp 版以及谷歌的.kmz 版(名称分别为: VillagesChina2666.kmz + VillagesChina2666.rar), 包括 7 个数据文件, 数据大小为 4.50 MB (压缩为 2 个文件, 258 KB)。

利用该数据集能方便地了解该批次传统村落的空间分布关系, 图 4 是在 Google Earth 地图的 2,666 个村落点数据, 可见空间上的村落分布同前几批, 依然不均, 呈现东南部多而西北部少的分布特性。村落数量上多的序列是湖南的 401 个、福建的 265、安徽的 237 和浙江的 235, 最少序列的是仅 1 个的北京、天津、宁夏和新疆, 宁夏几个批次均少, 新疆的递减较明显, 上海市连续 4 批无村落被收录, 城市化作用对传统村落的减少影响显著。

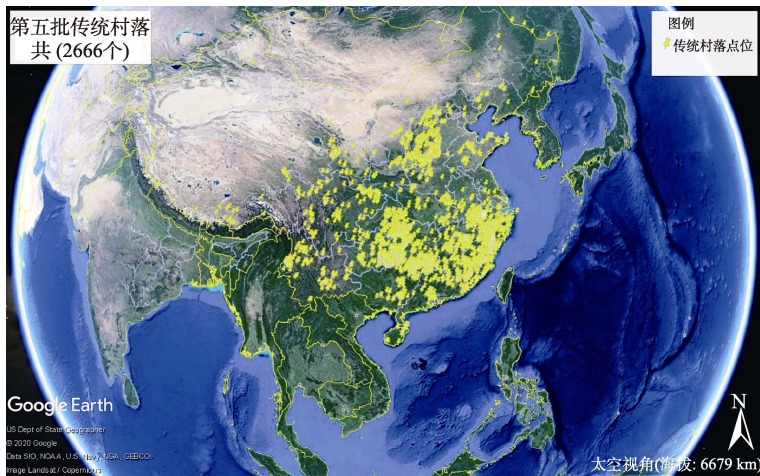


图 4 第五批 2666 个村落的空间分布

(来源: Google Earth)

作者分工: 余亮对数据集获取开发做了总体设计, 撰写了数据论文等; 唐铭婕和刘智涛整理验证了关键数据; 付蒙, 邱雨晨, 曹玲莉, 杨旭辉和沈军鑫进行了数据定位处理。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 余亮, 刘佳, 丁雨倩等. 中国 2555 个传统村落空间分布数据集[DB/OL]. 全球变化数据仓储, 2018. <https://doi.org/10.3974/geodb.2018.04.06.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2018.04.06.V1>.
- [2] 余亮, 丁雨倩, 唐铭婕等. 中国新增 1598 个传统村落空间分布数据集[DB/OL]. 全球变化数据仓储, 2019. <https://doi.org/10.3974/geodb.2019.01.19.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2019.01.19.V1>.
- [3] 余亮, 唐铭婕, 付蒙等. 中国再增 2666 个传统村落空间分布数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2020. <https://doi.org/10.3974/geodb.2020.03.22.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2020.03.22.V1>.
- [4] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [5] 宁夏回族自治区住房和城乡建设厅门户网站[OL]. <http://jst.nx.gov.cn/info/1077/30836.htm>.
- [6] 何曼丽, 丁旭华, 俞凯旋. 地貌视角的诸暨乡村地名空间分布研究[J]. 测绘科学, 2020, 45(11): 147–153.