

中国 2555 个传统村落空间分布数据集

余亮^{1*}, 刘佳², 丁雨倩¹, 曹倩颖¹, 廖庆霞¹, 唐铭婕¹, 付蒙¹

1. 苏州大学建筑学院, 苏州 215123;

2. 中国农业科学院农业资源与农业区划研究所, 北京 100081

摘要: 根据 2012 年 12 月、2013 年 8 月和 2014 年 11 月国家住房和城乡建设部、文化部、财政部先后公布的三批中国传统村落名录 (2,555 个), 参考百度地图、Google Earth 影像, 从村落名称的文字提示或者传统村落影像图形的几何中心, 提取村落坐标。对于地图上查不到该村落, 也没有参考影像的, 则以传统村落所在的上一级行政单位的坐标作为该村落的地理坐标。最后, 得到中国 2,555 个传统村落空间分布数据集。该数据集存储为.shp 和.kmz 格式, 由 8 个数据文件组成, 数据量为 6.88 MB (压缩为 2 个文件, 416 KB)。该数据集的分析研究成果发表在《地理科学进展》2016 年 35 卷 11 期。

关键词: 中国; 传统村落; 空间分布; 地理科学进展

DOI: 10.3974/geodp.2018.02.03

1 前言

传统村落是乡村居住和建筑在一定空间的聚合形式, 遍布于我国大江南北的广阔土地, 是因地制宜、自发生长并呈现丰富多样建筑形态的居住空间, 具有鲜明的地域特色 (图 1)^[1]。作为宝贵的历史文化遗产, 传统村落 (除必要的叙述外, 以下省略“传统”) 反映记录了真实朴素的民俗民风 and 民情, 其丰富的形态不仅受到当时经济文化和传统观念等因素的影响而致, 还受地形、地貌等自然因素的影响^[2], 对村落的分析研究不仅需要借助大量有效的微观数据, 还应强化村落宏观层面的空间分布梳理, 目前已有不少类似的村落分布研究^[3-5]。中国 2555 个传统村落空间分布数据集, 是分析解读村落地理特征及空间发展布局的基础数据, 在关联论文 (本数据集未介绍) 的数据整理归纳中还应用地理格网的分级方法, 形成了数据分级使用以及为多元数据的融合提供了方便统一的框架^[6-7], 并且因便于空间数据的分析比较以及可视化表达等, 地理格网系统在国内国外有不少应用^[8-9]。

2 数据集元数据简介

中国 2,555 个传统村落空间分布数据集^[12] (VillagesChina2555) 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。

收稿日期: 2018-03-16; 修订日期: 2018-06-15; 出版日期: 2018-06-25

基金项目: 国家自然科学基金面上项目 (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, 刘佳, 丁雨倩等. 中国 2555 个传统村落空间分布数据集[J]. 全球变化数据学报, 2018, 2(2): 144-150. DOI: 10.3974/geodp.2018.02.03.

[2] 余亮, 刘佳, 曹倩颖等. 中国 2555 个传统村落空间分布数据集[DB/OL]. 全球变化科学研究数据出版系统, 2018. DOI: 10.3974/geodb.2018.04.06.V1.

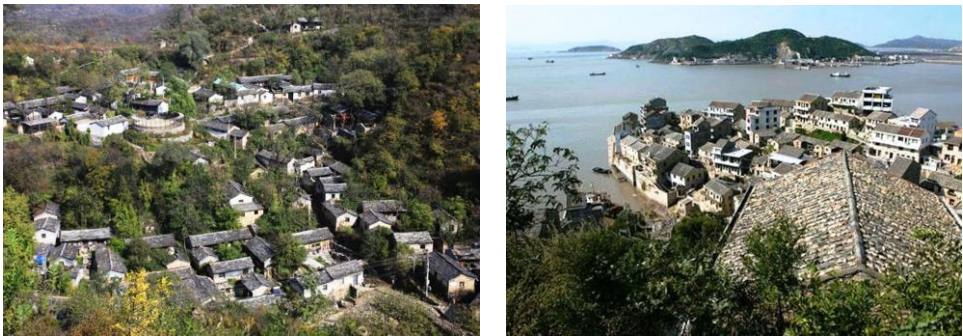


图 1 地域特色鲜明的传统村落（左：北京房山水峪村^[10]；右：台州市温岭市石塘镇里箬村^[11]）

表 1 中国 2,555 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 2555 个传统村落空间分布数据集
数据集短名	VillagesChina2555
作者信息	余亮 F-8099-2018, 苏州大学建筑学院, yuliang_163cn@163.com 刘佳 O-6279-2018, 中国农业科学院农业资源与农业区划研究所, liujia06@caas.cn 曹倩颖 O-6478-2018, 苏州大学建筑学院, 604459330@qq.com 丁雨倩 O-6445-2018, 苏州大学建筑学院, 442801422@qq.com 廖庆霞 O-6442-2018, 苏州大学建筑学院, 719197729@qq.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
地理区域	中国, 覆盖 31 个省级行政区（不包括港澳台）
数据年代	2012–2014 年
数据格式	.shp、.kmz
数据量	6.88 MB
数据集组成	2 个文件（VillagesChina2555.kmz、VillagesChina2555.rar）
基金项目	国家自然科学基金面上项目（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）、实体数据（中英文）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 ^[13]

3 数据提取及形成方法

数据是研究的基础和核心，但公布村落的名录中，无位置数据显示，且村落名称又不易辨别，故需对村落坐标和村落名称进行进一步辨别整理。从一定高度俯瞰，民居是村落的最大和最突起的人工物，是人类日常起居生活工作必需的场所，因有遮蔽隔绝外界的空

间,可以认为村落是民居的集合和自然载体,民居和村落类似水与河流的依赖关系(图2)。村落数据主要通过村落名称辨别以及认知村落空间的民居形体特征来提取村落的位置点数据,同时利用村落的名称辨别,确定村落及上级的行政隶属关系,以便归入不同的数据列。



图2 民居在村落和与村落的依存 (左:北京房山南窖乡水峪村;
右:台州市温岭市石塘镇里箬村;余亮拍摄)

3.1 数据集范围和村落原始数据

数据集范围:中国,覆盖31个省级行政区(不包括港澳台等)。村落原始数据:2012年起国家住房和城乡建设部、文化部、财政部等先后公布的第三批全国传统村落名录^①(第一批646个、第二批915个和第三批994个,不包含港澳台等地区,不附带坐标),数据集以上述名录的2555个村落为对象,通过赋予地理坐标,村落的行政隶属关系确认等程序形成村落的空间点数据。

3.2 村落点位置数据的提取

依据研究的对象特点,民居是空间组成的重要因子并是数据影像截取的依据,村落的位置关系根据村落坐标及其附带的文字信息进行判断与配对提取。首先利用百度地图先期(关联论文^[6])提取的村落点数据,考虑数据在Google Earth的共享应用要求,结合Google Earth的线上影像特点进行了村落的位置微调(图3),形成.kmz格式后,再转成.shp格式存储应用。

先期利用百度地图提取的村落点数据由地理坐标和影像数据组成,考虑到村落内民居多于一栋,且有屋顶的面状分布,以村落影像图形的几何中心为村落坐标,提取村落坐标的同时,还截取影像作为实证补充,作为目测的判断操作,提取时还考虑以下3种情况:

- (1) 从地图的文字提示中直接获得村落的坐标信息,不需影像辅助。
- (2) 从地图中获得村落较清晰的遥感影像,通过判断村落的几何中心获取坐标,易产生以下2种情况:

1)对于民居分布较均衡且形体较协调规整的小村落,提取平面几何中心为该村落坐标,图4的a-1和a-2表示了杭州市桐庐县富春江镇石舍村的提取判断过程。

2)对于村落较大、各种建筑类型多而复杂使之难以判断的情况,则结合民居的整体协调程度判断民居的大约营造年代,将坐标点落在较古老而形体统一的建筑几何中心上,这

① 来源:住房和城乡建设部、文化部、财政部等发布的“建村[2012]189号、[2013]124号和[2014]168号”文件。

是因为受当时的技术等条件影响，一定空间范围内的民居形体会相对统一协调，图 4 的 b-1 和 b-2 为东莞市寮步镇西溪村的提取过程。

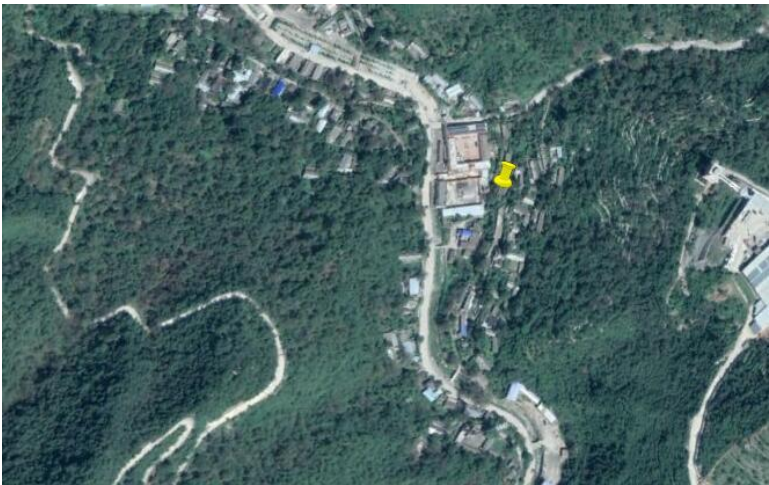


图 3 依据 Google Earth 影像的村落位置微调（北京房山南窖乡水峪村）

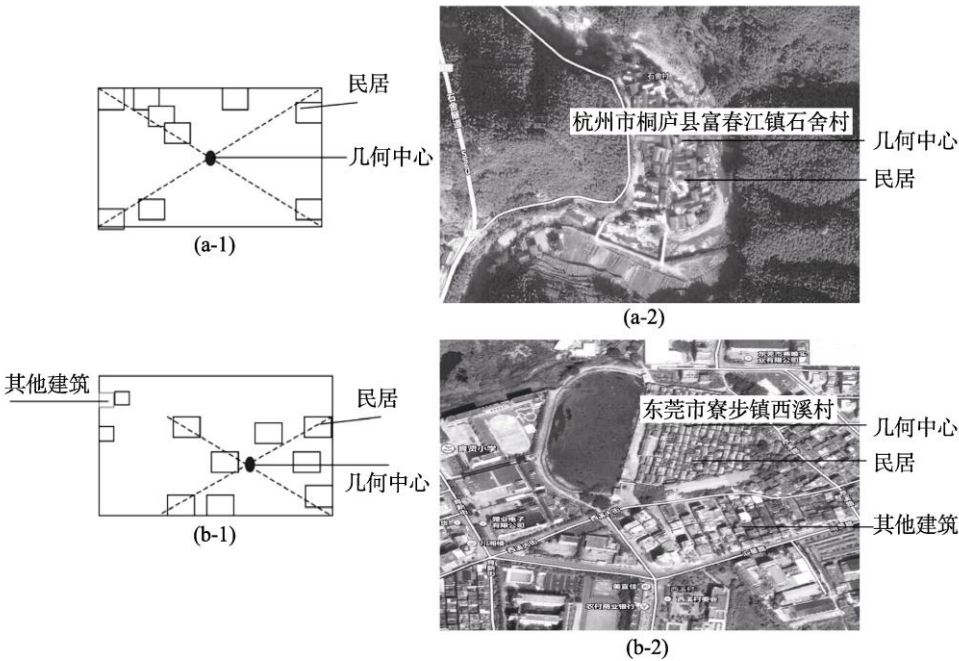


图 4 村落坐标点选取示意图

（3）地图上难查到该村落的坐标信息，也无参考的遥感影像，则以村落所在的上一级行政单位的坐标为该村落的地理坐标。

根据以上方法提取的村落坐标数据见表 2。从中可看出，第一、二、三批传统村落中直接查到坐标和取几何中心获取坐标的村落数分别占当批公布村落总数的 79%、64%、82%；

用上级行政单元坐标代替的村落数占分别为 21%、36%和 18%，三批的平均值约为 25%，数量相对较少，可认为对研究的整体结果影响有限。

表 2 三批传统村落的获取坐标统计表

批次与类型	上级坐标代替的村落	直接查到坐标的村落	取几何中心坐标的村落
第一批	134	464	48
第二批	332	519	64
第三批	177	799	18

3.3 村落名称的辨别整理

原始公布的村落名称使用文字，一村一行，通过一连串的文字表述基本可辨别村落及其上下级的行政隶属关系，作业需根据各行村落及附带的隶属关系，以分列格式，整理形成隶属结构关系明确的各列段，方便以后的数据下载使用，如第一批传统村落名录的北京市第一行的村落：房山区南窖乡水峪村，在略去前置的区县情况下，有乡或镇的行政标记，从省、直辖市到村的 4 级，较规范清晰，一目了然。但实际的操作并非如此，不少村落表述上的辨别难度较大，除 4 级行政隶属的关系不甚清晰外，还有不少村落概念的理解与辨别的差异，几种典型的现象罗列如下：1）行政级别提升，改革开放前的隶属，县级单位一般清楚地标示“县”，而现在有不少地方称县级市，后缀因标“市”后难定准，如“丽水市龙泉市龙南乡蛟垵村”的“龙泉市”虽称市，实际是“县”，故归在“县”的列段；2）村落的名称与功能的不清，如“阳泉市郊区义井镇小河村”的“郊区”，一般“郊区”指紧邻中心城区，计划经济时期主要为城市提供农副产品的地域范围，是功能定位，又是名称，但这里列段需要的是名称，因缺少“县”级名称而归在了“县”段中；类似的还有“泸州市泸县兆雅镇新溪村”中的“泸县”，“县”原本是“泸县”名称的一部分，不能去掉，故在列段中直接使用了“泸县”；3）名称尾部虽大多以村为标记，但不少以组、村委会、堡或寨等的名称出现，这些是“村”或下一级组织，甚至别名，均处在标记末尾，按村级来辨别定准，对个别表示不清的末尾，则通过进一步检索查验，予以确认。

4 数据集构成及村落分布

4.1 数据集构成

数据集由 2 个子集构成，分别是中国 2,555 个传统村落空间分布点数据的 ArcGis 的.shp 格式和谷歌的.kmz 格式（名称分别为：VillagesChina2555.kmz，VillagesChina2555.rar），8 个数据文件组成，数据量约为 6.88 MB（压缩为 2 个文件，416 KB）。

4.2 村落分布与村落数据调整

省市层面的村落数经统计后可知，不同地区的村落数不均匀，云南、贵州的数量较多，占 10%以上；辽宁、吉林、黑龙江和上海等的数量较少，均不足 10 个，第一批中辽宁和吉林为零，第二批中天津、辽宁、上海和海南为零，第三批中天津、上海和宁夏为零（图 5），

三批共 2555 个村落点数据见图 6（Google Earth）。

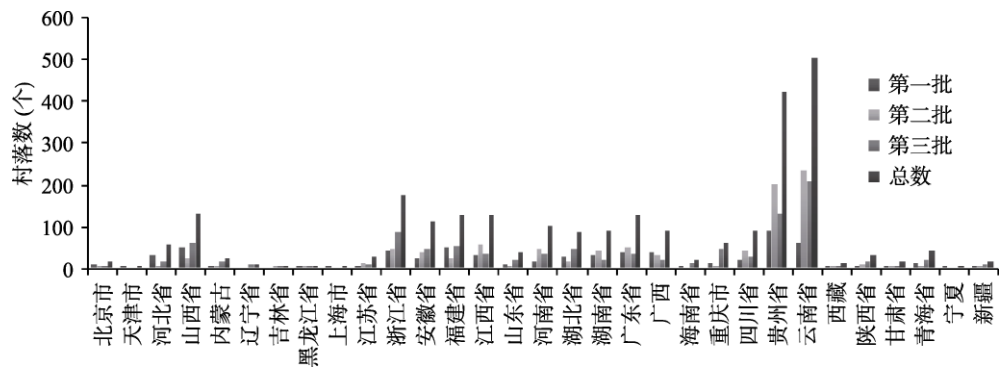


图 5 各省市传统村落分布

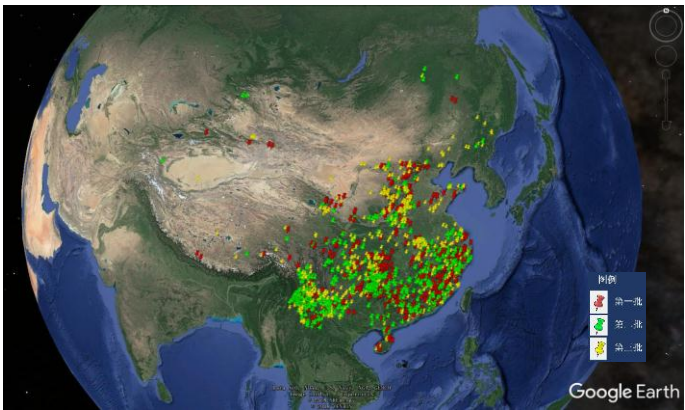


图 6 三批共 2,555 个村落分布（Google Earth）

村落的位置调整：关联论文先期使用百度地图进行了位置匹配，由于百度地图本身的差异，加之使用 Google Earth 后的位置差异，需要调整，先大致对 100 条村落的位置进行了校对或微调，可知 90% 以上的位置存在或大或小的错位需要调整。将属性表格关联到一起后，会有错位，如在 Google Earth 下打开时有村落点落在耕地的可能性，因而需要微调他们的位置，然后再转成.shp 格式存储。

5 讨论和总结

数据集的村落点坐标提取利用了百度地图和 Google Earth 影像，通过坐标提取使原本不显示坐标的村落名录能够显示空间分布状态，利用该数据集可整体把握中国传统村落的空间关系，为地理特征的研究提供了重要的基础依据。数据集分三种情况提取了村落坐标，直接查到和以几何中心获取坐标的村落数占其总数的 75%。另外，从图可知实际的村落空间分布不均匀，有些区域无村落，由于村落受不同的地理条件制约，现在主要以行政区划确定村落数量有一定的局限性，如在决定村落名录时，是否可适当考虑自然区划的特点？诸如此类，应为后续的研究提供更多的思考。

作者分工：余亮对数据集的开发做了总体设计，撰写了数据论文等；刘佳指导处理了关键的数据；曹倩颖、丁雨倩、廖庆霞、唐铭婕和付蒙等重新采集和处理了本数据集。

致谢：瞿颖、赵雨婷、冯雪、雷悦和袁萧奕等本科生参与了先期村落数据的提取，在此深表谢意！

参考文献

- [1] 陈甲全, 张义丰, 陈美景. 古村落研究综述[J]. 安徽农业科学, 2008, 36(23): 10103-10105.
- [2] 武启祥, 韩林飞, 朱连奇等. 江西婺源古村落空间布局探析[J]. 规划师, 2010, 26(4): 84-89.
- [3] 严赛. 中国传统村落分布的特点及其原因分析[J]. 大理学院学报, 2014, 13(9): 25-29.
- [4] 刘大均, 胡静, 陈君子等. 中国传统村落的空间分布格局研究[J]. 中国人口 资源与环境, 2014, 24(4): 157-162.
- [5] 康璟瑶, 章锦河, 胡欢等. 中国传统村落空间分布特征分析[J]. 地理科学进展, 2016, 35(7): 839-850.
- [6] 余亮, 孟晓丽. 基于地理格网分级法提取的中国传统村落空间分布[J]. 地理科学进展, 2016, 35(11): 1388-1396.
- [7] 周成虎, 欧阳, 马廷. 地理格网模型研究进展[J]. 地理科学进展, 2009, 28(5): 657-662.
- [8] 姚永慧, 张百平, 罗扬等. 格网算法在空间格局分析中的应用: 以贵州景观空间格局分析为例[J]. 地球信息科学, 2006, 8(1): 73-78.
- [9] Goodchild, M. F., Guo, H., Annoni, A., *et al.* Next-generation digital earth [J]. *Proceedings of the National Academy of Sciences*, 2012, 109(28): 11088-11094.
- [10] 古岛的博客: 北京房山南窖乡水峪村[OL]. <http://blog.sina.com.cn/u/3253570163>.
- [11] 温岭新闻网: 台州市温岭市石塘镇里箬村[OL]. <http://wlnews.zjo.com.cn/wlrbsystem/>.
- [12] 余亮, 刘佳, 曹倩颖等. 中国 2555 个传统村落空间分布数据集[DB/OL]. 全球变化科学研究数据出版系统, 2018. DOI: 10.3974/geodb.2018.04.06.V1.
- [13] 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]. DOI: 10.3974/dp.policy.2014.05 (2017 年更新).