

# The Spatial Distribution Dataset of 2666 Chinese Traditional Villages

Yu, L.\* Tang, M. J. Fu, M. Liu, Z. T. Qiu, Y. C. Cao, L. L. Yang, X. H. Shen, J. X.

School of Architecture, Soochow University, Suzhou 215123, China

**Abstract:** Following the list of Chinese traditional villages released by the Ministry of Housing and Urban-Rural Development of P. R. China, Ministry of Culture of P. R. China, and Ministry of Finance of P. R. China from 2012 to 2016, 2,666 Chinese traditional villages were released in June 2019. Using GIS methodology, the dataset was developed based on the information about the names of the new listed villages or geo-location of villages from Baidu Map and Google Earth images. Furthermore, more historical documentation or images helped to determine village geo-location. In a few cases, if a village was not found on the map, nor Google Earth image, the upper-level administrative resident village nearby was adopted. The dataset was archived in .shp and .kmz data formats with a data size of 7.48 MB in 6 data files (compressed to 362 KB in two data files).

**Keywords:** China; traditional villages; list; spatial distribution; the fifth batch

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

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

## Dataset Availability Statement:

The dataset supporting this paper was published and is accessible through the *Digital Journal of Global Change Data Repository* at: <https://doi.org/10.3974/geodb.2020.03.22.V1> or <https://cstr.escience.org.cn/CSTR:20146.11.2020.03.22.V1>.

## 1 Introduction

Traditional villages are spread all over the world, depict various forms of gathering and living since the birth of human beings, and are valuable historical and cultural heritage. They exhibit different characteristics that correlate to different regions. Due to the needs of long-term life and activities, a large number of traditional dwellings exist in traditional villages, which form inseparable interdependent relationships between them. For over ten years, the author has investigated dozens of villages in China, been impressed by the vivid regional characteristics of villages and dwellings and the exquisite skills of builders, and has empathized with the collapse of some villages and dwellings due to failure to improve and reconstruct on time. This is coming at a time when the country is aggressively advocating

---

**Received:** 26-02-2020; **Accepted:** 11-10-2021; **Published:** 25-03-2022

**Foundation:** National Natural Science Foundation of China (41371173)

\***Corresponding Author:** Yu, L. F-8099-2018, School of Architecture, Soochow University, [yuliang\\_163cn@163.com](mailto:yuliang_163cn@163.com)

**Data Citation:** [1] Yu, L., Tang, M. J., Fu, M., *et al.* The spatial distribution dataset of 2666 Chinese traditional villages [J]. *Journal of Global Change Data & Discovery*, 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] Yu, L., Tang, M. J., Fu, M., *et al.* The spatial distribution dataset of 2666 additional Chinese traditional villages [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2020.03.22.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2020.03.22.V1>.

“Lucid waters and lush mountains are invaluable assets” and rural revitalization, timely and sustained protection, and reconstruction work are very essential (Figure 1).

The spatial distribution dataset of 2,666 additional Chinese traditional villages includes the reorganization and publishing of the 2,555 village spatial distribution dataset and the 1,598 village spatial distribution dataset<sup>[1,2]</sup> in the previous period. This data source does not only identify and interpret the administrative division and subordination relationship at the grassroots level of villages and above, such as from towns and townships to counties but also compares and confirms the geographical location of the corresponding villages, and analyses the spatial distribution characteristics, thus illustrating the outstanding rural culture and stimulating the spiritual essence of traditional villages.



**Figure 1** The fifth batch of traditional villages: preservation and restoration of dwellings (Left: Dongpu village, Dongpu street, Yuecheng district, Shaoxing city, Zhejiang province; Right: Beilei village, Fotang town, Yiwu county, Jinhua city, Zhejiang province, by Yu, L., in 2017 and 2020)

## 2 Metadata of the Dataset

The metadata of The spatial distribution dataset of 2666 additional Chinese traditional villages<sup>[3]</sup> is summarized in Table 1. It includes the full name, authors, data format, data size, data files, data publisher, and data sharing policy, etc.

**Table 1** Metadata summary of the Spatial distribution dataset of 2666 additional Chinese traditional villages

Items	Description
Dataset full name	The spatial distribution dataset of 2666 more Chinese traditional villages
Dataset short name	VillagesChina2666
Authors	Yu, L. F-8099-2018, School of Architecture, Soochow University, yuliang_163cn@163.com Tang, M. J. O-6467-2018, School of Architecture, Soochow University, 361988267@qq.com Fu, M. O-6455-2018, School of Architecture, Soochow University, 821064405@qq.com Liu, Z. T. ABH-4639-2020, School of Architecture, Soochow University, 785025073@qq.com Qiu, Y. C. ABH-5207-2020, School of Architecture, Soochow University, 375284315@qq.com Cao, L. L. ABI-1416-2020, School of Architecture, Soochow University, 1083748619@qq.com Yang, X. H. ABH-6245-2020, School of Architecture, Soochow University, 654712015 @qq.com Shen, J. X. ABH-7152-2020, Tongji Zhejiang College, 409209740 @qq.com
Geographical region	China, 31 provincial-level administrative regions (Hong Kong, Macao, and Taiwan without data)
Year	2012–2019
Data format	.shp, .kmz
Data size	4.50 MB
Data files	Two files (VillagesChina2666.kmz + VillagesChina2666.rar)
Foundation	National Natural Science Foundation of China (41371173)
Data publisher	Global Change Research Data Publishing & Repository, <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>
Address	No. 11A, Datun Road, Chaoyang District, Beijing 100101, China
Data sharing policy	<b>Data</b> from the Global Change Research Data Publishing & Repository includes metadata, datasets (in the <i>Digital Journal of Global Change Data Repository</i> ), and publications (in the <i>Journal of Global Change Data &amp; Discovery</i> ). <b>Data</b> sharing policy includes: (1) <b>Data</b> are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use <b>Data</b> subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute <b>Data</b> subject to written permission from the GCdataPR Editorial Office and the issuance of a <b>Data</b> redistribution license; and (4) If <b>Data</b> are used to compile new datasets, the ‘ten per cent principal’ should be followed such that <b>Data</b> records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset <sup>[4]</sup>

Communication and search system DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

### 3 Point Data Processing of Village Location

The list of villages does not show spatial locations; however, it displays a line of name texts with administrative subordination as well as location names. The location of village space point data is based on the extraction of text name information, and it was located and obtained by comparing and identifying the spatial form characteristics of villages. For the spatial cognition of villagers that reside in a compact community, the main constituent element is a residential building, with fields, rivers, and infrastructure such as roads. Dwellings not only have a large horizontal plane on rooftops in space, but are distinct from other object features, for example, the natural elements of mountains, plants, and water surfaces, as well as artificial elements such as roads, bridges, etc. The space point data exploits the characteristics that the roofs have significant differences between other objects, and their graph-structured forms can easily be recognized.

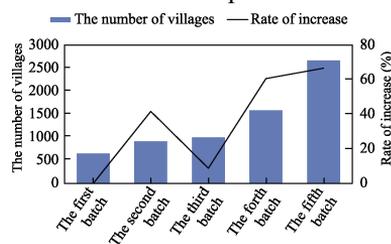
#### 3.1 Data Sources

Raw data of villages: The quantity of the fifth batch of villages listed in the list of traditional Chinese villages (Finally published) is 2,666 (Compared with the initial figure, 22 new villages were added and 2 villages were deleted)<sup>[5]</sup>. The dataset spans 30 provincial administrative regions, and in addition to the three batches of national traditional villages that were published, the total number of villages in the list reached 6,819. The increment is obvious, especially the rise from 60% to 65% in the fourth batch to the fifth (Figure 2).

#### 3.2 Location Data and Names of Villages

Data acquisition of the village space point is highly similar to the previous two occasions<sup>[1,2]</sup>. There are two steps: first, the space point data positioning, which requires distinguishing the differences in space between the villages and natural or artificial objects and focusing on the relatively orderly protruding individuals with small volume; second, sorting village names, in a bid to draft a regular attribute sheet using the semi-automatic method combined with the manual one. Semi-automatic means using Excel-related tools and methods to improve efficiency. Tools applied include left, right, mid-function, and advanced classification to directly identify the data segments of village names. The macro of the Excel tool is used to translate the characters into pinyin. Then, a new module is inserted into the visual basic editor. After inputting a code and defining the function meticulously, the Chinese characters are converted and the initial case is processed. Without any doubt, these methods are different from the previous two. The manual method targets unusual names, such as segments of autonomous regions, streets, communities, towns, and villages, which are executed after semi-automatic processing, as detailed as follows.

The positioning of the village space point data is inputted into the Baidu Map to obtain the point location of the village after which it is imported to Google Earth Map. The coordinate difference of the space point location is adjusted online by the manual visual method to determine whether the point falls on the roofs of ancient buildings in the village and on the geometric center position, which represents the spatial characteristics of the village. The positioning of the data at each point is not obvious and requires several manual adjustments. The main issue with the positioning, including on the first two occasions, is that it is not easy



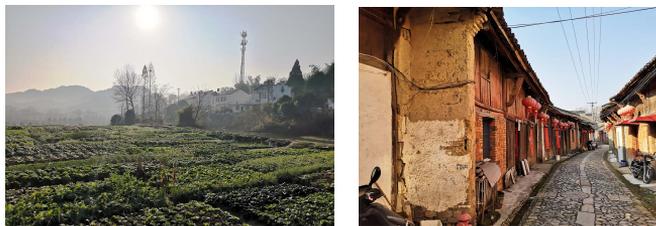
**Figure 2** Trends of the number of first to fifth batch of villages

to judge. After the Baidu Map imports data to Google Earth Map, changing the original position becomes much easier. For example, the point located on the roof will be shifted to the field or pond, etc., and with the significant increase in data size between the fifth batch and the fourth batch, the workload of adjustment increases heavily. In addition, for names that are difficult to identify and lack clues, based on inquiries made for relevant information or direct contact with the village for confirmation, if it is really difficult to position the village, it should be assigned to the administrative unit up to one level<sup>[1]</sup>.

### 3.3 Sorting out the Name of the Villages

Villages are grass root level organizations, and a series of names in the list are direct clues to obtain the location of the villages, especially the marks at the end. This implies that the sign of names at the end of the village is varied, but not uniform, which may be related to the Chinese multi-ethnic and multi-climatic characteristics<sup>[6]</sup>. The processing method is mainly to confirm the administrative levels and subordinate relations of the attribute sheet, segments from provinces, cities, and counties to towns and villages. Five administrative levels exist because the five administrative levels of general provinces are different from the four administrative levels of municipalities directly under the central government. This dataset is based on five columns, sets a municipality directly under the central government and the general municipal repeated column segments, such as the column “Beijing” is repeated in level one and level two of the column “Heilongguan village, Fozizhuang town, Fangshan district, and Beijing”.

Secondly, by observing the administrative sequence from province to village, it can be seen that many villages have sixth-level endings after the fifth level of administrative affiliation relationship, reflecting that multiple spatial positions need to be identified and located. According to the roof image, if it is clear that the sixth level can be identified, and will be at the sixth level. If it were not obvious, it would be at the fifth level. The “village” is the most frequent, but it doesn’t necessarily end with it; there are 2,598 villages that end with “village” in the fifth batch of five-level villages (Table 2), next is 34 villages in the “community” section, 27 in “cuncun”, 3 in “old street” and “tun”, and 1 in “gacha” (Equivalent to an administrative village). There are no “neighborhood committees and village committees”. At the fifth and sixth levels, the administrative village is generally connected to the natural village, such as “Hongyan old village-tun, Zhushan village committee, Lianhua town, Gongcheng Yao autonomous county, Guilin city, and Guangxi province.” “Zhushan village committee” is the 5<sup>th</sup> village level, “Hongyan old village-tun” is the six-level natural village, and should be positioned on the “Hongyan old village-tun”. If it is difficult to decipher, it will be located in the upper “Zhushan village committee”. Generally, an administrative village mostly corresponds to one natural village, and there are multiple corresponding villages as well, such as “Liuxiang tun, Langchong tun, Shangguchen tun” in “Liuxiang village, Liuxiang town, Jinxu Yao autonomous county, Laibin city, and Guangxi province”. 401 villages have sixth-level labels in the fifth batch of the village (15.04% of



**Figure 3** The development relationship between towns and villages (The fifth patch, Jinyan village, Maotanchang town, Jin’an district, Liu’an city, and Anhui province, left: Jinyan village; right: Maotanchang old street, by Yu, L., 2020)

2,666), the tails are “village, group, zhai, tun, zhaicun, ditch, slope, street, zhuang, and bay”. The most common is “village”. There are 262 villages, followed by the 45 in “group”, 43 in “zhai”, and “zhuang” and “bay” are just one. The form is diverse, the number of combinations differs as a whole, and the name vividly reflects the local natural landscape features.

**Table 2** Village level 5 and 6 and the suffix of “ancient” and “old”

No	The fifth level	Amount	The sixth level	Amount	Combine “ancient” and “old” endings	Amount
1	Village	2,598	Village	262	Ancient village	6
2	Community	34	Group	45	Ancient zhai	3
3	Cuncun	27	Zhai	43	Ancient zhaicun	1
4	Old street	3	Zhaicun	10	Old street	3
5	Tun	3	Tun	29	Old-street village	3
6	Gacha	1	Ditch	4		
7	Neighborhood committee	0	Slope	4		
8	Village committee	0	Street	2		
9			Zhuang	1		
10			Bay	1		
Total		2,666		401		16

Additionally, there are endings with “ancient village”, “ancient zhai”, and “old street” derived from the combination of “ancient” and “old”, both at levels 5 and 6. We do not want to pay attention to when the endings “ancient village” and “old street” should be used, or what should be intended in any “ancient-old” combination. At least, there are more clues for attention when it comes to spatial positioning. When positioning the fifth batch of villages, there is no spatial difference between the combination of “ancient” and “old”. For example, if without further investigation, it is difficult to distinguish whether “Gu” in “Baigu village, Puli village, Zhongshan township, Luoping county, and Qujing city” means “ancient”. There are six “ancient villages” and three “ancient zhai” in the villages. In addition, there are three “old streets”. Among them, although “Beizha old street in Zhegao town” and “Tongyang old street in Tongyang town” in Chaohu city, Anhui province are village level, according to the field investigation, there is a close relationship between the village and the town. The development of the town has traces of the village. The expansion of the village provides the basis for the development of the town. In space, it exhibits the characteristics of integration. These spatial phenomena can be inferred and supported by remote sensing images and applied in positioning. Generally speaking, towns have more commercial and administrative functions than villages. Some villages are “surrounded” by the new towns in space, while some are neighbors. Through the epoch, many villages exited the former state, and their structure, boundary, and mechanism changed tremendously. For example, “Jinyan village, Maotanchang town, Jin’an district, Liu’an city, and Anhui province” (Figure 3) in the fifth batch of villages, the village is near the town, and some locals refer to the old street of the town as the old buildings of the village. This highlights the close spatial relationship between them. Generally, if it were to be a village in space, we would position the point on the village. Otherwise, it would be positioned on the old street or the old buildings in the town. Similar characteristics are also reflected in the old street of Tongyang town, Chaohu city, and Anhui province, and the old street of Beizha town, Zhegao town, Chaohu city, and Anhui province.

Finally, the spatial location points of the sixth levels (401 in total) are shown in Table 3. It can be seen that the sixth level is the most common. Also, the locating point of the fifth level and the fourth level follow the rules, which is in line with the aforementioned method that if the location of the level is difficult to find, the administrative unit of any one level above will be assigned.

**Table 3** Locate numbers of the sixth level village

Locate point	The sixth level	Village	Group	Zhai	Stockaded village	Tun	Gou	Slope	Street	Zhuang	Bay	Amount
The sixth level		196	11	18	6	15	1	1	1	1	1	253
The fifth level		58	28	23	4	12	1	3	1	0	0	130
The Fourth level (Town)		5	5	2	0	1	0	0	0	0	0	9
The Fourth level (Countryside)		3	5	0	0	1	0	0	0	0	0	9
Total		262	45	43	10	29	4	4	2	1	1	401

## 4 Datasets Results and Discussion

The dataset consists of two files: 1) VillagesChina2666.rar, which is composed of seven data files in ArcGIS with a data size of 4.50 MB; 2) VillagesChina2666.kmz, which is two files with a data size of 258 KB in Google Earth.

This dataset can be used to understand the spatial distribution of the traditional villages easily. The data of the 2,666 villages are shown in Figure 4 (Google Earth). It can be seen that the spatial distribution of villages remains uneven, same as the previous batches, with more villages in the southeast and fewer in the northwest. The largest number in Hunan is 401, with 265 in Fujian, 237 in Anhui, and 235 in Zhejiang. The least are Beijing, Tianjin, Ningxia, and Xinjiang.

The several batches in Ningxia are fewer and the decline in Xinjiang is transparent. There aren't any traditional villages collected in the list for the four consecutive batches in Shanghai, it seems that urbanization has influenced the declining number of traditional villages.



**Figure 4** Spatial distribution of the fifth batch of 2,666 villages (Google Earth)

### Author Contributions

Yu, L. made an overall design for the acquisition and development of datasets, and wrote data papers; Tang, M. J., and Liu, Z. T. compiled the key data; Fu, M., Qiu, Y. C., Cao, L. L., Yang, X. H., and Shen, J. X. collected and processed the dataset.

### Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Yu, L., Liu, J., Ding, Y. Q., *et al.* The spatial distribution dataset of 2555 Chinese traditional villages [DB/OL]. Global Change Data Repository, 2018. <https://doi.org/10.3974/geodb.2018.04.06.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2018.04.06.V1>.
- [2] Yu, L., Ding, Y. Q., Tang, M. J., *et al.* Spatial distribution dataset of 1598 more traditional villages in China [DB/OL]. Global Change Data Repository, 2019. <https://doi.org/10.3974/geodb.2019.01.19.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2019.01.19.V1>.
- [3] Yu, L., Tang, M. J., Fu, M., *et al.* Spatial distribution dataset of 2666 more traditional villages in China [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2021. <https://doi.org/10.3974/geodb.2020.03.22.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2020.03.22.V1>.
- [4] GCdata PR Editorial Office. GCdata PR data sharing policy [OL]. <https://doi.org/10.3974/dp.policy.2014.05> (Updated 2017).
- [5] Ningxia Hui Autonomous Region, The Ministry of Housing and Urban-Rural Development [OL]. <http://jst.nx.gov.cn/info/1077/30836.htm>.
- [6] He, M. L., Ding, X. H., Yu, K. X. Spatial distribution characteristics of place-names in Zhuji from the perspective of geomorphology [J]. *Science of Surveying and Mapping*, 2020, 45(11): 147–153.