

# A Spatio-temporal Dataset for the Novel Coronavirus (2019-nCoV) and Its Typical Applications

**Authors:** Lin Hao, Bao Junya, Lin Hao

**Date:** 2020-02-19T00:00:00+00:00

## Abstract

Currently, the novel coronavirus (2019-nCoV) epidemic is receiving widespread attention from researchers worldwide. However, there is currently no official channel for real-time open-sourcing of 2019-nCoV epidemic data. To facilitate scientific research related to this epidemic, this study aims to provide authoritative, open, and multi-scale spatiotemporal datasets of the novel coronavirus (2019-nCoV) for the research community, serving as an important data source for epidemic monitoring, prevention and control, prediction, and early warning. Furthermore, this dataset can also be applied to multi-scale, multi-temporal mapping and visualization of the 2019-nCoV epidemic, offering guidance for spatial distribution, evolution, trend analysis, and simulation forecasting of the epidemic.

## Full Text

### Preamble

#### Spatiotemporal Dataset of Novel Coronavirus (2019-nCoV) and Its Typical Applications

**Hao Lin, Junya Bao**

School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China

**Objective:** To provide authoritative, open-access, and multi-scale spatiotemporal datasets of the novel coronavirus (2019-nCoV) for the broader research community. **Methods:** Data were collected, cleaned, organized, produced, and updated based on epidemic announcements from national, provincial, and municipal health commissions. **Results:** The 2019-nCoV spatiotemporal dataset has been shared and is updated every five days, accessible at <https://github.com/Estelle0217/COVID-19-Epidemic-Dataset>. **Limitations:**

Statistical attributes are constrained by information disclosed in official epidemic announcements, with some prefecture-level regions lacking corresponding statistical data. **Conclusion:** This dataset can provide an essential data source for epidemic monitoring, prevention, prediction, and early warning.

**Keywords:** 2019-nCoV; novel coronavirus; dataset; multi-scale; multi-temporal

**Document Code:** A

**Classification Number:** K901.2

## 1. Introduction

Since December 2019, the first cases of unexplained pneumonia emerged in Wuhan, Hubei Province (diagnosed as infections with the novel coronavirus, 2019-nCoV). As of 24:00 on February 5, 2020, a cumulative total of 28,060 confirmed cases and 564 deaths had been reported across 25 countries worldwide. The 2019-nCoV epidemic has attracted widespread attention from biologists [1-3], medical scientists [4, 5], epidemiologists [6], and mathematicians [7] globally. Viral transmission typically exhibits temporal and spatial characteristics [8-10]. Due to the incubation period of the virus [11], infected patients often require time to develop symptoms after infection, and their physical movement during this period accelerates viral spread and diffusion across geographic space. Recording epidemic information that changes over time and space (i.e., spatiotemporal information) helps understand transmission speed and scope, thereby providing crucial data sources for monitoring, predicting, and controlling epidemics [12].

Currently, 2019-nCoV epidemic data in China are primarily released by authoritative institutions, including the National Health Commission (NHC) and provincial health commissions. However, data released by national and provincial authorities exist only as text announcements without association with geographic spatial data, making spatial analysis inconvenient. Although platforms such as DXY.cn<sup>1</sup> and People's Daily<sup>2</sup> have implemented real-time dynamic visualization of the 2019-nCoV epidemic by crawling data from national and provincial health commissions, these platforms only display real-time data, making it difficult to obtain historical spatiotemporal data on epidemic development. Therefore, this study aims to provide authoritative, open-access, and multi-scale spatiotemporal datasets of the novel coronavirus (2019-nCoV) for the broader research community, facilitating contributions from global scientists to combat this epidemic.

The structure of this paper is as follows: Chapter 2 introduces the data sources and characteristics of the 2019-nCoV spatiotemporal dataset, while Chapter 3 presents typical applications of the dataset.

## 2. Data Description

The organizational structure of the 2019-nCoV spatiotemporal dataset is illustrated in Figure 1 [Figure 1: see original paper]. The dataset encompasses three different scales: global scale, national scale (provincial), and national scale (prefecture-level city). Each scale includes both vector and attribute data. Vector data refer to administrative polygon features of different countries or regions, comprising global countries, Chinese provincial administrative divisions, and Chinese prefecture-level city administrative divisions. The national prefecture-level city administrative divisions were created by merging, clipping, reprojecting, and dissolving county-level administrative polygon features from the National Fundamental Geographic Database (current as of 2015, scale 1:1,000,000) released by the National Geomatics Center of China. Provincial administrative divisions were directly adopted from the same database, while global country administrative divisions were extracted from high-resolution polygon data in GADM Ver. 3.6 (released May 6, 2018). Vector data employ the WGS-84 geodetic coordinate system with latitude and longitude coordinates in shapefile format, directly compatible with GIS platforms such as ESRI ArcGIS. Attribute data refer to epidemic data released daily by authoritative institutions for each country or region, including global epidemic data, domestic epidemic data (provincial), and domestic epidemic data (prefecture-level). Each dataset contains six statistical tables: newly confirmed cases, newly cured cases, newly deceased cases, total confirmed cases, total cured cases, and total deceased cases. Each statistical table includes fields for Region\_CN, Region\_EN, and T+YYYY+MM+DD, where YYYYMMDD represents the statistical date of the epidemic data, and the field attribute represents the daily epidemic data (-1 indicates no data available for that day). Data are stored in .txt text format, directly compatible with Microsoft Excel and ESRI ArcGIS.

### 2.1 Data Sources

Vector data include both national and global datasets. The national vector dataset comprises administrative boundary vector data for all Chinese provinces and prefecture-level cities, provided by the National Geomatics Center of China and freely available through the National Geographic Information Resources Directory Service System (<http://www.webmap.cn/commres.do?method=dataDownload>). Global vector data include administrative boundary vector data for countries and regions worldwide, obtained from the Database of Global Administrative Areas (GADM: [https://gadm.org/download\\_world.html](https://gadm.org/download_world.html)), which is a free, high-precision global administrative boundary database [13].

Attribute data include both national and global datasets. National attribute data refer to epidemic data for all Chinese provinces and prefecture-level cities (including six fields: newly confirmed, newly cured, newly deceased, total confirmed, total cured, and total deceased), obtained from 2019-nCoV epidemic announcements released by national and provincial health commissions. Specifically, provincial and prefecture-level case data were collected from provincial

health commission announcements, while data for Hong Kong, Macau, and Taiwan were obtained from the National Health Commission. Global attribute data refer to epidemic data for various countries, sourced from People's Daily epidemic announcements. Specific sources for attribute data are listed in Table 1.

**Table 1** Sources of attribute data for the 2019-nCoV spatiotemporal dataset

Region	Data Source
Beijing Municipal Health Commission	<a href="http://wjw.beijing.gov.cn/xwzx_20031/wnxw/">http://wjw.beijing.gov.cn/xwzx_20031/wnxw/</a>
Tianjin Municipal Health Commission	<a href="http://wsjk.tj.gov.cn/col/col14/index.html">http://wsjk.tj.gov.cn/col/col14/index.html</a>
Hebei Provincial Health Commission	<a href="http://wsjkw.hebei.gov.cn/index.do?templet=new_list&amp;cid=14">http://wsjkw.hebei.gov.cn/index.do?templet=new_list&amp;cid=14</a>
Shanxi Provincial Health Commission	<a href="http://wjw.shanxi.gov.cn/wjyw102/index.hrh">http://wjw.shanxi.gov.cn/wjyw102/index.hrh</a>
Inner Mongolia Autonomous Region Health Commission	<a href="http://wjw.nmg.gov.cn/xwzx/xwfb/index.shtml">http://wjw.nmg.gov.cn/xwzx/xwfb/index.shtml</a>
Liaoning Provincial Health Commission	<a href="http://wsjk.ln.gov.cn/wst_wsjskx/">http://wsjk.ln.gov.cn/wst_wsjskx/</a>
Jilin Provincial Health Commission	<a href="http://wjw.jlcity.gov.cn/gsgg/">http://wjw.jlcity.gov.cn/gsgg/</a>
Heilongjiang Provincial Health Commission	<a href="http://wsjkw.hlj.gov.cn/index.php/Home/Zwgg/all/typeid/42">http://wsjkw.hlj.gov.cn/index.php/Home/Zwgg/all/typeid/42</a>
Shanghai Municipal Health Commission	<a href="http://wsjkw.sh.gov.cn/xwfb/index.html">http://wsjkw.sh.gov.cn/xwfb/index.html</a>
Jiangsu Provincial Health Commission	<a href="http://wjw.jiangsu.gov.cn/col/col7290/index.html">http://wjw.jiangsu.gov.cn/col/col7290/index.html</a>
Zhejiang Provincial Health Commission	<a href="http://www.zjwjw.gov.cn/col/col1202101/index.html">http://www.zjwjw.gov.cn/col/col1202101/index.html</a>
Anhui Provincial Health Commission	<a href="http://wjw.ah.gov.cn/news_list_477_1.html">http://wjw.ah.gov.cn/news_list_477_1.html</a>
Fujian Provincial Health Commission	<a href="http://wjw.fujian.gov.cn/ztlz/gzbufk/yqtb/">http://wjw.fujian.gov.cn/ztlz/gzbufk/yqtb/</a>
Jiangxi Provincial Health Commission	<a href="http://hc.jiangxi.gov.cn/xwzx/wjxw/index.shtml">http://hc.jiangxi.gov.cn/xwzx/wjxw/index.shtml</a>
Shandong Provincial Health Commission	<a href="http://wsjkw.shandong.gov.cn/ztlz/rdzt/qlzhfkz/tzgg/">http://wsjkw.shandong.gov.cn/ztlz/rdzt/qlzhfkz/tzgg/</a>
Henan Provincial Health Commission	<a href="http://wsjkw.henan.gov.cn/channels/854.shtml">http://wsjkw.henan.gov.cn/channels/854.shtml</a>
Hubei Provincial Health Commission	<a href="http://wjw.hubei.gov.cn/fbjd/dtyw/">http://wjw.hubei.gov.cn/fbjd/dtyw/</a>
Hunan Provincial Health Commission	<a href="http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw_1/index.html">http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw_1/index.html</a>

Region	Data Source
Guangdong Provincial Health Commission	<a href="http://wsjkw.gd.gov.cn/zwyw_yqxx/index.html">http://wsjkw.gd.gov.cn/zwyw_yqxx/index.html</a>
Guangxi Zhuang Autonomous Region Health Commission	<a href="http://wsjkw.gxzf.gov.cn/zhuantiqu/ncov/ncovyqtb/">http://wsjkw.gxzf.gov.cn/zhuantiqu/ncov/ncovyqtb/</a>
Hainan Provincial Health Commission	<a href="http://wst.hainan.gov.cn/swjw/rdzt/yqfk/index.html">http://wst.hainan.gov.cn/swjw/rdzt/yqfk/index.html</a>
Chongqing Municipal Health Commission	<a href="http://wsjkw.cq.gov.cn/yqxyqtb/">http://wsjkw.cq.gov.cn/yqxyqtb/</a>
Sichuan Provincial Health Commission	<a href="http://wsjkw.sc.gov.cn/scwsjkw/gzbd01/ztwzlmgl.shtml">http://wsjkw.sc.gov.cn/scwsjkw/gzbd01/ztwzlmgl.shtml</a>
Guizhou Provincial Health Commission	<a href="http://www.gzhfpc.gov.cn/xwzx_500663/zwyw/">http://www.gzhfpc.gov.cn/xwzx_500663/zwyw/</a>
Yunnan Provincial Health Commission	<a href="http://ynswsjkw.yn.gov.cn/wjwWebsite/web/col?id=UU157976428326282067">http://ynswsjkw.yn.gov.cn/wjwWebsite/web/col?id=UU157976428326282067</a>
Tibet Autonomous Region Health Commission	<a href="http://wjw.xizang.gov.cn/xwzx/wsjkdt/">http://wjw.xizang.gov.cn/xwzx/wsjkdt/</a>
Shaanxi Provincial Health Commission	<a href="http://sxwjw.shaanxi.gov.cn/col/col9/index.html">http://sxwjw.shaanxi.gov.cn/col/col9/index.html</a>
Gansu Provincial Health Commission	<a href="http://wsjk.gansu.gov.cn/channel/10910/index.html">http://wsjk.gansu.gov.cn/channel/10910/index.html</a>
Qinghai Provincial Health Commission	<a href="https://wsjkw.qinghai.gov.cn/ztbd/yqjk/yqtb/index.html">https://wsjkw.qinghai.gov.cn/ztbd/yqjk/yqtb/index.html</a>
Ningxia Hui Autonomous Region Health Commission	<a href="http://wsjkw.nx.gov.cn/yqfkd/yqsd1.htm">http://wsjkw.nx.gov.cn/yqfkd/yqsd1.htm</a>
Xinjiang Uyghur Autonomous Region Health Commission	<a href="http://www.xjhfp.gov.cn/ztzl/fkxgzbdfygz/yqtb.htm">http://www.xjhfp.gov.cn/ztzl/fkxgzbdfygz/yqtb.htm</a>
National Health Commission	<a href="http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml">http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml</a>
People' s Daily	<a href="http://society.people.com.cn/GB/369130/431577/index.html">http://society.people.com.cn/GB/369130/431577/index.html</a>

## 2.2 Data Characteristics

This section elaborates on the characteristics of the 2019-nCoV spatiotemporal dataset from four perspectives: authority, completeness, temporal precision, and topological consistency.

**2.2.1 Authority** National epidemic data were obtained from authoritative state institutions—the National Health Commission and provincial health commissions. Global epidemic data (for other countries) were sourced from People' s Daily, a state-authoritative internet media platform.

**2.2.2 Completeness** Vector data cover global land areas, including 253 countries, regions, and islands worldwide; 34 provinces, municipalities, and special

administrative regions in China; and 367 prefecture-level cities, prefectures, or regions. These are organized into three scales: global (country), national (provincial), and national (prefecture-level city). Vector data at each scale record Chinese (field name: Region\_CN) and English (field name: Region\_EN) names for different regions. Attribute data cover 2019-nCoV epidemic data for all countries, Chinese provinces, and cities from December 31, 2019, to the present, comprising 18 tables including newly confirmed, newly cured, newly deceased, total confirmed, total cured, and total deceased cases. As the 2019-nCoV epidemic remains in an outbreak phase, data released by the National Health Commission exhibit certain lags and errors—total confirmed, cured, and deceased cases may be adjusted upward or downward, resulting in discrepancies between the sum of new cases and total confirmed cases. Additionally, since some provinces report newly confirmed cases with unspecified prefecture-level details, all attribute statistics are based on the latest data released by health commissions.

**2.2.3 Temporal Precision** The temporal resolution of this spatiotemporal dataset is one day, counting epidemic data released by national and provincial health commissions for the 0-24 hour period each day. The dataset is updated daily with data released the previous day until health commissions cease publishing 2019-nCoV epidemic information. It is important to note that during the early epidemic stage (December 31, 2019 to January 11, 2020), only Hubei Province (the primary source of the outbreak) released epidemic announcements, so attribute data for other provinces during this period are null values recorded as -1.

**2.2.4 Topological Consistency** To ensure boundary consistency between global and national administrative division data, this study used ArcGIS software's "Erase" tool to replace global administrative boundaries with national borders from the National Fundamental Geographic Database, preventing spatial adjacency analysis errors caused by topological inconsistencies.

### 3. Typical Applications

#### 3.1 Attribute Data Analysis

Figure 2 [Figure 2: see original paper] illustrates the evolutionary trend based on attribute data from the spatiotemporal dataset between January 11 and February 9, 2020. The analysis reveals several key patterns.

Domestically (Figure 2a), total confirmed cases began increasing gradually from January 26 and remained at high levels with continued growth through February 9, reflecting that the domestic epidemic was still in an outbreak phase [14]. Globally (outside China), the top five countries by total confirmed cases (Figure 2b) were Japan, Thailand, Singapore, South Korea, and Vietnam. Except for Japan, other countries showed stable and slow growth in total confirmed cases,

primarily because Japan's numbers increased more rapidly due to the mass infection event on the Diamond Princess cruise ship, while most other countries had sporadic imported cases [15] with few cluster infections, resulting in overall 平缓 growth.

In Hubei Province (Figure 2c), total confirmed cases far exceeded those in other provinces. Although the curve showed signs of moderation by February 9, it maintained a relatively rapid growth rate. Nationally (excluding Hubei), the top five provinces by total confirmed cases (Figure 2d) were Guangdong, Zhejiang, Henan, Hunan, and Anhui, all showing consistent growth trends with gradually flattening rates compared to the initial outbreak phase. This suggests that the city lockdown measures in 13 severely affected regions of Hubei and disease prevention efforts in other provinces have achieved preliminary effectiveness.

In Wuhan (Figure 2e), total confirmed cases exceeded 50% of the provincial total and continued to grow rapidly, indicating that the epidemic situation in Wuhan remained severe and required continued focus. Within Hubei prefecture-level cities (Figure 2f), the top five by total confirmed cases were Xiaogan, Huanggang, Suizhou, Jingzhou, and Xiangyang. The curves for Suizhou, Jingzhou, and Xiangyang gradually stabilized after February 2, while Xiaogan and Huanggang continued to see increases, though their growth rates began slowing on February 4 and February 6, respectively.

### 3.2 Spatiotemporal Data Analysis

Using ArcGIS software's "Join" tool, attribute data from the dataset can be directly combined with vector data to produce multi-scale, multi-temporal mapping and visualization of epidemic distributions. This study used attribute data from January 22 to February 5, 2020 as examples to create epidemic distributions at three scales: global (Figure 3 [Figure 3: see original paper]), national provincial (Figure 4 [Figure 4: see original paper]), and national prefecture-level (Figure 5 [Figure 5: see original paper]). The results demonstrate distinct patterns at each scale.

Globally (Figure 3), the transmission pattern of 2019-nCoV shows China as a radiation source spreading rapidly to neighboring countries and developed nations. In the early stage (Figure 3a), only a few countries such as South Korea, Japan, Thailand, and the United States reported confirmed cases. As the epidemic spread, additional countries including Singapore, Russia, Canada, Australia, and Germany reported cases. This pattern arises because, on one hand, proximity increases correlation between locations [16], making neighboring countries severely affected; on the other hand, although developed countries like Australia, the United States, and European nations are geographically distant from China, their close connections with China also resulted in severe outbreaks.

Nationally at the provincial level (Figure 4), the epidemic occurred primarily in provinces east of the Hu Huanyong Line, with Guangdong, Zhejiang, Hunan, and Anhui being the most severely affected. In the early stage (Figure

4a), the outbreak was concentrated in Beijing, Shanghai, Zhejiang, and Guangdong. Subsequently, the epidemic began spreading from Hubei to neighboring provinces, with significant increases in confirmed cases in Jiangxi, Hunan, Chongqing, Anhui, and Henan. In the middle and later stages (Figure 4c), severely affected areas were mainly distributed in central and eastern China.

At the prefecture-level city scale (Figure 5), considering Wuhan's lockdown on January 23, this study divides the spatiotemporal evolution into three levels: Wuhan, cities within Hubei, and areas outside Hubei. At the Wuhan level, the epidemic changed extremely rapidly, with total confirmed cases exceeding 1,000 by January 27. Within Hubei Province, spatiotemporal transmission primarily showed a radiation pattern from Wuhan to other cities. Nationally, Beijing, Shanghai, Guangzhou, and Shenzhen were the most severely affected areas outside Hubei. These patterns occurred because, on one hand, over 5 million people had left Wuhan before the lockdown, with most flowing to counties within Hubei Province, resulting in more cases in Hubei's prefecture-level cities than in other provinces; on the other hand, first-tier cities like Beijing, Shanghai, Guangzhou, and Shenzhen have larger populations and higher mobility, leading to more confirmed cases than other prefecture-level cities outside Hubei.

#### 4. Conclusions and Outlook

This study provides an authoritative, open-access, and multi-scale 2019-nCoV spatiotemporal dataset for the broader research community, addressing the lack of an official channel for real-time open-source epidemic data. The dataset includes three scales—global, national (provincial), and national (prefecture-level)—with each scale recording multi-temporal epidemic information including newly confirmed, newly cured, newly deceased, total confirmed, total cured, and total deceased cases. Based on data source analysis, this study demonstrates that the dataset not only provides essential data for epidemic monitoring, prevention, prediction, and early warning but also supports multi-scale, multi-temporal mapping and visualization of the 2019-nCoV epidemic, offering valuable resources for analyzing spatial distribution, evolution, trends, and simulation forecasting.

#### References

- [1] Li X, Zai J, Wang X, et al. Potential of large 'first generation' human-to-human transmission of 2019-nCoV[J]. *Journal of Medical Virology*, 2020.
- [2] Corman V M, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR[J]. *Eurosurveillance*, 2020, 25(3).
- [3] Benvenuto D, Giovannetti M, Ciccozzi A, et al. The 2019-new coronavirus epidemic: evidence for virus evolution[J]. *Journal of Medical Virology*, 2020.
- [4] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China[J]. *The Lancet*, 2020.
- [5] Chan J F, Yuan S, Kok K, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a

- study of a family cluster[J]. The Lancet, 2020.
- [6] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia[J]. New England Journal of Medicine, 2020.
- [7] Zhao S, Musa S S, Lin Q, et al. Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak[J]. Journal of Clinical Medicine, 2020, 9(2): 388.
- [8] Carroll M W, Matthews D A, Hiscox J A, et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa[J]. Nature, 2015, 524(7563): 97-101.
- [9] Zhong S. Application of GIS and Remote Sensing in Infectious Disease Epidemiology[D]. Graduate School of the Chinese Academy of Sciences (Institute of Remote Sensing Applications), 2006.
- [10] Zulu L C, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010[J]. BMC infectious diseases, 2014, 14(1): 285.
- [11] Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany[J]. New England Journal of Medicine, 2020.
- [12] Guo F, Lu Z. Progress in Disease Geography Research Based on Spatial Analysis Methods[J]. Geographic Information World, 2009, 7(06): 22-26.
- [13] Hijmans R, Garcia N, Wiczorek J. Global Administrative Areas Database (GADM) Version 3.6[J]. UN: New York, NY, USA, 2018.
- [14] Chen Z, Zhang W, Lu Y, et al. From SARS-CoV to Wuhan 2019-nCoV: Will History Repeat Itself?[J]. bioRxiv, 2020.
- [15] Read J M, Bridgen J R, Cummings D A, et al. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions[J]. medRxiv, 2020.
- [16] Tobler W R. A computer movie simulating urban growth in the Detroit region[J]. Economic geography, 1970, 46(sup1): 234-240.

**Corresponding Author:** Hao Lin

**E-mail:** cug\_bird@cug.edu.cn

**Author Contributions:**

Hao Lin: Conceived research ideas and designed the study;

Junya Bao: Collected, cleaned, organized, produced, and updated data;

Hao Lin: Conducted data application and wrote the manuscript.

---

<sup>1</sup> <https://ncov.dxy.cn/ncovh5/view/pneumonia>

<sup>2</sup> <http://society.people.com.cn/GB/369130/431577/index.html>

<sup>3</sup> <https://github.com/Estelle0217/COVID-19-Epidemic-Dataset>

<http://qianxi.baidu.com/>

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv – Machine translation. Verify with original.*