# ReadMe File for China's Surface Transport System Database over 1993-2020

Steven J. Davis and Meijun Qian
February 2025

This ReadMe file introduces our GIS database: China's Surface Transport System Database. It has comprehensive coverage for the period from 1993 to 2020, as described in Davis, Qian and Zheng (2025) and includes data for earlier periods developed by other researchers. The database contains three types of datasets:

(1) GIS datasets include a series of maps in the Geographic Information System (GIS) for all surface transport modes each year. These GIS maps are suitable for anyone who wants to directly access the geocoded data using ArcGIS in order to replicate and extend our measures of transport features or construct their own transport quality measures. They run from 1993 to 2020.

(2) Excel files contain rich statistical and categorical data on transport system features. We generate these files from the GIS maps and make it convenient for anyone unfamiliar with ArcGIS. The data in these files pertains to the period from 1993 to 2020.

(3) Data on roads and railroads for earlier periods from Baum-Snow et al. (2017), reproduced with permission.

For more information, see Davis and Qian (2021) and Davis, Qian and Zeng (2025abc). For more information on the data for roads and railrods before 1993, see Baum-Snow et al. (2017).

**Our China Surface Transport System Database is freely accessible. Database users should cite Davis, Qian and Zeng (2025a) as the source. Researchers who make use of the pre-1993 data on roads and railroads should cite Baum-Snow et al. (2017).**

We thank the Booth School of Business at the University of Chicago, the University of Chicago Library's map section, and the Becker Friedman Institute for financial support and technical assistance during the years-long development of the Database.

## A. Definitions of Surface Transports in Our GIS Database

**Motorways** refer to roads in an enclosed and access-controlled high-speed road system. A motorway has at least two lanes in each direction of traffic and a speed limit of 80 to 120 km/h. According to the prefectural administration, motorways are classified as national or provincial highways.

**Highways** are roads within the highway network that are not motorways. Technically, highways include national primary, secondary, and tertiary roads. Most highways in the Eastern and coastal regions have 2-2 or 3-3 lanes and a speed limit of 60-80 km/h in China. In China's Central and Western regions, highways are 2-2 or 1-1 tertiary roads with a speed limit as low as 20 km/h.

**Regular railways** encompass standard and fast railways, serving passenger and freight lines. In recent years, standard railways have maximum speeds below 160 km/h, and fast railways have maximum speeds of 160-250 km/h. The classification criteria of regular railways varied in earlier years, with six railway speed-up campaigns conducted from 1997 to 2007. Our identification and classification of railways follow the documents of railway additions and speed improvements disclosed by the National Railway Administration. We then further verify the official information with geocoded maps of the railway system.

**High-speed railways** are passenger-only railways with maximum speeds of 250-350 km/h in most areas and 200-250 km/h for some regional connecting lines.

**Waterways** are navigable waterway routes based on the waterway map in 1994. They include all graded inland and coastal lines navigable for ships with six capacity levels: more than 10,000 tons, 1,000-10,000 tons, 500-1,000 tons, 300-500 tons, 100-300 tons, and 50-100 tons. Compared to other surface transport modes, we do not track year-to-year changes in features for waterways since the waterway system has had little changes over the past decades.

**Historical roads and railroads** are the relevant transports derived from road and railroad maps in China in 1962, 1980, and the 1990s. Credits go to Baum-Snow, Brandt, Henderson, Turner, and Zhang (2017) for generously sharing the data on their website.

**Access points** refer to exits and entrances for motorways and stations for railways and high-speed railways.

**Terrain** measures its ruggedness based on remote-sensing data collected by satellite images. The data are from the National Geographic Information database managed by the Ministry of National Resources in China.

**B. GIS Files and Study Period**

The GIS dataset contains one shapefile in ArcGIS for each transport mode in each year. The geographical information of entrance and exit ramps, regular railway stations, and high-speed railway stations is stored in separate shapefiles. The GIS dataset also includes data on locations and boundaries at county, city, province, region, and nation levels, which could be used to plot county-level statistics, make heat maps, and compute transport costs between locations.

There are a total of 2,462 files in our GIS database named "**China Surface Transport Database**": 1,575 transport shapefiles in the folder "**GIS Transport China Shapefiles**", 42 transport shapefiles in the folder "**Old Maps**", 12 shapefiles and 1 Excel file in folder "**Boundary**", and 7 transport shapefiles in folder "**Terrain**", 816 transport shapefiles in folder "**Single Lines**", and 9 Excel files in folder "**Transport Measures**". All shapefiles can be layered onto each other in the GIS and accompanied by other supporting files in GIS. Some shapefiles have two lines for directions of tracks or lanes for part of transport networks built in recent years. The "**Single Line**" folder contains re-generated shapefiles in which transport routes are combined into a single line. While the folder and file names for these shape files are self-explanatory, the tables below illustrate how data are organized into the folders and files.

We conduct analyses using another file that contains geocoded locations with longitudes and latitudes of all 35,798 zip codes in China, 34,889 (97.5%) of which have manufacturing firms. This file helps merge socioeconomic, ecological, political, and environmental data with address information. Please contact us (meijunqian@gmail.com) for access to the data.

**Table R1. The categories of data/folders in "China Surface Transport Database"**

| Data category | Folder/File name | Content |
|---|---|---|
| Transport | GIS Transport China Shapefiles | (1) transportation routes<br>(2) access points: exit/entrances and stations |
| | Single Lines | Multiple lanes/tracks on the routes combined into single lines |
| Historical transport | Old Maps | Historical routes of railway and roads in 1962, 1980, and 1990s |
| Boundary | Boundary | Boundaries at different administrative levels |
| Terrain | Terrain | Elevation information at the county level |
| Statistical data | Transport Measures | (1) Transport quality measures calculated based on transport data |

| | | |
|---|---|---|
| | | (2) Socioeconomic variables used for analyses in three working papers |
| Manuscript | Working Papers | Three working papers that support the development of the GIS database |
| ReadMe | Readme | Explanatory notes for the GIS database |
| Technical note | Technical Details of Digitalization and Rectification | Notes for the technical details during the digitalization and rectification of transport data |

*Note*: This table lists all data categories, folder/file names, and their contents in our GIS database.

**Table R2. Transport information in the folder of "Single Line"**

| Transport modes | Folder/File name | Information | Year |
|---|---|---|---|
| Motorway | Motorway | Transport routes | 1993-2020 |
| | MotorwayAccess | Entrances and exits | 1993-2020 |
| Highway | Highway | Transport routes | 1993-2007, 2013-2020 |
| Regular railway | RegularRailway | Transport routes | 1995-2020 |
| | RailwayStation | Railway Stations | 1995-2020 |
| | RailwaySpeedup | Transport routes with speed attributes | 1997, 1998, 2000, 2011, 2004, 2007 |
| High-speed railway | HighspeedRailway | Transport routes | 2003-2020 |
| | RailwayStation | Railway Stations | 2003-2020 |
| Waterway | Waterway | Transport routes | 1993-2020 |
| Historical road | Old Maps | Road routes | 1962, 1980, 1990 |
| Historical railway | Old Maps | Railway routes | 1962, 1980, 1990 |

*Note*: This table lists all transport modes, folder/file names, transport information, and periods included in the folder "Single Line" in our GIS database.

### C. Digitization and Rectification

We implement a multi-step process to create reliable and ready-to-use GIS datasets. The first three steps involve the map digitalization. The following three steps rectify errors inherited from the maps and other data sources. The last step is to merge with additional information, such as road, railroad speed, and types and grades of stations. The technical notes of detailed steps for this process are in the "**Technical Details of Digitization and Rectification**" file**.**

### D. Statistical Data in Excel Files

The data files are in the "**Transport Measures**" folder. The definitions of variables are in the file "**Variable Definitions Transport Statistical Data**". There are three sets of variables: (a) terrain feature measures, (b) distances to transport access points, and (c) transport quality measures: density, connectivity, and centrality. Some statistical measures of transport quality are introduced

by Davis et al. (2025a). More details about methods and statistical descriptions can be found in Davis et al. (2025a).

**Table R3: Excel files in the folder of "Transport Measures"**

| File name | Variable group | Coverage |
|---|---|---|
| Zipcode Panel Access and Discrepancy | Access, distance, and distance discrepancy across different measurements | A panel at zipcode-year level over 1993-2020 |
| County Panel Transport Quality | Transport quality measures: density, connectivity, and centrality | A panel at the county-year level over 1993-2020 |
| County Terrain | Terrain measures: mini, max, mean, and variance of terrains | One cross-section for all counties |
| City Panel Transport Quality | Transport quality measure: density, connectivity, and centrality | A panel at the city-year level over 1993-2020 |
| Province Panel Transport Quality | Transport quality measure: density, connectivity, and centrality | A panel at a province-year level over 1993-2020 |
| Region Panel Transport Quality | Transport quality measures: density, connectivity, and centrality | A panel at a region-year level over 1993-2020 |
| Country Over Years Transport Quality | Transport quality measure: density, connectivity, and centrality | Time series data at the national level over 1993-2020 |
| Country Average over year Access Discrepancy | Distance discrepancy between different measurements | Time series data at the national level over 1993-2020 |

*Note*: This table lists all Excel files, variable groups, and their coverage, which are included in the folder "Transport Measures" in our GIS database.

# References


Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics,* 99(3), 435-448.

Davis, Steven J., Meijun Qian, and Wen Zeng (2025a). "A Comprehensive GIS Dataset for China's Surface Transport System with Implicaiton sfor Transport and Sociconomics Research," working paper, Stanford Univesity.

Davis, Steven J., Meijun Qian, and Wen Zeng (2025b). "Technical Details of Digitization and Rec," technical note, Stanford Univesity.

Davis, Steven J. Jiajun Lu, Meijun Qian, and Wen Zeng (2025c). "Understanding the Development of China's Transport System," working paper.

Davis, Steven J. and Meijun Qian (2021). "Transport Infrastructure and Productivity: The China Experience," working paper, University of Chicago.