

Technical Details of Digitization and Rectification

By Steven J. Davis, Meijun Qian, and Wen Zeng

February 2025

Database: China Surface Transport Database, 1993-2020

We detail a seven-step data processing procedure to generate a reliable and ready-to-use GIS database for all surface transports in China. The first three steps complete digitalization, followed by another three steps to rectify various errors inherited from data sources of digital open street maps. The last step is to add additional information such as road, railroad speed, and types and grades of stations. The entire procedure is introduced below:

Step 1: Collect physical and digital maps for transport modes for each year

- (i) Obtain the raw GIS data in 2013 and 2020 from the *OpenStreetMap*.
- (ii) Use China Road Atlas published by China Communications Press and Atlas of China from Sino-maps Press For year-by-year physical maps of roads over 1993-2020.
- (iii) Obtain other maps from online bookstores, such as Kongfuzi and Jiushujie, and second-hand book vendors on the E-commerce platform, such as Taobao¹.

Step 2: Merge physical maps of all transport modes into GIS

- (i) Scan and import each physical map into the ArcGIS interface and georeferenced 20 controlling points with known geographic coordination for positioning.
- (ii) Adjust the coordinate system *GCS_WGS_1984* without projection in *OpenStreetMap*.
- (iii) Layer these imported maps before 2013 with the GIS map 2013 and delete lines and links in 2013 maps but not in the scanned maps. We assume historically existing routes have the same road features if overlapped with the current ones.
- (iv) Manually add road lines and links that are in the scanned maps but not in the 2013 maps.
- (v) Repeat the process for each transport mode and 20 years over 1993-2012, as well as six railway speed-up campaigns.

¹ Kongfuzi: www.kongfz.com; Jiushujie: www.jiushujie.com; Taobao: www.taobao.com

- (vi) Repeat the above process (iii)-(v) for maps during 2014-2020 based on the digital raw data in 2020. We also verify the results by working forward from 2013 map by adding each year's new routes.
- (vii) Process railways over 2007-2020 based on the digital raw data in 2020 and verify new additions with railroads launching announcements by China's National Railway Administration (NRA). For 1995-2007, we build from the physical map in 2005, add the yearly official announcements of route changes, and use the above process for digitalization.
- (viii) The timing and routes of railway speed-up are available in publications by the NRA of China.

Step 3: Extract and rectify transport data from class attributes

- (i) Extract “*motorway*” and “*motorway_link*” data for the motorway system.
- (ii) Extract “*trunk*” and “*trunk_link*” in *OpenStreetMap* data for highway.
- (iii) Extract “*rail*”, “*funicular*”, “*monorail*”, “*narrow gauge*”, “*railway station*”, and “*railway halt*” data for railway.
- (iv) To separate high-speed railways from other railways, first, identify high-speed railways in the year 2020 based on the China Road Atlas in 2020 (SinoMap Press) and verify with the Mid-Term and Long-term Railway Network Development Program by the NRA in 2016.
- (v) Compile operation launching dates for high-speed railway lines based on official announcements and build the digital high-speed railway system over 2003-2020.
- (vi) Create the railway system (standard and fast) by excluding high-speed rail lines but keeping the lines that are shared by both systems.
- (vii) Keep all data for shapefiles that contain only a single map type, e.g., waterway and terrain.

Step 4: Rectify attribute selections to improve the accuracy of results from Step 3

- (i) Include attributes data of primary and secondary roads in the national highway system by referencing physical maps and BaiduMap for highways as “*trunk*” and “*trunk_link*”.
- (ii) Rectify disconnected roads in the national highway system to ensure smooth connections of roads.

- (iii) Rectify railways in four aspects: (1) exclude subways, light rails, and tram networks in metropolitan cities; (2) add back the railway connections lost in separating high-speed railways from railways; (3) assure the appearance in both railway and high-speed railway shapefiles for rail lines that allow both high-speed and fast trains; (4) add into the high-speed railway shapefiles the standard rail lines where high-speed trains travel.

Step 5: Rectify entrances, exits, and stations

First, some intersections without exits/entrances may also appear as additional points in the *OpenStreetMap*. We refer to actual maps and Baidu maps to delete these wrongly recorded points. Second, the “*motorway_link*” in the class attribute of motorway layers captures the shape of ramps for entrance and exit from highways. The ramp shape is troublesome for computing transport access, distance, and density when applying the GIS map data for network analyses. Therefore, we modify the entrance/exits ramp data into points in the transport network. Nevertheless, we keep the shapes of the original ramps in separate shapefiles as ramps contain helpful information for other research purposes. For railway stations, we obtain information from the National Railway Administration and 12306 China Railway Administration, delete the wrongly indicated station points on maps, and add back stations that were missed. The same is done for the high-speed railway stations.

Step 6: Rectify the connectivity accuracy

The ArcGIS software can detect the topology dangles automatically. We verify actual maps and identify whether the dangles are (i) actual disconnections, (ii) interruptions of road categorization, and (iii) errors by the digital map. Accordingly, in case (ii), primary or secondary roads at relevant sections are added to the database. In case (iii), we add lines to the data and rectify the connectivity. We do the same for motorways, highways, regular railways, and high-speed railways.

Step 7: Add speed attributes for roads and railroads

To facilitate computing transport cost/time, we use the legally allowed maximum speed of 80 km/h for the national highway network. Based on the official Railway Passenger/Freight Transport Price and Odometer information, we find the legally allowed railway speed limits in 2020: 120 km/h, 140 km/h, 160 km/h, 200 km/h, 250 km/h, 300 km/h and 350 km/h for each section. The speed

information is adjusted based on the six-speed-up campaign over 1995-2007. Finally, we add the available railway station types and grades in the attributes using official categorizations from the National Railway Administration (NRA) and 12306 China Railway Administration.

Caveats in utilizing the database:

Attributes for routes before 2013 are assumed to be the same as in 2013, as the road system in GIS before 2013 was built on the GIS data in 2013. The difference over time is detected if we find available attribute information from historical maps, the National Road Administration, or the National High Network Program. It is similar for the maps between those in 2014-2019 and those in 2020. We use the geographical coordinate system of *GCS_WGS_1984* in our GIS database, which is the most common one by *OpenStreetMap*. ArcGIS has its algorithm to compute distances. If projecting is needed for computing other variables, we use *Krasovsky_1940_Albers* projection, a conical projection that is relatively ideal for large areas. Applying the dataset to other studies may choose different or no projections based on specific research purposes. We also manually add designated legal speed limits to road and railroad sections to facilitate computing transport costs and time. The speed information is indicative only. The railway speed from 2007 onwards has an upward bias because of its proxy with 2020 speed. Therefore, the transport costs/time computed with this information are the lower-bound estimations.