NBER WORKING PAPER SERIES

A COMPREHENSIVE GIS DATABASE FOR CHINA'S SURFACE TRANSPORT NETWORK WITH IMPLICATIONS FOR TRANSPORT AND SOCIOECONOMICS RESEARCH

Steven J. Davis
Meijun Qian
Wen Zeng

Working Paper 33515
http://www.nber.org/papers/w33515

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2025

We gratefully acknowledge financial support from the Booth School of Business at the University of Chicago, the Australian National University, and the Australian Research Council (ARC DP190103511). We thank Sam Brown in the Map Section at the University of Chicago Library for expert assistance in the early development of our GIS database. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33515

A Comprehensive GIS Database for China's Surface Transport Network with Implications for Transport and Socioeconomics Research
Steven J. Davis, Meijun Qian, and Wen Zeng
NBER Working Paper No. 33515
February 2025
JEL No. L92, N75, O18, R40

ABSTRACT

We build a granular Geographic Information System (GIS) database that covers China's national highways, modern motorways, traditional railways, high-speed railways, and waterways at an annual frequency from 1993 to 2020. After describing the database and its construction, we characterize the development of China's surface transport network. Overall network length more than tripled after 1993. Modern motorways and high-speed railways, nearly non-existent in the early 1990s, account for half the increase in network length. The average distance between county centroids and transport access points fell by more than half, from 13.28 km in 1993 to 5.81 km in 2020. Common county-level measures of connectivity and centrality rose by 22% and 115%, respectively. We also show that discrepancies between distance to motorway access and straight-line distance to motorway routes are often large, and the discrepancies correlate with calendar time, terrain features, and economic development. Because motorway access is vital to local economic development in China, this finding raises the need to re-assess previous research that uses straight-line distance to proxy actual distance. Our GIS database is freely available on an open-access basis, creating an empirical laboratory for new research in multiple directions.

Steven J. Davis
Hoover Institution
Stanford University
434 Galvez Mall
Stanford, CA 94305
and NBER
stevend5@stanford.edu

Meijun Qian
Research Institute,
People's Bank of China
meijunqian@gmail.com

Wen Zeng
Australian National University
Wen.Zeng.001@gmail.com

I. Introduction

Transport networks shape the spatial distribution of economic activity and influence productivity, trade flows, prices, urbanization patterns, geographic mobility, and more.[1] China's surface transport network has grown tremendously in scale and quality in recent decades. To track this development, we build a granular Geographic Information System (GIS) database that covers China's modern motorways, regular railways, high-speed railways, highways, and waterways at an annual frequency from 1993 to 2020.

Partial snapshots of China's transport network in previous studies show its growing scale and complexity over the past century (Wang et al., 2009; Jiao et al., 2014; Hu et al., 2015; Jin et al., 2019). While less relevant for freight and low-wage workers, high-speed railways and air travel growth in recent decades have greatly improved connectivity between Chinese cities (Lao et al., 2016; Jiao et al., 2017). Relative to earlier work, our GIS database is more granular, more comprehensive in its coverage of surface transport modes, more attentive to route access locations, and better suited for tracking year-to-year changes in motorway, railway, and highway transports. We also incorporate travel speed estimates by travel modes and route segments.

After describing the construction of our database, we characterize China's surface transport network and quantify its development in multiple aspects. Overall network length rose from 184,000 km in 1993 to 563,000 km in 2020. Half the increase took the form of modern motorways and high-speed railways, which were nearly nonexistent in the early 1990s. The average straight-line distance between zip code centroids and their nearest access points fell from 13 km to 5 km over 1993-2020, while the average county-level connectivity nearly doubled in this period.[2] In addition, route density rose tremendously in recent decades, as with other transport quality measures.

We also show that straight-line distances from zip code centroids to the nearest route deviate considerably from distances to actual access points for motorway transports. The

[1] Studies on how transport networks affect trade, productivity, worker mobility, income, and economic growth include Eaton and Kortum (2002), Redding and Venables (2004), Faber (2014), Allen and Arkolakis (2014), Donaldson and Hornbeck (2016), Donaldson (2018), and Banerjee et al. (2020). The World Bank (2000, 2005), Sachs (2005), and Smith (2005) devote particular attention to market access and poverty alleviation. Studies that focus on land use, resource access, ecosystems, biodiversity, and environmental degradation include Kahn et al. (2007), Chapman (2007), Weinhold and Reis (2008), Benítez-López et al. (2010), Laurance (2013, 2014), and Ibisch et al. (2016). In addition, see Baum-Snow (2007), Duranton and Turner (2012), and Zheng & Kahn (2013) for urban development and regional integration. See Woodcock et al. (2009), Hystad et al. (2013), and Jiang et al. (2017) for public health effects. See Litman (2002), Lucas (2012), and Grieco and Urry (2016) for social equality and inclusiveness. See Weiss et al. (2018) and Rodrigue (2020) on transport system development.

[2] We define access points as the motorway entrances/exits and railway stations.

distance discrepancy correlates with calendar time, terrain features, and economic development, suggesting potential bias in estimating the causal effects of transport improvements on the local economy using distance-to-route measures.

Many studies have investigated the effects of transport treatment on socioeconomic outcomes in China. Zheng and Khan (2013), Qin (2017), and Lin (2017) find that high-speed railways facilitate market integration and urban employment, raise house prices near passenger stations, but reduce economic output along upgraded railway routes. Faber (2014) finds that China's national trunk highways affect the concentration of economic activity and lower growth in non-targeted peripheral counties. Zhang and Ji (2019) discover that the development of railways and roads could either raise or lower local GDP, depending on how new transports alter local competition with nearby areas. Banerjee et al. (2020) find proximity to railways and waterways has modest positive effects on local GDP per capita but not on its growth rate. Baum-Snow et al. (2020) argue that better access to national highways raises economic output and population in regional urban centers at the expense of hinterland areas. Our own earlier study suggests that better access to high-quality transport raises the productivity of manufacturing plants but lowers productivity dispersion between plants in China (Davis and Qian, 2021).

There is much scope for further research into the effects of China's transport network on the spatial distribution of production, supply chains, productivity, geographic mobility, poverty, and environmental outcomes. Given the large size of its economy, transport in China also influences global trade patterns and the economies of other countries. Therefore, the role of economic and political forces in driving transport development warrants much attention. Our paper and open-access database serve as inputs into future studies on these topics. The database covers multiple transport modes with high geospatial granularity over 28 years, thus making it well-suited for long-span longitudinal studies. It can be integrated with other GIS data sources of local streets and facilities, as in Davis and Qian (2021).

Section II defines transport modes, describes data sources, and explains the digitization of spatial information. Section III offers graphical and statistical analyses of the temporal evolutions of transports in China. Section IV documents the distinction between transport routes and access points and the growth of "pass-through" counties with motorway and high-speed railway routes but no access points. Section V sketches possible applications of our database in future research, and Section VI concludes.

II. Database Construction

Table 1 summarizes the main elements of our database. All shapefiles can be layered onto each other in the GIS and used for analysis.

II.1 Transport Modes and Coverage

(1) Motorway

A motorway in our database refers to a high-speed and access-controlled roadway with at least two traffic lanes in each direction and speed limits ranging from 80 to 120 km/h.[3] Some documents refer to motorways as expressways or freeways, and they are often regarded as parts of a national and provincial highway system from an administrative perspective.[4] Our GIS database records all operational motorways over 1993-2020. The shapefiles contain information on motorway routes, their maximum speeds, and entry and exit ramp locations. The first operating motorway opened in 1988 in China, from Shanghai to Jiading, and other major motorway projects began in the mid-1990s.[5] By the end of 2020, the total length of motorway transport reached 161,000 km in China, making it the longest national motorway network in the world (Ministry of Transport, 2020).

(2) Highway

Highways are a mix of primary, secondary, and tertiary roads that fill out the national highway system in China, including provincial and county roadways. Most highways in the Eastern region have 2-by-2 or 3-by-3 lanes with a speed limit of 60-80 km/h. In the Central and Western regions, highways are 2-by-2 or even 1-by-1 lanes, with a speed limit as low as 20 km/h. Highways usually connect smaller towns and are freely accessible, whereas motorways connect larger prefectural cities and ports and typically require tolls. Our GIS database digitizes all operational highways and manually collects speed limits for highway segments from 1993 to 2020. Since additions to highways primarily occurred after 2012, we adopt a GIS map in 1993 to represent the highway networks over 1993-2012, combined with year-to-year changes over 2013-2020. Our database suggests that motorways and

---

[3] Some motorway segments fall short of meeting these technical requirements in regions with a less developed road system. Typically, these segments involve a short primary road that completes a portion of the motorway network. These instances are infrequent.

[4] Regarding administration, roads in China are divided into national, provincial, and county highways. From the technical aspect, roads include motorways, primary, secondary, tertiary, and fourth roads. Thus, the administrative categorization of roads may only partially align with their technical and functional characteristics. *OpenStreetMap* also refers to roads as “trunk-ways”, which are primarily primary and secondary roads, as well as some tertiary roads in the Western region.

[5] For instance, the "7918" network project launched in 2004, and the "71118" project commenced in 2013 (Zhang et al., 2023).

highways have a total linear distance of 346,062 km in 2020[6] and that the proportion of counties covered by motorway or highway segments increased from 70% in 1993 to 96% in 2020 in China.

(3) Regular (Standard and Fast) Railway

Railways in China include regular (standard and fast) and high-speed railways. A standard railway has a maximum speed of 160 km/h, while a fast railway has a speed from 160 to 250 km/h, covering both passenger and freight routes.[7] We digitize routes of all operational standard and fast railways over 1995-2020 in one shapefile and information on stations in another. Speed limits for each route are also included in data attributes and categorized into six levels of stations for passenger and freight lines. Large-scale railway constructions began in the 1980s after economic reforms. After that, there were six "speed-up" campaigns in China from 1997 to 2007. In 2001, an "Eight Vertical and Eight Horizontal" railway network was launched, substantially expanding railway transport. In addition, we manually collect data on new railway additions and speed improvements from official documents released by the National Railway Administration (NRA).

(4) High-Speed Railway

Unlike regular railways, high-speed railways serve passengers only and have a much higher maximum speed of 250-350 km/h on main lines and 200-250 km/h on regional segments in China.[8] The construction of high-speed railways was first launched in 2003, followed by a large expansion that built high-speed railway segments parallel to the "Eight Vertical and Eight Horizontal" network in 2016. By 2020, there were more than 38,000 km of high-speed railways in China, more than any other country. Our GIS database contains annual maps of all operating high-speed railways over 2003-2020. As with other transports, we digitize GIS information of routes, stations, and speed limits in shapefiles. Combining regular and high-speed railways, the total length of railway transports reached 146,000 km in 2021, secondary only to the United States.[9] However, the extent of the railway network is less impressive in China if its geographic area or population size scales it. For instance, 84 countries have a per-capita railway length that exceeds China.

---

[6] The number computed by our database is very close to a linear distance of 370,700 km reported by the Statistical Bulletin on the Development of the Transport Industry in 2020 based on an administrative classification of roadways. See: http://wap.china-railway.com.cn/ (Last accessed on 2024/06/29)

[7] We manually collect railway speed information from official documents and high-speed railway launch announcements posted by the National Railway Administration (NRA). See: http://www.nra.gov.cn/

[8] China has tested higher speeds since 2014. For instance, the world's fastest maglev train, with a speed of up to 600 km/h, rolled off the assembly line in Qingdao, China, in 2021 but has yet to be placed in operation.

[9] https://en.wikipedia.org/wiki/List_of_countries_by_rail_transport_network_size

(5) Waterway

In China, navigable waterway routes have expanded modestly over the past three decades.[10] Using waterway maps from 1994 that include all inland and coastal routes navigable by ships, we identify six tonnage capacity levels as grade criteria: over 10,000 tons, 1,000-10,000 tons, 500-1,000 tons, 300-500 tons, 100-300 tons, and 50-100 tons. Our database contains one shapefile for waterway routes and another for capacity grades along each route.[11]

(6) Earlier Data on Roadway and Railway

Baum-Snow et al. (2017) study how the growth of roads and railways influenced the urban form of Chinese cities since the early 1990s using data on roadways and railways in 1962 to implement their identification strategy. They generously gave us permission to incorporate their data in 1962 into our database. As of 1962, roadways that formed highway networks typically had two lanes in each direction and operated in an open-access manner in China. Many roadways were designed for only some weather conditions or even paved (Lyons, 1985), and one of their main functions was to move agricultural goods to local markets back then. More than two-thirds of railways in 1962 were built before 1949 and were upgraded with engineering support from the Soviet Union from 1949 to 1962. During this period, the resource-rich Western region was connected with Eastern manufacturing centers, which shipped raw materials and manufactured goods between large urban cities (Baum-Snow et al., 2017). According to China Statistical Yearbooks, China had a total length of 34,600 km in railways and 143,442 km in roadways in 1962.[12]

II.2 Other Elements in the Database

In addition to transport modes, our database also includes information on terrain features, county boundaries, city locations, and postal areas as follows:

Terrain Feature: We obtain data on county-level terrain features from the National Geographic Information database managed by the Ministry of Natural Resources in China.[13]

---

[10] The inland waterway system consists of two horizontal trunk waterways (Yangtze River and Xijiang River), one vertical trunk waterway (Beijing-Hangzhou Grand Canal), two high-grade waterway networks (Yangtze River Delta and Pearl River Delta), and 18 high-grade mainstreams and tributary waterways. The south-north coastal line covers major cities on the east coast between Beihai in Guangxi province in the Southern region and Dangdong in Liaoning province in the Northern region. The total waterway length was 110,200 km in 1993, 127,000 km in 2015, and 127,300 km in 2019 in China.

[11] Since 1994, there have been improvements in navigability and cargo-handling capabilities of ports on the Yangtze River, Xijiang River, and Beijing-Hangzhou Grand Canal (Ministry of Transport, 2020).

[12] The statistics come from China Statistical Yearbooks by the National Bureau of Statistics (NBS). See: http://www.stats.gov.cn/tjsj/ndsj/

[13] The Ministry of Natural Resources is an executive department of the State Council in China, which is responsible for the country's natural resources. See: http://www.mnr.gov.cn/ (Last accessed on 2024/06/19)

Satellite images yield data on terrain elevation for each raster (grid square) in a given county. For each county, we record the number of grid squares and the cross-raster mean, median, standard deviation, minimum, and maximum elevations. We also measure the count of unique elevation values as elevation variety in a given county.

Boundary: Our database includes a shapefile that records 2,869 counties, 341 prefecture-level cities, and 31 provinces in China, along with their boundaries based on China's public administration structure in 2020. This file can be used to construct maps that exhibit how socioeconomic statistics in different administrative units vary with local transport measures and other data sources.

Geolocation: Our database includes geographical coordinates for 341 prefecture-level cities and provincial capitals in China. We record longitude and latitude information on the geographical centroids of city-level administrative units. Combined with our GIS data on transport modes, these data enable us to compute city-level transport quality measures, such as density, connectivity, and centrality, as well as travel time and distances between cities.

Distance and Area: We use the *GCS_WGS_1984* geographical coordinate system to define locations by longitude and latitude. This coordinate system is in wide use, including by *OpenStreetMap*. We then calculate distances between two points and straight-line distances between points and transport routes using *PointDistance* function in ArcGIS. We measure area sizes using the *Krasovsky_1940_Albers* projection.

Postal Area: China partitions its area into 35,798 postal areas, analogous to zip codes in the United States. For each "zip code" with one or more operational manufacturing plants over 1998-2013, we incorporate longitude and latitude values for the area's centroid from Davis and Qian (2021) into our database and provide geolocation data for about 97.5% of all postal areas.[14]. We calculate travel time and distance between each postal area and their nearest motorway access points, railway stations, highway routes, and navigable waterways.

[Table 1. An overview of the GIS database of surface transports in China]

II.3 Database Construction: An Overview

We summarize the main steps to construct a reliable and ready-to-use GIS database for all surface transports in China:[15]

[14] The missing areas have little commerce in the Western region, which is less important for our analysis.

[15] The online technical manual "ReadMe" file describes additional details of our database for all surface transports in China over 1993-2020 (Davis and Qian, 2024). The file named "Technical Details of Digitization and Rectification" introduces in more detail the database construction process.

1) Obtain raw GIS data from *OpenStreetMap* in 2013 and 2020 that cover all transport modes and include information on land features;
2) Collect physical maps for transport modes over 1993-2020 from the China Road Atlas published by China Communications Press, online bookstores, such as Kongfuzi and Jiushujie, and second-hand book vendors on E-commerce platform, such as Taobao;[16]
3) Scan each physical map and incorporate spatial information into ArcGIS interface using the *GCS_WGS_1984* coordinate system compatible with *OpenStreetMap*;
4) Merge the digital representations of physical maps in 2013 with *OpenStreetMap* to build shapefiles for transport routes and identify motorway access points and railway stations as "nodes":
    a. Include certain primary and secondary roads as part of the national highway system, following *OpenStreetMaps*;
    b. Exclude subway lines, light rail lines, and tram networks in metropolitan cities from our designation of standard and fast railways;
    c. Designate each motorway access point by the centroid of a rampway, which is a curve or butterfly shape;
    d. Delete false intersections in *OpenStreetMap* when one highway passes over another;
    e. Draw on official sources to correct missing and wrongly located railway stations;[17]
    f. Identify false dangles in *OpenStreetMap* using satellite data, correct these false dangles to indicate the presence of a throughway, and fix the errors for motorways, highways, and railways;
5) Repeat Step 4 for the data in 2020;
6) Create shapefiles of transport modes from 2012 back to 1993 using the *OpenStreetMap* in 2013 and new additions of transports obtained by digital representations of physical maps year-by-year over 1993-2012;[18]

[16] Kongfuzi: www.kongfz.com; Jiushujie: www.jiushujie.com; Taobao: www.taobao.com
[17] We obtain the official information on railway stations from the National Railway Administration (NRA) (http://www.nra.gov.cn/) and 12306 China Railway (https://www.12306.cn/mormhweb/kyyyz/).
[18] Since GIS transport data has the most extensive coverage in 2013 over the period of 1993-2013, it is easier to delete transport additions from the map of 2013 backward than to draw new sections forwards. For instance, we match the digital representations of physical map data in 2012 with *OpenStreetMap* in 2013 and delete the relevant parts to attain the GIS data in 2012 to create the shapefile of transport modes in 2012. Similarly, the creation of shapefiles of transport modes from 2019 back to 2014 is finished using *OpenStreetMap* in 2020 in the same way.

7) Create shapefiles of transport modes from 2019 back to 2014 using the *OpenStreetMap* in 2020 and verify with the approach by adding new routes from 2013 onwards to obtain digital representations of physical maps over 2014-2019.
8) Add speed attributes for motorways and highways using the legal maximum speeds by route *OpenStreetMap* in 2020;
9) Adjust the speed attributes for other years based on data from the six railway speed-up campaigns;[19]
10) Add data on types of railway stations and grades of railway routes.

II.4 What Measures Does Our Database Deliver?

Table 2 lists all transport measures delivered by our database, including location-specific, pairwise, and network-level statistics. Here, "locations" refer to county centroids, motorway exits/entrances, railway stations, and ports. We first calculate distances between county centroids and the nearest motorway exit/entrance, railway station, port, highway route, and waterway route. Our database also delivers centrality measures for each location, as well as travel distance and time between locations. In addition, transport density of length and node, centrality, and connectivity at national, regional, provincial level, and user-defined levels. These measures are widely used in studies on transport networks.[20]

[Table 2. Definition of transport measures in the database]

III. The Evolution of Surface Transports

III.1 Visual Depictions

Using our database, we conduct graphical analyses in this section. Figure 1 exhibits temporal evolutions of motorway transports over 1993-2020. The extraordinary expansion of the motorway network has been apparent and is highly concentrated in the eastern region of China. Figure 2 presents the changes in representative years of 1993, 2003, 2013, and 2020 to see the evolution more clearly. It shows that most new motorways were constructed after the early 2000s, even if the expansion was already underway by the late 1990s.

[Figure 1. The expansion of motorway transports over 1993-2020]

[19] Due to the "Eight Vertical and Eight Horizontal" launched by the Chinese government in 2001, there were six "speed-up" campaigns in the railway network from 1997 to 2007. The National Railway Administration posted the specific date, railway section, and new speed of all speed improvements. We manually collect data on them and draw six separate shapefiles based on these official documents (http://www.nra.gov.cn/).

[20] See Wasserman and Faust (1994) and Jackson (2010), among others, for broad treatments and many applications. See Kansky (1963), Taafee et al. (1996), and Rodrigue (2020) for a focus on transport networks. In previous work on China, Wang et al. (2009) measure the centrality and connectivity of city centroids based on railways only in various years from 1906 to 2000. Jin et al. (2010) quantify network attributes in 2006 based on data for railways, highways, ports, and airports.

[Figure 2. Motorway transports in 1993, 2003, 2013, and 2020]

The panels in Figure 3 show high-speed railways in 2003, 2010, 2015, and 2020, respectively. We see an extraordinary expansion after 2003. Figure 4 presents the temporal evolution of regular railways, indicating that China already had an extensive network of standard railways in place by 1995. We find remarkable further expansions and upgrades of these railways from 1995, especially after 2003. In 1995, the railway system had an average speed of 50 km/h but a maximum speed of 120 km/h in almost every railway route in 2013.

[Figure 3. High-speed railways in 2003, 2010, 2015, and 2020]

[Figure 4. Regular railways in 1995, 2003, 2013, and 2020]

Figure 5 shows the temporal evolution of national highways in 1995, 2003, 2013, and 2020. Given a large national highway system in place by 1993, China also experienced tremendous expansions in the national highway system over 1993-2020. Figure 6 depicts navigable waterways as of 1994 since there exists only a small increase of 15.8% in total inland waterway length over the past decades, from 110,200 km in 1993 to 127,700 km in 2020. We distinguish waterways by six capacity levels and drop the waterways with a capacity of less than 50 tons since they cannot be navigable channels used to transport goods, materials, or other movable objects.

[Figure 5. National highways in 1995, 2003, 2013, and 2020]

[Figure 6. Figure 6. Navigable waterways in 1993]

Figure 7 presents the county-level mean and variability of elevation levels in China. Panel (a) shows mean elevation levels rise from the Eastern to the Western region. The mean elevation is typically less than 300 meters in the Eastern part of China but exceeds 3,000 meters in the Western region. Not surprisingly, surface transport is highly developed in low-lying Eastern counties with elevation levels of less than 300 meters. Panel (b) presents county-level ruggedness measured as the standard deviation (variability) of within-county elevations. It shows the Eastern counties tend to have more rugged terrain. The Xinjiang and Southwest border of Yunnan Province feature the most rugged terrains.

[Figure 7. Mean and variability of terrain elevations by county]

III.2 Network Length, Access, and Coverage

Table 3 reports the total lengths of surface transports based on our GIS database in Panel A and official statistics in Panel B, respectively. Our GIS database shows in Panel A that the total length of motorways grew by 24 times from 6,300 km in 1995 to 153,400 km in 2020. The total length of high-speed railways was only 400 km in 2003 and grew to 13,600 km in 2013 and 39,000 km in 2020. The length of the highways also roughly doubled from

104,900 km in 1993 to 190,700 in 2020. Panel B reports the relevant statistics from China Statistical Yearbooks. The total lengths of high-speed railways and all railways in Panel B are close to those in Panel A, suggesting our GIS database accurately yields their length measures. Railway lengths are slightly greater than official statistics in earlier years since some railway routes that do not meet the grade requirements in China Statistical Yearbooks are also included in our GIS database. Similarly, the total length for high-speed railways reported in our GIS database is higher than official statistics because we include route segments of regular tracks on which high-speed trains travel occasionally. The total lengths of highways from official sources are also very close to what we calculated in GIS database.[21] Though not reported in Table 3, the total lengths of navigable waterways obtained from our GIS database are similar in value to official statistics.[22]

[Table 3. Total length of surface transports]

Figure 8 presents the number of access points for transport from 1993 to 2020. It shows that the number of railway stations increased by 31% from 6,684 in 1995 to 8,781 in 2020. Among all railway stations, high-speed railways expanded much faster than regular railways, and its number has risen by almost 107 times, from 10 in 2003 to 1069 in 2020. In the meantime, the number of motorway exits/entrances increased approximately 12 times, from 796 in 1993 to 9,972 in 2020.

[Figure 8. Number of access points to transports over 1993-2020]

Figure 9 shows the geographic coverage rate by transport network among all 2,819 counties over 1993-2020 in China. Panel (a) exhibits the proportion of counties through which a transport route passes, showing remarkably increasing coverage rates of all transport modes from 1993 to 2020. We find the most salient increases in motorways, whose coverage rate rises by 12 times from 6.7% in 1993 to 84.2% in 2020, and in high-speed railways from 0 in 2002 to 44.1% in 2020. Compared to high-speed railways, the coverage of regular railways increased less, from 58.3% in 1993 to 72.4% in 2020. In addition, 70.5% of counties in 1993 had at least one highway route passing through them, and this coverage ratio increased to 92.7% in 2020. Panel (b) plots the fraction of counties with access to motorway entrances/exits, railway stations, or high-speed railway stations within its geographical boundary. The percentage of counties with a high-speed railway station

[21] The China Statistical Yearbooks originally post the total lengths of national roadways, including highways and motorways, as defined in our GIS database. We subtract the total lengths of national roadways from motorways to attain the total lengths of highways in Panel B of Table 3.

[22] Our GIS database yields 64,000 km of navigable inland waterways in 1994 after dropping routes with graded capacity of less than 50 tons, nearly matching the length of 68,000 km in official sources.

increased dramatically, starting from 0.28% (8 counties) in 2003 to 29.1% (821 counties) in 2020, while the coverage rate of regular railway stations rose from 53.4% in 1993 to 65.1% in 2020. The counties with motorway entrances/exits grew 13.6 times, from 5.9% in 1993 to 80.0% in 2020.

[Figure 9. Transport coverage at the county level over 1993-2020]

III.3 Visualize the Statistics for Transport Features

We visualize the statistics for transport features computed in our GIS database in Figures 10 and 11. Figure 10 plots the average distance between the geographic centroids of 34,889 zip codes and their nearest access to motorways, regular railways, and high-speed railways from 1993 to 2020. It displays that access to motorways has significantly improved during this period, as the average distance from zip codes to the nearest motorway entrances/exits has decreased to less than 1/20, from 301.86 km in 1993 to just 14.59 km in 2020. The average distance to the nearest high-speed railway station also substantially dropped from 1,254 km in 2003 to 59 km in 2020.

[Figure 10. Average distance to the nearest transports]

On top of the distance to transport, we compute the transport density over 1993-2020 in China and present the average density *averaged* at the national level in Panel A and county level in Panel B, respectively, in Figure 11. Length Density equals the ratio of total length of all transports to a county's area, including motorway, high-speed rail, regular rail, highway, and waterway, in units of meter/km^2, and Node Density is the ratio of number of access points of motorway exit/entrances and railway stations to the county's area, in units of #/km^2. Panel A (B) shows length density at the national (county) level increased from 19 meter/km^2 (94) in 1993 to 59 meter/km^2 (177) in 2020. Similarly, node density measures are increasing, from 0.0008 #/km^2 (0.0042) in 1993 to 0.002 #/km^2 (0.0079) in 2020 at the nation (county) level in Panel A (B). We discover both density measures averaged at the national level are much smaller than those averaged at the county level. The differences are consistent with the geographical variations of transport and county features in China. Many counties have smaller areas but higher transport densities, while a few counties have larger areas but lower transport densities. The regional heterogeneity yields right-skewed distributions of transport densities.

[Figure 11. Transport density of length and access points over 1993-2020]

Figure 12 presents county-level *connectivity* of all transport modes in 1993, 2000, 2010, and 2020. We adopt Beta index to proxy for connectivity: the average number of links per node in a network.[23] The growth rates of connectivity vary across counties. They grew the fastest in counties in the Central region, such as those in Henan, Hebei, Shanxi, and Anhui provinces, reaching an average of 2.45 in 2020. However, in the Western region such as Xizang, Xinjiang, and Yunnan provinces, the connectivity grew the slowest, with an average value of 2.02 in 2020.

[Figure 12. County-level connectivity in 1993, 2000, 2010, and 2020]

On top of connectivity, we measure *centrality* that reflects the relative locational importance of a county to others in the entire transport network, taking geographical centroids of counties as nodes and transport network lines as links. A higher level of centrality implies a county has a more interconnected relationship with others. We adopt Betweenness to proxy for centrality, which equals the number of links of the shortest paths between pairs that pass through the location over the total number of links for all shortest paths between pairs in a network (See Table 2) (Jackson, 2010). Since the range and value of the raw Betweenness index measures are too large, we normalize it to [0, 1] in the classic way as follows (Freeman, 2002; Brandes, 2001):

$$B^n = \frac{2B}{(n-1)(n-2)}, \tag{1}$$

where B^n is the normalized index and B is the raw index. n is the number of vertices in the graph. The normalized Betweenness index becomes more suitable for making inter-regional and temporal comparisons and is thus used for subsequent graphical and statistical analyses. Figure 13 shows county-level centrality (Betweenness index) of all transport modes in years 1993, 2000, 2010, and 2020, respectively. The national average of Betweenness index grew by 16% from 0.0146 in 1993 to 0.0170 in 2020, suggesting counties become more interconnected via transport development in China. Grouping counties into five categories based on centrality, we discover that the highest centrality exists in counties along transport corridors connecting China's Northeast, East, and Central regions.

[Figure 13. County-level centrality in 1993, 2000, 2010, and 2020]

[23] Another alternative measure of connectivity, the Alpha index, which equals the actual number over the maximal number of circuits (paths that start and end at the same node) for a given county, is not calculated in our paper (Kansky, 1963). The underlying reason is that, given the dramatic expansion of transport networks, the maximum number of circuits increases faster than the actual connection of transport circuits newly constructed and thus cannot properly describe transport connectivity in China.

IV. Transport Route *versus* Access Point

Many studies use transport routes that pass through as a measure of transport coverage. However, Figure 9 shows a large discrepancy in coverage rates between route-passing and having actual access points to motorways, railways, and high-speed railways. It introduces potential errors and biases in analyses to use route-passing, rather than real access, to proxy for transport coverage. One advantage of our GIS database is the accurate information on both route-passing and access points, which allows us to examine these biases.

IV.1 The Growth of "Passing-Through" Counties

The expanding transport network leaves out many counties by passing through them but providing them with no transport access points in China. We find that the counties through which transports pass but without access points increased over time. By 2020, 1,243 out of 2,819 counties (44%) had at least one high-speed railway line passing through without access, and 421 counties had no high-speed railway stations. Therefore, using the shortest distance to transport routes by straight line to proxy distance to transport is problematic for these counties.

Figure 14 shows the discrepancy between the shortest distance to transport access points and to transport routes over 1993-2020. We first compute two distance measures between county centroid and motorways, railways, and high-speed railways at the county level and then take the differences between the two measurements. Panel (a) exhibits the simple average of distance discrepancies across counties, and Panel (b) shows the GDP-weighted average of distance discrepancies across counties. The two panels show similar trends for the distance discrepancy. However, the GDP-weighted distance discrepancies are smaller in magnitude than those with the simple average, indicating that less developed counties have more difficulty accessing transport networks and, thus, larger distance discrepancies. In Panel (a), the distance discrepancy for motorways steadily declines from 2.43 km in 1993 to 1.87 km in 2020. In contrast, the distance discrepancy for regular railways has increased, suggesting that they tend to skip more counties by passing through them without new stations in China. In both panels, the high-speed railways have the largest distance discrepancies that fluctuated during 2003-2020. Panel (b) shows that the GDP-weighted distance discrepancies for high-speed railways increased from 2.65 km in 2003 to 4.23 km in 2020.

[Figure 14. Distance discrepancy between access point and route to county centroid]

The large number of counties through which transports pass but without access points emphasize the importance of measuring distance by actual access points rather than the shortest distance to transport routes. Since the distance discrepancy remains until 2020, two concerns arise about the findings in existing studies. The first concern is the reliability of results in prior studies that rely on straight-line shortest distances. The second calls attention to the treatment effect of motorway transports. As Faber (2004) finds, new motorway constructions lead to decreased economic activities in counties through which transports pass but without access points. Therefore, the measurement bias in distance could be more severe in estimating the treatment effects of transport development on socioeconomic growth over time in China.

IV.2 Distance Discrepancy and Socioeconomics

Previous studies show proximity to transport significantly raises local income levels and GDP per capita but not economic growth rates (Banerjee et al., 2020). Given the concern over potential bias in empirical findings due to distance discrepancy, we examine how differences between "proximity" and actual "access" correlate with local geographical and socioeconomic characteristics and evaluate the economic bias in the previous studies. As before, we consider distance discrepancy as the difference between the shortest distance to a transport access point and the shortest distance to the nearest transport route from a county's centroid. Table 4 illustrates the correlation analysis between the two kinds of distance and distance discrepancy. Distance discrepancies positively correlate with two distance measures for motorways and regular railways. However, the distance discrepancies for high-speed railways are negatively correlated to their two measures of distances, implying that locations of high-speed railway stations are strategically chosen to make the actual distance to the stations shorter when the high-speed railway routes are far from county centroids. Since local geographical features could influence the construction strategies of transports, we also explore the regional heterogeneity by dividing the 2,819 counties into four regions in China, i.e., Eastern (785), Central (688), Northeastern (277), and Western (1,069). We discover that the positive correlations between distance discrepancies and two distance measures are consistent for motorways and regular railways across regions. However, the correlations flip the signs for high-speed railways and become significantly positive in the Northeastern region, even if the negative relations remain significant in the Eastern, Central, and Western regions. This pattern may be related to the fact that the strategic location selection for high-speed railway stations is easier to achieve in the developed region, i.e., the Eastern region, and regions where terrains have high ruggedness, i.e., Central and Western regions. To more clearly see the regional heterogeneity, we also divide the 2,819 counties into two regions, East (1,062) and West (1,757), and find significant variations between them in the correlations of distance discrepancy (See Table B2 in Appendix).[24]

[Table 4. Correlation analysis between two distances and distance discrepancy]

To examine how local socioeconomics affects distance discrepancy of transport, we manually collect local socioeconomic characteristics from Statistical Yearbooks of Counties and Statistical Yearbooks of Cities over 1995-2020 in China and conduct panel regression analyses. Panels A and B in Table 5 show the results at the county and prefectural city levels, respectively. Given the data availability, we conduct regressions for motorways and regular railways over 1999-2020 and high-speed railways over 2003-2020. The main explanatory variables include log GDP, GDP per capita, and GDP growth rate expressed as a percentage. We also include controls for commonly used local socioeconomic indicators, such as log population, annual government revenue and spending, aggregate household savings, loans, consumption, industrial companies, primary school students, and hospital beds. Year- and county-level fixed effects are also included as control variables. We find similar relationship patterns between city- and county-level regressions. A higher local GDP level is associated with reduced distance discrepancies for motorways, suggesting that the denser motorway transports are more likely to be constructed in wealthier counties/cities. However, a higher local GDP is related to increased distance discrepancy for regular and high-speed railways at both the county and city levels. This pattern could be explained by the construction strategy of high-speed railway stations, primarily driven by the central government's aim to balance economic development across regions in China. Therefore, unsurprisingly, denser high-speed railways are constructed in less economically developed areas. Furthermore, the distance discrepancy in high-speed railways is negatively associated with a higher GDP per capita growth rate. This pattern suggests that higher economic growth, most likely in poorer areas with lower initial GDP, could increase the likelihood of building a denser high-speed railway network within a county/city. In contrast, the distance discrepancy in regular railways is positively associated with GDP growth, indicating that the construction of denser regular railways is prioritized in counties/cities with a lower GDP growth. In another paper, we explore the endogeneity and causality between transport and economic development in more detail (Davis et al., 2024).

[Table 5. Effects of local socioeconomics on distance discrepancy]

[24] Here, we consider Eastern and Northeastern regions to be East and Central and Western regions to be West.

IV.3 Implications for Results in Previous Studies

Faber (2014) finds that new motorways reduce local productivity on the routes but have no access to new motorways, implying that counties with a larger distance discrepancy also have a lower GDP in China. Based on the empirical findings in Tables 4 and 5, the relations documented in Faber (2014) are likely a reverse causality since local socioeconomic conditions could also impact the construction of new motorway exit/entrances.[25] Moreover, significant anomalies might arise if a "theoretical" straight-line distance to transport routes rather than an "effective" distance via access points is used in the measurement. Specifically, counties with lower economic growth are mistakenly treated as gaining transport access when motorways just pass through them in reality. Banerjee et al. (2020) discover that proximity to transportation networks positively affects per capita GDP levels in China using historical transport routes as instrumental variables to address the endogenous placement of networks. However, they overlooked the endogeneity issue that arises from the distance measurement errors between transport access points *versus* transport passing-through.

V. Advantages of Our GIS Database

Previous studies on China's transportation typically focus on a single transport mode, a short period, or a point-in-time snapshot. They also lack accurate geospatial information on transport access points, i.e., railway stations and motorway exits/entrances, which is a significant advantage of our GIS database. Appendix Table B1 summarizes transport datasets in other studies, including other GIS data sources, digital and GIS data collected by different authors, China Statistical Yearbooks, and timetable data. We outline the data sources, transport mode, period, transport measures, and studies that use these data. Compared with existing databases, our GIS database offers four advantages.

Firstly, our panel database is the most comprehensive regarding period and transport modes (see Table 1). The panel covers a long period of 28 years in the post-opening up period, which is the longest and most recent among relevant studies (Faber, 2014; Jin and Chen, 2019; Banerjee et al., 2020; Dinlersoz and Fu, 2022). The long period in the panel data allows for the examination of transport development and its dynamic relationship to socioeconomic development. In addition to the long period, the year-by-year coverage in

[25] Specifically, Faber (2014) considered the route design and used the distance among major cities that aim to be connected by transport to address the endogeneity issue of route access points and pass-through counties. However, he did not analyze the reverse effect of the prospects of the local economy on transport constructions from a socioeconomic perspective.

our panel data improves causal inferences with temporal changes in a longitudinal study, panel data research, and a difference-in-difference analysis. In addition, our database covers all surface transport modes and makes it possible to control existing conditions of alternative transport modes when studying the effects of transport development. Beyond that, covering all surface transport yields a more accurate estimation of multiple transport modes and transport than one single transport.

Second, our GIS database contains geospatial information on motorway exits/entrances and railway stations, which helps differentiate route-passing from transport access points. Figure B1 presents an example of routes and access points of motorways and railways in our GIS database. Most extant studies use routes that pass through as a proxy for access points and the shortest distance to transport routes for distances to access points.[26] As earlier sections show, such proxies cause substantial measurement errors and biases in analyses. Moreover, the absence of information on exits, entrances, and stations could introduce biases in measuring the connectivity and centrality of any location within the transport network.

Thirdly, our GIS database significantly improves the accuracy of geographic information since it builds upon the *OpenStreetMap*, which has several advantages over digitized physical maps. Digitizing physical maps requires rectifying the scanned paper maps with known geographical coordinates. However, without a coordination benchmark for rectification, we cannot achieve the best match between scanned transport lines, open street road lines, and fitting coordinators. Besides, the original GIS data in *OpenStreetMap* has geographical information on locations and directions of transport routes, local streets, firms, and city and county boundaries. Furthermore, it is easy to keep updating and improving consistency and sustainability in our GIS database in future with technological development and joint efforts by the GIS community for the *OpenStreetMap*.

Finally, using our GIS database, other researchers can directly calculate travel distance, time, and other transport quality measures. We have gone through many topology verifications and rectifications in the GIS and put much effort into matching speed information for transport routes and modes in *OpenStreetMap*. However, other studies assume constant speed on all roads (Storeygard, 2016) and consider the same speed for the same type of transport mode to construct efficiency kilometers (Baum-Snow et al. (2016,

[26] One exception is Qin (2017), who employs information on high-speed railway stations.

2020). Our GIS database includes route-specific attributes of speed limitations for different transport modes, yielding a more accurate estimation of travel time between two locations.

VI. Applications of Our GIS Database

This section discusses potential applications of our GIS database for academic research, commercial decisions, policymaking, and general audiences and readers. We provide county-level, regional, and pairwise transport quality measures that are employed in applications. The summary statistics of transport features computed by our GIS databases show significant variations over the years and across regions in China. The panel enables us to examine the dynamics of their changes and how transport improvements influence socioeconomic development and environmental quality in China if combining our GIS database with other time series, cross-sectional, or panel data through locational identifiers. Moreover, the long panel allows for high-quality causal inferences, which has been increasingly demanded. We introduce potential applications of our GIS database as follows:

(I) **Economic activities:** Improved transports reduce transporting costs, broaden market access, and promote interregional and international competitiveness in productive output. Thus, more attention should be given to the impacts of transport infrastructure on resource allocation, labor mobility, entrepreneurial activities, investment, employment, productivity, return on capital, income growth, and interregional and international trade. We are yet to discover: 1) whether transport improvements causally affect economic outcomes; 2) how much transport conditions explain regional, industrial, or firm-level differences in economic activities; and 3) what transport features are more important than others for local socioeconomics.

(II) **Industry development:** Transports change the geographical distribution of industrial development and competition through their effects on the economic activities of firms and individuals. Though China has the world's largest and most comprehensive manufacturing industry, there is yet to be an estimation of the manufacturing economy contributed by transportation infrastructure in China. It is also critical to explore how to make strategic investment decisions and maintain its competitive advantages by improving transport.

(III) **Urbanization:** Industrial development and labor mobility substantially contribute to urbanization. Our GIS database encompasses all surface transport modes that supported domestic migration during the rapid urbanization period in China. Over the past decades, hundreds of millions of people moved from rural villages to large metropolitan areas and transferred a significant percentage of production from agriculture to the manufacturing sector in GDP. The rural population reduced from 85,344 million in 1993 to 55,162 million by 2020, but the urban population increased from 33,173 million in 1993 to 84,843 million in 2020.[27] Thus, it is crucial to examine how infrastructure investment, improvement in household access to transport, and reduction in transport costs contribute to urbanization and vice versa.

(IV) **Social mobility and poverty reduction:** Building roads to get rich has become a widespread slogan in rural areas. Resource allocations, economic activities, and urbanization lead to socioeconomic changes, poverty reduction, and social mobility. The Chinese government always prioritizes improvements in transport infrastructure to alleviate poverty. Over the past decades, the impoverished population has decreased notably in rural China, and the criterion of poverty increased from an annual income of 2,300 CNY ($335) in 1995 to 4,000 CNY ($584) in 2020 per year.[28] Since China has eliminated poverty for the largest population in the world, it is critical to investigate: 1) how much has transport improvement contributed to the large scale of poverty reduction? 2) Whether labor mobility through transport contributes to reducing inequality? 3) Which transport mode is more effective than others in achieving it? 4) How are the changes conditional on prior local socioeconomic status? 5) What is the regional heterogeneity in this process? 6) and most importantly, what lessons can be applied to other less developed economies?

(V) **Environment**: Transportation infrastructure facilitates the mobility of resources and technology diffusion, which supports income and economic growth. However, it may also aggravate pollution emissions and global environmental degradation.[29] The development of transport transforms the land and surrounding areas. Urbanization, industrialization, and city extensions transform agricultural and forest lands into industrial. Do these effects differ across transport modes or the structure of Access and Density? Are the effects dynamic related to the prior transport conditions, natural endowment, or economic development stage? What arrangement might achieve the best possible balance between development and preservation?

---

[27] The data comes from the China Statistical Yearbook released by the National Bureau of Statistics (NBS) of China. See: http://www.stats.gov.cn/sj/ndsj/2021/indexch.htm (Last accessed on 2024/06/19)
[28] The data is from Xinhua News. See: http://www.xinhuanet.com/fortunepro/2021-02/25/c_1127137706.htm (Last accessed on 2024/07/10)
[29] Road constructions aggravate illegal colonization, mining, hunting, and land speculation in nature reserves. Laurance (2009) documents that over 95% of deforestation, fires, and atmospheric carbon emissions in the Brazilian Amazon occurred within 50 kilometers of a road.

(VI) **Policymaking, education, and commercial use**: Our GIS database facilitates a comprehensive understanding of the evolvement of transportation in China over a long period. Directly visualizing transport networks can provide an informative overview of transport development in China. The long-time series also helps local governments understand and predict regional development pathways. Moreover, transport information can combine with industrial census, survey data, and market data as valuable inputs for investment, marketing, and strategic commercial decisions. Thus, our database is beneficial to a broad range of readers, individuals or organizations in education, tourists, corporations, and policymakers.

(VII) **Sustainability**: A readily applicable database encourages and facilitates studies of topics that resolve rising issues in growth and thus offer deep insights into policymaking. Our transport dataset helps examine the applicability of China's experience to other developing countries and what policies are essential for sustainability. For example, labor mobility during "Spring Festival Transport Season" leads to the largest seasonal population migration in human history. Therefore, an adequate accommodation for such large-scale internal migration is critical to people's welfare and social stability.

VII. Concluding Remarks

We construct and introduce a new GIS database for transportation in China and use it to shed new light on measurement issues that pertain to the impact of transport networks on local economies. Our database covers all surface transport modes, i.e., motorways, national highways, regular railways, high-speed railways, and waterways at multiple administrative levels from 1993 to 2020. We conduct extensive rectifications and verifications in the GIS database to ensure accuracy and reliability. The statistics on transport lengths derived from our database match well with those in official documents. The panel in our GIS database has the broadest coverage among currently available GIS databases and digital maps on China's transport network. Our database also provides information not included in other databases, including geocoded information on motorway exits/entrances and railway stations, lane directions, railway tracks, and speed limitations in some routes. These key attributes and geospatial information yield more accurate transport quality measures, like distance, access, density, connectivity, centrality, and travel time.

Using our GIS database, we examine the discrepancy in transport coverage between route-passing and actual access. We find large disparities between distance to transport access points and straight-line distance to transport routes, raising concerns about the findings in some previous studies. Our data facilitate more attention to the effects of China's transport expansion on peripheral regions, including "passing-through" regions. Our data also creates new opportunities for long-span longitudinal analyses, better controls for alternative transport modes, and better identification in studies of trade, productivity, income distributions, labor and population mobility, land usage, and environmental quality.

Finally, we use our database to quantify several aspects of transport scale, scope, density, connectivity. Based upon our high-quality GIS data, we can calculate these measures at the zip code, county, city, provincial, regional, and national levels. Our summary statistics and graphical presentation show large variations over the years and across regions in China. We use these measures to explore the relationship between transport and economic development in another paper (Davis et al., 2024).

References

[1] Allen, T., & Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. The Quarterly Journal of Economics, 129(3), 1085-1140.

[2] Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. Journal of Development Economics, 102442.

[3] Baum-Snow, N. (2007). Did highways cause suburbanization? The quarterly journal of economics, 122(2), 775-805.

[4] 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.

[5] Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2016). Highways, market access and urban growth in China. SERC, Spatial Economics Research Centre.

[6] Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? Journal of Urban Economics.

[7] Benítez-López, A., Alkemade, R., & Verweij, P. A. (2010). The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis. Biological conservation, 143(6), 1307-1316.

[8] Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2), 163-177.

[9] Chapman, L. (2007). Transport and climate change: a review. Journal of transport geography, 15(5), 354-367.

[10] China Communications Press. China Highway Atlas. Beijing: China Communications Press. 1993-2020.

[11] Davis, S., & Qian, M. (2021). Transport Infrastructure and Productivity: The China Experience. Working Paper.

[12] Deng, T., Shao, S., Yang, L., & Zhang, X. (2014). Has the transport-led economic growth effect reached a peak in China? A panel threshold regression approach. Transportation, 41(3), 567-587.

[13] Diao, M. (2018). Does growth follow the rail? The potential impact of high-speed rail on the economic geography of China. Transportation Research Part A: Policy and Practice, 113, 279-290.

[14] Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. American Economic Review, 108(4-5), 899-934.

[15] Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. The Quarterly Journal of Economics, 131(2), 799-858.

[16] Dong, X., Zheng, S., & Kahn, M. E. (2020). The role of transportation speed in facilitating high-skilled teamwork across cities. Journal of Urban Economics, 115, 103212.

[17] Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. Review of Economic Studies, 79(4), 1407-1440.

[18] Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. Econometrica, 70(5), 1741-1779.

[19] Emran, M. S., & Hou, Z. (2013). Access to markets and rural poverty: evidence from household consumption in China. Review of Economics and Statistics, 95(2), 682-697.

[20] Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. Review of Economic Studies, 81(3), 1046-1070.

[21] Fan, J. (2019). Internal geography, labor mobility, and the distributional impacts of trade. American Economic Journal: Macroeconomics, 11(3), 252-88.

[22] Freeman, L. C. (2002). Centrality in social networks: Conceptual clarification. Social network: critical concepts in sociology. Londres: Routledge, 1, 238-263.

[23] Grieco, M. & Urry, J. (2016). Mobilities: new perspectives on transport and society. London and New York: Routledge.

[24] He, G., Xie, Y., & Zhang, B. (2020). Expressways, GDP, and the environment: The case of China. Journal of Development Economics, 102485.

[25] Hu, H., Wang, J., Jin, F., & Ding, N. (2015). Evolution of regional transport dominance in China 1910–2012. Journal of Geographical Sciences, 25(6), 723-738.

[26] Huang, Y., & Xiong, W. (2017). Geographic Distribution of Firm Productivity and Production: A "Market Access" Approach. Harvard University. Working Paper.

[27] Huang, Y., & Zong, H. (2020). The spatial distribution and determinants of China's high-speed train services. Transportation Research Part A: Policy and Practice, 142, 56-70.

[28] Huang, Y., & Zong, H. (2020). Spatiotemporal evolution of land transportation networks and accessibility in inland mountainous areas 1917–2017: A case study of Southwest China. Journal of Mountain Science, 17(9), 2262-2279.

[29] Hystad, P., Demers, P. A., Johnson, K. C., Carpiano, R. M., & Brauer, M. (2013). Long-term residential exposure to air pollution and lung cancer risk. Epidemiology, 762-772.

[30] Ibisch, P. L., Hoffmann, M. T., Kreft, S., Pe'er, G., Kati, V., Biber-Freudenberger, L., & Selva, N. (2016). A global map of roadless areas and their conservation status. Science, 354(6318), 1423-1427.

[31] Jackson, M. O. (2010). Social and economic networks. Princeton university press.

[32] Jiang, B., Liang, S., Peng, Z. R., Cong, H., Levy, M., Cheng, Q., & Remais, J. V. (2017). Transport and public health in China: the road to a healthy future. The Lancet, 390(10104), 1781-1791.

[33] Jiao, J., Wang, J., & Jin, F. (2017). Impacts of high-speed rail lines on the city network in China. Journal of Transport Geography, 60, 257-266.

[34] Jiao, J., Wang, J., Jin, F., & Dunford, M. (2014). Impacts on accessibility of China's present and future HSR network. Journal of Transport Geography, 40, 123-132.

[35] Jin, F., Wang, C., Li, X., & Wang, J. E. (2010). China's regional transport dominance: Density, proximity, and accessibility. Journal of Geographical Sciences, 20(2), 295-309.

[36] Kahn R, Kobayashi SS, Beuthe M, et al. Transport and its infrastructure. In: Metz B, Davidson OR, Bosch PR, Dave R, Meyer LA, eds (2007). Climate change 2007: mitigation contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change. Cambridge and New York: Cambridge University Press.

[37] Kansky, K.J., (1963). Structure of Transportation Networks: Relationships between Network Geometry and Regional Characteristics. PhD Dissertation, Department of Geography, University of Chicago.

[38] Lao, X., Zhang, X., Shen, T., & Skitmore, M. (2016). Comparing China's city transportation and economic networks. Cities, 53, 43-50.

[39] Laurance, W. F., & Balmford, A. (2013). A global map for road building. Nature, 495(7441), 308-309.

[40] Laurance, W. F., Clements, G. R., Sloan, S., O'connell, C. S., Mueller, N. D., Goosem, M. & Arrea, I. B. (2014). A global strategy for road building. Nature, 513(7517), 229-232.

[41] Laurance, W. F., Goosem, M., & Laurance, S. G. (2009). Impacts of roads and linear clearings on tropical forests. Trends in ecology & evolution, 24(12), 659-669.

[42] Lean, H. H., Huang, W., & Hong, J. (2014). Logistics and economic development: Experience from China. Transport Policy, 32, 96-104.

[43] Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from China's high speed railway system. Journal of Urban Economics, 98, 98-123.

[44] Litman, T. (2002). Evaluating transportation equity. World Transport Policy & Practice, 8(2), 50-65.

[45] Liu, D., Sheng, L., & Yu, M. (2017). Highways and Firms' Exports: Evidence from China. Peking University. Working Paper.

[46] Liu, X. (2019). Characterizing broken links on national and local expressways in Chinese city-regions. Regional Studies, 53(8), 1137-1148.

[47] Lucas, K. (2012). Transport and social exclusion: Where are we now? Transport policy, 20, 105-113.

[48] Lyons, T. P. (1985). Transportation in Chinese Development, 1952-1982. The Journal of Developing Areas, 19(3), 305-328.

[49] Ministry of Transport (1993-2020). Road and Waterway Transportation Industry Development Statistics Bulletin (1993-2020).

[50] National Bureau of Statistics of China. China Statistical Yearbook 2000, 2001. Beijing: China Statistics Press

[51] Qin, Y. (2017). 'No county left behind?' The distributional impact of high-speed rail upgrades in China. Journal of Economic Geography, 17(3), 489-520.

[52] Redding, S., & Schott, P. K. (2003). Distance, skill deepening and development: will peripheral countries ever get rich? Journal of development economics, 72(2), 515-541.

[53] Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. Journal of international Economics, 62(1), 53-82.

[54] Roberts, M., Deichmann, U., Fingleton, B., & Shi, T. (2012). Evaluating China's road to prosperity: A new economic geography approach. Regional Science and Urban Economics, 42(4), 580-594.

[55] Rodrigue, J. P. (2020). The Geography of Transport Systems, Fifth edition. Oxon and New York: Routledge.

[56] Sachs, J. (2005). The End of Poverty: Economic Possibilities of Our Time. New York: Penguin Group.

[57] Sinomaps Press. Zhongguo Dituji (Atlas of China). Beijing: Sinomaps Press. Miscellaneous years.

[58] Smith, S. (2005). Ending Global Poverty: A Guide to What Works. London: Palgrave Macmillan.

[59] Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. The Review of economic studies, 83(3), 1263-1295.

[60] Taaffe, E. J., Gauthier, H. L. & O'Kelly, M.E. (1996). Geography of Transportation. New Jersey: Prentice Hall.

[61] Tong, T., & Yu, T. E. (2018). Transportation and economic growth in China: A heterogeneous panel cointegration and causality analysis. Journal of Transport Geography, 73, 120-130.

[62] Wang, J., Jin, F., Mo, H., & Wang, F. (2009). Spatiotemporal evolution of China's railway network in the 20th century: An accessibility approach. Transportation Research Part A: Policy and Practice, 43(8), 765-778.

[63] Wang, J., Du, D., & Huang, J. (2020). Inter-city connections in China: High-speed train vs. inter-city coach. Journal of Transport Geography, 82, 102619.

[64] Weinhold, D., & Reis, E. (2008). Transportation costs and the spatial distribution of land use in the Brazilian Amazon. Global Environmental Change, 18(1), 54-68.

[65] Wasserman, Stanley and Katherine Faust, (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

[66] Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., & Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature, 553(7688), 333-336.

[67] Woodcock, J., Edwards, P., Tonne, C., Armstrong, B. G., Ashiru, O., Banister, D., & Roberts, I. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. The Lancet, 374(9705), 1930-1943.

[68] World Bank (2000). World development report 2000: Attacking poverty. The World Bank.

[69] World Bank. (2005). World development report 2006: Equity and development. The World Bank.

[70] Yang, Y (2017). Transport Infrastructure, City Productivity Growth and Sectoral Reallocation: Evidence from China. University of California, Los Angeles. Working Paper.

[71] Yang, Z., Li, C., Jiao, J., Liu, W., & Zhang, F. (2020). On the joint impact of high-speed rail and megalopolis policy on regional economic growth in China. Transport Policy, 99, 20-30.

[72] Yu, Nannan, et al. "Does the expansion of a motorway network lead to economic agglomeration? Evidence from China." Transport Policy 45 (2016): 218-227.

[73] Zhang, Y. F., & Ji, S. (2019). Infrastructure, externalities and regional industrial productivity in China: a spatial econometric approach. Regional Studies.

[74] Zhang, W., Gai, Q., Zhu, X., & Shi, Q. (2023). Expressways and Poverty Reduction: Evidence from Rural China. The Singapore Economic Review, 1-30.

[75] Zheng, S., & Kahn, M. E. (2013). China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceedings of the National Academy of Sciences, 110(14), E1248-E1253.

Figures and Tables

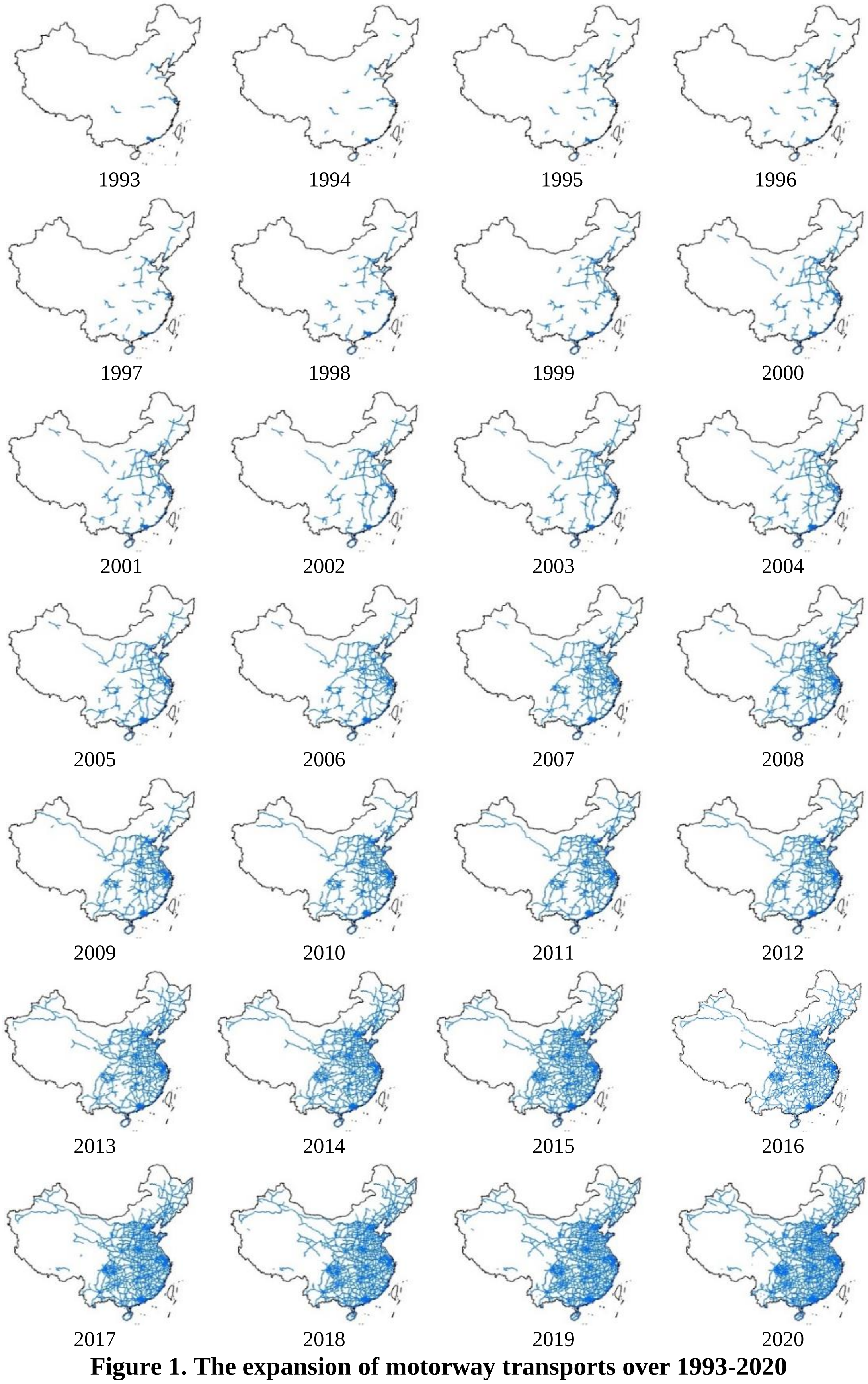
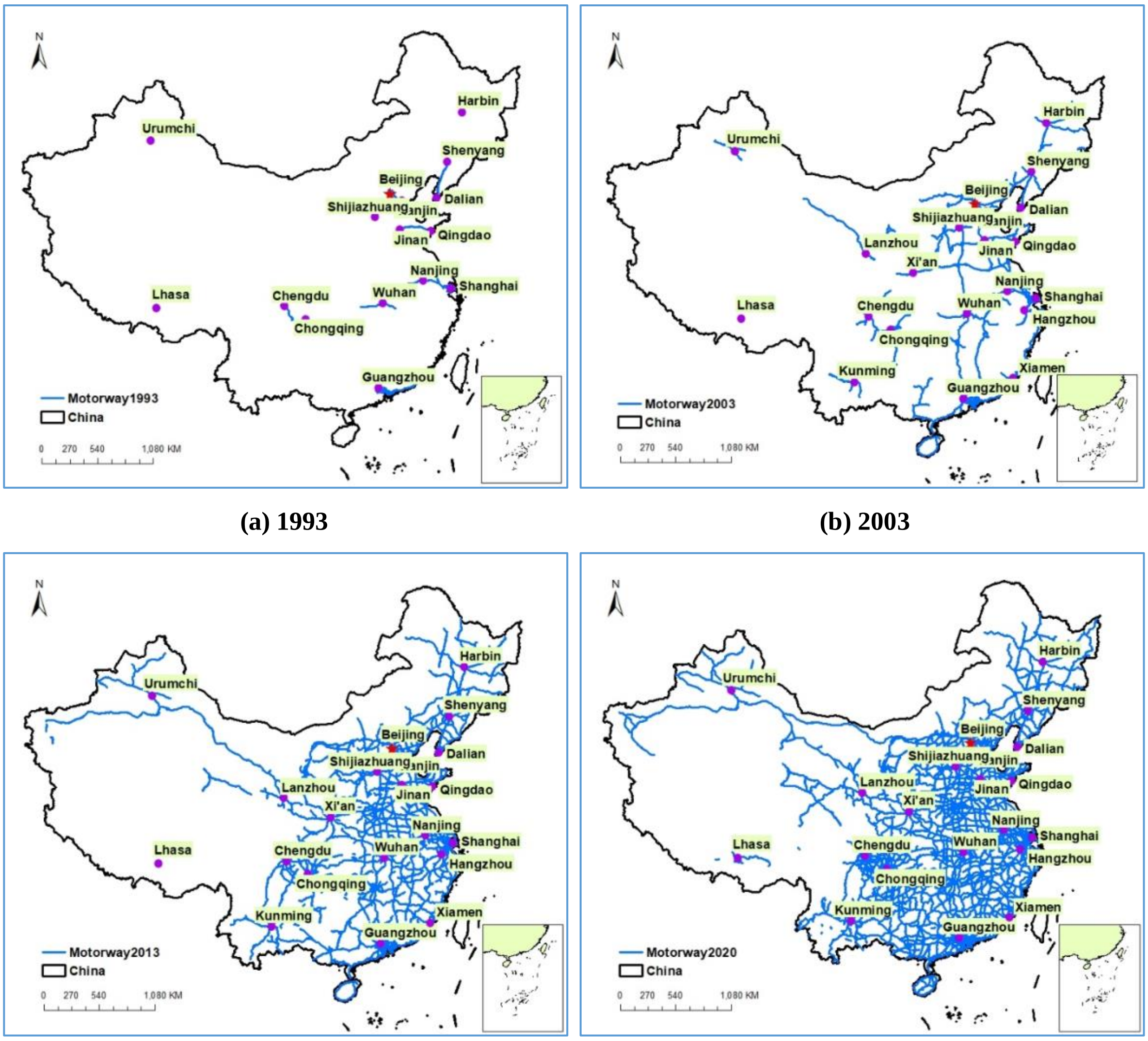


Figure 1. The expansion of motorway transports over 1993-2020

Note: The panels show the temporal evolution of motorway transports in China. *Data source*: GIS database

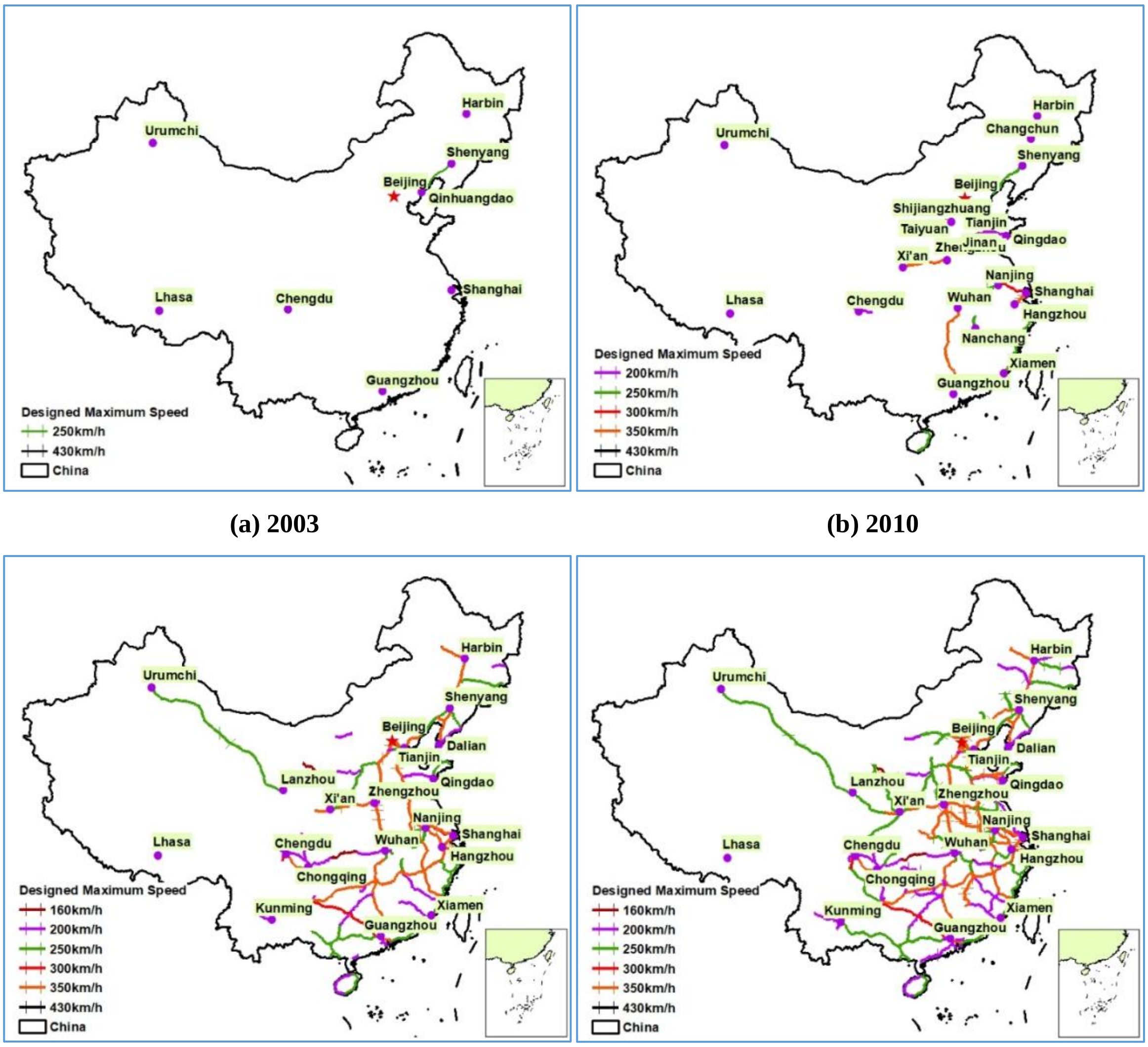


(a) 1993 **(b) 2003**

(c) 2013 **(d) 2020**

Figure 2. Motorway transports in 1993, 2003, 2013, and 2020

Note: The panels show the temporal evolution of motorway transports in China. *Data source*: GIS database

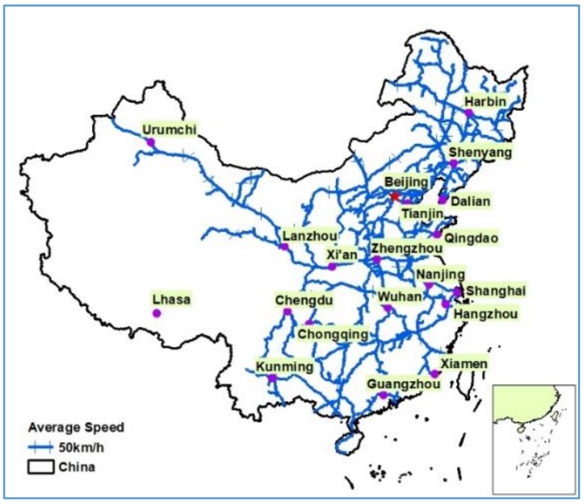


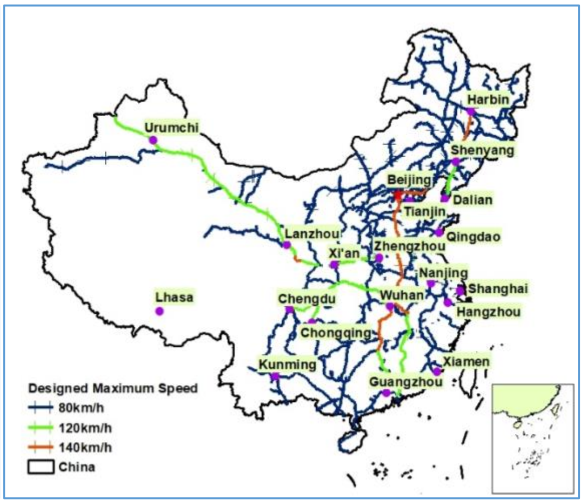
(a) 2003 (b) 2010

(c) 2015 (d) 2020

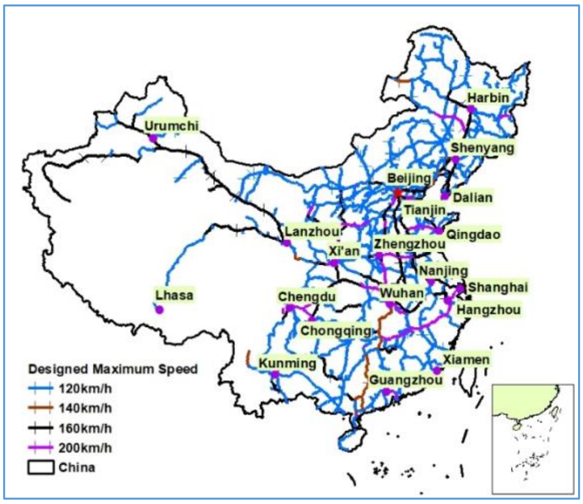
Figure 3. High-speed railways in 2003, 2010, 2015, and 2020

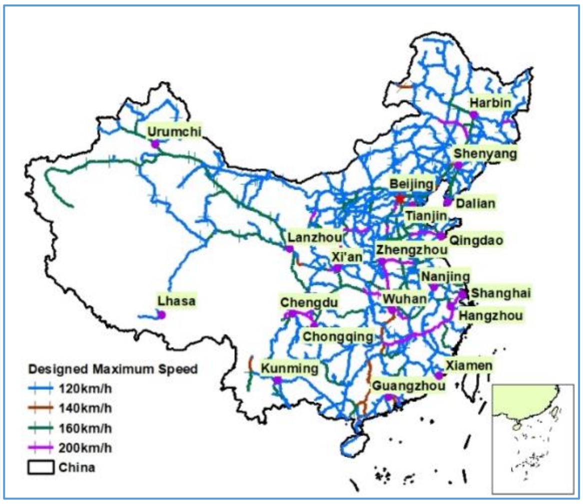
Note: The panels show high-speed railways in 2003, 2010, 2015, and 2020, respectively, including some lines with speed limits below 200 km/h. This is because some connections along high-speed rail lines have lower speed limits, and some lines are shared between high-speed and slower regular trains. For instance, part of the high-speed rail lines between Yichang and Enshi in Hubei province were designed as fast railways. Since high-speed trains also travel along this line, they reduce their speed in line with the lower maximum speed limit on these lines. *Data source*: GIS database





(a) 1995 **(b) 2003**

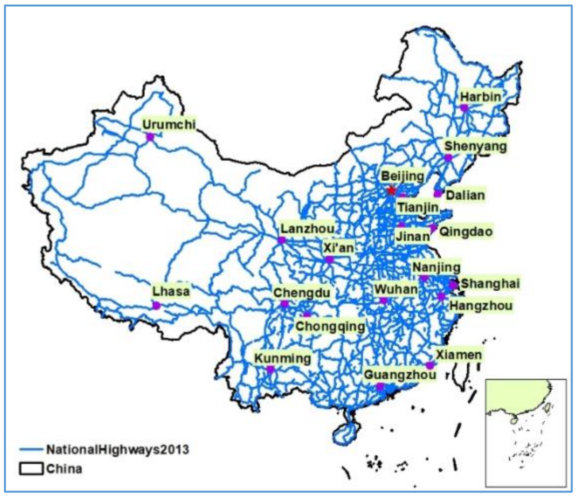




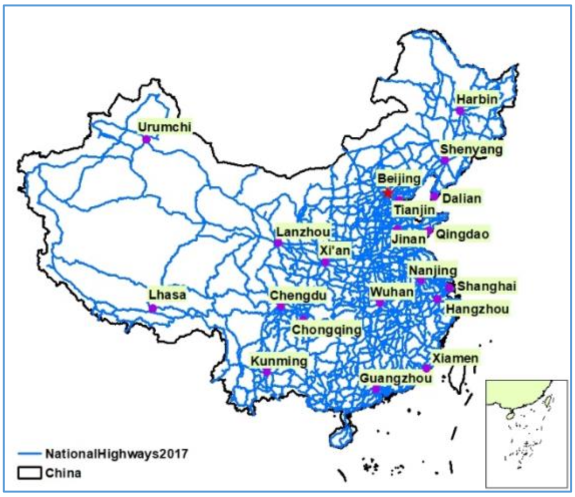
(c) 2013 **(d) 2020**

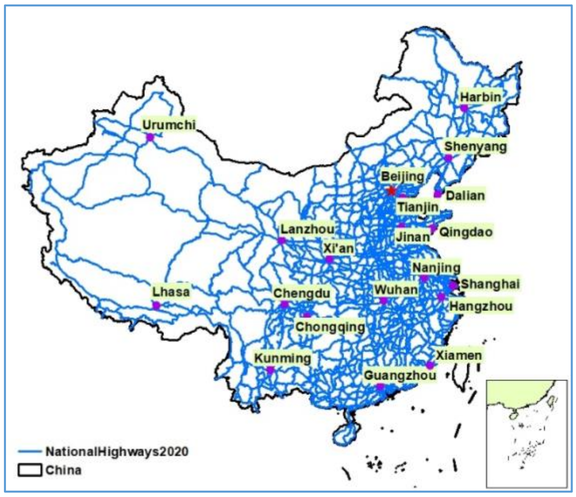
Figure 4. Regular railways in 1995, 2003, 2013, and 2020

Note: The panels show the temporal evolutions of regular railways in 1995, 2003, 2013, and 2020 in China. Regular railways include standard and fast railways. *Data source*: GIS database



(a) 1993 **(b) 2013**





(c) 2017 **(d) 2020**

Figure 5. National highways in 1995, 2003, 2013, and 2020

Note: The panels show the temporal evolutions of national highways in 1995, 2003, 2013, and 2020 in China.
Data source: GIS database

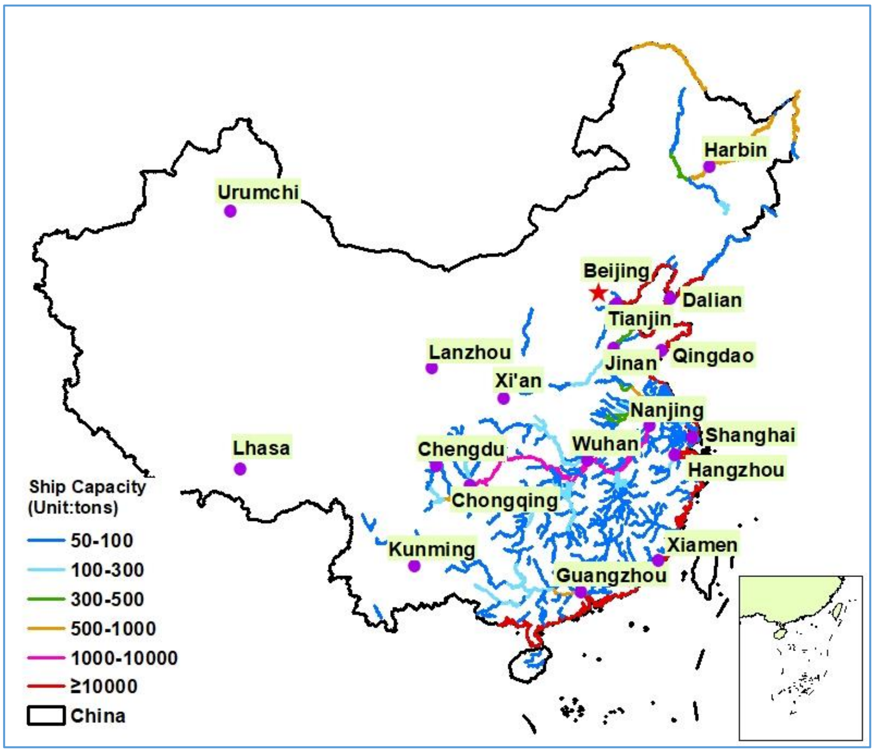
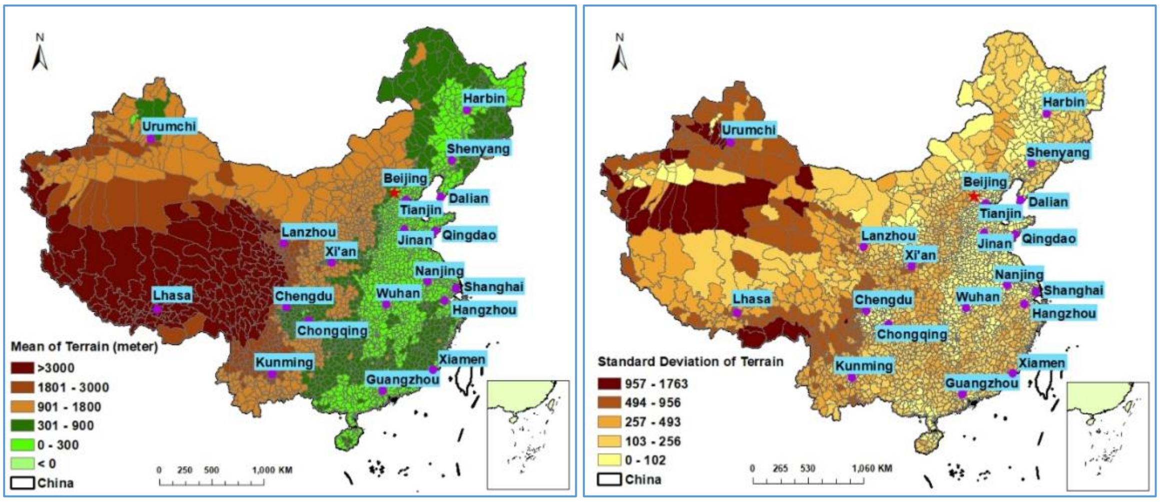


Figure 6. Navigable waterways in 1993

Note: The panels show the geographical distributions of navigable waterways in 1993 in China, with six grades of transporting capacities. Since there exists only a small increase of 15.8% in total inland waterway length over the past decades, from 110,200 km in 1993 to 127,700 km in 2020, we plot the navigable waterways in one year. We drop the waterways with less than 50 tons of capacity since they cannot be navigable channels used to transport goods, materials, or other movable objects. *Data source*: GIS database



(a) Mean elevation above sea level (b) Variability of elevation

Figure 7. Mean and variability of terrain elevations by county

Note: The panels show China's county-level mean and variability terrain elevations. Panel (a) presents the average elevation (meter) above the sea-level. The darkest brown denotes the highest elevation, such as Tibet, which is greater than 3,000 meters above sea-level, while the darkest green represents the lowest areas in the Eastern region. Panel (b) presents county-level ruggedness measured as the standard deviation (variability) of within-county elevations. The darkest brown denotes the most rugged counties with a standard deviation of elevation over 957-1,750 meters, such as Xinjiang and the Southwest border of Yunnan Province. *Data source*: GIS database

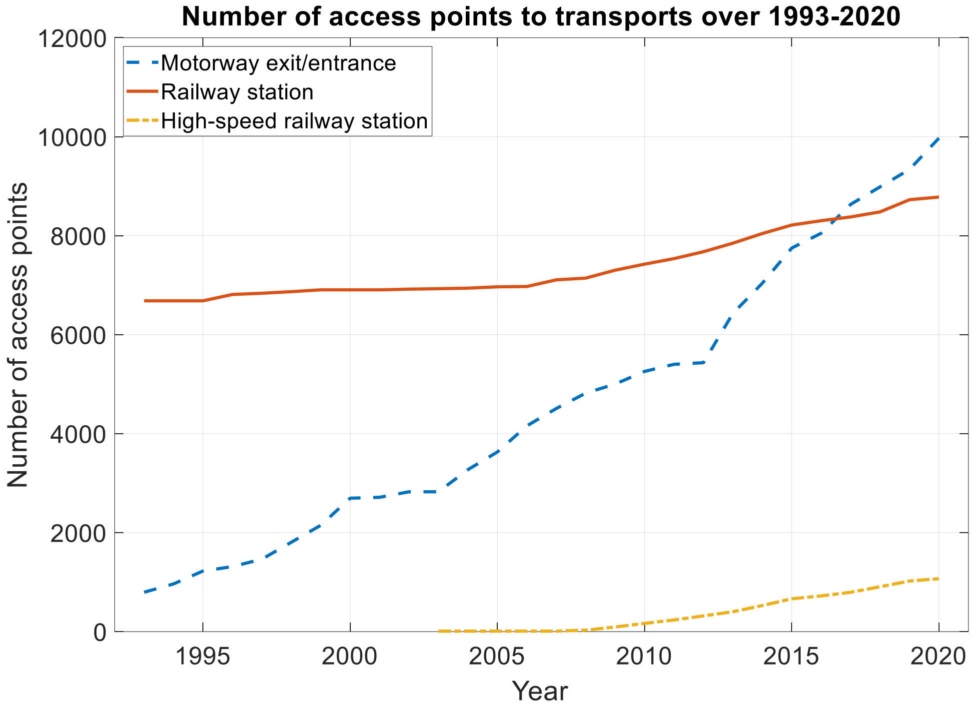
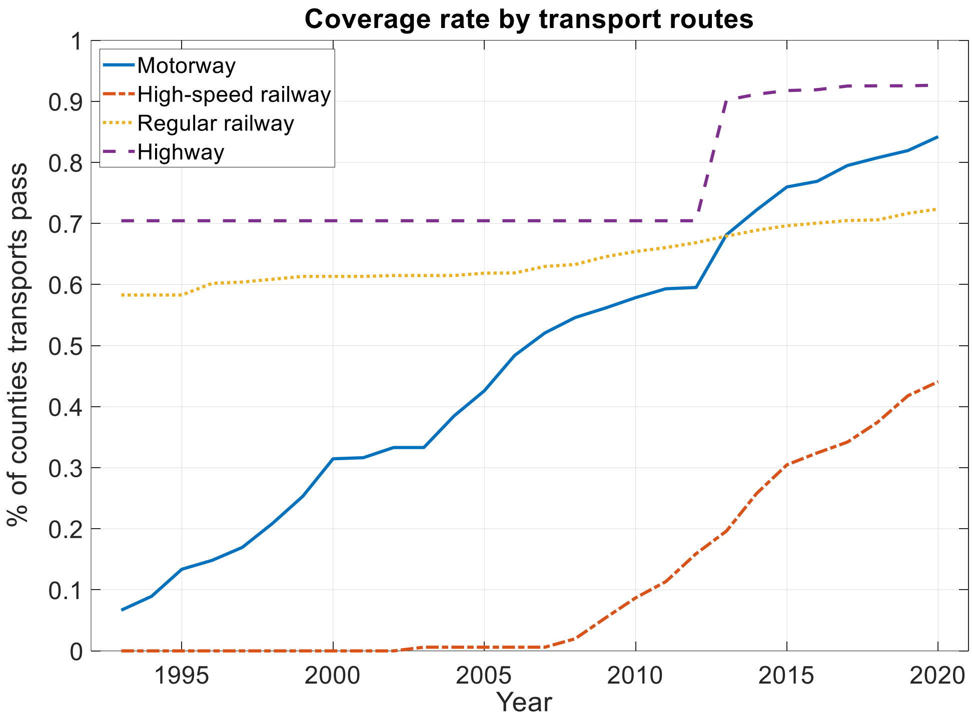
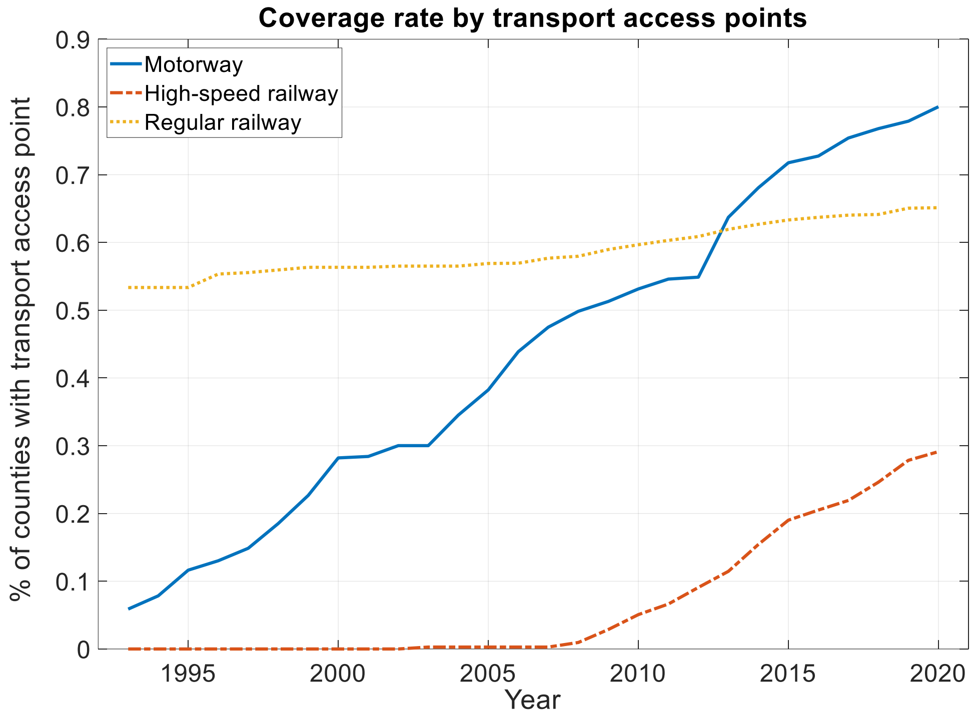


Figure 8. Number of access points to transports over 1993-2020

Note: This graph presents the number of access points to motorways, railways and high-speed railways from 1993 to 2020. The number of motorway exits/entrances increased by more than 12 times in the period, while the number of railway stations increased by 31%. High-speed railway stations initiated in 2003 and expanded to 1,069 in 2020. *Data source*: GIS database



(a) Coverage rate by transport routes



(b) Coverage rate by transport access points

Figure 9. Transport coverage at the county level over 1993-2020

Note: This graph presents the fraction of counties covered by transport networks from 1993 to 2020 in China. In Panel A, we define the counties as covered if transport routes pass through them, while we plot the fraction of counties with access to motorway entrances/exits, railway stations, or high-speed railway stations within its geographical boundary in Panel B. *Data source*: GIS database

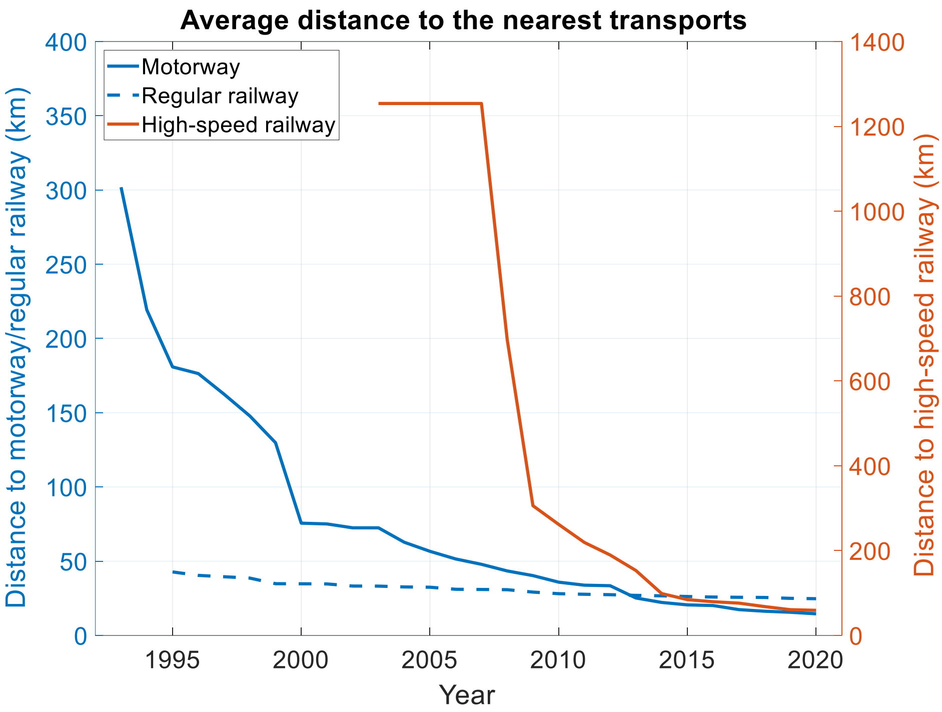
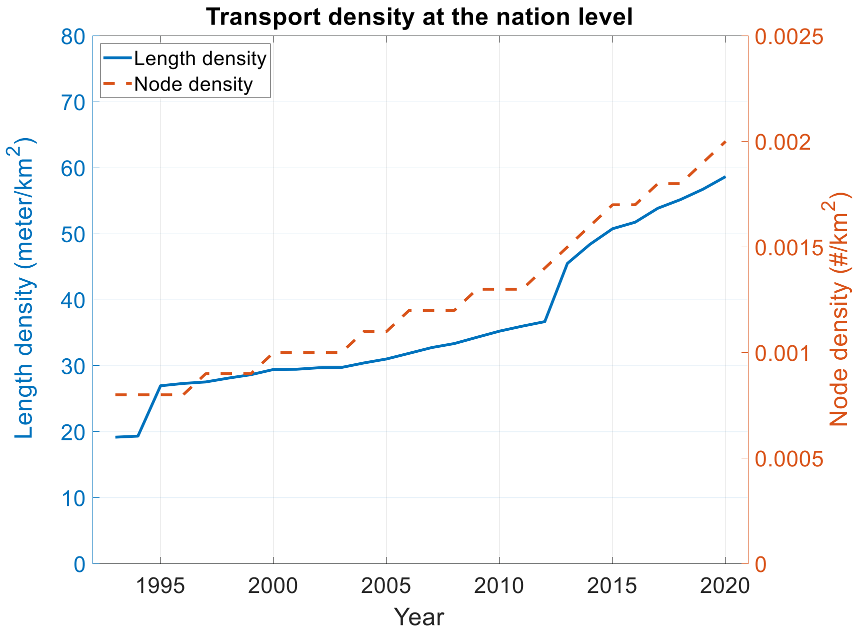
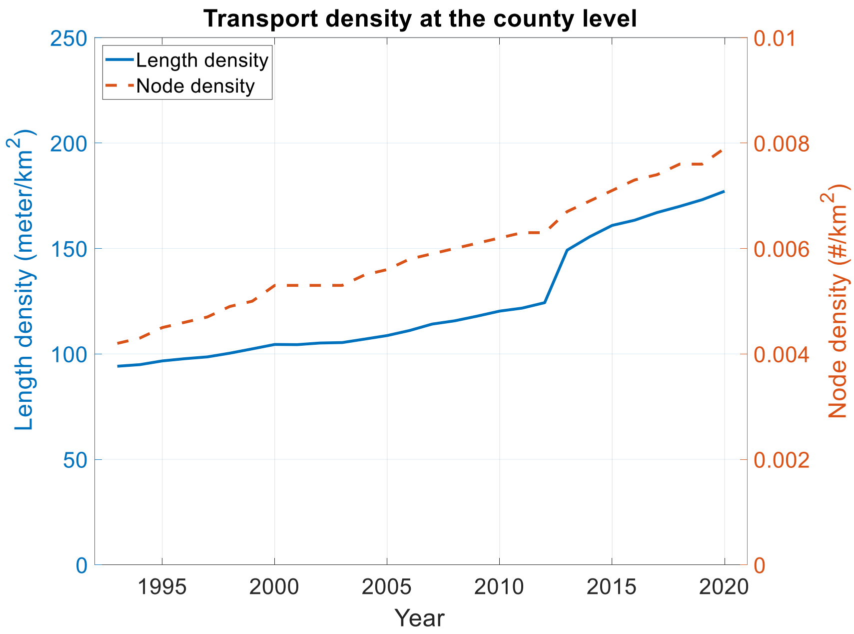


Figure 10. Average distance to the nearest transports

Note: The graph presents the average distance between 34,889 zip code centroids to the nearest motorway, regular railway, and high-speed railway in kilometers over 1993-2020 in China.



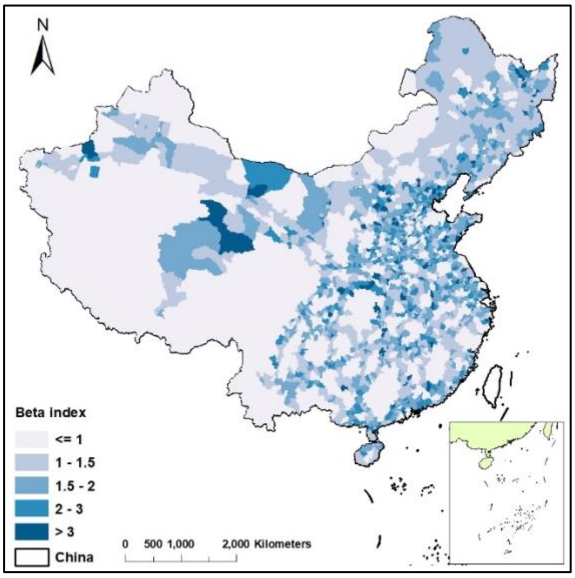
Panel (a): Transport density at the national level



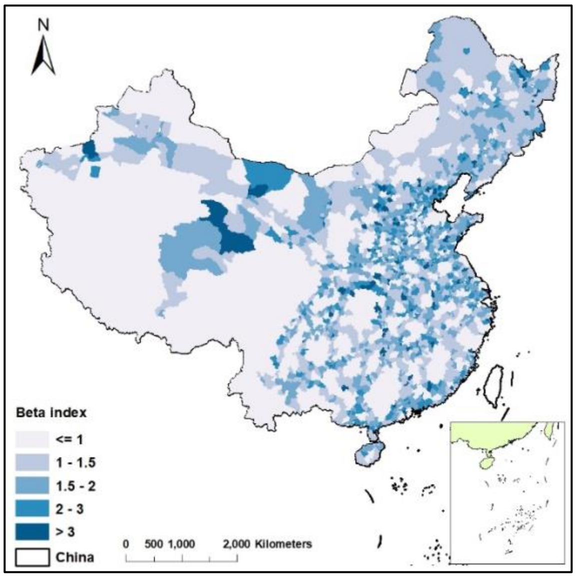
Panel (b): Transport density at the county level

Figure 11. Transport density of length and access points over 1993-2020

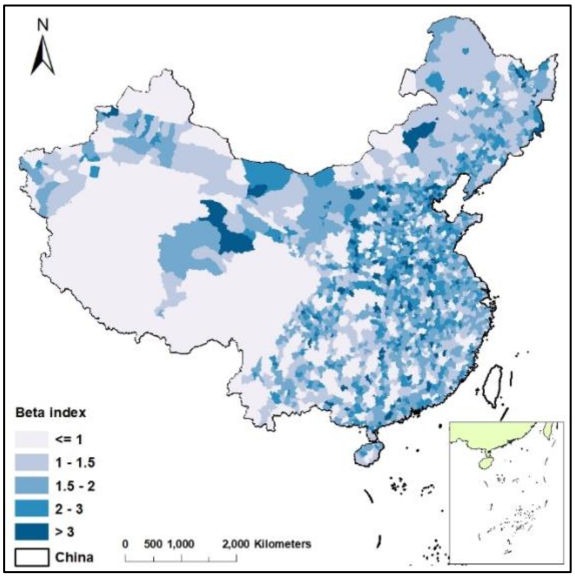
Note: The graphs present the transport density of length and access points in China from 1993 to 2020. Panel (a) shows the transport densities computed at the nation level, and Panel (b) exhibits the average densities at the county level. Length Density equals the ratio of total length of all transports to a county's area, including motorway, high-speed rail, regular rail, highway, and waterway, in units of meter/km^2, and Node Density is the ratio of number of access points to the area, for motorway exit/entrance and railway stations, in units of #/km^2.



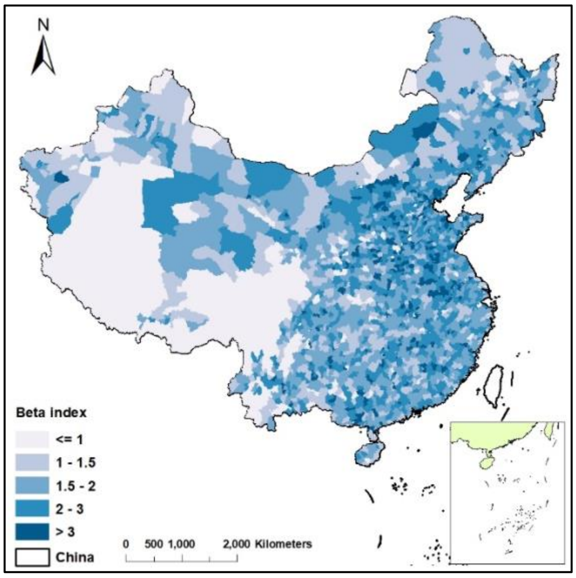
(a) 1993



(b) 2000



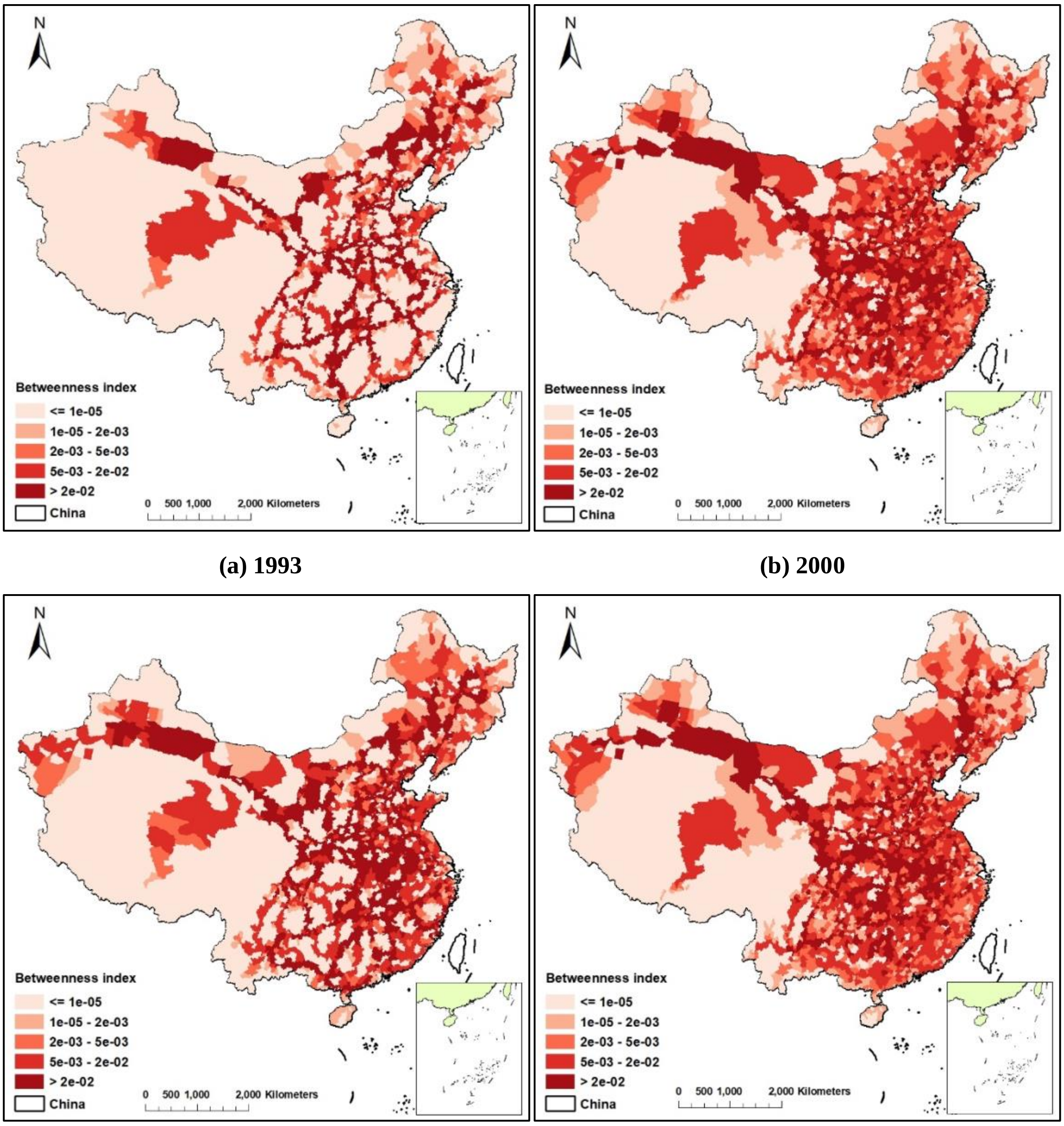
(c) 2010



(d) 2020

Figure 12. County-level connectivity in 1993, 2000, 2010, and 2020

Note: The panels show county-level connectivity (Beta index) of all transport modes in 1993, 2000, 2010, and 2020.

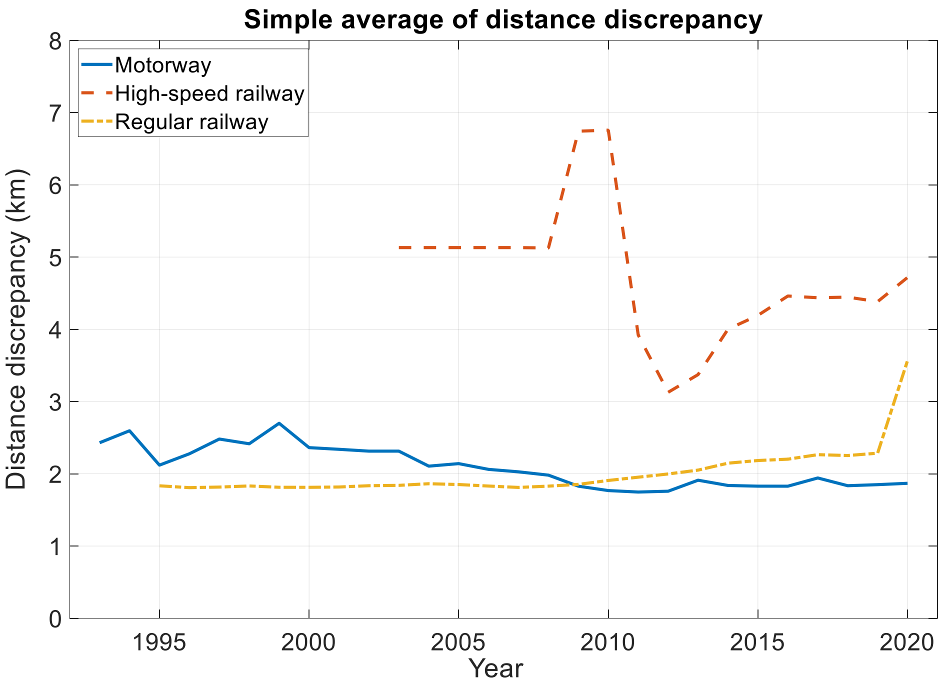


(a) 1993 (b) 2000

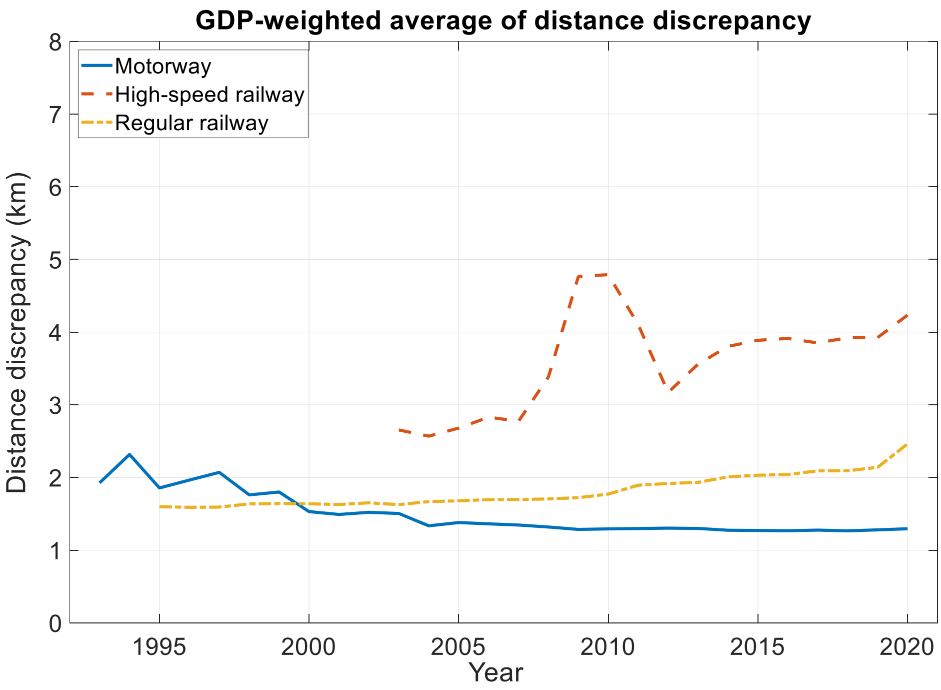
(c) 2010 (d) 2020

Figure 13. County-level centrality in 1993, 2000, 2010, and 2020

Note: The panels show county-level centrality (Betweenness index) of all transport modes in 1993, 2000, 2010, and 2020.



Panel (a): Simple average of distance discrepancy across counties



Panel (b): GDP-weighted average of distance discrepancy across counties

Figure 14. Distance discrepancy between access point and route to county centroid

Note: The panels show the discrepancy between the shortest distance to transport access points and routes over 1993-2020. We first compute two distance measures between county centroid and motorways, railways, and high-speed railways at the county level and then take the differences between the two distances. Panel (a) exhibits the simple average of distance discrepancies across counties, while Panel (b) shows the GDP-weighted average of distance discrepancies across counties.

Table 1. An overview of the GIS database of surface transports in China

Panel A: Surface Transport Modes			
Transport Mode	File Organization	Year	Shapefile
Motorways	Motorway network	1993-2020	Routes and max speed by route segment
	Motorway enter/exits	1993-2020	Motorway entry and exit points
Highways	Highway network	1993-2007, 2013-2020	Routes and max speed by route segment
Railways	Railway network	1995-2020	Routes and max speed by route segment
	Railway stations	1995-2020	Station locations and types
	Speed-up campaigns	1997, 1998, 2000, 2011, 2004, 2007	Railway additions and speed improvements
High-speed railways	High-speed railway network	2003-2020	Routes and max speed by route segments
	High-speed railway stations	2003-2020	Station locations
Waterways	Waterway network	1993	Waterways (inland and coastal) and tonnage grades
Panel B: Historical Data on Roadways and Railways			
Roadways	Road network	1962	Routes by route segment
Railways	Railway network	1962	Routes and stations
Panel C: Other Datasets			
Terrain features	Terrian	2002	Raster counts, mean, median, standard deviation, min, max, and the sum of elevations at the county level
Administrative boundaries	Boundary	2002	County and provincial boundaries
Cities and postal areas	Postal codes	2002	Coordinates of prefectural cities, provincial capitals, and postal areas

Note: This table summarizes the main elements of our database, including surface transport modes, historical data on roadways and railways, and some other datasets. All shapefiles can be layered onto each other in the GIS and used for analysis.

Table 2. Definition of transport quality measures in the database

Measure	Definition
Panel A: Location-specific measures	
Access indicator	Whether a location has access to transport within a distance
Access distance	The shortest distance to the nearest access point, i.e., motorway entrance/exit, railway station, or highway route
Centrality (Betweenness index)	Ratio of the number of links summed over all shortest paths in the network that pass through a county to the total number of links on the shortest paths summed over all node pairs in the network
Centrality (Degree index)	Ratio of the number of nodes connecting a county to the total number of nodes in the network
Centrality (Closeness)	Ratio of the total number of nodes in a network to the number of links along the shortest paths between a county and all other nodes
Panel B: Pairwise measures	
Travel Distance	Travel distance between locations via transport mode from one city to another or from a postal code centroid to a port
Travel Time	Shortest travel time between geolocations with a transport mode
Panel C: Network-level measures	
Centrality	Aggregations of location-specific centrality measures in a region
Route density	Transport length divided by its area in a region
Node density	Number of access points divided by its area in a region
Connectivity (Beta index)	Average number of links per node in a network (total number of links divided by total number of nodes
Connectivity (Gamma index)	Number of links in a network divided by the maximum feasible number

Note: This table lists all location-specific, pairwise, and network-level transport measures, as well as their definitions.

Table 3. Total length of surface transports

	(1) Railways	(2) High-speed rails	(3) Motorways	(4) Highways	(5) Railways	(6) High-speed rails	(7) Motorways	(8) Highways[1]
Year	Panel A: Statistics from our database				Panel B: Statistics from official documents			
1993	69.9		6.3	104.9	58.6		1.1	
1994	69.9		7.8	104.9	59.0		1.6	
1995	69.9		10.8	104.9	62.4		2.1	
1996	72.3		11.9	104.9	64.9		3.4	
1997	72.6		13.8	104.9	66.0		4.8	
1998	73.5		18.2	104.9	66.4		8.7	
1999	73.9		22.7	104.9	67.4		11.6	
2000	73.9		30.2	104.9	68.7		16.3	
2001	73.9		30.5	104.9	70.1		19.4	102.6
2002	74.4		32.2	104.9	71.9		25.1	99.9
2003	74.9	0.4	32.2	104.9	73.0		29.7	98.2
2004	75.0	0.4	38.6	104.9	74.4		34.3	95.5
2005	75.6	0.4	43.8	104.9	75.4		41.0	91.7
2006	75.8	0.4	51.8	104.9	77.1		45.3	88.1
2007	77.7	0.4	58.1	104.9	78.0		53.9	83.2
2008	79.2	1.0	62.4	104.9	79.7	0.7	60.3	95.0
2009	85.2	3.3	65.5	104.9	85.5	2.7	65.1	93.4
2010	89.3	5.0	70.2	104.9	91.2	5.1	74.1	89.9
2011	93.1	7.3	73.6	104.9	93.2	6.6	84.9	84.5
2012	99.0	10.5	74.2	104.9	97.6	9.4	96.2	77.2
2013	106.0	13.6	91.7	164.2	103.1	11.0	104.4	72.4
2014	114.0	19.6	102.3	173.2	111.8	16.5	111.9	67.3
2015	121.2	24.2	113.9	177.0	121.0	19.8	123.5	61.8
2016	123.8	26.3	118.6	179.1	124.0	23.0	131.0	223.8
2017	127.4	28.5	129.5	184.6	127.0	25.1	136.4	222.0
2018	131.4	32.3	136.0	186.6	131.7	29.9	142.6	220.4
2019	138.2	36.5	142.1	188.7	139.9	35.3	149.6	216.5
2020	143.2	39.0	153.4	190.7	146.0	38.0	161.0	209.7

Notes: This table lists and compares the total lengths of surface transports based on our database (Panel A) and those officially reported by China Statistical Yearbooks (Panel B). We treat all routes as single lines to compute lengths in our GIS database. [1]The China Statistical Yearbooks release total national roadways lengths that include highways and motorways defined in our GIS database. Thus, we subtract the total lengths of national roadways from motorways to attain the total lengths of highways. All lengths are measured in units of 1,000 kilometers. *Data source*: China Statistical Yearbooks and our GIS database, 1993-2020

Table 4. Correlation analysis between distance measures and distance discrepancy

	Distance to access points	Shortest straight-line to routes
Distance discrepancy	Full sample	
Motorway	0.0968***	0.0855***
	(27.3671)	(24.1474)
High-speed railway	-0.0957***	-0.1101***
	(-21.6939)	(-24.9956)
Regular railway	0.0792***	0.0428***
	(21.5458)	(11.6175)
Distance discrepancy	Eastern region	
Motorway	0.1565***	0.1228***
	(23.5455)	(18.3868)
High-speed railway	-0.2294***	-0.2389***
	(-28.1010)	(-29.3336)
Regular railway	0.0626***	-0.0677***
	(8.9830)	(-9.7181)
Distance discrepancy	Central region	
Motorway	0.2377***	0.1978***
	(34.0363)	(28.0658)
High-speed railway	-0.2805***	-0.2893***
	(-32.5796)	(-33.6935)
Regular railway	0.0601***	-0.0766***
	(8.0709)	(-10.2984)
Distance discrepancy	Northeastern region	
Motorway	0.0795***	0.0524***
	(7.0227)	(4.6205)
High-speed railway	0.6853***	0.6380***
	(66.4328)	(58.4923)
Regular railway	0.1302***	-0.0749***
	(11.1427)	(-6.3734)
Distance discrepancy	Western region	
Motorway	0.0684***	0.0598***
	(11.8740)	(10.3754)
High-speed railway	-0.2062***	-0.2128***
	(-29.2609)	(-30.2412)
Regular railway	0.0695***	0.0390***
	(11.6280)	(6.5142)

*Note***:** This table presents the correlation analysis at the county-year level between distance discrepancy and two kinds of distance, i.e., the distance to the nearest transport access point and the shortest straight-line distance to transport, using the whole dataset. Access points are exits/entrances for motorways and stations for railways. We consider distance discrepancy the difference between the shortest distance to a transport access point and the shortest distance to the nearest transport route from a county's centroid. We also divide the 2,819 counties into four regions in China, i.e., Eastern (785), Central (688), Northeastern (277), and Western (1,069), to explore the regional heterogeneity. The t statistics are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5. Effects of local socioeconomics on distance discrepancy

Panel A: Panel regression at the county level						
Distance discrepancy	Motorway		High-speed railway		Regular railway	
Log GDP	-.3735***	-.3951***	1.5112***	1.1561***	0.5441***	0.6057***
	(0.0639)	(0.0648)	(0.3086)	(0.3017)	(0.1321)	(0.1416)
GDP Growth (%)		-0.0087		-.1402***		0.0227*
		(0.0061)		(0.0310)		(0.0129)
GDP per Capita	0.0049	0.0051	-0.0030	0.0234	0.0732***	0.0712***
	(0.0118)	(0.0119)	(0.0413)	(0.0412)	(0.0193)	(0.0198)
GDP per Capita		0.0088		0.1414***		-0.0228**
Growth (%)		(0.0065)		(0.0310)		(.0111)
Controls	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	37,653	37,365	32,598	32,425	37,653	37,365
R^2	0.490	0.490	0.517	0.512	0.492	0.491
Panel B: Panel regression at the city level						
Distance discrepancy	Motorway		High-speed railway		Regular railway	
Log GDP	-.3011***	-.2937***	4.2581***	3.7041***	.1072	.2095
	(.0932)	(.0928)	(.7867)	(.8057)	(.1862)	(.2186)
GDP Growth (%)		-.0015		-.0514***		.0018
		(.0011)		(.0131)		(.0015)
GDP per Capita	.0266*	.0119	.2477***	.1879**	.0767*	.0738*
	(.0159)	(.0157)	(.0743)	(.0855)	(.0419)	(.0404)
GDP per Capita		.002565*		.07003***		-.0052**
Growth (%)		(.0013)		(.0162)		(.0021)
Controls	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	7,089	6,722	5,791	5,765	7,089	6,722
R^2	0.477	0.495	0.421	0.420	0.374	0.368

Note: This table presents the relationship between distance discrepancies and local socioeconomic factors through panel regressions. Panel A shows the results at the county level (excluding counties in the prefectural cities, while Panel B shows the results at the prefectural city level. Given the data availability, we conduct regressions for motorway and regular railways over 1999-2020 and high-speed railways over 2003-2020. The dependent variable is the distance discrepancy for the motorway, high-speed railway, and regular railway, measured in kilometers. The main explanatory variables include log GDP, GDP per capita, and GDP growth rate expressed as a percentage. We also include controls for commonly used local socioeconomic indicators, such as log population, annual government revenue and spending, aggregate household savings, loans, consumption, number of industrial companies, number of primary school students, and number of hospital beds. Year- and county-level fixed effects are also included as control variables. Local terrain characteristics, including elevation (average above sea level in the county/city), ruggedness (standard deviation of above sea level in the county/city), and other geographical factors, are excluded when incorporating county-level fixed effects. The robust standard errors are shown in the parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix

A. Existing Transport Databases in Other Studies

We extensively searched existing databases on China's transportation, including GIS data, digital maps, and statistical data sources in other academic studies. These studies fall into either the strand on the impact of transportation on various socioeconomic activities or geospatial analyses of transports themselves. Table B1 in the Appendix summarizes the search results and introduces other studies' transport datasets.

Panel (a) first presents other GIS data sources. The Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN) at Griffith University has been the most frequently used among all extant databases. The panel data covers a shorter period than ours and ends in 2011. Based on published atlases, the ACASIAN includes data on motorways over 1992-2011, national and provincial highways over 2007-2010, railways over 1997-2000, and waterways, but not on motorway exits/entrances and railway stations in China. The China Historical Geographic Information System (CHGIS) from Harvard University includes high-speed railways and stations in 2016 and road data that reflect 1990 conditions. The China in Time and Space (CITAS) database from the University of Washington includes roads, railroads, drainage systems, populated places, and urbanized areas and contours. It is historical data covering discrete periods from 1765 to 1994, with a large volume of local maps. It is also accessible through the Center for International Earth Science Information Network (CIESIN) at Columbia University. The National Fundamental Geographic Information System of China (NFGIS) from the Chinese Academy of Sciences has railway data in 2012.

We also introduce some studies on Chinese transport that build their own datasets in Panel (b). For example, Baum-Snow et al. (2015; 2016) obtained highway and railway data by digitalizing published maps from SinoMaps Press and Planet Maps Press. Jin et al. (2010) digitized transport lines using a remote sensing map. These studies cover some annual snapshots in a short timeframe and usually only one or two transport modes. In Panel (c), some other studies use statistics, rather than maps, of transports from Statistical Yearbooks at national, provincial, and regional levels. Others use the train timetable data in Panel (d).

In addition to those listed in Table B2, the China Transport Statistical Yearbooks have many statistics cited in many studies. The National Catalogue Service for Geographic Information (NCSG) data managed by the Ministry of Natural Resources of China includes accurate information on roads, e.g., motorways and high-speed railways, for 2015 in GIS. Another is the Resource and Environment Science and Data Center (RESDC) data from the

Chinese Academy of Sciences, which includes the road data in 1995 and GIS. However, to date, we have found no studies using these two data sources.

B. Additional Tables and Figures

Table B1. The summary of transport datasets in other studies

Data Source	Transport Mode	Time Span	Used in Studies	Transport Measure
Panel (a): Other GIS data sources				
(1) ACASIAN Data Center from Griffith University: https://acasian.com/price.html#china				
	Motorway	1992-2011	Faber (2014)	Whether a county is within 10 km of the route of the National trunk highway, 1998-2007
	Highway	2007-2010	Yang (2017)	Travel distance between cities by highway, railway, waterway, 1995-2005
	Railway	1997-2000	Liu et al. (2017)	Access to expressway and expressway density within the firm's radius, 2000-2006
	Waterway	1965-2010	Lin (2017)	Travel time to port through high-speed railway and highways, 2003-2014
			Huang and Xiong (2017)	Travel distance and time between prefecture-level cities by expressway, 1998-2007
			He, Xie, and Zhang (2020)	Whether any highway pass through the county, 1992-2010
			Fan (2019)	Travel time on roads used for market access, 1999, 2010
(2) CHGIS from Havard University: https://dataverse.harvard.edu/dataverse/chgis				
	Highway, railroad, and high-speed railway	1990	Banerjee et al. (2012)	The shortest distance to network route, straight-line, railroad, highway, and navigable waterways, 1990
	Waterway	2012		
(3) CITAS from University of Washington: https://citas.csde.washington.edu/				
	Road, railway, River, and lake	1765-1994, 2000, 2004	Emran and Hou (2013)	Access to transportation network through road, railway, navigable waterways, 1990, 2000, 2004
(4) NFGIS from Resource and Environment Science and Data Center (RESDC) from Chinese Academy of Sciences: https://www.resdc.cn/				
	Railway, road	2012	Jiao et al. (2014)	Travel time between cities via expressway, highway, and high-speed railway, 2012
Panel (b): Digital and GIS data collected by other authors				
(5)	Railway, motorway, national highway, port, airport	2006	Jin et al. (2010)	Density, proximity, accessibility, travel distance between cities, through railway, motorway, national highway, port, and airport, 2006
(6)	Railway and highway	1924, 1962, 1980, 1990, 1999, 2005, and 2010	Baum-Snow (2017)	Number of railroads and highway lines 5 and 10 km from CBD 1990, 2000, and 2010
			Baum-Snow (2020)	Travel time used for market access through highway and railway, 1990, 1999, 2005, and 2010
(7)	Railway, motorway, national highway	1911, 1935, 1953, 1981 and 2012	Hu et al. (2015)	Density, proximity, and accessibility of railway, motorway, and national highway, 1911, 1935, 1953, 1981 and 2012
(8)	Railway	11 discrete years from 1906 to 2000	Wang et al. (2009)	Connectivity, travel distance between cities through railway, 1906, 1911, 1925, 1937, 1949, 1957, 1965, 1974, 1981, 1988, and 2000

(9)	Motorway and national highway	2013, 2015	Liu (2019)	Compare planned and finished motorway and national highway, 2013, 2015
(10)	High-speed railway	2006-2010	Zheng and Kahn (2013)	Travel time and market access through high-speed railway, 2006-2010
		2004, 2007	Qin (2017)	Whether high-speed railway was upgraded, 2004, 2007
Panel (c): Data from China Statistical Yearbooks				
Transport mode: railway, high-speed railway, highway, navigable river, and port				
(11) China Statistical Yearbook from National Bureau of Statistics (NBS): http://www.stats.gov.cn/tjsj/ndsj/				
		1981-2020	Lean et al. (2014)	Length of railway, highway, number of deep-water berths, 1980-2009
			Tong and Yu (2018)	Freight transportation per capita, 2000-2015
(12) China City Statistical Yearbook from National Bureau of Statistics (NBS): https://data.cnki.net/Yearbook				
		1985-2020	Yu et al. (2016).	Density of motorway in each city, 2000-2010
			Yang et al. (2020)	Whether a city is connected to a high-speed railway, 1996-2017
(13) Regional Economic Statistical Yearbook from National Bureau of Statistics (NBS): https://data.cnki.net/Yearbook				
		2000-2014	Roberts et al. (2012)	Expenditure on national motorways, 2007
(14) China Provincial Statistical Yearbooks from National Bureau of Statistics (NBS): https://data.cnki.net/Yearbook				
		1983-2020	Deng et al. (2014)	Density of highway, 1987- 2010
Panel (d): Train Timetable Data				
(15) National railway passenger train schedules from National Railway Administration: http://www.nra.gov.cn/ and http://www.tielu.org				
	Regular railway and high-speed railway	1957-2016	Jiao et al. (2017)	Connectivity and centrality of high-speed railway, 2003-2014
			Diao (2018)	Travel time and accessibility through high-speed railway, 2009, 2010, 2012, 2013
(16) Train timetable data from official website: www.12306.cn				
	Railway, high-speed railway	Periodical	Huang and Zong (2020)	Connectivity and centrality of high-speed railway, 2016, 2019
			Wang et al. (2020)	Connectivity and centrality of high-speed railway, 2018

Note: The table presents the summary of transport datasets in other studies, including other GIS data sources, digital and GIS data collected by other authors, China Statistical Yearbooks, and timetable data. We outline the data sources, transport mode, time span, transport measures, and studies that use these data.

Table B2. Correlation analysis in distance discrepancy between the East and West

	Distance to access points	Shortest straight-line to routes
Distance discrepancy	Full sample	
Motorway	0.0968***	0.0855***
	(27.3671)	(24.1474)
High-speed railway	-0.0957***	-0.1101***
	(-21.6939)	(-24.9956)
Regular railway	0.0792***	0.0428***
	(21.5458)	(11.6175)
Distance discrepancy	East region	
Motorway	0.1565***	0.1275***
	(27.3228)	(22.1664)
High-speed railway	0.0697***	0.0399***
	(9.6598)	(5.5207)
Regular railway	0.0800***	-0.0626***
	(13.3357)	(-10.4222)
Distance discrepancy	West region	
Motorway	0.0792***	0.0692***
	(17.6217)	(15.3852)
High-speed railway	-0.2062***	-0.2131***
	(-37.4742)	(-38.7868)
Regular railway	0.0787***	0.0464***
	(16.8728)	(9.9277)

Note: This table presents the difference between East and West for correlations between distance discrepancy and two kinds of distance, i.e., the distance to the nearest transport access point and the shortest straight-line distance to transport, at the county-year level. To explore the regional heterogeneity, we divide the 2,819 counties into Eastern (1,062) and Western (1,757) regions. Here, we consider Eastern and Northeastern regions to be *East*, and Central and Western regions to be *West*. The *t* statistics are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

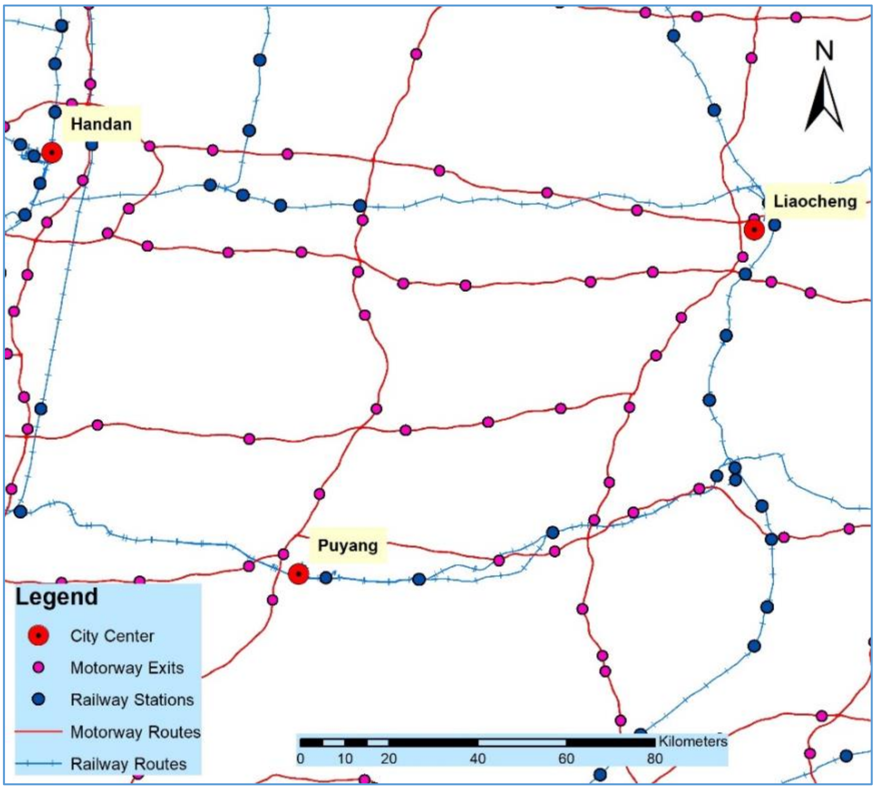


Figure B1. An example of motorways and railways in our GIS database

Note: The figure is a map scaled at 1:2,000,000. Blue lines are railways, and red lines are motorways. Points represent railway stations and entrances/exits. *Data source*: GIS database