www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

A multi-regional input-output database linking Chinese subnational regions and global economies

Xiaoqing Hou[1,9], Jiali Zheng[1,2,9] ✉, Minna He[1], Gengzhong Feng[1,3], Kehan He[4], Tao Ma[2] ✉, D'Maris Coffman[5,6], Zhifu Mi[5] ✉ & Shouyang Wang[7,8]

Recent efforts of nation states to enhance resilience and restructure global industry, supply chains, and value chains have intensified the dual economic structure that shapes flows at both domestic and international levels. Multi-regional input–output (MRIO) tables provide a quantitative approach to capture inter-regional and inter-sectoral economic flows. Existing MRIO databases mainly consist of international MRIO and subnational MRIO, respectively. However, they lack coupling MRIO tables and thus fail to adequately represent the dual structure of subnational and international trade. In this study, we develop a subnational and international coupling (SIC) MRIO in 2017, covering 30 sectors, 30 provinces in China and 66 countries & economies. We employ macro-level aggregated data, micro-level structural data and large language models (LLMs) to construct the SIC method. Our SIC MRIO captures the interplay between international and subnational economic flows, reflecting the emerging dual structure of production and consumption networks in the global landscape. It can serve as a baseline for future studies of the economic and environmental implications of chains under the dual structure.

Background & Summary

Recently, economic conflicts such as the US-China trade war have intensified a transition from the hyper-globalization era toward greater regionalization, digitalization, and strategic nationalism. While cross-border flows of data, services, and capital remain robust, trade in goods and physical capital is fragmenting, giving rise to a more multipolar, risk-aware, and locally oriented global economy. Industry, supply chains, and value chains are facing challenges to resilience and restructuring, with trends toward localization, regionalization and diversification[1,2]. Against this backdrop, a dual structure covering economic flows within and among nations at both the subnational and international levels, such as provinces in China, states in the U.S. and countries in the EU, is becoming increasingly important, and with substantial implications for both the economy and the environment[3]. The motivation for our work stems from the observed discrepancies between coupled multi-regional input–output (MRIO) tables and separately subnational and international MRIOs[4]. The coupling MRIO provides higher accuracy, which can reveal key issues and uncover findings that separate tables cannot reflect. As a result, the subnational and international coupling (SIC) database offers critical implications for the study of supply chains, global value chains (GVC), and embodied carbon emissions, facilitating more robust policy analysis and targeted strategies.

[1]The School of Management, Xi'an Jiaotong University, Xi'an, 710049, China. [2]Business School, Harbin Institute of Technology, Harbin, 150001, China. [3]Key Laboratory of Process Control & Efficiency Engineering (Xi'an Jiaotong University), Ministry of Education, Xi'an, 710049, China. [4]Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong, Hong Kong SAR, China. [5]The Bartlett School of Sustainable Construction, University College London, London, WC1E 7HB, UK. [6]Department of Earth System Sciences, Tsinghua University, Beijing, 100084, China. [7]State Key Laboratory of Mathematical Sciences, Chinese Academy of Sciences, Beijing, 100190, China. [8]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China. [9]These authors contributed equally: Xiaoqing Hou, Jiali Zheng. ✉e-mail: zhengjiali@amss.ac.cn; matao@hit.edu.cn; z.mi@ucl.ac.uk

Table name	Level	Year	Country/Province	Sector
GTAP[40]	International	2004, 2007, 2011, 2014, 2017	141	65
OECD[41]	International	1995–2020	71	45
EXIOBASE[42]	International	1995–2022	49	163
EMERGING[43]	International	2015–2019	245	135
Li *et al.*[44]	Subnational	1997, 2002, 2007, 2012, 2017	30 (31 in 2012)	33 (42 in 2012)
Mi *et al.*[45]	Subnational	2012	30	30
CEADs[46]	Subnational	2012, 2015, 2017	31	42
Meng *et al.*[18]	International and Subnational	2005, 2007	5 countries, 4 regions in China, 9 regions in Japan	8

Table 1. Summary of MRIO table studies.

MRIO tables at international and subnational levels differ substantially in their sector coverage, regional scope and time span, posing challenges for integration[5]. As quantitative tools for examining interregional economic structures, MRIOs are widely used in regional economics. International MRIO databases such as the global trade analysis project (GTAP), the organization for economic co-operation and development (OECD), EXIOBASE, the world input-output database (WIOD), The EMERGING database (EMERGING), and the EORA are prominent. GTAP covers 160 countries; EXIOBASE includes 163 sectors; OECD offers long time-series data (1995–2018); WIOD emphasizes value chains and environmental extensions; EMRGING contains 135 sectors across 245 economies; EORA provides high spatial and sectoral resolution (Table 1). MRIO tables play an important role in capturing economic flows[6] and their environmental effects[7] in the dual structure. The advantage of using international MRIO data lies in its global perspective, which enables macro-level analysis of ecological systems[8], climate systems[9], technological supply chains[10], etc. This approach describes and analyses international interrelationships between the consumption and production sectors[11]. However, the lack of coupling with subnational MRIO leads to a loss of important details of subnational regions. In contrast to international MRIO tables, subnational MRIO data provide detailed insights into structural economic changes and environmental policy impacts within countries, including China[12], the EU[13], Japan[14], etc. Some studies have extended their scope to inter-city input-output relationships among a few adjacent cities[15], and even across entire national cities[16]. However, subnational MRIO databases typically include only a single column for exports, offering limited information on international trade. This makes it difficult to evaluate the role and contribution of subnational regions or cities in the global economy, as represented by the division of labor in GVC. Some databases provide both international and subnational MRIO, such as the carbon emission accounts and datasets (CEADs). The linkage between subnational and international MRIO tables has been advanced through the development of region-extended Inter-Country Input-Output (ICIO) tables[17], in which linear programming is applied to harmonize macroeconomic data from customs and the Broad Economic Categories (BEC)[18]. However, the advent of frontier technologies, including machine learning, deep learning, large language models (LLMs), etc., has enabled the exploration of advanced methods such as model coupling in this field[19,20].

Regarding the measurement of the dual structure, existing research is generally divided into macro-level and micro-level approaches. Macro-level studies often assess the dual structure under the assumption that subnational and national trade structures are identical. This approach, however, would overlook subnational heterogeneity and compromise the accuracy of the data. For example, existing research has disaggregated the Chinese portion of the GTAP database using provincial shares from China's provincial MRIO tables[21,22], or integrated the Asian International Input-Output (IOT) database with China's MRIOs under the assumption of uniform regional import structures[23]. Due to the unavailability of subnational import and export data, existing studies commonly assume that subnational trade patterns resemble those at the national level, thus failing to capture the provincial distinction of economic development and participation in the initernatonal division. Micro-level studies typically rely on firm-level trade data to measure either domestic or foreign trade in isolation. For instance, foreign trade studies have used multinational enterprises' investment data[24], value-added firm evidence[25], etc., and domestic trade studies have integrated datasets such as capital and tax information[26,27]. This approach fails to capture the structural linkages between subnational and international trade within the dual structure.

While artificial intelligence techniques such as machine learning, deep learning and large language models (LLMs) have been widely applied in data-intensive fields[28,29], their application to input-output estimation remains limited. Recent methodological advances in input-output table estimation have introduced innovative computational approaches that substantially improve upon traditional techniques. Hybrid algorithms have substantially improved the accuracy of interregional input-output table estimation. These include combinations of the RAS method with a real-coded genetic algorithm[30] and a boundary tightening algorithm[31]. In addition, advanced balancing algorithms that can handle unequal net row and column sums[32] have also shown strong performance. Similarly, deep neural network architectures have shown superior performance in forecasting input-output tables. They perform well in both short-term and long-term scenarios. This approach has substantially reduced prediction error rates[33]. More recent evidence suggests that intelligent methods make a difference in the future inference of the IO table; however, the potential of artificial intelligence in MRIO table compilation remains largely untapped.

Data	Source
Inter-country input-output tables (ICIO) 2017	OECD
China MRIO 2017	EST_MRIO[47]
Chinese customs import and export data	General Administration of Customs of the People's Republic of China[34]
The service trade import and export table (STIE)	China Business Yearbook
China's balance of payments	State Administration of Foreign Exchange of China[48]
Conversion tables between the HS 2002 and Classification by broad Economic Categories	United Nations Statistics Division
Industry and HS code comparison table	Sheng *et al.*[35]
Firm-level import and export logistics data	Alayun Global Trade Logistics Technology Service Platform (Alayun Platform)[49]
American firm-level import and export logistics data	ImportYeti Platform[50]
China's non-competitive input-output table	State Statistics Bureau[38]
Sector classification corpus	Appendix II of China's Input-Output Table 2007: China's Input-Output Sector Classification Explanation and Code

Table 2. The list of data.

This study compiles a subnational and international coupling MRIO (SIC-MRIO) covering 30 provinces in China, 66 countries & economies worldwide, and 30 sectors, thereby linking international MRIO with subnational MRIO. The main contributions of this study are as follows. First, our methodology incorporates LLMs into the compilation of MRIO tables, to compensate for missing critical information and frameworks. Second, in addition to the applications of mathematical and planning methods, we use a large amount of real data to fill in the microstructure between the international and subnational sectors. Third, we take advantages of both the macro-level and the micro-level research perspectives, our model integrates macro-level aggregate data (e.g., national customs import–export data and service trade data) with micro-level structural data (e.g., firm-level import and export logistics data).

Methods

Data collection. The SIC is constructed based on a comprehensive data framework, integrating extensive trade information to capture subnational and international economic linkages. The SIC is compiled using a partial-survey method. It describes the main industrial input-output relationships through more than 2 million subnational and international trade data points. This study focuses on 30 Chinese provinces and 66 other countries & economies using 2017 as the base year. The trade data include more than 5,200 commodities, which are classified into 30 economic sectors following the Harmonized System (HS) Codes. Table 2 provides an overview of the raw data used in SIC construction. In this table, we use web crawlers on the customs website to obtain "Chinese customs import and export data". We collect "Firm-level import and export logistics data" and "American firm-level import and export logistics data" from the Alayun and ImportYeti websites respectively, and "China MRIO 2017" is based on our previously published work in *Environmental Science & Technology* (EST). And we account all the MRIO tables to reflect values at producers' prices.

The basic structure of SIC. The MRIO tables link multiple domestic input-output tables to improve data consistency within and across economies. The standard MRIO structure consists of three main components: I intermediate matrix (Z) section, II final demand (FD) section, III value added (VA) section. The subnational MRIO and international MRIO share similar structures, as shown in Fig. 1. In the figure, Ctry denotes a country; Prov, a province; S-MRIO and I-MRIO represent the subnational and international MRIOs, respectively. SIC-MRIO_ZExport refers to China's exports in the intermediate matrix section of the SIC, while SIC-MRIO_ZImport represents China's imports in the same section. SIC-MRIO_FDExport and SIC-MRIO_FDImport denote China's exports and imports in the final demand section, respectively. Figure 1a,b illustrate their layouts. However, integrating them is challenging. Figure 1c shows the subnational-international dual structure (grey squares). It connects international trade relations (green squares) with subnational trade relations (red squares). As we have mentioned in the first section, there are severe data gaps in the coupling process. To address this, we construct two key datasets: the subnational-international import table (SIC_Import) and the subnational-international export table (SIC_Export). In addition, the MRIO_Z section adopts EST_MRIO developed by our research team, while the ICIO_Z section uses OECD_ICIO.

The main work is divided into macro-level and micro-level components. We compile the SIC_Import and the SIC_Export using both macro and micro frameworks, as shown in Fig. 2. We calculate SIC_Import and SIC_Export by applying the transnational sector–sector trade ratio matrix, derived from a micro perspective, to the province–country total trade values estimated from a macro perspective. The macro dataset includes over 740,000 records from the General Administration of Customs, while the micro dataset contains more than 1,050,000 records from the Alayun Platform and 500,000 records from the ImportYeti Platform.

Data processing from a macro perspective. We calculate the province-country total value (goods) and province-country total value (services) through the steps shown in Fig. 2.

The provincial goods trade import and export data provided by General Administration of Customs of the People's Republic of China, include the following indicators: "HS code", "product name", "registration location

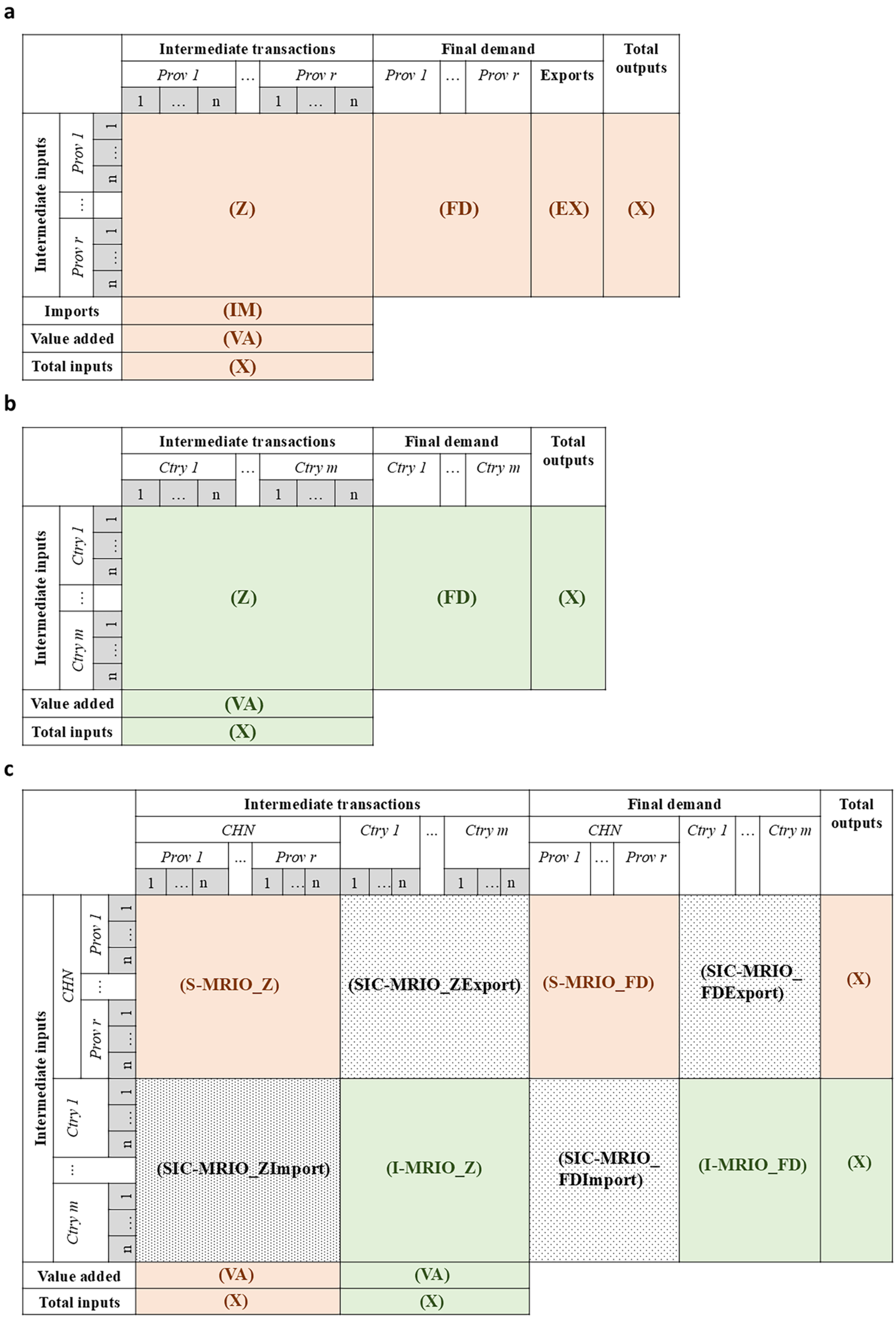

Fig. 1 Main structure of MRIO tables. (**a**) International MRIO; (**b**) Subnational MRIO; (**c**) SIC MRIO.

code", "registration location name", "trade partner code", "trade partner name", and "transaction value in RMB"[34]. Provinces are identified using registration location names, and trade partner countries are identified using the trade partner name. It is worth noting that the trade partner category for imports includes China itself, indicating the presence of domestic (China-to-China) trade flows. Supplementary Information 2 provides further details. The HS code, administered by the United Nations Statistics Division, serves as the global standard for product classification in international trade. This HS code employs a six-digit numerical framework to systematically categorize traded commodities, enabling uniform identification and description of products. We compile the goods sectors and HS code comparison table by extending the classification from industry to goods sectors based on the industry and HS code comparison table[35]. In addition, we use HS code to differentiate economic

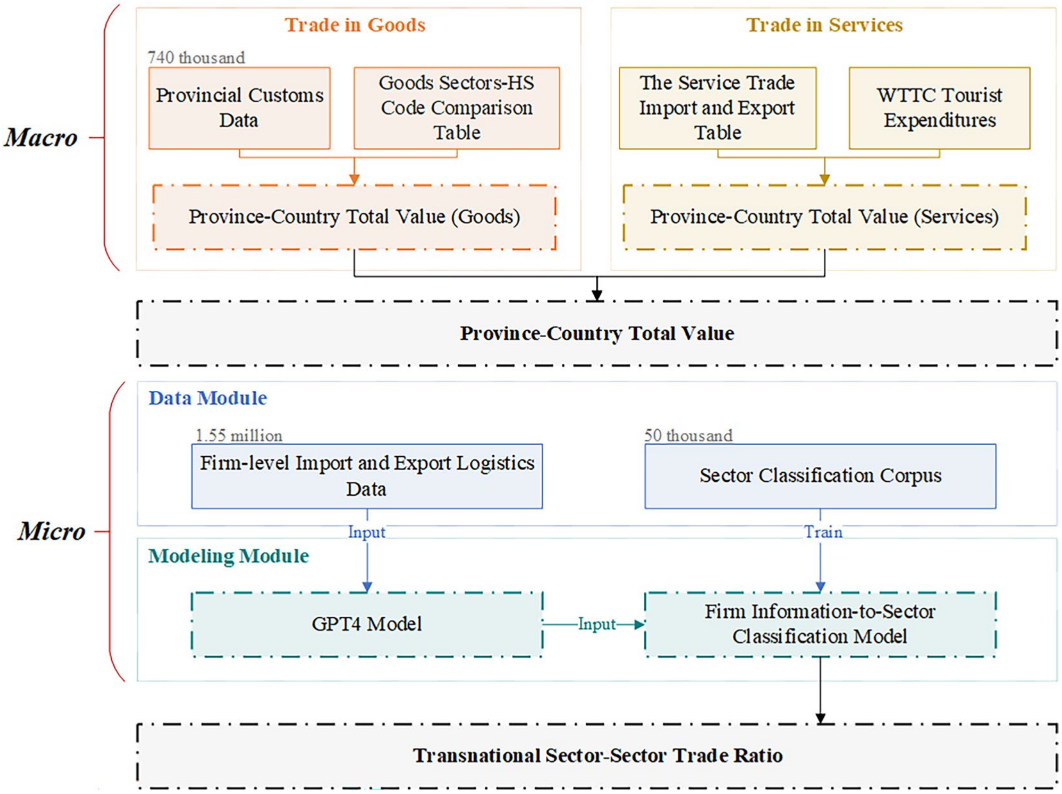

Fig. 2 Overview of the framework for building the SIC.

The STIE account	SIC sector code
Manufacturing services on physical inputs owned by others	S30
Maintenance and repair services. i.e	S30
Transport	S22
Travel	S22, S23, S29, S30
Construction	S20
Insurance and pension services	S25
Financial services	S25
Charges for the use of intellectual property n. i.e	S27
Telecommunications, computer, and information services	S24
Other business services	S30
Personal, cultural, and recreational services	S29
Government goods and services n.i.e	S28

Table 3. Comparison between SIC service sectors and the STIE account.

sectors and transform customs trade data into subnational and international sectoral input-output relationships. This method enables the calculation of the province-country total value (goods).

The primary data source for the service sectors is the China Commerce Yearbook, specifically the provincial STIE (see Table 3 for comparison details). Additional data are collected from provincial bureaus of statistics, provincial commerce bureaus and other government websites to supplement the analysis (see Supplementary Information 3). For tourism-related service trade, we use data from the World Travel & Tourism Council (WTTC) to distribute tourist expenditures across sectors[36] including accommodation and catering, transportation, warehousing and postal services, culture, sports and entertainment, as well as other service sectors. However, STIE provides only aggregate provincial-level data without detailing trade flows between provinces and countries. To address this, we use the sectoral distribution of China's service exports from the OECD_ICIO database. By integrating these data sources, we derive the province-country total value (services).

Data processing from a micro perspective. From a micro perspective, we employ large language models to classify firms by sector and convert firm-level product logistics relationships into inter-industry trade linkages.

Province	Export	Import_US	Import_Others
Beijing	Beijing	Beijing	Anhui
Tianjin	Beijing	Beijing	Anhui
Hebei	Chongqing	Shandong	Anhui
Shanxi	Jilin	Shandong	Anhui
Inner Mongolia	Jilin	Liaoning	Shaanxi
Liaoning	Jilin	Liaoning	Jilin
Jilin	Jilin	Liaoning	Jilin
Heilongjiang	Jilin	Liaoning	Jilin
Shanghai	Shanghai	Zhejiang	Zhejiang
Jiangsu	Shanghai	Zhejiang	Zhejiang
Zhejiang	Shanghai	Zhejiang	Zhejiang
Anhui	Anhui	Shandong	Anhui
Fujian	Beijing	Guangdong	Zhejiang
Jiangxi	Shaanxi	Shandong	Hunan
Shandong	Shanghai	Shandong	Anhui
Henan	Chongqing	Shandong	Anhui
Hubei	Hunan	Shandong	Hunan
Hunan	Hunan	Shandong	Hunan
Guangdong	Guangdong	Guangdong	Zhejiang
Guangxi	Guangxi	Liaoning	Guizhou
Hainan	Guangxi	Guangdong	Zhejiang
Chongqing	Chongqing	Liaoning	Guizhou
Sichuan	Chongqing	Liaoning	Guizhou
Guizhou	Guizhou	Liaoning	Guizhou
Yunnan	Guizhou	Liaoning	Guizhou
Shaanxi	Shaanxi	Liaoning	Shaanxi
Gansu	Gansu	Liaoning	Shaanxi
Qinghai	Gansu	Liaoning	Shaanxi
Ningxia	Gansu	Liaoning	Shaanxi
Xinjiang	Shaanxi	Liaoning	Shaanxi

Table 4. Data imputation rules for missing provincial data.

We use data from the Alayun Platform, which provides province-level firm-level import and export logistics data. These data include key indicators such as "HS code", "purchaser country", "supplier country", "detailed product name", "quantity", "purchaser name", "supplier name", "transaction date", "product description", "purchase amount", and "gross weight". To extract firm-level trade relationships, we focus on the "purchaser name" and "supplier name" indicators. Using the GPT4 API, we retrieve firm business scope, core product information according to the purchaser and supplier name from e-commerce platforms and firm websites, summarizing each firm's primary business activities. Specifically, we use 500,000 trade records of U.S. companies from the ImportYeti Platform in 2017 to assist in identifying their primary business activities. To ensure the accuracy, we conduct a manual verification, which is provided in the Technical Validation section.

We employ LLM techniques to extract and refine sector classification descriptions from China's input-output sector classification explanation and code, constructing a sector classification corpus. We fine-tune the Llama 3 model[37] to train a firm information-to-sector classification model based on the sector classification corpus. This model achieves an accuracy of 0.96. We apply this firm-to-sector model to categorize the extracted firm business information into corresponding sectors. By integrating these sector classifications with purchase amount, country of origin and other trade-related attributes, we derive the transnational sector-sector trade ratio for China's international trade.

Owing to the limited number of trade partner countries available on the Alayun Platform, we classify 66 countries & economies from the OECD_ICIO database into three income groups according to World Bank standards: high income, upper middle income and lower middle income (see Supplementary Information 1 for details). For high-income countries, we use Russia and the US as reference points in our analysis. We incorporate geographical and economic characteristics to analyze the correlation between each country's trade structure and that of the US and Russia. This allows us to assess which country's trade pattern a high-income country more closely follows.

Due to severe data deficiencies for certain provinces on the Alayun Platform, we apply a regional or GDP-based adjacency principle. This method imputes missing values for provinces with substantial missing data, based on the three scenarios, i.e. reference provinces for data-missing provinces in Export, Import_US and Import_Others, outlined in Table 4 and Supplementary Information 8. As a result, we compute 1980

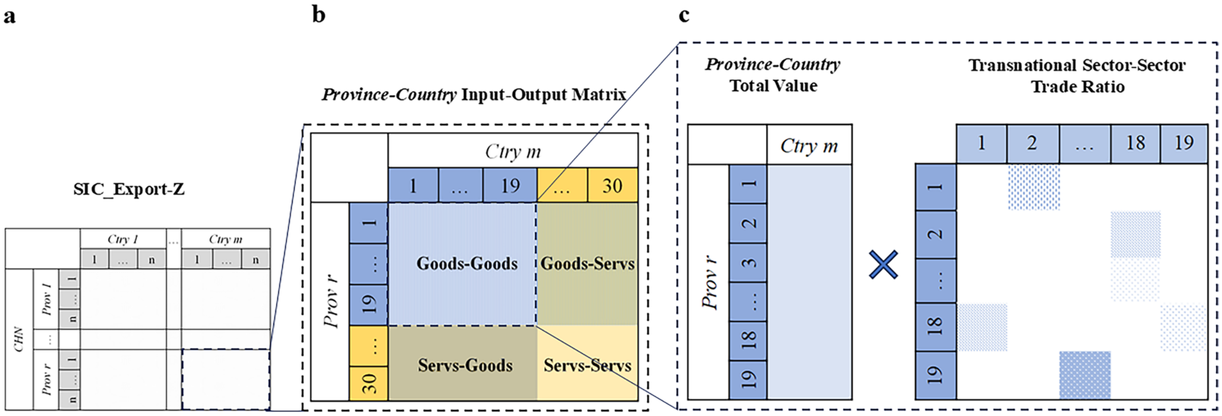

Fig. 3 Calculating methods of SIC_Z.

Data source	China's total import value	China's total export value
China's goods and services trade total value	2308.49	2491.49
SIC	2250.11	2364.72
EXIOBASE	1897.19	2088.82
OECD	2035.53	2214.37

Table 5. Comparison between SIC and other MRIO datasets for China's import & export total value, unit: trillion USD.

transnational sector-sector trade ratios, covering the trade flows between 30 Chinese provinces and 66 countries & economies.

Final demand section. Based on the aforementioned micro- and macro-level data, we proceed with the table compilation.

In the MRIO, the total output of an economy can be calculated as:

$$X = AX + FD \tag{1}$$

where X is the total output, A is the direct requirements matrix, AX or Z is the intermediate matrix, and FD is the final demand. To compile the SIC_Z and SIC_FD, we calculate the proportion of FD in goods sectors using the EST_MRIO, and the proportion of FD in service sectors using the OECD_ICIO database. Since the FD proportion of imports in China is not directly available in the OECD_ICIO database, we apply the FD proportion of imports from China's non-competitive input-output table instead[38].

Intermediate matrix section. Given the quantitative relationship between Z and FD components established in Eq. (1), the residual term (X – FD) mathematically corresponds to Z (Fig. 3). We take a province-country input-output matrix from SIC_ZExport as an example, focusing on export flows from province r to country m (Fig. 3a). This matrix contains 30 economic sectors, categorized into 19 goods and 11 services sectors. The interaction of these categories generates four distinct submatrices: goods-goods, goods-services, services-goods and services-services (as shown in Fig. 3b). First, the detailed calculation procedure for goods-goods is illustrated in Fig. 3c. It is calculated as the product of the Province-country total value (from Data processing from a macro perspective section) and the Transnational sector-sector trade ratio (from Data processing from a micro perspective section). The calculations for goods-services, services-goods and services-services are computed using a similar method, except that they rely on the Transnational sector-sector trade ratio from the OECD_ICIO database for China's trade with each country.

Value added section. In the MRIO, the total input of an economy can also be calculated as:

$$X = Z + VA \tag{2}$$

where X represents the total input, Z denotes the Intermediate matrix, and VA refers to the value added. The VA values of each province in China are sourced from our EST_MRIO table.

Table balancing. The table balancing approaches include RAS, generalised RAS (GRAS), entropy, constrained optimization, linear programming approaches, Monte Carlo, hybrid algorithms, etc., and they have different scopes of applicability and advantages. With imputed microdata, the chosen GRAS method can achieve

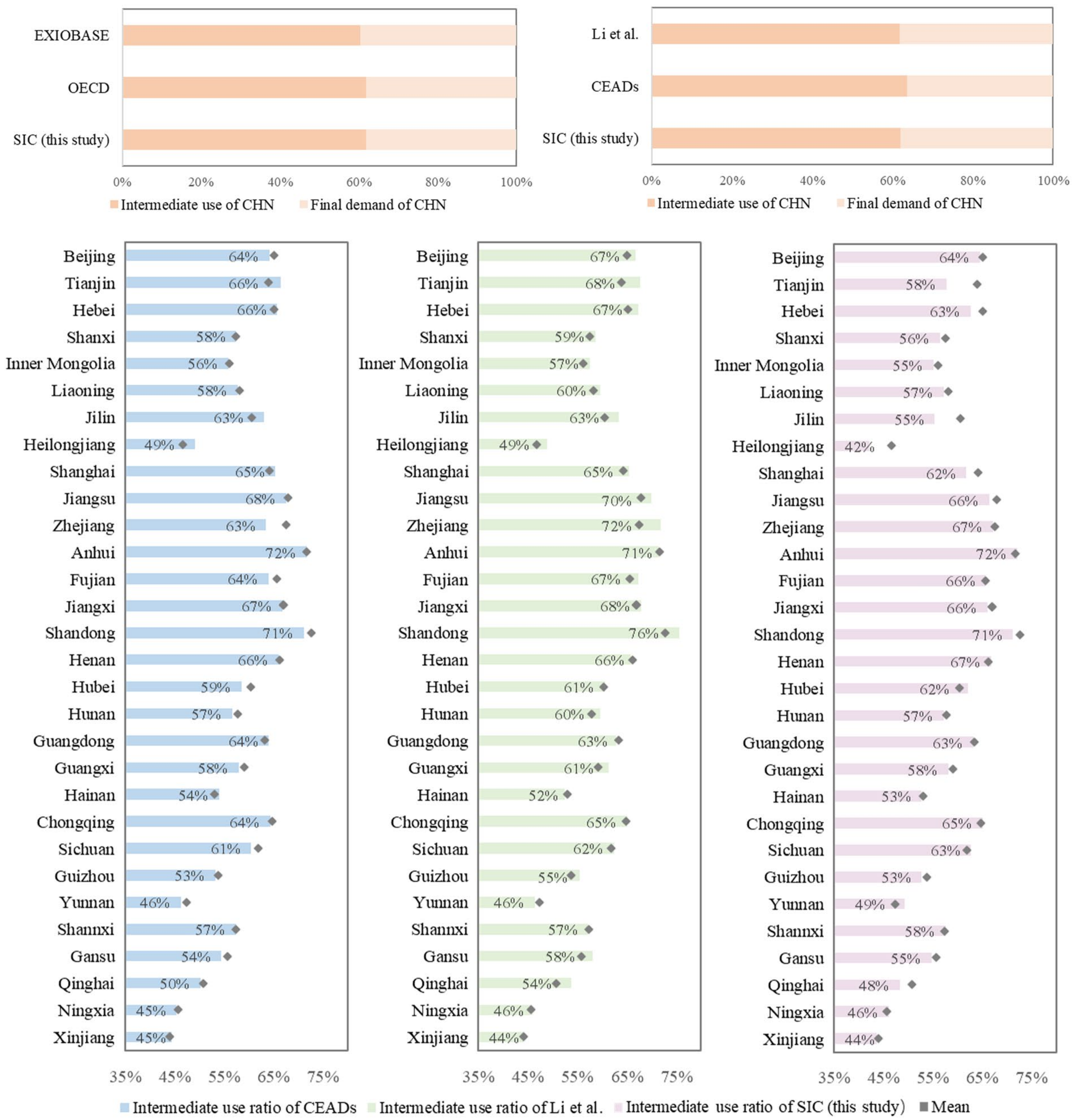

Fig. 4 Comparisons between SIC and other MRIO tables for the structure of intermediate use and final use. (**a**) Compares the structure of intermediate use and final demand in CHN; (**b**) Compares the provincial structure of intermediate use and final demand in CHN.

convergence at the 89th iteration. The GRAS method is an extension of the RAS approach, which aims to preserve the original matrix structure as much as possible while making minimal adjustments to align with new row and column totals, accommodating both positive and negative values in the matrix. Using the GRAS method, we ensure the input-output table satisfies the row-column relationships specified in Eqs. (1, 2).

$$\begin{cases} \sum_j r_i Z_{ij} s_j = u_i(\forall i) \\ \sum_i r_i Z_{ij} s_j = v_j(\forall j) \end{cases} \tag{3}$$

$$\min_{R,S} \sum_{i,j} |Z'_{ij} - Z_{ij}| \tag{4}$$

Source	Accuracy	Semantic precision
By human	98%	100%
By GPT4	98%	90%

Table 6. Summary of verification results.

Here, r_i denotes the row scaling multiplier for the i-th row, which may assume either positive or negative values; s_j represents the column scaling multiplier for the j-th column, also permitting both positive and negative values. The original intermediate matrix is denoted as Z_{ij}, while $Z_{ij}^{'}$ refers to the adjusted intermediate matrix. Furthermore, u_i and v_j correspond to the row and column totals of the adjusted matrix for the i-th row and j-th column, respectively.

Data Records

The SIC dataset is available at the figshare website[39]. We provide 6 processing folders containing source data and code: 1. Custom data processing, 2. Service trade data processing, 3. Micro data processing, 4. Table compiling_Z, 5. Table compiling_FD, 6. Table balancing and one final table of results: "SIC-MRIO_2017". This table contains an intermediate matrix (2820*2820), a final demand vector (2820*282) and a value added vector (2820*1). The SIC dataset can be freely downloaded from https://doi.org/10.6084/m9.figshare.28973396.v3.

Technical Validation

Comparison between SIC and other MRIO datasets for China's import & export total value. China's import and export statistics show major discrepancies across key datasets, including those from the OECD and CEADs. These differences arise mainly from limited data availability and inconsistent source data. As evidenced in Table 5, our estimates show minimal deviations of merely 2.53% for imports and 5.09% for exports when benchmarked against official China Customs statistics. This represents a marked improvement in accuracy relative to alternative datasets, with EXIOBASE exhibiting deviations of 17.82% (imports) and 16.16% (exports), and the OECD showing variances of 11.82% (imports) and 11.12% (exports).

Comparison between SIC and other MRIO datasets for the structure of intermediate use and final use. In MRIO tables, the structure of intermediate use and final demand serves as another crucial indicator. Comparison between SIC and other major domestic and international MRIO databases—including OECD, EXIOBASE, CEADs and Li *et al.*—reveals generally consistent structural patterns. Among these, CEADs shows the highest proportion of intermediate use at 63.54%, while EXIOBASE records the lowest at 60.47%. When compared specifically with Chinese regional input-output tables, more pronounced discrepancies emerge in the provinces of Tianjin, Jilin and Heilongjiang, shown in Fig. 4. We conduct a detailed verification of the ratio of the structure of intermediate use and final use in Supplementary Information 5.

Manual annotations and checks. We undertake a systematic process of manual annotation and verification of the data from a micro perspective. First, the firm-level logistics data are cross-validated with China Customs data to ensure consistency and reliability, shown in Supplementary Information 6. Second, we assess the accuracy of company-level business information collected through the GPT4 API by conducting a secondary manual verification, with results presented in Table 6. While both GPT4 and manual methods maintain high precision, GPT4 shows slightly lower semantic accuracy. Specifically, GPT4 occasionally generates vague descriptors like "Industry & Manufacturing", "Industrial" and "Technology development" for certain companies' primary businesses - terms too ambiguous for proper industry classification. These non-specific terms (detailed in Supplementary Information 7) are subsequently excluded from our analysis.

Data availability

The SIC dataset is available in the figshare repository[39] and is publicly accessible at https://doi.org/10.6084/m9.figshare.28973396. The shared output includes a final table of results titled "SIC-MRIO_2017", and 6 processing folders containing the source data and the Python code.

Code availability

The program used in this study is based on Python. Code is freely available on the figshare website.

Received: 16 May 2025; Accepted: 25 September 2025;
Published online: 07 November 2025

References

1. Zhang, Y., Tian, K., Li, X., Jiang, X. & Yang, C. From globalization to regionalization? Assessing its potential environmental and economic effects. *Applied Energy* **310**, 118642 (2022).
2. Zhao, R. *et al.* Cascade effects of global supply chains induced by regional conflicts increase global inequality. *Resources, Conservation and Recycling* **209**, 107740 (2024).
3. Haas, W., Krausmann, F., Wiedenhofer, D. & Heinz, M. How circular is the global economy?: An assessment of material flows, waste production, and recycling in the European Union and the world in 2005. *Journal of Industrial Ecology* **19**, 765–777 (2015).
4. Meng, B., Zhang, Y. & Inomata, S. Compilation and applications of IDE-JETRO's international input–output tables. *Economic Systems Research* **25**, 122–142 (2013).
5. Wang, Y., Geschke, A. & Lenzen, M. in *Proceedings of the 22nd International Input-Output Conference, Lisbon, Portugal*. 14–18.

6. Zheng, J. *et al*. The slowdown in China's carbon emissions growth in the new phase of economic development. *One Earth* **1**, 240–253 (2019).
7. Feng, K., Davis, S. J., Sun, L. & Hubacek, K. Drivers of the US CO_2 emissions 1997–2013. *Nature Communications* **6**, 7714 (2015).
8. Hamilton, H. A. *et al*. Trade and the role of non-food commodities for global eutrophication. *Nature Sustainability* **1**, 314–321 (2018).
9. Shan, Y. *et al*. Impacts of COVID-19 and fiscal stimuli on global emissions and the Paris Agreement. *Nature Climate Change* **11**, 200–206 (2021).
10. Han, X. *et al*. Comparative life cycle greenhouse gas emissions assessment of battery energy storage technologies for grid applications. *Journal of Cleaner Production* **392**, 136251 (2023).
11. Huo, J., Meng, J., Zheng, H., Parikh, P. & Guan, D. Achieving decent living standards in emerging economies challenges national mitigation goals for CO_2 emissions. *Nature Communications* **14**, 6342 (2023).
12. Yang, Y. *et al*. Mapping global carbon footprint in China. *Nature Communications* **11**, 2237, https://doi.org/10.1038/s41467-020-15883-9 (2020).
13. Bruckner, B. *et al*. Ecologically unequal exchanges driven by EU consumption. *Nature Sustainability* **6**, 587–598, https://doi.org/10.1038/s41893-022-01055-8 (2023).
14. Long, Y. *et al*. Comparison of city-level carbon footprint evaluation by applying single-and multi-regional input-output tables. *Journal of Environmental Management* **260**, 110108, https://doi.org/10.1016/j.jenvman.2020.110108 (2020).
15. Wang, S. & Chen, B. Energy–water nexus of urban agglomeration based on multiregional input–output tables and ecological network analysis: A case study of the Beijing–Tianjin–Hebei region. *Applied Energy* **178**, 773–783, https://doi.org/10.1016/j.apenergy.2016.06.112 (2016).
16. Wang, Y. *et al*. An inter-city input-output database distinguishing firm ownership in the Greater China area during 2002–2017. *Scientific Data* **12**, 723, https://doi.org/10.1038/s41597-025-04996-9 (2025).
17. Meng, B., Wang, Z. & Koopman, R. How are global value chains fragmented and extended in China's domestic production networks? *IDE Discussion Paper* **424** (2013).
18. Meng, B. & Yamano, N. Compilation of a regionally extended inter-country input–output table and its application to global value chain analyses. *Journal of Economic Structures* **6**, 23 (2017).
19. Zheng, J., Hou, X., Duan, H. & Wang, S. Where we are and where we will go to achieve carbon neutrality: A review of quantitative socioeconomic modeling research. *Fundamental Research* **4**, 1696–1709 (2024).
20. Zheng, J. *et al*. The energy system transition pathway towards carbon reduction using a model-coupling approach. *Global Environmental Change* **89**, 102945 (2024).
21. Mi, Z. F. *et al*. Chinese CO_2 emission flows have reversed since the global financial crisis. *Nature Communications* **8** (2017).
22. Ou, J. *et al*. Role of export industries on ozone pollution and its precursors in China. *Nature Communications* **11**, 5492 (2020).
23. Su, B. & Ang, B. W. Input-output analysis of CO_2 emissions embodied in trade: A multi-region model for China. *Applied Energy* **114**, 377–384 (2014).
24. Zhang, Z. K. *et al*. Embodied carbon emissions in the supply chains of multinational enterprises. *Nature Climate Change* **10**, 1096–U1032 (2020).
25. Kee, H. L. & Tang, H. W. Domestic Value Added in Exports: Theory and Firm Evidence from China. *American Economic Review* **106**, 1402–1436 (2016).
26. Ye, Q. *et al*. Allocating capital-associated CO_2 emissions along the full lifespan of capital investments helps diffuse emission responsibility. *Nature communications* **14**, 2727 (2023).
27. Chen, Q. *et al*. An interprovincial input–output database distinguishing firm ownership in China from 1997 to 2017. *Scientific Data* **10**, 293, https://doi.org/10.1038/s41597-023-02183-2 (2023).
28. Zhang, G. *et al*. China's environmental policy intensity for 1978–2019. *Scientific Data* **9**, 75, https://doi.org/10.1038/s41597-022-01286-6 (2022).
29. Huang, C. *et al*. ORLM: A Customizable Framework in Training Large Models for Automated Optimization Modeling. *Operations Research* https://doi.org/10.1287/opre.2024.1233 (2025).
30. Fukui, S. Estimating Input Coefficients for Regional Input–Output Tables Using Deep Learning with Mixup. *Computational Economics* **65**, 2423–2448 (2025).
31. Serpell, M. C. Incorporating data quality improvement into supply–use table balancing. *Economic Systems Research* **30**, 271–288 (2018).
32. Dalgaard, E. & Gysting, C. An algorithm for balancing commodity-flow systems. *Economic Systems Research* **16**, 169–190 (2004).
33. Zhao, B., Shuai, C., Qu, S. & Xu, M. Using Deep Learning to Fill Data Gaps in Environmental Footprint Accounting. *Environmental Science & Technology* **56**, 11897–11906 (2022).
34. General Administration of Customs of the People's Republic of China. http://stats.customs.gov.cn/indexEn (2017).
35. Sheng, B. The Political Economy of Trade Protection in Chinese Industries. *China Economic Quarterly* **1** (2002).
36. WTTC. https://researchhub.wttc.org/product/global-retail-tourism-trends-insights-full-report (2023).
37. Grattafiori, A. *et al*. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783* (2024).
38. State Statistics Bureau. https://data.stats.gov.cn/files/html/quickSearch/trcc/trcc0801.html (2025).
39. Zheng, J. & Mi, Z. A multi-regional input-output database linking Chinese subnational regions and global economies. *figshare. Dataset*. https://doi.org/10.6084/m9.figshare.28973396.v3 (2025).
40. Aguiar, A., Chepeliev, M., Corong, E. & van der Mensbrugghe, D. The GTAP Data Base: Version 11. *Journal of Global Economic Analysis* **7**(2), 1–37, https://doi.org/10.21642/JGEA.070201AF (2022).
41. Yamano, N. & Webb, C. Future Development of the Inter-Country Input-Output (ICIO) Database for Global Value Chain (GVC) and Environmental Analyses. *Journal of Industrial Ecology* **22**, 487–488, https://doi.org/10.1111/jiec.12758 (2018).
42. Merciai, S. & Schmidt, J. Methodology for the Construction of Global Multi-Regional Hybrid Supply and Use Tables for the EXIOBASE v3 Database. *Journal of Industrial Ecology* **22**, https://doi.org/10.1111/jiec.12713 (2018).
43. Huo, J. *et al*. Full-scale, near real-time multi-regional input–output table for the global emerging economies (EMERGING). *Journal of Industrial Ecology* **26**, 1218–1232, https://doi.org/10.1111/jiec.13264 (2022).
44. Li, S. T. *et al*. *Multi-regional input-output model in China :1987-2017*. (Economic Science Press, 2023).
45. Mi, Z. F. *et al*. A multi-regional input-output table mapping China's economic outputs and interdependencies in 2012. *Scientific Data* **5**, https://doi.org/10.1038/sdata.2018.155 (2018).
46. Zheng, H. R. *et al*. Chinese provincial multi-regional input-output database for 2012, 2015, and 2017. *Scientific Data* **8**, https://doi.org/10.1038/s41597-021-01023-5 (2021).
47. He, K., Mi, Z., Zhang, J., Li, J. & Coffman, D. M. The polarizing trend of regional CO_2 emissions in China and its implications. *Environmental Science & Technology* **57**, 4406–4414, https://doi.org/10.1021/acs.est.2c08052 (2023).
48. State Administration of Foreign Exchange of China. https://www.safe.gov.cn/en/2019/0329/1496.html (2025).
49. Alayun global trade logistics technology service platform. https://pc.alayunchina.com (2025).
50. ImportYeti Platform. https://www.importyeti.com/ (2025).

Acknowledgements
This study was supported by the National Natural Science Foundation of China (72104197; 72474179; 71988101; 72274188), the Young Elite Scientists Sponsorship Program by CAST (2022QNRC001), and the Royal Society Newton International Fellowship (NIF\R1\192828). We hereby extend our sincere appreciation to the Joint Laboratory of Global Trade and Supply Chain Big Data, Xi'an Jiaotong University, for providing us access to the Alayun Platform.

Author contributions
J.Z. designed the study. X.H. compiled the SIC table. J.Z., X.H. prepared the manuscript. J.Z., X.H. and M.H. collected the data. All authors (X.H., J.Z., M.H., G.F., K.H., T.M., D.C., Z.M. and S.W.) contributed to writing the manuscript. S.W., T.M., and Z.M. coordinated and supervised the project.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-06040-2.

Correspondence and requests for materials should be addressed to J.Z., T.M. or Z.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.