www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A high-resolution multi-scale industrial water use dataset in China

Meng Li[1], Yuan Tong[2]✉, Junming Zhu[2]✉ & Shuntian Xu[2]

Water is crucial for achieving the UN Sustainable Development Goals, particularly SDG 6. As a major source of water use and pollution, industrial sector requires improved water management based on more systematic and refined analysis. Such analysis, however, is compromised by the accuracy, granularity, and coverage of industrial water data. Here we present an open dataset of China's industrial water use, compiled from 1,480,265 plant-level reports. This high-resolution multi-scale dataset offers unparalleled details, supporting multi-scale analysis at the province, city, and county levels, and across 2-digit, 3-digit, and 4-digit industrial classifications. It provides comprehensive information on water use, recycling, pollution, and wastewater processing. Such data enables further macro- and micro-level analysis, including multi-regional input-output analysis, structural decomposition analysis, statistical analysis, machine learning, as well as many other advanced analytical methods. This dataset can equip researchers and policymakers with a valuable tool to advance sustainable water management, fostering alignment with global sustainability goals.

Background & Summary

Water, as an indispensable resource for human being and the planet, is facing critical challenges from both anthropogenic and natural factors[1], leading to increased scarcity[2], more frequent droughts[3], and disruptions to hydrological cycles[4,5]. Water plays a pivotal role in the achievement of the United Nations Sustainable Development Goals (SDGs)[6]. Particularly, SDG 6 calls for ensuring availability and sustainable management of water and sanitation for all. Effective management of water resources stands as a cornerstone for the nexus between ecosystems, economies, and social communities[7]. Industrial water management not only constitutes a considerable share of total water usage, but is also closely related to problems like water pollution, resource depletion, and ecosystem disruptions[6,8,9]. Sustainable industrial water management requires accurate, comprehensive, and detailed data to evaluate the performance of water usage[10]. Although the number of water-related datasets is increasing, extensive research and practical needs require further efforts in dataset construction, especially with improved granularity, higher coverage, connections across multiple scales, and index on resource management and pollution control[11,12].

To contribute to this mission, we focus on industrial water use in China, a large developing country grappling with complex water resource management challenges[13,14], for both substantive and illustrative purposes[15]. China ranks the sixth in global freshwater resources, but its per capita water resources is only one-fourth of the global average[16]. The complex water-related issues in China, such as scarcity[17], recycling[18], and pollution[19,20], needs more evidence-based solution and mirrors the issues of other developing economies. The industrial sector accounts for 16.2% of China's total water usage and is a major cause of water pollution, with both domestic and worldwide impacts via embodied water. Revealing the industrial water cycle is thus of great importance to not only understanding of China's progress in SDG 6 but also formulating sustainable policies and practices in developing countries[21]. In addition, plant-level data in China has been well collected and applied in research of carbon emissions and air pollutions[22,23], but not on water-related issues. With the increasing need of transparency and accuracy in scientific data, building a high-resolution, plant-level, cross-scale industrial water usage dataset in China can provide a replicable framework for other countries.

[1]School of Environmental Science and Engineering, Shanghai Jiaotong University, Shanghai, China. [2]School of Public Policy and Management, Tsinghua University, Tsinghua, China. ✉e-mail: tongy22@mails.tsinghua.edu.cn; junming@tsinghua.edu.cn

Despite great efforts and progress, existing datasets cannot adequately capture micro-level water usage, recycling, and pollution across Chinese regions and sectors, due to several limitations[23,24]. To start, annual industrial water use is usually publicized at the provincial level, but it is not directly measured. Rather, it is calculated based on sector output and sector-specific water-intensity coefficient, i.e., water use per economic output, from a national standard–the General Principles for Industrial Water Use Quotas. Similar practices have been applied by most research to estimate, rather than measure, water usage. As the sector level water usage is not reported, most studies have to split the provincial level water usage to sectors based on the benchmark year, usually year 2008, where sectoral water usage data is available from the Second National Economic Census. While this method brings variation in water use given output changes, the unchanged water usage structure across sectors can lead to limited accuracy and applicability for more recent analysis and in dynamic settings.

Due to the data generation process (DGP) based on estimation, the available datasets cannot provide micro-level information finer than a city level, making them not applicable for more detailed or advanced analysis. Additionally, the current datasets suffer from the lack of critical water-related information[25], especially sustainable production practices like water recycling. Such limitations impede a thorough understanding of water management and constrain the development of effective strategies for industrial sustainability.

We fill the data gap by presenting an open dataset of industrial water use in China, compiled from plant-level water information reported individually across the country. This granular data from a unique DGP of micro-level plant reports differs from the macro-level estimation based on fixed sector-specific water intensity in all other research and even official statistics. Our dataset provides much-needed, evidence-based foundation for analyzing water scarcity, recycling practices, and associated risks. It encompasses 1,480,265 plant-by-year observations, offering unprecedented details and insights into water use patterns. This granular data is crucial for a more precise understanding of industrial water management, especially in regions and sectors where water scarcity is critical.

Another key novelty of our dataset is the integration of data across multiple spatial and sectoral levels. We carefully aggregated the water usage data to different spatial levels and to sectors. In particular, the high-resolution dataset can be processed to obtain water usage data at province-, city-, county-levels and for 4-digit, 3-digit, and 2-digit industrial classifications. This multi-scale and multi-sector approach bridges the gap between micro-level data and broader macro-level trends, offering a comprehensive tool for analyzing water management at various scales.

Further, this newly built dataset provides a robust foundation for analysis of water scarcity, recycling, and risks. Subsequent analysis of this data can be conducted using a variety of methods to address water usage issues. At the macro level, multi-regional input-output (MRIO) models can be applied, for example, to uncover the virtual water footprints and inter-regional water linkages; structural decomposition analysis and index decomposition analysis can be used to examine the factors driving changes in water use, including both quantity and structural dimensions. At the micro level, statistical analysis and machine learning techniques can be employed to evaluate shifts in industrial water efficiency, water recycling, and sustainable performance; quasi-experimental design, such as difference-in-differences, matching, and regression discontinuity design, can be performed to reveal causal effects of policy and managerial factors on water management.

By making this dataset openly available, we aim to empower researchers, policymakers, and industry practitioners with the information necessary for sustainable water management, aligning with both national priorities and global sustainability goals. Our open dataset on Chinese industrial water extraction, usage, and pollution covers up to 364,535 million tonnes of total water usage, 20,062 million tonnes of wastewater emission, and 518,284 tonnes of COD emissions per year.

Methods

Our research involved using the Enterprise Green Development Database (EGDD) through the EPS data platform's China Microeconomic Data Query System (http://microdata.sozdata.com/#/home) for plant-level details on water usage and pollution. The EGDD5 is also known as China's Industrial Environmental Statistics Database. This extensive database is derived from the China Environment Statistical Database, which is compiled by China's Ministry of Ecology and Environment. It spans a broad spectrum of industrial plants distributed across 31 provinces, autonomous regions, and municipalities–Beijing, Shanghai, Tianjin, and Chongqing. Combined with the Baidu Map's geocoding API, the EGDD offers thorough data on geographical locations and industry classifications, alongside a wide range of data on emissions, pollution, and environmental management techniques from 1998 to 2014. Following an initial phase of data processing, we created a plant-level industrial water usage dataset across China, with a total of 1,480,265 observations, and based on it, we further combined the Multi-Regional Input-Output framework to develop the WE-MRIO satellite data for diversified use. The entire dataset construction framework is illustrated in Fig. 1, offering a visual overview of our methodology (see Appendix Table A1 and Table A2 for the detailed description of the dataset).

Data processing.

1. Concordance of Industry Classification. There are three versions of National Economic Industry Classification, which are the 2002 (GB/T 4754-2002), 2011 (GB/T 4754-2011), and 2017 version (GB/T 4754-2017). To enable a detailed analysis at the 4-digit sub-sector level, we updated the sector codes from the two previous versions in the dataset to align with the newest 2017 version of classification standard.
2. Alignment of MRIO Sector Codes. The industry classifications in our dataset differ from those used in the Multi-Regional Input-Output table, which adheres to the industrial classification system for national economic activities. Therefore, we mapped the industry classifications in our dataset with the sector codes specified in the MRIO table, ensuring coherence and comparability (See Appendix, Fig. A1 for details).

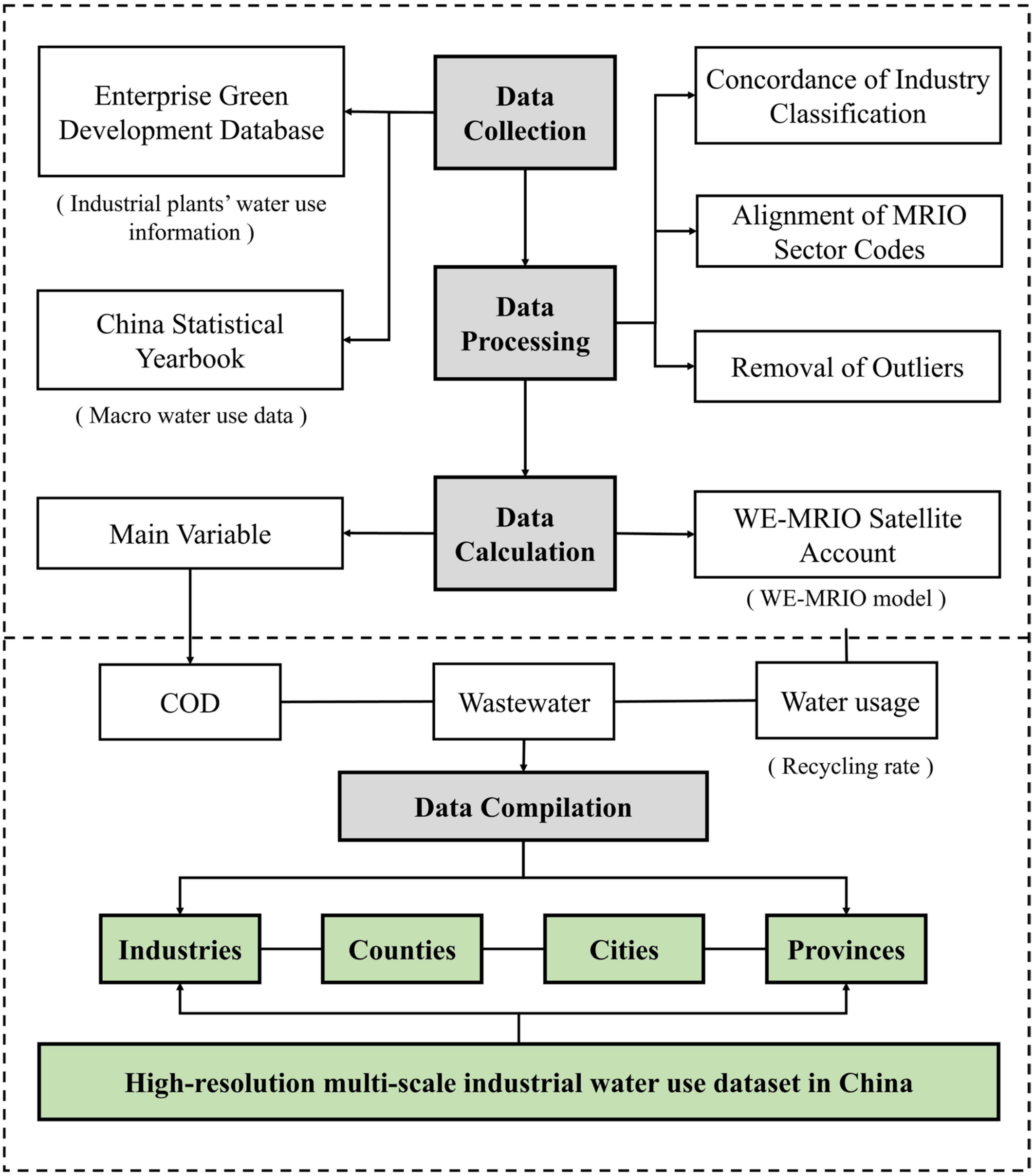

Fig. 1 Dataset construction framework.

Observations whose sector code could not be identified (classified as "other industries" in the EGDD, which are a miscellaneous collection of uncategorized industries) were removed.

3. Removal of Outliers. The data reported by plants may include inaccuracies and anomalies. To reduce the impact of such uncertainties, we removed outliers from our dataset according to each plant's annual statistics. In specific, we first removed observations with total output less than or equal to zero and water use less than zero, as such values were not feasible. In doing so, wastewater and COD data with negative values were also simultaneously removed, as they coincidentally appeared in the same records. After this initial cleaning step, we analyzed the distribution of both water usage intensity (defined as the amount of water use per unit of output) and total water usage, and identified the 99th percentile for each year. Plants surpassing the 99th percentile threshold for either metric in each year were first screened[23]. Then we manually checked their water usage and total output. We compared these values with reports of the same plant in other years, reports by other plants of similar output in the same sector, and the average water intensity of the same sector as references. Those observations abnormally different from the references were excluded. After removing outliers based on water usage and output, we also checked the corresponding wastewater and COD data to ensure consistency.

Major variable generation. In this step, we generate the major variables based on the processed dataset. We focused on five sets of key variables: (a) water usage, (b) water recycling, (c) wastewater, (d) chemical oxygen

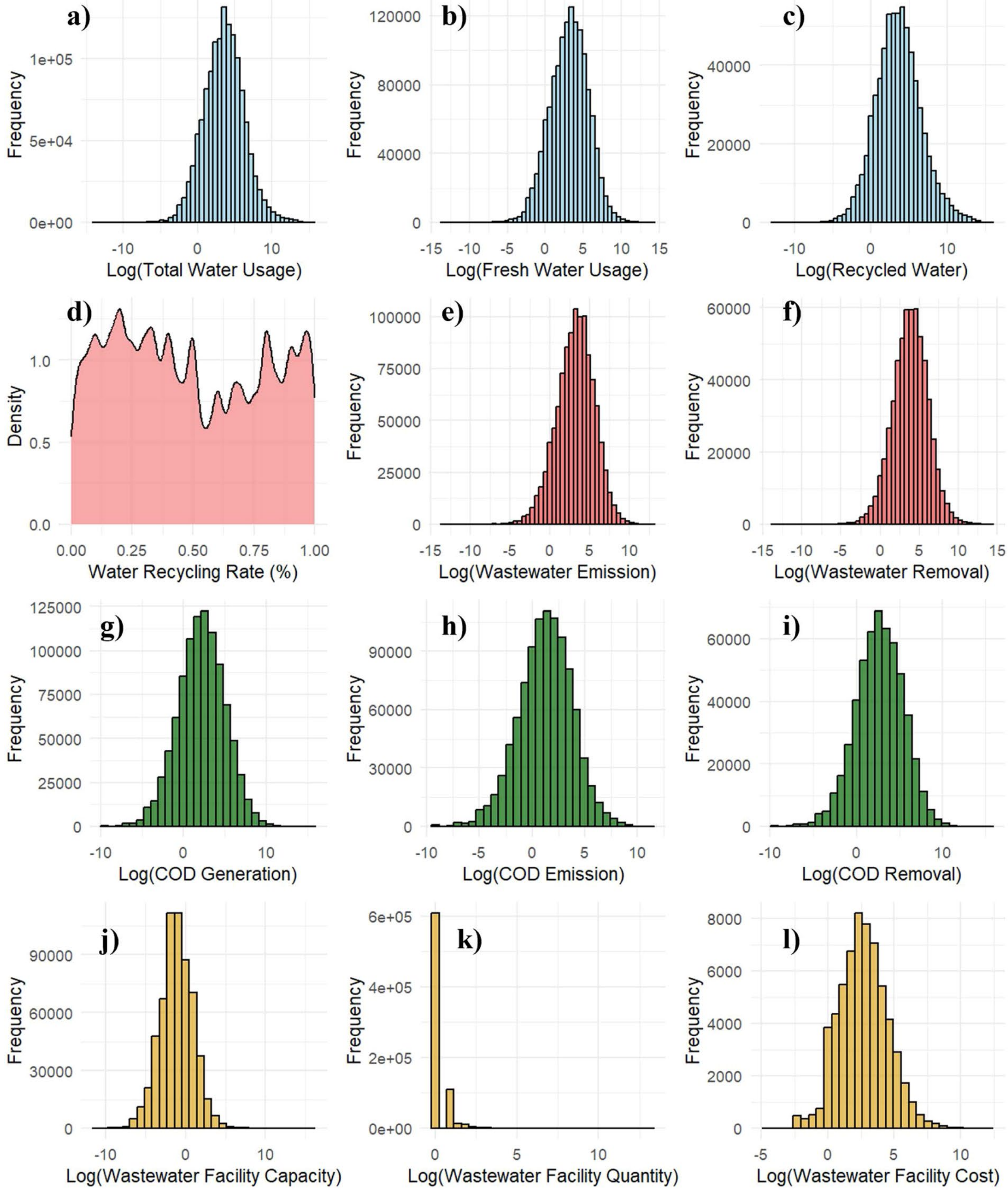

Fig. 2 Distribution of key variables. Panels (**a**–**c**) represent water usage in logarithmic form, and panel (**d**) further provides the water recycling rate. Panels (**e,f**) indicate wastewater, panels (**g**–**i**) indicate COD, whereas panels (**j**–**l**) indicate wastewater processing facility, all in logarithmic form.

demand (COD), and (e) wastewater processing facilities (Fig. 2). We generate water-related variables based on a "standardization-consistency check-correction" process. Where necessary, data discrepancies were corrected using cross-referenced industry benchmarks or plant-specific historical data.

Water usage. Water usage data were reported in kilotonnes, including total water usage, freshwater usage, and recycled water volume. All values were standardized to a uniform format, enabling accurate comparisons and calculations across the dataset. Further, we ensured that the total water usage was equal to the sum of freshwater usage and recycled water.

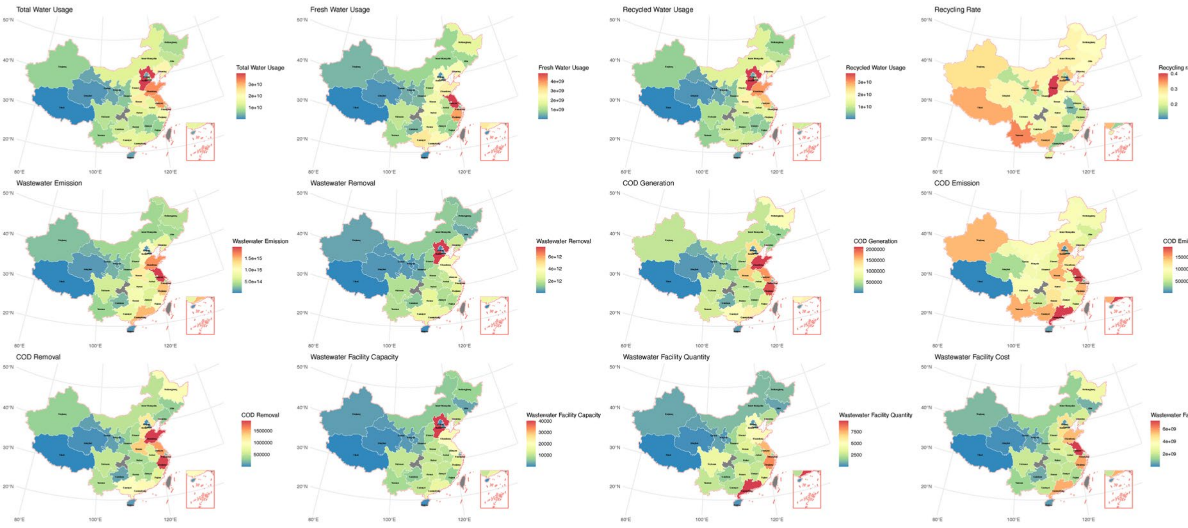

Fig. 3 Distribution of variables by province in the dataset, 2013.

Water recycling. The recycling rate of water was calculated as the ratio of recycled water to total water usage. This metric provided a clear indication of the efficiency of water reuse processes within each facility, allowing for the assessment of resource conservation efforts.

Wastewater. Wastewater data, also reported in kilotonnes, included both emission volume and removal volume. These quantities were standardized to align with water usage data, ensuring consistent measurement across all liquid outputs from the facilities.

COD. COD, a critical indicator of water pollution, was reported in tonnes, including COD generation, COD emission, and COD removal. We verified that the sum of COD removal and COD emission matched the total COD generation. Any inconsistencies were addressed by cross-referencing data from the same industry within the region or by examining historical data from the same plant.

Wastewater processing facilities. Data on wastewater processing facilities were also included, with key variables such as the number of facilities (set), their capacity (kilotonnes per day), and the associated costs (ten thousand yuan). This information allowed for a comprehensive evaluation of the capacity and effectiveness of industrial wastewater treatment processes.

WE-MRIO Satellite Account. To further support researchers with diversified data for studies such as input-output (IO) analysis and structural decomposition analysis (SDA)[12,26], we have expanded the application of the MRIO framework based on our water usage dataset. Specifically, we constructed a WE-MRIO (Water-Environment Multi-Regional Input-Output) model, and provided satellite data at the province, city, and county levels for further use (see Appendix for the model construction). It is important to note that we used the MRIO tables, available for the years 1997, 2002, 2007, and 2012, to align the data processing with our dataset period from 1998 to 2014. It ensures that the MRIO framework is closely aligned with the actual economic conditions of each respective year[27]. Besides, the WE-MRIO model proposed in this paper focuses solely on the industrial sector, with all other sectors set to zero. Researchers in relevant fields can adjust the spatial and sectoral resolution as needed and use this model as a foundation for their analyses. The method for using the WE-MRIO Satellite Account and the specific data settings for various sectors, is detailed in Appendix Tables A3 and A4.

Dataset compilation. After the data processing and calculation steps, we compiled a plant-level industrial water usage dataset in China. This compilation encompasses data on volumes of industrial water usage, industrial wastewater emission, and industrial emissions of chemical oxygen demand (COD). This dataset was further aggregated at the province, city, and county levels to provide a multi-scale view of water usage across different regions using 6-digit Administrative Division Codes. Meanwhile, we categorized the data based on 2-digit, 3-digit, and 4-digit sub-sector codes to facilitate more granular analysis. In addition, we also provided water usage satellite accounts at the province, city, and county levels for various purpose.

Figures 3–5 illustrate the landscape of key variables across provinces, cities, and counties in 2013, while Fig. 6 shows the corresponding distribution of plants at various spatial scales. Figure 7 provides a detailed view of water usage by industrial sectors in the top five provinces by water consumption in 2014. Additional visualizations of key variables for other years are presented in the Appendix: Fig. A2-A17 (province level), Fig. A18-A33 (city level), and Fig. A34-A49 (county level). The distribution of plants across different spatial scales is shown in Fig. A50-A65. Finally, Fig. A66-A81 offer a comprehensive explanation of the industry classification methodology and the results for other years. Fig. A82-98 present the flow of water between regions.

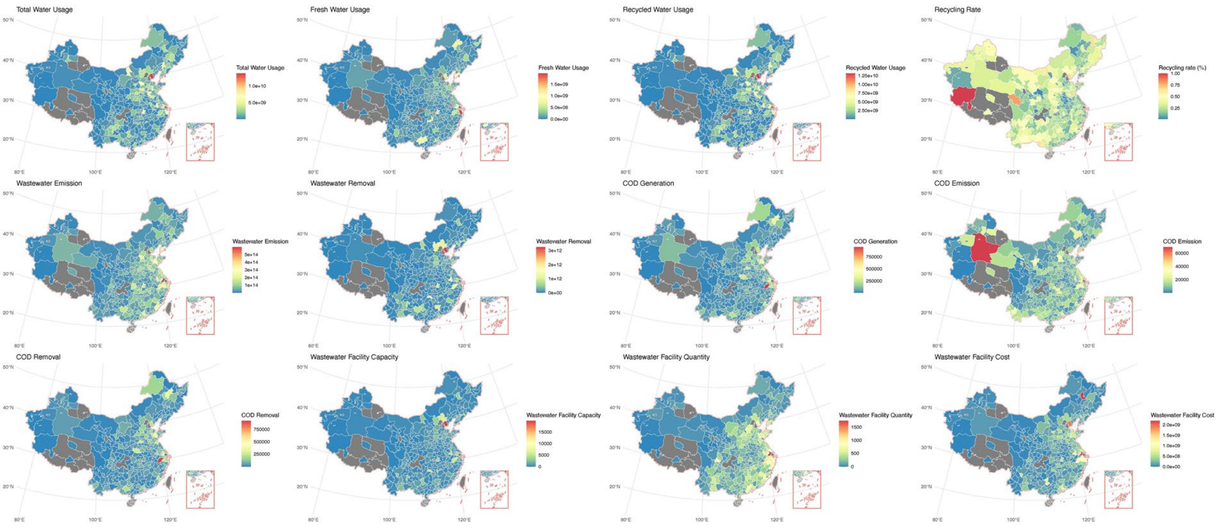

Fig. 4 Distribution of variables by city in the dataset, 2013.

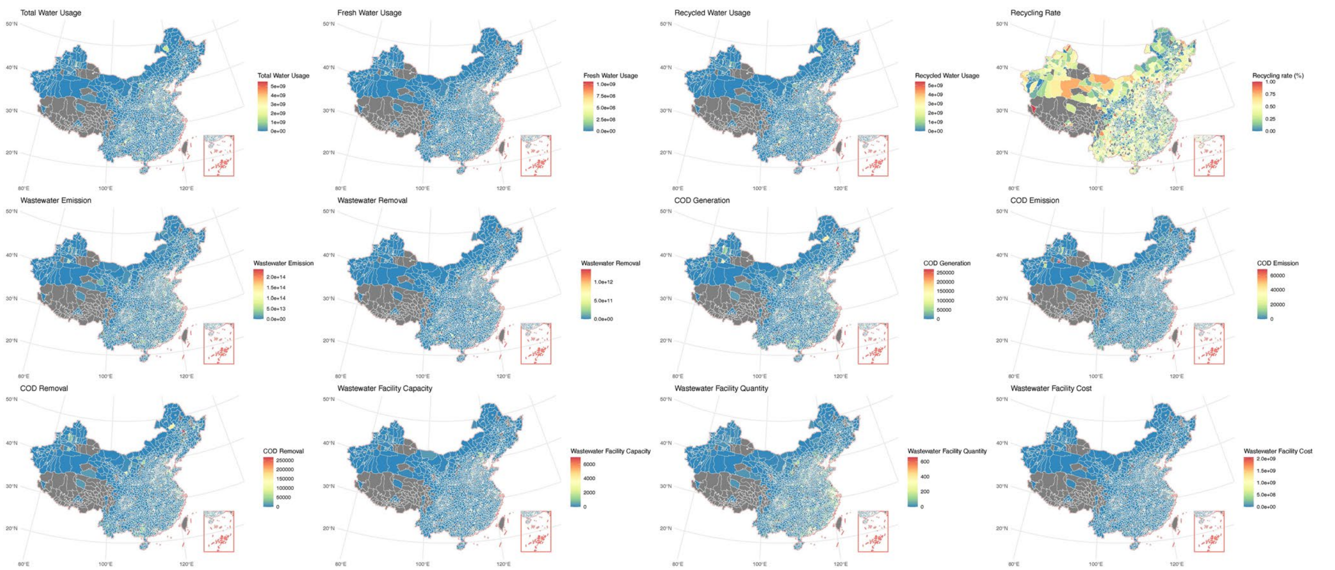

Fig. 5 Distribution of variables by county in the dataset, 2013.

Data Records

The dataset is available in Zenodo: https://doi.org/10.5281/zenodo.13822996[28]. The dataset has the potential to be applied in a wide range of aspects, including academic studies, policy formulations, and research reports.

The comprehensive dataset encompasses five distinct types of tables: "Water usage.dta", "Collapse.xlsx", "Macro data.xlsx", "Province-sectors.xlsx", "Water resources satellite account.xlsx", and "MRIO table.xlsx".

Water usage.dta: This dataset provides the fully processed data used in this study, including key variables such as year, total output, province, city, and county locations, 2-digit, 3-digit and 4-digit industrial sub-sector information, MRIO sector codes, and detailed micro-level information on water usage, recycling, and pollution. This comprehensive dataset is designed to facilitate further analysis by researchers, offering a rich foundation for studies on industrial water management and environmental impact.

Collapse.xlsx: Based on the fully processed data, this file aggregates micro-level plant water usage data by province, city, county, as well as industry. It aims to reveal water usage patterns across different industries and regions in China.

Macro Data.xlsx: This file contains macro-level water usage data obtained from the National Bureau of Statistics of China, allowing for verification and comparison with micro-level results.

Water Satellite Accounts.xlsx: This file contains data from 42 water usage satellite accounts in the MRIO tables from 1998 to 2014.

MRIO Tables.xlsx: This file includes the results of bilateral trade data between the 31 provinces in China from 1998 to 2014.

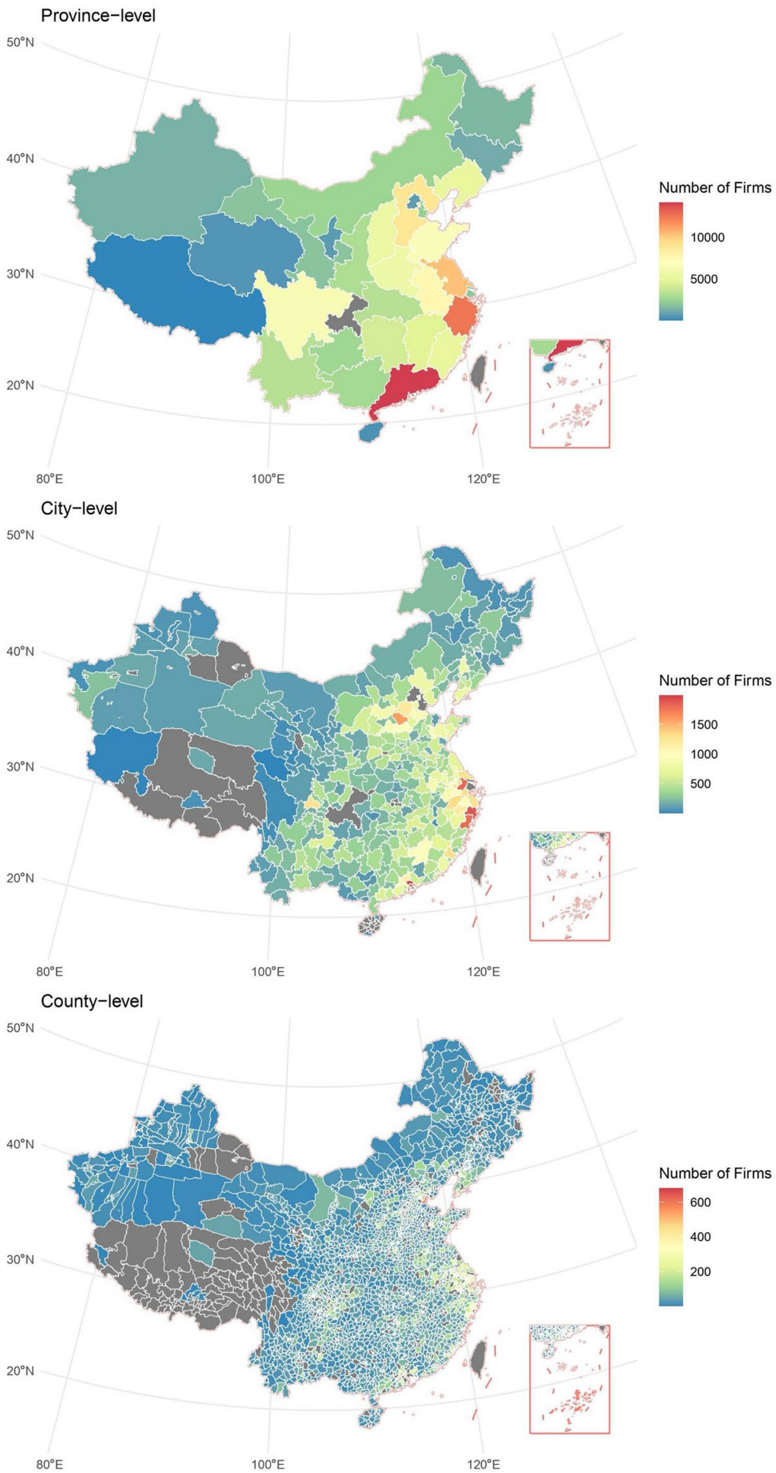

Fig. 6 Distribution of plants at province, city and county scales in the dataset, 2013. The grey area are regions without datapoints, which are regions with both a sparse population and few plants.

Technical Validation

Representativeness. The dataset is representative of Chinese water usage, waste water, and water recycling for the following reasons. First, our industrial water usage dataset has a high spatial coverage by providing data for almost all provinces in China from 1998 to 2014 (See Appendix Fig. A99 for details).

Second, our dataset covers the most water-intensive plants, as it is drawn from the Enterprise Green Development Database (EGDD), which focuses on environmentally critical plants. The purpose of the EGDD is to collect detailed information on industrial plants that have significant environmental impacts, particularly in terms of water usage and pollution, which is different from other data sources which mainly focus on economic

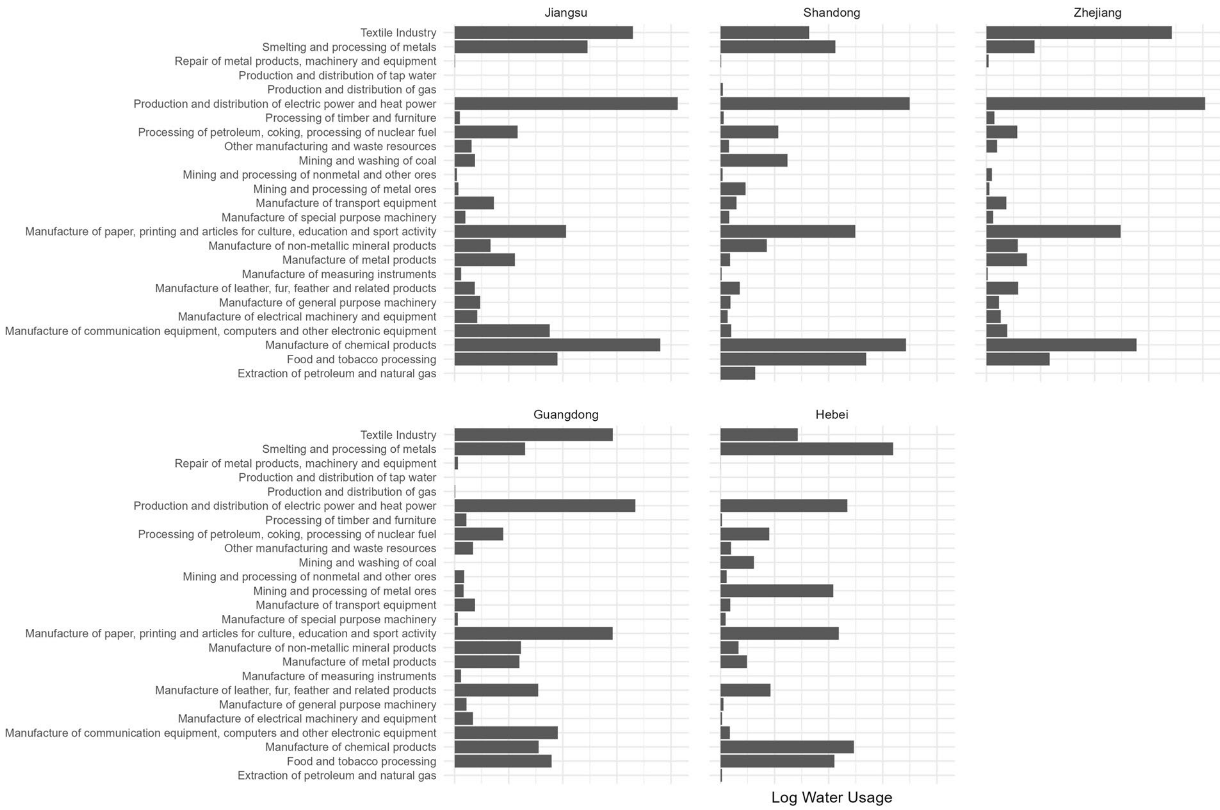

Fig. 7 Water usage in industrial sectors across the top five provinces by water use, 2014.

performance of enterprises or plants. This comprehensive coverage of environmentally significant plants allows future researchers to better understand the dynamics of water usage, recycling, and pollution control in industries that have the most substantial environmental footprints.

Third, as for the water volume of the dataset, while there is a certain level of discrepancy between our data and the estimates derived from official macro-level statistics, this can be attributed to the differences between two data DGPs. The provincial or sectoral water use data from the official statistics have been generated by estimation using fixed coefficients for sector-specific water intensity in 2008. They cannot fully capture the trends and fluctuations at fine scales as we present. Unlike macro-level estimations, our data is based on the aggregation of reports from industrial plants, providing a more precise representation of water usage at the ground level. Figure 8 presents the comparison results. In all provinces and across all years, our micro-level data tends to be lower than the macro-level estimates, suggesting efficiency improvement not fully captured by the macro coefficients. In some provinces, the two DGPs lead to close results, as in Tianjin, Hebei, Shanxi, Liaoning, Shandong, and Ningxia. In others, the two differ, with micro-level reports only a small portion of macro-level estimation, as in Shanghai, Jiangsu, Anhui, Fujian, Hubei, and Hunan.

Comparisons with other data sources. There are multiple sources of water-related statistics in China, as shown in Table 1. Most of the datasets are on provincial level and publicized annually. For instance, the National Water Resources Bulletins and the National Statistical Yearbooks all provide provincial water withdrawal data annually for three broad sectors—agriculture, industry, and service. The National Water Conservancy Statistical Yearbooks provides provincial water withdrawal data annually for agriculture and industry sectors. The Province and City Water Resources Bulletin provides provincial/city water withdrawal data annually for four broad sectors—agriculture, industry, household, and ecology sectors. The National Economic Census Yearbooks provide data on a more disaggregate level, which includes 42 sectors, but the dataset is only available in the economic census year, i.e., 2008. Both National Urban-Rural Construction Statistical Yearbooks and Province Statistical Yearbooks provide water consumption data for the household sectors on the provincial level. In 2024, the city-level water withdrawal and scarcity accounts of China has been released for year 2015, which covers detailed sectoral water withdrawal and scarcity data. The China Industrial Water Withdrawal Dataset was also released in 2024, using plant-level water usage data in 2008 to estimate long-term grid-level water withdrawal in China. Nevertheless, to date, a high-resolution dataset that covers water usage, pollution, and recycling on the plant or county level is still unavailable, which limits our understanding of the mechanisms of water management in China. The DGP in the current dataset enables researchers to trace the water usage in a bottom-up method, with consistent and comparable values across multiple scales—at province-, city-, and county-levels and by 4-digit, 3-digit, and 2-digit industrial classifications. In addition, there are multiple aspects of index provided in the

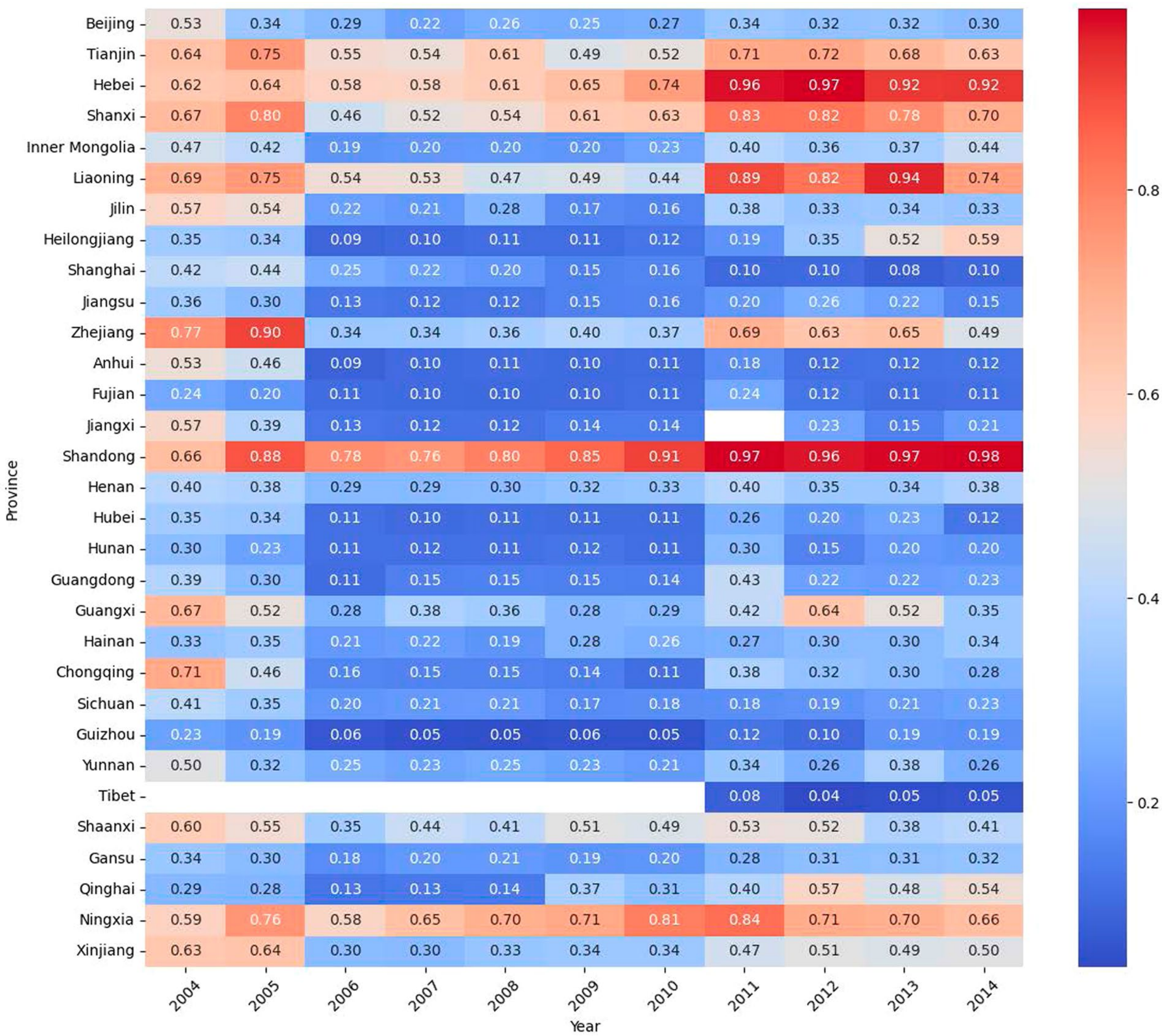

Fig. 8 Coverage of water usage compared to government released data: 2004–2014.

current dataset, in which variables on water withdrawal, water-related pollutions, water recycling, and waste water processing can be obtained.

To further examine the validity, accuracy, and reliability of the water usage data compiled in this study, we also compared the water usage structure in 2008 from our dataset with the water usage data gathered from the Second National Economic Census in 2008. The National Economic Census 2008 reports water usage across all 31 Chinese provinces and cities and 41 industries, including three large aggregate industries (mining; manufacturing; electricity, gas, and water supply) and 38 sub-industries. Based on large-scale plant census, it provides one of the most comprehensive water usage datasets in China, which has been adopted for most studies investigating the water-related studies. However, there are two key limitations to its use. First, although based on micro-level enterprise surveys, it only provides provincial and sectoral aggregated data. Second, since its data was collected only in 2008, using the 2008 water usage intensity to estimate other years could compromise accuracy. Thus, if a detailed dataset for 2008 aligns well with this published aggregated data while also providing granular details and data for other years, it would confirm the dataset's validity. We conduct a comparison between our compiled dataset and the National Economic Census 2008. Figure 9 indicates a high degree of correlation in most provinces, for instance, Yunnan, Inner Mongolia, Jiangsu, Zhejiang, Hainan, and Guizhou. Only a few provinces have a correlation index slightly lower than 0.50, for instance, Shanghai and Shaanxi. The congruence between our data and the National Economic Census 2008, despite some minor discrepancies, validates the robustness of our data collection and analysis methodologies.

Further discussions and limitations. Several opportunities for future dataset compilation exist. First, we do not have enough observations from the service sector, which is rapidly developing in China and could constitute an increasing share of water usage as well. Second, the post-2014 data is non-public and not available, making it unsuitable for inclusion in this dataset. Despite these limitations, the current dataset offers a high-resolution

Source	Data	Sector	Time	Spatial resolution	Temporal resolution
China Water Resources Bulletin	Water withdrawal	Broad agriculture, industry, and service sectors	1997–2022	Province	Year
China Statistical Yearbook	Water withdrawal	Broad agriculture, industry, and service sectors	1999–2023	Province	Year
China Water Conservancy Statistical Yearbook	Water withdrawal	Broad agriculture and industry sectors	2009–2021	Province	Year
Province and City Water Resources Bulletin	Water withdrawal and availability	Broad agriculture, industry, household, and ecology sectors	1991–2023	Province/City	Year
China Economic Census Yearbook	Water withdrawal intensity	42 sectors	2008	Province	Year
China Urban-Rural Construction Statistical Yearbook	Water consumption	Urban and rural households	2006–2022	Province	Year
Province or City Statistical Yearbooks	Water consumption	Urban and rural households	2002–2023	Province	Year
City level water withdrawal and scarcity accounts of China[29]	Water withdrawal and scarcity	63 economic-socio-environmental sectors	2015	City	Year
China Industrial Water Withdrawal Dataset[30]	Water withdrawal	Industry	1965–2020	Grid (0.1 and 0.25°)	Monthly

Table 1. Comparison with existing data sources.

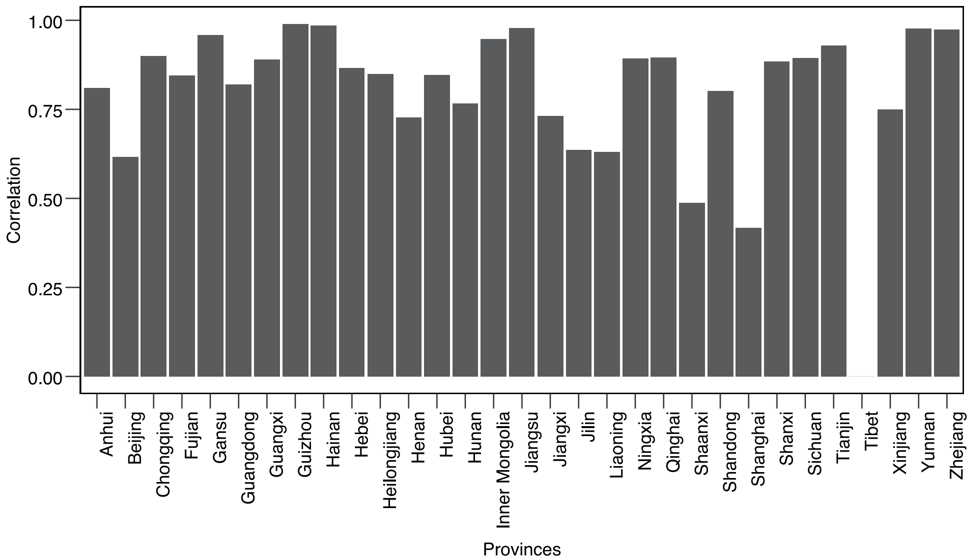

Fig. 9 Correlation between the dataset compiled and the National Economic Census 2008.

industrial water usage data in China. It provides a robust foundation for analyzing long-term trends in water usage, pollution, and recycling. Furthermore, the structure of our dataset is designed to support easy updates, so if more recent, publicly available data of sufficient quality is released, it can be quickly integrated into the dataset, ensuring consistency and completeness.

Usage Notes

The dataset described here is available from Zenodo: https://doi.org/10.5281/zenodo.13822996[28] under the Creative Commons Attribution, Non-Commercial, No Derivatives 4.0 International (CC-BY-NC-ND) license. The dataset includes all resources as described in Data Records. This dataset offers flexibility for researchers aiming to study industrial water usage, recycling, and pollution in China across various dimensions. The dataset includes geospatial information so that Researchers can use the spatial identifiers (province, city, county, and plant) to link the dataset with other external sources, such as economic census data or environmental policy records. The dataset supports sectoral comparisons, providing insights into water-intensive industries and their environmental impacts.

Code availability

The code is accessible in Zenodo: https://doi.org/10.5281/zenodo.13822996[28] and users can generate their own variables of interest based on their research objectives.

Water usage processing.do: Construction of the high-resolution industrial water use dataset in China, including all data cleaning, data merging and data calculation processes.

MRIO_Water.py: Construction of WE-MRIO model for 1998–2014. Processing MRIO tables and water resources satellite matrices for different years to produce virtual water resources trade among 31 provinces in China.

Variable_describe.R: Used to describe key variables statistical distribution in the industrial water use dataset in China.

Map distribution. R: Mapping the distribution of each key variable and the corresponding plant in China's industrial water use dataset at multiple scales such as province, city, and county from 1998–2014.

Province_sector. R: Mapping water usage in industrial sectors across the top five provinces by water use from 1998 to 2014.

Received: 23 June 2024; Accepted: 2 December 2024;
Published online: 05 December 2024

References

1. Vörösmarty, C. J., Green, P., Salisbury, J. & Lammers, R. B. Global Water Resources: Vulnerability from Climate Change and Population Growth. *Science* **289**, 284–288 (2000).
2. Mekonnen, M. M. & Hoekstra, A. Y. Four billion people facing severe water scarcity. *Sci. Adv.* **2**, e1500323 (2016).
3. Prudhomme, C. *et al*. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Natl. Acad. Sci. USA*. **111**, 3262–3267 (2014).
4. Durack, P. J., Wijffels, S. E. & Matear, R. J. Ocean Salinities Reveal Strong Global Water Cycle Intensification During 1950 to 2000. *Science* **336**, 455–458 (2012).
5. Haddeland, I. *et al*. Global water resources affected by human interventions and climate change. *Proc. Natl. Acad. Sci. USA*. **111**, 3251–3256 (2014).
6. Wang, M. *et al*. Accounting for interactions between Sustainable Development Goals is essential for water pollution control in China. *Nat Commun* **13**, 730 (2022).
7. Grafton, R. Q. *et al*. Global insights into water resources, climate change and governance. *Nature Clim Change* **3**, 315–321 (2013).
8. Ma, T. *et al*. Pollution exacerbates China's water scarcity and its regional inequality. *Nat Commun* **11**, 650 (2020).
9. Flörke, M. *et al*. Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environmental Change* **23**, 144–156 (2013).
10. Gong, P., Yin, Y. & Yu, C. China: Invest Wisely in Sustainable Water Use. *Science* **331**, 1264–1265 (2011).
11. Lin, C.-Y., Miller, A., Waqar, M. & Marston, L. T. A database of groundwater wells in the United States. *Sci Data* **11**, 335 (2024).
12. Mialyk, O. *et al*. Water footprints and crop water use of 175 individual crops for 1990–2019 simulated with a global crop model. *Sci Data* **11**, 206 (2024).
13. Liu, J. & Yang, W. Water Sustainability for China and Beyond. *Science* **337**, 649–650 (2012).
14. Yang, H., Flower, R. J. & Thompson, J. R. Sustaining China's Water Resources. *Science* **339**, 141–141 (2013).
15. Lin, J. *et al*. Making China's water data accessible, usable and shareable. *Nat Water* **1**, 328–335 (2023).
16. Guan, D. *et al*. Lifting China's Water Spell. *Environ. Sci. Technol.* **48**, 11048–11056 (2014).
17. Barnett, J., Rogers, S., Webber, M., Finlayson, B. & Wang, M. Sustainability: Transfer project cannot meet China's water needs. *Nature* **527**, 295–297 (2015).
18. Tao, T. & Xin, K. Public health: A sustainable plan for China's drinking water. *Nature* **511**, 527–528 (2014).
19. Pollution exacerbates China's water scarcity and its regional inequality | Nature Communications. https://www.nature.com/articles/s41467-020-14532-5.
20. Managing nitrogen to restore water quality in China | Nature. https://www.nature.com/articles/s41586-019-1001-1.
21. Huang, Y., Meng, F., Liu, S., Sun, S. & Smith, K. China's enhanced urban wastewater treatment increases greenhouse gas emissions and regional inequality. *Water Research* **230**, 119536 (2023).
22. Qian, H. *et al*. Air pollution reduction and climate co-benefits in China's industries. *Nat Sustain* **4**, 417–425 (2021).
23. Qian, H. *et al*. China industrial environmental database 1998–2015. *Sci Data* **9**, 259 (2022).
24. Zhang, C., Zhong, L., Fu, X., Wang, J. & Wu, Z. Revealing Water Stress by the Thermal Power Industry in China Based on a High Spatial Resolution Water Withdrawal and Consumption Inventory. *Environ. Sci. Technol.* **50**, 1642–1652 (2016).
25. Bakker, K. Water Security: Research Challenges and Opportunities. *Science* **337**, 914–915 (2012).
26. Long, Y. *et al*. Monthly direct and indirect greenhouse gases emissions from household consumption in the major Japanese cities. *Sci Data* **8**, 301 (2021).
27. Zheng, H. *et al*. Chinese provincial multi-regional input-output database for 2012, 2015, and 2017. *Sci Data* **8**, 244 (2021).
28. Li, M., Tong, Y., Zhu, J. & Xu, S. A high-resolution multi-scale industrial water use database in China. *Zenodo* https://doi.org/10.5281/zenodo.13822996 (2024).
29. Zhang, Z. *et al*. City level water withdrawal and scarcity accounts of China. *Sci Data* **11**, 449 (2024).
30. Hou, C. *et al*. High-resolution mapping of monthly industrial water withdrawal in China from 1965 to 2020. *Earth Syst. Sci. Data* **16**, 2449–2464 (2024).

Acknowledgements

This research is supported by National Natural Science Foundation of China (52370196, 72204159).

Author contributions

J.Z. and M.L. designed the study. Y.T. worked on the code. J.Z., Y.T. and M.L. contributed to data quality control. J.Z. contributed to data standard. J.Z., Y.T., M.L. and S.X. prepared the manuscript. J.Z. supervised the study and secured financial support to carry out the research.

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/s41597-024-04204-0.

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

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