



OPEN

DATA DESCRIPTOR

Chinese industrial air pollution emissions based on the continuous emission monitoring systems network

Ling Tang^{1,2}, Min Jia¹, Junai Yang¹, Ling Li³✉, Xin Bo^{4,5}✉ & Zhifu Mi⁶✉

As the world's largest industrial producer, China has generated large amount of industrial atmospheric pollution, particularly for particulate matter (PM), SO₂ and NO_x emissions. A nationwide, time-varying, and up-to-date air pollutant emission inventory by industrial sources has great significance to understanding industrial emission characteristics. Here, we present a nationwide database of industrial emissions named Chinese Industrial Emissions Database (CIED), using the real smokestack concentrations from China's continuous emission monitoring systems (CEMS) network during 2015–2018 to enhance the estimation accuracy. This hourly, source-level CEMS data enables us to directly estimate industrial emission factors and absolute emissions, avoiding the use of many assumptions and indirect parameters that are common in existing research. The uncertainty analysis of CIED database shows that the uncertainty ranges are quite small, within $\pm 7.2\%$ for emission factors and $\pm 4.0\%$ for emissions, indicating the reliability of our estimates. This dataset provides specific information on smokestack concentrations, emissions factors, activity data and absolute emissions for China's industrial emission sources, which can offer insights into associated scientific studies and future policymaking.

Background & Summary

China has been suffering from severe air pollution¹. Industrial sectors contributed the majority of China's air pollutant emissions, representing 72.8–86.1%, 74.3–91.0% and 40.7–79.1% of national anthropogenic particulate matter (PM, comprising all PM particle sizes)^{1,2}, sulphur dioxide (SO₂)^{1–3} and nitrogen oxide (NO_x)^{1–3} emissions, respectively, between 2010 and 2018. These air pollutants constituted the primary precursors of PM_{2.5} (PM with an aerodynamic diameter within 2.5 μm) pollution, which poses severe environmental problems and public health burden¹.

To control the industrial emissions, a nationwide, dynamic and up-to-date emission inventory is critical for accurately analysing industrial emission characteristics and targeted policymaking. There are some bottom-up emission inventories of atmospheric pollutants for China's industrial emissions, including the Multi-resolution Emission Inventory for China (MEIC)^{1,4}, the Regional Emission inventory in ASia (REAS)⁵, the Community Emissions Data System (CEDS)^{6–8} and other emission datasets^{9–14}. However, due to the lack of actual monitoring data, these databases resort to average emission factors (i.e., atmospheric pollutant emissions per unit of production or fossil fuel combustion) compiled by previous studies^{15,16} or official guidebooks^{17,18} (such as the National Pollution Source Census¹⁷ published by the Council of State Governments (CSG)), which bears several shortages. First, these average emission factors do not entail direct CEMS-monitored observations but are proxies for various assumptions or indirect parameters (about operational conditions and control measures), which result in high uncertainty¹⁹. Second, based on many indirect parameters and associated assumptions,

¹School of Economics and Management, Beihang University, Beijing, 100191, China. ²School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China. ³International School of Economics and Management, Capital University of Economics and Business, Beijing, 100070, China. ⁴Department of Environmental Science and Engineering, Beijing University of Chemical Technology, Beijing, 100029, China. ⁵BUCT Institute for Carbon-neutrality of Chinese Industries, Beijing, 100029, China. ⁶The Bartlett School of Sustainable Construction, University College London, London, WC1E 7HB, UK. ✉e-mail: lingli@cueb.edu.cn; boxin@buct.edu.cn; z.mi@ucl.ac.uk

the emissions factors employed in previous inventories are assumed invariable within a given province^{12,14}, region¹³ or nation^{1,4}, thereby failing to reflect individual heterogeneities throughout industrial facilities²⁰. Third, available emissions factors have been evaluated up to 2017²¹, and the effect of latest mitigation policies^{22,23} on industrial sectors through technology upgrading and operational adjustment has not been considered. Therefore, introducing direct and real-time CEMS-monitored observations can significantly reducing the estimation uncertainty due to the application of indirect and constant average emission factors.

China has started building a national continuous emissions monitoring system (CEMS) network (<http://www.envsc.cn/>) for high-emitting industrial stationary sources (such as operating units, machines, facilities or boilers for production) since 2007²⁴, to directly measure hourly and source-level PM, SO₂ and NO_x concentrations^{20,25–30}. During 2015–2018 period, the CEMS network involved comprehensive industrial sectors, particularly energy- and emission-intensive sectors such as thermal power industry (representing 57.7%–77.1% of plants and 95.9–97.4% of total national capacity), iron and steel industry (representing 62.9%–71.6% of plants and 74.2–88.3% of crude steel production) and cement industry (representing 63.5%–77.2% of plants and 78.9–87.6% of total clinker production). To improve the quality and reliability of CEMS system, China has implemented a number of policy actions: developing detailed specifications and technical guidelines for CEMS' proper operation, preservation and regulation^{27,31}; conducting quarterly random inspections to avoid data manipulation³²; and comparing monitoring values among emission sources to determine outliers³³. To date, some research has employed CEMS data to analyse industrial emissions from limited industries, including thermal power industry^{20,26,34}, iron and steel industry²⁸ and cement industry³⁵. However, these data based on actual monitoring measurements have not yet been extended to other industrial sectors, and a comprehensive analysis across different industry sectors has not been performed.

Here, we contribute to addressing above research gap by developing a new nationwide database of industrial emissions based on CEMS measurements, named Chinese Industrial Emissions Database (CIED). The CIED database considers comprehensive industrial sectors in China from 2015 to 2018, adding up to 10,933 plants and 19,032 facilities. In particular, the database introduces all available actual monitoring data of smokestack concentrations from CEMS network (exclusively provided by the China's Ministry of Ecology and Environment (MEE), <http://www.envsc.cn/>) for PM, SO₂ and NO_x from industrial plant stacks across China during 2015–2018, and estimates nationwide, real-time and dynamic industrial emission factors and absolute emissions on a source and monthly basis. The CEMS data can sufficiently provide a direct, simple estimates of nationwide, source-level and dynamic emission factors and absolute emissions for Chinese industrial sectors, which can address the above three limitations of using average emission factors. First, the CEMS database offers real-time measurements that avoid using diverse assumptions and indirect parameters employed in average emission factors of previous emission inventories and thus reduce the estimation uncertainty³⁶. Second, the hourly, source-level actual CEMS measurements enhance the spatio-temporal resolutions of emission factors and absolute emissions, which can effectively highlight the heterogeneous and dynamic characteristics of industrial emissions over periods^{26,37}. Third, the CEMS-monitored observations for the 2015–2018 period are applied, and the detailed, up-to-date emission factors and emission inventory are updated, directly reflecting the potential emission reduction effects of recent air pollution control policies^{22,23}. Moreover, the CIED dataset also encompasses other stack specific information derived from the MEE (<http://permit.mee.gov.cn/>), regarding geographic allocations (e.g., latitude and longitude), physical parameter (e.g., diameter, height and temperature) and so on. In addition, uncertainty analyses for total emissions of PM, SO₂ and NO_x across 2015–2018 show that our estimates are more robust (with 95% confidence interval (CI) of [−0.2%, 0.1%]) relative to prior studies (with 95% CI of [−76.0%, 136.0%])^{10,13,38–42}. This CEMS-based CIED dataset can be employed to conduct a more accurate analysis of overall, detailed and dynamic characteristics of industrial emissions, serve mitigation policy making for China, and offer insights for other countries looking to control industrial emissions^{43,44}.

Methods

Scopes and databases. The CIED database encompasses comprehensive industrial sources in mainland China from 2015 to 2018 in all the provinces and municipalities (totaling 26 and 5, respectively) in mainland China. According to the Industrial Classification for National Economic Activities (GB/T 4754–2017)⁴⁵, these industrial emission sources can be aggregated into 10 sectors or 33 subsectors (deposited at figshare⁴⁶). Thereafter, these sectors can be further divided into 170 subcategories (by fuel types, processes or products; deposited at figshare⁴⁶). Specifically, by fuel type, thermal power industry is classified into 5 subcategories (i.e., coal, gas, oil, biomass and other fuels-based burning thermal power industries), according to the varieties of fossil energies used in the power generation. By process, iron and steel industry is allocated to 7 production processes (i.e., sintering, pelletizing, coking, ironmaking, steelmaking via a basic oxygen furnace, steelmaking via an electric arc furnace and steel rolling) or 22 associated processes (e.g., sinter machine heads in sintering); and cement industry can fall into 2 processes of kiln heads and kiln tails. By product, other subsectors are classified into 141 subcategories.

The CIED dataset is a new dataset that offers nationwide, detailed, dynamic emission factors and total emissions of PM, SO₂ and NO_x between 2015 and 2018 for Chinese industrial sources across different fuel types, production processes or products categories. Compared with existing inventories^{1,4–14}, the CIED dataset has the unique advantage in reducing estimation uncertainty by using real CEMS-monitoring data rather than average emission factor (and many assumption and uncertain parameters thereof).

In particular, the CIED dataset incorporates two databases, i.e., the CEMS database and source-specific information. The CEMS database—actual, source-level and hourly monitoring data of smokestack concentrations of PM, SO₂ and NO_x for stationary industrial emission sources—are recorded by Chinese CEMS network

and released by the MEE. Overall, a total of 17,134 sources (associated with 7,708 generating plants) across different industrial sectors are encompassed in the CEMS network from 2015 to 2018.

Source-specific information is provided by national facility-level database for Chinese industrial sources. Specifically, the facility-specific information for each individual industrial source involves activity level (industrial production or power generation, inputs and fuel consumption; yearly), unit or facility property (geographical location, production processes involved, emission sources associated, facility type, facility scale and age), quality of fuel and raw material and pollution control technology (category and removal efficiency), which are derived from the MEE^{20,29,35}.

Pre-processing of CEMS data. The CEMS includes a sampling system (to filter and sample flue gases), an analysis system (to evaluate flue gas indicators, especially smokestack concentrations) and a data-processing system (to collect, process and report real-time measurements)^{24,27}. These three systems of CEMS should be carefully operated, maintained and examined, in order to prevent observation biases mainly in the sampling and analysis systems (in terms of zero drift, span drift and indication errors) and invalid data communication and data loss in data-processing system (leading to null and invalid values)²⁷.

To ensure the quality of CEMS data, Chinese governments have promulgated a series of measures to prevent systematic biases, including: formulating detailed specifications and technical requirements for local government agencies and industrial plants to perform and superintend the proper operation, maintenance and regulation of CEMS network^{27,31}; performing quarterly random examinations to prevent data manipulation³²; comparing data among emission sources to identify outliers³³; mandating plants to regularly calibrate, maintain and verify CEMS instruments^{24,47}; and employing third parties to conduct technical validation for the CEMS network²⁴. According to these official documents, all state-monitored firms are required to post their CEMS measurements to the local authorities via associated provincial online platforms. Then, the local governments randomly check the authenticity of the reported data at least once per quarter^{20,27,32}, and publicly disclose the relevant inspection results through the same online platforms^{20,48,49}. The firms that engage in data manipulation (including deletion, distortion and falsification of CEMS data, etc.) are subject to strict financial and criminal penalties^{50,51}.

Even with all the above efforts, there are still a small number of invalid values in the CEMS database (accounted for 2.6%–3.3% of the total data from 2015 to 2018, deposited at figshare⁴⁶), including missing data (nulls), zeros and abnormal (or extreme) observations, which should be seriously handled in accordance with relevant official documents and guidelines issued by the Chinese government. In general, nulls or zeros can be treated in three different methods depending on the duration²⁴. First, we treat the nulls or zeros that last below 5 day(s) as invalid values and set nulls or zeros successive from 1 to 24 hour(s) to the averages of the two closest valid values before and after them^{24,52}:

$$\hat{C}_{f,i,y,m,h} = \frac{C_{f,i,y,m,h-p} + C_{f,i,y,m,h+j}}{2} \quad (1)$$

where $C_{f,i,y,m,h}$ is the stack gas concentration (g m^{-3}) monitored by CEMS network, which denotes the real-time measurements of pollutant f produced by facility i in year y , month m and hour h ; $\hat{C}_{f,i,y,m,h}$ represents the estimated value for the nulls or zeros $C_{f,i,y,m,h}$; $C_{f,i,y,m,h-p}$ and $C_{f,i,y,m,h+j}$ indicate the closest last valid values (p hour(s) before) and next valid values (j hour(s) after), respectively, of the nulls or zeros $C_{f,i,y,m,h}$. Second, the nulls or zeros consecutive for more than 24 hours but less than 5 days are interpolated with the effective monthly averages near the time²⁴:

$$\hat{C}_{f,i,y,m,h} = \bar{C}_{f,i,y,m,\bullet} \quad (2)$$

where $\bar{C}_{f,i,y,m,\bullet}$ means the average of the hourly known values for the same atmospheric pollutant, production facility, operating year and month as $C_{f,i,y,m,h}$. Conversely, we consider nulls or zeros successive for more than 5 days as an overhaul and ignore them²⁴, in the light of estimation regulation. Furthermore, we use a data visualization to identify extreme data (in terms of the values outside the CEMS instruments' measurement range) as outliers and treat those data in a same way as nulls (or zeros) in accordance with the authoritative regulations²⁴.

Estimation of emission factors and emissions. Using actual CEMS-monitored observations for nationwide industrial sources, we can directly measure the emission factors for PM, SO₂ and NO_x on a source and hourly basis, which is the main contribution of this work and can enhance the estimate accuracy and avoids the use of various assumptions or indirect parameters that are common in existing research^{20,26,29,35}:

$$EF_{f,i,y,m,h} = C_{f,i,y,m,h} V_{i,y} \quad (3)$$

where $EF_{f,i,y,m,h}$ stands for the emission factor ($\text{g per activity data}$), expressed as the emissions mass per unit of production or fuel consumption; $V_{i,y}$ denotes the theoretical flue gas rate (m^3 per activity data), defined as the volume of flue gas per unit of product or fuel consumption. Given that the CEMS equipment installed at smokestacks are required to monitor the abated smokestack concentrations after the effect of pollution control technology (if any), the abated emission factors can be estimated in a direct way without considering the removal efficiency-related parameters²⁴.

Since the clean air policies and relevant regulations mainly focus on emission concentrations, vast quantities of other monitoring observations (especially flue gas rates) are missing from the CEMS database. Therefore, the application of theoretical flue gas rates in our estimation can significantly prevent serious underestimation

Sectors	Variables contributing to uncertainties of emission factors			Variables contributing to uncertainties of emissions			
	CEMS data	Theoretical flow gas rates	Altogether	CEMS data	Theoretical flow gas rates	Activity data	Altogether
Mining and quarrying	±5.85%	±3.80%	±7.13%	±4.06%	±3.03%	±2.26%	±5.08%
Manufacture of foods and tobacco	±5.76%	±3.69%	±6.92%	±5.62%	±3.59%	±1.98%	±6.79%
Light industry	±5.80%	±3.83%	±7.10%	±4.26%	±3.02%	±1.46%	±5.40%
Processing of petroleum, coking and processing of nuclear fuel	±5.70%	±3.76%	±6.92%	±4.49%	±3.26%	±1.77%	±5.42%
Manufacture of chemical products	±5.82%	±3.77%	±7.03%	±3.49%	±2.90%	±1.98%	±4.69%
Manufacture of non-metallic mineral products	±5.77%	±3.79%	±7.07%	±3.00%	±2.94%	±1.46%	±3.74%
Smelting and pressing of ferrous metals	±0.23%	±0.99%	±1.02%	±0.19%	±0.09%	±0.25%	±0.31%
Smelting and pressing of nonferrous metals	±5.82%	±3.76%	±6.99%	±4.26%	±3.05%	±2.61%	±5.62%
Manufacture of equipment	±5.84%	±3.88%	±7.23%	±5.65%	±3.63%	±1.29%	±6.82%
Production and distribution of electric power	±3.63%	±5.57%	±6.75%	±0.30%	±0.21%	±0.16%	±0.37%
Overall	±5.84%	±5.57%	±7.23%	±3.23%	±2.50%	±0.89%	±3.99%

Table 1. Percentage uncertainty of the emission factors and emissions in CIED.

of the actual flue gas volume due to these missing data^{20,26,29} and flue gas leakage^{26,53}. Such theoretical values are estimated according to the systematic field measurements and analogy method conducted by the CSG^{17,21} and MEE^{54,55}, with values determining by detailed products, process, scale, raw material, technologies, and fuel types. Accordingly, the actual flue gas rate can be obtained by multiplying the theoretical flue gas rate with the real industrial production or fuel consumption. Furthermore, we examine the theoretical flue gas rates based on the actual flue gas volume from CEMS monitoring samples for thermal power industry, iron and steel industry and cement industry (covering 1516, 210 and 919 facilities, respectively). Our estimates indicate that the actual values of flue gas rates generally approach their corresponding theoretical ones, within the uncertainty range (defined as the lower and upper bounds of a 95% confidence interval around the central estimates⁴²) of ±10.1%, ±12.1% and ±6.7% respectively, at the 95% confidence level (deposited at figshare⁴⁶). The results are consistent with the finding of existing studies^{20,26,29} and confirm the application of theoretical flue gas rates.

Then, we estimate the total emissions of PM, SO₂ and NO_x for Chinese industrial sectors by multiplying the emission factors by the activity data, on a source and monthly basis¹⁹:

$$E_{f,i,y,m} = EF_{f,i,y,m} A_{i,y,m} \quad (4)$$

where $E_{f,i,y,m}$ indicates the absolute atmospheric pollutant emissions (g); $A_{i,y,m}$ means the activity level, defined as the total amount of production (e.g., kg for crude steel in iron and steel industry) or fossil fuel consumption (kg for solid or liquid fuels and m³ for gas fuels).

In the CIED dataset, we calculate the total emissions on a monthly scale, in which emission factors are aggregated from hourly values to monthly values. Notably, the comprehensive annual, facility-level activity data is only available for three industrial subsectors (i.e., thermal power industry^{20,26}, iron and steel industry²⁹ and cement industry³⁵) for 2015–2018. Therefore, we need to use the production data on a monthly basis and a provincial scale as the weights to assign the yearly facility-level activity data²⁶:

$$A_{i,y,m} = \frac{A_{p_i,s_i,y,m}}{\sum_{m=1}^{12} A_{p_i,s_i,y,m}} A_{i,y} \quad (5)$$

where the subscript s_i indicates the industrial subcategory s to which facility i belongs; p_i means the province p where facility i is located; $A_{p_i,s_i,y,m}$ denotes the monthly provincial production of industrial subcategory s in province p , which is derived from the official statistical yearbook (i.e., *Chinese Energy Statistics Yearbooks*⁵⁶ and *China Statistical Yearbooks*⁵⁷) and reports (available at <http://www.cementchina.net/>). Given that the lack of comprehensive facility-level activity data for other 30 subsectors (covering 74 types of industrial products), we directly use monthly province-level activity data (derived from *China Statistical Yearbooks*⁵⁷), i.e., $A_{i,y,m} = A_{p_i,s_i,y,m}$ or scale annual data (from *China Statistical Yearbooks*⁵⁷, *China Mineral Resources*⁵⁸, *China's Building Materials Industry Yearbook*⁵⁹ and the association of China refractories industry) to monthly levels using the proxies of monthly production of counterpart products.

Uncertainties. We consider the uncertainties stemming from the volatility in the CEMS-monitored observations, theoretical flue gas rates and estimated monthly activity data, assuming that uncertainties in these three parameters are independent. Using Monte Carlo approach, we perform the systematic uncertainty analyses to examine the reliability and robustness of our estimated results introducing actual CEMS measurements. The detailed analysis steps are as follows: (a) assume the probability distributions for each tested model variable (CEMS-based smokestack concentrations, theoretical flue gas rates or activity levels) and obtain the related distribution parameters (e.g., mean and the standard deviation) as inputs to the Monte Carlo simulation; (b) produce

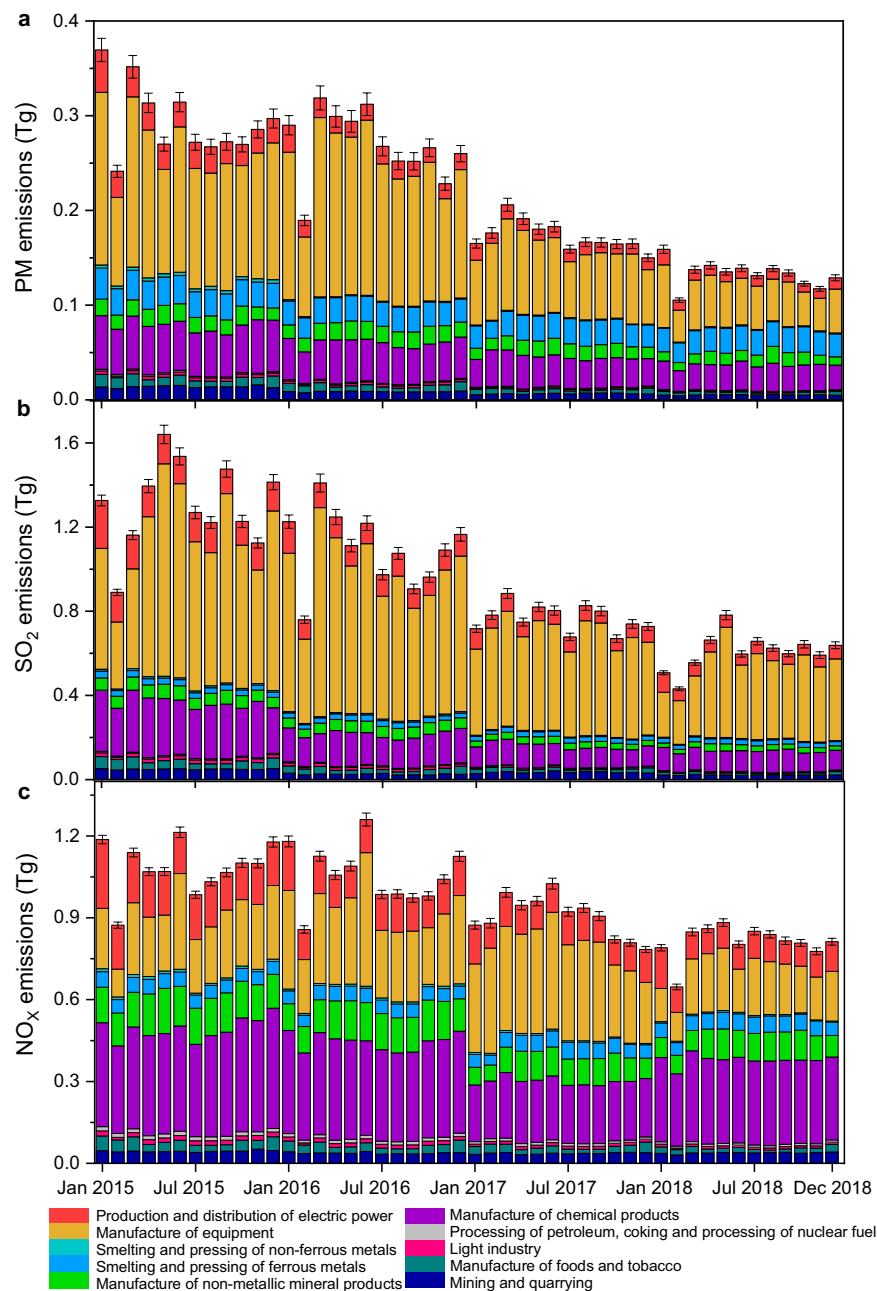


Fig. 1 Estimated emissions of Chinese industrial sectors from 2015 to 2018. (a–c), Estimated monthly industrial emissions (Tg) of PM (a), SO₂ (b) and NO_x (c). The error bars represent the uncertainty ranges.

random values following their respective probability distributions through Monte Carlo approach; (c) input random values to Eqs. (3–5) to generate a new group of emission factors and absolute emissions; and (d) run steps (b) and (c) for 10,000 times to obtain the range of uncertainty in our estimations in terms of 2 standard deviations (s.d.) of the above 10,000 sets of results^{19,42,60}. Our results indicate that the estimates are relatively stable, with 2 s.d. compared with the associated mean (in %; reflecting the uncertainty ranges of our estimates) within $\pm 7.2\%$ for emission factors and $\pm 4.0\%$ for absolute emissions (Table 1). In particular, based on the detailed source-level activity data, uncertainty ranges in our estimates for three subsectors (i.e., thermal power industry, iron and steel industry and cement industry) are relatively small ($\pm 6.8\%$ for emission factors and $\pm 0.2\%$ for emissions), compared to that for other subsectors ($\pm 7.2\%$ and $\pm 4.8\%$, respectively).

Uncertainties in CEMS data. To examine the volatilities in the high frequency CEMS data, we assume probability distributions (in a uniform distribution) for source-specific and monthly concentrations of each atmospheric pollutant, according to the tolerance ranges issued by the official regulation (HJ/T75-2007)²⁴. In detail, a set of legal CEMS measures are mandated to control the systematic errors within $\pm 15\%$, $\pm 5\%$ and $\pm 5\%$ for PM, SO₂ and NO_x concentrations, respectively. Regarding the emission sources without CEMS, we use

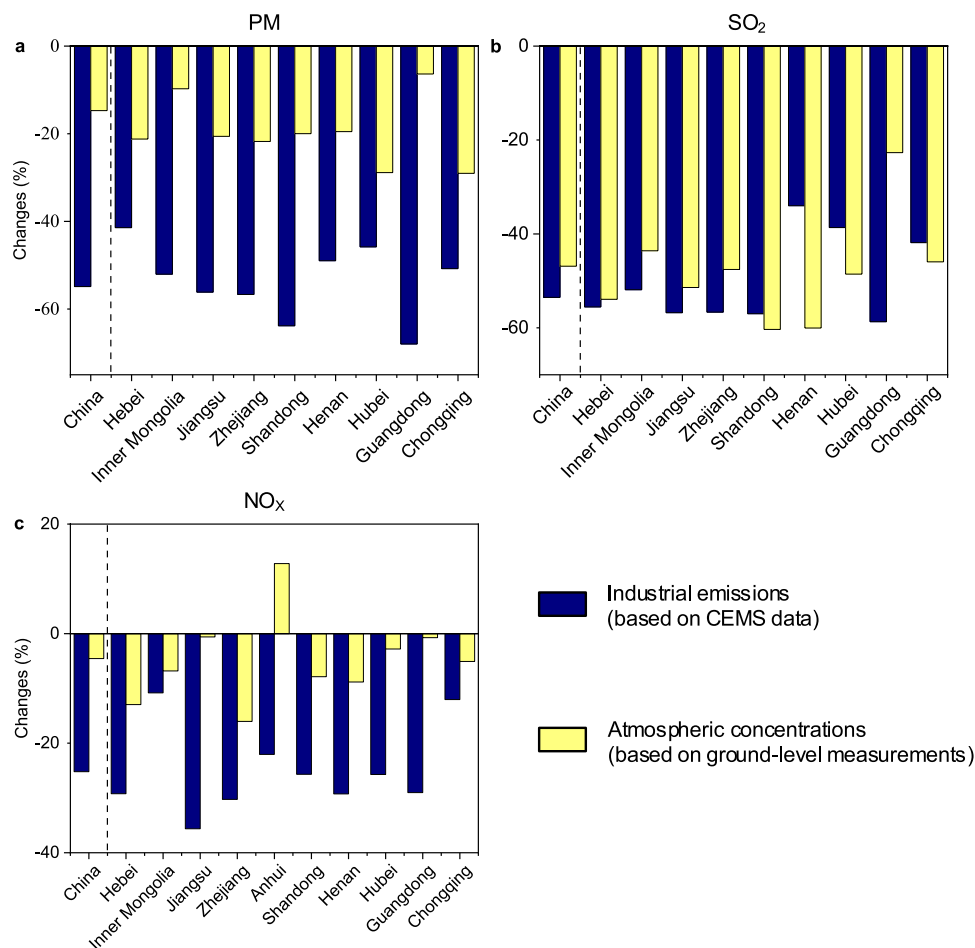


Fig. 2 Independent verification against atmospheric concentrations. **a–c**, Changes in industrial emissions (blue bars) for PM (**a**), SO₂ (**b**) and NO_x (**c**) and the associated ambient concentrations from 2015 to 2018, at the national level and China's top 10 provinces of the largest atmospheric emissions as of 2018. To verify the atmospheric impact of the emission changes for PM, SO₂ and NO_x (based on the CEMS data), ground-level PM₁₀, SO₂ and NO₂ concentration observations by national air-quality monitoring stations (yellow bars) are employed.

bootstrap simulation to randomly select samples from facilities with CEMS that in the same region, over the same period, and of similar emission source, fuel type and production process. Then, a Monte Carlo method is employed to generate random samples of pollutant concentrations following the associated distributions, and the simulations are conducted for 10,000 times to calculate the uncertainty ranges for emissions factors and absolute emissions (in terms of 2 s.d.). Our estimates indicate that the uncertainty ranges for emission factors and total emissions are within $\pm 5.8\%$ and $\pm 3.2\%$, respectively (Table 1).

Uncertainties in theoretical flue gas rates. In our estimation, the introduction of theoretical flue gas rates might arise uncertainty due to the large amount of missing data on real-monitoring flue gas rates from CEMS networks. Although this approach has the advantage of preventing severe underestimation and flue gas leakage, uncertainties may be attributed to the regardless of heterogeneities among individual facilities in production technologies, operational conditions and feedstocks, etc. Under such background, we calculate the uncertainty ranges based on the CEMS-monitored samples for 1,373, 210 and 919 facilities of thermal power industry, iron and steel industry and cement industry, respectively; and perform a single-sample two-tailed *t*-test (deposited at figshare⁴⁶) for each subcategory of these three industrial sectors. The results demonstrate that the actual CEMS monitoring flue gas rates mostly close to their theoretical values in our estimates, within a likely range of $\pm 12.1\%$ at a 95% confidence level. Then, the Monte Carlo technique is employed to generate random values of flue gas rates uniformly distributed in the relevant uncertainty ranges. In addition, we use the maximum ranges for the industrial sectors without uncertainty ranges (e.g., $\pm 10.07\%$ for thermal power facilities burning oil). With 10,000 simulations, our analysis indicate that uncertainty ranges, represented by 2 s.d., are quite small, within $\pm 5.6\%$ and $\pm 2.5\%$ for emission factors and emissions, respectively (Table 1).

Uncertainties in activity levels. To explore the uncertainty generated in the allocation of facility- and province-level activity data from yearly to monthly, we set a normal distribution with a 5% coefficient of variation

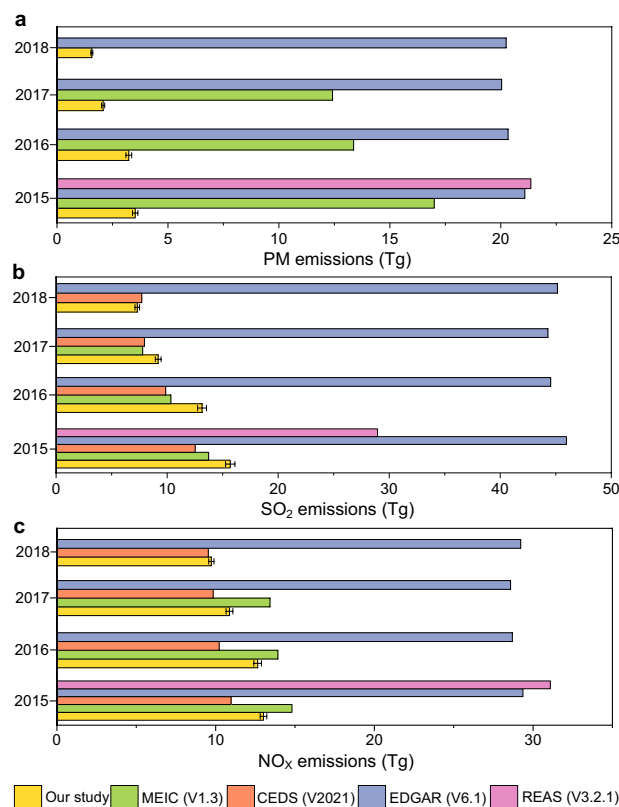


Fig. 3 Comparison of estimated Chinese industrial emissions between 2015 and 2018. (a–c), The estimated industrial emissions in China (Tg) for PM (a), SO₂ (b) and NO_x (c) in our dataset (yellow bars) and in existing datasets (MEIC (www.meicmodel.org); REAS (<https://www.nies.go.jp/REAS/>); CEDS (<https://github.com/JGCRI/CEDS>); the Emissions Database for Global Atmospheric Research (EDGAR) (<https://edgar.jrc.ec.europa.eu/>); non-yellow bars). The error bars denote the related uncertainties.

(CV, the standard deviation divided by mean) for three subsectors of thermal power industry, iron and steel industry and cement industry with comprehensive facility-specific activity level^{20,26,29,35}, and 10% CV for other industrial sectors according to existing literature⁶¹. Besides, the Monte Carlo method is conducted to produce random monthly activity values for estimated each facility or province. Relying on a total of 10,000 simulations, the uncertainty range in terms of 2 s.d. of all simulation results for absolute emissions is only $\pm 0.9\%$ from 2015 to 2018 (Table 1).

Data Records

The CIED datasets⁴⁶ are available at <https://doi.org/10.6084/m9.figshare.c.6269295>. It is organized as a set of excel datasets according to indicator and date. The indicators including emission concentrations, emission factors, activity data, absolute emissions and additional descriptions (including subcategories description, flue gas rates, comparison of uncertainty, invalid data and CEMS ranges); the date covers four years from 2015 to 2018. After merging, there are 5 excel datasets with 22 sheets provided at Figshare. In particular, the CIED database provided the high-resolution data at a source and monthly basis.

The CIED dataset introduces actual systematic smokestack concentration measurements from China's CEMS network and source-specific activity level from the MEE to directly estimate Chinese industrial emissions. In particular, the dataset presents systematic, dynamic, detailed emission factors and total emissions for PM, SO₂ and NO_x from China's industrial sources during 2015–2018, by region (including 26 provinces and 4 municipalities) and sector (33 subsectors and 170 subcategories; Fig. 1).

Technical Validation

Independent verification. The estimates drawn from the CEMS data need careful verification against other independent data, which can also provide insight as to how the large emission reductions in industrial sectors (based on CEMS data) translate to trends in regional atmospheric concentrations. Therefore, we conduct an independent verification against an atmospheric dataset (i.e., ground-level measurements by air-quality monitoring stations). In particular, we compare the changes in industrial emissions from 2015 to 2018 (based on CEMS data) with those in regional atmospheric concentrations, both at the national level and in the top 10 provinces with the largest atmospheric emissions as of 2018. This experimental design was also used in existing studies^{62,63}. In each of the 10 provinces, the industrial sectors has already been shown to have a large contribution to air pollution^{64–71}.

These provinces include Anhui, Guangdong, Jiangsu, Zhejiang, Shandong, Hubei, Chongqing, Henan, Hebei and Inner Mongolia (ranked by atmospheric emissions).

The ground-level PM₁₀, SO₂ and NO₂ concentrations measured by national air-quality monitoring stations are employed to verify the atmospheric impact of the changes in PM, SO₂ and NO_x emissions, respectively; these data are obtained from China National Environmental Monitoring Center (<http://www.cnemc.cn/>). The large reductions in PM, SO₂ and NO_x emissions from China's industrial sectors largely correlated with ground-level monitoring data. As shown in Fig. 2, the changes in all PM, SO₂ and NO_x associated atmospheric concentrations (yellow bars of Fig. 2) are generally similar to the changes in emissions from industrial sectors (blue bars).

Comparisons with existing emission databases. For verification, we compare our estimates for Chinese industrial emissions to previous datasets, as illustrated in Fig. 3. The results show that our estimates (based on the real measurements) are generally 85.15%, 20.32%, 23.21% below previous estimates. This is because existing studies resort to utilizing indirect average emission factors that were estimated up to 2017, overlooking the latest mitigation effects, especially associated with the upgraded pollution control technologies^{5–9} for PM, SO₂ and NO_x respectively. In addition, the uncertainty analysis shows that our estimation exhibits a relatively low uncertainty level (with 95% CI of [−0.2%, 0.1%]) compared to existing studies (with 95% CI of [−76.0%, 136.0%]; deposited at figshare⁴⁶)^{10,13,38–42}, by using real, hourly and facility-level CEMS measurements.

Usage Notes

The CIED dataset is subject to several limitations. First, China's CEMS network has not yet covered all industrial emission sources, and these samples can be collected and incorporated to extend a complete CIED database in the future. Second, besides air pollutants, the CIED dataset can also introduce the real measurements of greenhouse gases (particularly CO₂) and water pollutants, to support a comprehensive analysis of climate change policies and clean air policies for Chinese industrial emissions. Third, to enhance the accuracy of the estimation, future work can incorporate comprehensive high-frequency operational data (including activity data and flue gas rates) for each facility. Fourth, although Chinese governments have issued a range of stringent regulations³⁰ to guarantee the reliability of the CEMS system, the careful verification, for example, comparing CEMS data with satellite data²⁷ or ground-level monitoring data²⁸, is valuable for verifying the results drawn from the CEMS data. Given that, we would update our database in the future if data are available.

Code availability

The CIED datasets are available in the form of XLSX files. No custom code is used in the construction of the datasets.

Received: 2 November 2022; Accepted: 7 March 2023;

Published online: 22 March 2023

References

- Zheng, B. *et al.* Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos. Chem. Phys.* **18**, 14095–14111, <https://doi.org/10.5194/acp-2018-374> (2018).
- Ministry of Ecology and Environment of the People's Republic of China. *China Environmental Statistics Yearbooks* <https://www.mee.gov.cn/hjzl/sthjzk/sthjtnb/> (2010–2020).
- Tong, D. *et al.* Dynamic projection of anthropogenic emissions in China: Methodology and 2015–2050 emission pathways under a range of socio-economic, climate policy, and pollution control scenarios. *Atmos. Chem. Phys.* **20**, 5729–5757, <https://doi.org/10.5194/acp-2019-1125> (2020).
- Zheng, B. *et al.* Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020. *Earth Syst. Sci. Data* **13**, 2895–2907, <https://doi.org/10.5194/essd-13-2895-2021> (2021).
- Kurokawa, J. & Ohara, T. Long-term historical trends in air pollutant emissions in Asia: Regional Emission inventory in ASIA (REAS) version 3. *Atmos. Chem. Phys.* **20**, 12761–12793, <https://doi.org/10.5194/acp-2019-1122> (2020).
- Fan, T. *et al.* Comparison of the anthropogenic emission inventory for CMIP6 Models with a country-level inventory over China and the simulations of the aerosol properties. *Adv. Atmos. Sci.* **39**, 80–96, <https://doi.org/10.1007/s00376-021-1119-6> (2022).
- O'Rourke, P. R. *et al.* CEDS v-2021-02-05 Emission Data 1975–2019 (Version Feb-05-2021) <https://github.com/JGCRI/CEDS> (2021).
- Hoesly, R. M. *et al.* Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geosci. Model Dev.* **11**, 369–408, <https://doi.org/10.5194/gmd-2017-43> (2018).
- European Commission Joint Research Centre. *Emission Database for Global Atmospheric Research (EDGAR) (Version 6.1)*, <https://edgar.jrc.ec.europa.eu/> (2022).
- Zhao, Y., Zhang, J. & Nielsen, C. P. The effects of recent control policies on trends in emissions of anthropogenic atmospheric pollutants and CO₂ in China. *Atmos. Chem. Phys.* **13**, 487–508, <https://doi.org/10.5194/acpd-12-24985-2012> (2013).
- Jiang, P., Chen, X., Li, Q., Mo, H. & Li, L. High-resolution emission inventory of gaseous and particulate pollutants in Shandong Province, eastern China. *J. Clean. Prod.* **259**, 120806, <https://doi.org/10.1016/j.jclepro.2020.120806> (2020).
- Zhou, M. *et al.* Anthropogenic emission inventory of multiple air pollutants and their spatiotemporal variations in 2017 for the Shandong Province, China. *Environ. Pollut.* **288**, 117666, <https://doi.org/10.1016/j.envpol.2021.117666> (2021).
- An, J. *et al.* Emission inventory of air pollutants and chemical speciation for specific anthropogenic sources based on local measurements in the Yangtze River Delta region, China. *Atmos. Chem. Phys.* **21**, 2003–2025, <https://doi.org/10.5194/acp-21-2003-2021> (2021).
- Zhou, Y. *et al.* Development of a high-resolution emission inventory and its evaluation and application through air quality modeling for Jiangsu Province, China. *Atmos. Chem. Phys.* **17**, 211–233, <https://doi.org/10.5194/acp-17-211-2017> (2017).
- Li, M. *et al.* Anthropogenic emission inventories in China: A review. *Natl. Sci. Rev.* **4**, 834–866, <https://doi.org/10.1093/nsr/nwx150> (2017).
- Saikawa, E. *et al.* Comparison of emissions inventories of anthropogenic air pollutants and greenhouse gases in China. *Atmos. Chem. Phys.* **17**, 6393–6421, <https://doi.org/10.5194/acp-17-6393-2017> (2017).
- China Pollution Source Census. *Manual of Emission Coefficient of Industrial Pollution Products of the First National Pollution Source Census* (China Environmental Science, 2011).

18. European Environment Agency. *EMEP/EEA Air Pollutant Emission Inventory Guidebook 2013* <https://op.europa.eu/en/web/general-publications/publications> (2013).
19. Liu, Z. *et al.* Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature* **524**, 335–338, <https://doi.org/10.1038/nature14677> (2015).
20. Tang, L. *et al.* Substantial emission reductions from Chinese power plants after the introduction of ultra-low emissions standards. *Nat. Energy* **4**, 929–938, <https://doi.org/10.1038/s41560-019-0468-1> (2019).
21. China Pollution Source Census. *Manual of Emission Coefficient of Industrial Pollution Products of the Second National Pollution Source Census* (China Environmental Science, 2017).
22. State Council of the People's Republic of China. *Notice of the State Council on Issuing the Three-Year Action Plan for Winning the Blue-Sky Defence Battle* http://www.gov.cn/zhengce/content/2018-07/03/content_5303158.htm (2018).
23. State Council of the People's Republic of China. *Air Pollution Prevention and Control Action Plan* http://www.gov.cn/zhengce/content/2013-09/13/content_4561.htm (2013).
24. Ministry of Ecology and Environment of the People's Republic of China. *Specifications for Continuous Emissions Monitoring of Flue Gas Emitted from Stationary Sources HJ/T 75-2007 (on trial)*. <http://sthjj.cnbz.gov.cn/ztlz/wrykjsjxgk/jkbz/14461831.html> (2007).
25. Zhang, Y., Bo, X., Zhao, Y. & Nielsen, C. P. Benefits of current and future policies on emissions of China's coal-fired power sector indicated by continuous emission monitoring. *Environ. Pollut.* **251**, 415–424, <https://doi.org/10.1016/j.envpol.2019.05.021> (2019).
26. Tang, L. *et al.* Air pollution emissions from Chinese power plants based on the continuous emission monitoring systems network. *Sci. Data* **7**, 1–10, <https://doi.org/10.1038/s41597-020-00665-1> (2020).
27. Zhang, X. & Schreifels, J. Continuous emission monitoring systems at power plants in China: Improving SO₂ emission measurement. *Energy Policy* **39**, 7432–7438, <https://doi.org/10.1016/j.enpol.2011.09.011> (2011).
28. Karplus, V., Zhang, S. & Almond, D. Quantifying coal power plant responses to tighter SO₂ emissions standards in China. *Proc. Natl. Acad. Sci. USA* **115**, 7004–7009, <https://doi.org/10.1073/PNAS.1800605115> (2018).
29. Bo, X. *et al.* Effect of strengthened standards on Chinese ironmaking and steelmaking emissions. *Nat. Sustain.* **4**, 811–820, <https://doi.org/10.1038/s41893-021-00736-0> (2021).
30. Bo, X. *et al.* The spatial-temporal pattern of sintered flue gas emissions in iron and steel enterprises of China. *J. Clean. Prod.* **266**, 121667, <https://doi.org/10.1016/j.jclepro.2020.121667> (2020).
31. Ministry of Ecology and Environment of the People's Republic of China. *Measures for Environmental Monitoring Management*, http://www.gov.cn/ziliao/flfg/2007-08/07/content_708389.htm (2007).
32. Ministry of Ecology and Environment of the People's Republic of China. *Notice of Issuing the Statistics and Monitoring Methods for Total Emission Reductions of Major Pollutants in the 12th Five-Year*, http://www.mee.gov.cn/gkml/hbb/bwj/201302/t20130204_245884.htm (2013).
33. China National Environmental Monitoring Centre. *Technical Regulations for Comparing Monitoring Results of Automatically Monitoring Devices for Pollution Sources (on trial)*. <http://www.cnemc.cn/jcdt/201008/W020181010643502189858.pdf> (2010).
34. De Gouw, J. A., Parrish, D. D., Frost, G. J. & Trainer, M. Reduced emissions of CO₂, NO_x, and SO₂ from US power plants owing to switch from coal to natural gas with combined cycle technology. *Earths. Future* **2**, 75–82, <https://doi.org/10.1002/2013EF000196> (2014).
35. Tang, L. *et al.* Plant-level real-time monitoring data reveal substantial abatement potential of air pollution and CO₂ in China's cement sector. *One Earth* **5**, 892–906, <https://doi.org/10.1016/j.oneear.2022.07.003> (2022).
36. Tong, D. *et al.* Targeted emission reductions from global super-polluting power plant units. *Nat. Sustain.* **1**, 59–68, <https://doi.org/10.1038/s41893-017-0003-y> (2018).
37. Liu, F. *et al.* High-resolution inventory of technologies, activities, and emissions of coal-fired power plants in China from 1990 to 2010. *Atmos. Chem. Phys.* **15**, 13299–13317, <https://doi.org/10.5194/acp-15-13299-2015> (2015).
38. Liu, J. *et al.* Carbon and air pollutant emissions from China's cement industry 1990–2015: trends, evolution of technologies, and drivers. *Atmos. Chem. Phys.* **21**, 1627–1647, <https://doi.org/10.5194/acp-2020-631> (2021).
39. Liu, H. *et al.* A regional high-resolution emission inventory of primary air pollutants in 2012 for Beijing and the surrounding five provinces of North China. *Atmos. Environ.* **181**, 20–33, <https://doi.org/10.1016/j.atmosenv.2018.03.013> (2018).
40. Wang, X. *et al.* A unit-based emission inventory of SO₂, NO_x and PM for the Chinese iron and steel industry from 2010 to 2015. *Sci. Total Environ.* **676**, 18–30, <https://doi.org/10.1016/j.scitotenv.2019.04.241> (2019).
41. Chen, L. *et al.* Unit-based emission inventory and uncertainty assessment of coal-fired power plants. *Atmos. Environ.* **99**, 527–535, <https://doi.org/10.1016/j.atmosenv.2014.10.023> (2014).
42. Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B. & Hao, J. Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China. *Atmos. Chem. Phys.* **11**, 2295–2308, <https://doi.org/10.5194/acp-11-2295-2011> (2011).
43. Li, C. *et al.* India is overtaking China as the world's largest emitter of anthropogenic sulfur dioxide. *Sci. Rep.* **7**, 1–7, <https://doi.org/10.1038/s41598-017-14639-8> (2017).
44. Mir, K. A., Purohit, P., Cail, S. & Kim, S. Co-benefits of air pollution control and climate change mitigation strategies in Pakistan. *Environ. Sci. Policy* **133**, 31–43, <https://doi.org/10.1016/j.envsci.2022.03.008> (2022).
45. National Bureau of Statistics. *Industrial Classification for National Economic Activities GB/T 4754-2017* http://www.stats.gov.cn/tjsj/tjbz/201709/t20170929_1539288.html (2017).
46. Tang, L. *et al.* Chinese industrial air pollution emissions based on the continuous emission monitoring systems network. *figshare* <https://doi.org/10.6084/m9.figshare.c.6269295> (2023).
47. Ministry of Ecology and Environment of the People's Republic of China, *Specifications for Continuous Emissions Monitoring of SO₂, NO_x and Particulate Matter in the Flue Gas Emitted from Stationary Sources HJ 75-2017* https://www.mee.gov.cn/ywyz/fgbz/bz/bzwb/jcfffz/201801/t20180108_429327.shtml (2018).
48. Ministry of Ecology and Environment of the People's Republic of China. *Measures of Supervisory Monitoring and Information Disclosure for Pollution Sources in China's Key State-Monitored Enterprises (on trial)* <https://www.mee.gov.cn/gkml/hbb/bwj/201308/W020130801520697946114.pdf> (2013).
49. Ministry of Ecology and Environment of the People's Republic of China. *Measures of Self-monitoring and Information Disclosure for China's Key State-Monitored Enterprises (on trial)* <https://www.mee.gov.cn/gkml/hbb/bghz/201305/W020130509639333091992.pdf> (2013).
50. General Office of the Communist Party of China & General Office of the State Council of the People's Republic of China. *Suggestions to Deepen the Reform of Environmental Monitoring and Improve the Quality of Environmental Monitoring Data* http://www.gov.cn/xinwen/2017-09/21/content_5226683.htm (2017).
51. Supreme People's Court of the People's Republic of China & The Supreme People's Procuratorate of the People's Republic of China. *Explanations for Some Questions about Applicable Laws for Criminal Cases of Environmental Pollution* https://www.spp.gov.cn/zdgg/201612/t20161227_176817.shtml (2016).
52. Engels, J. M. & Diehr, P. Imputation of missing longitudinal data: A comparison of methods. *J. Clin. Epidemiol.* **56**, 968–976, [https://doi.org/10.1016/S0895-4356\(03\)00170-7](https://doi.org/10.1016/S0895-4356(03)00170-7) (2003).
53. Gilbert, A. Q. & Sovacool, B. K. Benchmarking natural gas and coal-fired electricity generation in the United States. *Energy* **134**, 622–628, <https://doi.org/10.1016/j.energy.2017.05.194> (2017).

54. Ministry of Ecology and Environment of the People's Republic of China. *Technical Specification for Application and Issuance of Pollutant Permit Iron and Steel Industry HJ846-2017* <http://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/pwxk/201708/W020170802615487081097.pdf> (2017).
55. Ministry of Ecology and Environment of the People's Republic of China. *Technical Specification for Application and Issuance of Pollutant Permit Cement Industry HJ847-2017* <https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/pwxk/201708/W020170802617396676231.pdf> (2017).
56. National Bureau of Statistics. *China Energy Statistics Yearbook* http://www.stats.gov.cn/tjsj/tjcbw/201303/t20130318_451533.html (2016–2019).
57. National Bureau of Statistics, *China Statistical Yearbook* <http://www.stats.gov.cn/tjsj/ndsj/> (2015–2018).
58. Ministry of Natural Resources, *China Mineral Resources* https://www.mnr.gov.cn/sj/sjfw/kc_19263/zgkcybg/ (2016–2019).
59. Editing Committee of China Construction Materials Industry Yearbook, *China Construction Materials Industry Yearbooks* <https://web05.cnki.net/KNavi/yearbook/Detail/CLHY/YJZCL> (2016–2019).
60. Zhao, Y. *et al.* Quantifying the uncertainties of China's emission inventory for industrial sources: From national to provincial and city scales. *Atmos. Environ.* **165**, 207–221, <https://doi.org/10.1016/j.atmosenv.2017.06.045> (2017).
61. Shan, Y. *et al.* China CO₂ emission accounts 1997–2015. *Sci. Data* **5**, 1–14, <https://doi.org/10.1038/sdata.2017.201> (2018).
62. Lu, Z. & Streets, D. Increase in NO_x emissions from Indian thermal power plants during 1996–2010: unit-based inventories and multisatellite observations. *Environ. Sci. Technol.* **46**, 7463–7470, <https://doi.org/10.1021/es300831w> (2012).
63. Li, C. *et al.* Recent large reduction in sulfur dioxide emissions from Chinese power plants observed by the Ozone Monitoring Instrument. *Geophys. Res. Lett.* **37**, L08807, <https://doi.org/10.1029/2010GL042594> (2010).
64. Chen, Y., Xie, S., Luo, B. & Zhai, C. Particulate pollution in urban Chongqing of southwest China: Historical trends of variation, chemical characteristics and source apportionment. *Sci. Total Environ.* **584**, 523–534, <https://doi.org/10.1016/j.scitotenv.2017.01.060> (2017).
65. Liu, X. *et al.* Chemical characteristics, sources apportionment, and risk assessment of PM_{2.5} in different functional areas of an emerging megacity in China. *Aerosol Air Qual. Res.* **19**, 2222–2238, <https://doi.org/10.4209/aaq.2019.02.0076> (2019).
66. Luo, R. *et al.* Investigating the influence of the implementation of an energy development plan on air quality using WRF-CAMx modeling tools: A case study of Shandong province in China. *Atmosphere* **10**, 660, <https://doi.org/10.3390/atmos10110660> (2019).
67. Qiao, X. *et al.* Source apportionment of PM_{2.5} for 25 Chinese provincial capitals and municipalities using a source-oriented Community Multiscale Air Quality model. *Sci. Total Environ.* **612**, 462–471, <https://doi.org/10.1016/j.scitotenv.2017.08.272> (2018).
68. Wang, L. *et al.* Source apportionment of PM_{2.5} in top polluted cities in Hebei, China using the CMAQ model. *Atmos. Environ.* **122**, 723–736, <https://doi.org/10.1016/j.atmosenv.2015.10.041> (2015).
69. Yin, X. *et al.* Source contributions to PM_{2.5} in Guangdong province, China by numerical modeling: Results and implications. *Atmos. Res.* **186**, 63–71, <https://doi.org/10.1016/j.atmosres.2016.11.007> (2017).
70. Zhang, L. *et al.* Pollution characteristics and source apportionment of PM_{2.5} in Baotou City. *Environ. Sci. Technol.* **41**, 65–71, <https://doi.org/10.19672/j.cnki.1003-6504.2018.03.010> (2018).
71. Zhang, X. *et al.* Characteristics of the chemical composition and source apportionment of PM_{2.5} for a one-year period in Wuhan, China. *J. Atmos. Chem.* **79**, 101–115, <https://doi.org/10.1007/s10874-022-09431-6> (2022).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (71971007, 72004144 and 72174125), the Beijing Natural Science Foundation (JQ21033), the Fundamental Research Funds for the Central Universities (buctrc202133).

Author contributions

L.T., L.L., X.B. and Z.M. designed the research. M.J., J.Y., L.L. and X.B. coordinated the data processing. L.T. and M.J. performed the experimental work. L.T., M.J., J.Y., L.L. and Z.M. wrote the paper. All authors contributed to developing, writing and polishing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.L., X.B. 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023