Air pollution emissions from Chinese power plants based on the continuous emission monitoring systems network

Ling Tang, Xiaoda Xue, Jiabao Qu, Zhifu Mi, Xin Bo, Xiangyu Chang, Shouyang Wang, Shibei Li, Weigeng Cui, Guangxia Dong

To meet the growing electricity demand, China's power generation sector has become an increasingly large source of air pollutants. Specific control policymaking needs an inventory reflecting the overall, heterogeneous, time-varying features of power plant emissions. Due to the lack of comprehensive real measurements, existing inventories rely on average emission factors that suffer from many assumptions and high uncertainty. This study is the first to develop an inventory of particulate matter (PM), SO2 and NOX emissions from power plants using systematic actual measurements monitored by China's continuous emission monitoring systems (CEMS) network over 96–98% of the total thermal power capacity. With nationwide, source-level, real-time CEMS-monitored data, this study directly estimates emission factors and absolute emissions, avoiding the use of indirect average emission factors, thereby reducing the level of uncertainty. This dataset provides plant-level information on absolute emissions, fuel uses, generating capacities, geographic locations, etc. The dataset facilitates power emission characterization and clean air policy-making, and the CEMS-based estimation method can be employed by other countries seeking to regulate their power emissions.

Background & Summary

China has become the top power producer globally, and it had the largest share (19.5–26.7%) of global power generation from 2010 to 2018. The majority (70.4–82.5% during 2010–2018) of China's power generation came from thermal power plants that combusted coal, oil plus natural gas, biomass or other fossil energy (accounting for 60.2–73.4% of the total capacity). Accompanying with large amounts of fossil energy combustion, China's thermal power plants have become major sources of air pollutants, emitting 5.0–23.5%, 15.7–38.7% and 19.1–51.5% of China's anthropogenic particulate matter (PM, defined as microscopic solid or liquid matter suspended in the atmosphere), SO2 and NOX, respectively, from 2010 to 2017. These air pollutants (representing 5.9%, 23.1% and 21.5% of China's anthropogenic PM, SO2 and NOX emissions, respectively, in 2015), through a series of physical processes and chemical reactions in the atmosphere, contributed to 7.6% of China's population-weighted PM2.5 concentration as of 2015, leading to severe haze events and human health damage nationwide.

To control power emissions, an emission inventory at high spatiotemporal resolutions is needed as the foundation for an analysis of power emission characteristics and specific policy designs. There are some detailed...
The CEAP dataset comprises all the thermal power plants operating in China, totalling 2,714 plants (or 6,267 units), from 2014 to 2017 in 26 provinces and 4 municipalities (except Hong Kong, Macao, Taiwan and Tibet; Table 1). The thermal power plants produce electricity by combusting a variety of fossil energies, which fall into 4 categories: coal, gas plus oil, biomass and others (detailed in Table 2).

The CEAP dataset integrates two databases, i.e., the CEMS data and unit-specific information. The CEMS data—the direct, real-time measurements of stack gas concentrations of PM, SO2 and NOX from China’s power plant stacks—are monitored by China’s CEMS network and reported to the China Ministry of Ecology and Environment (MEE; http://www.envsc.cn/). The CEMS data are recorded on a source and hourly basis. In total, the CEMS dataset covers 4,622 emission sources (i.e., power plant stacks) associated with 5,606 units (accounting for 98% of China’s thermal power capacity), 35,064 hours from 2014 to 2017, and 3 air pollutants (i.e., PM, SO2 and NOX) for each source-hour sample (Table 3). The MEE has also provided stack-specific information (regarding latitude and longitude, heights, temperature, diameter, etc.; http://permit.mee.gov.cn/).

Unit-specific information is also derived from the MEE, involving activity levels (energy consumption and power generation), operating capacities, geographic allocations and pollution control equipment (particularly the types and removal efficiencies) at a yearly frequency. Due to data availability, the unit information is available only until 2016, and the activity levels for 2017 are projected following the overall trends in provincial thermal power generation between 2016 and 2017 (which are available in the China Energy Statistical Yearbook[29]), under the assumption that new units constructed in 2017 have the same structures of installed capacities, energy uses and regions as those of the existing units in 2016.

With a combination of the two datasets, the CEAP dataset provides nationwide, plant-level, dynamic PM, SO2 and NOX emissions from China’s thermal power plants from 2014 to 2017. Relative to existing inventories, the

| Year | Number of units | Number of plants | Total capacity (MW) | Unit average capacity (MW) |
|------|-----------------|------------------|---------------------|---------------------------|
| 2014 | 5,943           | 2,583            | 878,240             | 147.78                    |
| 2015 | 6,267           | 2,714            | 958,308             | 152.91                    |
| 2016 | 6,015           | 2,597            | 983,857             | 163.57                    |

Table 1. China’s thermal power plants in CEAP.
CEAP dataset is innovative in that it incorporates comprehensive real CEMS-measured emission data, avoiding the use of average emission factors and the associated operational assumptions and uncertain parameters.

**Pre-processing of CEMS data.** We have been exclusively granted access to the data from China’s CEMS network. Generally, the CEMS consists of a sampling system (for filtering and sampling flue gas), an online analytical component (for monitoring flue gas parameters, particularly emission concentrations) and a data processing system (for collecting, processing and reporting monitoring data)\(^{27,28}\). According to the GB13223-2003 regulation\(^{29}\), the CEMS network should cover all power plant furnaces that burn coal (except stoker and spreader stoker) and oil and generate >65 tons of steam each hour, as well as those that burn pulverized coal and gas. Thus, some power plants have not yet been incorporated into the CEMS network (accounting for 3–4% of the total thermal power capacity from 2014 to 2017) because their furnaces did not meet the requirements necessary to install a CEMS. For the power plants outside the CEMS network, we assume their stack concentrations are similar to the averages of the units with similar fuel types and similar regions within the CEMS network.

To guarantee the reliability of CEMS data, China’s government has made great efforts in developing specific regulations and technical guidelines for power plants and local entities to follow and supervise,
respectively. These official documents elaborate on all the processes required to regulate the CEMS network, including not only CEMS installation, operation, inspection, maintenance and repair but also CEMS data collection, processing, reporting, analysis and storage. Since 2014, all state-monitored companies have been mandated to report their CEMS data to the local governments through a series of online platforms for different provinces (listed in Supplementary Table 1). Local entities have random onsite inspections to check the truthfulness of the reported results on at least a quarterly basis; this system enables a comparison of CEMS data across different firms to explore potential outliers and abnormalities and prevent data manipulation. Then, the governments release the inspection results to the public through the same online platforms (listed in Supplementary Table 1). Severe financial penalties and criminal punishments can be imposed on firms that adopt data manipulation (in terms of deleting, distorting and forging CEMS data, for example).

The malfunction of CEMS monitors may also introduce large uncertainty to CEMS data during the processes of operation (indication errors, span drift, zero drift, etc.), maintenance (particularly the failure to perform calibration and reference tests) and data reporting (invalid data communication, data missing, etc.). Accordingly, each power plant is required to make at least one A-, B- and C-grade overhaul for 32–80, 14–50 and 9–30 days per 4–6, 2–3 and 1 year(s), respectively, as well as one D-grade overhaul (if needed) for 5–15 days per year, to check, maintain and upgrade its technologies, thereby reducing measurement uncertainty. During these overhauls, CEMS operators conduct CEMS calibration (i.e., zero and span calibration), maintenance procedures (e.g., examining and cleaning major CEMS components and replacing or upgrading parts, if necessary, such as optical lens, filter and sampling meter) and a reference test (i.e., relative accuracy test audit). Furthermore, third-party operators examine CEMS operation and maintenance routines, to guarantee standardized CEMS operation and facilitate improvement in CEMS data accuracy. All the related activities should be documented according to standardized requirement contents. Even with the aforementioned efforts, there is still a small proportion of missing data accounting for 34.62%, 31.91%, 29.97% and 42.96% of the total samples in 2014, 2015, 2016 and 2017, respectively. Severe financial penalties and criminal punishments can be imposed on firms that adopt data manipulation (in terms of deleting, distorting and forging CEMS data, for example).

The introduction of real CEMS-monitored measurements provides a direct estimation for emission factors on a source and hourly basis, avoiding the use of average emission factors with many assumptions and uncertain parameters. In Eq. (3), the emission factor, defined as the amount of emissions per unit of fuel use (in g kg⁻¹ for solid or liquid fuel and in g m⁻³ for gas fuel), and \( V_{iy, y} \) is the theoretical flue gas rate, defined as the expected volume of flue gas per unit of fuel use under standard production conditions (m³ kg⁻¹ for solid or liquid fuel and m³ m⁻³ for gas fuel), which was estimated by the China Pollution Source Census (2011) based on sufficient field measurements (detailed in Table 5). Based on Eq. (3), abated emission factors can be directly obtained even without the use of removal efficiencies and the relevant parameters, because CEMS monitors the gas concentrations at stacks after the effect of control equipment (if any).

Notably, recent clean air policies (particularly different emissions standards) target stack concentrations, such that a large proportion of missing data exist regarding other measurements (particularly flue gas rates, with missing data accounting for 34.62%, 31.91%, 29.97% and 42.96% of the total samples in 2014, 2015, 2016 and 2017, respectively). Accordingly, we introduce theoretical flue gas rates into the estimation to avoid significant underestimation of the actual volume when there are too many missing data values. In addition, the adoption of theoretical flue gas rates can address flue gas leakage, a common problem in power plants that greatly distorts...
the real flue gas volume. The theoretical flue gas rates are derived from the China Pollution Source Census, with values varying across operating capacities, fuel types and boiler types. Thus, the actual volume of flue gas is computed in terms of the theoretical flue gas rate times actual fuel consumption.

The absolute emissions of PM, SO2 and NOx from individual power plants can be estimated in terms of the emission factors times the activity levels:

\[ E_{i,y,m} = EF_{i,y,m} A_{i,y,m} \]  

where \( E_{i,y,m} \) represents the air pollution emissions (g); \( EF_{i,y,m} \) is the activity data, i.e., the amount of fuel use (kg for solid or liquid fuel and \( m^3 \) for gas fuel). In the CEAP dataset, power plant emissions are estimated on a monthly basis (the smallest scale for activity data), in which the yearly unit-level activity data are allocated at the monthly scale using the monthly province-level thermal power generation as weights:

\[ A_{i,y,m} = \frac{F_{p,y,m}}{\sum_{m=1}^{12} F_{p,y,m}} A_{i,y} \]

where \( F_{p,y,m} \) denotes the thermal power generation by province \( P \), which is obtained from the *Chinese Energy Statistics Yearbook*, and \( p_i \) indicates the province where unit \( i \) is located.

**Data Records**

A total of 12 data records (emissions and plant/unit information inventories) are contained in the CEAP dataset, which have been uploaded to public repository figshare. Of these:

- 4 are emission inventories for China’s power plants (2014–2017) [“CEAP-Absolute emissions, 2014–2017”];
- 4 are stack gas concentration inventories for China’s power plants (2014–2017) [“CEAP-Stack gas concentrations, 2014–2017”];
- 4 are summary descriptions for China’s power plants (2014–2017) [“CEAP-Summary descriptions, 2014–2017”].

The CEAP dataset introduces systematic real measurements by China’s CEMS network to directly estimate the PM, SO2 and NOx emissions from China’s power plants during 2014–2017 (Fig. 1). In particular, the dataset provides plant-level information about absolute emissions, fuel uses, generating capacities and geographic allocations for 2,583, 2,714, 2,596 and 2,596 power plants from 2014 to 2017, respectively. In addition, the CEAP dataset presents dynamic stack concentrations by region and fuel type and describes the overall structures of operating units, capacities, ages, emission factors, emissions and CEMS coverage for China’s thermal power plants.

**Technical Validation**

**Uncertainties.** The CEMS-based estimates are subject to uncertainties arising from volatilities in the CEMS data, the introduction of theoretical flue gas rates and the projection of activity data. Thus, uncertainty analyses are performed to verify the robustness of our estimates. Generally, the uncertainty analysis on each examined model variable or parameter (emission concentrations, theoretical flue gas rates or activity data) includes five main steps: (a) estimate the probability distributions by fitting data with an given distribution as the input of the Monte Carlo simulation; (b) generate random values based on the probability distributions via Monte Carlo simulation; (c) put the random values into Eqs. (3–5) to replace the original values and obtain a new set of estimates for emission factors and total emissions; (d) repeat steps (b) and (c) 10,000 times and obtain 10,000 sets of results; and (e) yield the uncertainty ranges of our estimates in terms of 2 standard deviations of the new 10,000 set of results. Table 6 reports the related results and reveals that the uncertainties can be controlled within a small range (i.e., ±9.03% and ±2.47% for emission factors and absolute emissions, respectively).

---

| Type | Descriptions | Treatment method | Supporting official documents |
|------|--------------|------------------|-----------------------------|
| 1    | Successive nulls for >5 days | Consider them as downtime for maintenance and omit them in emission. | a. According to the regulation, a power plant should go through at least one long maintenance shutdown per year, with one lasting for at least 5 days. |
| 2    | Successive nulls for 1–5 days | Assume them around the levels of valid values near the time (in terms of monthly averages). | b. The estimated downtime (corresponding to successive nulls for at least 5 days) accounted for 17.11% of the time for 2015, which are generally consistent with the official statistics (19.41%)(considering that 3–4% of plants do not have CEMS). |
| 3    | Nulls lasting for 1–24 hour(s) (involving non-successive nulls) | Set them to the arithmetic mean of the two nearest valid points before and after them. | Chinese government published Specifications for Continuous Emissions Monitoring of Flue Gas Emitted from Stationary Sources (HJ/T 75–2007). It suggests no interpolation for successive missing data of emission concentrations lasting for above 24 hours during operation, assuming them at similar levels to the points near the time and not to largely deviate from the average values. |

Table 4. Treatment methods for nulls and the relevant official documents.
| Fuel type     | Boiler                                                                 | Unit capacity (MW) | CPSC value a (m³ ton⁻¹) | CEMS-based value b (m³ ton⁻¹) | Uncertainty ranges b | t-test c |
|--------------|------------------------------------------------------------------------|--------------------|-------------------------|-----------------------------|----------------------|----------|
| Coal         | Pulverized coal-fired boiler                                           | ≥ 750              | 8,271                   | 8,376                       | ±5.73%               | P = 0.67, n = 48 |
|              |                                                                        | 450–749            | 10,150                  | 10,690                      | ±4.75%               | P = 0.04, n = 332 |
|              | Pulverized coal-fired boiler & Circulating fluidized bed boiler       | 250–449            | 9,713                   | 9,790                       | ±3.08%               | P = 0.62, n = 541 |
|              |                                                                        | 150–249            | 9,305                   | 9,806                       | ±5.61%               | P = 0.08, n = 113 |
|              |                                                                        | 75–149             | 8,178                   | 8,043                       | ±6.56%               | P = 0.62, n = 58  |
|              |                                                                        | 35–74              | 7,558                   | 8,030                       | ±6.87%               | P = 0.10, n = 57  |
|              |                                                                        | 20–34              | 7,729                   | 8,038                       | ±6.71%               | P = 0.27, n = 77  |
|              | Pulverized coal-fired boiler, circulating fluidized bed boiler & stoker-fired boiler | 9–19               | 7,958                   | 8,452                       | ±4.88%               | P = 0.02, n = 83  |
| Bituminous coal | Stoker-fired boiler                                                   |                   | 10,290                  |                             |                      |          |
|              | Pulverized coal-fired boiler                                           |                   | 9,186                   |                             |                      |          |
|              | Circulating fluidized bed boiler                                      |                   | 9,415                   |                             |                      |          |
| Anthracite   | Stoker-fired boiler                                                   |                   | 10,197                  | 13,494                      | ±9.21%               |          |
|              | Circulating fluidized bed boiler                                      |                   | 11,034                  |                             |                      |          |
| Lignite      | Stoker-fired boiler                                                   |                   | 5,915                   |                             |                      |          |
|              | Pulverized coal-fired boiler                                           |                   | 5,915                   |                             |                      |          |
| Coal gangue  | Circulating fluidized bed boiler                                      | —                  | 4,806                   | 6,718                       | ±10.07%              | P < 0.00, n = 43 |
| Solid waste  | Incinerator                                                           | —                  | 6,722                   |                             |                      |          |
| Solid waste & Coal | Incinerator                                                                | —                  | 7,774                   |                             |                      |          |
| Gas          | Turbine                                                               | —                  | 24.55                   | 24.9                        | ±9.91%               | P = 0.79, n = 21 |
| Oil          | Boiler & Turbine                                                      | —                  | 11,152                  |                             |                      |          |
| Petroleum coke | Circulating fluidized bed boiler                                     | —                  | 11,665                  |                             |                      |          |

Table 5. Theoretical flue gas rate. Notes: aThe values are derived from the China Pollution Source Census (CPSC) (2011) and used in our estimation; bThe results are estimated using the CEMS-monitored samples; cA single-sample two-tailed t-test is conducted for each type with the null hypothesis that the mean CEMS-monitored flue gas rates deviate from the CPSC value.

Uncertainties in CEMS data. The volatility in stack gas concentrations (the key model inputs in our estimation) should be considered in the uncertainty analysis. As the hourly CEMS measurements are recorded as an average over an hour time period, the associated volatility well reflects real variability in the emissions (as power demand rises and falls throughout the day, for example). We assume normal distributions for stack concentrations for each unit on a monthly basis and then draw the related parameters of distributions (e.g., the mean and the standard deviation) through data fitting based on the associated daily averages of the CEMS measurements. For a unit without CEMS, the bootstrap method is used to select samples from the units of the same fuel type and the same region in the CEMS network at an equal probability. Then, the Monte Carlo simulation is performed to generate random stack concentrations based on the associated distributions. With 10,000 simulations, the uncertainty ranges of the estimates are assessed to be small, i.e., ±8.65% and ±1.09% for the emission factors and absolute emissions, respectively.

Measurement uncertainties lead to a certain level of CEMS-monitored stack concentration deviations. According to the official regulation, a qualified CEMS instrument should control the error tolerance within ±15%, ±5% and ±5% for PM, SO₂ and NOₓ concentrations, respectively. Accordingly, we assume uniform distributions within the allowed tolerance ranges for all stack concentrations on the hourly, unit and pollutant basis. Then, random stack concentrations are generated using the Monte Carlo technique and put into Eq. (3) replacing the associated original values. A total of 10,000 simulations are run to estimate the uncertainty ranges of our estimates (in terms of 2 standard deviations). The results show that the final uncertainties fall within ±10.38% for emission factors and ±0.59% for total emissions.

Uncertainties in theoretical flue gas rates. Given that a large proportion of measurements of actual flue gas rates are missing in CEMS data (29.97–42.96% from 2014 to 2017), we introduce theoretical flue gas rates (fourth column of Table 5) in the estimation. Even though this method can prevent significant underestimations and flue gas leakage, uncertainties might arise due to the heterogeneity across units in factors such as technological, operational situations and feedstocks. We assess the uncertainty ranges of flue gas rates (defined as the lower and upper bounds of a 95% confidence interval around the central estimates) using the real samples in the CEMS database for 1,373 units that have different unit capacities, fuel types and boiler types and are
located throughout mainland China (fifth column). A single-sample two-tailed $t$-test is conducted, and the results (last column) indicate that the mean CEMS-monitored flue gas rates (fifth column) are at similar levels to the theoretical values that we used (fourth column). In the uncertainty analysis, Monte Carlo simulation is conducted to produce random flue gas rates following a uniform distribution on the associated uncertainty ranges. For the unit types without uncertainty ranges (e.g., those burning solid waste, oil and petroleum coke), the largest range (i.e., $\pm 10.07\%$) is employed. Relying on 10,000 simulations, the results show that uncertainty ranges can be well controlled within $\pm 6.90\%$ and $\pm 0.23\%$ for the emission factors and absolute emissions, respectively.

**Uncertainties in activity data.** The unit-specific activity data are available only up to 2016, and the 2017 values are projected using the monthly provincial data for 2017. This approach assumes that the growth rates in the activity levels of different units in a province are uniform from 2016 to 2017, which somewhat contradicts reality and brings about uncertainties. To assess such uncertainties, a bootstrap method is used to generate 10,000 samples of the growth rates from the previous values from 2014 to 2016, and statistical analysis is employed to fit these samples in a normal distribution. The Monte Carlo simulation is performed to generate random growth rates and thence the growth of activity levels from 2016 to 2017 for individual units, and the total provincial growth is allocated into each unit using the random growth as weights. With 10,000 simulations, the uncertainty range of total emissions is estimated to be quite small (within $\pm 0.03\%$).

**Comparison with existing databases.** We compare our estimates with existing databases, finding that our estimates of Chinese power emissions (using the real CEMS measurements for 2014–2017; purple bars in Fig. 2) are 18.62–91.86%, 54.98–69.77% and 17.55–67.76% below previous estimates (based on average emission factors that were evaluated up to 2012 without considering the recent mitigation effect particularly attributable to the ULE standards policy promulgated in 2014) for PM, SO$_2$ and NO$_X$, respectively. Furthermore, using the detailed measurements on the source and hour basis, the uncertainty of our estimates can be controlled at a relatively low level (error bars).

![Fig. 1 Estimated power emissions in China from 2014 to 2017. (a–c). Monthly estimates for the total and regional (coloured bars) emissions (Gg) of PM (a), SO$_2$ (b) and NO$_X$ (c) from Chinese power plants. The error bars indicate the uncertainty ranges.](image-url)
Limitations and future work. The CEAP dataset can be improved and extended from the following perspectives. First, some power plants have not yet been incorporated into the CEMS network, which account for an average of 3.8% of the total thermal capacity for 2014–2017. Therefore, collecting and incorporating these samples is needed to extend the CEAP dataset. Second, apart from air pollutants from power plants, the CEMS network monitors both air and water pollutants from various industries, totalling over 30,000 emission sources. Based on these data, the CEAP database can be extended into multisector datasets for both air and water pollutants in the future. Third, due to the data availability, the estimation does not use high-frequency information about activity data, such that CEMS data majorly drive the power emissions on a monthly scale. Future research involves incorporating hourly operational data (especially fuel consumption and flue gas rates) for each unit to improve the reliability of emissions estimates. Fourth, although great efforts have been made to guarantee the reliability of CEMS data, serious verification works (such as aerial concentration measurements) are still needed to check the data quality of the CEMS system41.

Code availability
There is no custom code in the generation of the CEAP dataset. In this study, Microsoft Excel is employed to process all the data and Origin is used to draw the figures. Three model inputs have been used in the construction of this dataset, i.e., the measurements from China’s CEMS network, theoretical flue gas rates and activity data. First, the CEMS-monitored data are released by the Ministry of Ecology and Environment of China through online platforms for different provinces, and we have documented all the links to these platforms in Supplementary Table 1. Second, theoretical flue gas rates are available in Table 5. Third, activity data are exclusively offered by the Ministry of Ecology and Environment of China.

Received: 13 January 2020; Accepted: 24 August 2020; Published online: 05 October 2020

References
1. British Petroleum. *Statistical Review of World Energy* (British Petroleum, 2019).
2. China Electricity Council. *China Power Industry Annual Development Report 2019* (China Market Press, 2019).
3. U.S. Environmental Protection Agency. Particulate Matter (PM) Pollution. [https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM](https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM).
44. U.S. Environmental Protection Agency. Plain English Guide to the Part 75 Rule. https://www.ecf.gov/cgibin/retrieveECFR?gp=&SID=73db81c17ec425ce1e9589b5930be&mc=true&n=pt40.18.75&r=PART&ty=HTML (2020).
45. China Pollution Source Census. Manual of the First National Pollution Source Census on Emission Factors from Industrial Pollution Sources (China Environmental Science Press, 2011).
46. Gilbert, A. Q. & Sovacool, B. K. Benchmarking natural gas and coal-fired electricity generation in the United States. Energy 134, 622–638 (2017).
47. Tang, L. et al. Air pollution emissions from Chinese power plants based on the continuous emission monitoring systems network. figshare https://doi.org/10.6084/m9.figshare.c.4813653.v3 (2020).
48. 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 (2011).
49. Zhao, Y., Zhou, Y. D., Qiu, L. P. & Zhang, J. Quantifying the uncertainties of China's emission inventory for industrial sources: From national to provincial and city scales. Atmos. Environ. 165, 207–221 (2017).
50. Frey, H. C. & Zheng, J. Quantification of variability and uncertainty in air pollutant emission inventories: method and case study for utility NOx emissions. J. Air Waste Manag. Assoc. 52, 1083–1095 (2002).
51. Streets, D. et al. An inventory of gaseous and primary aerosol emissions in Asia in the year 2000. J. Geophys. Res. Atmos. 108, 1984–2012 (2003).
52. Tang, L., Wu, J., Yu, L. & Bao, Q. Carbon emissions trading scheme exploration in China: A multi-agent-based model. Energy Policy 81, 152–169 (2015).
53. Li, M. & Patiño-Echeverri, D. Estimating benefits and costs of policies proposed in the 13th FYP to improve energy efficiency and reduce air emissions of China’s electric power sector. Energy Policy 111, 222–234 (2017).
54. Tong, D. et al. Current Emissions and Future Mitigation Pathways of Coal-Fired Power Plants in China from 2010 to 2030. Environ. Sci. Technol. 52, 11350–11357 (2018).
55. 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, 12905–12914 (2020).

Acknowledgements
This work was supported by grants from the National Science Foundation for Outstanding Young Scholars (71622011), the National Natural Science Foundation of China (71971007, 71988101 and 11771012), the National Programme for Support of Top Notch Young Professionals, the National Research Programme for Key Issues in Air Pollution Control (DQGG0209-07), the National Key Research and Development Program of China (2019YFE0194500) and the Appraisal Center for Environment and Engineering Ministry of Ecology and Environment (No. 2019-10). We thank Dr. Fahua Zhu (Engineer of State Power Environmental Protection Research Institute, China) for helpful discussions on data processes.

Author contributions
L.T., Z.M., X.B. and S.W. led the project. X.B. and S.L. compiled and processed CEMS data. G.D. compiled and processed the unit information. L.T., X.X., J.Q. and X.C. integrated the two databases and performed the experimental work. L.T., Z.M., X.B. and S.W. wrote the paper. All authors contributed to developing, writing and polishing the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-00665-1.

Correspondence and requests for materials should be addressed to Z.M. or X.B.

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 Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2020