scientific data

OPEN

DATA DESCRIPTOR

Global daily CO_2 emissions from 1970 to 2024

Tao Li[1], Lixing Wang[1], Zihan Qiu[2], Philippe Ciais[3], Steven J. Davis[4], Zhu Deng[5], Yufei Zhao[6], Glen P. Peters[7], Piyu Ke[1], Matthew W. Jones[8], Robbie M. Andrew[7], Ye Hao[9], Taochun Sun[1], Xiaoting Huang[1], Robert B. Jackson[4], Pierre Friedlingstein[10,11], Chenxi Lu[12], Duo Cui[13] & Zhu Liu[14,15,16] ✉

As extreme temperature events become increasingly frequent, there is a growing need for daily CO_2 emissions data to quantify their impacts. However, such data are available only from 2019 onward. To address this gap, we compiled over two million near-real-time observations of electricity generation, traffic activity, natural gas consumption or heating degree days (HDD), and industrial output since 2019, and used these high-frequency data to construct a daily CO_2 emissions dataset for 2019–2024. We then applied machine-learning models and degree-day methods to disaggregate non-residential and residential monthly CO_2 emissions for 1970–2018 to a daily basis. The historical dataset was then merged with the 2019–2024 dataset to produce a global daily CO_2 emissions dataset spanning 1970 to 2024 for 14 countries and regions, covering four sectors: power, industry, residential, and transport (including ground transport and aviation). The resulting long-term dataset will enable robust analyses of extreme-temperature impacts on emissions and enhance the accuracy of chemical transport model inversions of carbon fluxes.

Background & Summary

High temporal resolution emission data are essential for evaluating the impact of short-term temperature fluctuations, such as heat waves lasting several days to weeks, on CO_2 emissions. High temporal resolution estimates of CO_2 emissions provide a more detailed picture and offer insights into driving factors that annual data may overlook. Utilizing high temporal resolution CO_2 emission estimates, the effects of extreme temperature events, COVID-19, as well as blackouts on CO_2 emissions can be quantified, enhancing the understanding of emission change drivers[1]. High temporal resolution CO_2 emission inventories can be used in atmospheric inversion models to quantify the impact of extreme temperature events on natural carbon fluxes[2]. These assessments are vital for formulating policies aimed at achieving net-zero greenhouse gas emissions through forest carbon sinks.

Scale factors from the Temporal Improvements for Modeling Emissions by Scaling (TIMES)[3] or the Emissions Database for Global Atmospheric Research (EDGAR) temporal profiles library (hereafter "EDGAR_profile")[4] are commonly used to estimate daily emissions. International agencies and organizations such as the

[1]Department of Earth System Science, Tsinghua University, Beijing, 100084, China. [2]Tsinghua University Institute for Interdisciplinary Information Sciences (IIIS) FIT Building, Tsinghua University, Beijing, 100084, China. [3]Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CEA/CNRS/UVSQ, Université Paris-Saclay, Paris, 91191, France. [4]Department of Earth System Science, Stanford University, Stanford, CA, 94305, USA. [5]Department of Geography, The University of Hong Kong, Hong Kong SAR, 999077, China. [6]School of Earth Sciences, Zhejiang University, Hangzhou, 310058, China. [7]CICERO Center for International Climate Research, Oslo, 0349, Norway. [8]Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. [9]College of Land Science and Technology, China Agricultural University, Beijing, 100193, China. [10]Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK. [11]Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, CNRS, École Normale Supérieure, Université PSL, Sorbonne Université, École Polytechnique, Paris, France. [12]Mercator Research Institute on Global Commons and Climate Change, Berlin, Germany Sustainability Economics of Human Settlements, Technical University Berlin, Berlin, 10623, Germany. [13]Department of Earth Science, The University of Hong Kong, Hong Kong SAR, 999077, China. [14]International Research Center of Big Data for Sustainable Development Goals, Beijing, China. [15]Department of Earth System Science, Institute for Carbon Neutrality, State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, 100084, China. [16]Yellow River Laboratory (Henan), Zhengzhou, China. ✉e-mail: zhuliu@tsinghua.edu.cn

International Energy Association (IEA), Carbon Dioxide Information Analysis Center (CDIAC)[5], Open-source Data Inventory for Anthropogenic CO_2 (ODIAC)[6], EDGAR[7], Global Carbon Budget (GCB)[8–12], and National greenhouse gas inventories (NGHGIs) submitted to UNFCCC[13], typically provide fossil CO_2 emission data on annual or monthly basis. These data can be disaggregated into daily values using TIMES and EDGAR_profile. TIMES assigns fixed scale factors for different days of the week by country. For example, in the United States, the scale factors for Sunday, Saturday, and weekdays are 0.91, 0.96, and 1.02, respectively. These values indicate that emissions on Sundays and Saturdays are 91% and 96% of the average daily value in a week, while emissions on weekdays are 2% higher than the average. EDGAR_profile also provides similar scale factors for each day of week by sector. However, due to the downscaling measure from monthly or yearly dataset, such approaches are still unable to reflect the fine temporal resolution environmental influences such as temperature changes. The methods also often use generic scale factors across countries and years. For example, fluctuations in temperatures within a few days can alter heating requirements in winter[14] and cooling demands in summer[1], both of which substantially affect CO_2 emission levels but are difficult to be captured by TIMES or EDGAR_profile. Therefore, they cannot be used to assess the impact of short-term temperature events on fossil CO_2 emissions.

Recently, with various sources of activity data and satellite measurements, obtained at daily or even hourly or sub-hourly frequency, it has become possible to estimate CO_2 emissions with high temporal resolution. For example, researchers have estimated the decline in CO_2 attributable to COVID-19 based on forced confinement policies[15,16] and data on mobility changes[17]. Extending this further, with the high temporal resolution data collected and processed in low latency[18,19], Carbon Monitor used the amount and dynamics of activity in numerous sectors to construct the low-latency CO_2 emission estimates at high temporal resolution[20]. Satellite NO_2 measurements can be used to build top-down methods for estimating low latency CO_2 emissions for China[21,22] and constrain the CO_2 emission reduction from a bottom-up dataset[23–25]. These high temporal resolution datasets present a valuable opportunity to investigate the impact of extreme temperatures on fossil CO_2 emissions, but they are only available for years after 2019[26–31].

By collecting daily activity data since 2019 and incorporating the effects of short-term meteorological fluctuations on emissions through machine learning, this study develops a two-step method to construct a long-term daily CO_2 emissions dataset spanning from 1970 to 2024. In the first step, we use daily observations of electricity generation, industrial output, traffic counts, and natural gas consumption (or heating degree days, HDD) to disaggregate annual CO_2 emissions into daily values, generating a daily emissions dataset for 2019–2024. In the second step, this 2019–2024 daily dataset is used to train a machine-learning model that captures the nonlinear effects of meteorology on emissions and their interactions with temporal surrogate variables (e.g., day of week and month of year)[32], enabling the allocation of non-residential monthly CO_2 emissions for 1970–2018 onto a daily basis. Residential CO_2 emissions for historical years are disaggregated directly using HDD. Finally, the daily datasets for 1970–2018 and 2019–2024 are merged to produce the complete 1970–2024 daily CO_2 emissions time series. This dataset provides globally consistent daily CO_2 emissions data, addressing the current limitations in such high-resolution datasets. The long-term global daily CO_2 emissions dataset is valuable for assessing the impacts of extreme temperatures on CO_2 emissions and supporting policy-making. Additionally, it offers critical data for chemical transport models, which can be used for evaluating the effects of extreme temperatures on land carbon fluxes while fully accounting for fossil fuel emissions.

Method

We utilized a two-step method to build a long-term daily CO_2 emissions dataset for the world and 14 major regions. First, following the approach used in our previous publications[18,20], we disaggregate the annual emissions from EDGAR version 8.0 into daily values by using activity data with daily or near daily resolution[33]. This study incorporates newly obtained high frequency activity indicators across key sectors including electricity generation, industrial production, traffic congestion, aviation flight distances, and residential gas consumption. These updated activity records, collected from national statistical offices, energy system operators, transportation databases, and international industry associations, allow us to generate more accurate bottom-up daily emission estimates that capture real short-term variations driven by energy demand, mobility patterns, and industrial activity.

Second, we extend our daily CO_2 emissions dataset back to 1970, beyond the 2019 starting point of our previous work[18], by using either machine learning models or a degree day based approach depending on the sector. This approach enables the construction of long-term daily emission datasets even in the absence of high-resolution activity data over such extended periods. For the power, industry, and transportation sectors, we train country- and sector-specific machine learning models using 2019–2024 observations to learn how daily emission patterns respond to meteorology, weekly cycles, seasonal structure, and public holidays. These trained models are then applied to historical meteorology and calendar data to reproduce daily variability for 1970–2018. In contrast, the residential sector is reconstructed using population-weighted HDD rather than machine learning, reflecting the temperature-driven nature of residential emissions and ensuring methodological consistency across regions. The reconstructed daily variations for all sectors are then merged with the updated 2019–2024 estimates to produce the final global daily CO_2 emissions dataset spanning 1970–2024.

Data. *Annual and monthly CO_2 emission estimates.* We used the annual and monthly CO_2 emission estimates from EDGAR as the baseline for deriving daily CO_2 emissions[33]. EDGAR, developed by the Joint Research Centre of the European Commission and the Netherlands Environmental Assessment Agency, provides a comprehensive global inventory of anthropogenic greenhouse gases and air pollutants at both annual and monthly scales. In producing its monthly estimates, EDGAR uses monthly power generation data where available to distribute annual emissions for the power sector, and it applies traffic count data to allocate annual emissions to monthly values for

Sector	Region/Country	Previous data type	Previous data source	Current data type	Current data source
Power	China	Thermal production / coal consumption	China's National Bureau of Statistics (https://data.stats.gov); Zhedian Company	Thermal production / coal consumption	National Bureau of Statistics (https://data.stats.gov); China Coal Transportation and Distribution Association (https://www.cctd.com.cn)
	Mexico	Not Applied	Not Applied	Thermal production	Gobierno De Mexico (https://www.cenace.gob.mx/Paginas/SIM/Reportes/EnergiaGeneradaTipoTec.aspx)
	South Africa	Not Applied	Not Applied	Thermal production	Eskom (https://www.eskom.co.za/dataportal/supply-side/stationbuild-up-for-the-last-7-days/), Mexico)
	Australia	Not Applied	Not Applied	Thermal production	OpenNEM (https://opennem.org.au/energy/nem/?range=7d&interval=30m)
	Chile	Not Applied	Not Applied	Thermal production	Coordinador Eléctrico Nacional (https://www.coordinador.cl/operacion/graficos/operacion-real/generacion-real/)
Industry	World countries	Industrial Production Index	Trading Economics (https://tradingeconomics.com)	Industrial Production Index; Iron steel production	Trading Economics (https://tradingeconomics.com); World Steel Association (https://worldsteel.org/)
	India	Industrial Production Index	Ministry of Statistics and Programme Implementation (http://www.mospi.nic.in); Trading Economics (https://tradingeconomics.com)	Industry production index by product	Office of the Economic Advisor (https://eaindustry.nic.in/default.asp); Ministry of Statistics and Programme Implementation (http://www.mospi.nic.in); Trading Economics (https://tradingeconomics.com)
	European Union	Industrial Production Index	Eurostat (https://ec.europa.eu/eurostat)	Industry production index by product	Eurostat (https://ec.europa.eu/eurostat)
Residential	EU27 & UK	Heating Degree day	ECMWF Reanalysis v5 (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5)	Gas consumption	ENSTO-G (https://www.entsog.eu/)
Ground Transportation	United States	Prime supplier sales volumes of motor gasoline and diesel; Trips by Distance	EIA's Prime Supplier Report; Bureau of Transportation Statistics (https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv)	Congestion levels	Tomtom (https://www.tomtom.com/en_gb/traffic-index/)
Aviation	United States	Prime supplier sales volumes of kerosene-type jet fuel; Flight distances	EIA's Prime Supplier Report; Flightradar24 (https://www.flightradar24.com/)	Flight distances	Flightradar24 (https://www.flightradar24.com/)

Table 1. Summary of activity data updates for improving the estimates of daily sectoral CO_2 emissions.

the transport sector. These monthly inventories serve as the foundation upon which we reconstruct consistent daily CO_2 emissions across all regions and sectors.

Meteorology data. Meteorological variables at the daily scale, including surface air temperature at 2 meters, surface solar radiation downward, total cloud cover, and wind speed at 100 meters, were obtained from the fifth generation ECMWF reanalysis dataset (ERA5), which has an original spatial resolution of 0.25 degrees[34]. Surface air temperature is used to capture heating and cooling driven changes in energy demand, which directly influence fossil fuel emissions. In contrast, surface solar radiation downward, total cloud cover, and wind speed are included to reflect meteorological conditions that affect the availability of solar and wind power. Variations in renewable energy generation can offset part of the demand for fossil fuel-based electricity, and these meteorological factors therefore help represent the potential substitution effect of renewable energy on fossil energy use.

Population data. Population data are used to represent the temperature actually experienced by people rather than the geographic mean temperature of a country[18]. The population data were sourced from the first edition of the Global Population Count Grid Time Series Estimates[35] and the fourth edition of the Global Grid Population Dataset (GPWv4)[36]. The Global Population Count Grid Time Series Estimates provided population distribution data for 1970–2000 at 10-year intervals, while GPWv4 provided population distribution data for 2000–2020 at 5-year intervals.

Activity data. The activity data serve as indicators of daily emission variations and are used to disaggregate annual CO_2 emission estimates into daily values[18]. These activity indicators include electricity generation for 38 countries, traffic congestion levels across 416 cities, and industrial output data from 76 countries. Detailed descriptions of these data sources can be found in our previous publications[18–20,30,37–40]. In this study, we further update and expand several key datasets to improve the accuracy of daily CO_2 emission estimates. The major updates and new data sources are summarized in Table 1.

For the power sector, we used national daily heat production to estimate daily CO_2 emissions from electricity generation[19]. Compared to our previous publication in 2020[18], we have newly included power generation data sources for South Africa (https://www.eskom.co.za/dataportal/supply-side/stationbuild-up-for-the-last-7-days/), Mexico (https://www.cenace.gob.mx/Paginas/SIM/Reportes/EnergiaGeneradaTipoTec.aspx), Chile (https://www.coordinador.cl/operacion/grafcos/operacion-real/generacion-real/), and Australia (https://opennem.org.au/energy/nem/?range=7d&interval=30m). Previously, for China, the monthly power generation data by energy type from the National Bureau of Statistics of China was disaggregated to daily data using the daily coal consumption of Zhedian Company. Here, we instead used daily coal consumption data from the China Coal Transportation and Distribution Association (https://www.cctd.com.cn/) because it covers a broader range of power plants across 24 provinces (including eight coastal provinces and sixteen inland provinces) in China[41].

In the industrial sector, aside from the United States and China, we primarily use the industrial production index data as activity data to calculate monthly CO_2 emissions for the entire industry sector, which are then disaggregated to daily values using daily power generation data[18–20]. Here, we updated the estimation of CO_2 emissions for key industries, i.e., the iron and steel sector. For the iron and steel industry, we obtained monthly iron and steel production data from the World Steel Association (https://worldsteel.org/). The World Steel Association provides monthly global steel production updates, covering 85% of global steel production.

For industrial sector emissions in the European Union and India, we used data from separate sources. For the European Union (EU), we utilized production indexes for various products, such as essential oils, iron and steel, cement, and fertilizers, which were sourced from Eurostat (https://ec.europa.eu/eurostat). By utilizing the production index for each category of product, we estimated the monthly CO_2 emissions for each product category in accordance with Eq. (2). For India, we obtained production indexes for petroleum refinery products, cement, steel, and fertilizers from the Office of the Economic Advisor (https://eaindustry.nic.in/default.asp). These production indexes were used to estimate CO_2 emissions from these products. For CO_2 emissions from other industries (e.g., paper production) not specified above, the industrial production index the whole industrial sector was applied for estimation.

Except for EU's residential sector, the activity data sources for the ground transportation, and aviation sectors remained consistent with those in our previous publication[18–20,30,37–40]: traffic congestion levels from TomTom for ground transport sector and flight distances from FlightRadar24 for aviation sector, respectively. For EU's residential sector, the gas consumption records from ENSTO-G (https://www.entsog.eu/) were used as activity data[42].

The daily CO_2 emissions of the United States were estimated separately by state through a two-step process: 1) disaggregating annual state-level emissions into monthly estimates using monthly consumption data for key fuels, including motor gasoline, diesel, aviation fuel, and natural gas[37], and 2) further disaggregating these monthly estimates into daily emissions using state-level data on daily electricity generation for the power sector, daily distance traveled for the transport sector, HDD for residential and commercial sectors, and daily flight kilometers from FlightRadar24 for the aviation sector. However, EIA's Prime Supplier Report used for disaggregating annual CO_2 emissions of ground transport and aviation sectors to monthly ceased updates starting in October 2022 and has not resumed. Therefore, the TomTom congestion levels were used to disaggregate the annual CO_2 emissions from ground transport into daily values. Meanwhile, the daily flight distances from FlightRadar24 were used for the aviation sector.

Global daily CO_2 emissions from 2019 to 2024. We constructed global daily CO_2 emissions data for the period 2019 to 2024 within the framework of Carbon Monitor[18,20]. The CO_2 emissions are typically calculated using the following formula[43]:

$$E_{c,s} = AD_{c,s} \times EF_{c,s} \tag{1}$$

where $E_{c,s}$ represents CO_2 emissions of sector s in country c; $AD_{c,s}$ is the activity data of sector s in country c; $EF_{c,s}$ is the emission factor of sector s in country c. Equation (1) is included to provide the foundational emission-accounting principle recommended by the IPCC[44], which underpins all subsequent disaggregation and reconstruction steps in this study. We assume that the emission factor remains constant throughout the year. The daily CO_2 emissions were then estimated according to the following equations:

$$E_{c,s,y,i} = \frac{AD_{c,s,y,i}}{\sum_i AD_{c,s,y,i}} \times E_{c,s,y} \tag{2}$$

where $E_{c,s,y,i}$ is the CO_2 emissions of sector s in country c on the day i of year y; $E_{c,s,y}$ is the CO_2 emissions of sector s in country c for year y, sourced from version 8.0 of EDGAR[7]; $AD_{c,s,y,i}$ is the activity data of sector s in country c on the day i of year y.

Estimate global daily CO_2 emissions from 1970 to 2018. We applied machine learning to build the relationship between daily CO_2 emissions and predictor variables for reconstructing daily CO_2 emissions (excluding the residential sector) from 1970 to 2018. The residential CO_2 emissions were separately estimated using population-weighted HDD. This is because residential daily CO_2 emissions in 2019 to 2024 were already primarily derived from HDD in most regions, except for the EU27 and the United Kingdom[18]. Using a machine learning model for the residential sector would essentially emulate the behavior of a degree-day model, adding methodological complexity without providing any meaningful benefit. Although natural gas consumption data were used for the EU27 and the United Kingdom in the recent period, we still applied the degree-day model to derive residential daily emissions for all regions in the historical reconstruction. This choice maintains methodological

consistency across regions and avoids unnecessary complexity, ensuring a simple and coherent approach for reconstructing residential emissions over the full historical period.

Given the substantial differences in current CO_2 emission levels and the sector distribution compared to historical periods, using recent (from 2019 to present) daily CO_2 emissions directly as the response variable in models could lead to concept drift[45,46]. To mitigate this, we computed the ratio of daily CO_2 emissions to their monthly averages as the response variable. This emission ratio can reflect the daily CO_2 emission variations due to the effects of meteorology and public holidays. The formula used for this calculation is expressed as follows:

$$R_{c,s,y,m,i} = \frac{DE_{c,s,y,m,i,\,Carbon\ Monitor}}{MAE_{c,\,s,y,m,\,Carbon\ Monitor}} \tag{3}$$

$$MAE_{c,s,y,m,\,Carbon\ Monitor} = \frac{\sum_{i=1}^{N_m} DE_{c,s,y,m,i,\,Carbon\ Monitor}}{N_m} \tag{4}$$

where Daily Emissions, $DE_{c,s,y,m,i,\,Carbon\ Monitor}$, represents the CO_2 emissions of country c in sector s on day i of month m in year y estimated by Carbon Monitor; Monthly Average Emissions, $MAE_{c,s,y,m,\,Carbon\ Monitor}$, is monthly average CO_2 emission of country c in sector s for month m of year y estimated by Carbon Monitor; $R_{c,s,y,m,i}$ is emission ratio of country c in sector s on day i of month m in year y; N_m is the number of days in month m.

Based on the emission ratio, we built machine learning models to estimate the relationship between the emission ratio and predictor variables (i.e., meteorological factors and time surrogate variables). We applied eXtreme Gradient Boosting (XGBoost)[47] as the workhorse due to its good interpretability and performance advantages over neural networks and Random Forests when handling tabular data[48–52]. Furthermore, considering the variations in meteorology levels and public holidays between countries, we constructed models for each country or country group and each sector:

$$\hat{R}_{c,s,y,m,i} = f_{c,s}\left(Met_{c,y,m,i},\ X_{c,y,m,i}\right) \tag{5}$$

where $f_{c,s}$ is the XGBoost model for country c in sector s. The reconstruction was implemented for the power, industry, and transport sectors. The transport sector encompasses CO_2 emissions from ground transport, domestic aviation, and international aviation. $\hat{R}_{c,s,y,m,i}$ is the predicted emission ratio of country c in sector s on day i of month m in year y; $X_{c,y,m,i}$ are the time surrogate variables (i.e., day of week, day of month, month of year, and whether it is a public holiday). The dates of public holidays were obtained using the Python Holidays package. Specifically, in China, the actual Chinese New Year holiday typically extends beyond the officially published period[53]. Therefore, the days from the 23rd day of the 12th lunar month to the 15th day of the first lunar month were also considered public holidays. For all countries/regions, at least New Year's Day and Labor Day are universally recognized as public holidays, in addition to country-specific holidays such as national or independence days. Some religious holidays, like Christmas, are primarily observed in countries with corresponding religious affiliations, such as the United States and Europe. For some non-traditional holidays, whether they are considered public holidays varies by country; for example, Children's Day is a public holiday in Japan but not in China. Note that our predictive variables do not include measures of economic activity, such as GDP, because no high-temporal-resolution GDP dataset with long historical coverage is available. However, the omission of such variables is unlikely to introduce substantial bias. This is because the machine-learning model is designed to capture daily variations relative to the monthly mean, under the assumption that economic activity remains relatively stable within each month.

$Met_{c,y,m,i}$ represents the population- or area-weighted meteorological factors (i.e., temperature, solar radiation, cloud cover, and wind speed) of country c on day i of month m in year y, which is calculated according to following equation:

$$Met_{c,y,m,i} = \frac{\sum_j Met_{c,j,y,m,i} \times P_{c,j,y}}{\sum_j P_{c,j,y}} \tag{6}$$

where $Met_{c,j,y,m,i}$ represents the daily meteorology of j^{th} grid cell in country c on day i of month m in year y; $P_{c,j,y}$ represent the population count or area in the j^{th} grid cell of country c in year y. $P_{c,j,y}$ represents the area when the meteorological factor is solar radiation, cloud cover, or wind speed, and represents the population count otherwise. Solar radiation, cloud cover, and wind speed were used to represent the variation in renewable energy, which can be considered independent of population distribution. Therefore, these variables are directly weighted by area for averaging.

To compute population-weighted temperatures, we used the Global Population Count Grid Time Series estimates for the years 1970–1999, and GPWv4 for 2000–2024. Specifically, the population grid for 1970 is applied to the years 1970–1979, the grid for 1980 to 1980–1989, and so on; similarly, the GPWv4 2000 grid is applied to 2000–2004, the 2005 grid to 2005–2009, and so forth. Both population datasets have a spatial resolution of 30 arc-seconds, approximately 1 km at the equator. To ensure consistency with these population grids, we resampled the ERA5 surface temperature fields to the same 30 arc second resolution using bilinear interpolation. We

IPCC 1996 code	Description	Carbon Monitor
1A1a	Public electricity and heat production	Power
1A1bc	Other Energy Industries	Industry (incl. Cement Process)
1A2	Manufacturing Industries and Construction	Industry (incl. Cement Process)
2A1	Cement production	Industry (incl. Cement Process)
1A3a	Domestic aviation	Domestic aviation
1A3b	Road transportation no resuspension	Ground Transport
1A3c	Rail transportation	Ground Transport
1A3d	Inland navigation	Ground Transport
1A3e	Other transportation	Ground Transport
1A4	Residential and other sectors	Residential
1A5	Other Energy Industries	Residential
1C2	Memo: International navigation	International shipping
1C1	Memo: International aviation	International aviation

Table 2. The correspondence between EDGAR's emission categories, specified by the IPCC 1996 code, and Carbon Monitor's emission categories.

note that it appears that EDGAR's monthly CO_2 emissions estimates represent the sum of fossil and biogenic emissions.

Based on the reconstructed emission ratios and monthly average CO_2 emissions estimated by EDGAR for the period 1970–2018, we calculate the daily CO_2 emissions from 1970 to 2018:

$$DE_{c,s,y,m,i} = MAE_{c,s,y,m,EDGAR} \times \hat{R}_{c,s,y,m,i} \tag{7}$$

$$MAE_{c,s,y,m,EDGAR} = \frac{MTE_{c,s,y,m,EDGAR}}{N_m} \tag{8}$$

where $DE_{c,s,y,m,i}$ is the daily CO_2 emissions of country c in sector s on day i of month m in year y; $MAE_{c,s,y,m,EDGAR}$ is monthly average CO_2 emissions of country c in sector s for month m of year y estimated by EDGAR; $MTE_{c,s,y,m,EDGAR}$ is monthly total CO_2 emissions of country c in sector s for month m of year y estimated by EDGAR. To align with Carbon Monitor, only the CO_2 emissions from those emission categories in EDGAR that correspond to those in Carbon Monitor were used for calculating the monthly averages of CO_2 emissions. Please refer to Table 2 for the correspondence of emission categories between EDGAR and Carbon Monitor[54–56].

We applied both sample-based and time-based 10-fold cross-validation methods to evaluate the model's predictive performance. In the sample-based 10-fold cross-validation, the CO_2 emission dataset was randomly divided into ten approximately equal-sized groups without considering the temporal order of the data. In contrast, the time-based 10-fold cross-validation respects the temporal structure by dividing the dataset into ten approximately equal-sized groups based on time, with each group containing a continuous time period. For both methods, in each iteration, nine groups were used for training and the remaining group was used for prediction. The predicted and observed values were then compared using the R^2 metric to assess the model's predictive skill. This procedure was repeated ten times so that every group was used once for prediction. Since the remaining 10% of data points are not used in model training and are solely used for prediction, R^2 is calculated by comparing the predicted values with the corresponding observed values to evaluate the model's predictive performance. Therefore, this should be considered as out-of-sample cross-validation. The R^2 was calculated according to following equation:

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \overline{y})\cdot(\hat{y}_i - \overline{\hat{y}})\right)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2 \cdot \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})^2} \tag{9}$$

where n is the total number of samples; y_i is the observation of sample i; $\hat{y}_i$ is the model prediction of sample i; $\overline{y}$ is the average of observations; $\overline{\hat{y}}$ is the average of predictions. Note that grid search was used to optimize the hyperparameters of the XGBoost model[57], selecting hyperparameters with relatively higher predictive performance and lower computational cost[58,59]. For clarity, we note that while Eq. (9) is not the classical coefficient of determination, it is a widely used metric for both evaluating predictive performance and quantifying the similarity between time series[22,24,25,32,60,61], and is applied consistently throughout this study.

When the input meteorological variables include temperature, cloud cover, solar radiation, and wind speed, the average R^2 values from the time-based 10-fold cross-validation approach are 0.82 (power), 0.67 (industry), and 0.78 (transportation), which are slightly lower than the R^2 values from the sample-based 10-fold cross-validation results (0.84, 0.72, and 0.79 for power, industry, and transportation, respectively; Table 3). When temperature is the only meteorological input variable, the R^2 values from the time-based cross-validation are also slightly lower than those from the sample-based cross-validation, except for the power sector. This underscores the importance of considering the temporal ordering of data during model validation. Therefore, in this study,

Meteorology specification	Sector	R^2	
		Sample-based	Time-based
Only temperature	Power	0.79	0.80
	Industry	0.66	0.64
	Transportation	0.80	0.79
Temperature, cloud cover, solar radiation, and wind speed	Power	0.84	0.82
	Industry	0.72	0.67
	Transportation	0.79	0.78

Table 3. Average R^2 values for power, industry, and transportation sectors using both sample-based and time-based 10-fold cross-validation methods. The "Meteorology Specification" column indicates which meteorological variables were used as predictors to the model.

we applied the time-based 10-fold cross-validation method during grid search for hyperparameter optimization instead of the sample-based 10-fold cross-validation method. During the grid search for optimal hyperparameters, we found that the hyperparameter combination with the poorest performance resulted in time-based 10-fold cross-validation R^2 values of 0.72, 0.53, and 0.66 for the power, industry, and transportation sectors, respectively. In contrast, the best performing hyperparameter combination produced average R^2 values of 0.82, 0.67, and 0.78 for the same sectors. This underscores the importance of optimizing the hyperparameter combination.

For comparison, we also evaluated the performance of TIMES[3] and EDGAR_profile[4]. However, TIMES and EDGAR_profile do not build models to disaggregate monthly CO_2 emissions to a daily basis; instead, they directly use fixed scale factors. Therefore, we compared the daily CO_2 emissions obtained from TIMES and EDGAR_profile with those calculated based on activity data (e.g., electricity, industrial output, and traffic counts) to evaluate performance[18]. The daily CO_2 emissions derived from these activity data are closer to real daily CO_2 emissions, as they are not allocated by a fixed ratio based on the date, and thus serve as a benchmark for comparison. The R^2 values for power, transportation, and industry are 0.33, 0.32, and 0.13, respectively. We observe that these R^2 values are relatively low, which indicates that allocating emissions based on a fixed daily ratio does not capture the daily variations well, highlighting the value of high-temporal-resolution data for estimating daily CO_2 emissions. Moreover, the industrial sector has the lowest R^2, which is consistent with the cross-validation results and may be due to the complexity of process emissions in the industrial sector[43].

The residential CO_2 emissions from 1970 to 2018 was estimated according to following formula:

$$DE_{c,residential,y,m,i} = \frac{HDD_{c,y,m,i}}{\sum_i HDD_{c,y,m,i}} \times MTE_{c,residential,y,m,EDGAR} \times RH_{c,y,m} + \frac{MTE_{c,residential,y,m,EDGAR}}{N_m} \times (1 - RH_{c,y,m}) \quad (10)$$

$$HDD_{c,y,m,i} = \frac{\sum_j P_{c,j,y} \times (18 - Temp_{c,j,y,m,i}) \times I(Temp_{c,j,y,m,i} < 18)}{\sum_j P_{c,j,y}} \quad (11)$$

where $DE_{c,residential,y,m,i}$ is the daily residential CO_2 emissions of country c on day i of month m in year y; $RH_{c,y,m}$ is the percentage of residential emissions from heating demand of country c on month m in year y; $HDD_{c,y,m,i}$ is the HDD of country c on day i of month m in year y; $MTE_{c,residential,y,m,EDGAR}$ is monthly total residential CO_2 emissions of country c for month m of year y estimated by EDGAR; $Temp_{c,j,y,m,i}$ represents the daily temperature of j^{th} grid cell in country c on day i of month m in year y; $P_{c,j,y}$ represent the population count in the j^{th} grid cell of country c in year y; The indicator function $I(Temp_{c,j,y,m,i} < 18)$ equals to 1 when $Temp_{c,j,y,m,i} < 18$ otherwise 0.

Data Records

The reconstructed daily CO_2 emissions dataset in this study covers 14 countries/regions: the United States, EU27 & UK, China, Russia, India, Germany, Spain, France, the United Kingdom, Italy, Japan, Brazil, and rest of world (ROW). For each country/region, CO_2 emissions are provided for four sectors: power, industry, transportation (ground transportation and aviation), and residential. The dataset is published as a single CSV file on Figshare[62], containing four columns: country, sector, date, and value, where "value" represents the emissions.

Figure 1 illustrates daily CO_2 emissions from 1970 to 2024 for various countries and regions, including global, the United States, EU27 & the UK, China, India, Japan, Brazil, Germany, Italy, the United Kingdom, Spain, France, Russia, and ROW. Figure 2 shows the global daily CO_2 emissions for the four sectors from 1970 to 2024.

Technical Validations

Comparing daily CO_2 emissions with other datasets. We compared our daily CO_2 emissions with two emission distribution methods, TIMES[3] and EDGAR_ profile[4]. TIMES and EDGAR_profile provide scale factors that reflect daily variations in CO_2 emissions. These scale factors were used to convert EDGAR's monthly CO_2 emission estimates into daily values for further comparison.

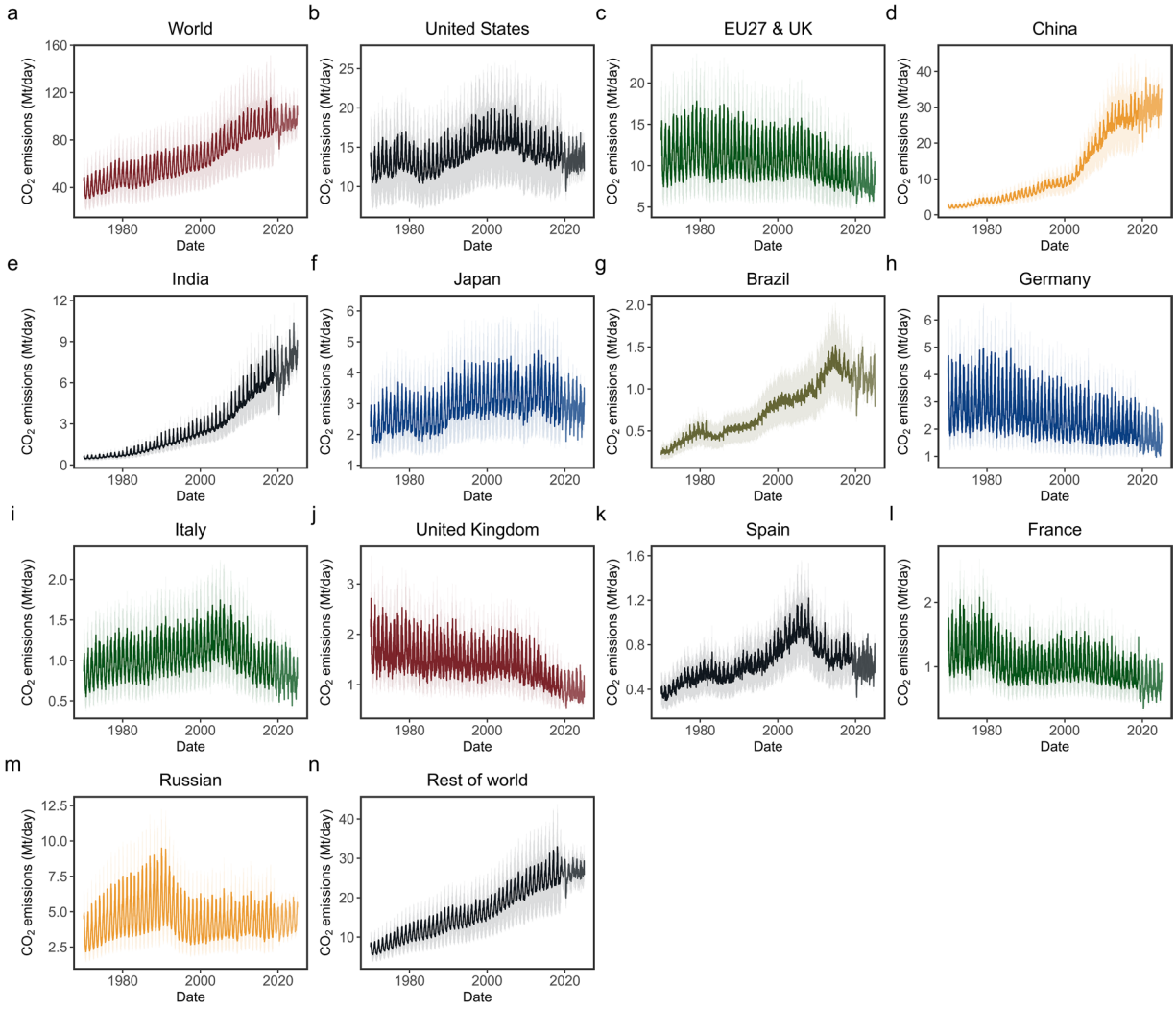

Fig. 1 Global daily CO_2 emissions from 1970 to 2024.

Figure 3 illustrates the agreement between daily CO_2 emissions estimated in this study and those derived from EDGAR_profile. The R^2 across all evaluated countries/regions ranges from 0.71 to 0.98, with a mean value of 0.88, indicating strong linear consistency between the two datasets. The highest R^2 value is observed for China, India, and the ROW aggregate, reflecting nearly perfect alignment in these cases. Conversely, the United Kingdom and the United States exhibits the lowest R^2 (0.71 for the United Kingdom and 0.76 for the United States), suggesting comparatively higher variability in emission estimates. In terms of the symmetric mean absolute percentage error (SMAPE), values span from 3.8% to 12%, with a mean of 7.2%, underscoring the robustness of the proposed methodology. The lowest SMAPE (3.8%) is achieved by the world aggregate, while the highest discrepancy (12%) occurs in Germany, likely attributable to regional data granularity or sectoral emission heterogeneity.

Table 4 compares daily CO_2 emissions from this study against estimates generated by the TIMES. The R^2 values range from 0.77 to 0.99, with a mean of 0.88, demonstrating excellent correlation across most regions. The highest R^2 is shared by China, India, ROW and the world aggregate. France again shows the lowest R^2 (0.77), highlighting potential discrepancies in temporal or sectoral allocation. SMAPE metrics vary between 2.8% and 12%, averaging 6.8%, which further validates the precision. The smallest SMAPE (2.8%) corresponds to world emissions, whereas the largest error (12%) persists in Germany, consistent with challenges noted in the comparison with EDGAR_profile.

Table 5 presents the comparative performance metrics (R^2 and SMAPE) between the daily sectoral CO_2 emissions estimated in this study and those derived by EDGAR_profile across various countries/regions and sectors. The R^2 values demonstrate strong consistency between the two datasets, with sector-specific ranges from 0.57–0.95 (industry), 0.70–0.97 (power), 0.65–0.93 (residential), and 0.83–0.98 (transportation). The highest R^2 values were observed in India's power sector (0.97), China's industry sector (0.95), and the transportation sector of China (0.98). Conversely, the lowest R^2 occurred in the Spain's industry (0.57), Brazil's residential (0.65), and Germany's transportation (0.84) sectors. The mean R^2 across all sectors was 0.83 (industry), 0.88 (power), 0.83 (residential), and 0.90 (transportation), indicating robust correlations.

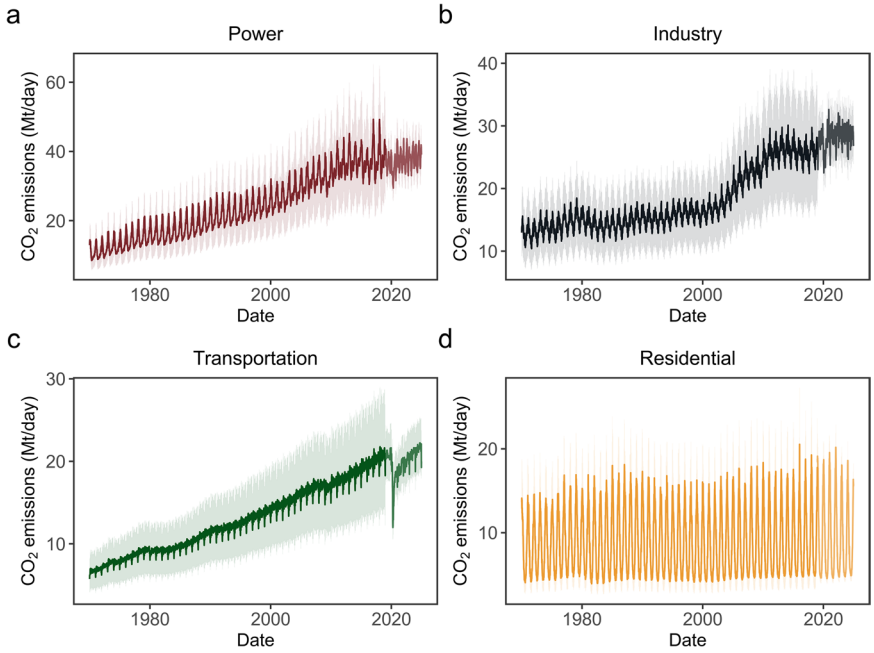

Fig. 2 Global daily CO_2 emissions for power, industry, transportation, and residential sectors from 1970 to 2024.

For error assessment, SMAPE ranged between 6.0%–17% (industry), 4.3%–18% (power), 11%–34% (residential), and 3.9%–14% (transportation). The lowest SMAPE values were achieved in ROW's industry (6.0%), ROW's power (4.3%), the residential sectors of Global and China (11%), and the transportation sectors of Global and Japan (3.9%). In contrast, the highest SMAPE corresponded to France's industry (17%), United Kingdom's power (18%), India's residential (34%), and Spain's transportation (14%). Sectoral SMAPE averages were 10% (industry), 9.3% (power), 18% (residential), and 6.3% (transportation), reflecting relatively higher uncertainties in residential emissions.

Uncertainties and limitations. Using error propagation rules and the IPCC 2006 uncertainty analysis method[44], the uncertainty was assessed for each sector and then aggregated to determine the total uncertainty using the following equation:

$$U_{c,y,m,i} = \frac{\sqrt{\sum_s (U_{c,s,y,m,i} \times DE_{c,s,y,m,i})^2}}{\sum_s DE_{c,s,y,m,i}} \tag{12}$$

where $U_{c,y,m,i}$ is the total uncertainty of country c on day i of month m in year y; $U_{c,s,y,m,i}$ represents the uncertainty of country c in sector s on day i of month m in year y. The uncertainty $U_{c,s,y,m,i}$ was evaluated separately for the periods 2019–2024 and 1970–2018.

From 2019 to 2024, uncertainty in the power sector primarily arises from inter-annual variability in coal emission factors and fluctuations in the fuel mix for thermal generation, with an uncertainty of ±14%. In the industrial sector, emissions were calculated from monthly data, and the uncertainty was assessed using Monte Carlo simulations, yielding a 68% confidence interval. For the ground transportation sector, emissions are estimated by assuming a similar relationship between car counts (and emissions) and the TomTom congestion index in Paris. The uncertainty in these estimates is quantified through the prediction interval from regression analysis. In the aviation sector, uncertainty is calculated as the average difference between daily emissions, based on flight route distances and the number of flights. For the residential sector, uncertainty is estimated by comparing daily emissions from actual fuel consumption data across several European countries, including France, Great Britain, Italy, Belgium, and Spain.

For the period 1970 to 2018, the uncertainty for each sector was determined with the following formula:

$$U_{c,s,y,m,i} = \sqrt{U^2_{EDGAR,c,s,y,m} + U^2_{R,c,s,y,m,i}} \tag{13}$$

where $U_{c,s,y,m,i}$ is the uncertainty for reconstructed daily CO_2 emissions of country c in sector s on day i of month m in year y; $U_{EDGAR,c,s,y,m}$ is the uncertainty for EDGAR monthly CO_2 emissions of country c in sector s on month m in year y; $U_{R,c,s,y,m,i}$ is the uncertainty for emission ratio of country c in sector s on day i of month m in year y.

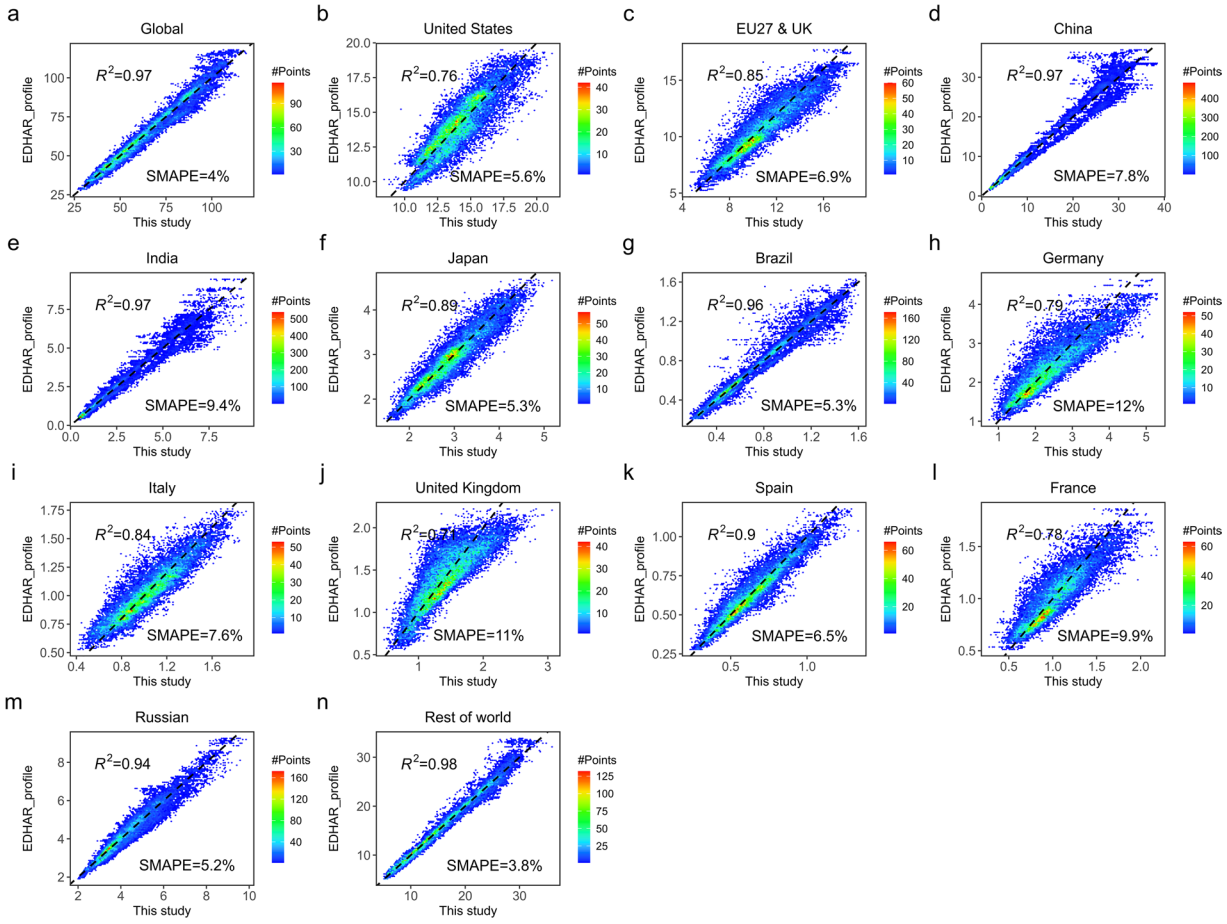

Fig. 3 Comparison between daily CO_2 emissions built in this study and those estimated according to EDGAR_profile. SMAPE: symmetric mean absolute percentage error.

Country/Region	R^2	SMAPE (%)
World	0.98	2.8
United States	0.83	4.1
China	0.98	5.6
EU27 & UK	0.84	7.7
India	0.98	7.7
Japan	0.90	5.0
Brazil	0.97	4.2
Germany	0.77	12
Italy	0.81	8.6
United Kingdom	0.70	12
Spain	0.88	7.6
France	0.77	10
Russia	0.96	4.0
Rest of world	0.99	3.1

Table 4. Comparison between daily CO_2 emissions built in this study and those estimated according to TIMES. SMAPE: symmetric mean absolute percentage error.

We employed the quantile regression[63] technique to quantify $U_{R,c,s,y,m,i}$. In quantile regression, the loss function $L(y, \hat{y})$ is expressed as follows:

$$L(y_i, \hat{y}_i) = q \times max^2(0, y_i - \hat{y}_i) + (1 - q) \times max^2(0, \hat{y}_i - y_i) \quad (14)$$

Country/Region	Industry	Power	Residential	Transportation
World	0.91/6.7	0.97/4.5	0.93/11	0.97/3.9
United States	0.63/8.2	0.86/7.5	0.85/12	0.85/4.5
China	0.95/11	0.97/8.1	0.92/11	0.98/4.4
EU27 & UK	0.89/8.7	0.85/8.3	0.86/19	0.88/6.7
Germany	0.84/15	0.71/16	0.81/27	0.83/6.5
Spain	0.57/9.1	0.91/8.7	0.83/15	0.84/14
France	0.82/16	0.87/18	0.78/22	0.88/6.3
United Kingdom	0.77/17	0.70/18	0.74/19	0.84/5.6
Brazil	0.92/8.6	0.83/8.3	0.65/12	0.92/8.5
India	0.95/11	0.97/7.5	0.72/34	0.96/5.3
Italy	0.89/9.9	0.85/10	0.86/17	0.88/7.8
Japan	0.80/7.0	0.93/5.8	0.89/16	0.95/3.9
Russia	0.76/9.0	0.94/6.2	0.93/17	0.84/6.0
Rest of world	0.94/6.0	0.97/4.3	0.89/15	0.97/5.1

Table 5. R^2/SMAPE between daily sectoral CO_2 emissions built in this study and those estimated according to EDGAR_profile. SMAPE: symmetric mean absolute percentage error.

where q indicates the quantile. In this study, q was set to 0.16 and 0.84 to obtain the 16th and 84th percentiles, respectively, forming the [16%, 84%] confidence interval for the predictions. These values were selected to ideally encompass 68% of the records within the prediction interval, corresponding to one standard deviation, assuming a Gaussian error distribution[63]. Subsequently, the U_R was obtained according to following equation:

$$U_R = \frac{(\hat{y}_{i,upper} - \hat{y}_{i,lower})}{2 \times \hat{y}_{i,middle}} \tag{15}$$

where U_R is uncertainty for emission ratio; $\hat{y}_{i,lower}$, $\hat{y}_{i,middle}$, and $\hat{y}_{i,lower}$ are values predicted for y_i where q is set to 0.16, 0.5, and 0.84, respectively.

The uncertainty of global daily CO_2 emission was within ±33% (1970–2018) and ±7.2% (2019–2024) based on the rule of error propagation and IPCC 2006 uncertainty analysis[44]. For the period 1970–2018, the main source of error arises from the EDGAR monthly CO_2 emissions, which are obtained using various proxies such as the HDD, but also survey data that has been extrapolated beyond its original geographical scope[4]. While HDD can explain emission changes in the residential sector, it is not applicable when biomass is primarily used for heating. A clear example is India, where fossil CO_2 emissions in the residential sector primarily come from cooking with petroleum-based fuels, while most household heating emissions originate from biomass combustion and therefore do not lead to fossil CO_2 emissions. Therefore, HDD is unsuitable for indicating emission variations from cooking. Future work should focus on generating improved sub-annual profiles that distinguish between fossil and biogenic CO_2 emissions. Furthermore, the uncertainty propagation in Eqs. (12), (13) assumes independence between uncertainty sources across different sectors due to the considerable challenge of quantifying potential correlations. Future research could aim to characterize these relationships for a more comprehensive uncertainty assessment.

Because the uncertainty in daily CO_2 emissions primarily arises from the monthly CO_2 emissions provided by EDGAR, the uncertainty in daily CO_2 emissions derived from TIMES and EDGAR_profile is also as high as 33%. The uncertainty of the scale factors for TIMES and EDGAR_profile is measured by the standard deviation[3], which is 1.4% for TIMES and 1.6% for EDGAR_profile. When these uncertainties are combined with the 33% uncertainty from EDGAR's monthly CO_2 emissions (as shown in Eq. (13)), the total uncertainty, to two significant digits, remains at 33%. This is due to the lack of detailed activity data, such as electricity generation, which leads EDGAR's monthly CO_2 emission estimates to be assigned a relatively conservative uncertainty value of 33%[4]. However, with more detailed activity data, such as that used by Carbon Monitor (covering electricity, industrial production, and natural gas consumption), the uncertainty in daily CO_2 emissions estimates could be reduced to 7.2%[18]. As more historical activity data become available in the future, the uncertainty in the monthly CO_2 emissions estimates is expected to decrease, which would, in turn, reduce the uncertainty in the reconstructed daily CO_2 emissions.

In addition, we only used meteorology and time variables for reconstructing fossil CO_2 emissions over the period from 1970 to 2018. Despite the good predictive performance achieved, applying the emission-meteorology and temporal relationships learned from the 2019–2024 period directly to earlier decades may introduce biases. The energy structure in the past differed substantially from today, especially with respect to the penetration of renewable energy[26]. As a result, the response of fossil energy consumption to temperature variations in earlier decades may not fully match the patterns observed in 2019–2024. This limitation is shared by other temporal disaggregation approaches, including TIMES[3] and EDGAR_profile[4], and represents an important source of uncertainty in the historical daily CO_2 emissions reconstructed in this study. In the future, the availability of more high-resolution energy consumption data, additional environmental indicators such as atmospheric NO_2 and $PM_{2.5}$ concentrations, and more high-temporal-resolution proxies for economic activity

(for example, stock indices) will help improve emissions estimation and enable more accurate reconstruction of high-frequency emissions time series.

This study does not directly provide daily CO_2 emissions at the gridded level, which would be required as input for chemical transport models. However, using existing monthly gridded CO_2 emissions datasets such as those from EDGAR[4,7], it is straightforward to generate gridded daily CO_2 emissions by applying the following two equations:

$$R_{c,s,y,m,i} = \frac{DE_{c,s,y,m,i}}{MTE_{c,s,y,m}} \tag{16}$$

$$DE_{c,j,s,y,m,i} = MTE_{c,j,s,y,m,EDGAR} \times R_{c,s,y,m,i} \tag{17}$$

where Daily Emissions, $DE_{c,s,y,m,i}$, represents the CO_2 emissions of country c in sector s on day i of month m in year y constructed in this study; Monthly Total Emissions, $MTE_{c,s,y,m}$, is monthly total CO_2 emission of country c in sector s for month m of year y constructed in this study; $MTE_{c,j,s,y,m,EDGAR}$ denotes monthly total CO_2 emissions of grid j in country c and sector s for month m of year y estimated by EDGAR; $DE_{c,j,s,y,m,i}$ represents the disaggregated CO_2 emissions of grid j in country c and sector s on day i of month m in year y.

Data availability

The dataset generated by the current study is available on Figshare at https://doi.org/10.6084/m9.figshare.28827932.

Code availability

This dataset was generated using R 4.4.2 and Python 3.8, and the code can be accessed on GitHub (https://github.com/PowderL/Global_daily_CO2_emissions) without any access restrictions.

Received: 12 May 2025; Accepted: 12 January 2026;
Published online: 24 February 2026

References

1. Zhao, W. *et al.* Reliance on fossil fuels increases during extreme temperature events in the continental United States. *Commun. Earth Environ.* **4**, 473, https://doi.org/10.1038/s43247-023-01147-z (2023).
2. van der Woude, A. M. *et al.* Temperature extremes of 2022 reduced carbon uptake by forests in Europe. *Nat. Commun.* **14**, 6218, https://doi.org/10.1038/s41467-023-41851-0 (2023).
3. Nassar, R. *et al.* Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets. *J. Geophys. Res.: Atmos.* **118**, 917–933, https://doi.org/10.1029/2012JD018196 (2013).
4. Crippa, M. *et al.* High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. *Sci. Data* **7**, 121, https://doi.org/10.1038/s41597-020-0462-2 (2020).
5. Gilfillan, D. & Marland, G. CDIAC-FF: global and national CO_2 emissions from fossil fuel combustion and cement manufacture: 1751–2017. *Earth Syst. Sci. Data* **13**, 1667–1680, https://doi.org/10.5194/essd-13-1667-2021 (2021).
6. Oda, T., Maksyutov, S. & Andres, R. J. The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): a global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions. *Earth Syst. Sci. Data* **10**, 87–107, https://doi.org/10.5194/essd-10-87-2018 (2018).
7. Crippa, M. *et al.* Insights into the spatial distribution of global, national, and subnational greenhouse gas emissions in the Emissions Database for Global Atmospheric Research (EDGAR v8.0). *Earth Syst. Sci. Data* **16**, 2811–2830, https://doi.org/10.5194/essd-16-2811-2024 (2024).
8. Friedlingstein, P. *et al.* Global Carbon Budget 2024. *Earth Syst. Sci. Data* **17**, 965–1039, https://doi.org/10.5194/essd-17-965-2025 (2025).
9. Friedlingstein, P. *et al.* Global Carbon Budget 2023. *Earth Syst. Sci. Data* **15**, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023 (2023).
10. Friedlingstein, P. *et al.* Global Carbon Budget 2022. *Earth Syst. Sci. Data* **14**, 4811–4900, https://doi.org/10.5194/essd-14-4811-2022 (2022).
11. Friedlingstein, P. *et al.* Global Carbon Budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005, https://doi.org/10.5194/essd-14-1917-2022 (2022).
12. Friedlingstein, P. *et al.* Global Carbon Budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340, https://doi.org/10.5194/essd-12-3269-2020 (2020).
13. Deng, Z. *et al.* Global greenhouse gas reconciliation 2022. *Earth Syst. Sci. Data* **17**, 1121–1152, https://doi.org/10.5194/essd-17-1121-2025 (2025).
14. Ciais, P. *et al.* Impact of Lockdowns and Winter Temperatures on Natural Gas Consumption in Europe. *Earth's Future* **10**, e2021EF002250, https://doi.org/10.1029/2021EF002250 (2022).
15. Le Quéré, C. *et al.* Fossil CO_2 emissions in the post-COVID-19 era. *Nat. Clim. Change* **11**, 197–199, https://doi.org/10.1038/s41558-021-01001-0 (2021).
16. Le Quéré, C. *et al.* Temporary reduction in daily global CO_2 emissions during the COVID-19 forced confinement. *Nat. Clim. Change* **10**, 647–653, https://doi.org/10.1038/s41558-020-0797-x (2020).
17. Forster, P. M. *et al.* Current and future global climate impacts resulting from COVID-19. *Nat. Clim. Change* **10**, 913–919, https://doi.org/10.1038/s41558-020-0883-0 (2020).
18. Liu, Z. *et al.* Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* **11**, 1–12, https://doi.org/10.1038/s41467-020-18922-7 (2020).
19. Liu, Z. *et al.* Global patterns of daily CO_2 emissions reductions in the first year of COVID-19. *Nat. Geosci.* **15**, 615–620, https://doi.org/10.1038/s41561-022-00965-8 (2022).
20. Liu, Z. *et al.* Carbon Monitor, a near-real-time daily dataset of global CO_2 emission from fossil fuel and cement production. *Sci. Data* **7**, 392, https://doi.org/10.1038/s41597-020-00708-7 (2020).

21. Zheng, B. *et al*. Satellite-based estimates of decline and rebound in China's CO_2 emissions during COVID-19 pandemic. *Sci. Adv.* **6**, eabd4998, https://doi.org/10.1126/sciadv.abd4998 (2020).
22. Wu, Q. *et al*. Evaluation of NOx emissions before, during, and after the COVID-19 lockdowns in China: A comparison of meteorological normalization methods. *Atmos. Environ.* **278**, 119083, https://doi.org/10.1016/j.atmosenv.2022.119083 (2022).
23. Weir, B. *et al*. Regional impacts of COVID-19 on carbon dioxide detected worldwide from space. *Sci. Adv.* **7**, eabf9415, https://doi.org/10.1126/sciadv.abf9415 (2022).
24. Li, T. *et al*. Enhancing Space-Based Tracking of Fossil Fuel CO_2 Emissions via Synergistic Integration of OCO-2, OCO-3, and TROPOMI Measurements. *Environ. Sci. Technol.* **59**, 1587–1597, https://doi.org/10.1021/acs.est.4c05896 (2025).
25. He, C. *et al*. Deriving full-coverage and fine-scale XCO_2 across China based on OCO-2 satellite retrievals and CarbonTracker output. *Geophys. Res. Lett.* **49**, e2022GL098435, https://doi.org/10.1029/2022GL098435 (2022).
26. Deng, Z. *et al*. Global carbon emissions and decarbonization in 2024. *Nat. Rev. Earth Environ.* **6**, 231–233, https://doi.org/10.1038/s43017-025-00658-x (2025).
27. Liu, Z., Deng, Z., Davis, S. J. & Ciais, P. Global carbon emissions in 2023. *Nat. Rev. Earth Environ.* **5**, 253–254, https://doi.org/10.1038/s43017-024-00532-2 (2024).
28. Liu, Z., Deng, Z., Davis, S. & Ciais, P. Monitoring global carbon emissions in 2022. *Nat. Rev. Earth Environ.* **4**, 205–206, https://doi.org/10.1038/s43017-023-00406-z (2023).
29. Liu, Z., Deng, Z., Davis, S. J., Giron, C. & Ciais, P. Monitoring global carbon emissions in 2021. *Nat. Rev. Earth Environ.* **3**, 217–219, https://doi.org/10.1038/s43017-022-00285-w (2022).
30. Ke, P. *et al*. Carbon Monitor Europe near-real-time daily CO_2 emissions for 27 EU countries and the United Kingdom. *Sci. Data* **10**, 374, https://doi.org/10.1038/s41597-023-02284-y (2023).
31. Ke, P. *et al*. Low latency carbon budget analysis reveals a large decline of the land carbon sink in 2023. *Natl. Sci. Rev.*, nwae367, https://doi.org/10.1093/nsr/nwae367 (2024).
32. Li, T., Lu, Y., Deng, X. & Zhan, Y. Spatiotemporal variations in meteorological influences on ambient ozone in China: A machine learning approach. *Atmos. Pollut. Res.* **14**, 101720, https://doi.org/10.1016/j.apr.2023.101720 (2023).
33. Crippa, M. *et al*. Fossil CO_2 and GHG emissions of all world countries. *Publication Office of the European Union: Luxemburg*, 1-251 (2019).
34. Hersbach, H. *et al*. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049, https://doi.org/10.1002/qj.3803 (2020).
35. CIESIN. Global Population Count Grid Time Series Estimates, https://doi.org/10.7927/H4CC0XNV (2017).
36. CIESIN. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11, https://doi.org/10.7927/H49C6VHW (2020).
37. Hong, C. *et al*. Near-real-time and state-level monitoring of US CO_2 emissions (2021).
38. Wu, S. *et al*. Global CO_2 uptake by cement materials accounts 1930–2023. *Sci. Data* **11**, 1409, https://doi.org/10.1038/s41597-024-04234-8 (2024).
39. Lu, C. *et al*. Health co-benefits of post-COVID-19 low-carbon recovery in Chinese cities. *Nat. Cities* **1**, 695–705, https://doi.org/10.1038/s44284-024-00115-8 (2024).
40. Shi, Q. *et al*. Comparison of High-Resolution Gridded Emission Maps of Anthropogenic Carbon Dioxide in Europe: GRACED & CAMS-REG. *Environ. Sci. Technol.* **59**, 4926–4937, https://doi.org/10.1021/acs.est.4c07289 (2025).
41. Zhu, B. *et al*. CarbonMonitor-Power near-real-time monitoring of global power generation on hourly to daily scales. *Sci. Data* **10**, 217, https://doi.org/10.1038/s41597-023-02094-2 (2023).
42. Zhou, C. *et al*. Natural gas supply from Russia derived from daily pipeline flow data and potential solutions for filling a shortage of Russian supply in the European Union (EU). *Earth Syst. Sci. Data* **15**, 949–961, https://doi.org/10.5194/essd-15-949-2023 (2023).
43. 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).
44. Eggleston, H., Buendia, L., Miwa, K., Ngara, T. & Tanabe, K. 2006 IPCC guidelines for national greenhouse gas inventories (2006).
45. Wu, Y. *et al*. A robust approach to deriving long-term daily surface NO_2 levels across China: Correction to substantial estimation bias in back-extrapolation. *Environ. Int.* **154**, 106576, https://doi.org/10.1016/j.envint.2021.106576 (2021).
46. Zhang, S. *et al*. A data-augmentation approach to deriving long-term surface SO_2 across Northern China: Implications for interpretable machine learning. *Sci. Total Environ.* **827**, 154278, https://doi.org/10.1016/j.scitotenv.2022.154278 (2022).
47. Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).
48. Shwartz-Ziv, R. & Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **81**, 84–90, https://doi.org/10.1016/j.inffus.2021.11.011 (2022).
49. Pan, B. Application of XGBoost algorithm in hourly $PM_{2.5}$ concentration prediction. *IOP Conf. Ser.: Earth Environ. Sci.* **113**, 012127, https://doi.org/10.1088/1755-1315/113/1/012127 (2018).
50. Zhou, Z. & Feng, J. Deep forest. *Natl. Sci. Rev.* **6**, 74–86, https://doi.org/10.1093/nsr/nwy108 (2019).
51. Liu, X. *et al*. Interpretable regional meteorological feature extraction enhances deep learning for extended 120-h $PM_{2.5}$ forecasting. *J. Cleaner Prod.* **483**, 144287, https://doi.org/10.1016/j.jclepro.2024.144287 (2024).
52. Huang, X. *et al*. Improved Consistency of Satellite XCO_2 Retrievals Based on Machine Learning. *Geophys. Res. Lett.* **51**, e2023GL107536, https://doi.org/10.1029/2023GL107536 (2024).
53. Jiang, Q., Sun, Y. L., Wang, Z. & Yin, Y. Aerosol composition and sources during the Chinese Spring Festival: fireworks, secondary aerosol, and holiday effects. *Atmos. Chem. Phys.* **15**, 6023–6034, https://doi.org/10.5194/acp-15-6023-2015 (2015).
54. Dou, X. *et al*. Near-real-time global gridded daily CO_2 emissions 2021. *Sci. Data* **10**, 69, https://doi.org/10.1038/s41597-023-01963-0 (2023).
55. Dou, X. *et al*. Near-real-time global gridded daily CO_2 emissions. *The Innovation*, 100182, https://doi.org/10.1016/j.xinn.2021.100182 (2021).
56. Huo, D. *et al*. Carbon Monitor Cities near-real-time daily estimates of CO_2 emissions from 1500 cities worldwide. *Sci. Data* **9**, 533, https://doi.org/10.1038/s41597-022-01657-z (2022).
57. Li, J. *et al*. Application of XGBoost algorithm in the optimization of pollutant concentration. *Atmos. Res.* **276**, 106238, https://doi.org/10.1016/j.atmosres.2022.106238 (2022).
58. Tang, D. *et al*. Harmonizing low-cost and regulatory air quality monitoring networks with interpretable semi-supervised learning: Reducing exposure misclassification in underrepresented communities. *J. Hazard. Mater.* **491**, 137893, https://doi.org/10.1016/j.jhazmat.2025.137893 (2025).
59. Zhao, Z. *et al*. Long-term spatiotemporal variations in surface NO_2 for Beijing reconstructed from surface data and satellite retrievals. *Sci. Total Environ.* **904**, 166693, https://doi.org/10.1016/j.scitotenv.2023.166693 (2023).
60. Zhan, Y. *et al*. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environ. Pollut.* **233**, 464–473, https://doi.org/10.1016/j.envpol.2017.10.029 (2018).
61. Zhan, Y. *et al*. Satellite-based estimates of daily NO_2 exposure in China using hybrid random forest and spatiotemporal kriging model. *Environ. Sci. Technol.* **52**, 4180–4189, https://doi.org/10.1021/acs.est.7b05669 (2018).
62. Li, T. *et al*. Global sectoral daily CO_2 emissions from 1970 to 2024. *Figshare* https://doi.org/10.6084/m9.figshare.28827932 (2025).
63. Tavazza, F., DeCost, B. & Choudhary, K. Uncertainty Prediction for Machine Learning Models of Material Properties. *ACS Omega* **6**, 32431–32440, https://doi.org/10.1021/acsomega.1c03752 (2021).

Acknowledgements
This paper is supported by the Open Research Program of the International Research Center of Big Data for Sustainable Development Goals (Grant No. CBAS2023ORP06), National Natural Science Foundation of China (Grant No. 72394402), and China Manned Space Program through its Space Application System (Grant No. SCP03-01-04).

Author contributions
T.L.: Writing - Original Draft, Methodology, Software, Validation, Visualization; L.W.: Writing - Methodology, Visualization, Review & Editing; Z.Q.: Writing - Methodology, Software, Visualization, Review & Editing; P.C.: Writing – Methodology, Review & Editing; S.D.: Writing - Review & Editing; Z.D.: Writing - Visualization, Review & Editing; Y.Z.: Writing – Visualization, Review & Editing; G.P.: Writing – Methodology, Validation, Review & Editing; P.K.: Writing – Visualization, Review & Editing; M.J.: Writing – Methodology, Visualization, Review & Editing; R.A.: Writing – Methodology, Review & Editing; Y.H.: Writing – Visualization, Review & Editing; T.S.: Writing – Methodology, Visualization, Review & Editing; X.H.: Writing – Validation, Review & Editing; R.J.: Writing – Review & Editing; P.F.: Writing – Review & Editing; C.L.: Writing – Review & Editing; D.C.: Writing – Review & Editing; Z.L.: Writing - Review & Editing, Conceptualization, Supervision, Funding acquisition.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Z.L.

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.