Building sector: Estimating Global, High-resolution Onsite Building Emissions

Paul Markakis[1,2,4], Trey M. Gowdy[1,4], Zion Sheng[3,4], Brittany Lancellotti[1,4], Jordan M. Malof[3,4], and Kyle Bradbury[1,2,4]

1. Nicholas Institute for Energy, Environment & Sustainability, Duke University; 2. Dept. of Electrical & Computer Engineering, Duke University; 3. Dept. of Electrical Engineering and Computer Science, University of Missouri; 4. Climate TRACE

Abstract

Globally, buildings account for approximately 26% of energy sector greenhouse gas (GHG) emissions by end-use and 8% of all direct emissions (International Energy Agency). Yet, the building sector lacks high spatial and temporal resolution emissions estimates which could help to drive action by informing planning at the local and subnational level through emissions inventories. We present a methodology that super-resolves the gridded EDGAR v8 dataset's building emissions estimates from a spatial resolution of approximately 11km-by-11km down to a 1km-by-1km resolution: a level that enables building data to be included in comprehensive municipal-level emissions inventories. We perform this disaggregation of the higher resolution EDGAR v8 emissions data using building floor area as a spatial proxy and subdivide emissions data into residential and non-residential subsectors, both based on satellite-derived data products from the European Commission's Global Human Settlement Layer. Lastly, these data are provided quarterly using heating degree days to allocate the time-varying portion of emissions. These data are provided from 2015-2023.

1. Introduction

1.1 Existing emissions inventories are inadequate

Accurate and up-to-date greenhouse gas (GHG) emissions inventories are a valuable tool for climate change mitigation and emissions reduction goal-setting; for example, such inventories underpin the effectiveness of the Paris Agreement (Umemiya and White 2023). However, multiple studies have found the state of emissions inventories at the national and subnational levels to be wanting (Umemiya and White 2023, Luers et al. 2022). Luers et al. went as far as to state that the "existing patchwork of greenhouse-gas inventories is woefully inadequate. From governments to businesses, information on these emissions is inconsistent, incomplete, and unreliable" (Luers et al. 2022). In many cases, emissions inventories at the national level are outdated, with a gap of 10 years not being uncommon (Pearce 2024). In other cases, the reliability of those inventories has come under scrutiny, and in a study evaluating promptness, accuracy, and consistency, 69 countries were deemed to be struggling in one or more of those three categories, representing roughly half of developing countries (Umemiya and White 2023).

1.2 The buildings sector GHG share is substantial and hard to estimate accurately

For a GHG emissions inventory to be valuable, each sector included needs to be accurately represented so that priorities for GHG emissions reduction targets can be accurately set. Additionally, the resolution of the information has to be sufficiently high for establishing actionable emissions mitigation plans. The buildings sector accounts for 8% of direct global emissions and an additional 18% of indirect emissions from the production of electricity and heat used in buildings (IEA); a substantial component, globally. This sector's emissions are perniciously difficult to estimate accurately, however, due to the highly distributed nature of building energy consumption and the general lack of ground truth for model calibration.

Some existing GHG inventory approaches include building sector information, but those approaches have limitations around spatial resolution and the availability of subsector information. Existing inventories include IEA's energy consumption data (which can be translated into emissions), the Open-source Data Inventory for Anthropogenic CO2 (ODIAC) (Oda et al. 2016), the Community Emissions Data System (CEDS) (McDuffie et al. 2020), the Emissions Database for Global Atmospheric Research (EDGAR) (Crippa et al. 2024), the Global Carbon Grid (Tong et al. 2018), and the Global Gridded Daily CO2 Emissions Dataset (GRACED) (Dou et al. 2022). The highest spatial resolution of these datasets are EDGAR and GRACED which are available at a resolution of about 11 km x 11 km (0.1 degree by 0.1 degree), which for context means about 3 data points for all of the Island of Manhattan (as shown in Figure 1): an insufficient resolution for local decision-making, planning, and monitoring. Similarly, there is limited subsector information, with most inventories grouping all of the building subsectors into one category rather than breaking down the estimates into, for example, residential and non-residential buildings.

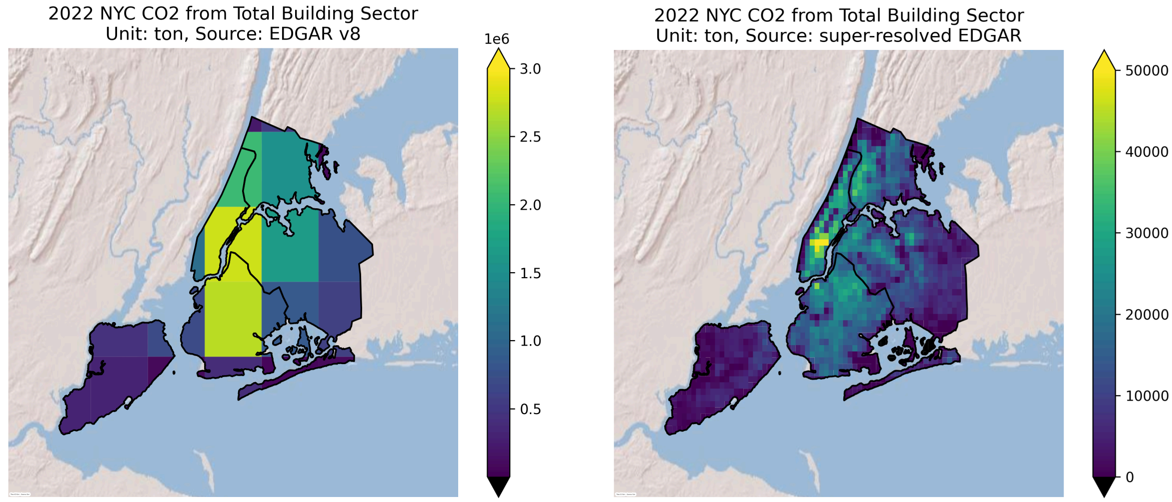

Figure 1. Visualization of EDGAR (left subplot), 0.1 degree by 0.1 degree (~11km-by-11km at the equator) building CO_2 emissions data centered on New York City and the island of Manhattan. In contrast, super-resolved EDGAR (right subplot) is provided at the 30 arc second (~1km-by-1km at the equator) resolution.

To make building sector information more actionable, the ability to identify larger sources of emissions is important. Commercial buildings vary greatly in terms of their energy use intensity (i.e., energy use per unit area), and therefore emissions, and yet in current data, it's not possible to distinguish residential from nonresidential building emissions. Hospitals, grocery stores, and restaurants may consume 2 to 8 times as much energy as residential facilities (Energy Star 2023), but there is very little by way of a means of identifying those diverse energy consumers and their resulting emissions. Providing higher resolution information separately for residential and non-residential buildings, would make these building sector data more directly actionable.

Direct (a.k.a. Scope 1) emissions from residential and commercial buildings result from onsite fuel use such as natural gas, oil, and kerosene consumption. Residential buildings primarily use these fuels for thermal comfort (space and water heating), cooking, and other equipment and appliances; in commercial buildings, space and water heating generally remain the largest end-use. The magnitude of these end uses is correlated with, for example: building size, climate conditions, and the number of occupants (Gonzalez 2022). However, timely and accessible high-spatial-resolution data on these are limited.

Existing activity-based methods to estimate emissions leverage data typically not accessible at scale, such as information on a building's insulation, appliances in a building, and information on occupant behavior (Kavgic 2010). These approaches may rely on building physics (DOE, 2014) or statistical approaches, but the data requirements for these approaches prevent them from being scaled globally.

Some activity-based approaches use spatial disaggregation techniques to reallocate emissions estimates to finer (i.e., sub-national) scales. Inventories that use these techniques include the Open-source Data Inventory for Anthropogenic CO2 (ODA 2018), the Community Emissions Data System (McDuffie 2020), the Emissions Database for Global Atmospheric Research (EDGAR) (Janssens 2019), the Global Carbon Grid (Tong 2018), and the Global Gridded Daily CO2 Emissions Dataset (GRACED) (Dou 2022). While GRACED data are published near-monthly, most other key data are produced with a year or more of latency. Additionally, the highest resolution of these data is 0.1 decimal degrees or roughly 11 km near the equator. Lastly, few of these models break down emissions into residential and commercial subsectors as well as separate emissions estimates into individual greenhouse gasses. These differences are summarized in Table 1. The lack of spatial resolution in these existing approaches makes it difficult to attribute emissions to building sources with the specificity required to inform prioritization at the local level that would ultimately drive global emissions reduction actions.

Table 1. Summary of building emissions data sources.

Emissions Data Source	**Spatial Resolution**	**Aggregated Residential and Non-Residential**	**GHGs Covered**	**Update Frequency**
UNFCCC	Country-Level	Disaggregated	CO2, N2O, CH4	Annual
CAIT	Country-Level	Aggregated	CO2, N2O, CH4	Annual
EDGAR	0.1 deg x 0.1 deg	Aggregated	CO2, N2O, CH4	Annual
GRACED	0.1 deg x 0.1 deg	Aggregated	CO2	Monthly

The methodology we present here works towards increasing the spatial and temporal resolution of building emissions data while providing a breakdown into residential and nonresidential subsectors. We introduce a method that results in ~1km-by-1km global grid cells for each primary GHG (CO2, CH4, and N2O) with a decomposition into residential and non-residential building sub-sectors as well as quarterly GHG emissions based on global temperature data.

2. Materials and Methods

The methodology we present works towards increasing the spatial and temporal resolution of the data while providing a breakdown into residential and nonresidential subsectors. To accomplish this, we perform three stages of disaggregation of the data: (1) disaggregate each ~11km-by-11km EDGAR v8 grid cell for each GHG (CO_2, CH_4, and N_2O) into ~1km-by-1km grid cells by using an estimate of building energy consumption, (2) further decompose each grid cell into residential and non-residential emissions using Global Human Settlement Layer (GHSL) data building type categorizations, and (3) allocate the time-varying component of annual GHG emissions to four quarters of the year based on heating degree days derived from Copernicus ERA5 globally gridded temperature data.

At the time of publication, EDGAR v8 data is available through 2022, we use backed-out 2022 emissions factors combined with extrapolated building area data to estimate emissions for 2023.

In this section, we first introduce the datasets used to complete this process and then describe each step of this process in detail.

2.1 Datasets employed

2.1.1 Energy and emissions datasets

EDGAR v8

EDGAR v8.0 uses spatial data on global human presence and air temperature to allocate country-level residential and non-residential building emissions estimates to a 0.1° by 0.1° resolution global grid. These datasets include the GHSL and data on heating degree days from the ERA5 climate reanalysis (Crippa et al., 2023). In this work, we super-resolve the EDGAR v8 building emissions dataset from the provided resolution of 0.1-degree-by-0.1-degree which near the equator is close to 11km by 11km onto a 30 arc second (~1km-by-1km) grid using the GHSL building area dataset. To accomplish this, for every nonzero EDGAR estimate of CO_2, CH_4, and N_2O, we allocate those emissions proportional to the fraction of building floor area in each 1km-by-1km grid cell region. This ensures the sum of the original EDGAR data is equivalent to the super-resolved dataset at a global and national level, while providing additional resolution to enable municipal-level inventories more effectively. We exclude the EDGAR emissions from biomass for CO_2 as those are accounted for in other sectors within Climate TRACE. However, it is not specified in the EDGAR v8 tool whether the sector-specific gridded data for N_2O and CH_4 contain or exclude biomass.

Energy Use Intensity data from the World Bank

Energy Use Intensity (EUI) is an indicator of energy use relative to building size. EUI data originates from the CURB Tool (World Bank 2016) which contains data for about 400 cities around the world. The EUI values by end-use and by city are presented as kWh/m^2/year. The data are provided for eight residential and non-residential building types. We include all end uses (including those typically associated with electricity) at this stage of the model, accounting for the removal of electricity consumption later in the model, as electricity emissions are accounted for elsewhere in the Climate TRACE inventory. We labeled the locations using identification numbers from the Geonames database to assign the EUI value to geographic coordinates closest in proximity. Using residential and non-residential building floor area data from the United

States, we developed a weighted average of EUIs across building types to develop a single residential and single non-residential estimate for each city (see supplementary materials (S.2, S.4.1) for more details on this process).

2.1.2 Remote sensing datasets

Building Area

The Global Human Settlement Layer (GHSL) is provided openly by the European Commission's Joint Research Centre (European Commission 2023) and we use this as our source of building-related data. These data, derived from Landsat and Sentinel-2 satellite imagery, offer global coverage of key building-related factors at a desirable spatial resolution. GHSL provides building-related datasets associated with surface areas, volumes, and heights at varying levels of resolution and across several different years. To generate these data, GHSL used machine learning to infer various building-related attributes from multiple distinct sources of satellite imagery collected across the entire surface of the earth. Additional detail on the GHSL data and methodology is available via GHSL (European Commission 2023).

We rely upon the built-up volume area data product, BUILT-V (Pesaresi 2023c), at 30 arc second resolution (~1km-by-1km at the equator) and convert it to total floor area by assuming a 3m average height for a typical story of a building. This allows us to create an estimate of the total floor area contained within the building volume in each region. As the source data are already subdivided into residential and non-residential building categories based on GHSL modeling, we directly use the GHSL categorization of the data to assign building-related emissions to residential and non-residential buildings. GHSL offers select years of data including 2015 and 2020. For the remaining years, we linearly interpolate and extrapolate at a grid-cell level to estimate building floor area for 2016-2019, and 2021-2023. See the supplemental materials (S.1) for more details about building data preprocessing.

Copernicus ERA5 3m temperature data

The European Commission's Copernicus program provides gridded climate reanalysis data through its suite of ERA5 data products. We use ERA5 2m temperature data as our primary source of daily temperature for computing heating and cooling degree days to allocate emissions quarterly. These data provide a highly accurate temperature estimate at an hourly cadence, globally. The data are available on a lat-lon grid with 0.25deg x 0.25deg spatial resolution.

2.1.3 Validation data

True ground truth of building emissions data would be measurements of the emissions emitted from individual buildings or groups of buildings. While there are measurements of *energy* consumed by individual buildings, there are not, to our knowledge at the time of this research, any known measurements of actual emissions from individual buildings or neighborhoods of buildings, and certainly none that we are aware of at a wider scale. At the national level, there are emissions inventories from the United Nations Framework Convention on Climate Change (UNFCCC) (UNFCCC, 2024) and the latest version of the Climate Action Indicators Tool (CAIT) curated by the World Resources Institute (Washington, DC: World Resources Institute, 2024). However, these two data sources both rely upon member nations to respond to surveys related to energy consumption and emissions so we do not treat them as empirical ground truth. At the municipality level, we rely upon region-specific estimates of building sector emissions data at the municipality level from the Data Portal for Cities (DPFC) tool as aggregated by the Global Covenant of Mayors for Climate & Energy (Global Covenant of Mayors for Climate & Energy, 2019). While also not empirically measured ground truth, since these estimates are performed at the municipality-level, we use the DPFC data as a source of local comparison and evaluation for our estimates at a high spatial resolution. These comparisons are further described in Section 2.4.

2.2 Models

Activity-based approaches for estimating building emissions generally follow the 2006 Intergovernmental Panel on Climate Change (IPCC) guidelines, which combine information on energy consumption and emissions factors for a given region (IPCC, 2006). Of course, neither energy consumption nor emissions at the individual or neighborhood level are directly observable at scale. What is observable, however, are buildings themselves, and we leverage this to create a three-part activity based model which combines building floor area (in square meters), or A; building energy use intensity (Joules per meter), or EUI; and emissions factors (tonnes of gas per Joule), or EF. If each of these quantities is accurate in a given region, we can estimate tonnes of emissions, E, as $E = (A)(EUI)(EF)$.

Since this work is focused on super-resolving EDGAR's gridded emissions estimates to a higher spatial and temporal resolution, we assume a regional (country-level) value of EUI (World Bank 2016) and assume that regionally (within a ~11km-by-11km EDGAR grid cell), that emissions factors are approximately constant. Therefore the variation of emissions is due to the differences in building energy consumption; that assumption we can use for disaggregating the emissions.

The GHSL data we use for building floor area also provides a classification of building volume into residential and non-residential classes. We use these classifications from the GHSL data and adopt the same definition of residential buildings as the GHSL, which is the European Commission's INSPIRE definition (INSPIRE 2024) of the residential sector:

Residential: "Areas used dominantly for housing of people. The forms of housing vary significantly between, and through, residential areas. These areas include single family housing, multi-family residential, or mobile homes in cities, towns and rural districts if they are not linked to primary production. It permits high density land use and low density uses. This class also includes residential areas mixed with other non-conflicting uses and other residential areas." (INSPIRE 2024)

While an explicit definition of non-residential is not provided, we define it as follows to be mutually exclusive and collectively exhaustive of buildings and assume that the emissions from **non-residential** buildings are any direct building emissions not included within the residential sector and include emissions from commercial and municipal buildings.

One challenge with the residential definition is that it will encompass some commercial building activity by including mixed-use and other building types that happen to exist within an area of predominantly residential buildings. This will lead to an overestimation of residential building emissions and an underestimation of non-residential emissions, which we do see in our data.

2.3 Methods

The process we use for generating the emissions estimates involves estimating activity data (energy consumption), disaggregating EDGAR emissions to a higher spatial resolution using the activity data, allocating emissions into residential and nonresidential categories, extrapolating data to estimate 2023 emissions, and generating higher temporal estimates of the data by allocating the variable component of emissions using heating degree days.

Before we begin detailing the steps of this methodology, we will include some definitions that we use for describing the emissions estimation procedure. In Table 1, we introduce the key indices that we use to reference our variables, and in Table 2 introduce the variables themselves.

Table 1. Definition of key indices used in the methodology.

Variable	**Definition**	**Possible values**
r	A region in space, as represented by the index of any grid cell in the output emissions dataset (representing a ~1km x 1km region at the equator)	The index of a grid cell in the output emissions dataset
R	The region covered by an EDGAR grid cell (representing a ~11km x 11km region at the equator)	The index of a grid cell in the EDGAR emissions gridded dataset
c	Building type category	residential (res), nonresidential (nres)

Variable	Definition	Possible values
g	Greenhouse gas	Carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O)

Table 2. Definition of key variables used in the model.

Variable	Units	Description
$A(r)$	m^2	Building floor area *Source: Global Human Settlement Layer, Building Volume data (BUILD-V)*
$F(r, c)$	*unitless*	Fraction of building type *Source: Global Human Settlement Layer, Building Volume data (BUILD-V)*
$EUI(r, c)$	J/m^2	Energy use intensity *Source: World Bank CURB tool*
$\hat{EF}(r, c, g)$	tonnes/J	Emissions factor
$\hat{EN}(r, c)$	J	Energy
$E^{EDGAR}(R, g)$	tonnes	EDGAR emissions *Source: EDGAR v8*
$\hat{E}(r, c, g)$	tonnes	Super-resolved emissions

Step 1. Calculate the activity data.

Here, we calculate the product of building floor area data (converted from building volume assuming 3m floor heights), $A(r)$, dividing it up into residential and non-residential floor area based on GHSL BUILT-V data classifications, $F(r, c)$, and multiply that product by the EUI value for the given region and building type $EUI(r, c)$, which are unique for each country, computed as the average of EUI values within that country. Additional assumptions and processing for the EUI data can be found in the supplementary materials. This process yields estimates of energy consumption for every ~1km-by-1km region of the world and by building type category (residential, nonresidential), $\hat{EN}(r, c)$.

$$\hat{EN}(r, c) = A(r)F(r, c)EUI(r, c)$$

This process is shown in Figure 2A.

Step 2. Disaggregate emissions.

Using the activity data, the EDGAR v8 gridded data, $E^{EDGAR}(R, g)$, are allocated proportionally based on the fraction of energy (activity) in each grid cell. This estimation process is completed for each gas and each category of buildings:

$$\hat{E}(r, c, g) = \left[\frac{\hat{EN}(r,c)}{\sum_{r\in R} \hat{EN}(r,c)}\right] E^{EDGAR}(R, g)$$

This process is shown in Figure 2B.

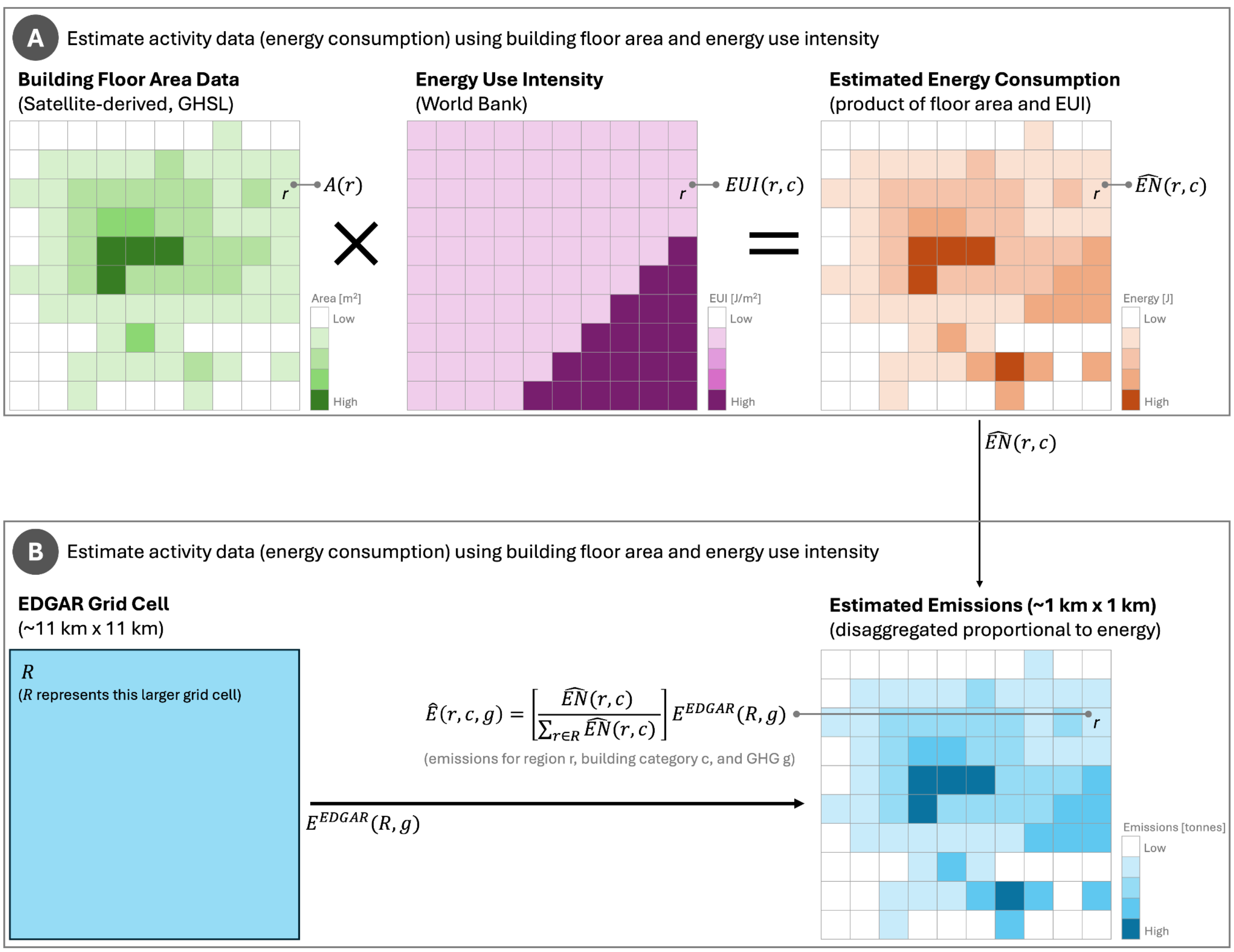

Figure 2. EDGAR disaggregation methodology.

Step 3. Back out emissions factors.

Through steps 1 and 2, we do not need emission factors to calculate the disaggregated emissions. However, EDGAR generally has a latency of more than a year, therefore, we back out emissions factors for each year so that we can use them for extrapolating emissions estimates to 2023. Our estimates of emissions factors are the emissions estimates divided by the energy estimates for each building category and GHG:

$$\hat{EF}(r, c, g) = \frac{\hat{E}(r,c,g)}{\hat{EN}(r,c)}$$

Step 4. Extrapolate to 2023.

To create estimates for 2023, we linearly extrapolate the building floor area data from the 2015 and 2020 estimates derived from the GHSL. Using the extrapolated building floor area, we can calculate emissions as the product of the extrapolated building floor area, EUI (which we assume is unchanged from 2022), and the emissions factors in Step 3 which we estimated from 2022 data.

Step 5. Estimate quarterly emissions.

As space heating is a major driver of energy consumption in buildings, we make the assumption that heating needs are the primary driver of variable energy consumption in buildings. Globally, variable loads (loads that change seasonally, space heating and cooling) represent around 37.8% of consumption (International Energy Agency 2021). We allocate 37.8% of load based on heating degree days for a given quarter, $HDD(q)$. Emissions for a given quarter, which we represent as $\hat{E}(r, c, g, q)$ are then allocated according to the relationship:

$$\hat{E}(r, c, g, q) = \alpha\left[\frac{HDD(q)}{\sum_{q=1}^{4} HDD(q)}\right]\hat{E}(r, c, g) + (1 - \alpha)\left[\frac{\hat{E}(r,c,g)}{4}\right]$$

Here, $\alpha = 0.378$, which is the fraction of variable load. This concept is demonstrated in Figure 3. The grey, baseload emissions are constant throughout the year and are simply $(1 - \alpha)\frac{\hat{E}(r,c,g)}{4}$ each quarter. The variable load emissions (the fraction represented by α) are distributed proportionally to the quantity of HDDs that occurred that quarter relative to the total for the year: $\left[\frac{HDD(q)}{\sum_{q=1}^{4} HDD(q)}\right]\hat{E}(r, c, g)$.

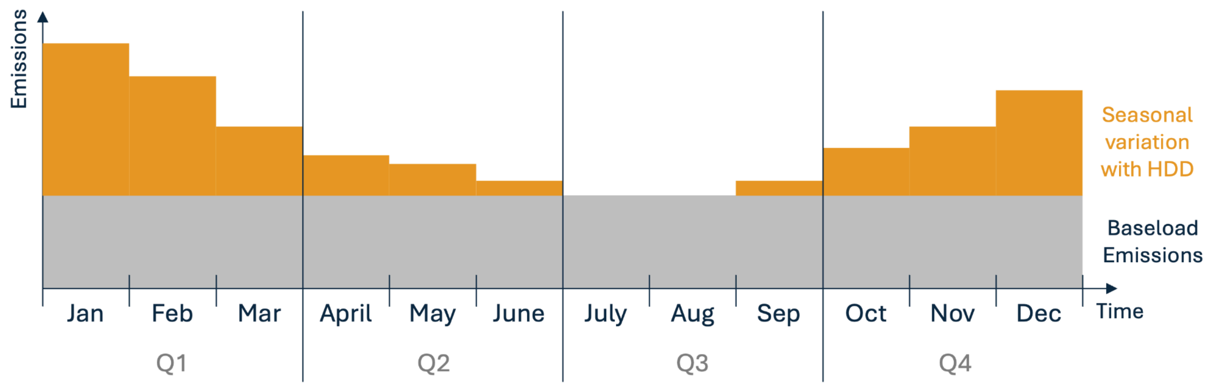

Figure 3. Quarterly emissions allocation.

Regions for which there are few HDDs per year may result in unrealistic estimates. For example, if a region only has 1 HDD per year, 37.8% of annual emissions would be shifted in full to the quarter containing that HDD. To prevent this outcome, we linearly scale up α from 0 to 0.378 based on the total number of HDDs. The value of 0.378 results if there are 1000 HDDs or more for a given region.

2.4 Verifying modeled emissions estimates

As there is no global empirical ground truth for emissions estimates, we instead compare our modeled emissions estimates to trusted inventories at both the national and municipality levels. At the national level, we aggregate our modeled emissions estimates by country and compare them against both the UNFCCC and the latest version of CAIT. At the municipality level, we aggregate our modeled emissions estimates by municipality boundary and compare them against emissions data from DPFC. Additional detail regarding the municipality boundaries used for the DPFC comparisons is included in Sections S.4.4 and S.4.5.

Across all comparisons we use three quantitative metrics to assess the quality of our modeled emissions estimates as compared to each of the three trusted inventories: Mean Absolute Percentage Error (MAPE), Weighted Absolute Percentage Error (WAPE), and Root Mean Squared Error (RMSE). These three metrics each consider error differently, and collectively provide a more comprehensive understanding of estimation model error than by using any one of the three metrics on its own. While MAPE considers error across all comparisons as equally important (e.g., percentage error estimating country-level emissions in the United States would be weighted equivalently to error estimating country-level emissions in Vatican City), the WAPE metric weights error in regions that have a greater magnitude of emissions more heavily than in regions with fewer emissions (e.g., percentage error estimating country-level emissions in the United States would be weighted more heavily than error estimating country-level emissions in Vatican City). The RMSE metric measures differences between the trusted inventory emissions values and the model-predicted emissions values, and the resulting error using this metric is in units of tonnes rather than the percentages of the MAPE or WAPE metric.

For our national-level comparisons, we aggregate modeled emissions estimates using country-level polygon boundaries for each country in UNFCCC and CAIT, respectively. To be included as part of the aggregation and ensure that there is no double-counting, the grid cell's centroid must exist within the country-level polygon boundary to have those emissions assigned to that country. Figure 4 shows maps denoting the presence of emissions data at the country level for UNFCCC and CAIT, respectively.

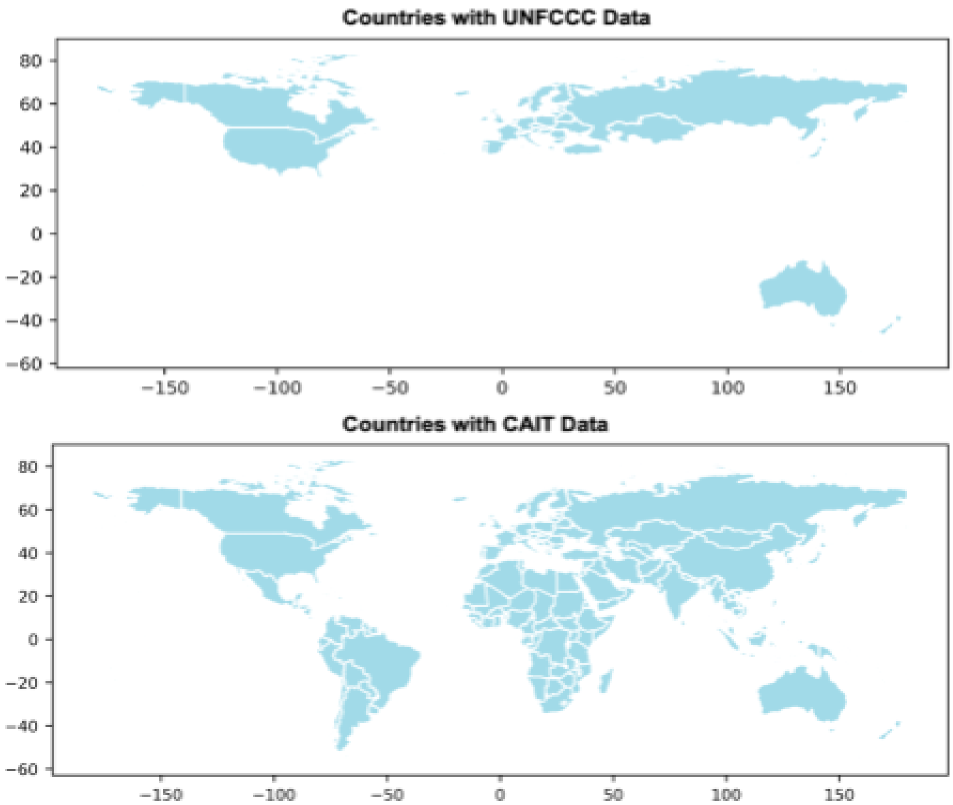

Figure 4. The two maps above show the countries containing emissions data from UNFCCC (above) and CAIT (bottom).

We can compare both the original EDGAR v8 data and our modeled emissions against both UNFCCC and CAIT for "total" buildings (i.e., aggregated residential and non-residential buildings). We are additionally able to compare our modeled emissions against UNFCCC data separately against disaggregated residential and non-residential buildings, as those disaggregated breakdowns are provided in the UNFCCC data; unfortunately, we must exclude the original EDGAR v8 from the subsector-specific comparisons, as those breakdowns are not included in the original EDGAR data. Similarly, we are unable to compare subsector-specific emissions against the CAIT data as only the “total” aggregated residential and non-residential buildings data are provided in the CAIT inventory.

3. Results

3.1 Validation Results

The figures below show the results of applying the country-level and municipal level validation strategies. First, we begin with the country-level comparison showing aggregate emissions for each country and compare four different emissions data sources: UNFCCC, CAIT, original EDGAR v8, and super-resolved EDGAR. This comparison is shown in Figure 5 for CO_2 emissions. We show results here for 2020, as this is the most recent year for which all comparison datasets have published emissions data. Figure 5 shows that at this level of aggregation, EDGAR and super-resolved EDGAR are nearly identical, with the exception of some small countries for which the original gridded EDGAR data was not sufficiently high resolution to produce a non-zero estimate. While all four emissions data sources are generally similar in country-level emissions magnitude, Figure 5 also shows that there still are some differences (to include differences between UNFCCC and CAIT). For other years and GHGs, refer to section “S.3. Additional Emissions Comparison Model Results”.

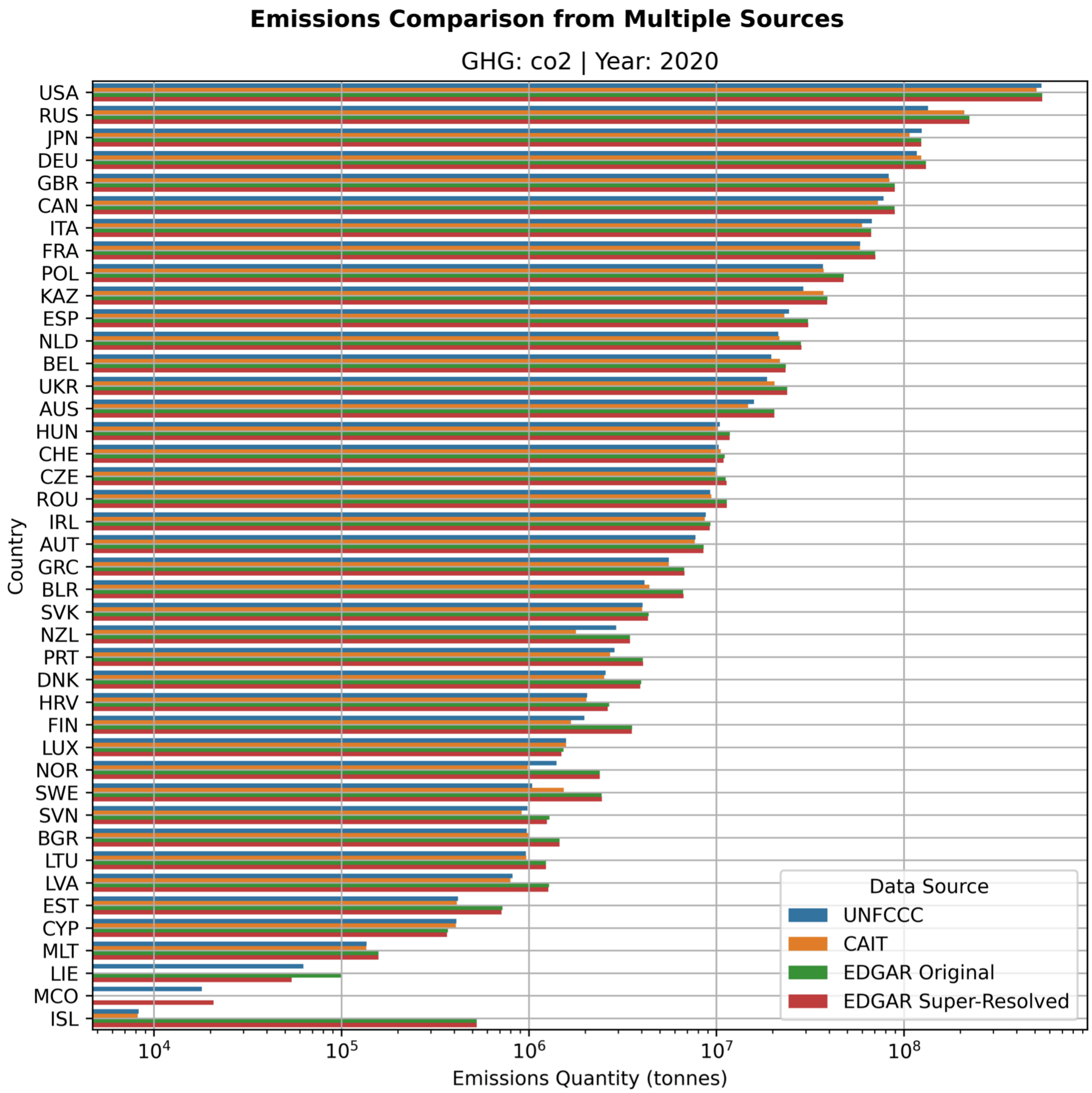

Figure 5. We compare the quantity of CO_2 emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2020. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

At the municipality-level, we use data from DPFC as a point of comparison against estimates from both the original EDGAR v8 data and our model. Because the grid cells in the original EDGAR v8 are often larger than the boundary of a municipality, we estimate the emissions contribution using a proportional allocation method. For this method, we first aggregate all EDGAR v8 grid cells that have any overlap with the municipality's boundary polygon and compute the sum of all emissions in those grid cells. We simultaneously compute a ratio of the municipality boundary area to the area of all grid cells touched by said municipality (note: we assume that each grid cell has an area of 121 km^2 for this computation). Finally, we multiply the emissions sum by the ratio to compute the estimated emissions associated with that municipality according to the original EDGAR's model. Computing the emissions contribution using our model's smaller 1km x 1km grid cells is far simpler; if the grid cell's centroid is contained within the municipality's boundary, that grid cell is considered as part of the municipality's emissions.

The DPFC emissions data for CO_2, N_2O, and CH_4 are disaggregated into residential and non-residential building sub-sectors, and each country's data is associated with a corresponding year. (Note that India's DPFC data are from 2013. We use 2015 data from EDGAR v8 and our model for DPFC validation, as 2015 data are the first that we consider as part of our current work.) DPFC data are also disaggregated into different contributing fuel types, so we can filter the data down to only include direct onsite fuel types and exclude sources such as electricity and district heating that are outside of the scope of our work as they are accounted for elsewhere in Climate TRACE. We filter DPFC data to only include municipalities that contain both residential and non-residential emissions. Because the exact geographic boundaries used by DPFC for each municipality are not provided, we curate boundaries for each region and exclude any municipalities for which the inferred population differs greatly from the DPFC reported population. The resulting comparison data once all of this filtering is complete contains a total of approximately 20,000 distinct municipalities across 8 different countries (Canada, Chile, Costa Rica, Denmark, India, Japan, Mexico, and the United States). Because DPFC data are provided in units of CO_2e, we also use the latest conversion factors associated with the 100-year Global Warming Potential to convert CO_2e into tonnes of each corresponding GHG.

Figure 6 below summarizes the WAPE, MAPE, and RMSE for each country in DPFC as compared to both the original EDGAR v8 and our model's estimates. Unlike the country-level comparisons shown in Figure 5, which showed that the original EDGAR v8 and our model's estimates were virtually identical, the municipality-level estimates are noticeably different for each validation region from DPFC. In many cases – particularly for CO2 – our super-resolved model is an improvement over the original EDGAR v8 data. For those cases in which the original EDGAR outperforms our model, we are able to use the municipality-level data from DPFC to understand why this difference occurs and thereby identify improvements to include in future iterations of our model.

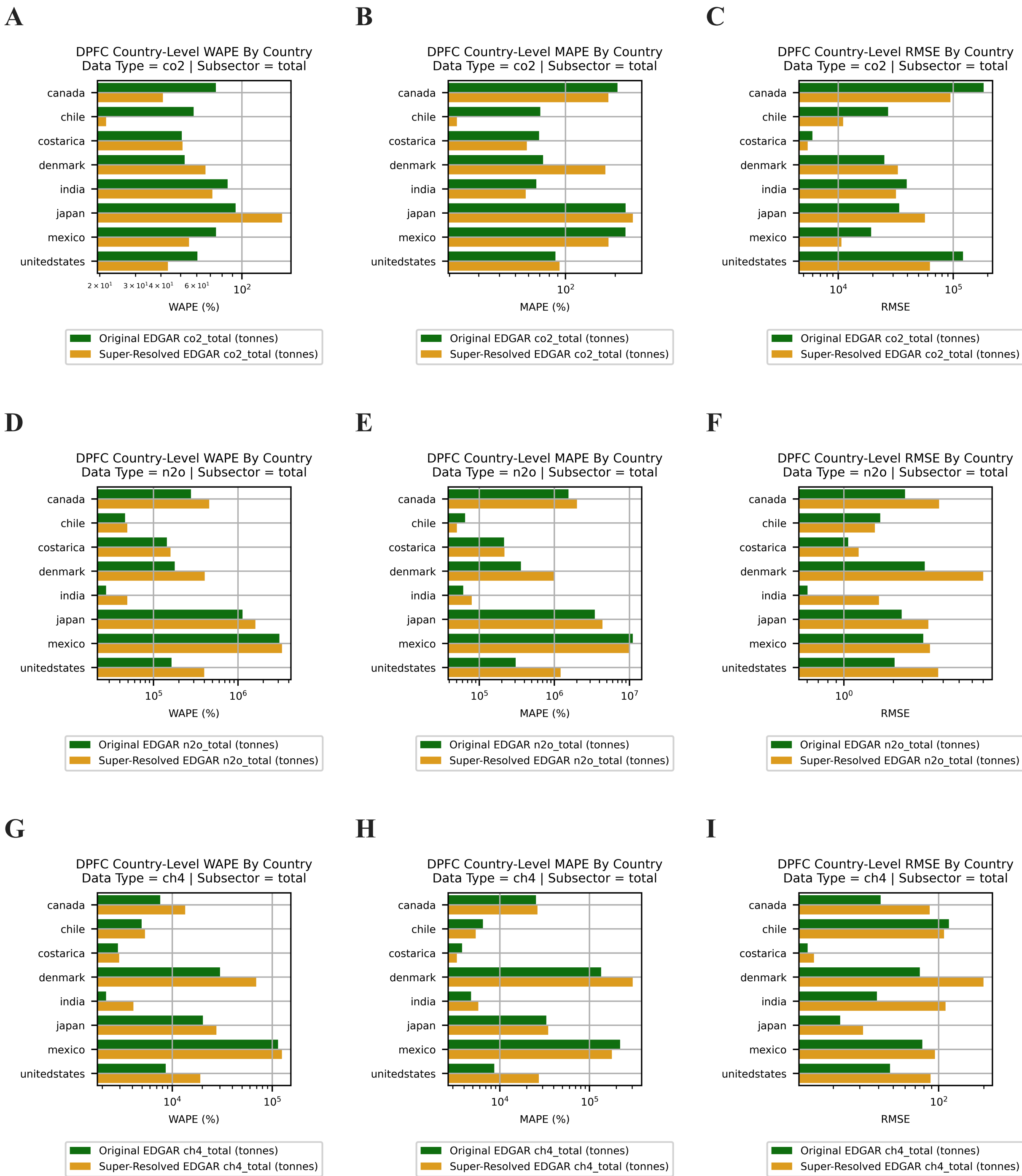

Figure 6. Comparison of WAPE (A, D, and G), MAPE (B, E, and H), and RMSE (C, F, I) between original EDGAR v8 (green) and our model (gold) for each DPFC validation region and the total buildings sector. These comparative metrics are all shown for CO_2 (A, B, and C, top row), N_2O (D, E, and F, middle row), and CH_4 (G, H, and I, bottom row).

3.2 Visualization of emissions

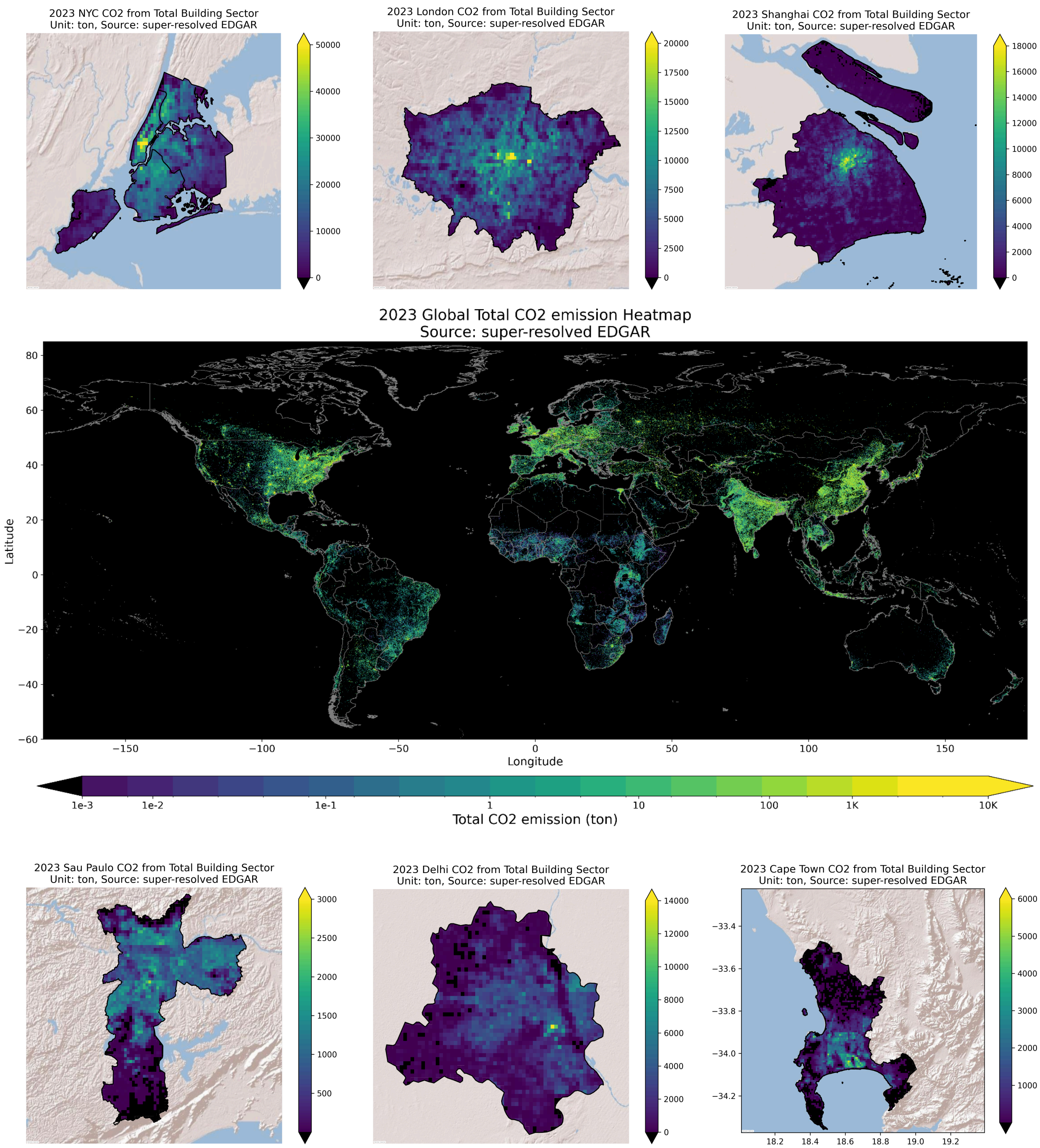

Figure 7. Global and city-level CO_2 emission heatmap. The value of each grid cell is the estimated CO_2 emission from the total building area in that 1km by 1km region in 2023 given by super-resolved EDGAR (unit: metric tonnes). As a shorthand, we use "Total CO_2" in the plots to denote the whole-year CO_2 emission from the total buildings (the sum of residential and non-residential buildings). From the top left to bottom right, selected cities are NYC, London, Shanghai, Sau Paulo, Delhi, and Cape Town. Places with zero building area (zero building-generated CO2 emission) are colored in black.

The approach presented here, which super-resolves EDGAR using high resolution building floor area data, provides global greenhouse gas emissions estimates at a higher spatial resolution (~1km by 1km) than any previous inventory for building emissions. The data shown in Figure 7 include both a global map of total building emissions and city-level data for six major cities across the globe.

The global plots show, as expected, the density of building emissions in some of the most populated regions of the world including the eastern United States, Europe, China, and India. What is new and different about this dataset is the ability to show differences within cities themselves. Emissions from Manhattan stick out clearly in New York, U.S., while the coastal regions of Cape Town, South Africa, clearly show a higher concentration of emissions. This level of spatial granularity enables prioritization in local and regional planning and decision-making when paired with other emissions sectors. See the supplemental materials (S.4.2) for details on known issues with final estimates.

4. Discussion

4.1 Contributions

Our work introduces a model that provides high-spatial-resolution global emissions for CO_2, N_2O, and CH_4. These emissions from our model are further disaggregated into residential and non-residential building subsectors and each year's emissions are reflected quarterly. Using activity data to spatially super-resolve the data from EDGAR v8 into 30-arcsec grid cells (roughly 1km x 1km grid cells) not only provides a two-order-of-magnitude increase in the information content for each 0.1 deg x 0.1 deg (roughly 11km x 11km) grid cell from the original EDGAR data, but it proportionally allocates the EDGAR emissions based on the underlying activity data most representative of where emissions are concentrated within each region. This work is also the first attempt to provide gridded data disaggregated into residential and non-residential subsectors. Finally, using temperature data, this work provides gridded data quarterly enabling the EDGAR emissions to be allocated temporally based on changing space heating needs.

While the results at the national level are the same as EDGAR v8 at the national level, there were municipal-level improvements in CO_2 estimates for the majority of countries for which municipal data are available. The differences—and in many cases, reduced error—from our super-resolved version of EDGAR are observable in the measurement of WAPE, MAPE, and RMSE at the municipality level compared to DPFC data.

4.1 Limitations

Comparison rather than validation. Although there are clear benefits to our model's temporal and spatial resolution improvements as compared to existing models, a significant challenge is the limited data available for validation to quantify model performance. In fact, there is no true ground truth data for building emissions, so we rely on comparisons to other inventories at the national and municipal level. UNFCCC and CAIT emissions data are both only available with annual values at a country level; these are insufficient to fully measure performance at subnational levels. While DPFC data are available at the municipality-level and thus can be used for sub-country-level analysis, they are only included in the source data for a subset of municipalities represented by eight countries in the world and only exist for a single year and annual set of emissions.

Quarterly emissions estimates. We currently do not have data available for validating or comparing quarterly emissions estimates.

Biomass exclusion is unclear in source and comparison data. Across inventories used for comparison and for CH_4 and N_2O for the gridded EDGAR data, biomass is aggregated with fossil fuels into a single category. This inconsistency complicates accurate emissions accounting and comparison, especially for N_2O and CH_4. See the supplemental materials (S.4.3) for more details.

Residential and non-residential definitions don't perfectly align with other inventories. By adopting the definitions of residential regions including the GHSL data, which categorizes all regions whose primary purpose is the housing of people as residential (this includes all mixed-use regions), more regions and buildings will be categorized as residential. This means the estimates for residential emissions will be a bit higher and non-residential a bit lower, than if alternative definitions of residential and non-residential buildings are used.

Each of these limitations has led to areas of active research to overcome these challenges in future iterations of this work. See the supplemental materials (S.4) for additional known issues.

5. Conclusion

This work introduced a methodology that builds on EDGAR v8 by using high-resolution building data derived from satellite imagery to create the highest spatial and temporal resolution estimates of building emissions presently available. This work also provides subsector disaggregation for residential and nonresidential buildings and quarterly emissions estimates. The data are compared at the national level to several other global emissions inventories and, for the first time, to municipal-level emissions estimates.

Our model's emissions data at the national level are virtually equivalent to the original EDGAR v8. Since we disaggregate the original gridded EDGAR v8 data into significantly smaller grid cells, when both versions of EDGAR are aggregated to a country level there is little difference between the two models' emissions data. The super-resolved EDGAR data, however, is closer to municipal emissions estimates from Data Portal for Cities emissions values for 5 of 8 countries included in the data across the major metrics of comparison for CO_2 data, with one additional country being approximately the same as EDGAR v8 for a comparison with DPFC. While there are greater disparities for CH_4 and N_2O, this is likely because EDGAR v8 contains aggregated fossil fuels and biofuels emissions estimates for CH_4 and N_2O, whereas DPFC only contains data for fossil fuels; because biofuels have disproportionately larger emissions factors for CH_4 and N_2O as compared to fossil fuel emissions factors, the discrepancy between these data sources is the likely cause of our performance for CH_4 and N_2O. Ongoing research is focused on overcoming this challenge.

These high-resolution data are the first to enable municipal-level building emissions inventories at a global scale. These data can also enable the development of a building emissions inventory for any arbitrarily defined region. We hope that these data can be an additional tool to guide planning and prioritization for climate change mitigation strategies at several organizational levels.

6. Acknowledgments

The authors would like to WattTime and the Climate TRACE coalition for their organizational support, and Climate TRACE funders – Al Gore, Generation Investment Management, Google.org, and the Patrick J. McGovern Foundation.

We also thank the following students who have assisted with this project - without their help this work would not have been possible: Pablo Salazar Armella, Ya-Yun Huang, Julia Kourelakos, Pragya Raghuvanshi, Nicholas Sommer, Nicholas Sorokin, Weilin Wang, and Yuanjing Zhu.

S. Supplementary materials

Supplemental section metadata

Covered Emissions and Available Data: Emissions from residential and non-residential buildings are covered in these data. The asset and country-level data that is freely available on the website can be found in Tables S1, S2 and S3 below. Additional information on building emissions by country for country-level emission estimates can be shared upon request.

Table S.1. General dataset information for buildings

General Description	Definition
Sector Definition	*Residential and Non-Residential (Commercial) Buildings*
UNFCCC sector equivalent	*1A4+1A5*

General Description	Definition
Temporal Coverage	*2015-2023*
Temporal Resolution	*Quarterly and annual; Monthly (on website)*
Data format(s)	*CSV*
Coordinate Reference System	*EPSG:4326, decimal degrees*
Number of assets/countries available for download and percent of global emissions (as of 2023)	*Data covers 251 countries (across ~924M grid cells)*
Total emissions for 2023	*3.4 billion tonnes of CO_2e*
Ownership	*N/A*
What emission factors were used?	*Emissions factors were backed out from EDGAR data through a process described in step 3 of Section 2.3*
What is the difference between a "NULL / none / nan" versus "0" data field?	*"0" values are for regions that we estimate as having zero emissions. If we know that the grid cell has emissions for that specific gas, but the gas was not modeled or the data were incomplete, this is represented by "NULL/none/nan".*
total_CO2e_100yrGWP and total_CO2e_20yrGWP conversions	*Climate TRACE uses IPCC AR6 CO2e GWPs. CO2e conversion guidelines are here: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full*

Table S.2. Asset level metadata description for building emissions.

Data attribute	Definition
sector	*buildings*
asset_sub-sector_name	*residential, nonresidential*
asset_definition	*Buildings within a 30 arcsecond by 30 arcsecond (~1km x 1km) grid cells*
start_date	*start date of record*
end_date	*end date of record*
asset_identifier	*Grid cell ID for 30 arcsecond by 30 arcsecond grids*
asset_name	*Grid cell ID for 30 arcsecond by 30 arcsecond grids*
iso3_country	*country code*
location	*N/A*
type	*N/A*
capacity_description	*building floor area*
capacity_units	*square meters*
capacity_factor_description	*energy consumption per unit area of building*
capacity_factor_units	*MJ per square meter*
activity_description	*energy consumption*
activity_units	*MJ per year*
CO2_emissions_factor	*Emissions factors were backed out from EDGAR data through a process described in step 3 of Section 2.3*
CH4_emissions_factor	*Emissions factors were backed out from EDGAR data through a process described in step 3 of Section 2.3*

Data attribute	Definition
N2O_emissions_factor	*Emissions factors were backed out from EDGAR data through a process described in step 3 of Section 2.3*
other_gas_emissions_factor	*not modeled*
CO2_emissions	*CO2 emissions estimate*
CH4_emissions	*CH4 emissions estimate*
N2O_emissions	*N2O emissions estimate*
other_gas_emissions	*not modeled*
total_CO2e_100yrGWP	*CO2e emissions, 100 year global warming potential*
total_CO2e_20yrGWP	*CO2e emissions, 20 year global warming potential*
other1_description	*N/A*
other1_units	*N/A*

Table S.3. Asset level metadata description confidence and uncertainty for buildings. Note, confidence estimates range from: "*Very low = estimates based upon assumption driven statistical models; empirical validation was not able to be performed due to a lack of alternative data for comparison*" to "*Low = Data have been compared to at least one other data source*".

Data Attribute	Confidence Definition	Uncertainty Definition
type	*N/A*	*N/A*
capacity_description	*confidence of capacity estimate*	*IPCC high and low ranges of capacity*
capacity_factor_description	*confidence of capacity factor estimate*	*IPCC high and low ranges of capacity factor*
capacity_factor_units	*N/A*	*MJ per square meter*
activity_description	*confidence of activity estimate*	*IPCC high and low ranges of activity*
CO2_emissions_factor	*confidence in emissions factor*	*IPCC high and low ranges*
CH4_emissions_factor	*confidence in emissions factor*	*IPCC high and low ranges*
N2O_emissions_factor	*confidence in emissions factor*	*IPCC high and low ranges*
other_gas_emissions_factor	*N/A*	*N/A*
CO2_emissions	*confidence in CO2 emissions*	*IPCC high and low ranges*
CH4_emissions	*confidence in CH4 emissions*	*IPCC high and low ranges*
N2O_emissions	*confidence in N2O emissions*	*IPCC high and low ranges*
other_gas_emissions	*N/A*	*N/A*
total_CO2e_100yrGWP	*country level confidence in emissions*	*IPCC high and low ranges*
total_CO2e_20yrGWP	*country level confidence in emissions*	*IPCC high and low ranges*

S.1. Building floor area

We also identified that the GHSL volume data, when converted to total floor area of buildings (by assuming an average floor height of 3m per story) and combined with energy use intensity to compute energy consumption data, tended to overestimate energy consumption at the national level, so a scaling factor of 0.12 was applied to the building floor area estimates based on a regression of our derived energy consumption estimates and national level energy consumption from IEA data. Since our goal is to estimate emissions, not energy consumption directly, what is most important is that the estimates of energy consumption are proportional to actual energy consumption and globally consistent.

S.2. Energy Use Intensity

The data sources used by the World Bank CURB Tool for these values are the 'DOE Model' and EDGE,' model year(s) unknown. The CURB Tool was released in 2016. Using building stock information from the United States Energy Information Administration's Commercial Building Energy Consumption Survey (CBECS) (U.S. Energy Information Administration 2023a) and Residential Energy Consumption Survey (RECS) (U.S. Energy Information Administration 2023b), we developed a proportional mix of values for residential vs. non-residential building categories to create a weighted average of building EUI totals. We then developed weighted averages across socioeconomic tiers for the residential sector (homes - houses and apartments/flats) to create a sole residential sector value for each location.

Table S.4. Building Floor Area Weighted Average Data Preparation

Building Category	**Percentages**	**Specific Building Type**	**Sq. Footage (US) - To inform percentage (Millions SF)**	**Percentages (based on US data)**	**Notes/Source**
Residential	71.12%	Home	237,400.00	71.12%	RECS 2015 HC10.1 - includes all Single Family Homes, Apartment Units, and Mobile Homes.
Non-residential	28.88%	Hospital	4,018.00	1.20%	CBECS 2018 B12 - categories based recategorization of CBECS data categories to align with CURB Tool building types
		Hotel	6,976.00	2.09%	
		Office	18,200.00	5.45%	
		Retail	19,412.00	5.82%	
		Warehouse	17,483.00	5.24%	
		Other	30,333.00	9.09%	

Table S.4. This table describes the weighted averages derived from U.S. Energy Information Administration (EIA) data to allocate floor area to the building types found in the World Bank

CURB Tool for energy use intensity values. The EIA values originate from the Residential Energy Consumption Survey (RECS) and Commercial Building Energy Consumption Survey (CBECS).

There are a few limitations with this process. The EUI values represent only a single point in time (the exact year for each location is unknown) and the values are associated primarily with major population centers only (potentially reducing accuracy for rural and suburban areas). Building-use percentages are based on a US average, and then applied globally. Identifying more localized detailed building use percentage estimates that can be applied globally will improve accuracy. Further refinement of EUI values by recency and spatial resolution (more locations overall, including rural and suburban areas) provide opportunity for improvement to this model. These data are then converted from kilowatt hours (kWh)/square meter to megajoules (MJ)/square meter.

Prior to that conversion, we clip outlying data points. We clip all values to be between 50 and 400 kWh/m^2/year to eliminate potential outliers, before then converting them MJ/m^2/year.

S.3. Additional Emissions Comparison Model Results

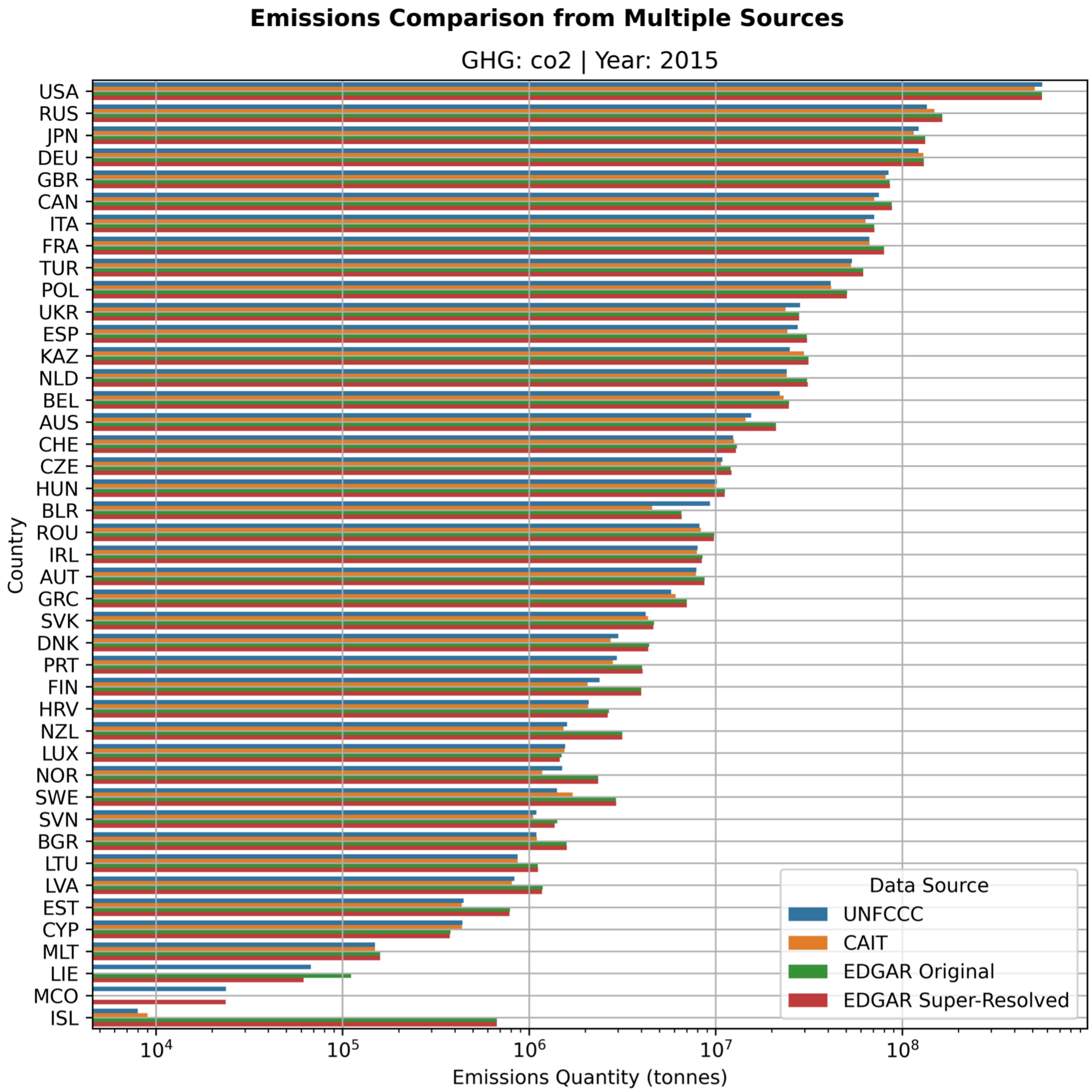

Figure S.1. We compare the quantity of CO_2 emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2015. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

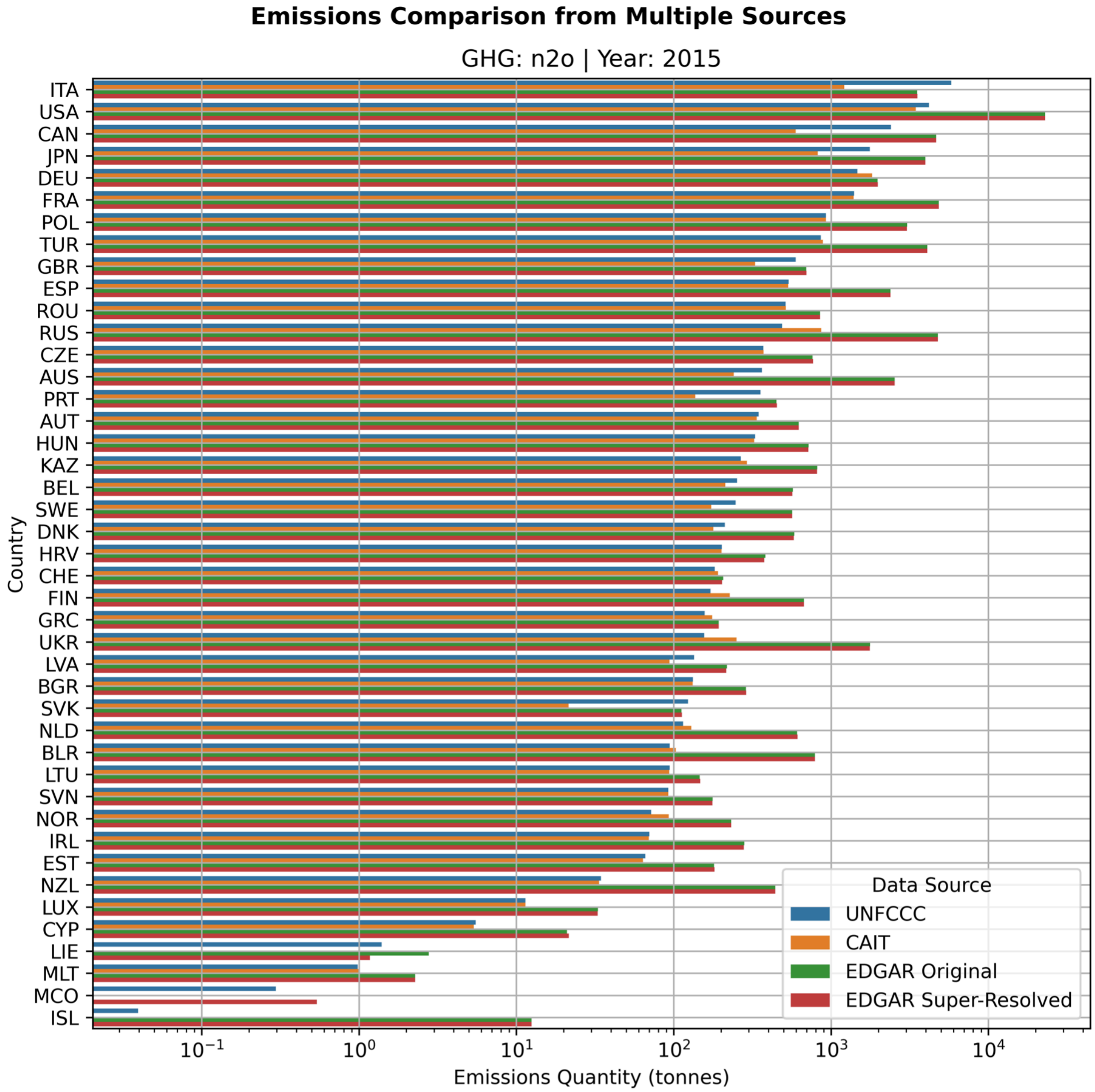

Figure S.2. We compare the quantity of N_2O emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2015. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

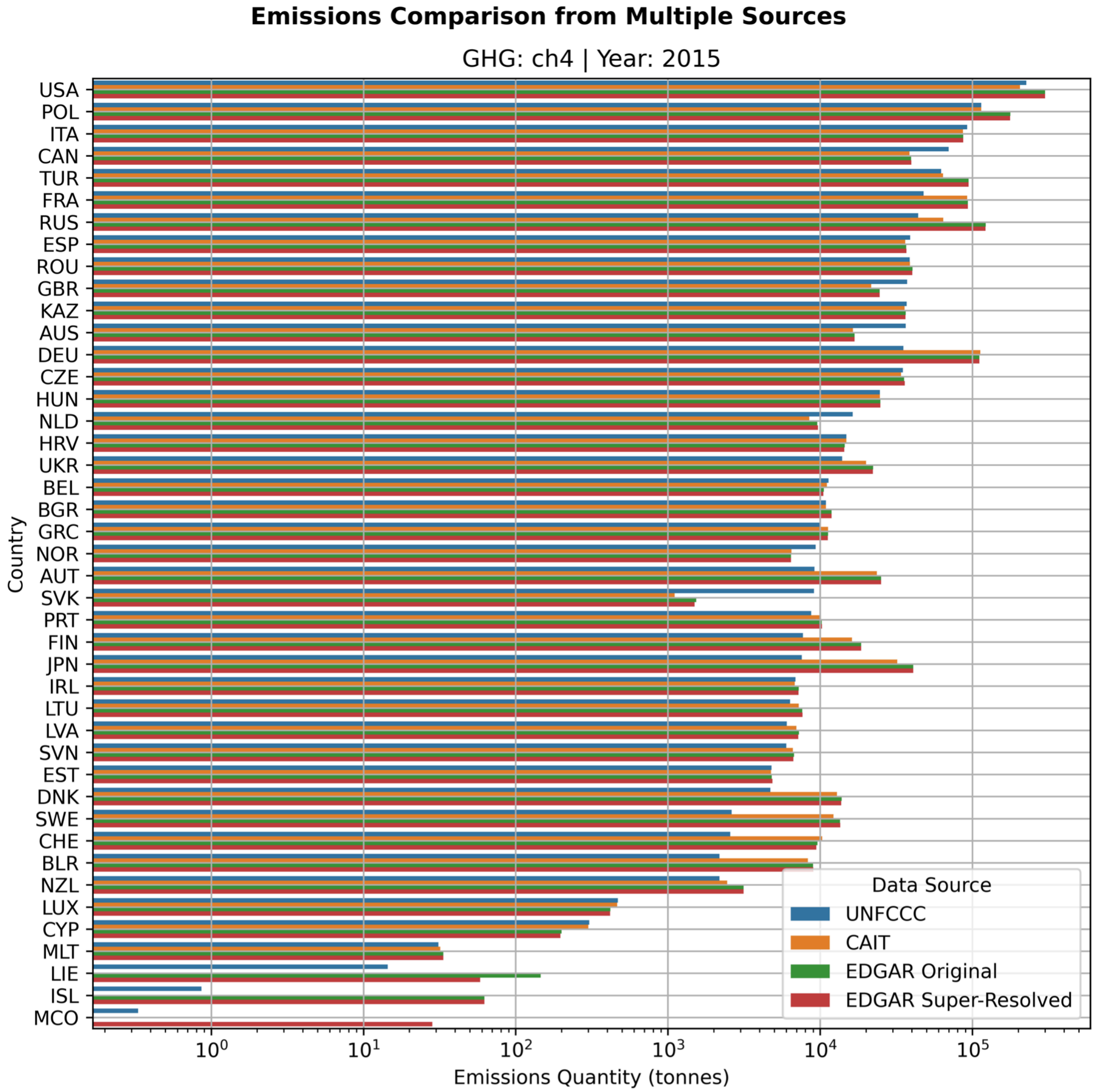

Figure S.3. We compare the quantity of CH_4 emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2015. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

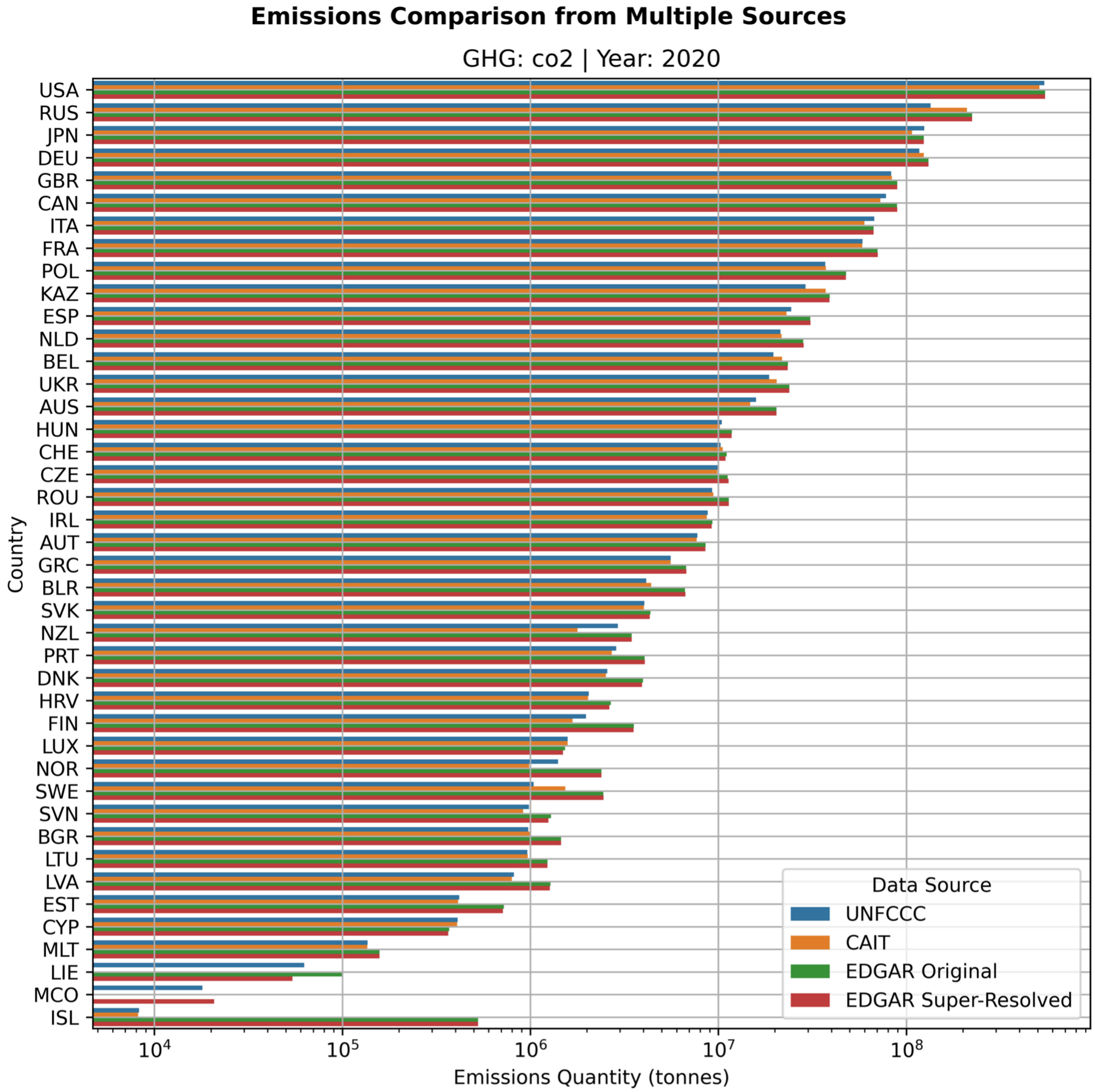

Figure S.4. We compare the quantity of CO_2 emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2020. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

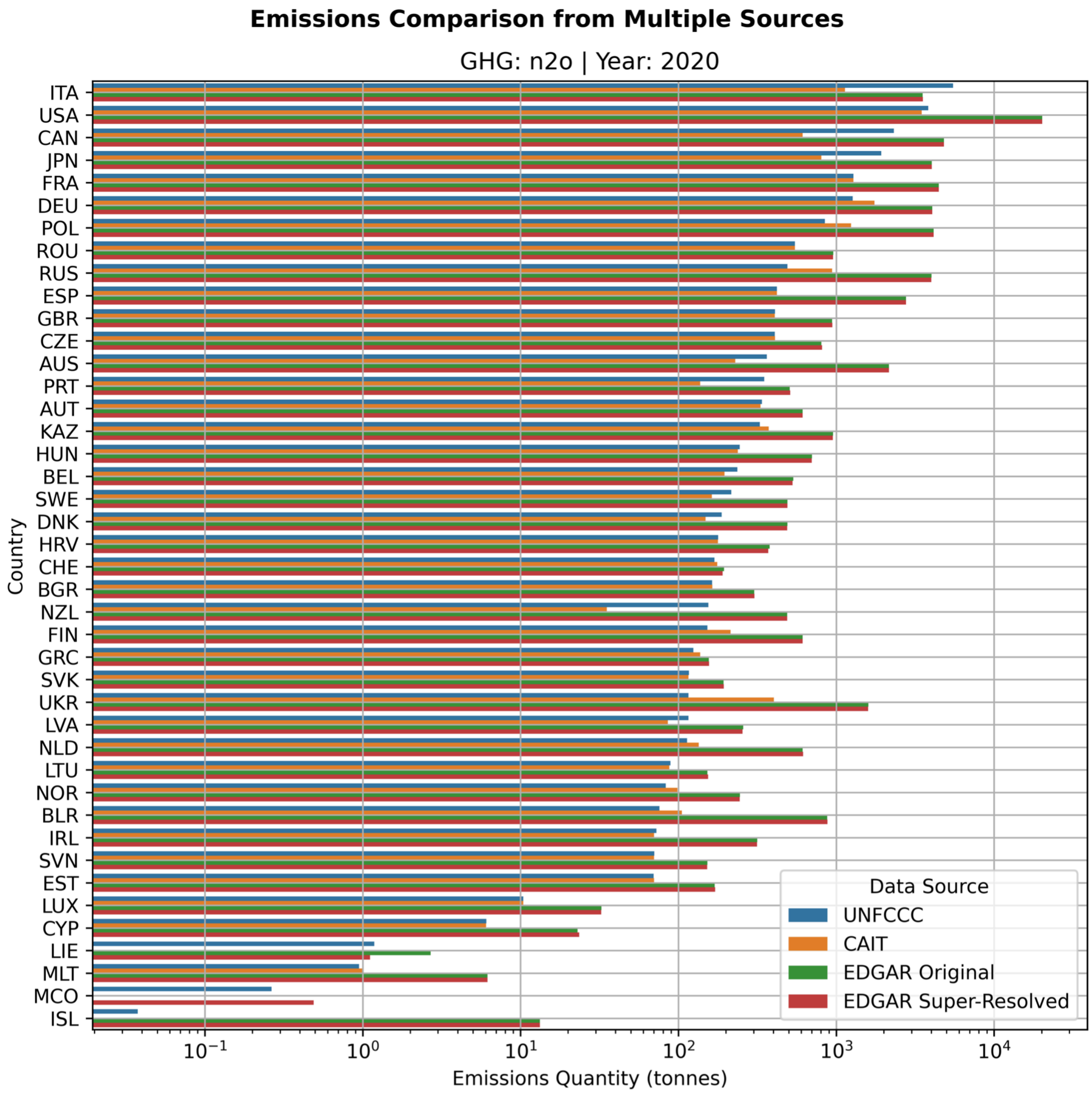

Figure S.5. We compare the quantity of N_2O emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2020. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

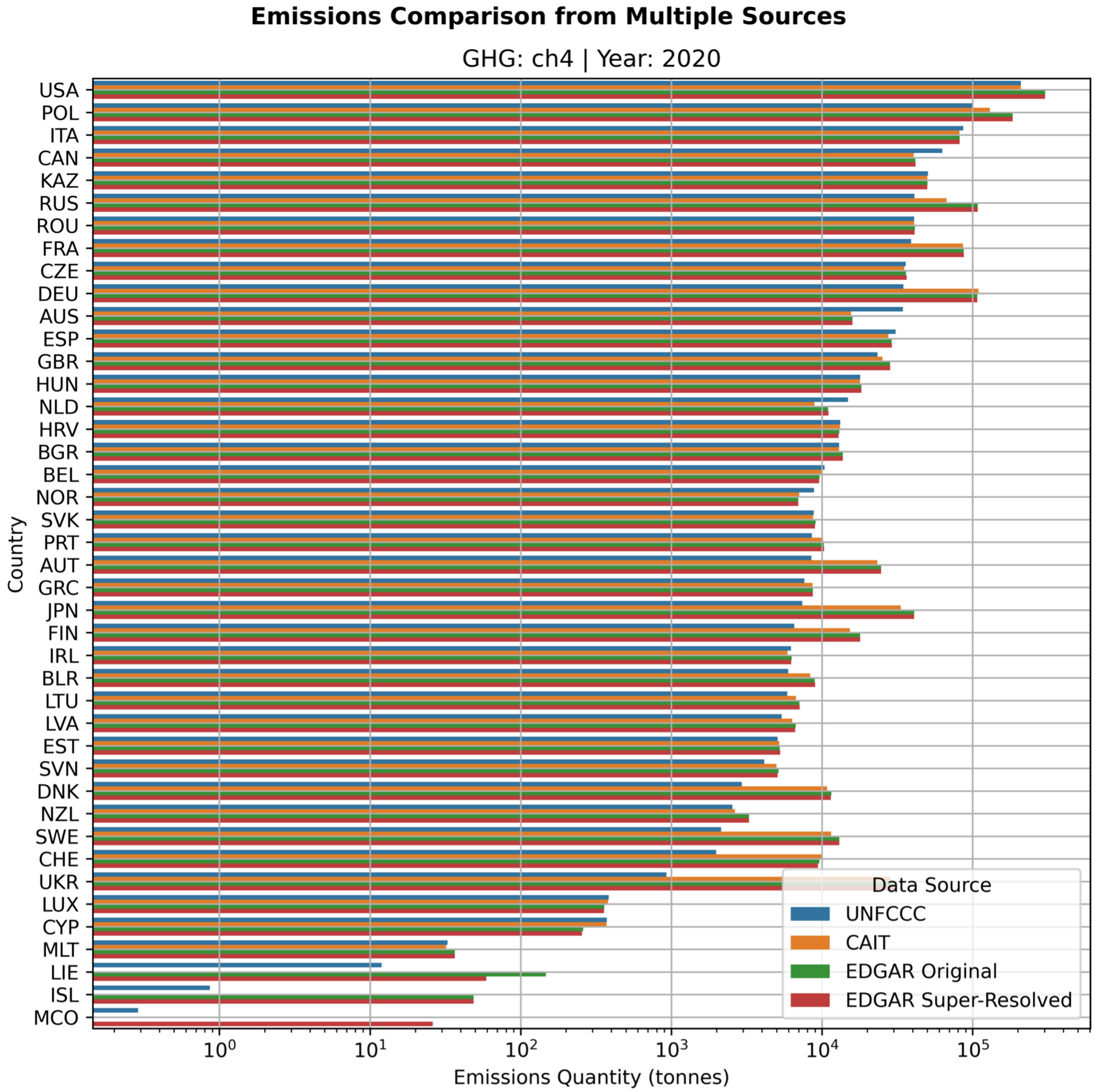

Figure S.6. We compare the quantity of CH_4 emissions (in tonnes) across each of the countries included in the UNFCCC inventory for 2020. We include emissions inventory data from UNFCCC and CAIT for each such country, as well as emissions estimates from the original EDGAR v8 as well as our model's super-resolved version of EDGAR.

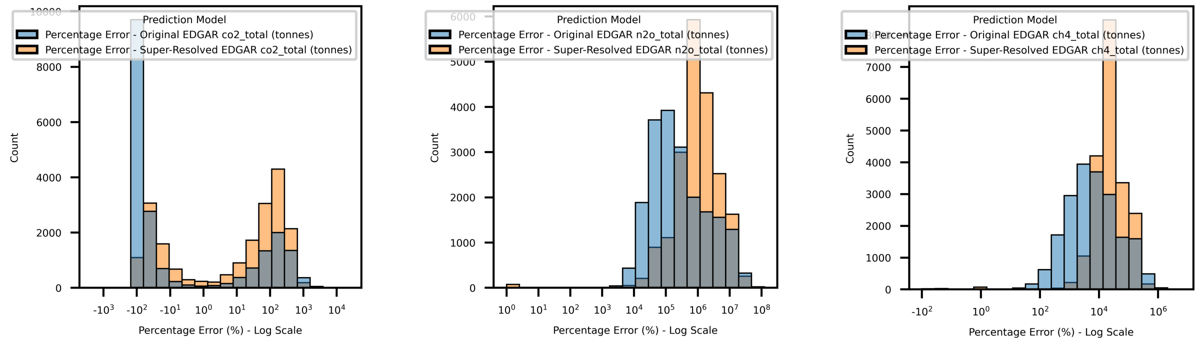

Figure S.7. Histograms showing the percentage error (on log-10 scale) of emissions estimation error between original EDGAR v8 and our model's super-resolved version of EDGAR.

S.4. Known issues

S.4.1. Energy Use Intensity

To address challenges with the CURB Tool data, Agartala (India) was removed since there are no residential sector values in the original data, duplicate values for Juba (labeled "Sudan" for "South Sudan") from CURB tool dataset were removed, and Panama City's (Panama) data was removed as there was a likely error with the hotel EUI datapoint that impacted the non-residential data totals.

S.4.2. Final Estimates

In the Democratic Republic of the Congo, we are aware of possible missing data from the EDGAR data that impacts our estimates and may lead to underestimation in this location in the EDGAR super-resolved version of the model.

S.4.3. Inclusion/Exclusion of Biomass

Biomass (e.g., wood) is a fuel type used across the building energy sector and - particularly for N2O and CH4 - has a disproportionately large emissions factor as compared to other fuel types. However, biomass is not consistently included (or excluded) by default in various inventories and emissions estimation models. Because these sources aggregate emissions from biomass and other fossil fuels into a single emissions estimate, we cannot choose to exclude biomass by default - nor can we add it to sources that contain no biomass-related data. This is especially evident as part of our comparison between DPFC and EDGAR: DPFC does not contain biomass data from building energy consumption, whereas the gridded sector-specific version of EDGAR v8 (and our model's super-resolved derivative of those data) both include biomass as part of an aggregated emissions estimate for N_2O and CH_4.

S.4.4. Verifying modeled emissions estimates

EDGAR's Energy for Buildings data includes not only residential and commercial buildings. As EDGAR's Energy for Buildings data uses data from the IPCC's 1996 (1A4) and 2006 codes (1A4+1A5) on 'Other Sectors,' this includes agricultural off-road vehicles (1.A.4.c.ii), fishing (1.A.4.c.iii), and marine vessels (1.A.5.b) that do not exactly align with our data estimates. From our review, these are small values that have a minor impact on estimates, but is key to note as we compare our model to EDGAR for validation.

S.4.5. Municipal-Level Validation Boundary Summary

For municipal-level validation, we discovered that there is not a consistent set of geospatial boundaries from a common source at a corresponding municipality level. GADM is an open-source project that "wants to map the administrative areas of all countries, at all levels of sub-division" (GADM, 2024) and our primary source for country-level boundaries represented by GADM Level 0. While GADM tends to have near-complete global coverage within their data at the Level 0 administrative subdivision representing each nation, their data was not globally comprehensive at the municipality level. For example, the GADM Level 2 boundaries for the United States were not consistent with the size of each municipality contained within the United States DPFC data, and GADM does not contain municipality-level boundaries for the United States at a level corresponding to a more detailed subdivision.

To conduct our model validation at the municipal level, we identified distinct geospatial boundary files for each municipal region under consideration using authoritative data sources for each region. We found that GADM boundaries were sufficient for Chile, Canada, Costa Rica, and Denmark. For Mexico, Japan, and India, we identified corresponding municipality-level boundaries from the Humanitarian Data Exchange (The Humanitarian Data Exchange, 2024). In the United States, we were required to identify a boundary file independently for each of the 50 states in the country from authoritative state government websites. The table below summarizes the source for each such set of boundaries.

Table S.5. Boundary Sources for Municipality-Level Validation

Country	Level for municipalities	Boundary Source
Mexico	Level 2	Humanitarian Data Exchange
Japan	Level 2	Humanitarian Data Exchange
Chile	Level 3	GADM
Canada	Level 3	GADM
Costa Rica	Level 2	GADM
India	Level 5	Humanitarian Data Exchange
Denmark	Level 2	GADM
United States	N/A	State Government Boundaries

Data citation format: Markakis, P., Gowdy, T.M., Shen, Z., Lancellotti, B., Malof, J., Bradbury, K. (2024). *Building Sector- Estimating Global, High-resolution Onsite Building Emissions.* Duke University and University of Missouri, USA, Climate TRACE Emissions Inventory. https://climatetrace.org [Accessed date]

Geographic boundaries and names (iso3_country data attribute): The depiction and use of boundaries, geographic names and related data shown on maps and included in lists, tables, documents, and databases on Climate TRACE are generated from the Global Administrative Areas (GADM) project (Version 4.1 released on 16 July 2022) along with their corresponding ISO3 codes, and with the following adaptations:

- HKG (China, Hong Kong Special Administrative Region) and MAC (China, Macao Special Administrative Region) are reported at GADM level 0 (country/national);
- Kosovo has been assigned the ISO3 code 'XKX';
- XCA (Caspian Sea) has been removed from GADM level 0 and the area assigned to countries based on the extent of their territorial waters;
- XAD (Akrotiri and Dhekelia), XCL (Clipperton Island), XPI (Paracel Islands) and XSP (Spratly Islands) are not included in the Climate TRACE dataset;
- ZNC name changed to 'Turkish Republic of Northern Cyprus' at GADM level 0;
- The borders between India, Pakistan and China have been assigned to these countries based on GADM codes Z01 to Z09.

The above usage is not warranted to be error free and does not imply the expression of any opinion whatsoever on the part of Climate TRACE Coalition and its partners concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its borders.

Disclaimer: The emissions provided for this sector are our current best estimates of emissions, and we are committed to continually increasing the accuracy of the models on all levels. Please review our terms of use and the sector-specific methodology documentation before using the data. If you identify an error or would like to participate in our data validation process, please contact us.

References

1. Crippa, M., Guizzardi, D., Pagani, F., Schiavina, M., Melchiorri, M., Pisoni, E., Graziosi, F., Muntean, M., Maes, J., Dijkstra, L. and Van Damme, M., 2023. 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 System Science Data, 16(6), pp.2811-2830.

2. Dou, X., Wang, Y., Ciais, P., Chevallier, F., Davis, S.J., Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Solazzo, E., Yan, F. and Huo, D., 2022. Near-real-time global gridded daily CO2 emissions. The Innovation, 3(1).
3. Energy Star, 2023. U.S. Energy Use Intensity by Property Type. https://portfoliomanager.energystar.gov/pdf/reference/US%20National%20Median%20Table.pdf
4. Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S. and Strano, E., 2017. Breaking new ground in mapping human settlements from space–The Global Urban Footprint. ISPRS Journal of Photogrammetry and Remote Sensing, 134, pp.30-42.
5. European Commission Joint Research Centre T1 Unit (2024) INSPIRE Registry. Available at: https://inspire.ec.europa.eu/codelist/HILUCSValue/5_ResidentialUse:1 (Accessed: 30 October 2024).
6. GADM maps and Data (2024) GADM. Available at: https://gadm.org/ (Accessed: 30 October 2024).
7. Geonames (2023): https://www.geonames.org/
8. Global Covenant of Mayors for Climate & Energy, "Data Portal for Cities." Accessed: Sep. 28, 2023. [Online]. Available: https://dataportalforcities.org/
9. IEA (no date) Buildings - Energy System, IEA. Available at: https://www.iea.org/energy-system/buildings (Accessed: 30 October 2024).
10. Intergovernmental Panel on Climate Change (IPCC). 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 2, Energy: Chapter 2, Stationary Combustion. Prepared by Gómez, D.R., Watterson, J.D., Americano, B.B., Ha, C., Marland, G., Matsika, E., Namayanga, L.N., Osman-Elasha, B., Kalenga Saka, J.D., and Treanton, K. IGES, Japan.
11. International Energy Agency (2021) Net Zero by 2050: Net Zero by 2050 Scenario - Data product - IEA; as modified by Duke University. https://www.iea.org/data-and-statistics/data-product/net-zero-by-2050-scenario#overview
12. Liu, E., Meng, C., Kolodner, M., Sung, E.J., Chen, S., Burke, M., Lobell, D. and Ermon, S., 2023. Building coverage estimation with low-resolution remote sensing imagery. arXiv preprint arXiv:2301.01449.
13. Luers, A., Yona, L., Field, C.B., Jackson, R.B., Mach, K.J., Cashore, B.W., Elliott, C., Gifford, L., Honigsberg, C., Klaassen, L. and Matthews, H.D., 2022. Make greenhouse-gas accounting reliable—build interoperable systems. nature, 607(7920), pp.653-656.
14. Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A. and Paganini, M., 2020. Outlining where humans live, the World Settlement Footprint 2015. Scientific Data, 7(1), p.242.

15. McDuffie, E.E., Smith, S.J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E.A., Zheng, B., Crippa, M., Brauer, M. and Martin, R.V., 2020. A global anthropogenic emission inventory of atmospheric pollutants from sector-and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS). Earth System Science Data, 12(4), pp.3413-3442.
16. Oda, T., Maksyutov, S. and Andres, R.J., 2018. 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 System Science Data, 10(1), pp.87-107.
17. Pearce, F., 2024. Nations Are Undercounting Emissions, Putting UN Goals at Risk. Yale Environment 360.
18. Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L. and Marin-Herrera, M.A., 2013. A global human settlement layer from optical HR/VHR RS data: Concept and first results. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(5), pp.2102-2131.
19. The Humanitarian Data Exchange (2024) Welcome - Humanitarian Data Exchange. Available at: https://data.humdata.org/ (Accessed: 30 October 2024).
20. Tong, D., Zhang, Q., Davis, S.J., Liu, F., Zheng, B., Geng, G., Xue, T., Li, M., Hong, C., Lu, Z. and Streets, D.G., 2018. Targeted emission reductions from global super-polluting power plant units. Nature Sustainability, 1(1), pp.59-68.
21. U.S. Energy Information Administration (2015). Residential Energy Consumption Survey (RECS). Accessed 2023. https://www.eia.gov/consumption/residential/.
22. U.S. Energy Information Administration (2018). Commercial Building Energy Consumption Survey (CBECS). Accessed 2023. https://www.eia.gov/consumption/commercial/.
23. Umemiya, C. and White, M.K., 2024. National GHG inventory capacity in developing countries–a global assessment of progress. Climate Policy, 24(2), pp.164-176.
24. UNFCCC, “GHG data from UNFCCC | UNFCCC.” Accessed: Jun. 26, 2024. [Online]. Available: https://unfccc.int/topics/mitigation/resources/registry-and-data/ghg-data-from-unfccc
25. Washington, DC: World Resources Institute, “Climate Watch.” 2024. Accessed: Jun. 26, 2024. [Online]. Available: https://www.climatewatchdata.org
26. World Bank (2016). CURB: Climate Action for Urban Sustainability tool. Accessed 2023. https://datacatalog.worldbank.org/search/dataset/0042029.