

# Transportation Sector: Global Road Emissions

Kevin Foster<sup>1,2</sup>, Patrick Sicurello<sup>1,2</sup>, Christine Piatko<sup>1,2</sup>,  
and Elizabeth Reilly<sup>1,2</sup>



<sup>1</sup>*The Johns Hopkins University Applied Physics Laboratory (JHU/APL)*

<sup>2</sup>*The Climate TRACE Coalition*

## 1 Introduction

Transportation contributed 27% of anthropogenic greenhouse gas (GHG) emissions in the United States of America (USA) for 2020, higher than any other sector, and 12.6% of all global GHG emissions in 2019 (US EPA 2023; World Resources Institute 2022). The primary source of transportation sector emissions are on-road vehicles, accounting for approximately 74% of global transportation emissions in 2018 (International Energy Agency 2019). Quantifying the distribution of on-road transportation emissions and creating timely emissions inventories are vital to identify trends, track mitigation efforts, and inform policy decisions.

Previous efforts have developed detailed bottom-up on-road emission inventories for the USA (C. Gately, L. Hutyra, and I. Wing 2019; Gurney et al. 2020), but do not easily extend globally due to the reliance on vehicle traffic and road data that is not always readily available. The Emissions Database for Global Atmospheric Research (EDGAR) provides a global inventory for transportation that uses road density as a proxy to spatially distribute emissions (M. Crippa et al. 2025). However, some emission estimates for urban centers in EDGAR deviated from other bottom-up inventories by 500%, indicating that road density is not a sufficient proxy for global high-resolution inventories (C. K. Gately, L. R. Hutyra, and I. S. Wing 2015). Carbon Monitor is a global emissions inventory that utilizes a variety of activity data to estimate daily GHG emissions, however the reliance on proprietary traffic data in the ground transportation sector limits the ability to extend to locations where this data is not available (Liu et al. 2020). Other methods have used machine learning (ML) to directly predict emissions, but their ability to generalize globally is unclear (Mukherjee et al. 2021; Scheibenreif, Mommert, and Borth 2021).

JHU/APL has developed an approach to estimate road transportation emissions, and applied this methodology globally. Our “hybrid” algorithm leverages the strengths of ML applied to remote sensing data, in addition to incorporating region-specific emissions factors (EFs) data to create scalable and transparent emissions estimations globally. A detailed description of our method, data sources, and validation results is contained herein.

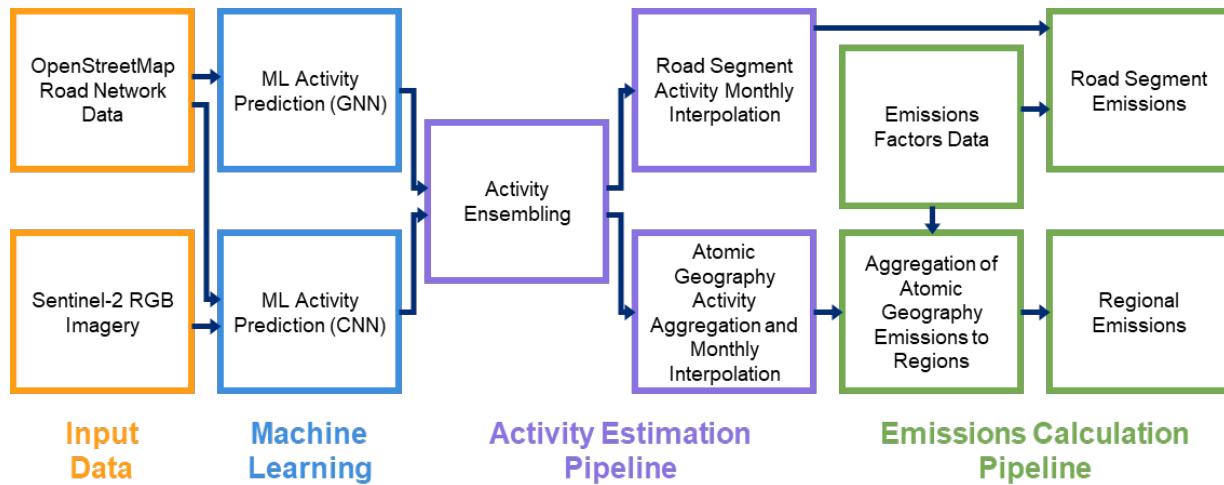
Additionally, an initial emission reduction strategy (ERS) for this sector was analyzed and quantified. A primary opportunity to reduce emissions within the road transportation sector is the adoption of battery-powered electric vehicles (BEVs). These vehicles, instead of burning fuel directly for motive power, store and use energy drawn from the local electrical grid, which can be composed of lower-emitting power sources (e.g., solar, wind, etc.) resulting in a net emission reduction. The analyzed strategy focuses on the effects if all current gasoline fueled internal combustion engines

(ICEs) passenger vehicles are switched to BEV. While gains can be realized by switching other vehicle types to BEV as well (e.g., buses and commercial vehicles), the technology for passenger vehicles is the most mature and widespread.

## 2 Data and Methods

### 2.1 Overview

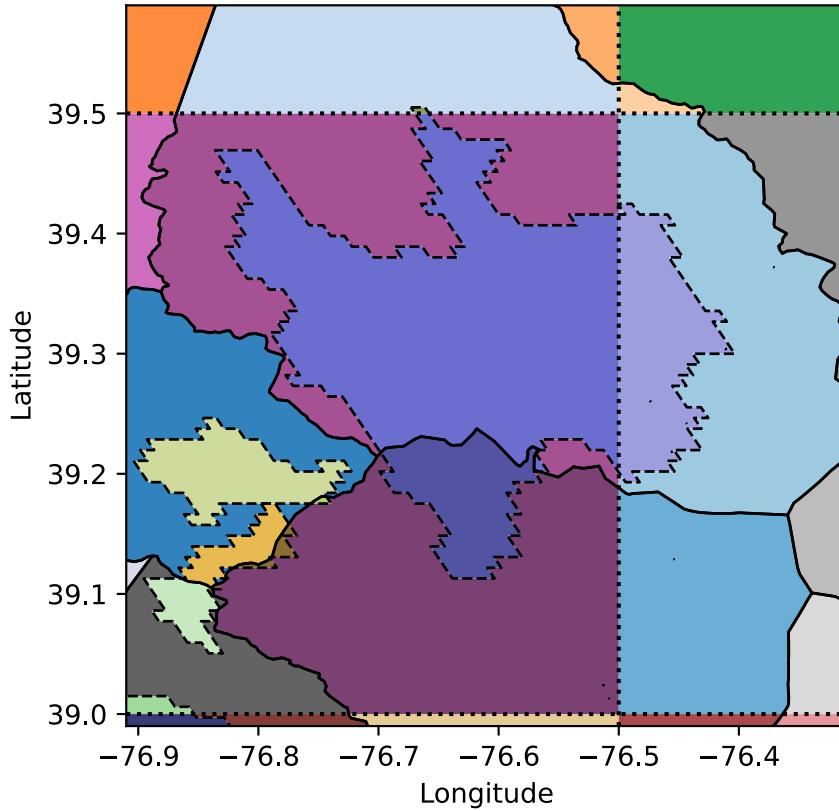
Our hybrid emissions estimation method was primarily composed of three parts: a set of ML models trained to predict road transport activity, an activity estimation pipeline that ensembles, aggregates, and interpolates the ML results, and an emissions factors (EFs) pipeline that converts activity predictions to emissions estimates. This approach combines the strengths of satellite imagery and ML with traditional “bottom-up” emissions inventories that directly incorporate vehicle fleet mix, fuel efficiency, and other EF data. Separating the machine learning and activity estimation from the emission factors pipelines affords continuous improvement of each as newer and better data become available. A high-level system architecture is shown below in Figure 1. A separate method is used to backproject these estimates to historical time periods (2015-2020; see Section 2.3.9).



*Figure 1* 2021-2024 emissions estimation architecture overview.

Quarterly remote sensing and geospatial data were fed to two different families of ML models, convolutional neural network (CNN) based models and graph neural network (GNN) based models, producing two separate average annual daily traffic (AADT) estimations on a per-road segment basis. Separate models from each family are used for rural and urban areas, as defined by the European Union Joint Research Center Global Human Settlement Layer Urban Centers Database (GHSL-UCDB) dataset for a globally consistent representation of urban center extent (Florczyk et al. 2023); see Section 2.2.1. Output values from the two different model families were ensembled via averaging within each quarter, and then averaged across quarters within a year to produce a yearly AADT estimate for each road segment (Section 2.4.5). The annual AADT estimate is

then converted to the activity value of total vehicle kilometers traveled (VKT) using the known length (capacity) of each road segment. For computational efficiency reasons, the pipeline is then split into two; in one pipeline, the activity estimates per-road segment are interpolated to monthly values, while in the other the activity estimates and capacities are summed per road type within each atomic geography prior to interpolation. Atomic geographies are non-overlapping geometries that cover the populated landmasses of the world, whose boundaries are the unique combinations of GADM Level-2 boundaries, GHSL-UCDB urban center boundaries, and a half degree latitude/longitude grid; an example of atomic geographies are shown in Figure 2. EFs were computed on a per-atomic geography basis a priori from environmental factors and a curated database of road and vehicle-related factors, assigning EF values to each type of road in an area; Sections 2.2.6 through 2.2.10 describe the data sources used for the factors. Activity values were then multiplied by the appropriate EF for each road type to produce emissions estimates. Emission estimates for the atomic geographies were then aggregated to regions (GADM Levels 0-2) to produce monthly emission estimates from 2021 onward. This process is explained in detail in Section 2.3. Additionally, estimates for 2015-2020 were made by leveraging EDGAR v2025 GHG data (M. Crippa et al. 2025) (Section 2.6) to backproject estimates as discussed in Section 2.3.9. ERS estimates were made by applying BEV efficiencies by market segment (e.g. small car, large SUV) to the per-market segment sales figures for gasoline-powered passenger vehicles; this is described more fully in Section 2.7.



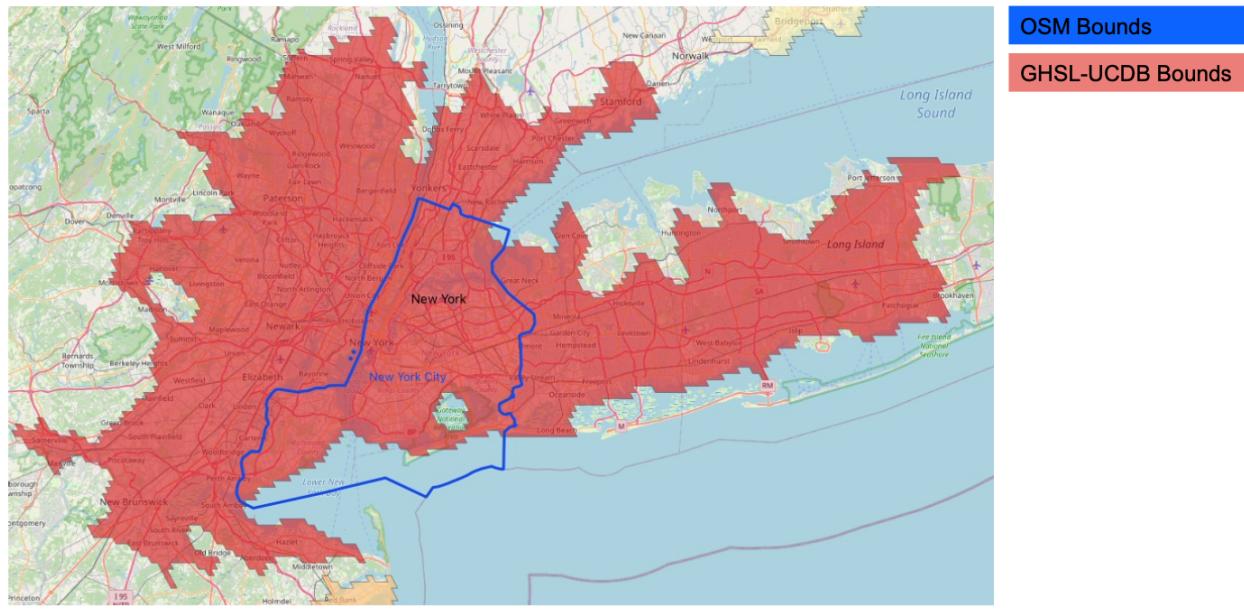
**Figure 2** Example atomic geographies in the Baltimore, Maryland region of the USA. Each color is a different atomic geography; solid lines are GADM Level-2 boundaries, dashed lines are GHSL-UCSB urban center boundaries, and dotted lines are half degree latitude/longitude grid boundaries.

## 2.2 Data

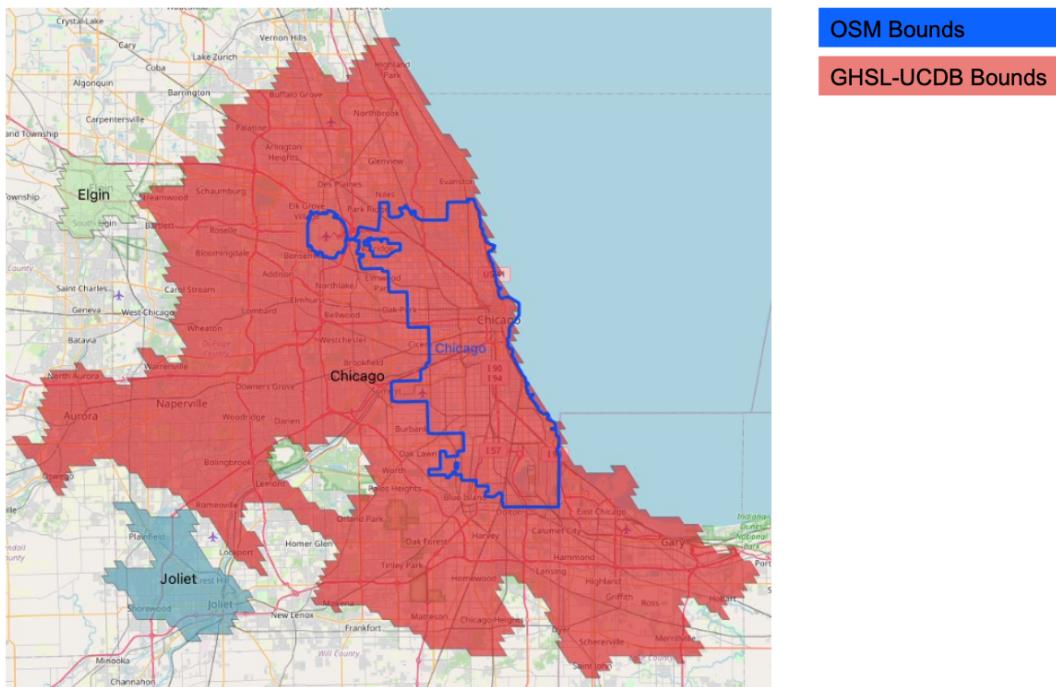
### 2.2.1 Urban Centers

We utilized the European Union Joint Research Center Global Human Settlement Layer Urban Centers Database (GHSL-UCDB) dataset for a globally consistent representation of urban center extent (Florczyk et al. 2023). This database contains approximately 13,000 urban centers worldwide, and utilizes a definition of urban center based on population density and built-up area. Specifically, an urban center was defined as “the spatially-generalized high-density clusters of contiguous grid cells of 1 km<sup>2</sup> with a density of at least 1,500 inhabitants per km<sup>2</sup> of land surface or at least 50% built-up surface share per km<sup>2</sup> of land surface, and a minimum population of 50,000.” (ibid.). Due to this definition, urban center geometries in UCDB often have significantly different shapes and sizes as compared to official administrative bounds of cities, e.g., from OpenStreetMap (OpenStreetMap Contributors 2020) or Global Administrative Areas (GADM) (Global Adminis-

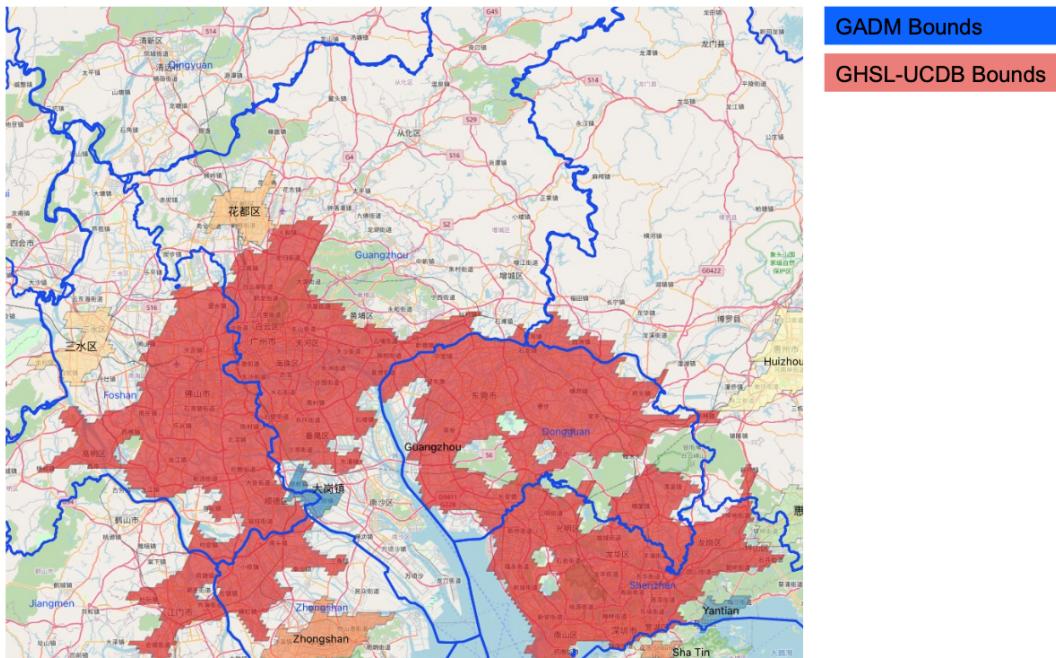
trative Areas n.d.). Examples of these differences are shown below in Figure 3-Figure 5.



**Figure 3** Comparison of OSM administrative bounds (blue boundary) and GHSL-UCDB city bounds (red area) for New York City, USA.



**Figure 4** Comparison of OSM administrative bounds (blue boundary) and GHSL-UCDB city bounds (red area) for Chicago, USA.



**Figure 5** Comparison of GADM level 2 administrative boundaries (blue boundaries) and GHSL-UCDB urban center bounds (red area) for Guangzhou, China.

## **2.2.2 Regional Administrative Boundaries**

The Climate TRACE coalition decided to assign emissions based on Global Administrative Areas (GADM) (Global Administrative Areas 2022) at the country (Level 0), state/province (Level 1) and county (Level 2) levels. Here, “county” is the USA based English-version name of the typical administrative area. In other countries, other English-translated names are used. The coalition used a slightly modified version of GADM v4.1; details are described in Section 7.3.

## **2.2.3 Visual Satellite Imagery**

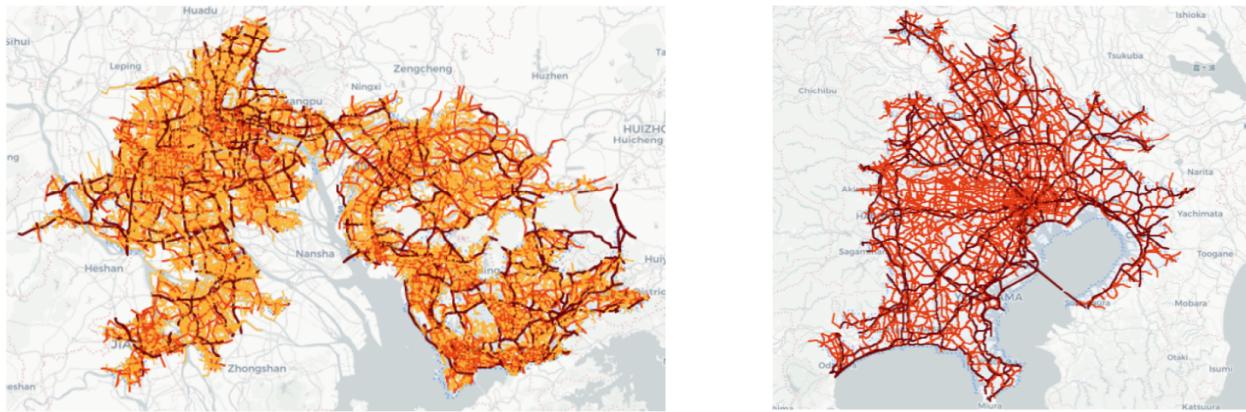
Remote sensing imagery from the Sentinel-2A/B satellites were used as input data in our ML modeling approach to predict road transportation activity (see Section 2.3 below). The European Space Agency’s (ESA) Sentinel-2 mission comprises two satellites- Sentinel-2A, launched in 2015, and Sentinel-2B, launched in 2017 (Main-Knorn et al. 2017). Each Sentinel-2 satellite has a 10-day revisit time with a 5-day combined revisit. Both satellites are equipped with a multispectral (MSI) instrument which provides 13 spectral band measurements, blue to shortwave infrared (SWIR) wavelengths ( 442 nm to 2202 nm) reflected radiance. We used the Sentinel-2 Level-2A product at 10 m x 10 m resolution, using bands 4 (red), 3 (green), and 2 (blue) (Drusch et al. 2012).

Sets of images are selected on a per Sentinel-2 Level-2A tile (or granule) basis every quarter, with preference given to unclouded images that best span the tile, with additional unclouded images selected as needed to bring the tile coverage to 90%. If there are insufficient unclouded images, an additional image with up to 10% cloud coverage may be used to improve the tile coverage.

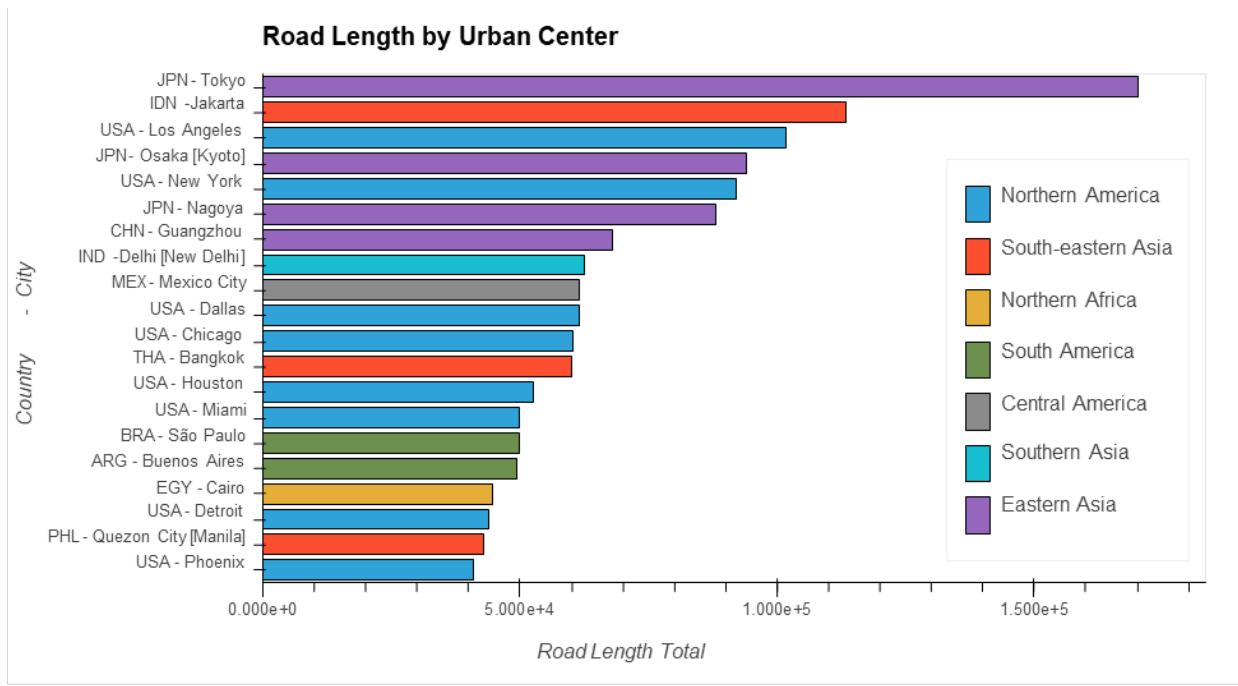
## **2.2.4 Road Network Data**

Using the open-source tools Osmium (Topf 2023) and OSMnx (Boeing 2017) and custom filters, OpenStreetMap data was filtered to the set of roads that carry normal vehicular traffic (see Section 2.2.6 below for details). A multi-directed graph of the road network was created from that data, and the total edge length (meaning the road length for each direction of traffic is counted separately) was computed to yield the capacity, or road network length, in kilometers. Example road networks for select GHSL UCDB urban centers are visualized in Figure 6. A plot of road network length for differing urban centers is provided in Figure 7; only the top 20 GHSL UCDB urban centers are displayed for visual clarity, though OSM produces road network data outside of urban centers, as well.

Roads and associated metadata (including coordinates, number of lanes, etc.) are constantly changing around the world. OpenStreetMap is updated continuously in an attempt to keep an atlas that reflects reality. While most road segments are consistent between quarters, some road segments appear, some road segments disappear, and some have properties that change. We attempt to deal with these changes by using one snapshot of data from OpenStreetMap per quarter per year. For our activity estimates, we use model inference ensembling to combine four quarters of data; more details are given in Section 2.4.5.



**Figure 6** Example road network data for Guangzhou, China (left) and Tokyo, Japan (right). Highways are in dark red, arterial roads in orange, and local roads in yellow. Local roads are intentionally not displayed for Tokyo due to their large count.



**Figure 7** Total road length for the top 20 GHSL UCDB urban centers (“cities”) colored by region.

## 2.2.5 Road Segment Type

Road segment types (categories) were derived from OpenStreetMap data for all public roads. The current supported road types are highway, arterial, and local, which were chosen to align with other similar emissions inventories and traffic-related databases. The mapping between these road types and their respective OSM tags is provided in Table 1. Road type categorization is important in the emissions factor calculation for a given road segment as other emissions factors variables, including vehicle fleet mix and fuel efficiency, could vary significantly across different types of roads. These classes are also used within the modeling effort to distinguish roads (see Section 2.4.2).

*Table 1* Road segment type association with OpenStreetMap tags

Road Class	OpenStreetMap Tags
Highway	motorway, motorway_link, trunk, trunk_link
Arterial	primary, primary_link, secondary, secondary_link
Local	tertiary, tertiary_link, residential, living_street, unclassified

## 2.2.6 Emissions Factors

Calculating emissions factors for a given area was a complex process, due to the fact that an area can actually contain potentially hundreds of thousands of individual road segments, which might be considered “sub-sources”. Transportation emissions factors are dependent on many variables, including (but not limited to) road category, vehicle type, fuel type, nominal fuel efficiency and environmental factors. Data collection for each of these variables across the globe was a significant undertaking. Thus, this version of estimated emissions factors focused on collecting data at the country level. A detailed description of the emissions factor calculation is provided in Section 2.3 below. Sources for each type of data required for the emissions factor calculation are shown below in Table 2, with descriptions of each data type provided in subsequent sections.

**Table 2 Primary emissions factors data sources used.**

Data Type	Source(s)
Road Segment Type	OpenStreetMap (OpenStreetMap Contributors 2020), see Section 2.2.5 above
Vehicle Fleet Mix	Various, see additional information in Section 2.2.7 and Section 6.1
Fuel Type	CURB (World Bank Group 2019), see Section 2.2.8 below
Nominal Fuel Efficiencies	CURB (World Bank Group 2019), see Section 2.2.8 below
Monthly Gridded Temperatures	Climatic Research Unit gridded Time Series v4.07 (Harris et al. 2020), See Section 2.2.9 below
GHG Emissions Factors	U.S. EPA GHG Emissions Factors Hub (US EPA 2024), see Section 2.2.10 below
Non-GHG Emissions Factors	U.S. DOT Bureau of Transportation Statistics (US BTS 2024), see Section 2.2.10 below

## 2.2.7 Vehicle Fleet Mix

Vehicle fleet mix refers to the distribution of total vehicles in a given country across various vehicle types. The supported vehicle types were: passenger cars, light duty trucks, single unit trucks, combination trucks, motorcycles, and buses. Country-specific vehicle distribution numbers do not always share these same categories. There might be one, two or more country-specific categories mapping to one supported type or there might be no country-specific category that matches a supported type. These differences were dealt with on a per-country basis. For example, in countries which only reported “trucks” as a single category, we used regional averages of the percentage of trucks in each category to interpolate the single reported category into the three standard truck types.

In general, there are two categories of information available on vehicle fleet mixes: registration data and proportions of kilometers traveled by vehicle category. Our vehicle fleet mix encompasses both types of data and does not distinguish between them in accuracy.

For example, the U.S. Federal Highway Administration (FHWA) provides estimates of both registered cars by state in Table MV-1 (US FHWA 2018) and vehicle-miles traveled in Table VM-4 (US FHWA 2020). These differ in the percentages of vehicles in each standard type. We would argue that the former (vehicle-miles traveled) is the better distribution of vehicle types to use, as we are attempting to estimate emissions from vehicles actually traveling on roads. However, we have not found such specific information for many countries. Instead, we use the following prioritized list to estimate the vehicle mix in each country:

1. Distribution of vehicle-km traveled categories: as described above
2. Distribution of registration types: as described above
3. Neighboring country average of vehicle fleet mix: use the average vehicle fleet mix from any countries whose GADM Level 0 border touches the target country
4. US FHWA estimate: when all else fails, use the US FHWA estimate for the USA

Country-specific vehicle fleet data was found for 118 countries. A full listing of the countries with country-specific data and their respective sources is provided in the supplementary material (see Section 6.1). Vehicle fleet mix values are currently the same across all supported road types but will be updated as sources of road type-specific data are identified.

For each country with specific vehicle fleet data, when sufficient samples were available, the data was linearly interpolated / extrapolated to cover the modeled time period; otherwise, the data is forward-/backward-filled to cover the time period.

For countries without vehicle specific fleet data, fleet estimates were averaged from any neighboring countries with available data or, failing that, from an average across the UN defined region for the country. For the UN regions of Micronesia and Polynesia, no country data was available, so vehicle fleet data from Melanesia was used as a substitute.

## 2.2.8 Fuel Type and Nominal Efficiencies

Due to the fact that different fuel types have different emissions factors, it is important to know the relative mix of fuel types for each type of vehicle traveling on a given road segment. The types of supported fuels are:

- Gasoline: no distinction is made between different sub-categories such as (in the US) 87, 89, and 91 octane gasoline or ethanol-free gasoline.
- Diesel: no distinction is made between sulfur-free diesel, bio-diesel or other sub-categories
- Compressed natural gas (CNG)
- Liquefied petroleum gas (LPG)
- Plug-in hybrid: no distinction is made between hybrids that run on different liquid fuels
- Battery electric vehicle (BEV)
- Other fuels (e.g., biogas, ethanol)

The primary source of this data is the Climate Action for Urban Sustainability (CURB) tool (World Bank Group 2019), which provides a global database of fuel type mix by country. Future updates may include updated country or city-specific fuel type data. CURB was also the primary source

of nominal fuel efficiency data for all countries. CURB fuel efficiency values are reported in units of kilometers per liter and were extracted for all supported fuel and vehicle types described above. Fuel efficiencies were the same across all supported road types (highway, arterial, and local) in this release, but may be continuously updated as better country or region-specific datasets are located.

## 2.2.9 Fuel Efficiency Environmental Adjustments

Fuel consumption varies under external environmental factors. A study by Lohse-Busch et al. 2013, conducted several tests investigating fuel consumption for several conventional vehicles. These tests included:

- Urban and highway driving tests at 72° F
- Cold urban driving tests at 20° F
- Hot urban driving tests at 95° F and “sunny” skies

These tests broadly found that fuel consumption increases at low temperatures due to heater use and cooler fluids, and fuel consumption also increases at high temperatures due to air conditioning. The results from the study were grouped by vehicle fuel type and road segment category and averaged to produce representative fuel consumption increases at 20° F and 90° F relative to nominal consumption at 72° F, as listed in Table 3 and Table 4.

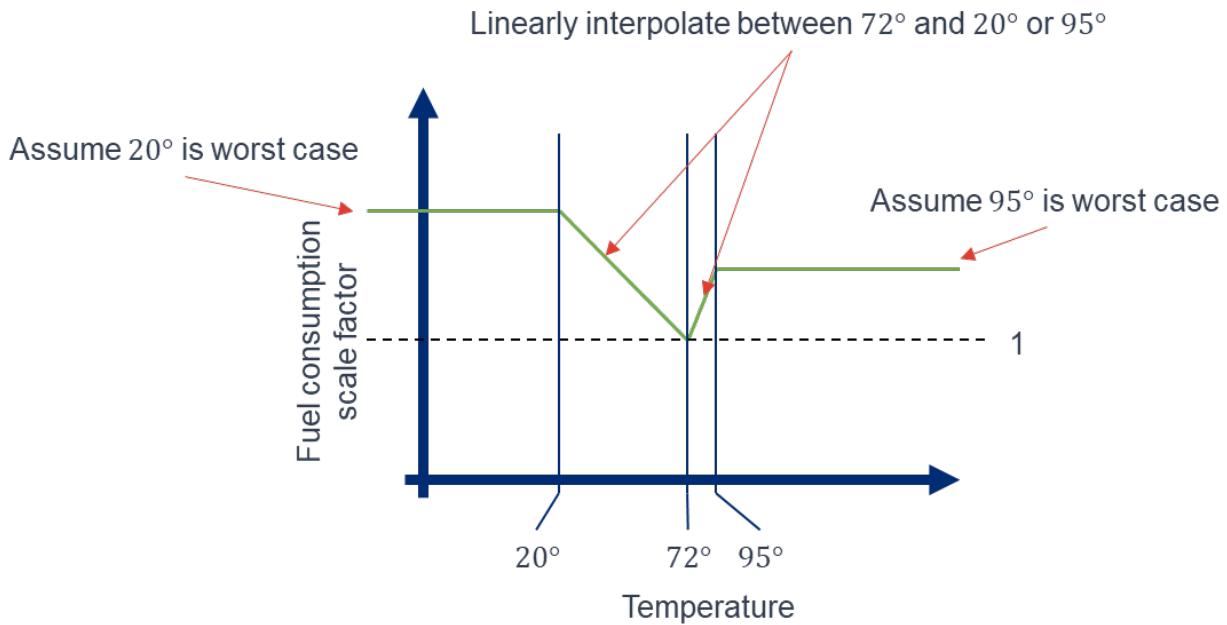
**Table 3** Increase in fuel consumption for arterial and local roads. Values were derived from Lohse-Busch et al. 2013, by averaging across UDDS and US06 tests, then averaging across vehicles with the same fuel type.

Vehicle Fuel Type	Increase at 20° F	Increase at 95° F
Battery Electric	59%	11%
Diesel	7%	25%
Gasoline	5%	20%
Plug-In Hybrid	23%	43%

**Table 4** Increase in fuel consumption for highways. Values were derived from Lohse-Busch et al. 2013, by averaging across HWY and US06 tests, then averaging across vehicles with the same fuel type.

Vehicle Fuel Type	Increase at 20° F	Increase at 95° F
Battery Electric	34%	3%
Diesel	4%	16%
Gasoline	5%	14%
Plug-In Hybrid	10%	22%

These fuel consumption increases were used to estimate a temperature-dependent fuel consumption modifier function, with the modifier linearly interpolated from 20° F to 72° and 72° F to 90° F, and values outside that range clipped to the values at the bounds, as shown in Figure 8.



**Figure 8** Design of temperature dependent fuel consumption scale factor function; implemented functions use scale factors found in Table 3 and Table 4

Temperature values on a half degree latitude/longitude grid and monthly basis were pulled from the Climatic Research Unit gridded Time Series (CRU TS) v4.09 dataset (Harris et al. 2020). We used this dataset due to its standard usage in the literature; the 2020 paper and the 2014 paper for v3.x of the dataset have been cited nearly 7,000 times together. The dataset includes temperatures from 1901 to 2024; the values from 2024 were repeated to provide temperature estimates for 2025.

## 2.2.10 Vehicle GHG and non-GHG Emissions Factors

GHG emissions factors refer to how much of a given greenhouse gas is emitted per unit of fuel burned and varies by fuel type. Our data focuses on carbon dioxide ( $\text{CO}_2$ ), nitrous oxide ( $\text{N}_2\text{O}$ ), and methane ( $\text{CH}_4$ ) emissions factors, using data from the U.S. Environmental Protection Agency (US EPA 2024).

For nitrous oxide and methane, the emissions factors for each gas were given in units of grams of each gas per mile driven. This led to a different emission factor calculation for these gases than for carbon dioxide; details are given in Section 2.3.

Emission factors for non-GHGs follow a similar methodology as nitrous oxide ( $\text{N}_2\text{O}$ ) and methane emissions using data immediately sourced from the U.S. DOT Bureau of Transportation Statistics (US BTS 2024), though ultimately derived from the U.S. EPA. The non-GHGs estimated include other nitrogen oxides ( $\text{NO}_x$ ), carbon monoxide ( $\text{CO}$ ), particulate matter ( $\text{PM}_{2.5}$ ), and other non-GHG hydrocarbons (HC).  $\text{PM}_{2.5}$  emissions are provided both as a total and broken out as separate contributions from brakewear, tirewear, and exhaust.

### **2.2.11 BEV Efficiency by Market Segment**

While direct (scope 1) greenhouse gas emissions from BEVs are zero, for the purposes of evaluating the proposed emissions reduction strategy the efficiency of BEVs (in terms of kWh of electricity used per vehicle-kilometer traveled) was required. Data collected by the Global Fuel Economy Initiative (GFEI) (Cazzola, Paoli, and Teter 2024) provides on an annual, per-country basis the specific energy consumption of light-duty BEVs across 7 different market segments (small car, medium car, large car, small SUV, large SUV, light commercial vehicle and unclassified) in units of Lge (litre of gasoline equivalent) per 100km; these are converted to kWh/km using the equivalent energy content of gasoline of 9.3kWh/Lge (Cazzola, Paoli, and Teter 2023). The light-duty class data provided by GFEI was used as a sufficiently close equivalent of our passenger car class (Section 2.2.7).

After restricting the provided dataset to the newest values between 2021 and 2024, removing outliers and values based on fewer than 1,000 sales, the dataset included values across 35 different countries. In a similar manner to the vehicle fleet mix (Section 2.2.7), missing values for countries were filled first by interpolating among neighboring countries, then within the United Nations defined region, then within the continent, and then over the globe if necessary. Values for the unclassified market segment were not filled, and only used if there was a matching value for the sales of gasoline-powered vehicles within the same, unclassified market segment (as the unclassified segment is not assumed to be equivalent across countries).

### **2.2.12 Vehicle Sales by Market Segment and Powertrain**

As efficiency of BEVs vary with vehicle size, it was necessary to be able to quantify the proportion of current gasoline-powered passenger vehicles for each market segment. In addition to the BEV efficiency data discussed in the previous section, the data collected by GFEI (Cazzola, Paoli, and Teter 2024) also provides on the same annual, per-country basis the sales numbers for light-duty vehicles among the same 7 different market segments per powertrain (internal combustion engine (ICE), mild hybrid, hybrid, plug-in hybrid, battery electric, and fuel cell hydrogen). As only gasoline-powered passenger vehicles are being considered for the emissions reduction strategy, only the sales values for the internal combustion engine powertrain were used. Furthermore, while the sales figures for ICE vehicles included both gasoline and diesel fuel types, since these values are only used to produce a relative proportion of the different market segments, the figures for the combined set are useful as a proxy for just gasoline-powered ICE vehicles.

Similar to in the BEV efficiency data (Section 2.2.11), the dataset was cleaned by restricting data values to the most recent year between 2021 and 2024 that had sales figures of at least 1,000 vehicles in all market segments (excluding the unclassified segment). This left a dataset with 28 different countries, which was filled out to cover other countries in the exact same manner as the BEV efficiency values.

## 2.3 Emissions Calculations

The road transportation sector reports emissions, emissions factor, activity, capacity, and capacity factor on a per month and per GHG basis for GADM Level 1 and GADM Level 2 regions, and on an annual and per greenhouse gas basis for GADM Level 0 regions. These variables are defined at different levels, from individual road segments to GADM Level 0, and are defined in the following sections such that the following general equations hold true:

- **Emissions = Activity \* Emissions Factor**
- **Activity = Capacity \* Capacity Factor**

Values in **bold** are considered the independent variables, with calculations defined in more detail in Sections 2.3.1 through 2.3.6, while dependent variables are described in more detail in Section 2.3.8.

### 2.3.1 Region Emissions

Road transportation emissions ( $E$ ) are reported on a monthly basis for GADM Level 1 (state/province) and GADM Level 2 (county) regions, and on an annual basis for GADM Level 0 (country) regions. Emissions are calculated by summing over the emissions for the atomic geographies within each region  $r$ , where atomic geographies are unique combinations of GADM Level 2 areas, GHSL UCDB urban areas (or no urban area), and a half degree global grid. The sum for GADM Level 0 regions is:

$$E_{r,g,y} = \sum_{a \in r, j \in y} AE_{a,g,j} \quad (1)$$

for GADM Level 0 regions, where  $AE_{a,g,j}$  is the road transportation emissions ( $AE$ ) for each emission type  $g$ , for month  $j$ , for an atomic geography area  $a$ , and  $y$  is the year. Emissions for GADM Level 1 and 2 regions are similar:

$$E_{r,g,j} = \sum_{a \in r} AE_{a,g,j} \quad (2)$$

Emissions ( $E_{r,g,j}$ ,  $E_{r,g,y}$ ) are reported in units of metric tonnes (per month or year).

### 2.3.2 Atomic Geography Emissions

Total road transportation emissions ( $AE$ ) for each GHG or non-GHG emission type  $g$  for month  $j$  for an atomic geography area  $a$  could be calculated as:

$$AE_{a,g,j} = \sum_{i \in a} SE_{g,i,j} \quad (3)$$

where  $SE_{g,i,j}$  is the “segment emissions” for a road segment  $i$ , and the sum is over any road segment which is contained by the atomic geography area  $a$ . Each  $SE_{g,i,j}$  is calculated as:

$$SE_{g,i,j} = \alpha \cdot N_j \cdot AADT_{i,j} \cdot l_{i,j} \cdot SEF_{g,i,j} \quad (4)$$

where:

- $\alpha$  is an activity scale factor to align emissions with EDGAR estimates (See Section 2.3.7)
- $N_j$  is the number of days in month  $j$
- $AADT_{i,j}$  is the average annual daily traffic of the road segment  $i$  for month  $j$ , in units of vehicles per day (See Section 2.4.5)
- $l_{i,j}$  is the length of the road segment  $i$  in month  $j$ , in units of kilometers
- $SEF_{g,i,j}$  is the segment emission factor for greenhouse gas  $g$ , road segment  $i$  and month  $j$  in units of mass per vehicle kilometer

However, this formulation is somewhat computationally inefficient, with a computational complexity of  $O(G * I)$ , where  $G$  is the number of gases (on the order of 10) and  $I$  is the number of road segments (on the order of 4e8 globally). A more computationally efficient approach  $O(I)$  calculates emissions as:

$$AE_{a,g,j} = \sum_c CACT_{a,c,j} * CEF_{a,c,g,j} \quad (5)$$

where  $CACT_{a,c,j}$  is the monthly activity within the atomic geography per road-segment category  $c$ , and  $CEF_{a,c,g,j}$  is the monthly road-segment category emission factor (Section 2.3.3). Each  $CACT_{a,c,j}$  is calculated as:

$$CACT_{a,c,j} = \alpha \cdot N_j \cdot \sum_{i \in a,c} AADT_{i,j} \cdot l_{i,j} \quad (6)$$

And  $CEF_{a,c,g,j}$  is equivalent to  $SEF_{g,i,j}$  where  $c$  is the road segment category of  $i$  and  $a$  is the atomic geography that contains  $i$ .

### 2.3.3 Segment Category Emission Factor

Emission factors are currently calculated in two different paths, as the source emission factors from the EPA are expressed in units of  $\langle \text{mass} \rangle / \langle \text{volume fuel} \rangle$  for carbon dioxide, and units of  $\langle \text{mass} \rangle / \langle \text{distance traveled} \rangle$  for other greenhouse and non-greenhouse gases.

For carbon dioxide, the emission factor for each road segment category is calculated as:

$$CEF_{a,g=CO_2,c,j} = \sum_f t_{a,c,j,f} \cdot \left( \sum_v \eta_{a,v,f} \cdot m_{a,j \in y,v,f} \cdot ev_{j \in y,v,f} \right) \quad (7)$$

- $\eta_{a,v,f}$  is the fuel efficiency, in units of liters of fuel per km, for a vehicle type  $v$ , fuel type  $f$ , and atomic geography  $a$ . Fuel efficiencies are currently estimated on a per-country basis (Section 2.2.8).
- $m_{a,j \in y,v,f}$  is the vehicle mix, as a fraction, typically present in atomic geography  $a$  based on the vehicle type  $v$ , fuel type  $f$ , and year  $y$  of month  $j$ . Specifically, we require that

- $\sum_{v,f} m_{a,j \in y,v,f} = 1$ . Vehicle mix is estimated on an annual per-country basis (Section 2.2.7).
- $egv_{j \in y,v,f}$  is the CO<sub>2</sub> emissions factor, in grams of CO<sub>2</sub> per liter of fuel, for the vehicle type  $v$ , fuel type  $f$ , and year  $y$  of month  $j$ . The CO<sub>2</sub> emissions factor is estimated on an annual basis (Section 2.2.10).
- $t_{a,c,j,f}$  is a unitless, temperature-dependent and road-segment category-dependent modifier to fuel efficiency. It is calculated on a monthly, half-degree grid basis (Section 2.2.9).

For other greenhouse and non-greenhouse gases, the emission factors are independent of road segment category and are calculated on an annual basis as follows:

$$CEF_{a,g \neq CO_2,c,j} = m_{a,j \in y,v,f} \cdot egd_{g,j \in y,v,f} \quad (8)$$

where  $egd_{g,j \in y,v,f}$  is the gas emissions factor, in grams of gas per vehicle kilometer, for the vehicle type  $v$ , fuel type  $f$ , and year  $y$  of month  $j$  (Section 2.2.10). The gas emissions factors are estimated on an annual basis.

### 2.3.4 Capacity

Capacity is simply defined as the total length of the road network in an asset area for month  $j$  for any region  $r$ :

$$CAP_{r,j} = \sum_{a \in r} ACAP_{a,j} \quad (9)$$

where  $ACAP_{a,j}$  is the atomic geography road segment capacity in units of kilometers, and is given as:

$$ACAP_{a,j} = \sum_{i \in a} l_{i,j} \quad (10)$$

Similarly, we can define a capacity for any given road segment  $i$  as  $SCAP_{i,j}$ :

$$SCAP_{i,j} = l_{i,j} \quad (11)$$

Note that because OSM data (Section 2.2.4) can change over time, it is possible for length (and therefore capacity) of a road segment to also change over time.

### 2.3.5 Activity

Activity ( $ACT$ ) for any region  $r$  is the number of vehicle kilometers traveled in month  $j$ :

$$ACT_{r,j} = \sum_{a \in r} AACT_{a,j} \quad (12)$$

where

$$AACT_{r,j} = \alpha \cdot N_j \cdot \sum_{i \in a} AADT_{i,j} \cdot l_{i,j} \quad (13)$$

For segment level activity, we similarly define  $SACT_{i,j}$ :

$$SACT_{i,j} = \alpha \cdot N_j \cdot AADT_{i,j} \cdot l_{i,j} \quad (14)$$

### 2.3.6 Segment Capacity Factor

Only the *segment* capacity factor is independent, the region capacity factor is defined in Equation (19) below. The segment capacity factor ( $SCF$ ) is defined as the ratio of activity to capacity:

$$\begin{aligned} SCF_{i,j} &= \frac{SACT_{i,j}}{SCAP_{i,j} \cdot N_j} \\ SCF_{i,j} &= \frac{\alpha \cdot N_j \cdot AADT_{i,j} \cdot l_{i,j}}{l_{i,j}} = \alpha \cdot N_j \cdot AADT_{i,j} \end{aligned} \quad (15)$$

This variable provides a direct view into the predicted AADT.

### 2.3.7 Activity Scale Factor

To better align our estimates with EDGAR v2025 (M. Crippa et al. 2025), an activity scale factor was introduced. While a portion of the difference between EDGAR and our estimates may ultimately be due to emission factors, due to biases in the training data (Section 2.4.1) our raw, region activity estimates are expected to be high, so the scale factor was chosen to be introduced to the activity component of the emissions estimation. The scale factor is calculated as follows:

$$\alpha = \frac{\sum_{r,y} G_{r,y}}{\sum_{r,y} \hat{E}_{r,CO_2,y}} \quad (16)$$

Where  $\hat{E}_{r,CO_2,y}$  are our annual GADM Level-0 CO<sub>2</sub> emission estimate calculated with an initial activity scale factor of 1, and  $G_{r,y}$  are the annual country level EDGAR v2025 road transportation CO<sub>2</sub> emission estimates. The sums are over the countries and years (2021 and 2022) in common between our estimates and EDGAR v2025. Using this methodology, the activity scale factor was calculated to be 0.7895.

### 2.3.8 Dependent Variables

The reported region emissions factors ( $EF$ ) are defined as the emissions divided by the activity for emission type  $g$  in month  $j$  for any aggregation area  $a$ :

$$EF_{r,g,j} = \frac{E_{r,g,j}}{ACT_{r,j}} \quad (17)$$

This can equivalently be viewed as a segment activity-weighted average emissions factor:

$$EF_{r,g,j} = \frac{\alpha \cdot N_j \cdot \sum_{i \in r} AADT_{i,j} \cdot l_{i,j} \cdot SEF_{g,i,j}}{\alpha \cdot N_j \cdot \sum_{i \in r} AADT_{i,j} \cdot l_{i,j}} \quad (18)$$

The reported region capacity factor ( $CF$ ) is defined as the activity in month  $j$  ( $ACT_j$ ) divided by capacity ( $CAP_j$ ) for any region  $r$ :

$$CF_{r,j} = \frac{ACT_{r,j}}{CAP_{r,j}} \quad (19)$$

This can equivalently be viewed as a segment length-weighted average of vehicles per month:

$$EF_{r,g,j} = \frac{\alpha \cdot N_j \cdot \sum_{i \in r} AADT_{i,j} \cdot l_{i,j}}{\sum_{i \in r} l_{i,j}} \quad (20)$$

### 2.3.9 Emissions Back-propagation

Our current methodology would have several challenges to overcome if it was to be applied to periods before 2021. Due to its crowd-sourced nature, OSM data coverage and quality decreases the farther back in time one looks. Additionally, the time period covering the response to the COVID-19 pandemic in 2020 is also expected to challenge our methodology, as no direct indicators of traffic volume during that time were used. Instead, in order to provide historical, country level (GADM Level-0) emission estimates back to 2015, EDGAR v2025 data was used as a basis to anchor emission estimates.

For those countries that were included in the EDGAR v2025 dataset (as identified by their ISO-3 country code, see Section 7.3), historical estimates are given as follows:

$$E_{r,g,y'} = \left( \frac{G_{r,y'}}{\frac{1}{2} \sum_{y=2021,2022} G_{r,y}} \right) \left( \frac{1}{2} \sum_{y=2021,2022} E_{r,g,y} \right) \quad (21)$$

Where the first term is the ratio between the EDGAR v2025 CO<sub>2</sub> emissions for the historical year  $y'$  and the average annual CO<sub>2</sub> emissions for 2021-2022, and the second term is the average estimated GHG or non-GHG emissions for gas  $g$  over the same time period.

For those countries that are omitted from the EDGAR v2025 dataset, historical estimates are instead given by:

$$E_{r,g,y'} = \left( \frac{\sum_r G_{r,y'}}{\frac{1}{2} \sum_{r,y=2021,2022} G_{r,y}} \right) \left( \frac{1}{2} \sum_{y=2021,2022} E_{r,g,y} \right) \quad (22)$$

Where the first term is now the ratio between the global total of EDGAR v2025 CO<sub>2</sub> emissions for the historical year  $y'$  and the average global annual CO<sub>2</sub> emissions for 2021-2022.

## 2.4 Machine Learning Models

As part of our hybrid modeling approach, machine learning (ML) models were trained to predict AADT from satellite imagery and road network data. These predictions were required in the absence of global, openly available, and high-quality traffic activity data and is a significant departure from existing methods.

### 2.4.1 Ground Truth Road Activity Data

To train our ML models, we utilized the U.S. Highway Performance Monitoring System Average Annual Daily Traffic (AADT) data from 2017 and 2018 (US FHWA 2017). This AADT data was recorded using road-side devices and was provided by each state independently. We utilized the total AADT measure in our work, measured in vehicles per day. AADT data was not available for every road segment and was typically only recorded on major highways and arterial (collector) roads. Where possible, the AADT values are aligned to OSM roads to use as ground truth when training the neural networks.

### 2.4.2 Convolutional Neural Network (CNN)

Our first machine learning approach used semantic segmentation CNNs to predict AADT. Specifically, we input visual satellite imagery in combination with rasterized road network data to predict AADT on a per-pixel basis (Figure 9). This approach was informed by previous work in directly regressing road transport CO<sub>2</sub> emissions (Mukherjee et al. 2021). Models were trained using the Sentinel-2 Level-2A product at 10 m x 10 m resolution, using bands 4 (red), 3 (green), and 2 (blue) (Drusch et al. 2012), and separate models were trained and used on urban and non-urban areas respectively (Section 2.2.1).

Road network data was retrieved from OpenStreetMap and rasterized for the corresponding extent of each visual image tile. Each standardized road type (*highway*, *secondary*, *local*, see Sections 2.2.4 and 2.2.5) was rasterized independently, and the resulting raster channels are concatenated together to form a three-channel image. This image was then combined with the visual image to form a six-channel input image that was input to the CNN. Thus, the CNN model is tasked with predicting the number of vehicles traveling on a road segment as a function of visual satellite image features and road location and type information. We primarily used MAnet-based architectures (Fan et al. 2020) for our segmentation models, based on the findings of similar previous work (Mukherjee et al. 2021).

When the model was used within the overall emissions calculation pipeline, pixel-based AADT



**Figure 9** Example Sentinel-2 visual (RGB) satellite image (left) and corresponding rasterized OSM road network data (right) for Baltimore, Maryland USA. The OSM raster colors represent different road types: highways (red), secondary roads (green), and local roads (blue).

predictions were post-processed to associate predicted values with their corresponding road segment. Ten evenly spaced locations along the length of each road segment were selected to measure the AADT prediction; locations that were not marked as cloudy by the Sentinel-2 Level-2A scene classification map were averaged to produce a single AADT value for that road segment within each image; these AADT values were then averaged across all images within the quarter to produce the final prediction for the CNN models.

The loss function used to train the model has two components, the first of which is a fully-supervised mean square log error term calculated on unclouded pixels containing nonzero truth AADT values. Log predictions and errors are used to account for the many orders of magnitudes AADT values span. The second component is a semi-supervised loss term to address locations where the truth AADT is zero. For this loss term, the same image is passed through the model multiple times using different flips/rotations, and the loss then seeks to reduce the variance between the pixel level AADT predictions after undoing those flips/rotations.

#### 2.4.3 Graph Neural Network (GNN)

Another type of ML model trained to predict AADT was graph neural networks (GNNs) (Bronstein et al. 2017). Road networks inherently take the form of a graph structure, and GNNs can capture road activity across a range of scales more easily than the image-based convolutional neu-

ral network (CNN) segmentation models. CNN-based solutions constrain the spatial area that can be covered during inference, making it difficult to capture potential dependencies on features in neighboring or distant locations. GNNs can easily leverage various features assigned to nodes and efficiently reason over the full road network graph to provide more robust predictions of on-road activity.

For this work, a GNN was trained using OpenStreetMap road network data, including a number of road features: the number of lanes, road length, road type, link or not (such as an exit ramp), and the directional angle between roads. Separate models were trained for both urban and non-urban areas (Section 2.2.1). The Graph Attention v2 (GATv2) network (Brody, Alon, and Yahav 2022) architecture was used as it allows for both edge and node input features, and was set up to predict log-AADT values. We note that the GNN does not use visual imagery as input and is able to learn a relationship between road types and configurations to the density of traffic on those roads.

The loss function used to train the GNN network, like the CNN network, also has two components; the first is a fully-supervised L1 loss on the log error of AADT values (for road segments with AADT truth values), and the second is a semi-supervised relative consistency loss. The consistency loss is computed per-intersection, and is formulated as the net traffic flow into an intersection (i.e., sum of predicted AADT for road segments entering the intersection minus the sum of predicted AADT for road segments exiting the intersection) divided by the average flow through the intersection (i.e., the average of the sum of predicted AADT for road segments entering and the sum for roads exiting the intersection).

#### 2.4.4 Model Training Datasets

Training data was increased compared to previous years to include rural areas for training GNN and CNN models targeted to rural areas. Table 5 shows the details of the various data sources used for training the GNN and CNN models.

**Table 5** Training details and comparison between the 2023 and 2025 release of the road transportation emissions datasets

Model	Training Data	2023 Dataset	2025 Dataset
GNN-Urban	<i>US HPMS AADT</i>	2017 and 2018	2017 and 2018
	<i>OpenStreetMap Dataset Date</i>	4 quarters each in 2017 and 2018	4 quarters each in 2017 and 2018
	<i>Areas</i>	308 urban areas for each quarter of 2017, and 307 urban areas for each quarter of 2018; total 2460 samples	305 urban areas for each quarter of 2017, and 311 urban areas for each quarter of 2018; total 2464 samples
GNN-Rural	<i>US HPMS AADT</i>	N/A	2017 and 2018
	<i>OpenStreetMap Dataset Date</i>	N/A	4 quarters each in 2017 and 2018
	<i>Areas</i>	N/A	1204 rural areas for each quarter of 2017 and 1071 rural areas for each quarter of 2017; areas are aligned to Sentinel-2 tiles

Continued on next page

**Table 5** (Continued)

<b>Model</b>	<b>Training Data</b>	<b>2023 Dataset</b>	<b>2025 Dataset</b>
<b>CNN-Urban</b>	<i>US HPMS AADT</i>	2017 and 2018	2017 and 2018
	<i>Sentinel-2 Imagery</i>	One image per urban area, per quarter for a total of 2028 images	Nominally one image per tile with urban areas per quarter (some locations have multiple images to ensure close to full coverage after taking into consideration clouds), for a total of 2438 images
	<i>OpenStreetMap Dataset Date</i>	4 quarters each in 2017 and 2018	4 quarters each in 2017 and 2018
	<i>Areas</i>	308 urban areas for each quarter of 2017, and 307 urban areas for each quarter of 2018; total 2460 samples	252 tiles covering 308 urban areas

Continued on next page

**Table 5** (Continued)

<b>Model</b>	<b>Training Data</b>	<b>2023 Dataset</b>	<b>2025 Dataset</b>
<b>CNN-Rural</b>	<i>US HPMS AADT</i>	N/A	2017 and 2018
	<i>Sentinel-2 Imagery</i>	N/A	Nominally one image per tile per quarter (some tiles have multiple images to ensure close to full coverage after taking into consideration clouds), for a total of 11994 images
	<i>OpenStreetMap Dataset Date</i>	N/A	4 quarters each in 2017 and 2018
	<i>Areas</i>	N/A	1019 unique areas; areas are aligned to Sentinel-2 tiles and cover the United States

#### 2.4.5 Model Inference & Ensembling

The data sources for inference has expanded greatly compared to previous years due to covering additional years and all global regions. In Table 6 we highlight the differences between the 2023 dataset release and the current 2025 dataset.

**Table 6** Inference details and comparison between the 2023 and 2025 road transportation emissions dataset releases.

<b>Model</b>	<b>Inference Data</b>	<b>2023 Dataset</b>	<b>2025 Dataset</b>
<b>GNN-Urban</b>	<i>OpenStreetMap Dataset Date</i>	4 quarters each in 2021 and 2022 covering 10k urban areas	19 quarters between 2021 Q1 and 2025 Q3 covering almost 13k urban areas

Continued on next page

**Table 6** (Continued)

Model	Inference Data	2023 Dataset	2025 Dataset
<b>GNN-Rural</b>	<i>OpenStreetMap Dataset Date</i>	N/A	19 quarters between 2021 Q1 and 2025 Q3, broken into approximately 15k rural road networks per quarter
<b>CNN-Urban</b>	<i>OpenStreetMap Dataset Date</i>	4 quarters each in 2021 and 2022 covering 10k urban areas	19 quarters between 2021 Q1 and 2025 Q3 covering almost 13k urban areas
	<i>Sentinel-2 Imagery</i>	A total of 37885 images covering 10k urban areas spread over the 8 quarters of 2021 and 2022 (multiple urban areas can be covered by the same image, and multiple images may be needed to cover a single urban area)	A total of 84596 images covering almost 13k urban areas over the 19 quarters between 2021 Q1 and 2025 Q3 (multiple urban areas can be covered by the same image, and multiple images may be needed to cover a single urban area)
<b>CNN-Rural</b>	<i>OpenStreetMap Dataset Date</i>	N/A	19 quarters between 2021 Q1 and 2025 Q3, broken into approximately 15k rural road networks per quarter, one per Sentinel-2 tile
	<i>Sentinel-2 Imagery</i>	N/A	A total of 394286 images covering almost 19k tiles over the 19 quarters between 2021 Q1 and 2025 Q3 (approximately 4k tiles do not actually contain roads)

For each area (urban or rural), for each year, we ended up with up to eight predictions of AADT for every road segment; 2025, having only three quarters of available data, has a maximum of six predictions. However, as discussed in Section 2.2.4, OpenStreetMap data is continuously updated,

and so not every road-segment is present across all eight predictions (four from the CNN, four from the GNN). Additionally, some CNN predictions were not possible due to persistent cloud cover. To create a more robust and predictive yearly AADT prediction model, ensembling was performed using the CNN and GNN models.

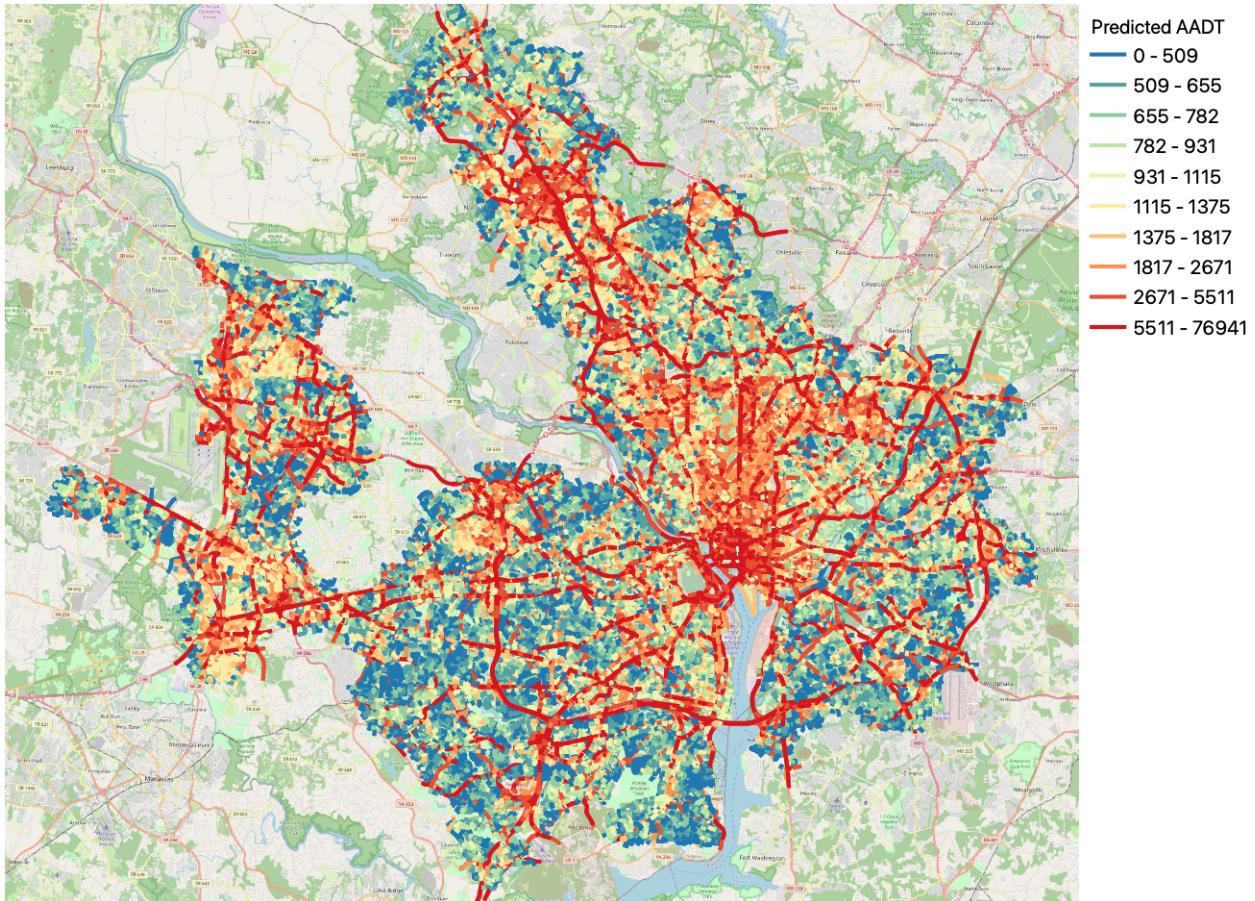
To ensemble the eight, possibly non-identical, sets of road-segment AADT predictions, we employed the following routine:

1. For each model, each AADT prediction for an OSM road segment was assigned to possibly multiple atomic road segments. Atomic road segments were OSM road segments that have been split at our atomic geographies; the unique ID of an atomic road will change if the underlying OSM road changes in either geometry or metadata.
2. For each model, the atomic road AADT predictions within a year were averaged together
3. The yearly atomic road AADT predictions for each model were then averaged together. For cases where a road is only present in the GNN due to clouds in the CNN image, this effectively sets the data for that segment to the mean GNN predicted value

The daily activity estimates (AADT \* road length) and capacities (road lengths) were then aggregated within each atomic geography by road segment category to produce annual estimates. In the aggregation, the contribution from each atomic road segment is weighted by the portion of each year each atomic road segments is valid for (e.g., a road segment that is only valid for the first quarter will only contribute about  $1/4$  of its values to the total). This effectively weighs the road segments that only appear in one quarter as contributing lower to our estimate of emissions than road segments that are present in all four quarters.

From these daily activity and capacity at yearly intervals, we linearly interpolate to produce estimates at monthly intervals. Months before July 2nd, 2021 or after the midpoint of the elapsed portion of the current year were backward/forward filled from the nearest estimate.

An example ensemble AADT output from the 2023 dataset can be seen below in Figure 10.



**Figure 10** Ensemble-predicted AADT for 2021, measured in vehicles per day, for Washington, D.C. Cooler colored road segments indicate relatively less AADT compared to warmer colored road segments.

## 2.5 Uncertainty & Confidence

The 2025 dataset includes an updated methodology for the estimates of uncertainty in emissions estimates. In this section, we discuss our methods for quantifying uncertainty in assets.

### 2.5.1 Uncertainty Overview

In addition to CO<sub>2</sub>, N<sub>2</sub>O, and CH<sub>4</sub> emissions estimates, we provide estimates of activity, capacity, and other values at aggregated geographies. A full dataset description is in Section 6.2. In Table 7 below, we summarize the approach taken to estimate the uncertainty in each value.

In the following subsections, we provide details on the estimation process in detail.

**Table 7** Summary of methods to estimate uncertainty per column of aggregated data

Column	Method
capacity	Bootstrapping using per-road-segment data
activity	Bootstrapping using per-road-segment data
capacity_factor	Propagation of error from capacity and activity
CO2_emissions_factor	Propagation of error from CO2_emissions and activity
CH4_emissions_factor	Propagation of error from CH4_emissions and activity
N2O_emissions_factor	Propagation of error from N2O_emissions and activity
other_gas_emissions_factor	Propagation of error from other_gas_emissions and activity
CO2_emissions	Bootstrapping using per-road-segment data
CH4_emissions	Bootstrapping using per-road-segment data
N2O_emissions	Bootstrapping using per-road-segment data
other_gas_emissions	Bootstrapping using per-road-segment data
total_CO2e_100yrGWP	Propagation of error from CO2_emissions, CH4_emissions, and N2O_emissions, including constants
total_CO2e_20yrGWP	Propagation of error from CO2_emissions, CH4_emissions, and N2O_emissions, including constants

## 2.5.2 Bootstrapping Approach

The machine learning inference process (see Section 2.4.5) provides estimates of AADT on a road-segment level (see Section 2.2.4) per segment category (see Section 2.2.5), and includes the road-segment length. These road segments are tied to specific aggregation areas (such as GADM 0, 1, 2; see (Section 2.2.1, Section 2.2.2, and Figure 2)). In order to estimate the uncertainty in the total activity, capacity, or emissions for an aggregated area, a bootstrapping approach is used (Efron and Tibshirani 1994). Specifically, we use the bootstrapping technique as implemented by the `scipy` package with the “basic” method option due to the number of road segments in many large aggregated regions. This replaces the prior Monte Carlo approach, removing reliance on distributional assumptions which break down in rural regions, while also reducing computational overhead.

In this approach, we first calculate the activity and emissions for each road segment as shown in Equations 14 and 4 respectively, using the inferred AADT per road segment. The bootstrap is calculated against the total activity, capacity, and emissions for each segment type and month for each aggregated area. The algorithm in brief is:

1. Choose a set of road segments, of the same size as the original dataset, with replacement
2. Calculate the total activity, capacity, and emissions
3. Repeat steps 1 and 2 one thousand times, and estimate the mean, median, standard deviation,

and the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles.

The Climate TRACE coalition has decided to use a single value to describe the uncertainty in a particular value. In order to estimate this single value instead of providing lower and upper confidence interval bounds, we use the following:

$$\sigma = \frac{CI_{97.5} - CI_{2.5}}{\mu} \quad (23)$$

where  $\mu$  is the bootstrap mean. While the bootstrap standard deviation can be helpful as a single value uncertainty estimate, the confidence interval will likely provide a better estimate of uncertainty with non-normal data.

The bootstrap technique provides estimates for the following properties:

- Activity
- Capacity
- Emissions (for each of 10 emissions types)

These are calculated per aggregate region, per month, per segment category.

### 2.5.3 Correlations

While the bootstrap calculation provides estimates for the activity, capacity, and emissions directly, there are many secondary properties, such as emissions factor and capacity factor, that require propagation of uncertainty from these three values. To more accurately calculate the propagation of errors, we estimate the correlations between these values. Specifically, using the same per-road-segment data for activity, capacity, and emissions, we estimate the Pearson  $\rho$  correlation coefficient between these two pairs of values for every aggregation area and every segment type:

- activity and emissions (for each emission type separately)
- activity and capacity

While the activity and capacity vary in correlation, activity and emissions are very strongly correlated. Although emissions depends on the multiplication of activity with the segment category emissions factor (SCEF), the activity varies significantly more than the SCEF. Therefore, activity and emissions are typically strongly correlated with a coefficient near 1.

### 2.5.4 Uncertainty Aggregation

The standard method for combining uncertainties for a given model (e.g., calculating emissions from emissions factors and activity) is to propagate the sources of uncertainties by any linear or non-linear differentiable function  $f$  of  $n$ -variables  $x_i$ :

$$\sigma_f^2 = J \Sigma^x J^T = \sum_i^n \sum_j^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \rho_{x_i x_j} \sigma_{x_i} \sigma_{x_j} \quad (24)$$

where  $\sigma_f$  is the uncertainty of the variable of interest, and  $\sigma_{x_i}$  are the uncertainties of the sources, and  $\rho_{x_i x_j}$  is the correlation between the variables. Additionally, we explicitly assume that errors are uncorrelated between the segment categories  $c$ . For variables that can be expressed as a linear sum of per-segment-category values, the per-segment-category errors can be added in quadrature to calculate the final per-region, per-month uncertainty estimate, for example as with activity:

$$\sigma_{ACT_{r,j}}^2 = \sum_{c=1}^3 \sigma_{ACT_{r,j,c}}^2 \quad (25)$$

## 2.5.5 Emissions Factors Uncertainty

For emissions factors  $EF_{r,g,j}$ , their definitions (Equation (17)) can be re-expressed as in terms of per-segment-category values for emissions ( $E_{r,g,j,c}$ ) and activity ( $ACT_{r,j,c}$ ):

$$EF_{r,g,j} = \frac{E_{r,g,j}}{ACT_{r,j}} = \frac{\sum_c E_{r,g,j,c}}{\sum_c ACT_{r,j,c}} \quad (26)$$

Under the same, previous assumption that errors are uncorrelated between the segment categories  $c$ , combining Equations 26 with 24 and simplifying gives the following formula for emissions factor uncertainties:

$$\begin{aligned} \sigma_{EF_{r,g,j}}^2 = \sum_c & \left[ \left( \frac{\sigma_{E_{r,g,j,c}}}{ACT_{r,j}} \right)^2 + \left( \sigma_{ACT_{r,j,c}} \frac{E_{r,g,j}}{ACT_{r,j}^2} \right)^2 \right. \\ & \left. - 2\rho_{E_{r,g,j,c}ACT_{r,j,c}} \left( \frac{\sigma_{E_{r,g,j,c}}}{ACT_{r,j}} \right) \left( \sigma_{ACT_{r,j,c}} \frac{E_{r,g,j}}{ACT_{r,j}^2} \right) \right] \end{aligned} \quad (27)$$

## 2.5.6 Capacity Factor Uncertainty

Similarly, capacity factor  $CF_{r,j}$  can be re-expressed in terms of segment categories:

$$CF_{r,j} = \frac{ACT_{r,j}}{CAP_{r,j}} = \frac{\sum_c ACT_{r,j,c}}{\sum_c CAP_{r,j,c}} \quad (28)$$

with an associated equation for uncertainty:

$$\begin{aligned} \sigma_{CF_{r,j}}^2 = \sum_c & \left[ \left( \frac{\sigma_{ACT_{r,j,c}}}{CAP_{r,j}} \right)^2 + \left( \sigma_{CAP_{r,j,c}} \frac{ACT_{r,j}}{CAP_{r,j}^2} \right)^2 \right. \\ & \left. - 2\rho_{ACT_{r,j,c}CAP_{r,j,c}} \left( \frac{\sigma_{ACT_{r,j,c}}}{CAP_{r,j}} \right) \left( \sigma_{CAP_{r,j,c}} \frac{ACT_{r,j}}{CAP_{r,j}^2} \right) \right] \end{aligned} \quad (29)$$

## 2.5.7 Total Equivalent CO<sub>2</sub>: 20 and 100 Year Baselines

These two columns are calculated from propagation of error of the total:

$$TOT_Z = CO2 + CH4 \cdot CO2e_{CH4,Z} + N2O \cdot CO2e_{N2O,Z} \quad (30)$$

where  $Z \in 20, 100$  are the two different baselines,  $CO2e_{CH4,Z}$  is the equivalent CO<sub>2</sub> emissions for methane, and  $CO2e_{N2O,Z}$  the equivalent CO<sub>2</sub> emissions for nitrous oxide. We use the propagation of error along with the standard IPCC definitions of error in the equivalent CO<sub>2</sub> emissions such that:

$$\sigma_{TOT_Z}^2 = \sigma_{CO2}^2 + CH4^2 \sigma_{CO2e_{CH4,Z}}^2 + CO2e_{CH4,Z}^2 \sigma_{CH4}^2 + N2O^2 \sigma_{CO2e_{N2O,Z}}^2 + CO2e_{N2O,Z}^2 \sigma_{N2O}^2 \quad (31)$$

## 2.5.8 Confidence

As described above, there are many assumptions and rough approximations made in order to estimate uncertainties of many different parameters calculated from our models. We judge, based on subject matter expertise and knowledge that this initial calculation while useful, can be improved in many ways. As such, all uncertainty parameters calculated here have a LOW confidence using standard IPCC confidence ranges.

## 2.6 International Emissions Inventories

We have performed comparison against one international emissions inventory globally, EDGAR v2025 (M. Crippa et al. 2025), for the years 2021-2022. Although we use EDGAR to get a global scale factor, this comparison is still valid to check the distribution of model performance and dependencies on input features. Specifically, the estimated global scale factor affects primarily the mean of our model output. EDGAR provides a global inventory for transportation that uses road density as a proxy to spatially distribute emissions. Other possible inventories, such as Google EIE or Carbon Monitor, for example, do not have open data sources that we could easily download across our entire dataset. As a public, gridded product, the EDGAR dataset is somewhat more amenable to comparisons against our datasets.

Our emissions estimates were compared against EDGAR v2025, both within the USA and globally. This comparison is performed for initial validation of our data; it is not used for uncertainty estimates. EDGAR v2025 separates CO<sub>2</sub> emissions by long-cycle (IEA-EDGAR CO<sub>2</sub>) and short-cycle (EDGAR CO<sub>2</sub>bio) sources; these are summed together for comparison to our data. EDGAR v2025 also reports CO<sub>2</sub> emissions on a 0.1 degree grid, while we report our estimates on a 0.5 degree grid. To provide a consistent estimate across the two datasets, we aggregate the 0.1 grids to the nearest 0.5 grid and compare results on that.

See Section 3.3 for a discussion of the results of the comparison.

## 2.7 Emission Reduction Strategy (ERS)

The initial ERS analyzed for the road transportation sector is the conversion of passenger gasoline-powered internal combustion engine (ICE) vehicles to battery-powered electric vehicles (BEVs). The overall change in emissions under this scenario for this sector are driven by the reduction in scope 1 emissions due to the removal of ICEs, but are partially negated by increases in scope 2 emissions due to the energy used by BEVs coming from the local electrical grid. An additional reduction in scope 2 emissions occurs due to the reduced need for gasoline production from the oil-and gas-refining sector. Unmodeled is the response within the ethanol fuel industry, which would see reduced demand from the road transportation sector and might respond by either reducing corn or sugarcane production or shifting production towards other sectors.

The ERS values as computed are based on aggregated road emissions and induced activity values across all of 2024, the last full year of data available as of writing. The net reduction  $R_{r,g}$  in road emissions in a given region  $r$  for a greenhouse gas  $g$  is defined by the following equation:

$$R_{r,g} = \sum_{j \in y=2024} E_{r,g,j} \cdot (1 - \rho_{r,g}) - ACT_{r,j} (\alpha_{1,r} MER_{1,r,g} + \alpha_{2,r} MER_{2,r,g}) \quad (32)$$

where:

- $E_{r,g,j}$  is the per-month emissions within the region as given by Equation (2)
- $\rho_{r,g}$  is the ratio between scope 1 emissions in the ERS scenario relative to the baseline emissions; see Section 2.7.1
- $ACT_{r,j}$  is the per-month activity within the region as given by Equation (12)
- $\alpha_{1,r}$  is the induced activity conversion rate for the electricity generation sector in units of MWh per vehicle kilometers traveled (VKT); see Section 2.7.2
- $\alpha_{2,r}$  is the induced activity conversion rate for the oil & gas production sector in units of barrels of oil (BBL) per VKT; see Section 2.7.3
- $MER_{1,r,g}$  and  $MER_{2,r,g}$  are the marginal emissions rates for the electricity generation and oil & gas production sectors respectively

It is important to note that the induced activity is at point-of-use; transmission or transportation losses are accounted for with the  $MER$  values for the respective sectors. While the derivation of the  $MER$  values are out-of-scope for this document, they can be found within the individual sectors' methodology documentation within the broader ClimateTRACE coalition.

Additionally, confidence values were selected to indicate the feasibility of the ERS within each region. While the strategy relies on currently available, proven technology, the financial burden of implementation, both in terms of replacement vehicles and the supporting electric grid, makes the strategy more attainable for countries with a higher income. As such, the country income classifications by the World Bank (The World Bank 2025) are used to set confidence (feasibility) levels; high-income countries are given a high confidence; upper- and lower-middle income countries a medium confidence, and low-income countries a low confidence.

### 2.7.1 Estimating Decrease in Scope 1 Emissions

Scope 1 emission changes are easily accounted for within the emission factor calculations by shifting the proportion of passenger vehicles that are gasoline powered to BEV, and computing the difference in the final emission tallies. Specifically, we adjust the vehicle mix values  $m_{a,j,v,f}$  in Equations 7 and 8 such that the adjusted values  $m'_{a,j,v=passenger\_cars,f}$  are given as:

$$m'_{a,j,v=passenger\_cars,f=gasoline} = 0 \quad (33)$$

$$m'_{a,j,v=passenger\_cars,f=bev} = m_{a,j,v=passenger\_cars,f=bev} + m_{a,j,v=passenger\_cars,f=gasoline} \quad (34)$$

The remainder of the emissions equations remain the same to produce an updated emission estimate  $E'_{r,g,j}$  under the ERS scenario for each region  $r$ . The emissions ratio  $\rho_{r,g}$  is then given as:

$$\rho_{r,g} = \frac{\sum_{j \in y=2024} E'_{r,g,j}}{\sum_{j \in y=2024} E_{r,g,j}} \quad (35)$$

### 2.7.2 Induced Increase in Electricity Generation

As a result of switching from ICE to BEV passenger vehicles, there will be an increase in electricity used to power the new BEVs. The rate of new, induced electrical energy usage per VKT of activity will vary from region to region, as there are clear preferences in different countries for differently sized vehicles, and there is a negative correlation between vehicle weight and efficiency (Cazzola, Paoli, and Teter 2023). Furthermore, the current average efficiency of BEVs within a country is not necessarily a good proxy for the expected efficiency under the ERS scenario, as the current fleets of BEVs are skewed by the preferences of early adopters. As such, we use the latest vehicle sales figures per market segment (Section 2.2.12) of the ICE passenger vehicles to compute a weighted average of the BEV efficiencies (Section 2.2.11):

$$\bar{\beta}_r = \frac{\sum_k \beta_{r,k} \cdot n_{r,k}}{\sum_k n_{r,k}} \quad (36)$$

where  $\bar{\beta}_r$  is the weighted average for region  $r$ ,  $\beta_{r,k}$  is the BEV efficiency for market segment  $k$ , and  $n_{r,k}$  is the vehicle sales number for ICE vehicles within the region's country.

In order to match the format of  $\alpha_1$  within Equation (32),  $\bar{\beta}_r$  needs to be scaled by the ratio of the activity of the affected gasoline-powered passenger cars relative to the total activity within a region. This is calculated as:

$$\alpha_{1,r} = \frac{\sum_{a \in r, j \in y=2024} AACT_{a,j} \cdot m_{a,j,v=passenger\_cars,f=gasoline}}{\sum_{a \in r, j \in y=2024} AACT_{a,j}} \bar{\beta}_r \quad (37)$$

where  $AACT_{a,j}$  is the activity within the atomic geography area in units of vehicle kilometers traveled (Equation (13)).

### 2.7.3 Induced Decrease in Gasoline Production

As switching passenger cars from gasoline-powered to BEVs reduces the amount of motor gasoline consumed, there is expected to be a corresponding decrease in the amount of gasoline produced from the oil & gas production sector. Using mostly previously defined emission factors, the nominal amount of gasoline consumed per unit of activity  $G_{a,c,j}$  is calculated as:

$$G_{a,c,j} = t_{a,c,j,f} \cdot \left( \sum_v \eta_{a,v,f} \cdot m_{a,j \in y, v, f} \right) \quad (38)$$

for  $f=\text{gasoline}$ , with a corresponding  $G'_{a,c,j}$  calculated when using the modified  $m'$  values (Equation (33)) under the ERS scenario.

Similar to the scale factor derived in Section 2.3.7, a scale factor  $s_G$  is introduced to ensure the estimated total motor gasoline consumed globally is consistent with other estimates. International Energy Agency 2025 reports a demand of 27,214 thousand barrels of motor gasoline per day in 2024. However, motor gasoline consumed at the fuel pump is not produced purely from petroleum; rather, it is often blended with ethanol. Ramsey et al. 2023 reports 27.6 billion gallons of fuel ethanol consumed globally relative to a total consumption of 403.3 billion gallons of motor gasoline in 2018, with the remaining  $(1 - 27.6/403.3) = 0.932$  fraction coming from petroleum-derived sources. Using a conversion factor of  $6.2898 \times 10^{-3}$  liters of motor gasoline per barrel (British Petroleum 2022), the scale factor is calculated as:

$$s_G = \frac{0.932 * 27214e3 * 365}{6.2898 \times 10^{-3} \sum_{a,c,j \in y=2024} G_{a,c,j} \cdot CACT_{a,c,j}} = 0.85 \quad (39)$$

The induced activity ratio can then be calculated as:

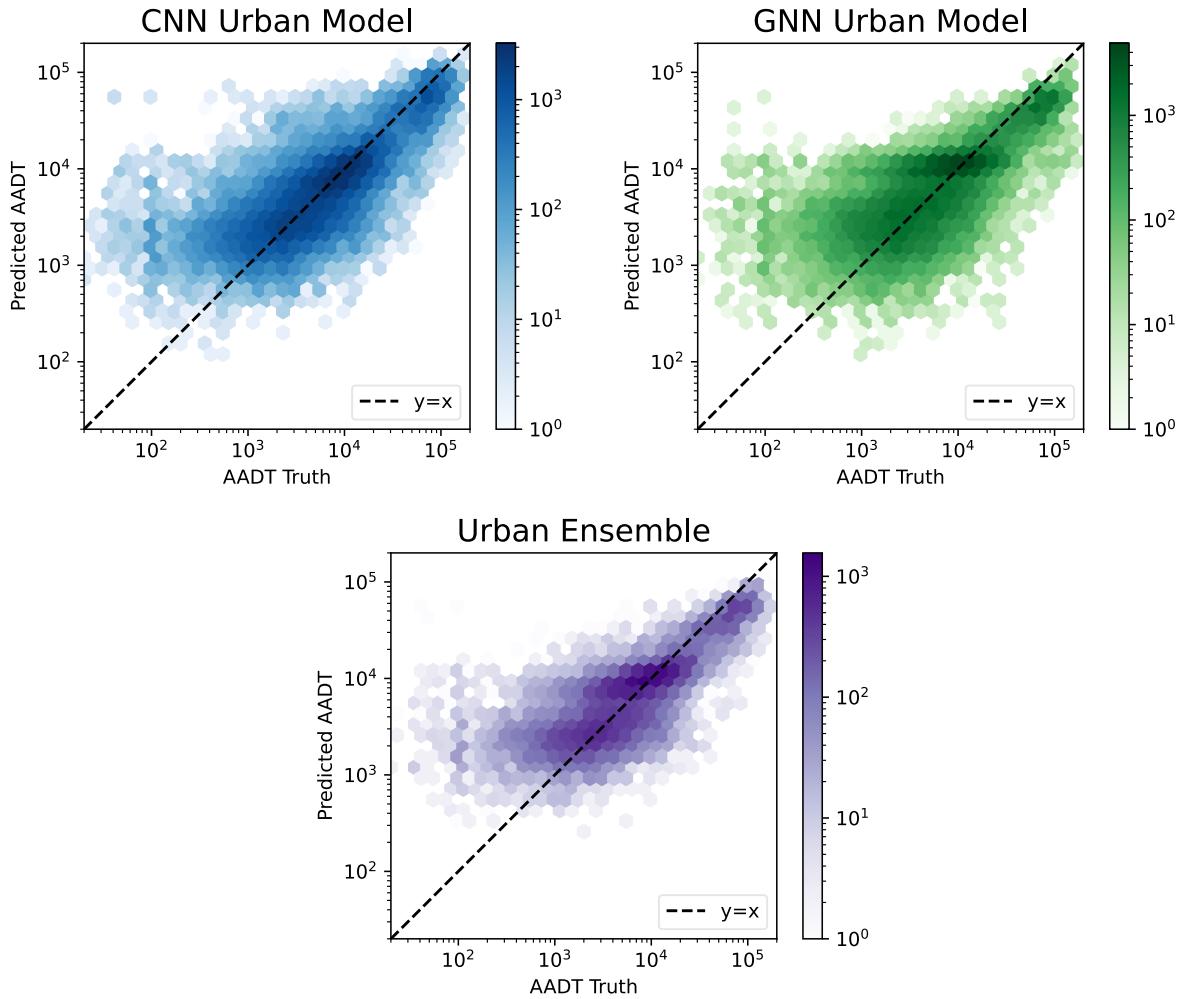
$$\alpha_{2,r} = s_G \cdot 6.2898 \times 10^{-3} \cdot \frac{\sum_{a \in r, c, j \in y=2024} (G'_{a,c,j} - G_{a,c,j}) \cdot CACT_{a,c,j}}{\sum_{j \in y=2024} ACT_{r,j}} \quad (40)$$

## 3 Results

In this section, we highlight top-level trends and differences between dataset release years. Specific information on per-urban-area emissions is available in the released dataset.

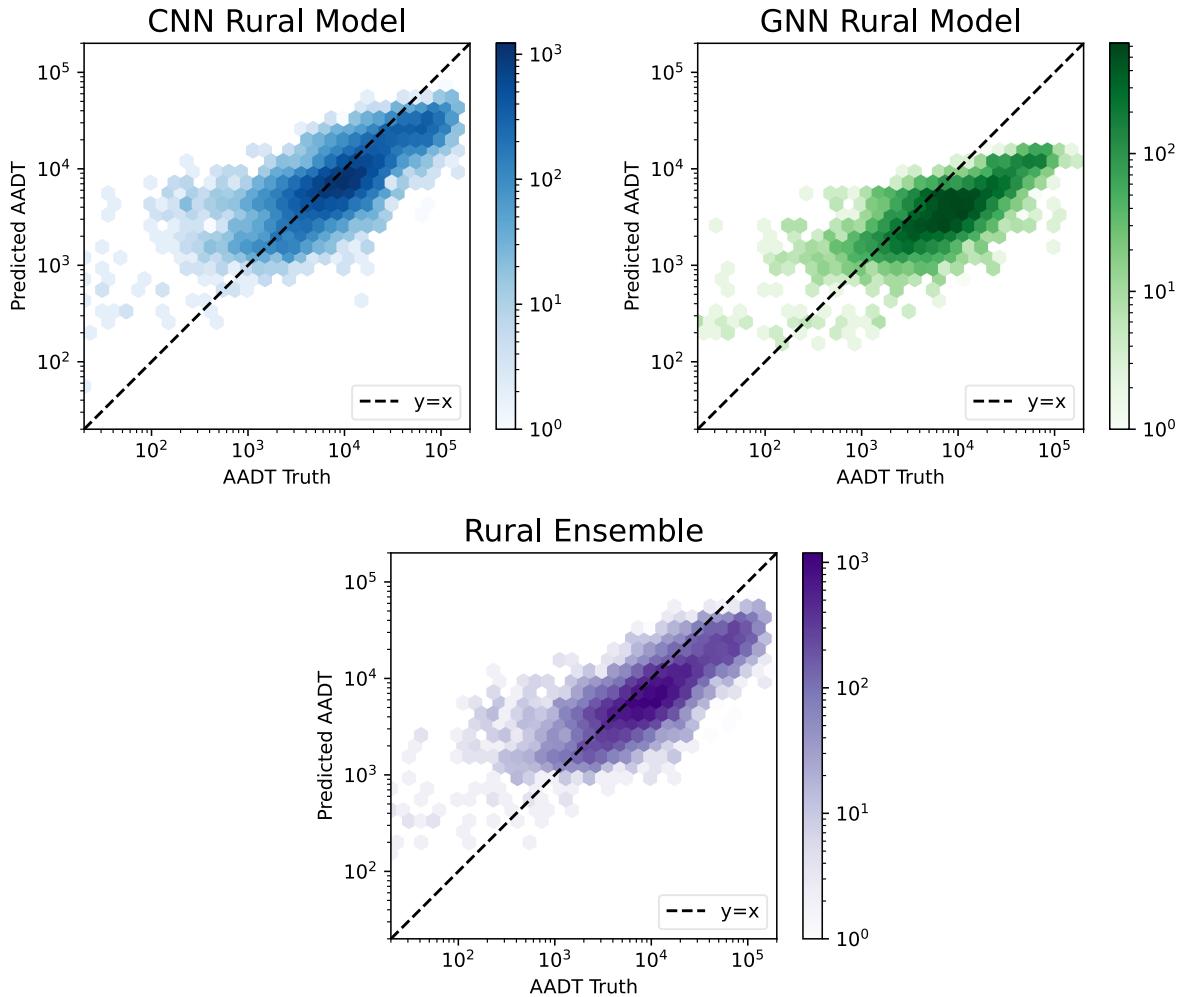
### 3.1 Machine Learning Model Ground Truth Comparison

While each trained ML model reduces the difference between the ground truth AADT values and the inferred values, it is worthwhile comparing a visual representation of the kinds of differences we see. Figures 11 and 12 below display a hexbin plot of predicted vs ground truth AADT values for the CNN Model, the GNN Model, and the ensemble for both the urban models (Figure 11) and the rural models (Figure 12).



**Figure 11** Comparison of predicted AADT vs. ground truth across a set of 14 test cities in the USA for 2017 and 2018.

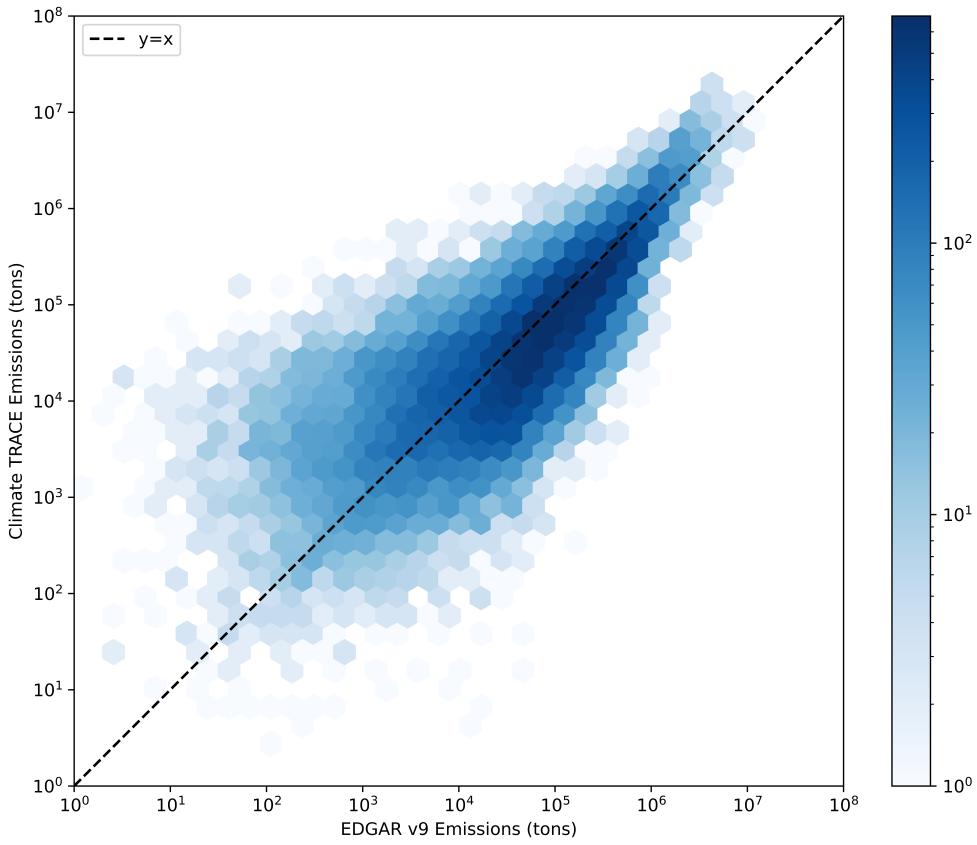
The dashed lines in the plots indicate the line along with predicted and ground truth AADT values are equal. Bins within the plot that have a darker color are places which have more instances in that bin according to the color scale presented. We note in general, all models tend to overestimate AADT for low true AADT values, and models tend to slightly underestimate at the highest true AADT values. The underestimation at low AADT values is expected to be largely due to the bias in available truth AADT values towards more highly trafficked roads. Of special note, the GNN rural model more significantly underestimates true AADT values, though this is largely alleviated through ensembling with the CNN model.



**Figure 12** Comparison of predicted AADT vs. ground truth across a set of 85 rural areas from 2017 and 77 rural areas from 2018 in the USA.

### 3.2 Global Emissions Inventories Comparison

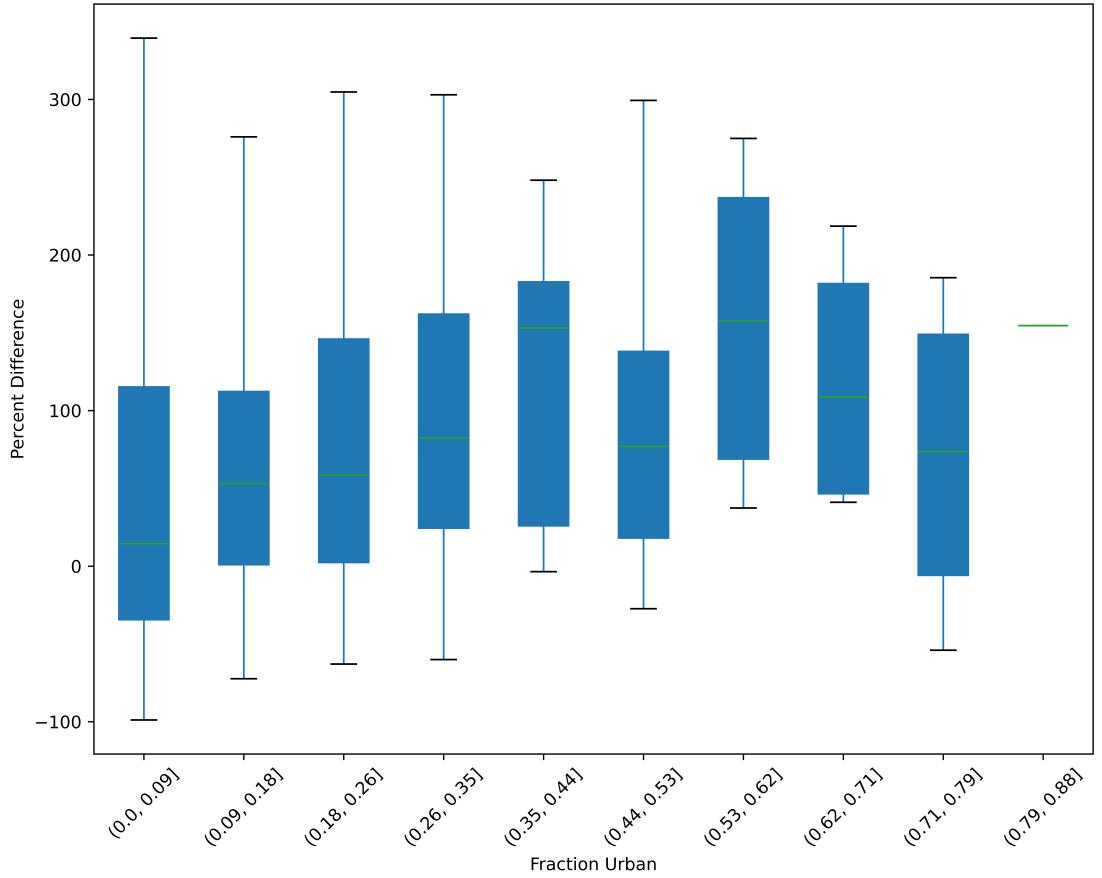
Figure 13 shows the comparison between the EDGAR v2025 dataset and our release of emissions estimates for the entire globe. EDGAR v2025 emissions data is aggregated from a 0.1 degree grid to a 0.5 degree grid for comparison. We compare only for year 2021 and aim to look at the dependencies of model performance on UN Region and on our rural/urban split. While there is a strong correlation across all emissions, there are also significant differences in areas with lower emissions. For example, where EDGAR broadly predicts  $10^1$  to  $10^3$  tons, our methodology predicts more. Where EDGAR broadly predicts  $10^4$  to  $10^6$  tons, our methodology also predicts less. For the purposes of this methodology document, we simply highlight that the two datasets provide comparable estimates when taken in aggregate.



**Figure 13** Comparison of EDGAR v2025 emissions and estimates described in this methodology across our global 0.5 degree grid. Darker blues indicate that the most common comparison between the two databases indicates that our methodology over-estimates emissions in lower values, and under-estimates larger values relative to EDGAR.

Figure 13 shows our scaled emissions estimates compared with EDGAR v2025 data. The global scale factor derived from EDGAR enforces that our estimates have a similar means, but we demonstrate that our emissions are similar in their dependency on other factors in the data as well. We further explore the discrepancies from EDGAR by determining whether the error is largely attributed to rural or urban areas. For each grid we calculate the fraction of the total urban area over the total area, with urban areas being defined using GHSL boundaries. Figure 14 shows the percent difference across various urban fraction bins. We primarily notice largest variance in our deviations from EDGAR at lower urban fractions, which is expected given the lack of data availability in rural areas.

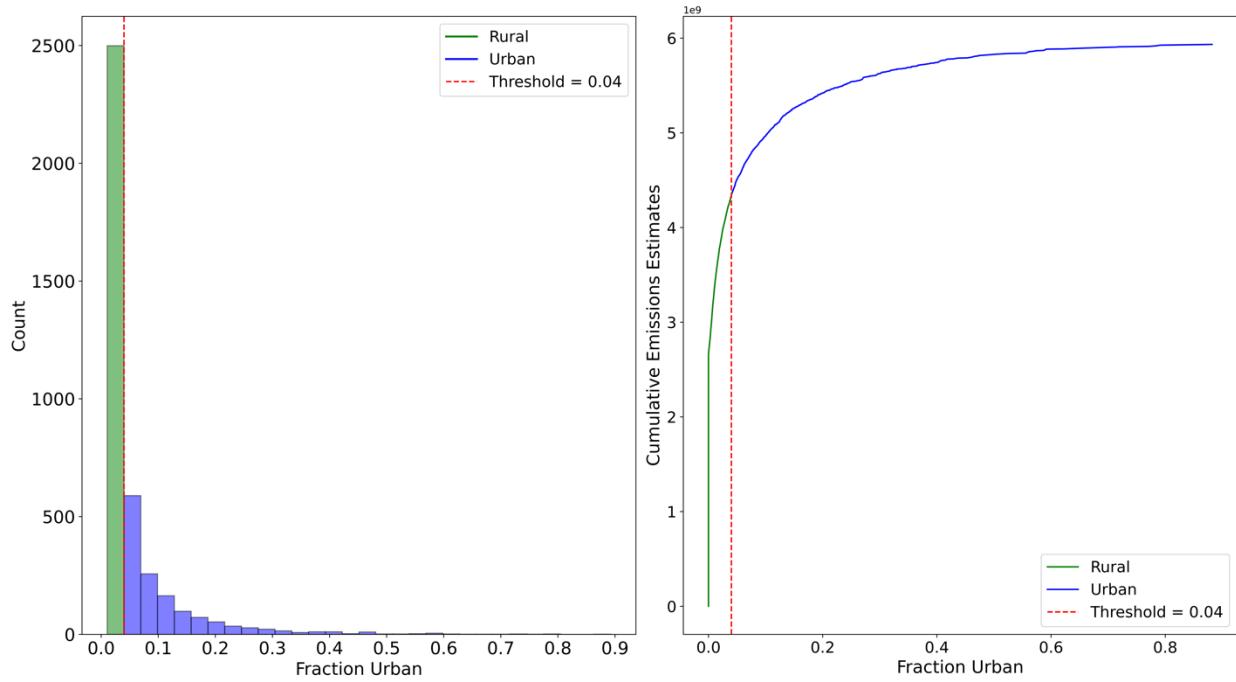
To further explore this difference in Figure 14, a Mann-Whitney U-test is performed between our



**Figure 14** Signed percent difference between EDGAR v2025 and estimates across various urban fraction values.

error distributions for urban and rural areas. For this analysis, we choose grid points with urban fractions one standard deviation above the mean to be “urban”, and all other points to be “rural”. Figure 15 shows the result of this split. According to this threshold, rural areas contribute the most to our emissions. We sample 100 points without replacement from both the distribution of the actual difference and percent difference between our model estimates and EDGAR. For the percent difference, there was a significant difference between urban ( $\mu = 124.93$ ,  $\sigma = 290.06$ ) and rural ( $\mu = 237.46$ ,  $\sigma = 1491.13$ ) distributions with  $u = 704977.0$ ,  $p < 1e - 57$ . For the actual signed difference there was a significant difference between urban ( $\mu = 1363765.20$ ,  $\sigma = 2336733.34$ ) and rural ( $\mu = -7479.13$ ,  $\sigma = 228008.06$ ) distributions with  $u = 817888.0$ ,  $p < 1e - 134$ .

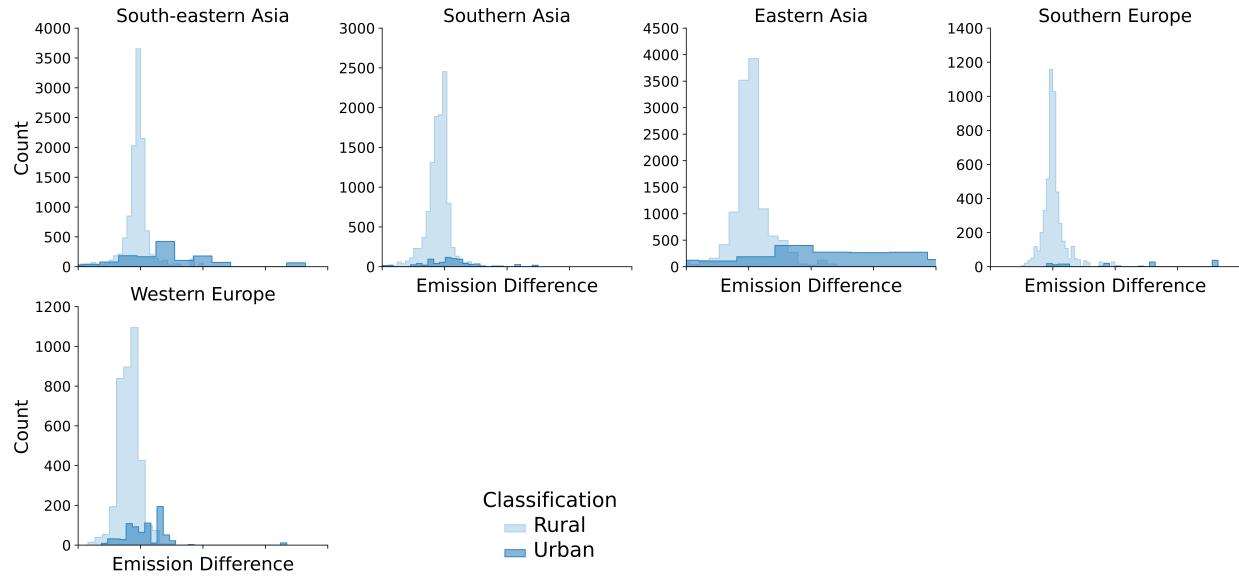
Lastly, we look at the difference between our estimates and EDGAR according to UN regions. Figure 16 below shows the absolute difference per each UN region, split up in rural and urban regions. The urban and rural regions are determined using the same method as above. For rural



**Figure 15** Urban/rural split according to a 1-standard deviation above the mean for a given grid's fraction of urban atomic geometries.

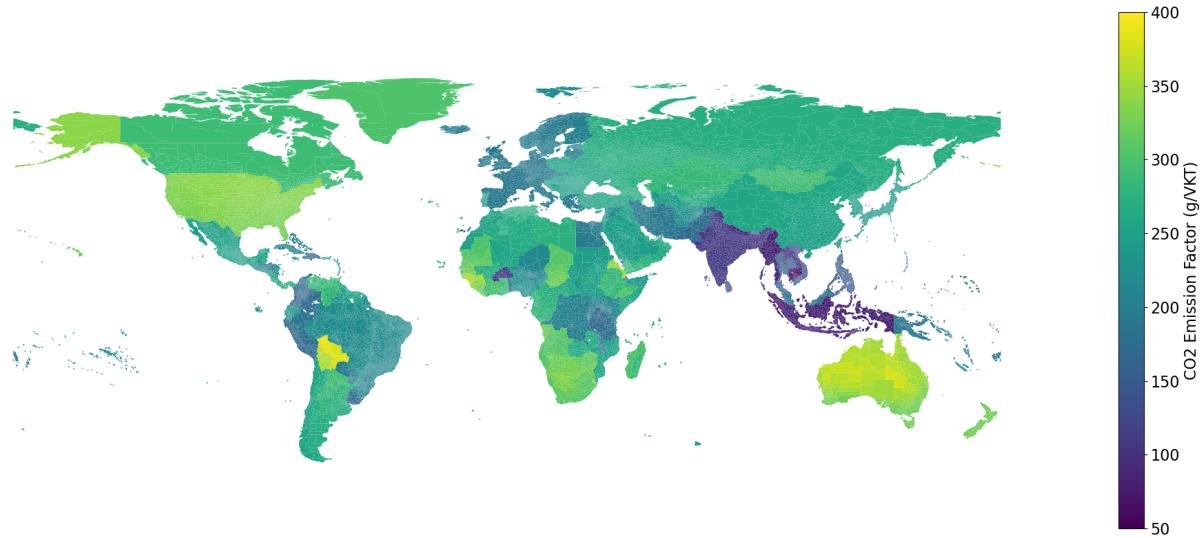
areas, our estimates differ the most in Eastern Asia, Southern Europe, and Western Europe. For urban areas, the largest difference is in Southern Asia and South-eastern Asia. Differences for all UN regions are provided in the appendix below (Figure 28).

Further international validation for both AADT and resulting emissions estimates is necessary, including a deeper understanding of all potential sources of differences (e.g., emission factors uncertainty, varying total road network length included, etc.). We aim to continue to incorporate international AADT datasets, where available, into our model training and validation process. The addition of more real-time data (e.g., traffic and mobility) will help address the temporal ambiguity of our estimates. Finding publicly available data of this type remains a significant challenge.



**Figure 16** Comparison of actual difference between estimates and EDGAR for key UN regions, split between rural and urban regions.

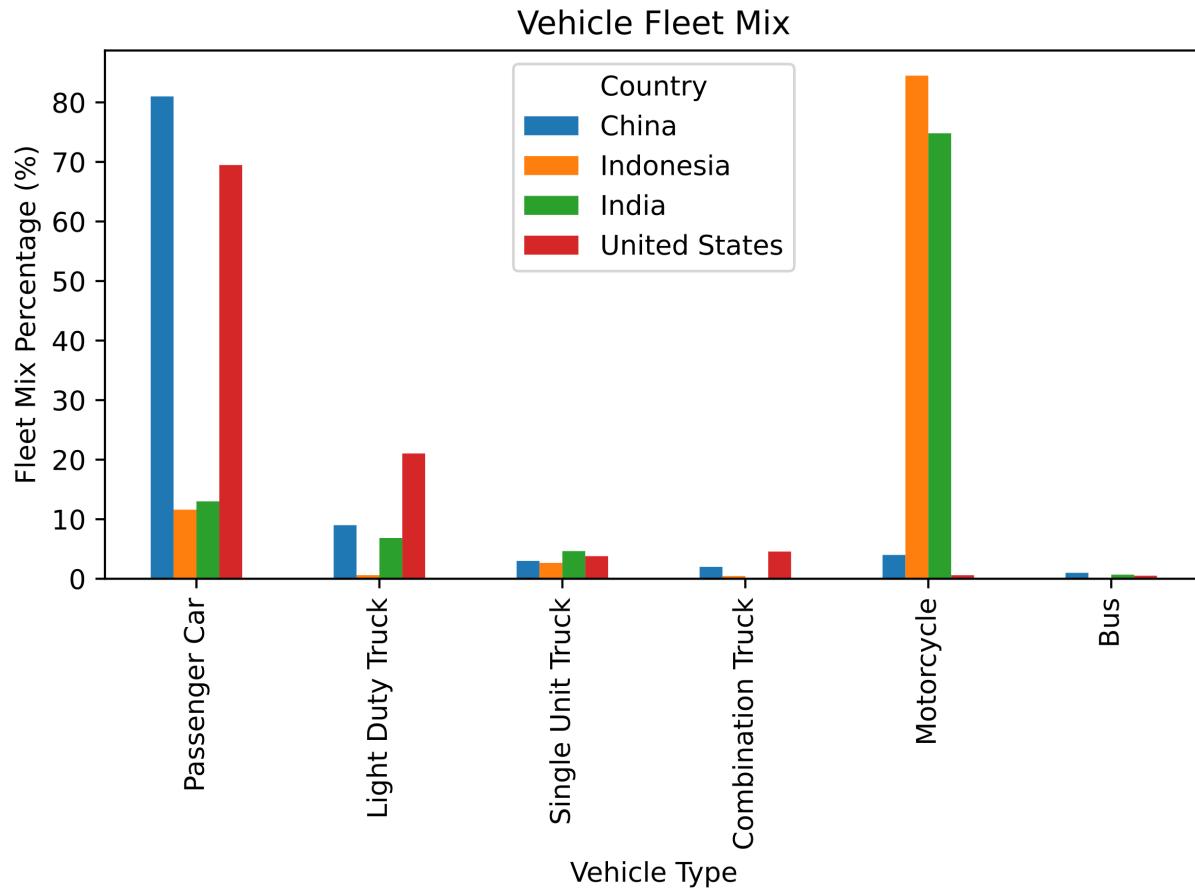
### 3.3 Worldwide Emissions Results



**Figure 17** Global map of the January 2024 emission factors in the released data set. Units are grams per VKT. Regions shown are GADM Level-2 when available, GADM Level-1 otherwise.

Figure 17 shows an overview of the emission factors for January, 2024 (EF, see Section 2.3.8). One interesting highlight from our data is the fact that areas in India and Indonesia have a much lower EF than countries such as China and the USA. This is mostly due to our data sources for vehicle fleet mix in those countries, which indicate that a vast majority of vehicles in those two

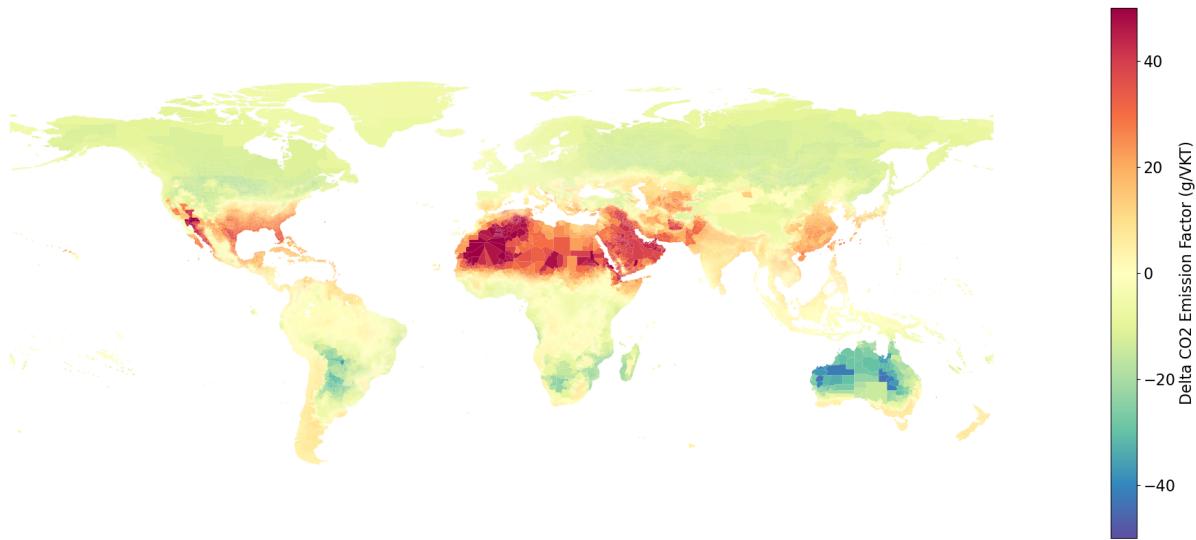
countries are motorcycles. This is highlighted in Figure 18 below: Indonesia and India have vehicle registrations around 75% for motorcycles.



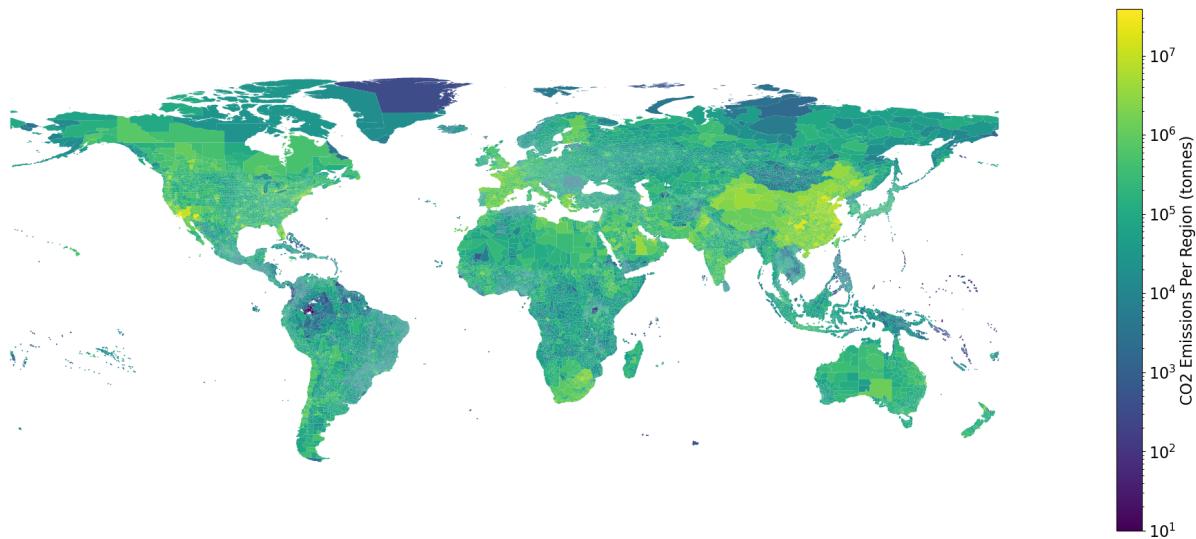
**Figure 18** Vehicle fleet mix for four countries highlighting the difference in motorcycle usage.

In Figure 19, the effects of the temperature-based adjustments to fuel efficiency are shown (Section 2.2.9) by plotting the difference between the emission factors between July and January 2024. In locations such as the Southern U.S., there is a clear increase in the emission factor as January (winter) temperatures are close to the nominal 72° F and July (summer) temperatures are much higher than nominal. In locations such as Northern Australia, with summer in January, the reverse is true. More subtly, locations on the southern extremes of the southern continents show a higher emission factor in January than July; this is because the January temperatures are close to nominal (72° F) and the July temperatures are much colder (see Figure 8 for more details).

Additionally, Figure 20 displays the estimated total CO<sub>2</sub> emissions for 2024. Not unexpectedly, some of the highest emitting regions are in the U.S., Europe, and China. In contrast to Figure 17, the emissions in India and Indonesia are much more typical, as the comparatively low emission factors are balanced by higher activity.



**Figure 19** Difference in CO<sub>2</sub> Emission Factors between July, 2024 and January, 2024. Units are grams per VKT. Regions shown are GADM Level-2 when available, GADM Level-1 otherwise.



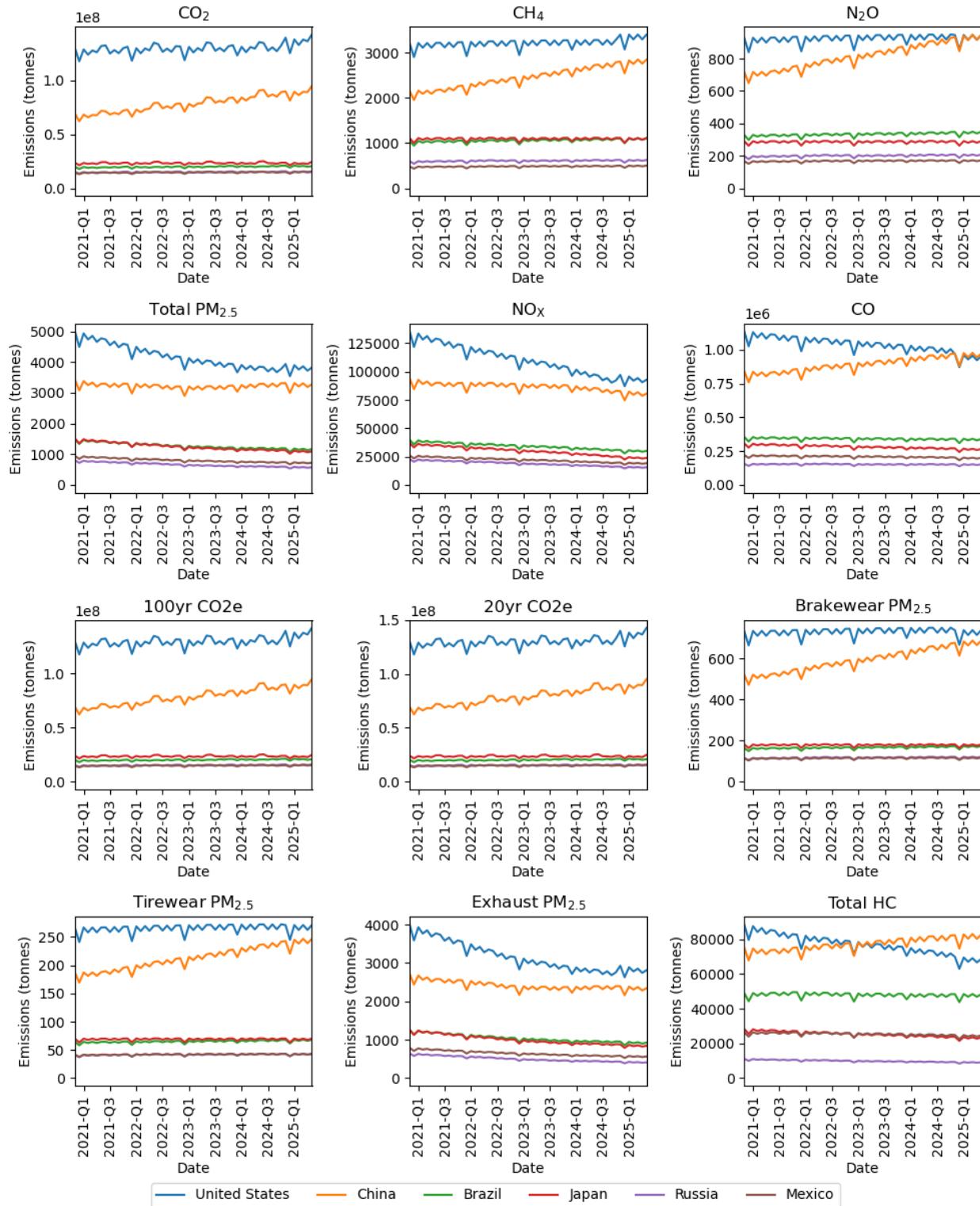
**Figure 20** Total estimated annual CO<sub>2</sub> emissions for 2024. Units are tonnes. Regions shown are GADM Level-2 when available, GADM Level-1 otherwise. Note emissions are shown per region; locations where an area is broken into many small GADM Level-2 regions will show much lower emissions than an equivalently emitting location of the same area contained within just one GADM Level-2 region.

Figure 21 shows the estimated emissions on a monthly basis for six of the top emitting countries for all twelve of the modeled GHG and non-GHG emissions. As these are plots of the total emissions for each month, there are month-to-month variations due to the number of days in each month, especially noticeable as a dip in February each year. CO<sub>2</sub> and CO<sub>2e</sub> are more dynamic than the other emissions as the model has a temperature-dependence for CO<sub>2</sub> but not the other factors

(Section 2.3.3). In general, increases are largely due to increased activity; noticeable decreases in emissions such as NO<sub>x</sub>, CO, and Exhaust PM<sub>2.5</sub> are driven by improved emissions controls. The largest individual country contributions to emissions are from the U.S. and China, with the exception of Brazil which due to its uniquely heavy use of ethanol fuel has a significant contribution to the total hydrocarbon (Total HC) emissions.

Additional highlights include:

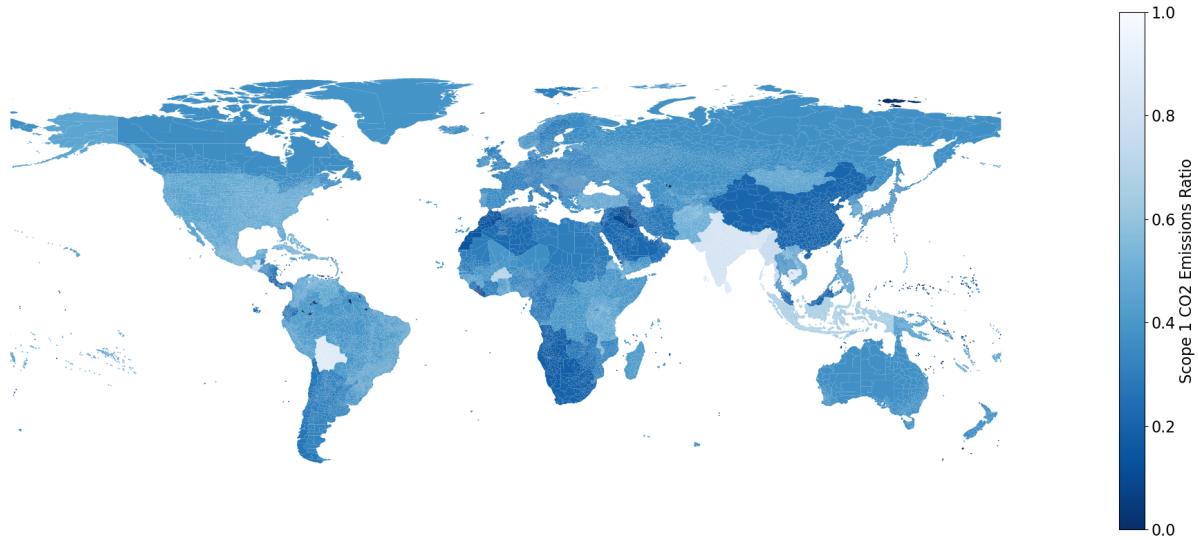
- We estimate that between 2023 and 2024, an extra  $177 \times 10^6$  extra tonnes of CO<sub>2</sub> were emitted from road transportation; this is a 2.7% increase
- Per country income class (as defined by the World Bank The World Bank 2025), between 2023 and 2024 we estimate High Income Countries (HIC) saw an increase of  $30 \times 10^6$  tonnes of CO<sub>2</sub>, while Upper-Middle Income Countries (UMIC) increased by  $102 \times 10^6$  tonnes, Lower-Middle Income Countries (LMIC) by  $36 \times 10^6$  tonnes, and Low Income Countries (LIC) by  $8 \times 10^6$  tonnes
- Of the 2024 estimated emissions, 50% is from HIC, 35% UMIC, 12% LMIC, and 3% LIC



**Figure 21** Monthly emissions for the top 6 emitting countries (by 100yr CO<sub>2</sub>e in 2024) for all estimated gases

### 3.4 Emission Reduction Strategy Results

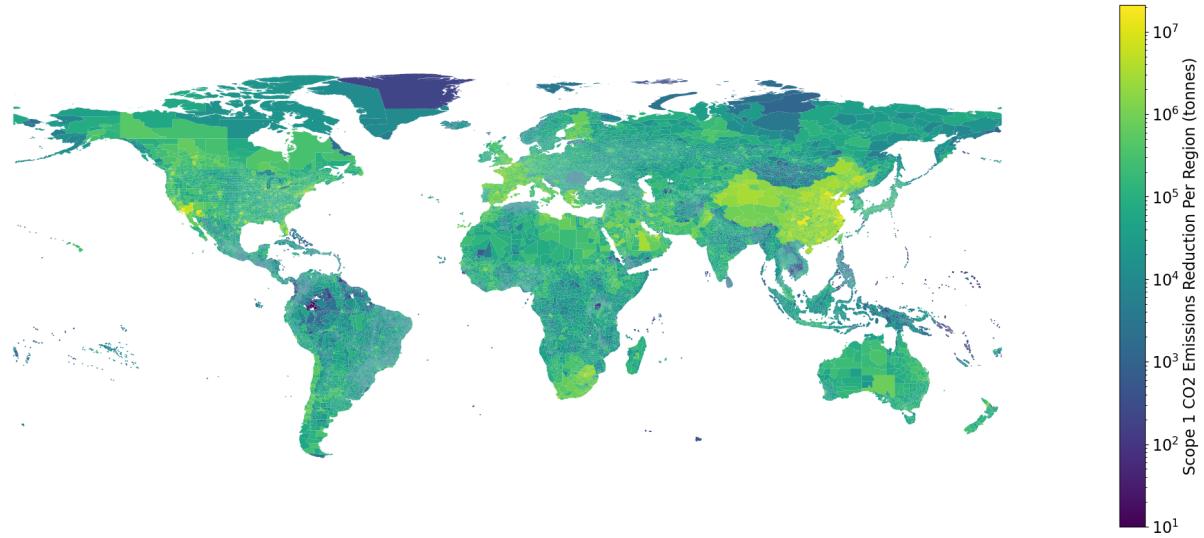
While a full understanding of the emission reduction potential of the proposed emission reduction strategy requires a cross-sector analysis, net emissions are projected to be reduced except in a small minority of countries where the current electrical grid is exceptionally dirty. Those regions are still expected to result in a net decrease however if the induced activity is met with new, clean electrical power sources. In this document, presented results will be focused on results solely within the road transportation sector, namely scope 1 emissions and the induced activity rates used for calculating scope 2 emissions.



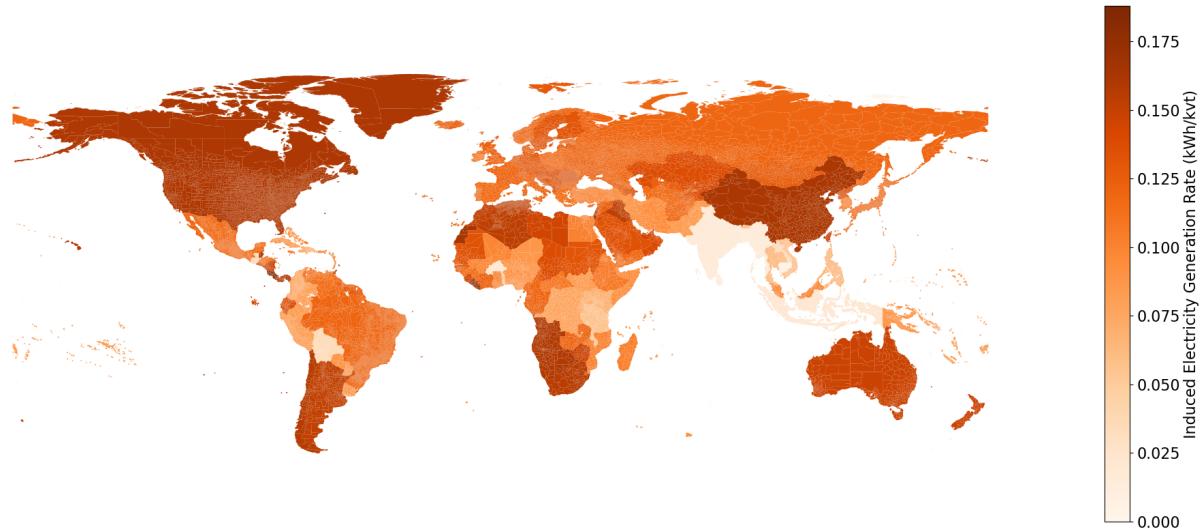
**Figure 22** Global map of the ratio  $\rho_{r,g=CO_2}$  (Equation (35)) of 2024 CO<sub>2</sub> emissions under the ERS strategy compared to baseline 2024 values; darker blue shades indicate greater relative emissions reduction potential. Regions shown are GADM Level-2 when available, GADM Level-1 otherwise.

Figure 22 shows the computed CO<sub>2</sub> emissions ratio  $\rho_{r,g=CO_2}$  for the year 2024. 2024 was used as it was the last full year as of writing, and the reference values for emissions and activity in Equation (32) are from 2024. Locations with extreme values largely match those in Figure 17 and for much the same reasons. Locations such as India and Indonesia have low proportions of passenger cars within their vehicle fleet, so there is comparatively less reductions available under the ERS scenario, and thus the emissions ratio is higher. In contrast, China for instance stands out with a low ratio (comparatively large reduction available) due to the high proportion of passenger vehicles within its vehicle fleet (Figure 18).

Multiplying 1 minus the emissions ratio by the total CO<sub>2</sub> emissions (Figure 20) yields the reduction in scope 1 road transportation CO<sub>2</sub> emissions, plotted in Figure 23. While there are some subtle differences in locations such as India, the figure is visually similar to Figure 20, indicating broad potential to reduce scope 1 emissions. Summing the emissions reduction across the entire globe suggests a total reduction in scope 1 CO<sub>2</sub> emissions of  $4.2 \times 10^9$  tonnes, or 63% by converting gasoline passenger cars to BEVs.



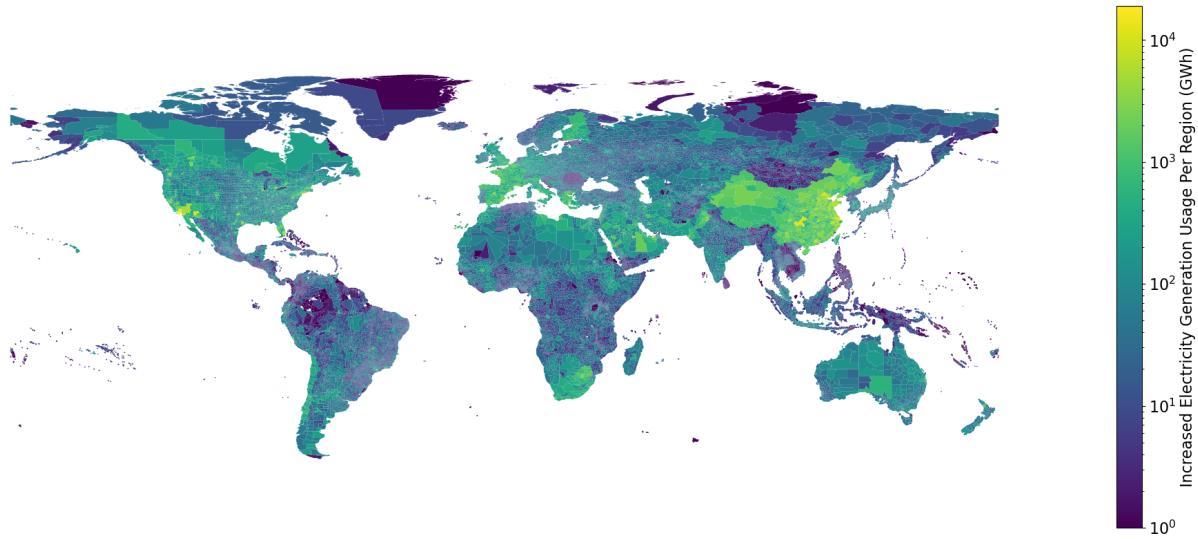
**Figure 23** Total estimated annual reduction per region in CO<sub>2</sub> emissions under the ERS strategy for 2024 values; lighter colors indicate higher emissions reduction potential. Units are tonnes per year.



**Figure 24** Global map of the induced electricity generation rate  $\alpha_{1,r}$  (Equation (35)) due to new BEVs under the ERS strategy; darker orange shades indicate higher induced electricity usage rates. Units are in kWh per kilometer vehicle traveled.

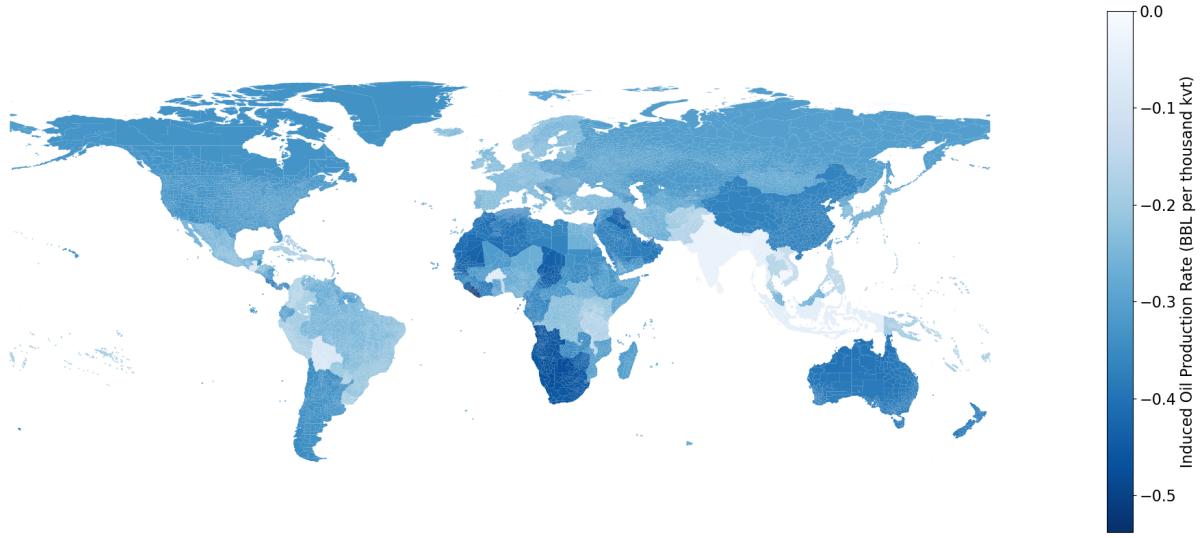
Figure 24 indicates the rate at which additional electricity generation is required per unit of activity (across all vehicle types). This data again largely tracks the proportion of vehicles that are passenger cars. The USA has a higher rate than China despite having a lower proportion of passenger cars because the passenger car fleet within the USA skews towards heavier and less efficient vehicles than in China (Cazzola, Paoli, and Teter 2023). Figure 25 shows the increase in electricity usage by multiplying the annual activity in each region by the induced rate  $\alpha_{1,r}$ . Globally, there

is an increased usage of electricity of approximately 3400 TWh annually as a result of the ERS, with the largest contributions occurring in the USA, China, and Europe where there are both high activity values and high values of the induced rate  $\alpha_{1,r}$ .

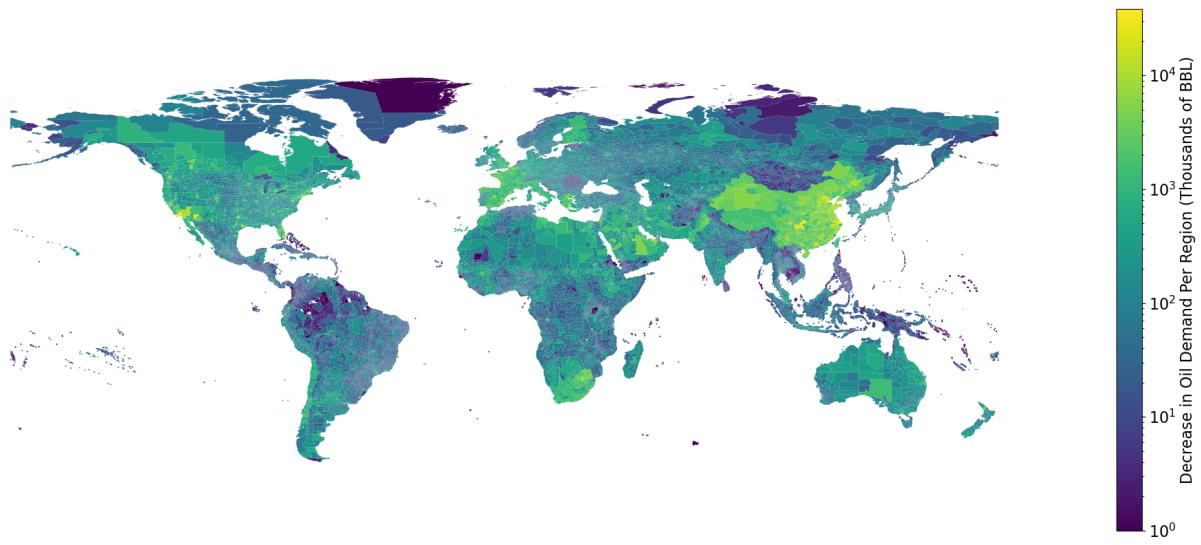


**Figure 25** Global map of the increase in electricity usage per region due to new BEVs under the ERS strategy; lighter colors indicate higher induced electricity usage. Units are in GWh per year.

Figure 26 shows the rate  $\alpha_{2,r}$  at which the demand for oil & gas production decreases within the ERS scenario. This largely tracks Figure 22 as expected. Figure 27 shows the decrease in total demand given by multiplying  $\alpha_{2,r}$  by the total activity within a region. Summing over the entire globe suggests a total reduction of  $7.7 \times 10^9$  barrels of oil per year as a result of the emissions reduction strategy.



**Figure 26** Global map of the induced oil & gas production rate  $\alpha_{2,r}$  (Equation (35)) due to the removal of gasoline-powered passenger cars under the ERS strategy; darker blue shades indicate greater reductions in oil demand rates. Values are negative indicating a decrease in oil demand, with units in barrels per thousand vehicle-kilometers traveled.



**Figure 27** Global map of the decrease in oil demand due to the removal of gasoline-powered passenger cars under the ERS strategy; lighter colors indicate greater reductions in oil demand. Units are in thousands of barrels per year.

#### 4 Discussion and Conclusions

We have presented a hybrid road transportation emissions estimation method that is detailed, scalable, and easy to update. The ability to calculate emissions per road segment can be further refined

to reach an unprecedented level of detail and global coverage. Where available, the integration of real-time traffic data would increase the temporal resolution and accuracy of our models. This type of actionable emissions monitoring data will be critical to ensuring we meet global emissions reduction targets and may inspire new ways of mitigating the effects of climate change. Finally, more local vehicle registration or usage data would improve our emissions estimates regardless of a change in the underlying model predictions of average traffic.

We have also developed an uncertainty quantification pipeline (see Section 2.5) that integrates IPCC's Approach 1 (propagation of uncertainty) and Approach 2 (Monte Carlo simulations) for combining uncertainties. Various areas for further exploration remain, including improving aspects of the data and the uncertainty quantification process. One opportunity for reducing uncertainty is to improve the spatial and temporal resolution of the variables contributing to emission factors, and extending emissions factors by vehicle years. In extending uncertainty analyses, we will continue exploring the uncertainties of urban and rural regions separately. We noted that the distribution of emissions factors varied by urban and region areas, with rural emissions factors often dominating the overall distribution of emissions factors and separate treatment of those distributions may be beneficial. Sensitivity analyses to identify the contributions to uncertainty of each variable will be helpful, as well as assessing uncertainty in trends.

An initial emission reduction strategy for the road transportation sector of converting gasoline-powered passenger cars to battery-electric vehicles has also been presented and analyzed, with estimations of effects to scope 1 and scope 2 emissions. While BEV adoption is occurring organically in various locales, challenges can be addressed and widespread adoption aided through policy; references from GFEI (Cazzola, Paoli, and Teter 2023, Cazzola, Teter, and Craglia 2025) and MIT Center for Energy and Policy Research (Knittel and Tanaka 2025) provide some guidance. How any policy shift is structured will impact the specific emission impacts; e.g. replacing all ICE vehicles immediately has a different net impact than if ICE vehicles are replaced at end-of-life, and policies can be structured to encourage smaller, more efficient BEVs. Furthermore, the nature of the global supply chain for this sector makes scope 3 estimations difficult, however some estimates for full life-cycle costs for specific use cases are available (Smith et al. 2025).

## 5 Acknowledgements

Thanks to Aaron Davitt and Lekha Sridhar from WattTime for their help in reviewing our methodology. We thank previous team members for their contributions to the road transportation emissions sector:

- Melat Ghebreselassie
- Marisa Hughes
- Tomek Kott
- Carlos Levy Capote
- Wayne Loschen
- Marisel Villafane-Delgado
- Kevin Zhang

## 6 Supplementary Materials

### 6.1 Vehicle Fleet Mix Sources

**Table S1** List of Vehicle Fleet Mix Sources for specific countries

<b>Reference:</b> ACEA dataset: (ACEA - European Automobile Manufacturers' Association 2022)	
<ul style="list-style-type: none"><li>• Russian Federation</li><li>• Republic of Turkiye</li><li>• United Kingdom of Great Britain and Northern Ireland</li><li>• Federal Republic of Germany</li><li>• French Republic</li><li>• Kingdom of Spain</li><li>• Italian Republic</li><li>• Republic of Poland</li></ul>	<ul style="list-style-type: none"><li>• Kingdom of the Netherlands</li><li>• Portuguese Republic</li><li>• Kingdom of Belgium</li><li>• Kingdom of Sweden</li><li>• Czech Republic</li><li>• Kingdom of Denmark</li><li>• Ireland</li><li>• Republic of Finland</li></ul>

Continued on next page

**Table S1** (Continued)

<b>Reference:</b> UNECE dataset: (UNECE - United Nations Economic Commission for Europe 2022)	
<ul style="list-style-type: none"><li>• Hellenic Republic</li><li>• Republic of Belarus</li><li>• State of Israel</li><li>• Republic of Azerbaijan</li><li>• Swiss Confederation</li><li>• Republic of Hungary</li><li>• Republic of Austria</li><li>• Republic of Bulgaria</li><li>• Republic of Croatia</li><li>• Republic of Lithuania</li><li>• Republic of Moldova</li></ul>	<ul style="list-style-type: none"><li>• Bosnia and Herzegovina</li><li>• Republic of Latvia</li><li>• Republic of Cyprus</li><li>• Republic of Estonia</li><li>• Republic of Slovenia</li><li>• Republic of Malta</li><li>• Republic of Iceland</li><li>• Grand Duchy of Luxembourg</li><li>• Montenegro</li></ul>
<b>Reference:</b> ASEANStat dataset: (ibid.)	
<ul style="list-style-type: none"><li>• Republic of Indonesia</li><li>• Republic of Philippines</li><li>• Kingdom of Cambodia</li><li>• Lao People's Democratic Republic</li><li>• Malaysia</li></ul>	<ul style="list-style-type: none"><li>• Singapore</li><li>• Thailand</li><li>• Vietnam</li><li>• Myanmar</li><li>• Brunei</li></ul>

Continued on next page

**Table S1** (Continued)

<b>Reference:</b> WHO dataset (World Health Organization 2018)
<ul style="list-style-type: none"><li>• Japan</li><li>• Arab Republic of Egypt</li><li>• Islamic Republic of Iran</li><li>• Republic of Colombia</li><li>• Republic of Iraq</li><li>• Bolivarian Republic of Venezuela</li><li>• Republic of the Sudan</li><li>• Kingdom of Morocco</li><li>• Republic of Peru</li><li>• Republic of Ghana</li><li>• Commonwealth of Australia</li><li>• United Republic of Tanzania</li><li>• Islamic Republic of Afghanistan</li><li>• Republic of Ecuador</li><li>• Republic of Cote d'Ivoire</li><li>• Republic of Senegal</li><li>• Syrian Arab Republic</li><li>• Republic of Kazakhstan</li><li>• United Arab Emirates</li><li>• Democratic Socialist Republic of Sri Lanka</li><li>• Dominican Republic</li><li>• Romania</li><li>• Plurinational State of Bolivia</li><li>• Federal Democratic Republic of Nepal</li><li>• Republic of Tunisia</li><li>• Burkina Faso</li><li>• Republic of Cuba</li><li>• Republic of Guinea</li><li>• Republic of Zimbabwe</li><li>• Republic of Benin</li><li>• State of Palestine</li><li>• Lebanese Republic</li><li>• Togolese Republic</li><li>• Republic of Paraguay</li><li>• State of Libya</li><li>• Republic of Honduras</li><li>• Republic of El Salvador</li><li>• Kingdom of Cambodia</li><li>• Republic of Madagascar</li><li>• Republic of Costa Rica</li><li>• Republic of Burundi</li><li>• Republic of Panama</li><li>• Oriental Republic of Uruguay</li><li>• Republic of South Sudan</li><li>• State of Eritrea</li><li>• Republic of Liberia</li><li>• Sultanate of Oman</li><li>• Republic of Rwanda</li><li>• Mongolia</li><li>• Kingdom of Norway</li><li>• Republic of the Gambia</li><li>• Slovak Republic</li><li>• Republic of Botswana</li><li>• Republic of Trinidad and Tobago</li><li>• Republic of Mauritius</li><li>• Republic of Namibia</li><li>• Democratic Republic of Timor-Leste</li><li>• Co-operative Republic of Guyana</li><li>• Republic of Suriname</li><li>• Republic of Maldives</li><li>• Barbados</li><li>• Belize</li></ul>

Other countries for which we have specific data:

- Albania (ALB): (Institute of Statistics (INSTAT) 2025)
- Argentina (ARG): (ADEFA n.d.)
- Australia (AUS): (Australian Bureau of Statistics 2021)
- Brazil (BRA): (*Fleet Size of the Auto Industry in Brazil by Type 2021 2022*)
- Canada (CAN): (Statistics Canada 2017)
- Chile (CHL)
- People's Republic of China (CHN)
- Guatemala (GTM)
- India (IND)
- Indonesia (IDN)
- Japan (JPN)
- Kuwait (KWT)
- Malaysia (MYS)
- Myanmar (MMR)
- Singapore (SGP)

## 6.2 Supplementary Materials Metadata

**Table S2** General dataset information

General Description	Definition
<b>Sector definition</b>	Emissions from all road-based transportation such as cars, buses, motorcycles, and trucks.
<b>UNFCCC sector equivalent</b>	1.A.3.b (“Road Transportation”)
<b>Temporal Coverage</b>	2015 – 2024
<b>Temporal Resolution</b>	Monthly
<b>Data format(s)</b>	CSV
<b>Coordinate Reference System</b>	EPSG:4326, decimal degrees

Continued on next page

**Table S2** (Continued)

General Description	Definition
<b>Number of sources available for download and percent of global emissions (as of 2024)</b>	Almost 48,000 GADM Level-2 regions covering the entire globe with the exception of Antarctica
<b>Total emissions for 2024</b>	$6.73 \times 10^9$ tonnes of 100yr CO2e
<b>Ownership</b>	N/A: estimates rely on modeling individual vehicles; no ownership is possible.
<b>What emission factors were used?</b>	Model based
<b>What is the difference between a “NULL / none / nan” versus “0” data field?</b>	“0” values are for true non-existent emissions. If we know that the sector has emissions for that specific gas, but the gas was not modeled, this is represented by ‘NULL/none/nan’
<b>total_CO2e_100yrGWP and total_CO2e_20yrGWP conversions</b>	Climate TRACE uses IPCC AR6 CO2e GWPs (CO2e conversion guidelines available from Climate Change (IPCC) 2021)

**Table S3** Asset level metadata description

Data Attribute	Definition
<b>sector</b>	Transportation
<b>asset_sub-sector_name</b>	Road transportation
<b>asset definition</b>	Emissions from on-road transportation by GADM based administrative regions. See Section 2.2.2.
<b>start_date</b>	Start date for time period of emissions estimation (YYYY-MM-DD format)

Continued on next page

**Table S3 (Continued)**

Data Attribute	Definition
<b>end_date</b>	End date for time period of emissions estimation (YYYY-MM-DD format)
<b>asset_identifier</b>	GADM Level-1 or GADM Level-2 GADM ID
<b>model_number</b>	String identifying model & source data version
<b>asset_name</b>	GADM Level-1 or GADM Level-2 name
<b>iso3_country</b>	ISO 3166-1 alpha-3 country code for city taken from GADM ADM_2 layer value
<b>location</b>	GADM Level-1 or GADM Level-2 GADM ID
<b>type</b>	Not used; N/A
<b>capacity_description</b>	Total road network length for each region. See Section 2.3.4
<b>capacity_units</b>	kilometers
<b>capacity_factor_description</b>	Monthly activity divided by capacity for each region. See Section 2.3.8
<b>capacity_factor_units</b>	vehicles per month
<b>activity_description</b>	Total vehicle kilometers traveled (VKT) per month for each region. See Section 2.3.5
<b>activity_units</b>	vehicle-kilometers per month
<b>CO2_emissions_factor</b>	tonnes CO <sub>2</sub> /VKT; see Section 2.3.8
<b>CH4_emissions_factor</b>	tonnes CH <sub>4</sub> /VKT; see Section 2.3.8
<b>N2O_emissions_factor</b>	tonnes N <sub>2</sub> O/VKT; see Section 2.3.8
<b>PM2_5_emissions_factor</b>	tonnes Total PM <sub>2.5</sub> /VKT; see Section 2.3.8
<b>NOX_emissions_factor</b>	tonnes Total NO <sub>x</sub> /VKT; see Section 2.3.8
<b>CO_emissions_factor</b>	tonnes Total CO/VKT; see Section 2.3.8
<b>CO2_emissions</b>	tonnes CO <sub>2</sub> ; see Section 2.3.1
<b>CH4_emissions</b>	tonnes CH <sub>4</sub> ; see Section 2.3.1
<b>N2O_emissions</b>	tonnes N <sub>2</sub> O; see Section 2.3.1
<b>PM2_5_emissions</b>	tonnes Total PM <sub>2.5</sub> ; see Section 2.3.1
<b>NOX_emissions</b>	tonnes NO <sub>x</sub> ; see Section 2.3.1
<b>CO_emissions</b>	tonnes CO; see Section 2.3.1

Continued on next page

**Table S3** (Continued)

Data Attribute	Definition
<b>total_CO2e_100yrGWP</b>	tonnes 100yr CO2e; see Section 2.3.1
<b>total_CO2e_20yrGWP</b>	tonnes CO2e; see Section 2.3.1
<b>other1_description</b>	The GADM ID of the parent region; for GADM Level-2 assets, this is the GADM Level-1 ID; for GADM Level-1 assets, this is the GADM Level-0 ID
<b>other1_units</b>	N/A
<b>other2_description</b>	The GADM ID of the asset
<b>other2_units</b>	N/A
<b>other3_description</b>	The GADM level of the asset
<b>other3_units</b>	N/A
<b>other4_description</b>	PM2.5 emissions from brakewear
<b>other4_units</b>	tonnes
<b>other5_description</b>	PM2.5 emissions from tirewear
<b>other5_units</b>	tonnes
<b>other6_description</b>	PM2.5 emissions from exhaust
<b>other6_units</b>	tonnes
<b>other7_description</b>	Total non-GHG HC emissions
<b>other7_units</b>	tonnes
<b>other8_description</b>	N/A
<b>other8_units</b>	N/A
<b>other9_description</b>	N/A
<b>other9_units</b>	N/A
<b>other10_description</b>	N/A
<b>other10_units</b>	N/A

**Table S4** Asset level metadata description confidence and uncertainty

Data Attribute	Confidence Definition	Uncertainty Definition
<b>type</b>	Not used; N/A	Not used; N/A

Continued on next page

**Table S4 (Continued)**

Data Attribute	Confidence Definition	Uncertainty Definition
<b>capacity_description</b>	See Section 2.5.8	See Section 2.5.2
<b>capacity_units</b>	See Section 2.5.8	Kilometers
<b>capacity_factor_description</b>	See Section 2.5.8	See Section 2.5.6
<b>capacity_factor_units</b>	See Section 2.5.8	N/A
<b>activity_description</b>	Section 2.5.8	See Section 2.5.2
<b>activity_units</b>	See Section 2.5.8	vehicle-kilometers
<b>CO2_emissions_factor</b>	See Section 2.5.8	See Section 2.5.5
<b>CH4_emissions_factor</b>	See Section 2.5.8	See Section 2.5.5
<b>N2O_emissions_factor</b>	See Section 2.5.8	See Section 2.5.5
<b>other_gas_emissions_factor</b>	See Section 2.5.8	See Section 2.5.5
<b>CO2_emissions</b>	See Section 2.5.8	See Section 2.5.2
<b>CH4_emissions</b>	See Section 2.5.8	See Section 2.5.2
<b>N2O_emissions</b>	See Section 2.5.8	See Section 2.5.2
<b>other_gas_emissions</b>	See Section 2.5.8	See Section 2.5.2
<b>total_CO2e_100yrGWP</b>	See Section 2.5.8	See Section 2.5.7
<b>total_CO2e_20yrGWP</b>	See Section 2.5.8	See Section 2.5.7

### 6.3 Emission Reduction Strategy (ERS) Tables

**Table S5** Strategy ERS Table: each row in the data file represents the same strategy of replacing gasoline-powered passenger cars with BEVs, only differing in specific values that vary by region

Column	Definition
<b>strategy_id</b>	String formatted as “[F]c[CCC]m[MMM]n[NNN]wh[EEE]g[GGG]” where the values in brackets are as follows: F : H, M, or C to match confidence attribute CCC : $10^3 \times co2\_emissions\_factor\_new\_to\_old\_ratio$ MMM : $10^3 \times ch4\_emissions\_factor\_new\_to\_old\_ratio$ NNN : $10^3 \times n2o\_emissions\_factor\_new\_to\_old\_ratio$ EEE : $10^6 \times induced\_sector\_1\_activity\_conversion\_rate$ GGG : $-10^6 \times induced\_sector\_2\_activity\_conversion\_rate$
<b>strategy_name</b>	“Transition to EVs”
<b>strategy_description</b>	“Replace gasoline passenger cars with electric vehicles”
<b>mechanism</b>	“retrofit”
<b>max_activity_affected_ratio</b>	1
<b>co2_emissions_factor_new_to_old_ratio</b>	$\rho_{r,g=CO_2} = 10^{-3} \cdot [CCC]$ ; see Equation (32)
<b>ch4_emissions_factor_new_to_old_ratio</b>	$\rho_{r,g=CH_4} = 10^{-3} \cdot [MMM]$ ; see Equation (32)
<b>n2o_emissions_factor_new_to_old_ratio</b>	$\rho_{r,g=N_2O} = 10^{-3} \cdot [NNN]$ ; see Equation (32)
<b>confidence</b>	One of “high”, “medium”, or “low”; see Section 2.7
<b>induced_sector_1</b>	“electricity-generation”
<b>induced_sector_1_activity_conversion_rate</b>	$\alpha_{1,r} = 10^{-6} \cdot [EEE]$ ; see Equation (37)
<b>induced_sector_1_activity_conversion_rate_units</b>	“MWh per KVT”
<b>induced_sector_2</b>	“electricity-generation”
<b>induced_sector_2_activity_conversion_rate</b>	$\alpha_{2,r} = -10^{-6} \cdot [GGG]$ ; see Equation (40)

Continued on next page

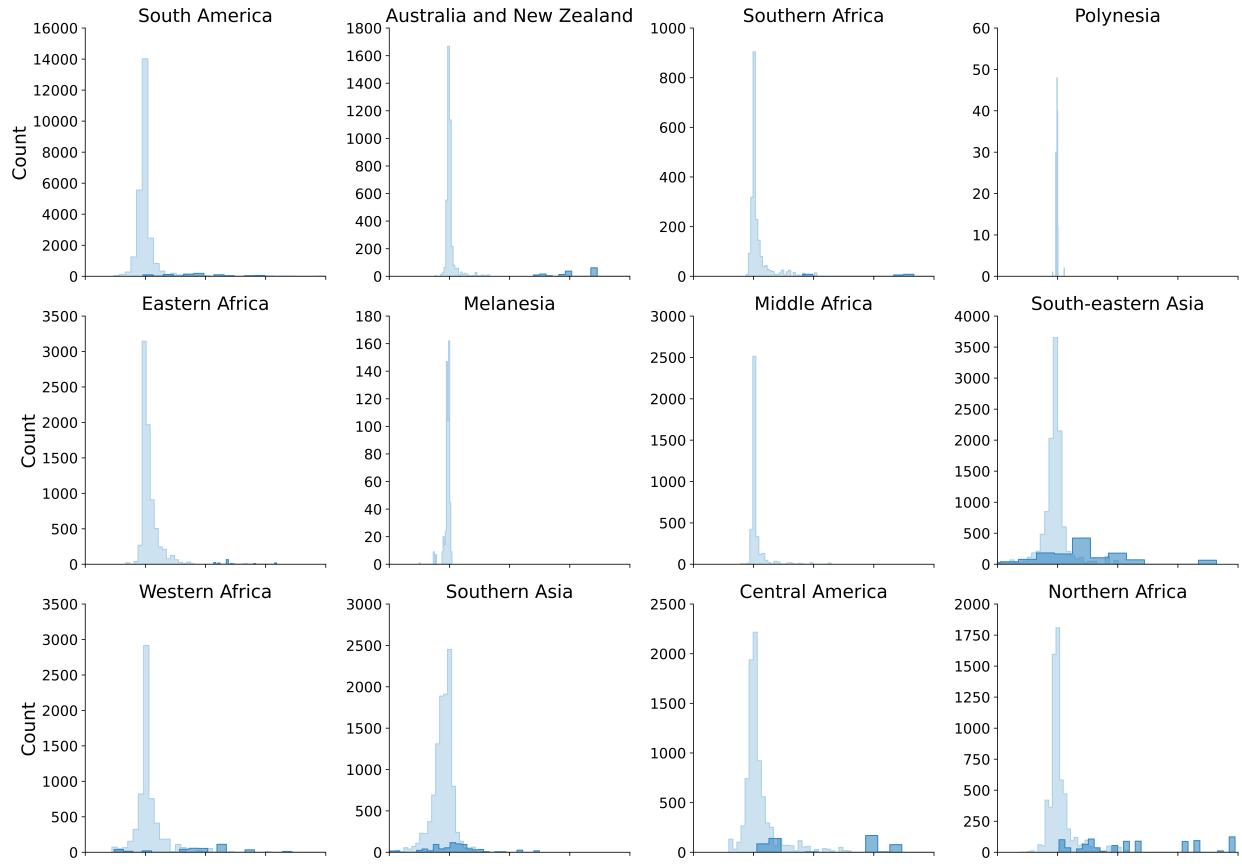
**Table S5** (Continued)

Column	Definition
<b>induced_sector_2_activity_conversion_rate_units</b>	“BBL per KVT”

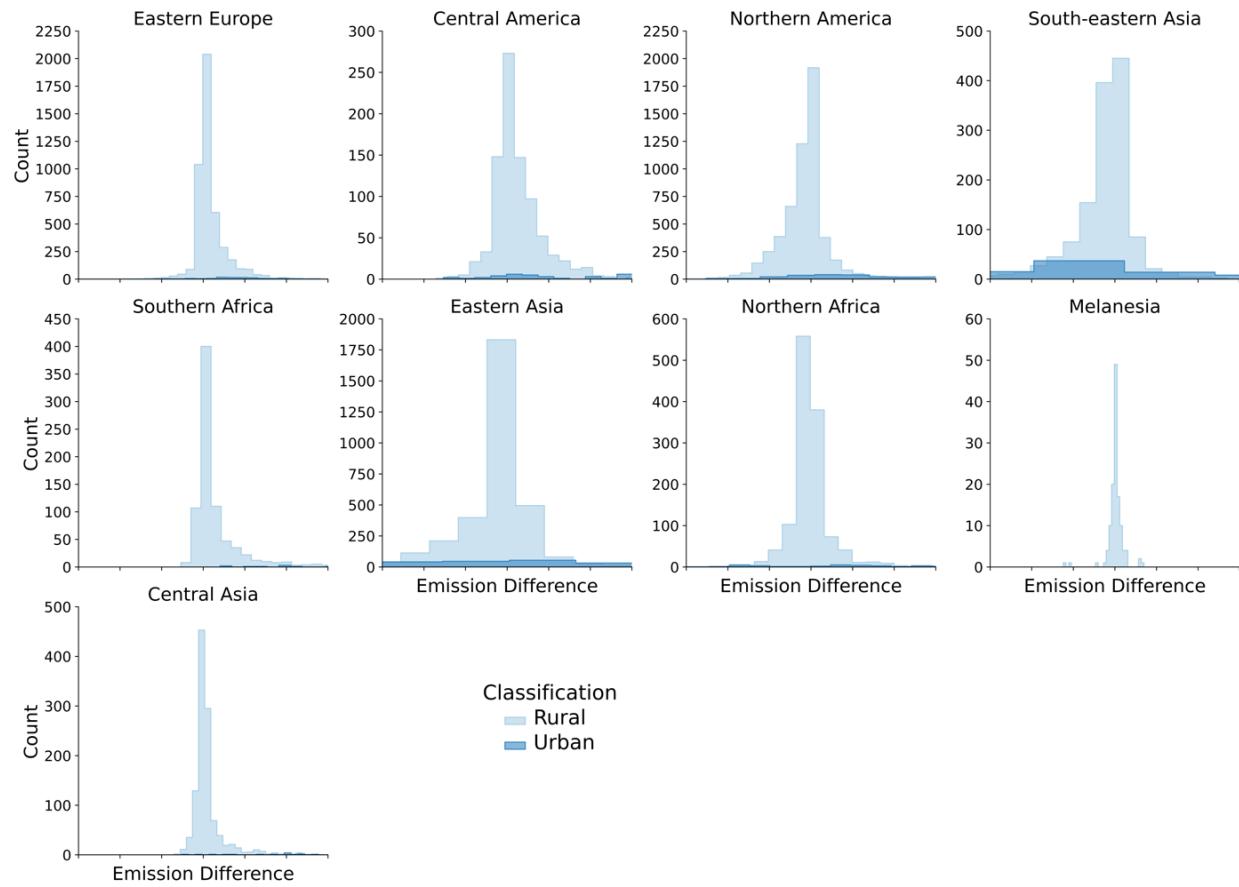
**Table S6** Example rows from Strategy Crosswalk Table. Table links asset\_identifiers (for either GADM Level-1, GADM Level-2, or GHS FUA assets) to specific strategy\_ids in the Strategy table. Currently, strategy\_rank values are always 1, as only one strategy is defined.

asset_identifier	strategy_id	strategy_rank
ARM.1_1	Mc315m702n586wh114g307	1
ARM.2_1	Mc315m702n586wh114g308	1
ARM.3_1	Mc315m702n586wh114g307	1
...	...	...
ghs-fua_10542	Mc822m952n959wh017g038	1

## 6.4 Difference Between Estimates and EDGAR



**Figure 28** Comparison of actual difference between estimates and EDGAR for all UN regions, split between rural and urban regions.



**Figure 28 (continued)** Comparison of actual difference between estimates and EDGAR for all UN regions, split between rural and urban regions.

## **7 Administrative Concerns**

### **7.1 Permissions and Use**

All Climate TRACE data is freely available under the Creative Commons Attribution 4.0 International Public License, unless otherwise noted below.

### **7.2 Citation format**

Kevin Foster, Patrick Sicurello, Christine Piatko, and Elizabeth Reilly (2025). Transportation Sector: Global Road Emissions. The Johns Hopkins University Applied Physics Laboratory (JHU/APL), Laurel, MD, USA, Climate TRACE Emissions Inventory. <https://climatetrace.org> [Accessed date]

### **7.3 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.

### **7.4 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.

## 8 References

1. ACEA - European Automobile Manufacturers' Association (2022). *Report - Vehicles in Use, Europe 2022*. URL: <https://www.acea.auto/files/ACEA-report-vehicles-in-use-europe-2022.pdf>.
2. ADEFA (n.d.). *Anuario 2020 — Estadísticas*. Accessed December 14, 2022. URL: <http://adefra.org.ar/es/estadisticas-anuarios-interno?id=55>.
3. ASEANstats Data Portal (2025). Indicators used: ASE.TRP.ROD.B.006; ASE.TRP.ROD.B.007; ASE.TRP.ROD.B.008; ASE.TRP.ROD.B.009; ASE.TRP.ROD.B.011; URL: <https://data.aseanstats.org>.
4. Australian Bureau of Statistics (2021). *Motor Vehicle Census, Australia, 31 Jan 2021*. URL: <https://www.abs.gov.au/statistics/industry/tourism-and-transport/motor-vehicle-census-australia/latest-release>.
5. Boeing, Geoff (2017). “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks”. In: *Computers, Environment and Urban Systems* 65, pp. 126–139. ISSN: 0198-9715. DOI: 10.1016/j.comenvurbssys.2017.05.004.
6. British Petroleum (2022). *bp Statistical Review of World Energy*. 71st ed. URL: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf>.
7. Brody, Shaked, Uri Alon, and Eran Yahav (2022). “How Attentive are Graph Attention Networks?” In: *International Conference on Learning Representations*. URL: <https://openreview.net/forum?id=F72ximsx7C1>.
8. Bronstein, Michael M. et al. (2017). “Geometric Deep Learning: Going beyond Euclidean data”. In: *IEEE Signal Processing Magazine* 34.4, pp. 18–42. DOI: 10.1109/MSP.2017.2693418.
9. Cazzola, Pierpaolo, Leonardo Paoli, and Jacob Teter (2023). *Trends in the Global Vehicle Fleet 2023: Managing the SUV and the EV Transition*. Global Fuel Economy Initiative (GFEI). DOI: 10.7922/G2HM56SV. URL: <https://www.globalfueleconomy.org/data-and-research/publications/trends-in-the-global-vehicle-fleet-2023>.
10. — (Nov. 2024). *Trends in global fuel economy of new vehicles: 2005 - 2022*. Zenodo. DOI: 10.5281/zenodo.14046031. URL: <https://doi.org/10.5281/zenodo.14046031>.
11. Cazzola, Pierpaolo, Jacob Teter, and Matteo Craglia (June 2025). *Small and electric - The international case to move away from combustion SUVs*. European Transport and Energy Research Centre, Institute of Transportation Studies, University of California, Davis. URL: <https://www.globalfueleconomy.org/resources/gfei-working-paper-25-small-and-electric-the-international-case-to-move-away-from-combustion-suvs>.
12. Climate Change (IPCC), Intergovernmental Panel on (June 2021). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Re-*

- port of the Intergovernmental Panel on Climate Change*. Cambridge University Press. ISBN: 9781009157896. DOI: 10.1017/9781009157896.
13. Crippa, M. et al. (2025). *GHG emissions of all world countries: 2025 Report*. JRC143227. Publications Office of the European Union. DOI: 10.2760/9816914.
  14. Crippa, Monica et al. (Apr. 2020). “High Resolution Temporal Profiles in the Emissions Database for Global Atmospheric Research”. In: *Scientific Data* 7.1. ISSN: 2052-4463. DOI: 10.1038/s41597-020-0462-2.
  15. Drusch, M. et al. (2012). “Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services”. In: *Remote Sensing of Environment* 120. The Sentinel Missions - New Opportunities for Science, pp. 25–36. ISSN: 0034-4257. DOI: 10.1016/j.rse.2011.11.026.
  16. Efron, Bradley and R.J. Tibshirani (May 1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC. ISBN: 9780429246593. DOI: 10.1201/9780429246593.
  17. Fan, Tongle et al. (2020). “MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation”. In: *IEEE Access* 8, pp. 179656–179665. DOI: 10.1109/ACCESS.2020.3025372.
  18. *Fleet Size of the Auto Industry in Brazil by Type 2021* (2022). Statista. URL: <https://www.statista.com/statistics/787327/automobile-industry-fleet-size-brazil-vehicle-type/>.
  19. Florczyk, Aneta et al. (2023). *GHS-UCDB R2019A - GHS Urban Centre Database 2015, Multitemporal and Multidimensional Attributes*. European Commission, Joint Research Centre (JRC). DOI: 10.2905/53473144-B88C-44BC-B4A3-4583ED1F547E.
  20. Gately, C., L.R. Hutyra, and I.S. Wing (2019). *DARTE Annual On-road CO2 Emissions on a 1-km Grid, Conterminous USA, V2, 1980-2017*. en. ORNL Distributed Active Archive Center. DOI: 10.3334/ORNLDAAAC/1735.
  21. Gately, Conor K., Lucy R. Hutyra, and Ian Sue Wing (2015). “Cities, traffic, and CO<sub>2</sub>: A multidecadal assessment of trends, drivers, and scaling relationships”. In: *Proceedings of the National Academy of Sciences* 112.16, pp. 4999–5004. DOI: 10.1073/pnas.1421723112.
  22. Global Administrative Areas (n.d.). *GADM Database of Global Administrative Areas, Version 4.1*. URL: <https://www.gadm.org>.
  23. Gurney, Kevin R. et al. (2020). “The Vulcan Version 3.0 High-Resolution Fossil Fuel CO<sub>2</sub> Emissions for the United States”. In: *Journal of Geophysical Research: Atmospheres* 125.19, e2020JD032974. DOI: 10.1029/2020JD032974.
  24. Harris, Ian et al. (Apr. 2020). “Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset”. In: *Scientific Data* 7.1. ISSN: 2052-4463. DOI: 10.1038/s41597-020-0453-3.
  25. Institute of Statistics (INSTAT) (2025). *Transport, Accidents and Characteristics of road vehicles — Statistical tables*. Tables used: Road vehicles by type. Institute of Statistics (INSTAT). URL: <https://www.instat.gov.al/en/themes/industry-transport-accidents-and-characteristics-of-road-vehicles-statistical-tables>.

trade-and-services/transport-accidents-and-characteristics-of-road-vehicles/#tab3.

26. International Energy Agency (Mar. 2019). *Transport Sector CO<sub>2</sub> Emissions by Mode in the Sustainable Development Scenario, 2000-2030 – Charts – Data & Statistics*. IEA. URL: <https://www.iea.org/data-and-statistics/charts/transport-sector-co2-emissions-by-mode-in-the-sustainable-development-scenario-2000-2030>.
27. — (Jan. 2025). *Oil Market Report - January 2025*. IEA. URL: <https://www.iea.org/reports/oil-market-report-january-2025>.
28. Knittel, Christopher R. and Shinsuke Tanaka (June 2025). “Challenges to expanding EV adoption and policy responses”. In: *Handbook on Electricity Regulation*. Edward Elgar Publishing, pp. 405–447. ISBN: 9781035314355. DOI: 10.4337/9781035314355.00025. URL: <http://dx.doi.org/10.4337/9781035314355.00025>.
29. Liu, Zhu et al. (Nov. 2020). “Carbon Monitor, a near-real-time daily dataset of global CO<sub>2</sub> emission from fossil fuel and cement production”. In: *Scientific Data* 7.1. ISSN: 2052-4463. DOI: 10.1038/s41597-020-00708-7.
30. Lohse-Busch, Henning et al. (Apr. 2013). “Ambient Temperature (20°F, 72°F and 95°F) Impact on Fuel and Energy Consumption for Several Conventional Vehicles, Hybrid and Plug-In Hybrid Electric Vehicles and Battery Electric Vehicle”. In: *SAE Technical Paper Series*. SAE International. DOI: 10.4271/2013-01-1462.
31. Main-Knorn, Magdalena et al. (2017). “Sen2Cor for Sentinel-2”. In: *Image and Signal Processing for Remote Sensing XXIII*. Ed. by Lorenzo Bruzzone. Vol. 10427. International Society for Optics and Photonics. SPIE, p. 1042704. DOI: 10.1117/12.2278218.
32. Microsoft Bing Maps (2023). *Road Detections*. Microsoft. URL: <https://github.com/microsoft/RoadDetections>.
33. Mukherjee, Ryan et al. (2021). “Towards Indirect Top-Down Road Transport Emissions Estimation”. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1092–1101. DOI: 10.1109/CVPRW53098.2021.00120.
34. OpenStreetMap Contributors (2020). *OpenStreetMap: User-Generated Street Maps*. URL: <https://planet.osm.org>.
35. Ramsey, Steven et al. (Feb. 2023). *Global Demand for Fuel Ethanol Through 2030*. BIO-05. U.S. Department of Agriculture, Economic Research Service.
36. Scheibenreif, Linus M., Michael Mommert, and Damian Borth (2021). “Estimation of Air Pollution with Remote Sensing Data: Revealing Greenhouse Gas Emissions from Space”. In: *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*. URL: <https://www.climatechange.ai/papers/icml2021/23>.
37. Smith, Elizabeth et al. (Aug. 2025). “Greenhouse Gas Reductions Driven by Vehicle Electrification across Powertrains, Classes, Locations, and Use Patterns”. In: *Environmental Sci-*

- ence & Technology* 59.37, pp. 19768–19780. ISSN: 1520-5851. DOI: 10.1021/acs.est.5c05406.
38. Statistics Canada (2017). *Vehicle Registrations, by Type of Vehicle*. Government of Canada. DOI: 10.25318/2310006701-ENG. URL: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310006701>.
  39. The World Bank (July 2025). *World Bank Country and Lending Groups*. URL: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>.
  40. Topf, Jochen (2023). *Osmium: Command Line Tool for Working with OpenStreetMap Data Based on the Osmium Library*. URL: <https://osmcode.org/osmium-tool/>.
  41. UNECE - United Nations Economic Commission for Europe (2022). *Road Vehicle Fleet at 31 December by Fuel Type, Type of Vehicle, Country and Year*. URL: [https://w3.unece.org/PXWeb2015/pxweb/en/STAT/STAT\\_\\_40-TRTRANS\\_\\_03-TRRoadFleet/03\\_en\\_TRRoadFuelFlt\\_r.px/](https://w3.unece.org/PXWeb2015/pxweb/en/STAT/STAT__40-TRTRANS__03-TRRoadFleet/03_en_TRRoadFuelFlt_r.px/).
  42. US BTS (2024). “Table 4-43 - Estimated U.S. Average Vehicle Emissions Rates per Vehicle by Vehicle Type Using Gasoline, Diesel, and Electric”. In: *National Transportation Statistics*. U.S. DOT Bureau of Transportation Statistics. URL: [https://www.bts.gov/sites/bts.dot.gov/files/2024-06/table\\_04\\_43\\_062724.xlsx](https://www.bts.gov/sites/bts.dot.gov/files/2024-06/table_04_43_062724.xlsx).
  43. US EPA (2022). *GHG Emissions Factors Hub*. URL: [https://www.epa.gov/system/files/documents/2022-04/ghg\\_emission\\_factors\\_hub.pdf](https://www.epa.gov/system/files/documents/2022-04/ghg_emission_factors_hub.pdf).
  44. — (2023). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2021*. URL: <https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2021>.
  45. — (2024). *GHG Emissions Factors Hub*. URL: <https://www.epa.gov/system/files/documents/2024-02/ghg-emission-factors-hub-2024.pdf>.
  46. US FHWA (2017). *Highway Performance Monitoring System (HPMS) Data*. URL: <https://www.fhwa.dot.gov/policyinformation/hpms>.
  47. — (2018). *State Motor-Vehicle Registrations*. URL: <https://www.fhwa.dot.gov/policyinformation/statistics/2018/mv1.cfm>.
  48. — (2020). *Distribution of Annual Vehicle Distance Traveled - Percentage by Vehicle Type Rural/Urban (Table VM-4)*. URL: <https://www.fhwa.dot.gov/policyinformation/statistics/2020/vm4.cfm>.
  49. World Bank Group (2019). *CURB: Climate Action for Urban Sustainability, Version 2.1*. URL: <https://datacatalog.worldbank.org/search/dataset/0042029>.
  50. World Health Organization (2018). *Global Status Report on Road Safety 2018*. Genève, Switzerland: World Health Organization. ISBN: 9789241565684. URL: <https://www.who.int/publications-detail-redirect/9789241565684>.
  51. World Resources Institute (2022). *Greenhouse Gas (GHG) Emissions — Climate Watch*. URL: <https://www.climatewatchdata.org/ghg-emissions>.