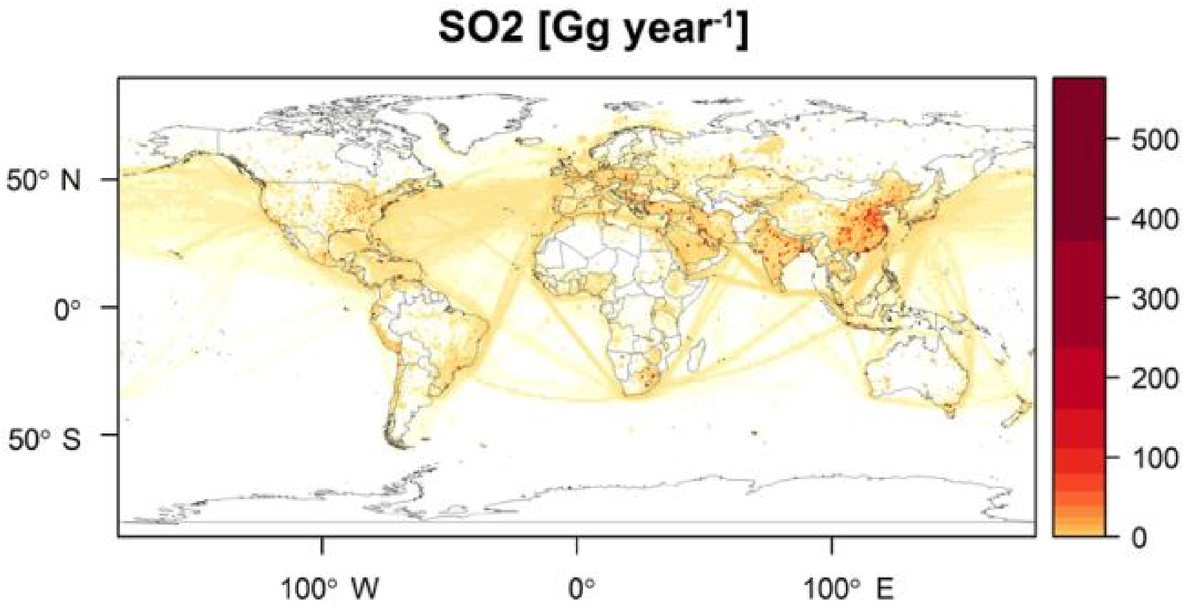
CEDS Point Source Gridding

Noah Prime, Hamza Ahsan, Steven J. Smith

Joint Global Change Research Institute
Pacific Northwest National Laboratory



April 4, 2023

Introduction

The Community Emissions Data System (CEDS) is a data-driven, open-source project that produces annual emission estimates for research and analysis. The data system produces historical emission estimates (currently 1750-2019) by country, sector, and fuel, and is readily updated every year.

CEDS has the capability to generate gridded output. Using spatial proxy, the CEDS global inventory is distributed over a global grid at 0.5° resolution. Seasonal cycle data is included to provide a monthly resolution.

For most CEDS gridding sectors, the spatial proxy used to distribute the CEDS inventory are gridded emissions estimates from the Emissions Database for Global Atmospheric Research (EDGAR) [Crippa et al. 2019]. This includes the energy sector, industrial process and product use, industrial combustion, and others. The EDGAR grids are pre-processed off-line and downloaded as CEDS input before performing a run.

The distribution of emissions is an additional source of uncertainty that CEDS had not yet addressed. Therefore, these distributions provided an obvious area for improvement that could lead to more accurate final gridded emissions.

In order to improve upon the spatial proxy from EDGAR, we introduce additional point source specific data. Point sources may be power plants, metal smelters, or anything that has large emissions from a single source. These sources may make up a large proportion of emissions from entire gridding sectors, making it high priority to model these as accurately as possible. Doing so could:

- Improve the spatial position of large sources,
- Provide more accurate emission estimates at the sources,
- Generally improve the spatial distribution of emissions in a given country/region,
    - note - each country in CEDS is designated by a unique iso code (see here)
- Allow for estimation of co-emitted species at a given source, and
- Enable down-scaling routines to provide finer resolution grids.

This document provides an overview of the methods used in incorporating point source data into the CEDS gridding system, and the resulting data product. This data product includes global grid estimates of all available CEDS emission species from 2000-2019, at 0.5° and 0.1° resolution.

1 Data Product

Emissions estimates are provided for the years 2000 to 2019 at both 0.5° and 0.1° resolution. Estimates are given as flux in units of kilograms of emissions per second per meter squared ($kg/s/m^2$). Each year is associated with a single NetCDF file which contains grids for each CEDS gridding sector. The sectors included are:

- Agriculture (`AGR`)
- Energy (`ENE`)
- Industrial (`IND`)
- Residential, Commercial, Other (`RCO`)
- International Shipping (`SHP`)
- Solvents Production and Application (`SLV`)
- Transportation (`TRA`)
- Waste (`WST`)

Sectors are split into monthly grids in each annual file using a 365-day calendar, using -180–180 longitude and -90–90 latitude bounds.

Data can be found on Zenodo in two separate data releases:

- primary CEDS emission species - 10.5281/zenodo.7526345
- speciated-VOC emission species - 10.5281/zenodo.7526534.

1.1 Initial Comparison

The result this work are new gridded estimates for CEDS that incorporate satellite data from the Ozone Monitoring Instrument (OMI) at many large point sources. Figure 1 provides an example of what these grids look like.

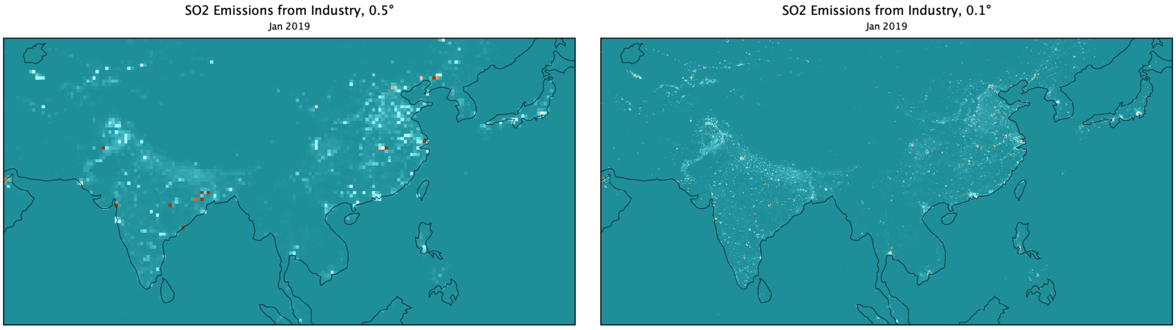


Figure 1: SO2 flux over a section of Asia at the 0.5° and 0.1° resolution.

The Energy and Industrial gridding sectors are the only CEDS sectors that will have been affected by this update, as these sectors contain all of the OMI point sources. There has also been updates to the CEDS sector definitions, with some sectors having been broken out since the latest release.

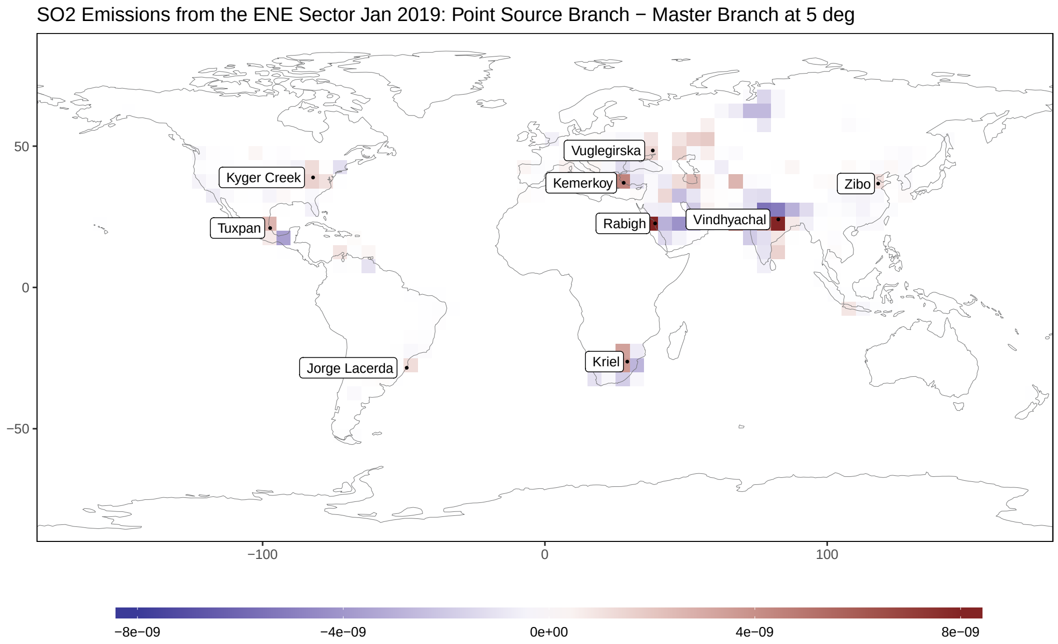


Figure 2: The change in SO2 flux grids resulting from the introduction of OMI point source data given for the month of January 2019 in the energy sector. Grids have been aggregated to 5° for viewing and some major power plants labeled.

For initial comparison of the gridded output to the previous release, we provide a plot of the difference in Figure 2. In this case we use the Energy sector to show the differences in SO2 emissions between the point source grids and the grids output using the old CEDS gridding system. With some of the major power plants labeled, we can see that this is where the largest changes occur. This is inline with what we would expect as we are making direct changes at these locations, and these locations are those with the most emissions.

We also looked at how this update changed the distribution of SO2 emissions in the US. It has been previously noted that CEDS produced grids biased towards Western US SO2 emissions [Yang et al. 2018]. In part this is because the default spatial allocation does not fully consider the differing regional sulfur content of coal in the US. While SO2 emissions still appear to be overestimated in the Western US, the amount of bias is reduced in this version. In Figure 3 we show how CEDS compares to inventories from the EPA [*Air Pollutant Emissions Trends Data - State Tier 1 CAPS Trends* 2022] and a 1985 study from Gschwandtner et al. [Gschwandtner et al. 1985], with and without the point source update.

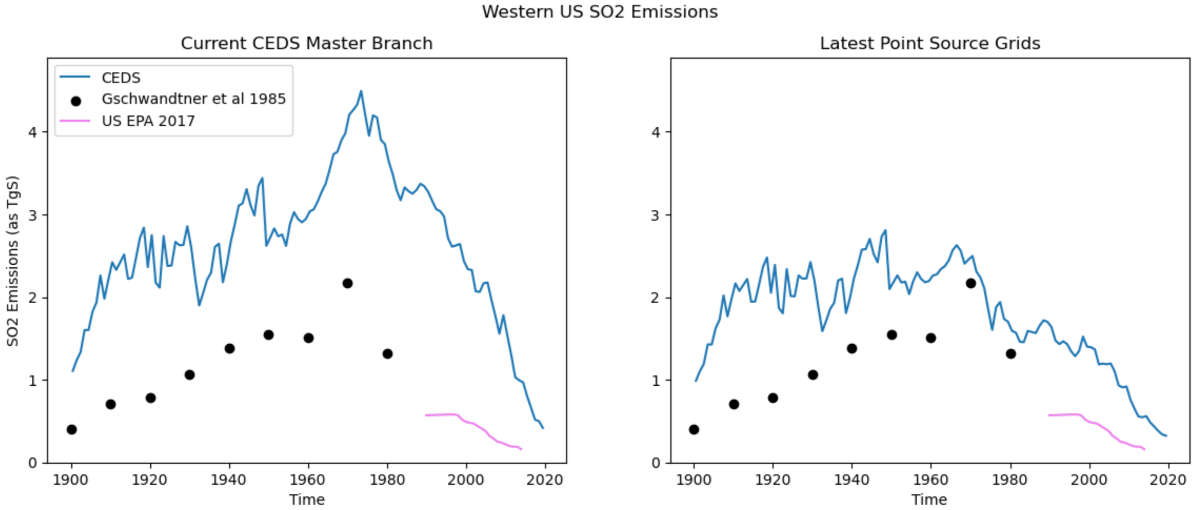


Figure 3: CEDS SO2 emissions in the western United States compared to estimates from Gschwandtner et al and US EPA, before (left) and after (right) point sources were introduced. Units have been converted to teragrams of sulfur to match the units of the these two additional datasets. Note that the geographic regions being compared are not fully compatible, with the Gschwandtner and EPA data based on state boundaries, and the CEDS data is based on data extraction using a spatial overlay.

2 Methods

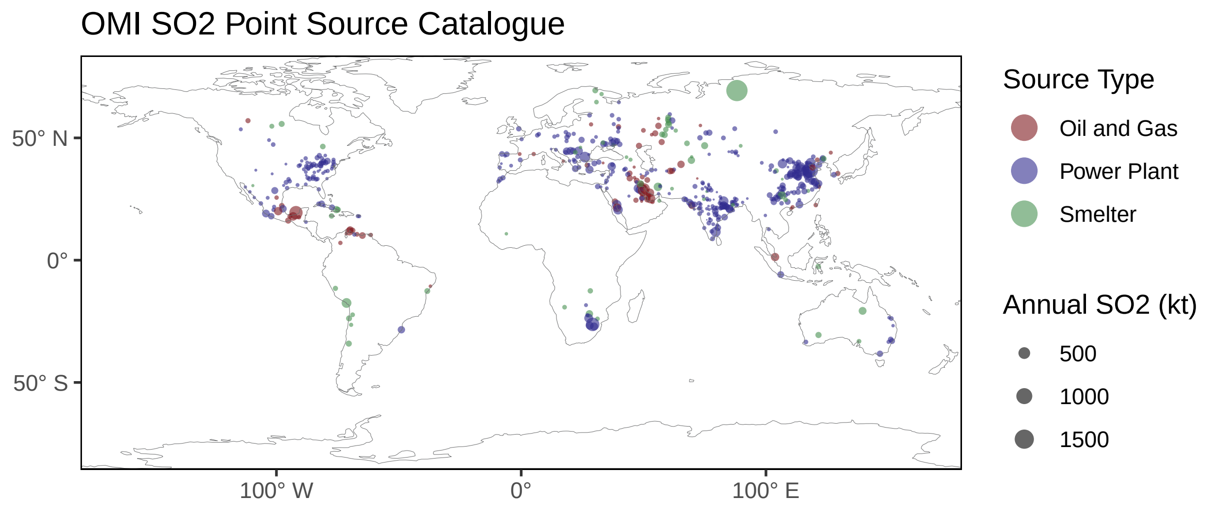
There are five broad steps in the generation of the final grids:

1. Processing point source data
2. Generating proxy grids
3. Distributing CEDS inventory
4. Downscaling (only for the 0.1° grids)
5. Adding point sources

Much of this functionality is not new to CEDS. As such, while each step is reviewed here, additional information on these processes can be found on the CEDS wiki.

2.1 Processing Point Sources

2.1.1 OMI



The point sources included in this data release come from NASA's Ozone Monitoring Instrument (OMI) [*Ozone Monitoring Instrument* 2021]. The OMI catalog contains **589** time series of SO2 emissions from **2005-2019** at localized points around the globe. Each time series is associated with a source type, either 'Power Plant', 'Smelter', 'Oil and Gas', or 'Volcano', the last of which is not included in CEDS. Additionally, each source is named, with latitude/longitude and country data.

While OMI sources are from localized satellite imaging, the resolution is not so fine as to guarantee that each OMI source is associated with just one power plant, smelter, etc. This is often not the case, and each OMI source must be disaggregated into its component point sources.

Power Plants

For power plants in OMI, CEDS uses WRI's Global Power Plant Database (GPPDB) [*Global Power Plant Databases* 2018] to determine which power plants are within a given radius of the OMI source. This radius has been set to 40km as default as suggested by Narayan, et al [Narayan et al. 2023]. The GPPDB also provides electricity capacity, fuel type and commissioning year for each plant. We attribute OMI emissions to the plants proportionally by emissions (calculated using plant capacities, fuel types, and CEDS emission factors) within the radius of the OMI source. When plant capacities or build dates are missing, values are added manually using supplementary data, often the Global Energy Observatory [*Global Energy Observatory* 2018].

The total annual emissions of a set of power plants associated with a given fuel, iso, and year are calculated using Equation 1,

$$T_{ijk} = \sum_{n} \text{Capacity}(P_n; i, j, k) \tag{1}$$

Where:

$P_1, ..., P_n$ is the set of power plants within 40km of an OMI source,

i is a CEDS supported fuel,

j is a year from 2005 to 2019, and

k is a CEDS supported iso.

The total OMI source emissions associated with each fuel, iso, and year is given by Equation 2,

$$\rho_{ijk} = \text{OMI}_j \cdot \frac{EF_{ijk} \cdot T_{ijk}}{\sum\limits_{i,k}(EF_{ijk} \cdot T_{ijk})} \tag{2}$$

Where:

EF_{ijk} is the corresponding CEDS emission factor,

OMI_j is the total emissions of an OMI source in year j,

i is a CEDS supported fuel,

j is a year from 2005 to 2019, and

k is a CEDS supported iso.

Finally, the emissions associated with each plant, P_n, is given by Equation 3 for the years 2005-2019.

$$\text{Emissions}(P_n; i, j, k) = \rho_{ijk} \cdot \frac{\text{Capacity}(P_n; i, j, k)}{\sum\limits_{n} \text{Capacity}(P_n; i, j, k)} \tag{3}$$

Where:

$P_1, ..., P_n$ is the set of power plants within 40km of an OMI source,

i is a CEDS supported fuel,

j is a year from 2005 to 2019, and

k is a CEDS supported iso.

For the years before OMI data is available, emissions are extended back using the ratio of emissions factors in consecutive years according to Equation 4. Here the year j goes from the commissioning year of plant, P_n, to 2004. Emissions are set to 0 for all years before the commissioning year, back to 1750.

$$\text{Emissions}(P_n; i, j, k) = \frac{EF_{ijk}}{EF_{i(j+1)k}} \text{Emissions}(P_n; i, j+1, k) \tag{4}$$

Smelters

For the OMI sources marked as smelters, WRI's GPPDB is used to identify power plants within the footprint of the OMI source. Here, the combined emissions from nearby power plant sources are estimated and then subtracted from the OMI emissions, attributing any remaining emissions to the smelter.

The ratio of the electric capacity of the nearby power plants to the entire electric capacity of the iso is estimated according to the GPPDB data. Using this ratio, emissions are proportionally allocated from the CEDS inventory to the nearby power plants. This process is performed by iso and fuel.

The CEDS inventory is also used to extend the smelter emissions. Commission dates for smelters, unlike power plants, is often unavailable, requiring a different approach for extension. Here, smelter emissions are extended back from OMI years in proportion to the CEDS inventory except where data has been determined for specific smelters.

Oil and Gas

The 'Oil and Gas' OMI sources are estimated similarly to smelter sources. The emissions from nearby power plants are estimated in the same way, removing the result from the OMI source. Like smelter emissions, Oil and Gas emissions are also extended in proportion to the CEDS inventory. Here, remaining emissions are attributed to the CEDS sector **1B2_Fugitive-petr**, CEDS intermediate gridding sector **FLR**, and associated CEDS gridding sector **ENE**.

2.1.2 Co-emitted Species

While OMI only provides estimates of SO2 emissions, we know that these sources can emit multiple species. We, therefore need to estimate emissions of any co-emitted species which will also be incorporated as point sources in the CEDS gridded data. This keeps consistency between gridded estimates, maintaining the location and relative size of emissions between species for each of the OMI point sources.

Two methods for estimating co-emitted species have been implemented so far.

1. Using emission factors. In this case we use the ratio of emission factors for SO2 and the co-emitted species, both from CEDS data for that iso and sector, to scale the SO2 time series. We of course use the emission factors respective to the given year, sector, iso, and species as before. Below, 'ces' is some co-emitted species; CO2 for example.
$$\text{Emissions}_{ces}(S_n, i, j, k, \ell) = \text{Emissions}_{SO2}(S_n, i, j, k, \ell) \cdot \frac{EF_{ijk}(SO2)}{EF_{ijk}(ces)}.$$
2. Using power plant capacity data from WRI. Since the GPPDB is a fairly complete database for electricity production, we can estimate the emissions of a given species using the relative capacity of the point source in the GPPDB. This option is only available when the OMI source exists in the GPPDB.

Here, we simply scale the total emissions in the source's iso (as estimated by CEDS) by the proportion of electric capacity of the source relative to the total electric capacity of the iso (according to the GPPDB). The GPPDB also provides the commissioning year for most power plants, allowing us to set the capacity to zero before the commissioning date.

2.1.3 Scaling

For consistency, we ensure that the total gridded data matches the CEDS inventory. It is possible that the point source data from OMI exceeds the CEDS inventory in aggregate for a particular iso and sector. When this occurs, point source emissions are scaled down proportionally to match the CEDS inventory. Point sources are aggregated and reconciled with the CEDS inventory by iso, CEDS sector, year and fuel. In cases where the point sources need to be scaled, the CEDS inventory is distributed to the point sources in proportion to their estimated emissions as shown in Equation 5.

Note that a diagnostic report for scaling is generated. Overall, in this dataset, the amount of scaling that occurs is small. The total change in all point sources before and after scaling is just 0.4% of CEDS SO2 emissions for the the sectors and country/regions present in OMI. In any instances of larger scaling, the underlying data is examined to evaluate if CEDS emissions in these cases may be underestimated, in which case, improved data can be incorporated.

$$\text{Emissions}_{final}(S_n; i, j, k, \ell) = I_{ijk\ell} \cdot \frac{\text{Emissions}_{initial}(S_n; i, j, k, \ell)}{\sum\limits_{n} \text{Emissions}_{initial}(S_n; i, j, k, \ell)} \tag{5}$$

Where:

S_n is a single point source,

i is a CEDS supported fuel,

j is a year from 2005 to 2019,

k is a CEDS supported iso,

ℓ is CEDS sector, and

$I_{ijk\ell}$ is the corresponding CEDS inventory.

The effect of the scaling process in the CEDS sector '1A1a_Electricity-public' is shown in in Figure 4. Note that in one case, for cze, there is significant scaling in the recent years. Because SO2 emissions data in this case is from the EMEP inventory, and the OMI data has larger errors for these smaller emission magnitudes at this point, we take the inventory data as the most accurate indication of total country emission magnitudes. We, therefore, make no changes the underlying CEDS estimate for this country.

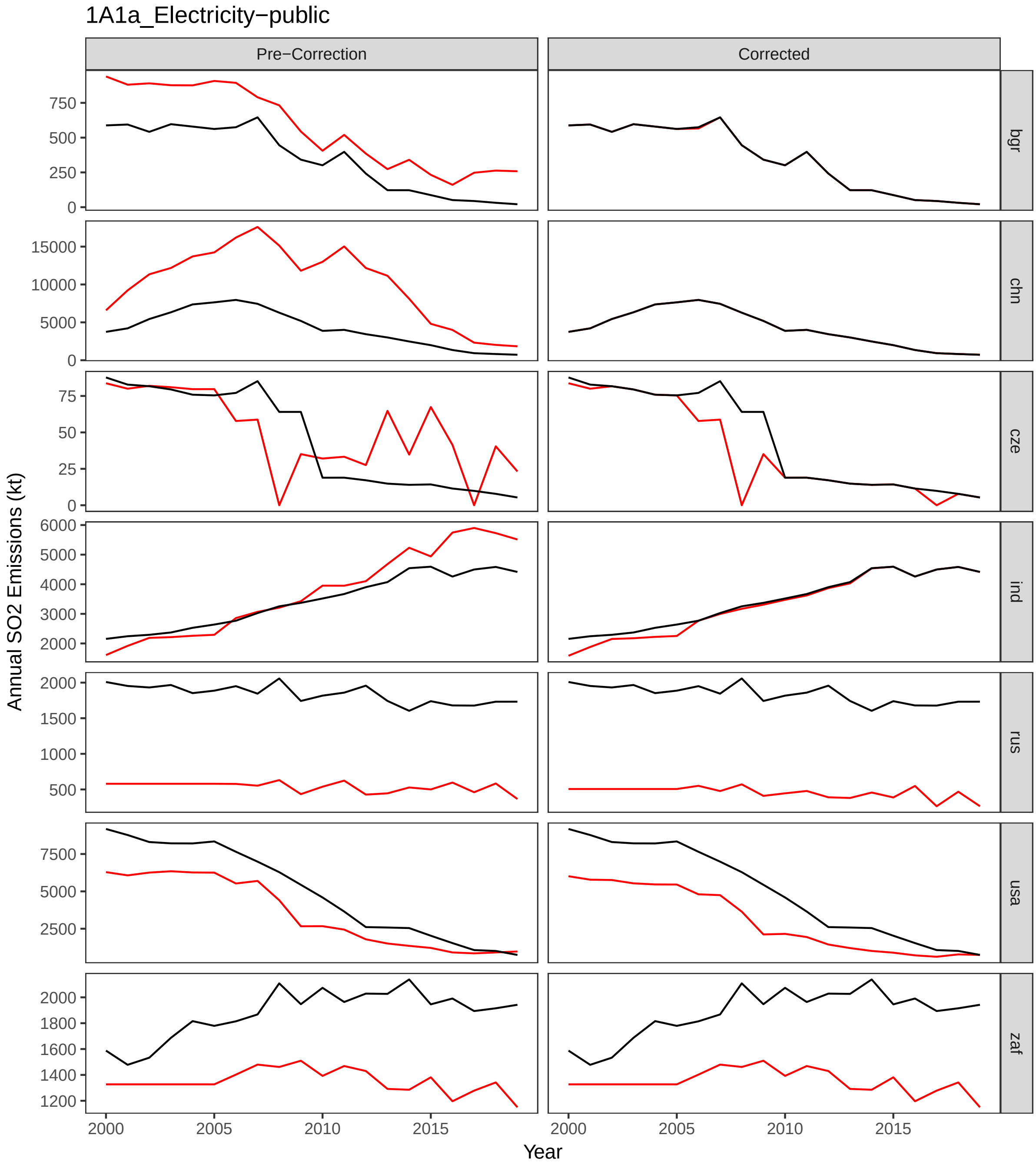


Figure 4: Comparison of emissions as suggested by the point source data, and CEDS inventory, in the electricity sector for select isos. In red is the aggregate of the point source time series' in a given iso, in the 1A1a_Electricity-public CEDS sector. In black is the CEDS inventory for the 1A1a_Electricity-public CEDS sector.

2.2 Proxy Generation

The addition of time series data specific to a point source enables those emissions to be added directly to a grid at any resolution. This requires that grids be generated with the point source emissions removed, which requires both emission inventories without the point sources and spatial proxy without the point sources.

The base spatial proxies used in CEDS primarily come from EDGAR v5.0 [Crippa et al. 2019], and ECLIPSE V5a for flaring emissions [Klimont et al. 2017]. When distributing the CEDS inventory without the point sources, all emissions go to cells not associated with a point source. This is equivalent to setting the spatial proxy to zero at the point source locations. First, point sources are identified in the proxy grids by year and sector. Cells with the largest emissions in a given sector and year are assumed to indicate point sources. The number of point sources assumed to be in each sector and year varies and is informed by the distribution of emissions as detailed in Figure 5.

Next, point sources identified in the proxy are matched to input point source time series. Sources are considered a match if they are within 0.5° of each other. When a point source is identified, it's proxy grid cell and all cells within 0.1° are set to zero, as there is a tendency for emissions to 'leak' into surrounding cells in the proxy data. If a match is not found, the cell at the coordinates of the input point source is zeroed. This is done at the resolution of the input proxy files, which is 0.1° for EDGAR, and 0.5° for ECLIPSE.

Finally, spatial proxies are aggregated to CEDS gridding resolution of 0.5°, and grids reformatted to have consistent latitude and longitude bounds.

2.3 Spatially Distributing CEDS Inventory

Grids are generated by distributing the CEDS inventory over the spatial proxy created in the previous step. As point sources are added directly to the final grids, point sources are first removed from the inventory by subtracting the final point source time series emissions from the total CEDS inventory.

Next the CEDS time series inventory (by iso, sector, fuel) with the point sources removed is distributed over the spatial proxy with the point sources removed in the same manner as is done in the old CEDS routine. This routine is detailed in the CEDS wiki page, Hoesly et al. 2018, and Feng et al. 2020.

2.4 Downscaling

The default CEDS grid is output at 0.5° resolution. Due to interest in grids at finer resolution, we are also supplying grids at 0.1° resolution produced through a downscaling routine.

The downscaling routine again uses spatial proxy files as reference, this time at 0.1° resolution. These are generated just as before to remove point sources from the grid, as those emissions

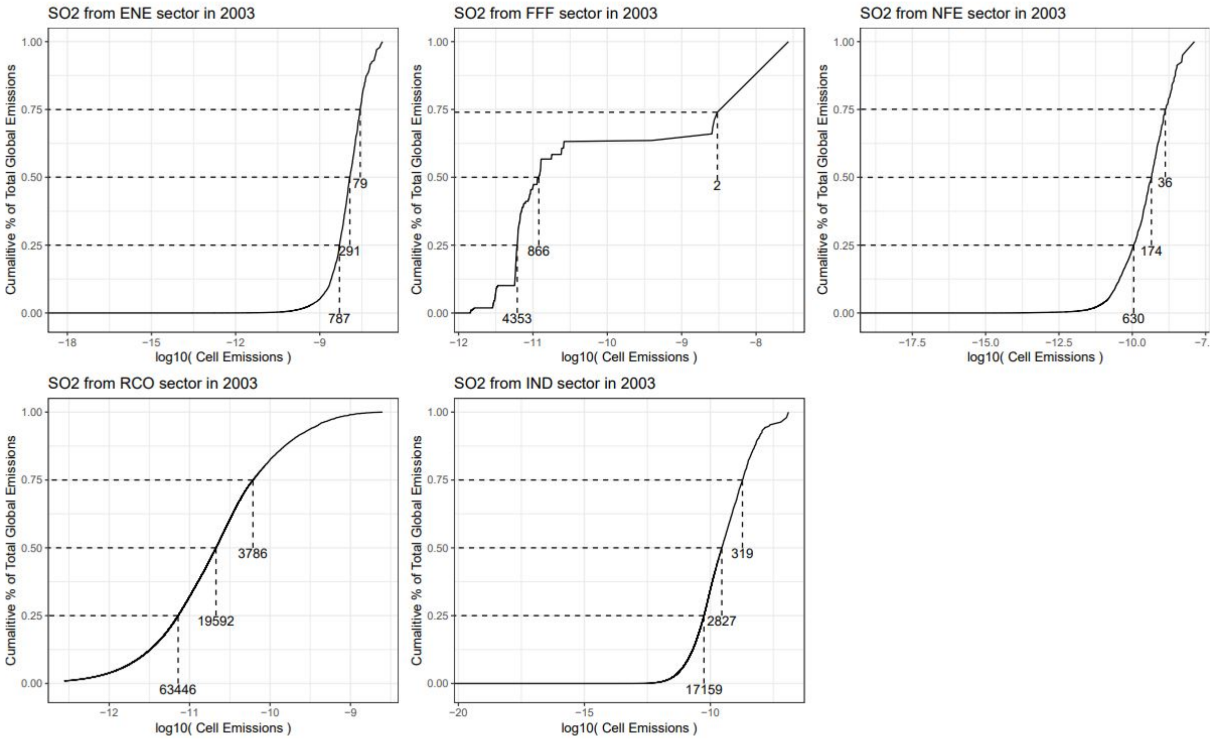


Figure 5: Plots of the cumulative percentage of global SO2 emissions in 2003 for selected EDGAR sectors, vs the emissions of a given cell. Markings along the plot indicate how many cells have equal or larger emissions at that point. Markings are provided at 25, 50, and 75 percent of total global emissions. One can interpret, for example, that 79 cells make up 25% of global SO2 emissions in the ENE sector in 2003. These large cells are assumed to indicate a point source.

are dealt with separately. For each CEDS gridding sector, corresponding proxy files are used to distribute coarse cells into fine cells. For a given fine cell i at 0.1° resolution, inside a coarse cell j at 0.5° resolution, the emission estimate in the downscaled CEDS grid, CEDS_{ij}, is allocated from the coarse CEDS grid, CEDS_j, according to the proportion of proxy cells, proxy_{ij}, defined in Equation 6.

$$\text{CEDS}_{ij} = \text{CEDS}_j \cdot \frac{\text{proxy}_{ij}}{\sum\limits_{i} \text{proxy}_{ij}} \tag{6}$$

Downscaling is applied to the 0.5° CEDS grids with the point sources removed. Point sources are then added at the fine resolution of 0.1°.

However, not all proxy files are available at 0.1° resolution, which causes the downscaling routine to miss emissions. This occurs, for example, in the FLR sector, as data from ECLIPSE is only available at 0.5° resolution.

To manage this problem, downscaling is done in three steps. First, emissions are downscaled using the EDGAR proxies as described. Second, any missed emissions are downscaled using population proxy data from HYDE [Klein Goldewijk et al. 2011]. Finally, any remaining emissions, uniformly distributed. This does not add any information beyond the 0.5° resolution though ensures consistency in global total emissions throughout the data products.

2.5 Adding Point Sources

At either data resolution, point sources are added in the final step. At this point in the gridding routine, grids only include emissions from the CEDS inventory that are not attributed to any point sources for which we have time series data. Each final point source time series is added to the corresponding grid cells. Each point source is associated with latitude and longitude coordinates, which are converted to cell indices for the given resolution. CEDS adopts EDGAR's convention of defining grid cells using the coordinates of the center of the cell. Each point source is also associated with a given gridding sector and iso. For each year, each point source in the given sector and iso is converted from annual emissions to flux and added to the corresponding cell.

References

Air Pollutant Emissions Trends Data - State Tier 1 CAPS Trends (2022). `https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data`. United States Environmental Protection Agency. URL: `%5Curl%7Bhttps://www.epa.gov/sites/default/files/2021-03/state_tier1_caps.xlsx%7D`.

Crippa, Monica et al. (2019). *EDGAR V5.0 Global Air Pollutant Emissions*. European Commission, Joint Research Centre (JRC). URL: `http://data.europa.eu/89h/377801af-b094-4943-8fdc-f79a7c0c2d19`.

Feng, L. et al. (2020). "The generation of gridded emissions data for CMIP6". In: *Geoscientific Model Development* 13.2, pp. 461–482. DOI: `10.5194/gmd-13-461-2020`. URL: `https://gmd.copernicus.org/articles/13/461/2020/`.

Fioletov, V. E. et al. (2016). "A global catalogue of large SO_2 sources and emissions derived from the Ozone Monitoring Instrument". In: *Atmospheric Chemistry and Physics* 16.18, pp. 11497–11519. DOI: `10.5194/acp-16-11497-2016`. URL: `https://acp.copernicus.org/articles/16/11497/2016/`.

Global Energy Observatory (Dec. 2018). `http://GlobalEnergyObservatory.org/`. Accessed July. 2022.

Global Power Plant Databases (2018). `http://resourcewatch.org/,%20https://earthengine.google.com/`. Accessed July. 2022.

Gschwandtner, G et al. (Apr. 1985). "Historic emissions of sulfur and nitrogen oxides in the United States from 1900 to 1980. Volume 1. Results. Final report, September 1982-December 1984". In: URL: `https://www.osti.gov/biblio/5486012`.

Hoesly, R. M. et al. (2018). "Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS)". In: *Geoscientific Model Development* 11.1, pp. 369–408. DOI: `10.5194/gmd-11-369-2018`. URL: `https://gmd.copernicus.org/articles/11/369/2018/`.

Klein Goldewijk, Kees et al. (2011). "The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years". In: *Global Ecology and Biogeography* 20.1, pp. 73–86. DOI: `https://doi.org/10.1111/j.1466-8238.2010.00587.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1466-8238.2010.00587.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1466-8238.2010.00587.x`.

Klimont, Z. et al. (2017). "Global anthropogenic emissions of particulate matter including black carbon". In: *Atmospheric Chemistry and Physics* 17.14, pp. 8681–8723. DOI: `10.5194/acp-17-8681-2017`. URL: `https://acp.copernicus.org/articles/17/8681/2017/`.

Li, C. et al. (2017). "New-generation NASA Aura Ozone Monitoring Instrument (OMI) volcanic SO_2 dataset: algorithm description, initial results, and continuation with the Suomi-NPP Ozone Mapping and Profiler Suite (OMPS)". In: *Atmospheric Measurement Techniques* 10.2, pp. 445–458. DOI: `10.5194/amt-10-445-2017`. URL: `https://amt.copernicus.org/articles/10/445/2017/`.

Narayan, Kanishka et al. (2023). "Evaluation of uncertainties in the anthropogenic SO2 emissions in the USA from NASA's OMI point source catalog". In: In Preparation.

Ozone Monitoring Instrument (2021). `https://so2.gsfc.nasa.gov/`. Accessed June. 2022.

Prime, N. et al. (Aug. 2022). *Gridded SO2 Emissions With Explicit Point Sources from the Community Emissions Data System, 2000-2019*. Zenodo. DOI: `(0.5°)10.5281/zenodo.6949566,(0.1°)10.5281/zenodo.6964915`.

Yang, Yang et al. (2018). "Source Apportionments of Aerosols and Their Direct Radiative Forcing and Long-Term Trends Over Continental United States". In: *Earth's Future* 6.6, pp. 793–808. DOI: `https://doi.org/10.1029/2018EF000859`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF000859`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EF000859`.