For additional information, contact Prof. Peter Adams ([peteradams@cmu.edu](mailto:peteradams@cmu.edu)).

**Overview**

This dataset contains source-resolved and speciated outdoor concentrations of PM2.5 across the contiguous U.S. for 2001 and 2010 (annually-averaged). We use geographically weighted regression to predict and then correct for the bias in PM2.5 concentrations (Hernandez et al., 2021) originally simulated from an air quality model (Skyllakou et al., 2021).

Observations from speciated ground-level monitors (IMPROVE and CSN) are used to correct simulated ammonium (NH4+), nitrate (NO3-), sulfate (SO42-) elemental carbon (EC) and organic aerosol (OA). The primary and secondary fractions of OA, as predicted by the air quality model, are applied to corrected OA to estimate primary organic aerosol (OA) and secondary organic aerosol (SOA). The air quality model also predicts concentrations of sodium, chloride, and mineral dust, but these are excluded from this dataset due to large uncertainty in the emissions inventory and because speciated monitors for these species are not readily available. Note that total PM2.5 values provided in this dataset are the sum of NH4+, NO3-, SO42-, EC and OA. To account for concentrations of sodium, chloride, and mineral dust, we recommend uniformly adding 2 and 1.5 μg/m3 to the total PM2.5 concentrations for 2001 and 2010, respectively.

The underlying air quality model simulations use a source apportionment module to attribute PM2.5 concentrations to six source categories: electricity-generating units (EGUs; includes all emissions in EPA’s Integrated Planning Model), non-EGU (all other industrial point sources), on-road mobile, off-road mobile, biogenic (i.e., vegetation), and “other” which includes on-road emissions from Canada and Mexico plus all other emissions (e.g., biomass burning, cooking, agriculture, etc.). PM2.5 concentrations attributed to the model’s boundary and initial conditions are also tracked as separate categories, but not made explicitly available in this dataset. The original source fractions, as predicted by the air quality model, are applied to corrected estimates of all PM2.5 chemical components. Note that total source values provided in this dataset are the sum of EGU, nonEGU, on-road, off-road, biogenic, other, and contributions from initial/boundary conditions.

The PM2.5 estimates are originally simulated in grids with a 36 km horizontal resolution. However, bias corrections are performed at census tract centroids for 2010 census geographies. Data is made available in its native resolution (tract level), as well as population-weighted averages at the county, state and national level. However, note that some sharp concentration gradients may exist at the tract level estimates. This is a consequence of the coarser resolution used in the original air quality model simulations, and impacts tracts near the borders of the air quality model’s grid cells. Therefore, we recommend the population-weighted county level estimates for most users.

**File** **Format**

|  |  |
| --- | --- |
| Parameter | Description |
| year | Prediction year for a given data point (2001 or 2010) |
| pollutant | Pollutant for a given data point (NH4+, NO3-, SO42-, EC, OA, POA, SOA & total PM2.5) |
| fips | Unique code for geographic unit with the following structure: SSCCCTTTTTT  (S = 1–2-digit state code; C = 3-digit county code; T = 6-digit tract code) |
| source | Source for a given data point (EGU, nonEGU, on-road, off-road, biogenic, other & total) |
| concentration | Population-weighted concentration (native resolution at the census tract level) with units in μg/m3. |

**Suggested Citations**

Hernandez, C., Skyllakou, K., Garcia Rivera, P., Dinkelacker, B., Marshall, J., Pope, A., Robinson, A., Pandis, S., & Adams, P. (2021) Bias corrections for speciated and source-resolved PM2.5 chemical transport model simulations using a geographically weighted regression. ChemRxiv. Cambridge: Cambridge Open Engage; This content is a preprint and has not been peer-reviewed. <https://doi.org/10.26434/chemrxiv-2021-h71p5>.

Skyllakou, K., Rivera, P. G., Dinkelacker, B., Karnezi, E., Kioutsioukis, I., Hernandez, C., Adams, P. J., & Pandis, S. N. (2021). Changes in PM2.5 concentrations and their sources in the US from 1990 to 2010. Atmospheric Chemistry and Physics, 21(22), 17115–17132. <https://doi.org/10.5194/acp-21-17115-2021>