

Replication Dataset for:

Air inequality: Global divergence in urban fine particulate matter trends

Joshua S. Apte^{1,2*}, Sarah Seraj³, Sarah E. Chambliss³, Melanie Hammer^{4,5}, Veronica A. Southerland⁶, Susan C. Anenberg⁶, Aaron van Donkelaar⁵, Michael Brauer^{7,8}, Randall V. Martin^{4,5}

¹Department of Civil and Environmental Engineering, University of California, Berkeley, USA

²School of Public Health, University of California, Berkeley, USA

³ Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX USA

⁴Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, USA

⁵Department of Physics, Dalhousie University, Halifax, Canada

⁶School of Public Health, George Washington University, Washington, DC, USA

⁷School of Population and Public Health, The University of British Columbia, Vancouver BC, Canada

⁸Institute for Health Metrics and Evaluation, University of Washington, Seattle WA, USA

* Correspondence to apte@berkeley.edu.

Main manuscript is available in preprint form at *doi*:10.26434/chemrxiv.14671908

This dataset contains:

- **Input data**
- **Custom MATLAB code from this analysis**
- **Links to supporting MATLAB libraries**
- **Output data**

Version: 24 May 2021

Input Data Sources

The key input data sources for this analysis are provided below. We provide here our imported MATLAB working files for these data here as a courtesy to those who may wish to reproduce our results from the same data we used. However, all rights to these underlying data sources belong to their respective owners, and we discourage use of these MATLAB files for any purpose other than validating our results.

1. **Gridded PM_{2.5} data** available as a MATLAB file: `PM.mat`, 8.33 GB with supporting latitude and longitude files `pmLat.mat` and `pmLon.mat`.

Format: 0.01° × 0.01° global grid, GWR-adjusted PM_{2.5} at 35% RH for entire global domain, 21 years (1998-2018), saved as MATLAB georeferenced raster

Dimensions: latitude × longitude × year

MATLAB Georeferencing Matrix:

```
R =  
      0      0.0100  
    0.0100      0  
 -179.9950 -54.9950
```

Citation: Hammer, M. S. *et al.* Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ. Sci. Technol.* **54**, 7879-7890 (2020).

Note: This file is similar to the canonical underlying V4.GL.03 NetCDF datasets available from the Atmospheric Composition and Analysis Group here: <https://sites.wustl.edu/acag/datasets/surface-pm2-5/>

2. **City-level demographic data**

File: `city_input_data.mat`

Format Dimensions: MATLAB file with 11 variables with 4231 entries for individual cities.

Variable	Description
<code>city_names</code>	Name of the city
<code>city_lat</code>	Latitude of approx. center of city
<code>city_lon</code>	Longitude of approx. center of city
<code>country_code</code>	ISO3 country code of city location
<code>country_names</code>	Country name
<code>GBD_region_city</code>	Regional grouping for GBD2019
<code>pop2000</code>	Urban agglomeration population in year 2000
<code>pop2010</code>	Urban agglomeration population in year 2010
<code>PopCat</code>	Membership in one of four equal-population groupings of cities based on ranking of year-2010 city size (“1” is smallest < 400k, “4” is largest > 5.6 M).
<code>country_GDPcat</code>	Country income grouping, World Bank
<code>city_nationalGDP_2018</code>	Country real GDP per capita, 2018, World Bank

The source of the underlying dataset of 4231 cities, their locations, names, and populations is:

Citation: Angel, S. *et al.* Atlas of urban expansion—2016 edition. *Areas and densities*. Cambridge, MA: NYU Urban Expansion Program at New York University, UN-Habitat, and the Lincoln Institute of Land Policy (2016). Last accessed: May 5, 2021.
<http://www.atlasofurbanexpansion.org/>

3. WHO global dataset of ground-based PM_{2.5} monitoring

File: WHO_monitor_data_201909.mat

Format Dimensions: MATLAB file based on a September 2019 version of the WHO Global Ambient Air Quality Database.

Citation: World Health Organization. *WHO Global Ambient Air Quality Database (Update 2020)*. (WHO, 2020).

4. UN World Urbanization Prospects 2018 country-level urban population counts

File: WUP2018F03UrbanPopulation.mat

Format Dimensions: MATLAB file based on UN's Urban Population Prospects

Data: Country-level estimates of urban population in 5-year intervals, indexed by ISO3 code

Citation: United Nations; Department of Economic and Social Affairs; Population Division. *World Urbanization Prospects: The 2018 Revision*. (2018).

https://www.who.int/phe/health_topics/outdoorair/databases/cities/en/

MATLAB Code

Custom code

`compute_city_statistics.m`

This routine iteratively computes city-level estimates of PM_{2.5} based on gridded data and saves results out as an intermediate dataset named `city_extracted_PM.mat`. This routine calls `extract_gridded_city_data` and `extract_grid_annular` to do so.

`extract_gridded_city_data.m`

For an inputted lat-long coordinate and inputted gridded dataset, find all pixels within a given distance of this centroid, and compute the average value of those pixels. Used to compute city mean PM_{2.5} concentration.

`extract_grid_annular.m`

For an input lat-long coordinate and inputted gridded dataset over a given set of search radii, compute average values of all pixels contained in the concentric annular regions that make up the full set of search radii. Used to compute the isotropic radial concentration distance-decay function of PM_{2.5} for each world city.

`analyze_city_data.m`

After executing the above code, this routine computes concentration time trends, compiles an exportable city-level dataset, and produces the core figures from the manuscript (Fig 1-3 and Extended Data Figures 1-5).

`trendcolors.mat`

This dataset contains the color scheme used for visualizing concentration trends.

Code dependencies

The following MATLAB libraries are required to execute this code. Custom code, where used, is referenced to MATLAB Central.

Mapping toolbox

Many of the geospatial operations in this codebase require the MATLAB mapping toolbox.

`coast.mat`

This polygonal map of the world's coastlines (variables: lat, long) is included in a default installation of MATLAB.

MOVING

Aslak Grinsted (2021). Moving averages / Moving median (<https://www.mathworks.com/matlabcentral/fileexchange/8251-moving-averages-moving-median-etc>), MATLAB Central File Exchange

BREWERMAT

Stephen Cobeldick (2021). ColorBrewer: Attractive and Distinctive Colormaps (<https://github.com/DrosteEffect/BrewerMap>), GitHub.
<https://www.mathworks.com/matlabcentral/fileexchange/45208-colorbrewer-attractive-and-distinctive-colormaps>

MAGMA, INFERNO and other colormaps

Stephen Cobeldick (2021). Matplotlib Perceptually Uniform Colormaps (<https://www.mathworks.com/matlabcentral/fileexchange/62729-matplotlib-perceptually-uniform-colormaps>), MATLAB Central File Exchange.

Output Dataset

The Excel file `UrbanPMDataset_20210524.xlsx` contains city-level demographic information and our results (PM_{2.5} data, best-fit trends, and other analysis products) for all 4231 cities.