Replication Package for 'Designing Air Quality Monitoring Systems for Measurement and Environmental Policy Evaluation in Data-Scarce Environments'

Overview

The code in this replication package i) constructs the analysis file from data ii) generates plots in Python iii) generates tables in Stata

Requirements

- Conda
- Stata
- QGIS
- Google Earth Engine Code Editor[1]

Folder Structure

1. **Outputs:**
   - **Figures:** Contains all the figures generated for the paper.
   - **Tables:** Contains all the tables generated for the paper.
2. **Data:**
   - **Raw:** Contains the original data.
   - **Final:** Contains processed data ready for analysis.
3. **Code:**
   - **Javascript:** Contains Javascript scripts to collect raw MERRA-2 satellite data.

[1] `https://code.earthengine.google.com/`

- **Python:** Contains Python scripts to generate the intermediate data and the figures in the paper.
- **Stata:** Contains Stata do-files to generate the tables in the paper.

Running the Code

Google Earth Engine

1. Open the script `load_merra2_data.js` in the directory `./reproducibility_package/Code/JavaScript/gee/` using Notepad or a similar text editor. Copy the script.
2. Open the Google Earth Engine Code Editor at `https://code.earthengine.google.com/` and paste the copied script.
3. Run the script, and, in the 'Tasks' tab, execute each dataset individually (in Google Earth Engine, export tasks must be manually initiated).
4. Manually transfer the generated datasets to `./reproducibility_package/Data/Raw/Satellite/merra-2`.

Python

1. Open the Anaconda Prompt and navigate to `./reproducibility_package/Code/Python` on your device.
2. Create a new environment 'test_env' using the 'environment.yml' file using the command `conda env create -f environment.yml -n test_env`.
3. Run `conda activate test_env` to activate the new environment.
4. Open the notebook `master.ipynb` on the activated environment.
5. Run the master.ipynb. The .png plots will be generated in 'reproducibility_package\Outputs\Plots'

Stata

1. Open the Master do-file at 'reproducibility_package\Code\Stata\master.do'
2. Change the global 'reproducibility_package_dir' to where the reproducibility package is stored on your computer
3. Run the master.do. The .tex tables will be generated in 'reproducibility_package\Outputs\Tables'

Runtime Requirements

Approximate time needed to reproduce the analyses on a standard 2024 machine: 3 hours.

List of Tables and Programs

The provided code reproduces:

- [ ] All numbers provided in the text of the paper.
- [ ] All tables and figures in the paper.
- [x] Selected tables and figures in the paper, as explained below.

Raw Data Catalogue

Dataset Name	Source	Publishable?	Date obtained
Satellite\merra-2\ [1]	Google Earth Engine	No	May-2024
cgqa_hourly.csv	Center for the Control of Air Quality of Senegal [2]	No	Mar-2024
cgqa_daily.csv	Center for the Control of Air Quality of Senegal	No	Mar-2024
humidity_new.csv	Own Data Collection [3]	No	June-2024
LPAOSF.csv	Own Data Collection	No	June-2024
PM0_3_um_hourly.csv	Own Data Collection	No	June-2024
PM0_5_um_hourly.csv	Own Data Collection	No	June-2024
PM10_cf_1_hourly.csv	Own Data Collection	No	June-2024
PM1_0_cf_1_hourly.csv	Own Data Collection	No	June-2024
PM2_5_atm_hourly.csv	Own Data Collection	No	June-2024
PM2_5_cf_1_hourly.csv	Own Data Collection	No	June-2024
PM2_5_um_hourly.csv	Own Data Collection	No	June-2024
PM5_0_um_hourly.csv	Own Data Collection	No	June-2024
temp_hourly.csv	Own Data Collection	No	June-2024
weather_visualcrossing.csv	Visual Crossing Weather Data [4]	No	Aug-2024
Senegal_ADM3_SHP	The Humanitarian Data Exchange [5]	Yes	Aug-2024
Dakar_SHP	The Humanitarian Data Exchange	Yes	Aug-2024

[1] This folder contains .tif files conatining data from the MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) product for Dakar. To gather the data, see instructions above on running Google Earth Engine Code. From Google Earth Engine, the data is consists of 5 individual components: black carbon mass (BCSMASS), dust mass (DUSMASS25), organic carbon mass (OCSMASS), sea salt mass (SSSMASS25), and sulfate mass (SO4SMASS), from 2012-23.

[2] Provided courtesy of the Center for the Control of Air Quality of Senegal [CGQA].

[3] For data marked as sourced from 'Own Data Collection', data will be available on the World Bank Microdata library in the future.

[4] Data for Visual Crossing is downloaded using the API set to 'Dakar, Senegal' and can be accessed through the link *https://www.visualcrossing.com/weather/weather-data-services/dakar,%20senegal?v=api*. Data is gathered from 2000-2024 on an hourly level.

[5] The shapefiles of Senegal at administrative 3 level is downloaded from (`https://data.humdata.org/dataset/cod-ab-sen?`) **Senegal administrative level 0-3 boundaries (COD-AB) dataset** (`sen_admbnd_anat_20240520_AB_SHP.zip`).

Table 1: List of Raw Datasets

Figures

Figure Label	Description	Generated from
Figure 1	Location of regulatory and low-cost monitors in Dakar, Senegal	plots.ipynb
Figure 2	Daily $PM_{2.5}$ from regulatory-grade monitors in Dakar, 2012-24	plots.ipynb
Figure 3	Daily monitor metrics, 2020-23	plots.ipynb
Figure 4	Regulatory and Satellite $PM_{2.5}$ data comparison	plots.ipynb
Figure 5	$PM_{2.5}$ estimates for March 2020 at the hourly level	plots.ipynb
Figure 6	Low-cost and Regulatory $PM_{2.5}$ comparison at Pikine station, May-December 2023	plots.ipynb
Figure 7	Low-cost monitor and Satellite $PM_{2.5}$ values, 2020-21	plots.ipynb
Figure 8	One day example of $PM_{2.5}$ across Dakar measured using different data sources	QGIS
Figure 9	$PM_{2.5}$ Measures before and after Covid mobility restrictions in 2020	plots.ipynb
Figure 10	Comparison of air pollution in 2020 and 2023, Weekly	plots.ipynb
A1	Dashboard showing $PM_{2.5}$ values for low-cost monitors over time	plots.ipynb
A2	Satellite $PM_{2.5}$ estimates with and without dust, 2020-2021	plots.ipynb
A3	One day example of $PM_{2.5}$ across Dakar measured using Low-cost monitors applying daily calibrations	QGIS
A4	Comparison of air pollution in 2020 and 2023, Daily	plots.ipynb

Tables

Table Label	Description	File
Table 1	Calibration Coefficients comparing Regulatory and nearby Purple Air monitors (Hourly)	calibrations-hourly.tex
Table 2	Comparison of PM2.5 Measures	correlation.tex & rmse.tex
Table 3	Impact of Covid Mobility Restrictions on Air Pollution Measures (Hourly Averages, 2020 & 2023)	reg-hourly-2020-2023
Table 4	Comparison of Low-cost Correction Regression Coefficients (Hourly, 2020 & 2023)	coefficient-comparison-hourly.tex
Table 5	Comparison of Different Low-cost Calibration Coefficients (Hourly, 2020 & 2023)	local_calibration.tex
Table A1	Air Pollution products from Satellites	Manually created
Table A2	Calibration Coefficients comparing Regulatory and nearby Purple Air monitors (Daily)	calibrations-daily.tex
Table A3	Impact of Covid Mobility Restrictions on Air Pollution Measures	max-daily-2020+2023.tex reg-daily-2020+2023.tex reg-hourly-2017.tex reg-daily-2017.tex

Instructions for Reproducing Figures in QGIS

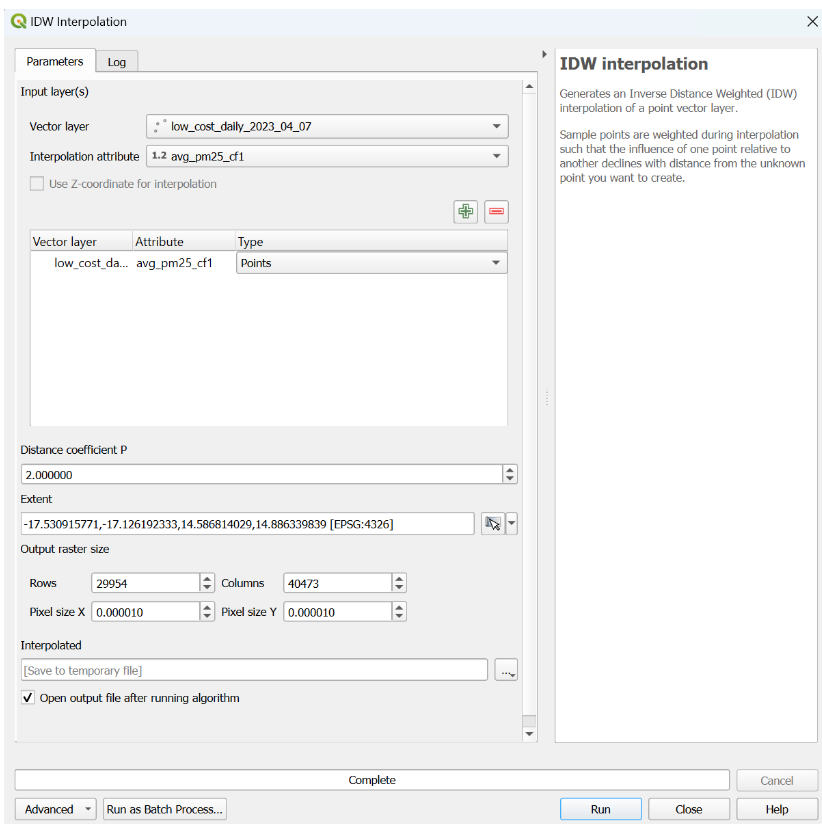
1. **Data Used**

   - **Shapefile of Dakar area:** the shapefiles in `Dakar_SHP` folder
     - the shapefiles for Dakar (`Dakar_SHP`) were extracted from the shapefiles of Senegal (`Senegal_ADM3_SHP`) at administrative 3 level by selecting only the regions where the administrative 3 name (`ADM3_FR`) is labeled as Dakar.
   - **Panel data for unadjusted PM2.5 and calibrated PM2.5:** `low_cost_daily_2023_04_07.csv` contains unadjusted Purple Air pm2.5, pm2.5 with local calibration, and pm2.5 with satellite calibration from 19 locations on April 7, 2023.
   - **Satellite data:** `satellite_daily_2023_04_07.csv` contains pm2.5 data from satellite on April 7, 2023.

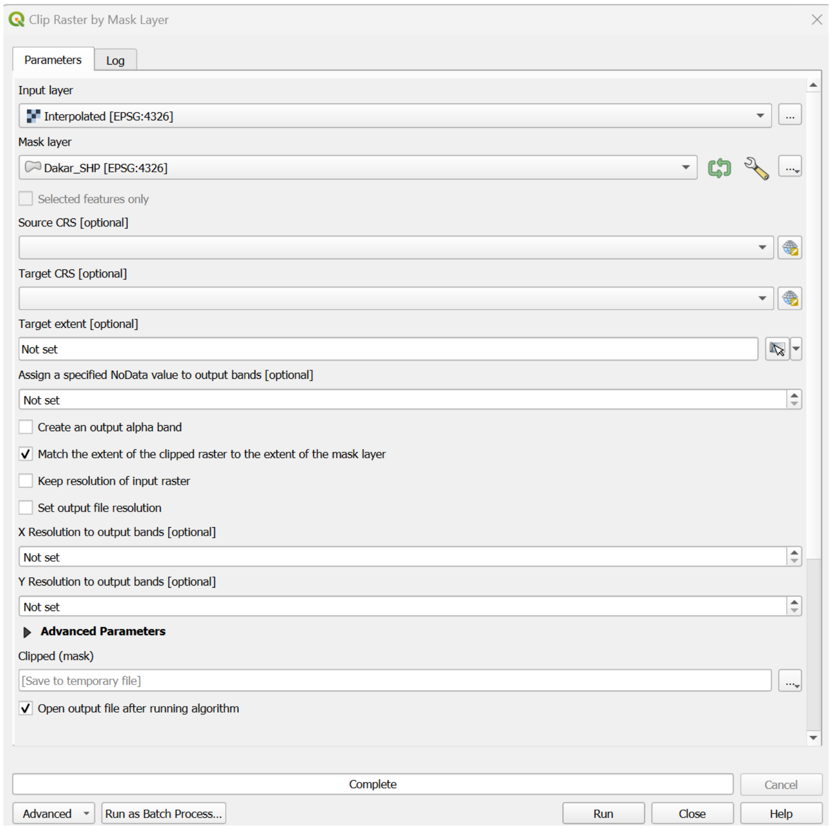
- **CGQA Pikine data:** `regulatory_daily_2023_04_07.csv` contains pm2.5 data from the regulatory monitor at Pikine in Dakar on April 7, 2023.

2. **Steps to Creating the Figure 8 and and A3**

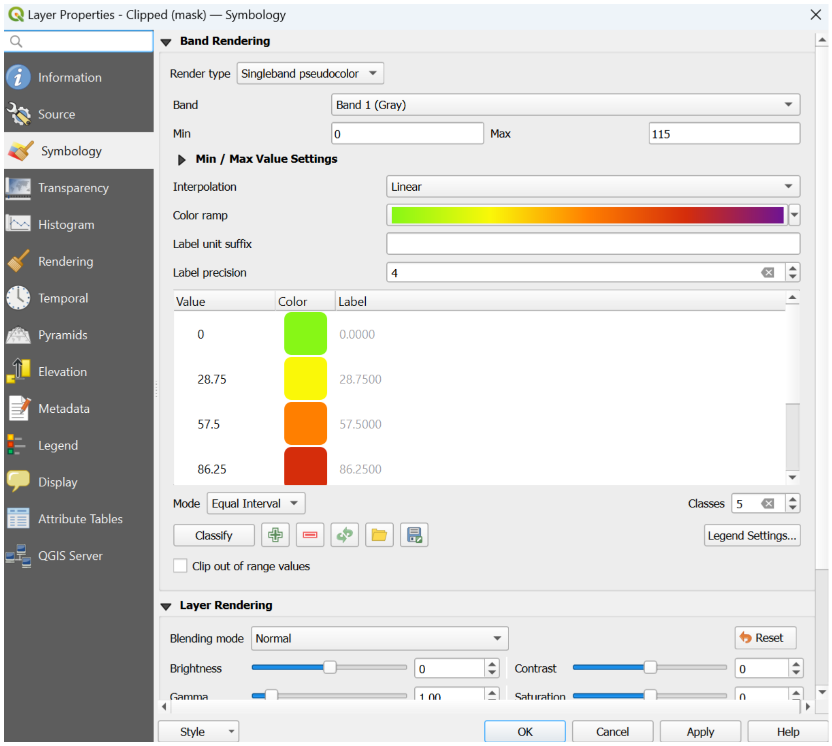
   - **Figure 8 (C): Low-cost monitors (unadjusted data)**
     - The input data consists of the geographic locations of 19 monitors and the daily average pm2.5 concentration (`avg_pm25_cf1`) in the `low_cost_daily_2023_04_07.csv`
     - Use the shape file of Dakar area as the base map
     - In the processing toolbox, locate the IDW (Inverse Distance Weighted) interpolation tool. Set the following parameters:
       * Interpolation Attribute: Select `avg_pm25_cf1`
       * Vector Layer: Choose `low_cost_daily_2023_04_07.csv`
       * Distance Coefficient: Leave this as the default value
       * Extent: Set to Dakar area
       * Output Raster Size: Set both Pixel Size X and Pixel Size Y to 0.00001



– Select Clip Raster by Mask Layer under Processing menu to crop the interpolated raster file within the Dakar area. The input layer is the interpolated raster file, and the mask layer is the Dakar shape file



– Change the color of the cropped raster file from single-band grayscale to single-band pseudocolor. Divide the values into 5 classes with equal intervals from 0 to 115, and assign the following colors to the classes, from lowest to highest values: green, yellow, orange, red, and purple. Please note that in order to ensure consistent interpretation, a standardized legend across all maps is provided.



- **Figure 8 (D): Low-cost monitors (local calibration applied)**
  - The input data consists of the geographic locations of 19 monitors and the pm2.5 concentration with local hourly calibration (`local_hourly`) in the `low_cost_daily_2023_04_07.csv`
  - Please refer to the steps outlined in **Figure 8 (C)** for the process of creating the cropped raster file
- **Figure 8 (E): Low-cost monitors (satellite calibration applied)**
  - The input data consists of the geographic locations of 19 monitors and the pm2.5 concentration with satellite hourly calibration (`satellite_hourly`) in the `low_cost_daily_2023_04_07.csv`
  - Please refer to the steps outlined in **Figure 8 (C)** for the process of creating the cropped raster file
- **Figure 8 (A): Regulatory monitors**
  - The input data is the geo-location of the regulatory monitor at Pikine in Dakar, and the daily average pm2.5 concentration (`PM2_5_20230407`) in the `regulatory_daily_2023_04_07.csv`
  - Use the shapefile of Dakar as the base map
  - Fill the shapefile of Dakar with colors corresponding to the class in which the pm2.5 concentration falls.

- **Figure 8 (B): Satellite**
  - The input data is the daily average pm2.5 concentration from satellite (`PM2_5_20230407`) in the `satellite_daily_2023_04_07.csv`
  - Use the shapefile of Dakar as the base map
  - Fill the shapefile of Dakar with colors corresponding to the class in which the pm2.5 concentration falls.
- **Figure A3 (A): Low-cost monitors (local calibration applied)**
  - The input data consists of the geographic locations of 19 monitors and the pm2.5 concentration with local daily calibration (`local_daily`) in the `low_cost_daily_2023_04_07.csv`
  - Please refer to the steps outlined in **Figure 8 (C)** for the process of creating the cropped raster file
- **Figure A3 (B): Low-cost monitors (satellite calibration applied)**
  - The input data consists of the geographic locations of 19 monitors and the pm2.5 concentration with satellite daily calibration (`satellite_daily`) in the `low_cost_daily_2023_04_07.csv`
  - Please refer to the steps outlined in **Figure 8 (C)** for the process of creating the cropped raster file