

Tracking Air Pollution from Space: High-Resolution Daily Estimates for PM2.5, O3, NO2 and CO

Contents

- 1. Overview
- 2. Data Availability
- 3. Instructions for Replicators
- 4. List of Exhibits
- 5. Requirements
- 6. Code Description
- 7. Folder Structure

Overview

The code in this replication package constructs the analysis file from the data sources on the World Bank Development Data Hub and local data using R. A main script called run all of the code to generate the data for the 7 figures and 1 table in the World Bank Policy Research Working Paper entitled, “Tracking Air Pollution from Space: High-Resolution Daily Estimates for PM2.5, O3, NO2 and CO” (No. X). The replicator should expect the code to run for about 6 hours and require approximately 500GB of data

Data Availability

They fall into three categories: (1) datasets included directly in the package, (2) datasets downloaded automatically by the code, and (3) datasets available only upon request. The table below provides a summary; detailed descriptions follow.

Other files included: temperature_2025-10-14.tif (raster dataset), FUA_All_Cell_IDs.csv (FUA cells), Global_Population_IDs.csv (pop cells), Koppen_Dummies.csv (climate) and VariableName_Source.csv (variable name with source)

#	Dataset	Availability	Included in Package
1	IQAir Ground Station Data	Restricted — upon request	Yes (two CSV files: IQAir_Combined_Stations_2_ID.csv and IQAir_Geographic_Database.csv)
2	Sentinel-5P TROPOMI Satellite Data	Publicly available	Downloaded by code
3	ERA5-Land Daily Reanalysis Data	Publicly available	Downloaded by code
4	MERIT DEM Elevation Data	Publicly available	Processed via GEE script

#	Dataset	Availability	Included in Package
5	GHS-POP Global Population Grid	Publicly available	Derived file included
6	EDGAR Emissions Database	Publicly available	Derived file included
7	Köppen-Geiger Climate Classification	Publicly available	Derived files included
8	World Bank Official Boundaries	Publicly available	Downloaded by code
9	GHSL Functional Urban Areas (FUA)	Publicly available	Derived file included

1. IQAir Ground Station Data — Restricted (available upon request) Description: Ground-level air quality measurements from IQAir’s global network of monitoring stations. Used to train and validate the machine-learning pollutant prediction models for PM2.5, O3, NO2, and CO.

Availability: Data must be requested directly from IQAir (<https://www.iqair.com/us/commercial-air-quality-monitors/api> (<https://www.iqair.com/us/commercial-air-quality-monitors/api>)). Academic and research access inquiries should be directed to IQAir’s data partnerships team.

Files excluded in the package:

IQAir_Combined_Stations_2_ID.csv — combined station-level measurements with geographic identifiers

IQAir_Geographic_Database.csv — station metadata including coordinates, country, and station type [October 16 to November 17, 2025].

2. Sentinel-5P TROPOMI Satellite Data — Publicly available, downloaded by code Description: Level-3 daily gridded satellite retrievals of tropospheric trace gases and aerosol optical depth from the TROPospheric Monitoring Instrument (TROPOMI) aboard the ESA Copernicus Sentinel-5P satellite. The four pollutant streams used are: carbon monoxide (CO), nitrogen dioxide (NO2), ozone (O3), and UV Aerosol Index (AI, used as a PM2.5 proxy). Daily files are stored in `aq_rp/wp/Data/TROPOMI_Daily_History/` under subdirectories `CO/`, `NO2/`, `O3/`, and `AI/`.

The download is automated by the pipeline scripts in each pollutant subdirectory (e.g., 1 Generate Monthly Prediction Databases CO.R).

Version/Date accessed: Daily data accessed for the period [start 09-2014] through [end 02-2026].

3. ERA5-Land Daily Reanalysis Data — Publicly available, downloaded by code Description: High-resolution daily climate reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Variables include (but are not limited to) surface temperature, wind speed, boundary layer height, and precipitation used as meteorological predictors in the pollutant estimation models. Daily CSV files are stored in `aq_rp/wp/Data/ERA5_Land_Daily_CSV/` (dated ERA5_2025-04-15 through ERA5_2025-11-22) and raster files in `ERA5_Land_Rasters/` (e.g., `temperature_2025-10-14.tif`).

The pipeline scripts automate extraction and aggregation; see the `Load data/` directory for loading utilities.

Version/Date accessed: ERA5-Land, accessed April–November 2025. CDS dataset DOI:

<https://doi.org/10.24381/cds.e2161bac> (<https://doi.org/10.24381/cds.e2161bac>).

4. MERIT DEM Elevation Data — Publicly available, processed via Google Earth Engine Description: The Multi-Error-Removed Improved-Terrain (MERIT) Digital Elevation Model provides globally consistent, high-accuracy elevation data at approximately 90 m resolution. Elevation is used as a geographic predictor in the pollutant models. The processed output (Merit_Dem_Global_ERA5.tif) is included in aq_rp/wp/Data/.

Availability: Publicly available. Processing was performed in Google Earth Engine using the script wp/GEE/merit_dem_at_0_1dd.js, which must be run first before executing the main R pipeline.

Navigate to <https://code.earthengine.google.com/> (<https://code.earthengine.google.com/>) and log in with a Google account registered for GEE access (<https://earthengine.google.com/signup/> (<https://earthengine.google.com/signup/>)). Open the script wp/GEE/merit_dem_at_0_1dd.js from the replication package. Run the script to aggregate MERIT DEM to 0.1° resolution and export the resulting raster to Google Drive or GEE Assets. Download the output TIF file and place it in aq_rp/wp/Data/Merit_Dem_Global_ERA5.tif. Original data source: Yamazaki et al. (2017). MERIT DEM. Available at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/ (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/).

5. GHS-POP Global Human Settlement Population Grid — Publicly available, derived file included Description: The Global Human Settlement Layer Population (GHS-POP) provides spatially disaggregated population counts at global coverage. It is used to construct population-weighted exposure estimates. The derived file Global_Population_IDs.csv (population cell identifiers linked to the analysis grid) is included in aq_rp/wp/Data/. The raw source rasters are stored in aq_rp/wp/Data/GHS_POP/.

Availability: Publicly available from the European Commission's Joint Research Centre (JRC).

Navigate to https://ghsl.jrc.ec.europa.eu/ghs_pop2023.php (https://ghsl.jrc.ec.europa.eu/ghs_pop2023.php) Select the desired epoch (e.g., 2020) and spatial resolution (100 m or 1 km). Download the global mosaic GeoTIFF file(s). Place raw files in aq_rp/wp/Data/GHS_POP/. Version/Date accessed: GHS-POP R2023A (epoch 2020), accessed [05-2026]. Dataset DOI: <https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE> (<https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE>).

6. EDGAR Emissions Database — Publicly available, derived file included Description: The Emissions Database for Global Atmospheric Research (EDGAR) provides gridded anthropogenic emission inventories for greenhouse gases and air pollutants by sector. Used as an emissions predictor in the pollutant estimation models. The derived file EDGAR.csv is included in aq_rp/wp/Data/.

Availability: Publicly available from the European Commission's Joint Research Centre (JRC).

7. Köppen-Geiger Climate Classification — Publicly available, derived files included Description: The updated Köppen-Geiger climate classification map (1991–2020) provides globally consistent climate zone assignments at 1 km resolution. Used to construct climate zone dummy variables as predictors in the pollutant models. Derived files Koppen_Dummies.csv and the raw classification rasters in aq_rp/wp/Data/Koppen_Geiger_1991_2020/ are included in the package.

Navigate to <https://doi.org/10.1038/s41597-023-02549-6> (<https://doi.org/10.1038/s41597-023-02549-6>) (Beck et al., 2023, Scientific Data). Download the 1991–2020 classification raster from the paper's associated data repository at https://figshare.com/articles/dataset/koppen_geiger_tif/21854202 (https://figshare.com/articles/dataset/koppen_geiger_tif/21854202). Place files in aq_rp/wp/Data/Koppen_Geiger_1991_2020/. Version/Date accessed: Beck et al. (2023) updated classification, 1991–2020 period, accessed [month-year].

8. World Bank Official Boundaries — Publicly available, downloaded by code Description: Official World Bank administrative boundary polygons (Admin 0) used for geographic mapping and spatial joins. Downloaded in the pipeline via load_wb_gad_adm0.R.

Availability: Publicly available via the World Bank Data Catalog. Navigate to <https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries> (<https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries>) Download the boundary shapefiles or GeoJSON directly from the page, or allow the pipeline script `load_wb_gad_adm0.R` to download them automatically.

9. GHSL Functional Urban Areas (FUA) — Publicly available, derived file included Description: The Global Human Settlement Layer Functional Urban Areas (GHSL-FUA) delineates urban agglomerations and their surrounding commuting zones globally. Used to aggregate pollutant predictions to city-level time series in Figures 3–7 and the appendices. The derived file `FUA_All_Cell_IDs.csv` (grid cell identifiers linked to each FUA) is included in `aq_rp/wp/Data/`.

Availability: Publicly available from the European Commission's Joint Research Centre (JRC).

Download the GHS-FUA 2019 dataset (GeoPackage or Shapefile). The raw FUA boundaries can be cross-referenced against the included `FUA_All_Cell_IDs.csv` to reproduce the cell-to-FUA linkage. Dataset DOI: <https://doi.org/10.2905/347C1B77-2D3F-4089-AE4B-B16B0B4B9816> (<https://doi.org/10.2905/347C1B77-2D3F-4089-AE4B-B16B0B4B9816>).

Instructions for Replicators

New users should follow these steps to run the package successfully:

Obtain restricted data. Request the IQAir datasets (`IQAir_Combined_Stations_2_ID.csv` and `IQAir_Geographic_Database.csv`) from IQAir directly and place them in `aq_rp/wp/Data/`.

Run the GEE elevation script. Open `wp/GEE/merit_dem_at_0_1dd.js` in Google Earth Engine (<https://code.earthengine.google.com/> (<https://code.earthengine.google.com/>)), run it, and export the output to `aq_rp/wp/Data/Merit_Dem_Global_ERA5.tif`. This step must be completed before running the R pipeline.

Ensure all other data are in place. Datasets that are automatically downloaded (TROPOMI, ERA5-Land, World Bank Boundaries) will be retrieved by the pipeline scripts. Verify internet connectivity and required API credentials are configured as needed.

Install required software. Ensure R version 4.5.0 and all required packages are installed as listed in the Requirements section. Also see Requirements section.

Create a `rtemp` directory on line 38 of the `0__Run_All_Programs.R` to help store and process temporary data.

Run the main script. Execute `wp/R Suites/0__Run_All_Programs.R`. This script orchestrates all downstream scripts in the correct order.

Check outputs. All figures and tables will be written to `wp/R Results/`.

List of Exhibits

Exhibit name	Output filename	Script	Note
Table 1	<code>Composite_Regression_Summary.docx</code>	<code>Generate Table 1 .R</code>	Found in R Results
Figure 1	<code>fig_01_pm25_24hr.png</code>	<code>Generate Figure 1.R</code>	Found in R Results

Exhibit name	Output filename	Script	Note
Figure 2	fig_02_climate.png	Generate Figure 2.R	Found in R Results
Figure 3	Figure_3.png	Generate Figures 3 4 5 6.R	Found in R Results
Figure 4	Figure_4.png	Generate Figures 3 4 5 6.R	Found in R Results
Figure 5	Figure_5.png	Generate Figures 3 4 5 6.R	Found in R Results
Figure 6	Figure_6.png	Generate Figures 3 4 5 6.R	Found in R Results
Figure 7	Figure_7_panel_cities.png	Generate Figure 7 FUA.R	Found in R Results
Figure 8	Pollutant_Correlation_Boxplots.png	Generate Figure 8 - Daily.. .R	Found in R Results
Appendix 1	Comparative_FUA_TimeSeries_Plots.docx	Generate Appendix 1.R	Found in R Results
Appendix 2	FUA_Full_Prediction_Plots.docx	Generate Appendix 2.R	Found in R Results

Requirements

Computational Requirements

Platform: x86_64-w64-mingw32/x64 (64-bit) Memory; 256GB

Software Requirements

R version 4.5.0 (2025-04-11 ucrt) Platform: x86_64-w64-mingw32/x64 Running under: Windows Server 2022 x64 (build 20348)

Matrix products: default LAPACK version 3.12.1

locale: [1] LC_COLLATE=English_United States.utf8 LC_CTYPE=English_United States.utf8
LC_MONETARY=English_United States.utf8 [4] LC_NUMERIC=C LC_TIME=English_United States.utf8

time zone: America/New_York tzcode source: internal

attached base packages: [1] parallel tools stats graphics grDevices utils datasets methods base

other attached packages: [1] future.apply_1.20.0 future_1.67.0 fs_1.6.6 tidyterra_1.0.0 officer_0.6.10
[6] lubridate_1.9.4 tibble_3.3.0 tidyverse_2.0.0 forcats_1.0.0 fasterize_1.1.0
[11] purrr_1.1.0 exactextractr_0.10.0 patchwork_1.3.1 flextable_0.9.9 colourvalues_0.3.9
[16] XML_3.99-0.18 colorspace_2.1-1 units_0.8-7 nngео_0.4.8 dplyr_1.1.4

[21] xfun_0.52 modelsummary_2.4.0 geojsonsf_2.0.3 ranger_0.18.0 ordinal_2023.12-4.1 [26] RColorBrewer_1.1-3
 rmapshaper_0.5.0 R.utils_2.13.0 R.oo_1.27.1 R.methodsS3_1.8.2
 [31] renv_1.1.5 tidyr_1.3.1 jsonlite_2.0.0 terra_1.8-60 foreign_0.8-90
 [36] caret_7.0-1 lattice_0.22-6 ggplot2_4.0.1 foreach_1.5.2 data.table_1.17.8
 [41] readr_2.1.5 Hmisc_5.2-3 stringr_1.5.1 httr2_1.2.1 httr_1.4.7
 [46] sf_1.0-21 raster_3.6-32 sp_2.2-0

loaded via a namespace (and not attached): [1] rstudioapi_0.18.0 magrittr_2.0.3 farver_2.1.2 rmarkdown_2.30
 ragg_1.5.0

[6] vctr_0.7.1 askpass_1.2.1 base64enc_0.1-3 htmltools_0.5.8.1 curl_6.4.0
 [11] Formula_1.2-5 pROC_1.19.0.1 parallelly_1.45.1 KernSmooth_2.23-26 htmlwidgets_1.6.4
 [16] plyr_1.8.9 uuid_1.2-1 lifecycle_1.0.5 iterators_1.0.14 pkgconfig_2.0.3
 [21] Matrix_1.7-3 R6_2.6.1 fastmap_1.2.0 digest_0.6.37 numDeriv_2016.8-1.1
 [26] textshaping_1.0.4 labeling_0.4.3 timechange_0.3.0 compiler_4.5.0 proxy_0.4-27
 [31] fontquiver_0.2.1 withr_3.0.2 htmlTable_2.4.3 S7_0.2.1 backports_1.5.0
 [36] DBI_1.2.3 MASS_7.3-65 lava_1.8.1 openssl_2.3.4 rappdirs_0.3.3
 [41] ucminf_1.2.2 classInt_0.4-11 ModelMetrics_1.2.2.2 otl_0.2.0 zip_2.3.3
 [46] nnet_7.3-20 glue_1.8.0 nlme_3.1-168 grid_4.5.0 checkmate_2.3.2
 [51] cluster_2.1.8.1 reshape2_1.4.4 generics_0.1.4 recipes_1.3.1 gtable_0.3.6
 [56] tzdb_0.5.0 class_7.3-23 hms_1.1.4 utf8_1.2.6 xml2_1.3.8
 [61] tables_0.9.31 pillar_1.11.1 splines_4.5.0 survival_3.8-3 tidyselect_1.2.1
 [66] fontLiberation_0.1.0 knitr_1.51 fontBitstreamVera_0.1.1 gridExtra_2.3 V8_6.0.5
 [71] stats4_4.5.0 hardhat_1.4.1 timeDate_4041.110 stringi_1.8.7 evaluate_1.0.5
 [76] codetools_0.2-20 gdtools_0.4.4 cli_3.6.5 rpart_4.1.24 systemfonts_1.3.1
 [81] Rcpp_1.1.0 globals_0.18.0 gower_1.0.2 listenv_0.9.1 viridisLite_0.4.2
 [86] ipred_0.9-15 scales_1.4.0 prodlim_2025.04.28 e1071_1.7-16 rlang_1.1.7

Memory and Runtime and Storage Requirements

The scripts are run on a computing environment with 256MB memory.

Code Description

The scripts are located in the wp/R Suites/ directory. The pipeline is organized into four pollutant-specific subdirectories (PM25/, O3/, NO2/, CO/) and a set of cross-cutting scripts for figures, tables, and appendices.

Execution order:

GEE preprocessing (run once, outside R): wp/GEE/merit_dem_at_0_1dd.js — extracts and aggregates MERIT DEM elevation data at 0.1° resolution for use as a model predictor. Main pipeline: wp/R Suites/0__Run_All_Programs.R — orchestrates all subsequent steps below. For each pollutant (PM25, O3, NO2, CO), scripts run in sequence: Script 1: Build monthly prediction databases from TROPOMI and ERA5 inputs. Script 2: Generate estimation (training) databases by merging satellite, meteorological, and ground station data. Script 3: Train the random forest estimator model. Script 4: Generate global predictions. Scripts 5–9: Aggregate predictions and monitor data to FUA-level time series and compute statistics/correlations. Figure and table generation scripts produce all exhibits listed in List of Exhibits. Supporting directories: Build_Databases/ (database construction utilities), Load data/ (data loading helpers), UDF/ (user-defined functions), Output_Data/ and Output_GIS/ (intermediate and final outputs).

The Build_Databases directory includes scripts to build the databases. The data directory has data. The figures_rp directory contains the figures. The gis directory has geospatial data. The Load data directory loads data. The Output_Data directory contains output data. The Output_GIS directory contains output geospatial data. The tables_rp contains tables. The UDF directory has user defined functions.

Folder Structure

```
=====
# Code and Reproducibility Package
=====
```

```
# Root:
```

```
wp/
```

```
├─ Data/
│   ├── ERA5_Land_Rasters/
│   │   └─ temperature_2025-10-14.tif
│   ├── GHS_POP/
│   │   ├── ghs_pop_dd01.tif
│   │   ├── GHS_POP_E2020_GLOBE_R2023A_4326_30ss_V1_0.tif
│   │   ├── GHS_POP_E2020_GLOBE_R2023A_4326_30ss_V1_0.tif.ovr
│   │   ├── GHS_POP_E2020_GLOBE_R2023A_4326_30ss_V1_0.zip
│   │   └─ Global_Population_IDs.csv
│   ├── EDGAR.csv
│   ├── FUA_All_Cell_IDs.csv
│   ├── Global_Population_IDs.csv
│   ├── IQAir_Combined_Stations_2_ID.csv
│   ├── IQAir_Geographic_Database.csv
│   ├── Koppen_Dummies.csv
│   ├── Merit_Dem_Global_ERA5.tif
│   └─ VariableName_Source.csv
├─ GEE/
│   └─ merit_dem_at_0_1dd.js
├─ R_Results/
│   ├── Comparative_FUA_TimeSeries_Plots.docx
│   ├── Composite_Regression_Summary.docx
│   ├── fig_01_pm25_24hr.png
│   ├── fig_02_climate.png
│   ├── Fig_8_corr_text.txt
│   ├── Figure_3.csv
│   ├── Figure_3.png
│   ├── Figure_4.csv
│   ├── Figure_4.png
│   ├── Figure_5.csv
│   ├── Figure_5.png
│   ├── Figure_6.csv
│   ├── Figure_6.png
│   ├── Figure_7_panel_cities.png
│   ├── FUA_Full_Prediction_Plots.docx
│   ├── Pollutant_Correlation_Boxplots_mod.png
│   ├── Pollutant_Correlation_Boxplots_mod2026-06-04.csv
│   └─ results_6_1_Daily_tracking.csv
└─ R_Suites/
    ├── CO/
    │   ├── 1_Generate_Monthly_Prediction_Databases_CO.R
    │   └─ 2_Generate_CO_Estimation_Databases.R
```

- └─ 3 Revised_CO_Estimator 3.R
- └─ 4 CO_Predictor_4.R
- └─ 5 Generate FUA Predicted CO Time Series.R
- └─ 6 Generate FUA Predicted CO Statistics Time Series.R
- └─ 7 Generate FUA CO Time Series From Monitors.R
- └─ 8 Generate FUA CO Monitor Statistics Time Series.R
- └─ 9 Generate FUA CO Statistics Correlations.R
- └─ N02/
 - └─ 1 Generate Monthly Prediction Databases N02.R
 - └─ 2 Generate N02 Estimation Databases.R
 - └─ 3 Revised_N02_Estimator 3.R
 - └─ 4 N02_Predictor_4.R
 - └─ 5 Generate FUA Predicted N02 Time Series.R
 - └─ 6 Generate FUA Predicted N02 Statistics Time Series.R
 - └─ 7 Generate FUA N02 Time Series From Monitors.R
 - └─ 8 Generate FUA N02 Monitor Statistics Time Series.R
 - └─ 9 Generate FUA N02 Statistics Correlations.R
- └─ O3/
 - └─ 1 Generate Monthly Prediction Databases O3.R
 - └─ 2 Generate O3 Estimation Databases.R
 - └─ 3 Revised_O3_Estimator 3.R
 - └─ 4 O3_Predictor_4.R
 - └─ 5 Generate FUA Predicted O3 Time Series.R
 - └─ 6 Generate FUA Predicted O3 Statistics Time Series.R
 - └─ 7 Generate FUA O3 Time Series From Monitors.R
 - └─ 8 Generate FUA O3 Monitor Statistics Time Series.R
 - └─ 9 Generate FUA O3 Statistics Correlations.R
- └─ PM25/
 - └─ 1 Generate Monthly Prediction Databases PM25.R
 - └─ 2 Generate PM25 Estimation Databases.R
 - └─ 3 Revised_PM25_Estimator 3.R
 - └─ 4 PM25_Predictor_4.R
 - └─ 5 Generate FUA Predicted PM25 Time Series.R
 - └─ 6 Generate FUA Predicted PM25 Statistics Time Series.R
 - └─ 7 Generate FUA PM25 Time Series From Monitors.R
 - └─ 8 Generate FUA PM25 Monitor Statistics Time Series.R
 - └─ 9 Generate FUA PM25 Statistics Correlations.R
- └─ 0_Run_All_Programs.R
- └─ aq_wp_global_libraries.R
- └─ Generate Appendix 1 - Monitored vs. Predicted Ambient Pollution.R
- └─ Generate Appendix 2 - Predicted Ambient Pollution.R
- └─ Generate Figure 1.R
- └─ Generate Figure 2.R
- └─ Generate Figure 7 FUA.R
- └─ Generate Figure 8 - Daily Tracking Correlations Boxplots.R
- └─ Generate Figures 3 4 5 6.R
- └─ Generate Figures 3 desc.R
- └─ Generate Figures 4 desc.R
- └─ Generate Figures 5 desc.R
- └─ Generate Figures 6 desc.R
- └─ Generate Geographic Coast.R
- └─ Generate Geographic Database.R

- └─ Generate Population.R
- └─ Generate small land area mask.R
- └─ Generate Table 1 Model Estimation Statistics.R
- └─ Generate Weather Database.R
- └─ load_data_era5.R
- └─ load_ddh_data.R
- └─ load_koppen.R
- └─ load_land_and_ocean.R
- └─ load_pop.R
- └─ load_tropomo_data.R
- └─ load_wb_gad_adm0.R

=====

Processing Folder (server-local – not included in package)

=====

aq_rp/

- └─ Data/
 - └─ Coast/
 - └─ Global Land Coastal Ocean Shapefiles.dbf
 - └─ Global Land Coastal Ocean Shapefiles.prj
 - └─ Global Land Coastal Ocean Shapefiles.shp
 - └─ Global Land Coastal Ocean Shapefiles.shx
 - └─ IQAir_Geographic_Coast_Mask_50km.tif
 - └─ IQAir_Geographic_LandFrac.tif
 - └─ IQAir_Geographic_LandGTE50pct.tif
 - └─ IQAir_Geographic_LandGTE50pct_Mask.tif
 - └─ IQAir_Geographic_LandLT50pct.tif
 - └─ IQAir_Geographic_LandLT50pct_Mask.tif
 - └─ GAD/
 - └─ WB_GAD_ADM0.cpg
 - └─ WB_GAD_ADM0.dbf
 - └─ WB_GAD_ADM0.prj
 - └─ WB_GAD_ADM0.shp
 - └─ WB_GAD_ADM0.shx
 - └─ WB_GAD_ADM0_complete.dbf
 - └─ WB_GAD_ADM0_complete.prj
 - └─ WB_GAD_ADM0_complete.shp
 - └─ WB_GAD_ADM0_complete.shx
 - └─ WB_GAD_ADM0_NDLA.cpg
 - └─ WB_GAD_ADM0_NDLA.dbf
 - └─ WB_GAD_ADM0_NDLA.prj
 - └─ WB_GAD_ADM0_NDLA.shp
 - └─ WB_GAD_ADM0_NDLA.shx
 - └─ WB_GAD_LandFrac.tif
 - └─ WB_GAD_ocean_mask.cpg
 - └─ WB_GAD_ocean_mask.dbf
 - └─ WB_GAD_ocean_mask.prj
 - └─ WB_GAD_ocean_mask.shp
 - └─ WB_GAD_ocean_mask.shx
 - └─ GHS_POP/

- └─ ghs_pop_dd01.tif
- └─ EDGAR.csv
- └─ ERA5_Weather_Database.csv
- └─ FUA_All_Cell_IDs.csv
- └─ GHS_FUA_UCDB2015_GLOBE_R2019A_54009_1K_V1_0.gpkg
- └─ Global_Population_IDs.csv
- └─ IQAir_Coastal_Location_50km.csv
- └─ IQAir_Combined_Stations_2_ID.csv
- └─ IQAir_Elevations.csv
- └─ IQAir_Geographic_Database.csv
- └─ IQAir_Geographic_Database_50km.csv
- └─ IQAir_Latitude_Bands.csv
- └─ IQAir_Weather_Database.csv
- └─ Koppen_Dummies.csv
- └─ VariableName_Source.csv
- └─ ERA5_Land_Daily_CSV/
 - └─ 2025-04/
 - └─ ERA5_2025-04-15.csv
 - └─ ...
 - └─ 2025-05/
 - └─ ERA5_2025-05-01.csv
 - └─ ...
 - └─ 2025-06/
 - └─ ERA5_2025-06-01.csv
 - └─ ...
 - └─ 2025-07/
 - └─ ERA5_2025-07-01.csv
 - └─ ...
 - └─ 2025-08/
 - └─ ERA5_2025-08-01.csv
 - └─ ...
 - └─ 2025-09/
 - └─ ERA5_2025-09-01.csv
 - └─ ...
 - └─ 2025-10/
 - └─ ERA5_2025-10-01.csv
 - └─ ...
 - └─ 2025-11/
 - └─ ERA5_2025-11-01.csv
 - └─ ...
- └─ ERA5_Land_Monthly_RAR/
 - └─ ERA5_Land_2025-04.rar
 - └─ ERA5_land_raster.tif
- └─ ERA5_Land_Rasters/
 - └─ temperature_2025-10-14.tif
- └─ FUA_Predictions/
 - └─ *CO_FUA_Correlations.csv
 - └─ *CO_FUA_Monitor_Statistics_Dates.csv
 - └─ *CO_FUA_Monitors_TimeSeries_2025.csv
 - └─ *CO_FUA_Statistics_Dates.csv
 - └─ *FUA_CO_TimeSeries_2025.csv
 - └─ *FUA_Monitored_vs_Predicted_PWMeans.csv

- ├─ *FUA_N02_TimeSeries_2025.csv
- ├─ *FUA_O3_TimeSeries_2025.csv
- ├─ *FUA_PM25_TimeSeries_2025.csv
- ├─ *N02_FUA_Correlations.csv
- ├─ *N02_FUA_Monitor_Statistics_Dates.csv
- ├─ *N02_FUA_Monitors_TimeSeries_2025.csv
- ├─ *N02_FUA_Statistics_Dates.csv
- ├─ *O3_FUA_Correlations.csv
- ├─ *O3_FUA_Monitor_Statistics_Dates.csv
- ├─ *O3_FUA_Monitors_TimeSeries_2025.csv
- ├─ *O3_FUA_Statistics_Dates.csv
- ├─ *PM25_FUA_Correlations.csv
- ├─ *PM25_FUA_Monitor_Statistics_Dates.csv
- ├─ *PM25_FUA_Monitors_TimeSeries_2025.csv
- ├─ *PM25_FUA_Statistics_Dates.csv
- ├─ FUA_Predictions_Graphs/
 - ├─ CO/
 - └─ *CO_[FUA].png
 - ├─ N02/
 - └─ *N02_[FUA].png
 - ├─ O3/
 - └─ *O3_[FUA].png
 - ├─ PM25/
 - └─ *PM25_[FUA].png
- ├─ GIS/
 - └─ Merit_Dem_Global_ERA5.tif
- ├─ Koppen_Geiger_1991_2020/
 - └─ koppen_geiger_0p1.tif
- ├─ Model_Predictions/
 - ├─ CO/
 - ├─ N02/
 - ├─ O3/
 - └─ PM25/
- ├─ Model_Predictions_Data/
 - ├─ *CO_Predictions_[Month_YYYY].csv
 - ├─ *N02_Predictions_[Month_YYYY].csv
 - ├─ *O3_Predictions_[Month_YYYY].csv
 - └─ *PM25_Predictions_[Month_YYYY].csv
- ├─ Model_Results/
 - ├─ CO/
 - ├─ *CO Adjusted Test Prediction Regression.rda
 - ├─ *CO Calibration Prediction Regression.rda
 - ├─ *CO Test Prediction Regression.rda
 - ├─ *CO Training Prediction Regression.rda
 - ├─ *CO_Adjustment_Coefficients.csv
 - ├─ *CO_Estimate_Combined_Group.csv
 - ├─ *rf_model_CO.rds
 - ├─ *Variable_Importance_Plot_CO.png
 - └─ *Variable_Importance_Plot_CO_withSource.png
 - ├─ N02/
 - ├─ *N02 Adjusted Test Prediction Regression.rda
 - └─ *N02 Calibration Prediction Regression.rda

```

├── *N02_Test_Prediction_Regression.rda
├── *N02_Training_Prediction_Regression.rda
├── *N02_Adjustment_Coefficients.csv
├── *N02_Estimate_Combined_Group.csv
├── *rf_model_N02.rds
├── *rf_predictor_vars.rds
├── *Variable_Importance_Plot_N02.png
├── *Variable_Importance_Plot_N02_withSource.png
├── 03/
│   ├── *03_Adjusted_Test_Prediction_Regression.rda
│   ├── *03_Calibration_Prediction_Regression.rda
│   ├── *03_Test_Prediction_Regression.rda
│   ├── *03_Training_Prediction_Regression.rda
│   ├── *03_Adjustment_Coefficients.csv
│   ├── *03_Estimate_Combined_Group.csv
│   ├── *rf_model_03.rds
│   ├── *Variable_Importance_Plot_03.png
│   └── *Variable_Importance_Plot_03_withSource.png
├── PM25/
│   ├── *PM25_Adjusted_Test_Prediction_Regression.rda
│   ├── *PM25_Calibration_Prediction_Regression.rda
│   ├── *PM25_Test_Prediction_Regression.rda
│   ├── *PM25_Training_Prediction_Regression.rda
│   ├── *PM25_Adjustment_Coefficients.csv
│   ├── *PM25_Estimate_Combined_Group.csv
│   ├── *rf_model_PM25.rds
│   ├── *Variable_Importance_Plot_PM25.png
│   └── *Variable_Importance_Plot_PM25_withSource.png
├── PPTs/
│   ├── *CO_FUA_TimeSeries_Plots.pptx
│   ├── *N02_FUA_TimeSeries_Plots.pptx
│   ├── *03_FUA_TimeSeries_Plots.pptx
│   └── *PM25_FUA_TimeSeries_Plots.pptx
├── Prediction_Databases/
│   ├── *CO_Prediction_Database_[Month_YYYY].csv
│   ├── *N02_Prediction_Database_[Month_YYYY].csv
│   ├── *03_Prediction_Database_[Month_YYYY].csv
│   └── *PM25_Prediction_Database_[Month_YYYY].csv
├── Predictions_Data/
├── Predictions_GIS/
│   ├── CO/
│   │   ├── *coded_predicted_co_[YYYY-MM-DD].tif
│   │   └── *predicted_co_[YYYY-MM-DD].tif
│   ├── N02/
│   │   ├── *coded_predicted_no2_[YYYY-MM-DD].tif
│   │   └── *predicted_no2_[YYYY-MM-DD].tif
│   ├── 03/
│   │   ├── *coded_predicted_o3_[YYYY-MM-DD].tif
│   │   └── *predicted_o3_[YYYY-MM-DD].tif
│   └── PM25/
│       ├── *coded_predicted_pm25_[YYYY-MM-DD].tif
│       └── *predicted_pm25_[YYYY-MM-DD].tif

```

- └─ R Results/
 - └─ Comparative_FUA_TimeSeries_Plots.docx
 - └─ Composite_Regression_Summary.docx
 - └─ fig_01_pm25_24hr.png
 - └─ fig_02_climate.png
 - └─ Fig_8_corr_text.txt
 - └─ Figure_3.csv
 - └─ Figure_3.png
 - └─ Figure_3_paragraph.txt
 - └─ Figure_4.csv
 - └─ Figure_4.png
 - └─ Figure_4_paragraph.txt
 - └─ Figure_5.csv
 - └─ Figure_5.png
 - └─ Figure_5_paragraph.txt
 - └─ Figure_6.csv
 - └─ Figure_6.png
 - └─ Figure_6_paragraph.txt
 - └─ Figure_7_panel_cities.png
 - └─ FUA_Full_Prediction_Plots.docx
 - └─ Pollutant_Correlation_Boxplots_mod.png
 - └─ Pollutant_Correlation_Boxplots_mod2026-06-04.csv
 - └─ results_6_1_Daily_tracking.csv
- └─ R Scratch Images/
 - └─ *CO_[FUA].png
 - └─ *NO2_[FUA].png
 - └─ *O3_[FUA].png
 - └─ *PM25_[FUA].png
- └─ Scratch_Data/
- └─ Scratch_GIS/
 - └─ Check_Coastal_50km.tif
 - └─ Check_Elevation_50km.tif
 - └─ Check_Latitude_Bands_50km.tif
 - └─ Check_TempMean_50km.tif
- └─ Scratch_Images/
 - └─ *CO_[FUA]_mon_vs_pred.png
 - └─ *CO_[FUA]_pred_only.png
 - └─ *NO2_[FUA]_mon_vs_pred.png
 - └─ *NO2_[FUA]_pred_only.png
 - └─ *O3_[FUA]_mon_vs_pred.png
 - └─ *O3_[FUA]_pred_only.png
 - └─ *PM25_[FUA]_mon_vs_pred.png
 - └─ *PM25_[FUA]_pred_only.png
- └─ Store/
 - └─ *_Active_Test.csv
 - └─ *_Active_Train.csv
 - └─ *_Store_Test.csv
 - └─ *_Store_Train.csv
- └─ TROPOMI_Daily_History/
 - └─ AI/
 - └─ AI_100_Means.csv

```

├── ...
│   ├── AI_332_Means.csv
├── CO/
│   ├── CO_100_Means.csv
│   ├── ...
│   └── CO_321_Means.csv
├── NO2/
│   ├── NO2_100_Means.csv
│   ├── ...
│   └── NO2_317_Means.csv
├── O3/
│   ├── O3_100_Means.csv
│   ├── ...
│   └── O3_320_Means.csv
├── ERA5_land_raster.tiff
├── FUA_All_Cell_IDs.csv
├── Global_Land_Coastal_Ocean.rar
├── IQAir_Combined_Stations_2_ID.csv
├── TROPOMI_AI
└── VariableName_Source.csv

```

* denotes files as the result of the script