

--- Data & File Descriptions ---

A) Raw data files

Raw data includes the data measured by measurement devices/sensors from the field/lab (all csv in this project) in “Mahdavi et al. (2021) Environmental Pollution” (particulate matter (PM) phase). Due to data privacy, only a representative sample of the raw data (meaning one from each type) is shared (in the “Raw Data Files” folder). See “Table 1.pdf” at the end of this document, “Data Pipeline and Code Configuration.jpg”, and “list_of_files.xlsx” files to understand more about raw data.

B) Code files

Codes include all “.py” or “.ipynb” that read raw data or previously processed data to generate processed data, figures, calculations, or statistical analyses. There are four different code types used in this project:

- **Processing codes:** which process from raw data (stored in the “Processing Codes” folder).
- **Generic code:** which does generic processing such as path file name correction or labelling categorical parameters (not shown in “Data Pipeline and Code Configuration.jpg”).
- **Calculation codes:** which present numeric or statistical results from the processed data (stored in the main “Code” folder).
- **Visualization codes:** which aim to plot figures. This may come with additional processing of data to make the dataframe compatible for plotting (stored in the “Plotting Codes” folder).

See “Data Pipeline and Code Configuration.pdf”, and “list_of_files.xlsx” files to see the list of code blocks. (Two generic codes titled “notion_correction.py” and “labels_all.py” is not shown in “Data Pipeline and Code Configuration.pdf” but it is called by many code blocks).

C) Processed data files

Processed data includes the data generated after processing in data pipelines in “Mahdavi et al. (2021) Environmental Pollution”. See “Data Pipeline and Code Configuration.pdf”, and “list_of_files.xlsx” files to check the processed data lists and blocks. “Data Pipeline & Code Configurations.jpg” illustrates the entire data pipeline processing that generates processed data using codes from the raw data or previously processed data. Table 2 (at the end of this document) has a list of all these files.

A separate code block (“df_summary.ipynb”) also presents a summary of some processed files. For data privacy purposes, not all the processed dataframes have been presented.

D) Plots

Plots include all the figures presented in “Mahdavi & Siegel (2021) IA” (PM Phase) from the processed data. See “Data Pipeline and Code Configuration.pdf”, and “list_of_files.xlsx” for more information. The plots aren’t presented in separate files (e.g., jpg) but illustrated in the same code file generating them (in Jupyter). See “Plotting Codes” for more information.

E) Other files

Any other file (mostly guidelines or descriptions) not classified above. A full list is available in “list_of_files.xlsx” (sheet “other_summary_files”) provided in the repository.

Table 1 – Raw Data Summary (Mahdavi et al. (2021) EP – Airborne Phase)

Item #	Raw Data Files & Name Conventions	Raw Data File Description	Variables/Columns	Variables/Columns Explanation
1	qff_eval_verisx_{sni}_{yyymmdd}_{in}.csv (×6)	Pressure sensor data for runtime calculations	Date Time	Date and time (yyddmm and hh:mm)
			Current, Amps	Current in amps from pressurized sensor tip
			Pressure, Pa	Pressure recorded from the pressure sensor tip
2	qff_eval_dtdrxx_{sni}_{yyymmdd}_{in}.txt (×6)	DustTrak DRX particle monitoring sensor raw data	Date	Date: yyymmdd
			Time	Time: hh:mm:ss
			PM1	Concentration of PM ₁
			PM2.5	Concentration of PM _{2.5}
			RESP	Concentration of respirable fraction
			PM10	Concentration of PM ₁₀
			Total	Total suspended particle concentration
3	qff_eval_dc1700_{sni}_{yyymmdd}_rk.txt (×6)	DC1700 particle monitoring sensor raw data	Date/Time	Date and time (yyddmm and hh:mm)
			Small	Small channel number concentration (> 0.5 µm)
			Large	Small channel number concentration (> 2.5 µm)
4	Filter_Master_v21_200206_am.xlsx ⁽¹⁾	Gravimetric analyses of all filters used (HVAC and small 37mm airborne samplers)	Filters S/N	Serial number of the filter
			Filter Position	Where filter was deployed: filter cassette, DustTrak, PEM, Cascade impactor.
			Metrics	What filter was used to measure: TSP, PM2.5, gravimetric size distribution.
			Visit #	Visit number: 1-6.
			Pre-deploy mass	Mass of filter prior to service.
			Launch Date	Date when the filter started service (yyymmdd).
			Launch Time	Time when the filter started service (mmhh).
			Initial Flow	Initial flow of the sampler in which the filter was deployed
			Stop Date	Date when the filter ended service (yyymmdd).
			Stop Time	Time when the filter ended service (mmhh).
			Final Flow	Final flow of the sampler in which the filter was deployed
			Average Flow	Average flow of the sampler ((Initial Flow+ Final Flows) /2)
			Post-deploy Mass	Mass of filter after the service (and PM collection)
			PM Mass	Mass of PM collected on the filter (Post-deploy Mass – Pre-deploy Mass)
5	Metals Results_Alireza.xlsx	ICP-MS results of the trace metals	SN	Filter serial number
			Location deployed	Where filter was deployed: filter cassette, DustTrak, PEM, Cascade impactor.
			Dust/tsp mass_mg	Mass of dust collected on the filter
6	Metals_UC.xlsx	Measurement uncertainties	Trace metal (× 26)	26 trace metals (Na, Pb, Cd, Ni, Cu etc.,)
7	visit_time_cutoff.xlsx		visit	Visit #
			date	Date where the visit started (yyymmdd)
			time	Time where the visit started (hh:mm)

(1) Only some columns are described

Table 2 – Processed Data Summary (Mahdavi et al. (2021) EP – Airborne Phase)

Item #	Processed File Name	Processed File Description	Variables/Columns	Variables/Columns Explanation
1	pm_master_all.xlsx	A master spreadsheet including all airborne PM data (pm_master_all_error.xlsx has some other columns not show here)	visit	Visit number: 1-6
2	pm_master_all_error.xlsx		TSP concentration	TSP concentration
			TSP Concentration Error	TSP Concentration Error
			PM10 SCI	PM ₁₀ measured by Sioutas Cascade Impactor (SCI)
			PM10 SCI Error	PM ₁₀ error measured by Sioutas Cascade Impactor (SCI)
			PM2.5 SCI	PM _{2.5} measured by Sioutas Cascade Impactor (SCI)
			PM2.5 SCI Error	PM _{2.5} error measured by Sioutas Cascade Impactor (SCI)
			PM1 SCI	PM ₁ measured by Sioutas Cascade Impactor (SCI)
			PM1 SCI Error	PM ₁ error measured by Sioutas Cascade Impactor (SCI)
			PM2.5 PEM	PM _{2.5} measured by PEM sampler
			PM2.5 PEM Error	PM _{2.5} error measured by PEM sampler
3	tm_concentration_master.xlsx	A master spreadsheet including all airborne trace metal data	visit	Visit number: 1-6
			Average Flow	Average flow measured in filter cassette during weekly sampling
			Pb, As, Cd etc., (12 trace metals)	Trace metal concentration (in ng/m ³)
			Pb_error, As_error etc., (× 12)	Trace metal concentration error (in ng/m ³)
4	Runtime_master.xlsx	Runtime and time-series system status	Time	Date and time (yymmdd hh:mm) (only in runtime_master.xlsx)
5	Runtime_sum_mode.xlsx		Pressure	Pressure recorded by the sensor
6	Runtime_sum.xlsx		Visit	Visit number: 1-6
			Mode	System status (Off, compressor, fan only)
			runtime	Runtime fraction value (only in runtime_sum.xlsx and runtime_sum_mode.xlsx)
4	dc_1700.xlsx	DC1700 time-series data	Similar to dc raw data + visit	Similar to raw data + visit number
5	Dt_drx.xlsx	Dt_drx time-series data	Similar to dt raw data + visit	Similar to raw data + visit number