--- 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) Pre-processed files

These files have been already processed in other projects (using STATA or Python). They have a format of either ".dta", ".csv", or ".xlsx" (depending on how stored from STATA or python). They play the role of raw files in this project, but still called as pre-processed file data as they don't come from purely raw data (e.g., sensors or lab sheet completions). These data files are "runtime_hr.xlsx", "tm_mass.xlsx", "tm_mdl_errors.xlsx", "pm_master_all_error.xlsx" aren't classified in a separate table

C) 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).

D) 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.

E) 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.

F) 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 – QFF Phase)

Item #	Raw Data Files & Name Conventions	Raw Data File Description	Variables/Columns	Variables/Columns Explanation
1	Pressure_Flow_v09_191114_am.xlsx	Pressure values and flow rate calculations from HVAC duct using True Flow Plate (TFP)	Visit Type	Visit type (initial, final, short visit)
			Pressure_Type	Pressure type by TFP (TFP, TFSOP (down and supply), NSOP (down and supply))
			System_Mode	Pressure recorded from the pressure sensor tip
			Pressure_1 (through _6)	Pressure recorded (in Pa) from six consecutive measurements
			Average_Pressure	Average of 6 pressures read from Pressure_1 (through _6)
			Flow_rate	Flow rate calculated using pressure values
2	mini_wras_time_cutoff.xlsx	Sampling cutoff times for Mini-WRAS tools	location	Location of Mini-WRAS (upstream, downstream)
			sn	Serial number of Mini-WRAS
			start	Start time of sampling (yymmdd hh:mm)
			end	End time of sampling (yymmdd hh:mm)
			visit	Visit number (1, 6)
3	qff_eval_visit{nn}_miniwr_{sni}_{yymmdd}_{in}-C.xlsx (Sheet “Mass values”) (× 4)	Mini_WRAS raw data of airborne size distribution	date & time	Date and time (yyddmm and hh:mm)
			total mass	Total mass of particles recorded (in ng)
			{size - nanometer} (× 41)	Channel bin (× 41)
4	qff_eval_visit{nn}_miniwr_{sni}_{yymmdd}_{in}-C.xlsx (Sheet “PM values”) (× 4)	Mini_WRAS raw data of airborne size distribution	date & time	Date and time (yyddmm and hh:mm)
			{PM - } (× 6)	6 columns: PM_{10}, $PM_{2.5,}$ $PM_{1,}$ Inhalable, Thoracic, Alveolic concentration
5	psd_qff_evaluation.csv	PSD raw data from laser diffraction particle sizer	Record Number	A number that tracks the measurements of LDPS (#)
			Size bins (× 92)	Bin sizes varying from 0.1 to 3500 (total of 92 bins) (µm)
			Dx(10), Dx(50), Dx(90)	10^{th}, 50^{th}, and 90^{th} percentile sizes of the samples PSDs (µm)
			Sample Name	Name of the sample at the measurement time given by the user
			Measurement Date Time	Date and time of measurement
			File Name	Name of the file (including all measurement observations) given by the user
			SOP File Name	Name of the Standard Operating Procedure (SOP) defined for the LDPS measurement
			Laser Obscuration	The obscuration of the sensor laser during measurement (%)
			Particle Refractive Index	Light Index of Refraction (IR) during the measurement (-)
			Particle Absorption Index	Light Index of Absorption (IR) during the measurement (-)
			Particle Density	Density of particle assumed in SOP (inputted not measured) (g/cm^3)
			Ultrasonication Duration (SOP)	The duration of ultrasonication to the sample prior to testing (to avoid particle agglomeration during PSD measurement) (s)
			Ultrasonication Mode	Mode of ultrasonication
			Original Record Number	The original number tracking LDPS measurement prior to PSD recalculation for changing input parameters (#)

			Specific Surface Area	Specific surface area of the particles (μm^2)
6	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 (yymmdd).
			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 (yymmdd).
			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)
7	lds_extraction_qff_F00_00_ _190927.xlsx	Lab data sheets collecting gravimetric analyses of the HVAC filter prior and after extraction.	M_filter_blank	Mass of blank filter (not loaded) (g)
			M_filter_pre	Mass of filter after loading prior to any extraction (g)
			M_dust	Mass of dust loaded in the tested filter (g)
			M_vd_empty	Mass of recovery container for after-sieve dust before all extraction cycles (g)
			M_vs_empty	Mass of recovery container for sieved dust prior to all extraction cycles (g)
			Cycle_N	Dust extraction cycle number over the same filter (#)
			M_filter_post	Mass of HVAC filter after an extraction cycle (g)
			M_ph_clean	Mass of the after-sieve HEPA sock prior to an extraction cycle (g)
			M_ph_full	Mass of the after-sieve HEPA sock after an extraction cycle (g)
			M_ph_dumped	Mass of the after-sieve HEPA sock after dust transfer in recovery container (g)
			M_vd	Mass of after-sieve dust recovery container after an extraction cycle (g)
			M_vs	Mass of sieved dust recovery container after an extraction cycle (g)
			M_filter_change	Mass extracted from filter after an extraction cycle (g)
			M_filter_change_cum	Cumulative mass extracted from filter after all extraction cycles over the same filter (g)
			M_d	Mass of dust recovered from filter after an extraction cycle (after-sieve) (g)
			M_d_cum	Cumulative mass of dust recovered from the filter after all extraction cycles over the same filter (after-sieve) (g)
			M_s	Mass of dust recovered from the filter after an extraction cycle (pre-sieve) (%)
			M_s_cum	Cumulative mass of dust recovered from the filter after all extraction cycles over the same filter (pre-sieve) (g)
			E	Suction efficiency after an extraction cycle (mass extracted / total mass loaded in the filter) (%)
			E_cum	Cumulative suction efficiency after all extraction cycles (cumulative mass extracted / total mass loaded in the filter) (%)

			C	After-sieve collection efficiency by the sampler (after-sieve mass collected / mass extracted) (%)
			D	Transfer efficiency after an extraction cycle (mass recovered into after-sieve recovery container / after-sieve mass collected) (%)
			CE	After-sieve recovery efficiency after an extraction cycle (after-sieve dust mass recovered / dust mass loaded in filter) (%)
			CE_cum	Cumulative after-sieve recovery efficiency after all extraction cycles (cumulative after-sieve dust mass recovered / dust mass loaded in filter) (%)
			M_C	Mass closure after an extraction cycle (dust recovered by sieve and after-sieve/ mass extracted from a filter) (%)
			M_C_cum	Cumulative mass closure after all extraction cycles (cumulative dust recovered by sieve + after-sieve/ cumulative mass extracted from a filter) (%)
8	Metals_Results_Alireza.xlsx	See Airborne repository Data & File Description for more info.	See Airborne repository Data & File Description for more info.	See Airborne repository Data & File Description for more info.

(1) Only some columns are described

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

Item #	Processed File Name	Processed File Description	Variables/Columns	Variables/Columns Explanation
1	Flow_rates.xlsx		Visit_#	Visit number: 1-6
			Flow Fan	Flow recorded when system is on Fan Only mode
			Flow Compressor	Flow recorded when system is on Compressor mode
			Flow Fan Error	Flow error uncertainty when system is on Fan Only mode
			Flow Compressor Error	Flow error uncertainty when system is on Compressor mode
2	Filtration_volume.xlsx		FV C	Filtration volume by cooling compressor mode
			FV F	Filtration volume by fan only mode
			FV All	All filtration volume
			FV C Error	Filtration volume error uncertainty by cooling compressor mode
			FV F Error	Filtration volume error uncertainty by fan only mode
			FV All Error	All filtration volume error uncertainty
3	mini_wras_eff.xlsx		Size	Particle size (in µm)
			Initial	Initial efficiency of the unloaded filter
			Initial Error	Initial efficiency error uncertainty of the unloaded filter
			Final	Final efficiency of the filter in service
			Final Error	Final efficiency error uncertainty of the filter in service
			Overall	Overall efficiency of the filter
			Overall_Error	Overall efficiency error uncertainty of the filter
4	psd_qff_evaluation.xlsx		Size	Similar to raw data + visit number
			Distribution metrics: min, max, mean, median, count, and stddev	min, max, mean, median, count, and stddev of five runs of dust sample
5 6	psd_ff.xlsx psd_qff.xlsx		Size	Particle size (in µm)
			Ff/qff psd	Size distribution of the channel (volume percentage)
			Ff/qff psd error	Size distribution error/uncertainty of the channel (volume percentage)
7	psd_sci_unstack.xlsx		FP	Fiter position in the SCI (A-E)
			PM Mass	Mass of particles collected on the 37mm or 25mm filter deployed in SCI.
			Error	Error associated with mass measurement of particles on 37/25mm filters
			Mass Percentage	Particle size distribution (mass based) in the SCI given range
			Size	Minimum size of the range in the SCI
			Max Size	Maximum size of the range in the SCI
8	tm_qff.xlsx		Element	Trace metal element: Pb, As, Cd, Cu, etc., (total of 12)
			Concentration	Concentration of the trace metals (in ng/m^3)
			Error	Error uncertainty in the concentration
			measure	Measurement technique: Airborne sampler (ab) or QFF
9	psd_sci_distributed.xlsx		Size	Particle size

			Mass Perc	Mass percentage in the particle size distribution measured by SCI
			FP	Fiter position in the SCI (A-E)
10	psd_above_1mic.xlsx		value	Value of PSD in the given range
			error	Error uncertainty associated with the value of PSD in the given range
			size	Minimum of particle size range in the
			Method	Measurement method: QFF or airborne
11	ab_pm_integ.xlsx		PM	PM metric: TSP, PM_{10}, $PM_{2.5}$ etc.,
12	qff_pm.xlsx		Concentration	Concentration of the PM metric
			Error	Error uncertainty of the concentration
			Measure	Measurement technique: QFF or airborne