README: The MultiWiSE Dashboard

Purpose

This dashboard uses multiyear time series of daily total (i.e., all-source) $PM_{2.5}$ to generate weekly estimates of wildfire smoke (WFS) $PM_{2.5}$ and calculate the 12 Multiyear Wildfire Smoke Exposure (MultiWiSE) metrics that characterize episodic exposure to WFS $PM_{2.5}$. The dashboard can be used to generate WFS exposure profiles and the MultiWiSE metrics for multiple identifiers (IDs), so long as a multiyear $PM_{2.5}$ time series is provided for each ID. An ID can represent a geographic location (e.g., postal or ZIP code) or an individual and is used to group the daily $PM_{2.5}$ concentrations into a single time series. For each ID, the dashboard will calculate the exposure profile and MultiWiSE metrics for the entire duration of the provided time series. Details on required input to the dashboard can be found below in 'Usage', and details on the outputs provided by the dashboard can be found in 'Outputs.' More information on the methods used to generate the MultiWiSE metrics can be found in the manuscript '*Multiyear Wildfire Smoke Exposure (MultiWiSE) metrics: A data-driven approach to characterizing episodic $PM_{2.5}$ exposures for epidemiologic research*.' **Figure 1** provides an overview of the step-by-step approach the dashboard uses to generate an WFS exposure profile and calculate the MultiWiSE metrics for a given ID.

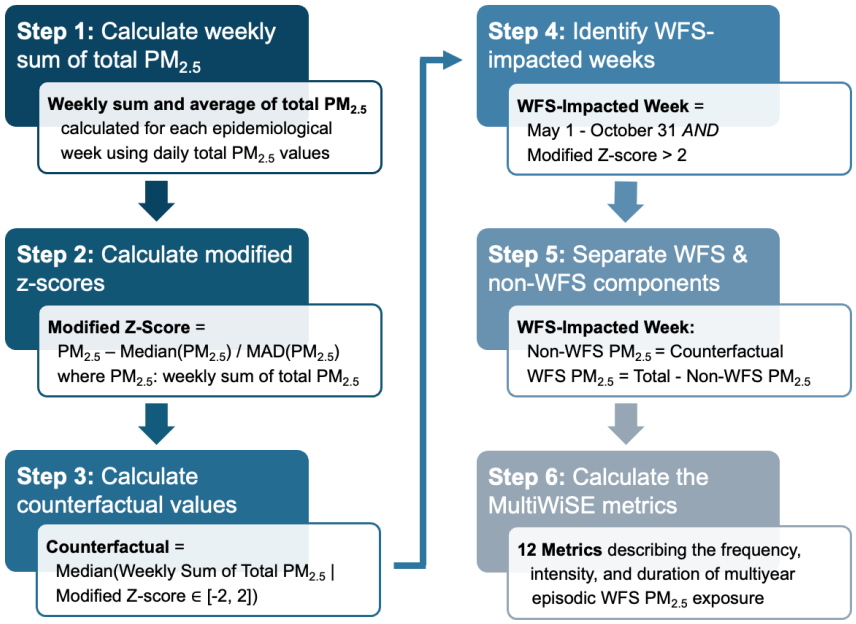


Figure 1. The step-by-step methodological approach used in the dashboard to generate estimates of weekly wildfire smoke (WFS) $PM_{2.5}$ and calculate the 12 Multiyear Wildfire Smoke Exposure (MultiWiSE) metrics. This approach is applied separately to each ID.

Usage

The dashboard relies on the user providing a multiyear time series of daily total (i.e., all-source) $PM_{2.5}$ for each ID. The time series for all IDs must be uploaded to the dashboard in a single file in .csv, .xlsx, or .xls format and include the following columns: `ID`, `Date`, `Total_PM25`. The start and end dates of the time series can be different for each ID, so long as the time series for each ID is at least 1 year in length. **Table 1** outlines the required fields and formatting of the file, and **Table 2** provides an example of what the uploaded file should look like.

In the dashboard, the user can also customize Metric 6 (Total $PM_{2.5}$ > 25 µg/m³) to use a different threshold (e.g., a regulatory value other than the 24-hour air quality objective in BC).

Table 1. Descriptions of the required variables

Variable	Type	Description
ID	Character string	The identifier that the $PM_{2.5}$ concentration corresponds to. The ID can represent a geographic region (e.g., ZIP or postal code) or an individual.
Date	Date (YYYY-MM-DD)	Date the $PM_{2.5}$ concentration corresponds to
Total_PM25	Numeric	Total $PM_{2.5}$ concentration (µg/m³)

Table 2. Example of file with multiyear time series of daily $PM_{2.5}$ for multiple IDs.

ID	Date	Total_PM25
ID_01	2012-01-01	5.6
ID_01	2012-01-02	7.2
ID_01	2012-01-03	6.3
...		
ID_01	2019-12-31	4.2
ID_02	2022-06-09	12.3
ID_02	2022-06-10	14.7
ID_02	2022-06-11	13.9
...		
ID_02	2024-11-23	6.1

For the dashboard to work as expected, the uploaded file should meet the following conditions:

- Column names must be exact: `ID, Date, Total_PM25`
- All dates must be written as YYYY-MM-DD (e.g., 2022-07-15)
- The $PM_{2.5}$ values provided in the `Total_PM25` column must represent $PM_{2.5}$ from all sources (i.e., not WFS-specific $PM_{2.5}$)
- The $PM_{2.5}$ values provided in the `Total_PM25` column must be ≥ 0 µg/m³
- One $PM_{2.5}$ concentration must be provided for each ID-Date combination (i.e., no duplicates)

- The time series for a given ID should be contiguous, with no missing dates or $PM_{2.5}$ values
- The time series for each ID must be at least 1 year in length and include at least 365 days of $PM_{2.5}$ data
- File provided must be in the following format: .csv, .xlsx, or .xls

If any of the above conditions are not met, the dashboard will either: 1) automatically handle the issue and store an associated warning message detailing how the issue was addressed, or 2) stop running entirely or skip the problematic ID(s) and store an error message detailing what the issue was. **Table 3** outlines how different issues are handled in the dashboard, and what actions can be taken to address the issue.

Table 3. Description of how different issues are handled by the dashboard, and what actions to take to resolve the issue.

Issue	How Issue is Handled	How to Resolve Issue
Incorrect file format provided	No IDs are processed. Error message is outputted to dashboard.	Ensure that the file is provided in .csv, .xls, or .xlsx format.
Incorrect column names	No IDs are processed. Error message is outputted to dashboard.	Ensure the following column names are used: `ID`, `Date`, `Total_PM25`
Incorrect date format	No IDs are processed. Error message is outputted to dashboard.	Ensure that all dates are formatted as follows: YYYY-MM-DD.
`Total_PM25` contains values < 0 g/m³	All negative values are replaced with `NA` prior to processing. Warning message is stored in log file.	Ensure that all `Total_PM25` values are ≥ 0 µg/m³.
An ID has duplicate `Date` values	Duplicated dates are removed. For each date with duplicates, `Total_PM25` is set to the mean value across all duplicates. Warning message is stored in log file.	Ensure that there are no duplicate ID-Date combinations.
Provided time series is less than 1 year in length	ID is not processed. Error message is stored in log file.	Ensure that the time series provided for each ID is at least 1 year in length.
Missing dates or $PM_{2.5}$ values	Handling is dependent on the amount and type of missingness: - ≤ 25% missingness & ≥ 365 days of data: - *All weeks have some data available*: ID is processed using available data. Warning message is stored in log file. - *Some weeks have no data available*: For weeks with no data, total and non-WFS $PM_{2.5}$ are set to the counterfactual value and WFS $PM_{2.5}$ is set 0. Warning message is stored in log file. - ≤ 25% missingness & < 365 days of data: ID is not processed. Error message is stored in log file. - > 25% missingness: ID is not processed. Error message is stored in log file.	For each ID, ensure that the time series has no more than 25% missingness and that there are at least 365 days of $PM_{2.5}$ data provided. When possible, ensure that missingness does not exist in periods ≥7 days in length.

Outputs

There are four outputs provided by the dashboard: **1)** the MultiWiSE metrics for all IDs, **2)** figures illustrating the $PM_{2.5}$ exposure profile for each ID, **3)** the weekly estimates of total, WFS, and non-WFS $PM_{2.5}$ and identification of WFS episodes for each ID, and **4)** a log file with the error and warning messages for each ID. Each output is calculated over the duration of the provided time series and described in additional detail below.

1. MultiWiSE metrics for each ID

The dashboard will generate the 12 MultiWiSE metrics for each ID (MultiWiSE_Metrics.csv), which characterize the frequency, intensity, and duration of episodic WFS exposure over the duration of the provided time series (**Figure 2**). Metric 6 will be calculated based on the user-provided threshold, and the threshold used will be reflected in the column name. In addition, the file will include values for 5 additional variables: the Cumulative Total $PM_{2.5}$ (mg/m^3), Average Total $PM_{2.5}$ ($\mu g/m^3$), Cumulative Non-WFS $PM_{2.5}$ (mg/m^3), Average Non-WFS $PM_{2.5}$ ($\mu g/m^3$), and Counterfactual Weekly $PM_{2.5}$ ($\mu g/m^3$). **Table 4** provides an example of the MultiWiSE metrics and variables produced by the dashboard.

2. Exposure profiles for each ID

The dashboard generates a series of figures for each ID, as shown in **Figure 3**, including:

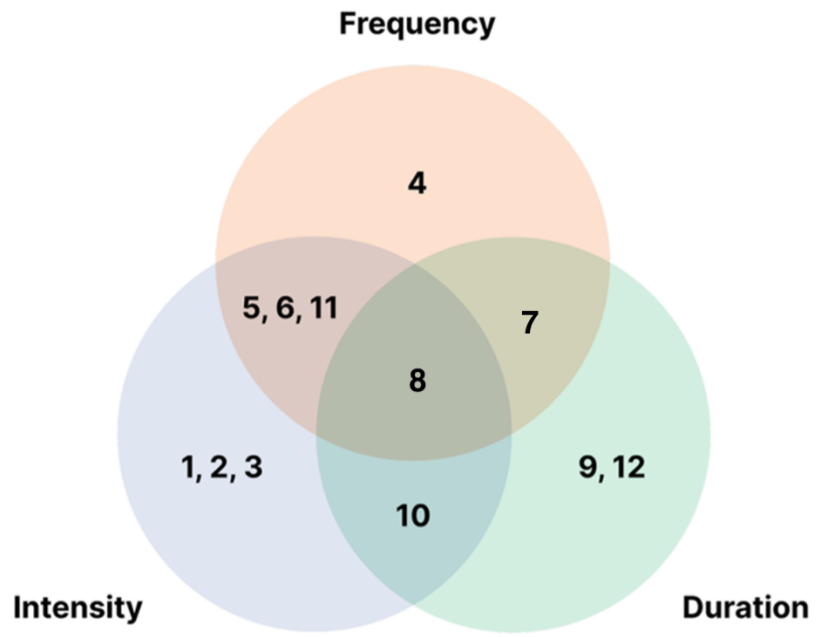
A) Cumulative $PM_{2.5}$ exposure, separated into total, wildfire smoke (WFS), and non-WFS components.
B) Weekly average total, WFS, and non-WFS $PM_{2.5}$ concentrations, with identification of WFS episodes and severe WFS episodes.
C) Distribution of the weekly sum of total $PM_{2.5}$ exposure, with indication of the range and median used to identify the counterfactual weekly value.

3. Weekly $PM_{2.5}$ estimates for each ID

The dashboard will generate ID-level estimates of weekly $PM_{2.5}$ over the duration of the provided time series (Weekly_PM25_Estimates.csv). This file includes information on weekly total, WFS, and non-WFS $PM_{2.5}$, as well as identification of WFS episodes and severe WFS episodes. These estimates are used to generate the MultiWiSE metrics and figures. A list of all variables provided can be found in **Table 5** and an example of the weekly estimates can be found in **Table 6**.

4. Log file with warning and error messages

The dashboard will generate a log file (Warning_Error_Log.csv) with all error and warning messages that occurred when processing each ID. For an ID that was processed without issue, the log file will contain a message stating “No issues.” This log file can be used to identify and resolve issues with the provided file. An example of the log file can be found in **Table 7**.



Metric	Name	Unit	Definition
1	Cumulative WFS $PM_{2.5}$	mg/m³	Total sum of WFS $PM_{2.5}$ over the multiyear period
2	WFS Fraction	%	Fraction of cumulative total $PM_{2.5}$ exposure from WFS
3	Average WFS $PM_{2.5}$	µg/m³	Average WFS $PM_{2.5}$ across all WFS-impacted weeks
4	Any WFS	# weeks	Total count of WFS-impacted weeks
5	WFS $PM_{2.5}$ > 5 µg/m³	# weeks	Count of WFS-impacted weeks where average WFS $PM_{2.5}$ exceeded 5 µg/m³
6	Total $PM_{2.5}$ > 25 µg/m³	# weeks	Count of WFS-impacted weeks where average total $PM_{2.5}$ exceeded 25 µg/m³
7	WFS Episodes	# episodes	Total count of WFS episodes
8	Severe Episodes	# episodes	Count of severe WFS episodes
9	Longest Episode	# weeks	Length of the longest WFS episode
10	Worst Episode	µg/m³	Average WFS $PM_{2.5}$ during the WFS episode with the highest average WFS $PM_{2.5}$
11	WFS from Severe Episodes	%	Fraction of cumulative WFS $PM_{2.5}$ from severe WFS episodes
12	Average Recovery	# weeks	Average time between episodes, including periods before the first and after the last episode
	Cumulative Total $PM_{2.5}$	mg/m³	Total sum of total (all source) $PM_{2.5}$ over the multiyear period
	Average Total $PM_{2.5}$	µg/m³	Average total $PM_{2.5}$ across all weeks in the exposure window
	Cumulative Non-WFS $PM_{2.5}$	mg/m³	Total sum of non-WFS $PM_{2.5}$ over the multiyear period
	Average Non-WFS $PM_{2.5}$	µg/m³	Average non-WFS $PM_{2.5}$ across all weeks in the exposure window
	Counterfactual Weekly $PM_{2.5}$	µg/m³	The counterfactual value, calculated as the median of total $PM_{2.5}$ during weeks with a modified Z-score between -2 and 2

Figure 2. The 12 MultiWiSE metrics, and 5 additional variables, which capture the frequency, intensity, and duration of wildfire smoke (WFS) exposure for each ID over the duration of the provided time series. Metric 6 uses 25 µg/m³ as the default threshold value, but this value can be customized by the user in the dashboard.

Table 4. Example of the 'MultiWiSE_Metrics.csv' file with the 12 MultiWiSE metrics and 5 additional variables for two IDs. The units and definitions of the metrics can be found in Figure 2.

ID	1_Cumulative_WFS_PM25	2_WFS_Fraction	3_Average_WFS_PM25	4_Any_WFS	5_WFS_PM25_exceeds_5	6_Total_PM25_exceeds_25	7_WFS_Episodes	8_Severe_Episodes	9_Longest_Episode	10_Worst_Episode	11_WFS_from_Severe_Episodes	12_Average_Recovery	Cumulative_Total_PM25	Average_Total_PM25	Cumulative_NonWFS_PM25	Average_NonWFS_PM25	Counterfactual_Value
ID_01	2.71	11.33	12.47	31	22	4	5	2	8	31.42	51.8	72.71	23.89	7.31	21.18	6.88	6.12
ID_02	3.24	13.79	13.92	34	25	6	6	3	9	34.85	58.6	75.24	26.13	8.06	22.89	7.18	6.42

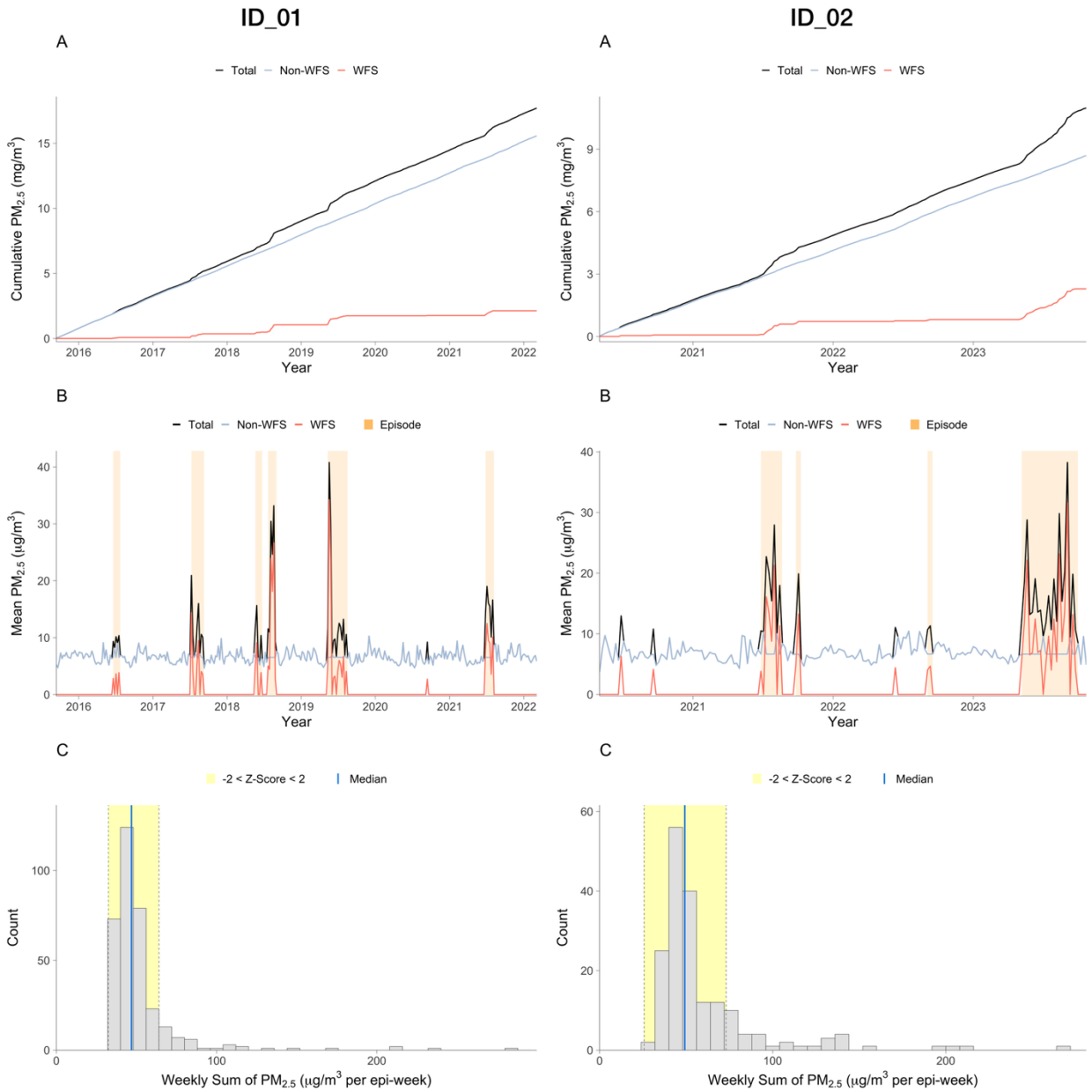


Figure 3. Example of provided figures. The (A) cumulative and (B) average weekly total, wildfire smoke (WFS), and non-WFS $PM_{2.5}$ and (C) distribution of the weekly sum of total $PM_{2.5}$ for each ID over the duration of the provided time series. The distribution plots highlight the range and median of values used to establish the counterfactual weekly value and extract the WFS and non-WFS contributions from total $PM_{2.5}$.

Table 5. Definition of variables provided in the 'Weekly_PM25_Estimates.csv' file

Variable	Unit	Description
ID	N/A	A unique identifier for a given location or individual
epiweek	N/A	The week number the exposure data corresponds to. The first week in the time series will be week 1. Weeks are represented as epidemiological weeks, which are standardized to start on a Sunday and end on a Saturday.
n_days	days	The number of days in the epiweek that had data available for analysis. A value of 7 indicates no missing data. Incomplete epiweeks (<7 days) that occur at the beginning and end of the time series are dropped and not included in analysis.
epiweek_start_date	weeks	The date on which the epiweek started (always a Sunday), shown in date format YYYY-MM-DD.
epiweek_end_date	weeks	The date on which the epiweek ended (always a Saturday), shown in date format YYYY-MM-DD.
weekly_pm25_avg	$\mu g/m^3$	The weekly average of total $PM_{2.5}$
weekly_wfs_pm25_avg	$\mu g/m^3$	The weekly average of WFS $PM_{2.5}$
weekly_non_wfs_pm25_avg	$\mu g/m^3$	The weekly average of non-WFS $PM_{2.5}$
weekly_pm25_sum	$\mu g/m^3$	The weekly sum of total $PM_{2.5}$
weekly_wfs_pm25_sum	$\mu g/m^3$	The weekly sum of WFS $PM_{2.5}$
weekly_non_wfs_pm25_sum	$\mu g/m^3$	The weekly sum of non-WFS $PM_{2.5}$
episode_id	N/A	Unique identifier for each WFS episode. All weeks in a given WFS episode will have the same identifier. A WFS episode is defined as either (1) two or more WFS-impacted weeks separated by no more than three non-WFS-impacted weeks, or (2) a single week where the sum of WFS $PM_{2.5} > 250\ mg/m^3$.
severe_episode_id	N/A	Unique identifier for each severe WFS episode. All weeks in a given severe episode will have the same identifier. A severe episode is defined as any WFS episode that includes a week with a sum of WFS $PM_{2.5} > 250\ mg/m^3$.

Table 6. Example of the 'Weekly_PM25_Estimates.csv' file with weekly $PM_{2.5}$ estimates and identification of WFS episodes for each ID.

ID	epiweek_index	num_days	epiweek_start_date	epiweek_end_date	weekly_pm25_avg	weekly_wfs_pm25_avg	weekly_non_wfs_pm25_avg	weekly_pm25_sum	weekly_wfs_pm25_sum	weekly_non_wfs_pm25_sum	episode_id	severe_episode_id
ID_01	1	7	2014-04-27	2014-05-03	6.89	0.00	6.89	20.68	0.00	20.68	0	0
ID_01	2	7	2014-05-04	2014-05-10	5.41	0.00	5.41	37.90	0.00	37.90	0	0
ID_01	3	7	2014-05-11	2014-05-17	5.51	0.00	5.51	38.56	0.00	38.56	0	0
ID_01	4	7	2014-05-18	2014-05-24	17.79	11.31	6.48	124.56	79.20	45.36	1	0
ID_01	5	7	2014-05-25	2014-05-31	20.13	13.65	6.48	140.93	95.57	67.75	1	0
…												
ID_02	1	7	2014-06-08	2014-06-14	4.17	0.00	4.17	29.16	0.00	29.16	0	0
ID_02	2	7	2014-06-15	2014-06-21	4.36	0.00	4.36	30.51	0.00	30.51	0	0
ID_02	3	7	2014-06-22	2014-06-28	5.15	0.00	5.15	36.07	0.00	36.07	0	0
ID_02	4	7	2014-06-29	2014-07-05	9.00	2.85	5.66	62.99	19.97	43.02	0	0
…												

Table 8. Example of the 'Warning_Error_Log.csv' file that logs the warning and error messages for each ID.

ID	Message
ID_01	No issues
ID_02	Error: Unable to process ID. The provided file only provides PM2.5 data for 68% of the days in the time series. ≥75% data coverage and at least 1 year of data is required for processing. Weeks with missing data can be identified using the 'n_days' column in the Weekly_PM25_Estimates.csv file.
ID_03	No issues
ID_04	Warning: The ID had 15 negative value(s) in the 'Total_PM25' column. All negative values were replaced with 0 before processing.

Support

If you require further support or have questions regarding this dashboard, please contact:

Stephanie Cleland, PhD | email: stephanie_cleland@sfu.ca
Sarah Henderson, PhD | email: sarah.henderson@bccdc.ca

Acknowledgement

This README and dashboard were the joint effort of Olivia Qiong (Research Assistant, BCCDC and VCHRI), Stephanie Cleland (Assistant Professor, SFU and VCHRI), and Sarah Henderson (Scientific Director, BCCDC and UBC). The research team acknowledges the contributions of Michael Brauer to the project.