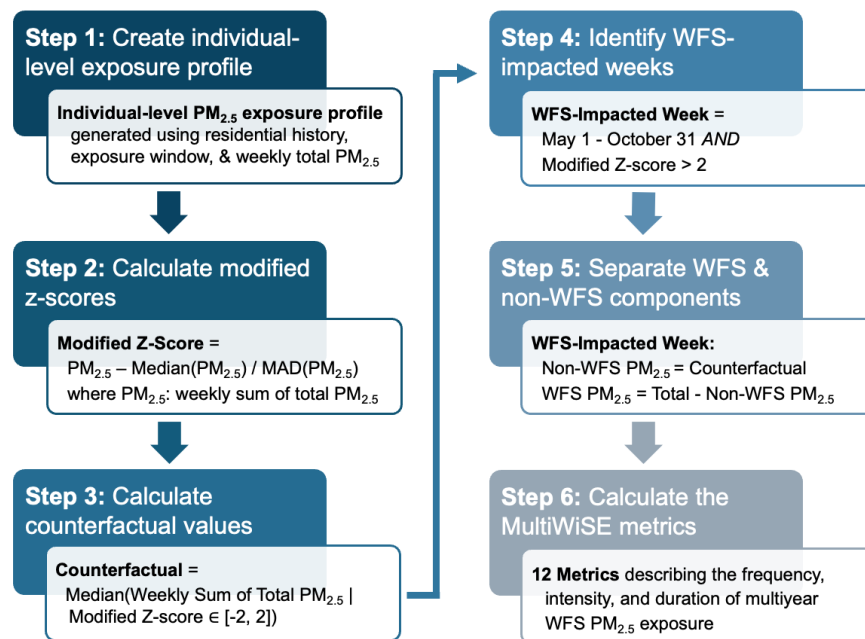


# README: Dashboard for Calculating Individual-Level Multiyear Wildfire Smoke Exposure (MultiWiSE) metrics in Canada

## Purpose

This dashboard uses information on individual-level residential histories and defined exposure windows to generate weekly estimates of wildfire smoke (WFS)  $PM_{2.5}$  and calculate the 12 Multiyear Wildfire Smoke Exposure (MultiWiSE) metrics. The MultiWiSE metrics characterize an individual's episodic exposure to WFS  $PM_{2.5}$  over the duration of the provided multiyear exposure window. The dashboard can be used to generate  $PM_{2.5}$  exposure profiles and the MultiWiSE metrics for multiple individuals residing in Canada for any period between 2010-2023. Details on the required inputs to the dashboard can be found below in 'Usage', and details on the outputs provided by the dashboard can be found below in 'Outputs.' More information on the datasets and 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 on chronic health effects.*'

**Figure 1** provides an overview of the step-by-step approach the dashboard uses to generate an individual's WFS exposure profile and calculate the MultiWiSE metrics.



**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 individual.

## Usage

This dashboard uses estimates of weekly total (i.e., all-source) PM<sub>2.5</sub> across Canada for 2010-2023 (publicly available via Dropbox - see 'Supplemental' below), in combination with user-provided residential histories and exposure windows, to generate individual-level PM<sub>2.5</sub> exposure profiles and MultiWiSE metrics.

To use the dashboard, the following information must be provided for each individual:

- The exposure window (i.e., time period over which you want to calculate the individual's exposure, which must be at least 1 year in length)
- The postal code(s) the individual lived in for the duration of the exposure window, and the date(s) they moved into and out of each postal code

This information should be provided as two separate files (.csv, .xlsx, or .xls format), one for residential history and one for exposure window. **Table 1** below outlines the required fields and formatting of the .csv files, and **Table 2** and **Table 3** provide examples of the required files:

- **File 1:** Individual-Level Residential History  
For each individual, one row per postal code they lived in during the exposure window, with information on the dates they moved into and out of each postal code. Required columns include: IndividualID, PostalCode, StartDate, EndDate.
- **File 2:** Individual-Level Exposure Window  
One row per individual with the start and end date of the exposure window of interest. This exposure window can be different for each individual. Required columns include: IndividualID, ExposureStart, and ExposureEnd

**Table 1.** Descriptions of the Required Variables

Variable	Type	Description
File 1: Individual-Level Residential History		
IndividualID	Character string	Unique identifier for individual
PostalCode	Character string	6-digit alphanumeric Canadian postal code with no spaces
StartDate	Date (YYYY-MM-DD)	Date individual moved into the postal code
EndDate	Date (YYYY-MM-DD)	Date individual moved out of the postal code
File 2: Individual-Level Exposure Window		
IndividualID	Character string	Unique identifier for individual
ExposureStart	Date (YYYY-MM-DD)	Start date of the exposure window
ExposureEnd	Date (YYYY-MM-DD)	End date of the exposure window

**Table 2.** Example file for residential histories

IndividualID	PostalCode	StartDate	EndDate
Person1	Y1A0J4	2012-01-27	2016-04-31
Person1	T8V8K9	2016-05-01	2022-11-15
Person2	C0A0A1	2010-03-15	2013-06-18
Person2	C0A0A2	2013-06-19	2018-12-31
Person2	B3M3K2	2019-01-01	2023-10-15
Person3	V0E0A3	2011-07-01	2023-07-01

**Table 3.** Example file for exposure windows

IndividualID	ExposureStart	ExposureEnd
Person1	2014-07-01	2022-07-31
Person2	2012-11-15	2023-08-01
Person3	2011-07-01	2023-07-01

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

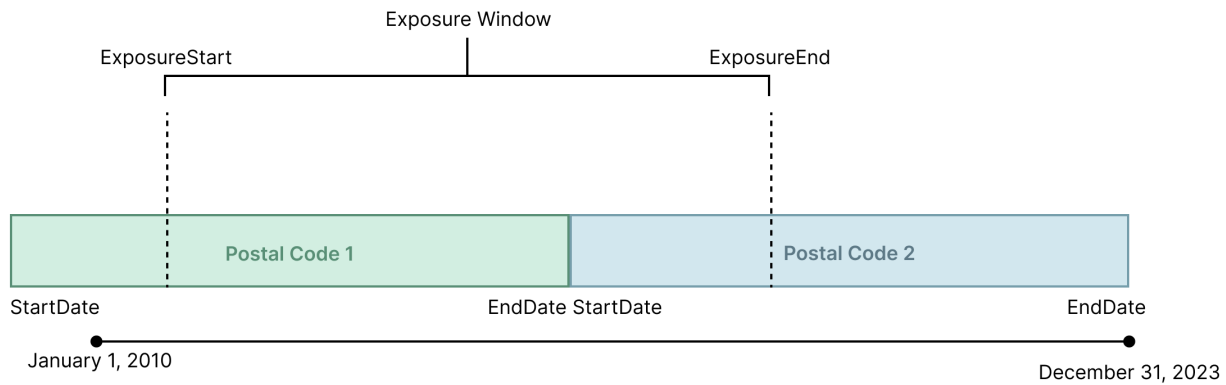
- Column names must be exact: IndividualID, PostalCode, StartDate, EndDate in the residential history file, and IndividualID, ExposureStartDate, ExposureEndDate in the exposure window file
- All dates must be written as YYYY-MM-DD (e.g. 2022-07-15)
- The provided postal codes must be valid and located in one of the 13 provinces or territories in Canada. The postal codes must be formatted to have 6 digits and no spaces (e.g., V5Z4R4)
- Each individual must appear in both the exposure window file and the residential history file
- The residential history for an individual should be contiguous (i.e., no gaps) and non-overlapping, and the start date for a given postal code must occur before the end date
- The exposure window for each individual must fall within the range of available data (January 1, 2010 – December 31, 2023), be at least 1 year in length, and the exposure window start date must occur before the end date
- The residential history for an individual must include residential information for the entirety of the exposure window
- Each individual must appear exactly once in the exposure window .csv file. Multiple exposure windows for the same IndividualID will not be processed
- There can be no blank cells
- Files provided as a .csv, .xls, or .xlsx

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 individual and store an error message detailing what the issue was. **Table 4** outlines how different issues are handled in the dashboard, and what actions can be taken to address the issue. **Figure 2** illustrates both how the individual-level residential history, exposure window, and data availability period can overlap and the importance of contiguous, non-overlapping residential history.

**Table 4.** 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 individuals are processed. Error message is outputted to dashboard.	Ensure that both the residential history and exposure window files are provided in .csv, .xls, or .xlsx format.
Incorrect column names in the residential history and/or exposure window file	No individuals are processed. Error message is outputted to dashboard.	Ensure the below column names are used: - Residential history file: IndividualID, PostalCode, StartDate, EndDate - Exposure window file: IndividualID, ExposureStartDate, ExposureEndDate.
Incorrect date format or missing date in the residential history and/or exposure window file	No individuals are processed. Error message is outputted to dashboard.	Ensure there are no missing dates and that all dates in both the residential history and exposure window files are formatted as follows: YYYY-MM-DD.
Differences in the IndividualIDs included in the residential history and exposure window files	No individuals are processed. Error message is outputted to dashboard.	Ensure that residential history and exposure window files have the same number of IndividualIDs, and that each IndividualID exists in both files.
An IndividualID is included more than once in the exposure window file	For each IndividualID, the first available exposure window is used, and all other exposure windows are removed. Individual is processed. Warning message is stored in log file.	Ensure that each IndividualID appears exactly once in the exposure window file.
Exposure window end date occurs before start date	Individual is not processed. Error message is stored in log file.	Ensure that for each exposure window, the end date occurs after the start date.
Exposure window is less than 1 year in length	Individual is not processed. Error message is stored in log file.	Ensure that each exposure window is at least 1 year in length.

Exposure window starts before or ends after the data availability period	<p>If start date occurs before 2010, set to January 1, 2010. If end date occurs after 2023, set to December 31, 2023. Handling is dependent on length of revised exposure window:</p> <ul style="list-style-type: none"> <li>- <math>\geq 1</math> year: Individual is processed. Warning message is stored in log file.</li> <li>- <math>&lt; 1</math> year: Individual is not processed. Error message is stored in log file.</li> </ul>	Ensure that the start and end dates for each exposure window occur between 2010-2023 and that the exposure window is at least 1 year in length.
Exposure window starts before or ends after the provided residential history	If the exposure window start date occurs before provided residential history, set to earliest available residential start date. If the exposure window end date occurs after the provided residential history, set to last available residential end date. Individual is processed with revised exposure window. Warning message is stored in log file.	Ensure that the residential history for an individual includes residential information for the entire exposure window.
Gap in residential history	<p>Handling is dependent on length of gap in residential history:</p> <ul style="list-style-type: none"> <li>- <math>&lt; 90</math> days: The gap is split evenly into two periods, and the first period is assigned to the pre-gap postal code and the second period is assigned to the post-gap postal code. Individual is processed. Warning message is stored in log file.</li> <li>- <math>\geq 90</math> days: Individual is not processed. Error message is stored in log file.</li> </ul>	Ensure there are no gaps in the residential history for an individual, or if there are gaps, ensure that the gaps are less than 90 days in length.
Overlap in residential history	<p>Handling is dependent on length of overlap in residential history:</p> <ul style="list-style-type: none"> <li>- <math>&lt; 90</math> days: The overlapping period is split evenly in two periods, and the first period is assigned to the first postal code, and the second period is assigned to the second postal code. Individual is processed. Warning message is stored in log file.</li> <li>- <math>\geq 90</math> days: Individual is not processed. Error message is stored in log file.</li> </ul>	Ensure that an individual's residential history is contiguous and non-overlapping, or if there are overlaps, ensure that the overlaps are less than 90 days in length.
Residential history end date occurs before start date	Individual is not processed. Error message is stored in log file.	For all postal codes in the residential history, ensure the start date occurs before the end date.
Invalid or incorrectly formatted postal code in residential history	Individual is not processed. Error message is stored in log file.	Ensure all postal codes included in the residential history file are correctly formatted (6 digits, no spaces) and located in one of the 13 provinces and territories in Canada.



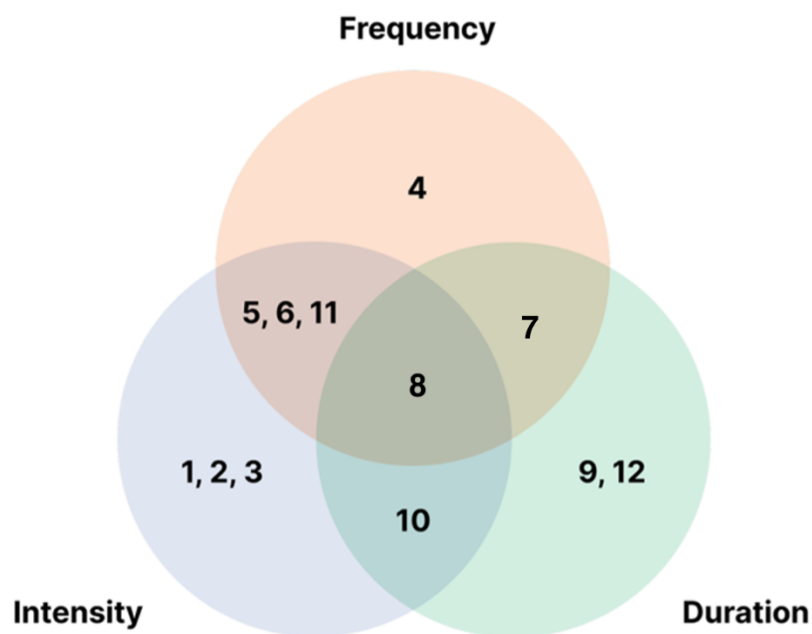
**Figure 2.** A conceptual diagram illustrating how an individual's residential history can overlap with the exposure window and the data availability period (2010-2023). No overlaps or gaps should exist between postal codes in an individual's residential history. The exposure window must fall within the available residential history and data availability period.

## Outputs

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

### 1. File of Individual-Level MultiWiSE Metrics

The dashboard will generate a .csv file (MultiWiSE\_Metrics.csv) with 12 MultiWiSE metrics for each individual, which characterize the frequency, intensity, and duration of episodic WFS exposure over the provided exposure window (**Figure 3**). 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 5** provides an example of the MultiWiSE metrics and variables produced by the dashboard.



Metric	Name	Unit	Definition
1	Cumulative WFS PM <sub>2.5</sub>	mg/m <sup>3</sup>	Total sum of WFS PM <sub>2.5</sub> over the multiyear period
2	WFS Fraction	%	Fraction of cumulative total PM <sub>2.5</sub> exposure from WFS
3	Average WFS PM <sub>2.5</sub>	µg/m <sup>3</sup>	Average WFS PM <sub>2.5</sub> across all WFS-impacted weeks
4	Any WFS	# weeks	Total count of WFS-impacted weeks
5	WFS PM <sub>2.5</sub> > 5 µg/m <sup>3</sup>	# weeks	Count of WFS-impacted weeks where average WFS PM <sub>2.5</sub> exceeded 5 µg/m <sup>3</sup>
6	Total PM <sub>2.5</sub> > 25 µg/m <sup>3</sup>	# weeks	Count of WFS-impacted weeks where average total PM <sub>2.5</sub> exceeded 25 µg/m <sup>3</sup>
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/m <sup>3</sup>	Average WFS PM <sub>2.5</sub> during the WFS episode with the highest average WFS PM <sub>2.5</sub>
11	WFS from Severe Episodes	%	Fraction of cumulative WFS PM <sub>2.5</sub> 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 <sub>2.5</sub>	mg/m <sup>3</sup>	Total sum of total (all source) PM <sub>2.5</sub> over the multiyear period
	Average Total PM <sub>2.5</sub>	µg/m <sup>3</sup>	Average total PM <sub>2.5</sub> across all weeks in the exposure window
	Cumulative Non-WFS PM <sub>2.5</sub>	mg/m <sup>3</sup>	Total sum of non-WFS PM <sub>2.5</sub> over the multiyear period
	Average Non-WFS PM <sub>2.5</sub>	µg/m <sup>3</sup>	Average non-WFS PM <sub>2.5</sub> across all weeks in the exposure window
	Counterfactual Weekly PM <sub>2.5</sub>	µg/m <sup>3</sup>	The counterfactual value, calculated as the median of total PM <sub>2.5</sub> during weeks with a modified Z-score between -2 and 2

**Figure 3.** The 12 MultiWiSE metrics, and 5 additional variables, that capture the frequency, intensity, and duration of an individual's exposure to wildfire smoke (WFS) over the duration of the provided exposure window.

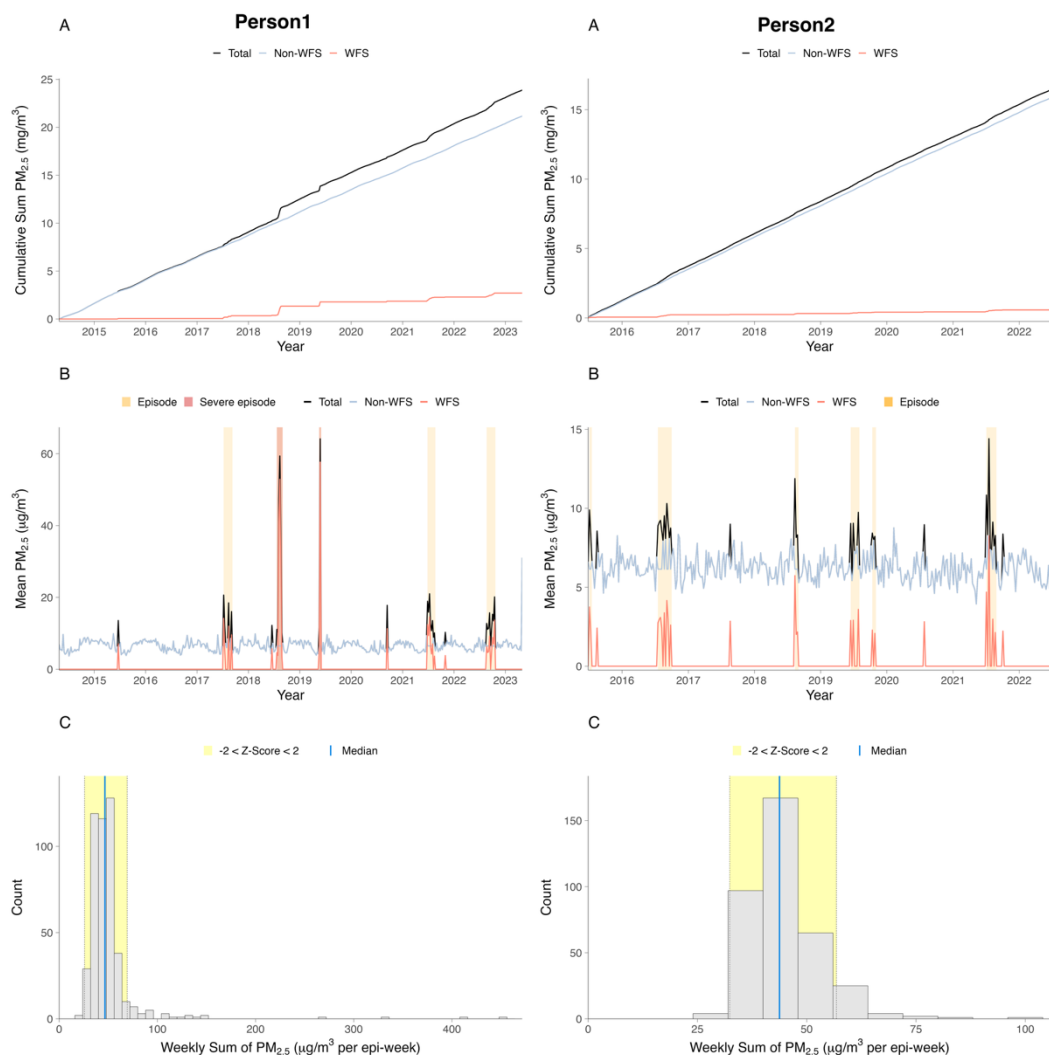
**Table 5.** Example of the 'MultiWiSE\_Metrics.csv' file with the 12 MultiWiSE metrics and 5 additional variables for two individuals. The units and definitions of the metrics can be found in Figure 3.

Individual 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
Person1	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
Person2	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

## 2. Figures of Individual-Level Exposure Profiles

The dashboard generates a series of figures for each individual, as shown in **Figure 4**, including:

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



**Figure 4.** 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 individual. 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  $PM_{2.5}$ .



### 3. File of Individual-Level Weekly PM<sub>2.5</sub> Estimates

The dashboard will generate a .csv file (Weekly\_PM25\_Estimates.csv) with individual-level data on the weekly PM<sub>2.5</sub> estimates during the defined exposure windows. This file includes information on weekly total, WFS, and non-WFS PM<sub>2.5</sub>, as well as identification of WFS episodes. These estimates are used to generate the MultiWiSE metrics and figures. A list of all variables provided can be found in **Table 6**, and an example of the outputted .csv can be found in **Table 7**.

**Table 6.** Definition of variables provided the 'Weekly\_PM25\_Estimates.csv' file

Variable	Unit	Description
IndividualID	N/A	A unique identifier for a given individual
PostalCode	N/A	6-digit alphanumeric Canadian postal code, indicating the postal code that the individual lived in during a given week. For weeks where an individual moved postal codes, the postal code with the most overlapping days is used.
epiweek	N/A	The week number the exposure data corresponds to. The first week in the exposure window 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 exposure window 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	µg/m <sup>3</sup>	The weekly average of total PM <sub>2.5</sub>
weekly_wfs_pm25_avg	µg/m <sup>3</sup>	The weekly average of WFS PM <sub>2.5</sub>
weekly_non_wfs_pm25_avg	µg/m <sup>3</sup>	The weekly average of non-WFS PM <sub>2.5</sub>
weekly_pm25_sum	µg/m <sup>3</sup>	The weekly sum of total PM <sub>2.5</sub>
weekly_wfs_pm25_sum	µg/m <sup>3</sup>	The weekly sum of WFS PM <sub>2.5</sub>
weekly_non_wfs_pm25_sum	µg/m <sup>3</sup>	The weekly sum of non-WFS PM <sub>2.5</sub>
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 <sub>2.5</sub> > 250 mg/m <sup>3</sup> .
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 <sub>2.5</sub> > 250 mg/m <sup>3</sup> .

**Table 7.** Example of the ‘Weekly\_PM25\_Estimates.csv’ file with weekly PM<sub>2.5</sub> estimates and identification of WFS episodes for each individual.

IndividualID	PostalCode	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
Person1	Y1A0J4	1	7	2014-04-27	2014-05-03	6.89	0.00	6.89	20.68	0.00	20.68	0	0
Person1	Y1A0J4	2	7	2014-05-04	2014-05-10	5.41	0.00	5.41	37.90	0.00	37.90	0	0
Person1	Y1A0J4	3	7	2014-05-11	2014-05-17	5.51	0.00	5.51	38.56	0.00	38.56	0	0
Person1	T8V8K	4	7	2014-05-18	2014-05-24	17.79	11.31	6.48	124.56	79.20	45.36	1	0
Person1	T8V8K	5	7	2014-05-25	2014-05-31	20.13	13.65	6.48	140.93	95.57	67.75	1	0
...													
Person2	Y1A0J4	1	7	2014-06-08	2014-06-14	4.17	0.00	4.17	29.16	0.00	29.16	0	0
Person2	Y1A0J4	2	7	2014-06-15	2014-06-21	4.36	0.00	4.36	30.51	0.00	30.51	0	0
Person2	C0A0A1	3	7	2014-06-22	2014-06-28	5.15	0.00	5.15	36.07	0.00	36.07	0	0
Person2	C0A0A1	4	7	2014-06-29	2014-07-05	9.00	2.85	5.66	62.99	19.97	43.02	0	0
...													

## 4. Log file with warning and error messages

The dashboard will generate a .csv file (Warning\_Error\_Log.csv) that logs all the error and warning messages that occurred when processing each individual. For an individual who 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 user-provided files. An example of the outputted .csv can be found in **Table 8**.

**Table 8.** Example of the ‘Warning\_Error\_Log.csv’ file that logs the warning and error messages for each individual.

IndividualID	Message
Person1	No issues
Person2	Error: Unable to process individual. In the residential history, the provided end date for T8V8K occurs prior to the start date. For all postal codes, please ensure the start date occurs before the end date, and that the residential history is contiguous and non-overlapping.
Person3	No issues
Person4	Warning: The provided start date of the exposure window is outside the data availability window (2010 – 2023). The exposure window start date was automatically set to January 1, 2010.

# Supplemental

## Data Source

The estimates of total (i.e., all-source) PM<sub>2.5</sub> come from the Canadian Optimized Statistical Exposure Model (CanOSSEM). CanOSSEM is a machine learning model that estimates daily average total PM<sub>2.5</sub> concentrations across all populated regions of Canada using multiple predictor variables. More information about CanOSSEM can be found here:

<https://www.sfu.ca/fhs/canossem>.

## The Data Repository

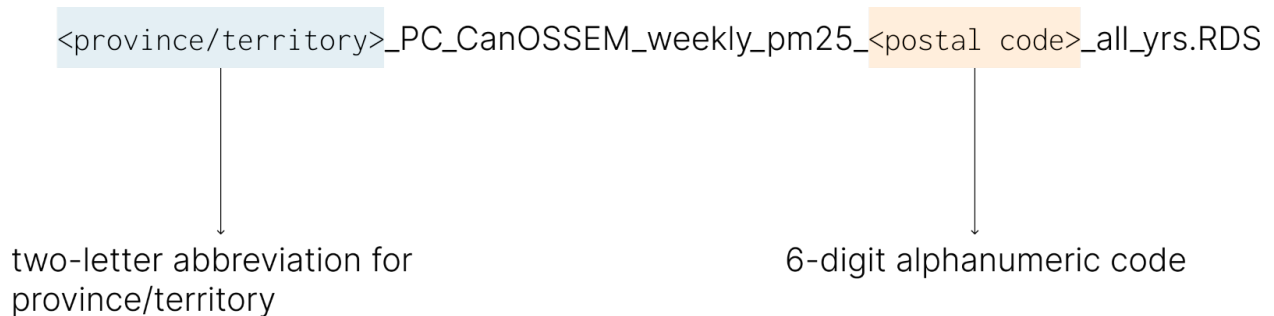
The postal code-level estimates of weekly total PM<sub>2.5</sub>, generated using the CanOSSEM daily estimates, are stored in Dropbox and are open access. However, to use the dashboard, users do **not** need to directly access the Dropbox.

### Link to data repository:

<https://www.dropbox.com/scl/fo/sli2n1whstc38p10yntio/AD6LBnflbC96LBlcRsk8hXs?rlkey=coocamx8hjb2wybkvbe13skt&e=4&st=d5s9ygtw&dl=0>

## About the Data Files in the Repository

Each file in the repository contains 14-year data (2010-2023) on the weekly averages and sums of total PM<sub>2.5</sub> concentrations. The file naming structure contains two distinguishing components: (1) the two-letter abbreviation for each province/territory and (2) the postal code:



## Support

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

Stephanie E. Cleland, PhD | email: [stephanie\\_cleland@sfu.ca](mailto:stephanie_cleland@sfu.ca)

Sarah B. Henderson, PhD | email: [sarah.henderson@bccdc.ca](mailto: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 Naman Paul and Michael Brauer to the project.