Particulate Matter Component Analyses in Relation to Public Health in Canada

A Thesis Submitted to the Committee on Graduate Studies

in Partial Fulfillment of the Requirements for the Degree of Master of

Science in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Applied Modeling and Quantitative Methods M.Sc. Program

January 2023

Abstract

Particulate Matter Component Analyses in

Relation to Public Health in Canada

Shannon Jarvis

This thesis explores the shot-term relationship between exposure to ambient air pollution and human health through metrics such as mortality and hospitalization in Canada. We begin by detailing the organization and interpolation of air pollution data from its partially quality-controlled source form. Analyses of seasonal, regional and temporal trends of all major components of $PM_{2.5}$, was performed, showing a seasonal variation across most regions and validating the dataset.

A one-pollutant statistical Generalized Additive Model was applied to the data, estimating the health risk associated with exposure to thirteen different components of $PM_{2.5}$. The selected components were based on those that compromised the majority of the mass and included: sulphate, nitrate, zinc, silicon, iron, nickel, vanadium, potassium, organic carbon, organic matter, elemental carbon, total carbon. Trends based on annual estimates of the association for $PM_{2.5}$, and its constituents, were compared, showing that carbonaceous compounds, sulphate and nitrate had similar estimates of association. Many estimates, as is common in population ecologic epidemiology, had association estimates statistically indistinguishable from zero, but with clear features of interest, including evident differences between cold and warm

season associations in Canada's temperate climate.

A method to model two correlated pollutants (in this case, $PM_{2.5}$ and O_3) was developed using thin plate splines. In this approach, the location of the response surface (after accounting for the temperature, a smooth function of time and day of week) that corresponds to the average pollutant concentration and the average plus one unit was used as the estimate of the joint contribution of pollutants due to a unit increase. The estimates from the thin plate spline (TPS) approach were compared to the single pollutant models, with large increases and decreases in $PM_{2.5}$ and O_3 being captured in the TPS estimates. However, this approach indicated significantly larger error in the estimates than would be expected, indicating a possible future area for refinement.

Acknowledgements

To Wesley Burr: for your support, ideas, encouragement, knowledge, time, and so much more. You are an exceptional supervisor and have helped me grow, both academically and personally, these past two years.

The research discussed in Chapters 2 and 3 was funded, in part, by Health Canada (under grant #859104), and with the additional direction of my co-supervisor, Dr. Hwashin Hyun Shin. Thank you for your support, access to the detailed data necessary for this work, and your guidance.

To my family; for your unconditional support and instilling a love of learning.

To Zander; for cheering me on every step of the way.

Table of Contents

Abstract **ii**

Acknowledgements **iv**

1 Introduction **1**

2 Monitoring Ambient Air Pollution in Canada **3**

2.1 Air Pollution Data 3

2.2 Continuous Data 5

2.2.1 Cleaning of Continuous Data 7

2.2.1.1 Step 1: Checking the Raw Data 7

2.2.1.2 Step 2: Interpolate Missing Hourly Data 8

2.2.1.3 Step 3: Aggregate Hourly Data for Each Census Division 9

2.2.1.4 Step 4: Interpolate Missing Daily Data 9

2.2.2 Data Issues 10

2.2.2.1 Negative Pollutant Concentrations 10

2.2.2.2 Interpolation Between Large Gaps 18

2.2.3 Analysis of Continuous Data 19

2.2.3.1 O3 Data 20

2.2.3.2 NO2 Data 22

2.2.3.3 SO2 Data 24

2.2.3.4 PM2.5 Data 25

2.2.3.5 PM10 Data 26
2.3 Integrated Data 27
2.3.1 Cleaning of Integrated Data 29
2.3.1.1 Step 1: Retrieving the Data 29
2.3.1.2 Step 2: Addressing Inconsistencies in Compound Naming 32
2.3.1.3 Step 3: Creating the Database 33
2.3.1.4 Step 4: Data Verification 33
2.3.1.5 Step 5: Compound Calculations 34
2.3.2 Analysis of Integrated Data 35
2.3.2.1 PM2.5 Mass Concentrations 37
2.3.2.2 Reconstructed PM2.5 Mass 41
2.3.2.3 Monthly Variations in the Major PM2.5 Components and Vapour Phase Species 46
2.3.2.4 Discussion and Conclusions 51
2.4 Temperature Data 54

3 Modelling Adverse Health Effects of Air Pollution 55

3.1 Literature Review 55
3.2 Generalized Additive Model 60
3.3 PM Component Models 62
3.3.1 Lagging variables 64
3.3.1.1 Discrete Lags 65
3.3.1.2 Synthetic Lags 65
3.4 Particulate Matter Risk Estimates 66
3.4.1 Exploring GAM Parameters for PM2.5 Risk Estimates 66
3.4.2 PM Component Risk Estimates 73

3.4.2.1 PM Constituents 73
3.4.2.2 Risk Estimates for Toronto - CD 3520 76
3.4.2.3 Risk Estimates by Census Divisions 79
3.5 Discussion and Conclusions 81

4 Modelling Adverse Health Effects of Two Pollutants 84
4.1 Literature Review 84
4.2 Multiple Pollutant Model 86
4.2.1 Brief Introduction to Thin Plate Splines (TPS) 87
4.2.2 The TPS Bivariate Model 87
4.3 Comparison of Single and Two Pollutant Models 89
4.4 Discussion 91

5 Conclusions 92
5.1 Future Work 94

Bibliography 96

Appendix A:
Data Summaries 110

Appendix B:
Ozone Risk Estimates 121

Appendix C:
Bivariate Risk Estimates 123

List of Figures

2.1 Particulate matter size comparison [18]. 4

2.2 Locations of NAPS stations. Stations represented by red stars are locations that are highlighted in subsequent analyses, while locations in grey are not analyzed in this work. 5

2.3 Count of negative raw continuous pollutant concentrations by province, organized from west to east. 12

2.4 Count of negative daily interpolated continuous pollutant metric concentrations by province, organized from west to east. 13

2.5 Hourly CATNAPS interpolated SO_2 concentration for NAPS stations in Lambton, ON from 1990-2019. Highlighted points (in maroon) are negative values. 16

2.6 Daily average SO_2 concentration for Lambton, ON after CATNAPS interpolation from 1990-2019. Highlighted points (in maroon) are negative values. 17

2.7 Daily average SO_2 concentration for Lambton, ON after CATNAPS interpolation, from June 2016 to December 2016. Highlighted points (in maroon) are negative values. 17

2.8 Example of error when interpolating large gaps of data using plots of mean 24-hour PM_{10} concentrations for CD 2466, Montreal QC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 19

2.9 Plots of mean 8-hour maximum O_3 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. . 21
2.10 Plots of mean 24-hour O_3 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 22
2.11 Plots of mean 24-hour NO_2 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 23
2.12 Plots of mean 24-hour SO_2 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 24
2.13 Plots of mean 24-hour $PM_{2.5}$ concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 26
2.14 Plots of mean 24-hour PM_{10} concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations. 27
2.15 Schematic detailing the organization and file naming of NAPS raw data 29
2.16 Snapshot of contents of speciation file for station 60427, Toronto ON, `2009_S60427_IC.XLS` 30
2.17 Snapshot of contents of speciation file for station 60427, Toronto ON, `S60427_PM25_2010.xlsx` 32
2.18 Example of an abnormal measured value (in red) for pollutant concentrations measured in Division No. 11 (Edmonton), AB for 2018. (a)Nitrate Concentrations. (b) $PM_{2.5}$ Concentrations. 34

2.19 Map of Canada with location of NAPS speciation stations indicated in red 36
2.20 Timeline of data availability at NAPS stations for speciation (grey), dichotomous (light blue) and manual (dark blue) air samplers from 2003–2019. 37
2.21 Measured $PM_{2.5}$ monthly average mass in μg m^{-3}, by location, organized west to east from 2010–2019 38
2.22 Monthly mean $PM_{2.5}$ concentration in μg m^{-3}, with the error bars representing 90% percentiles. 39
2.23 Daily $PM_{2.5}$ concentration in μg m^{-3}, for January with outliers ($> 3\sigma$) indicated in red 40
2.24 Mean organic carbon concentration, in μg m^{-3}, by cartridge and site 41
2.25 Reconstructed $PM_{2.5}$ mass from 2010–2019 by mean compound for the cold season (October - March, left) and the warm season (right, April - September). 42
2.26 Reconstructed $PM_{2.5}$ mass from 2003 – 2009 by mean compound for the cold season (October - March, left) and the warm season (right, April - September). 44
2.27 Reconstructed $PM_{2.5}$ mass by mean compound for the 10 highest mass concentration days for the cold season (October - March, left) and the warm season (right, April - September) from 2010–2019. 45
2.28 Ammonium sulphate and ammonium nitrate concentrations, in μg m^{-3}, by site and month (median and interquartile range). Note that the scale for Fraser Valley is an order of magnitude larger than the others. 47
2.29 Elemental carbon (EC) and organic matter (OM) concentration, in μg m^{-3}, by site and month (median and interquartile range) 48

2.30 Soil and sodium chloride concentration, in μg m^{-3}, by site and month (median and interquartile range) 49
2.31 Ammonia mixing ratio, in ppb, by site and month (median and interquartile range) 50
2.32 Sulphur dioxide mixing ratio, in ppb, and nitric acid concentration, in μg m^{-3}, by site and month (median and interquartile range). 50

3.1 Annual **mortality** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year. No temperature lag is used. 67
3.2 Annual **morbidity** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year. No temperature lag is used. 68
3.3 Annual **mortality** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for four lags on the pollutant data (0 days; 1 day; 2 days; and synthetic) 69
3.4 Annual **morbidity** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for four lags on the pollutant data (0 days; 1 day; 2 days; and synthetic) 70

3.5 Annual health risk estimates, by cause, for $PM_{2.5}$ in Toronto, ON, for three temperature lags (0 days; 1 day; and 2 days) for all year synthetically lagged pollutant concentrations. 71

3.6 Proportional composition of components in $PM_{2.5}$ mass by season and census division. 74

3.7 Correlation coefficient between $PM_{2.5}$ and its components, by season, for CD's organized from east to west. Note that the lines are visual aids, and do not signify series. 76

3.8 Annual scaled model estimates for census division 3520, Toronto, for response variable pulmonary and cardiac **mortality**. 77

3.9 Annual scaled model estimates for census division 3520, Toronto, for response variable pulmonary and cardiac **morbidity**. 78

3.10 Annual scaled mortality risk estimates for $PM_{2.5}$ and components, across five census divisions, for three seasonal breakdowns (all months; cold season; and warm season). 80

3.11 Annual scaled morbidity risk estimate for $PM_{2.5}$ and components, across five census divisions, for three seasonal breakdowns (all months; cold season; and warm season). 81

4.1 Daily ozone (top), $PM_{2.5}$ (middle) and temperature (bottom) measurements for Toronto, ON census division. 86

4.2 Thin Plate Spline residual mortality and cross-sectional plots at the average pollutant concentrations for Toronto, ON throughout 2014. Each of the single pollutant plots shows the surface and the prediction interval along that pollutant's axis, with the other pollutant fixed at its arithmetic overall average value. The response visualized here is the residual log mortality after accounting for temperature, DOW and the smooth function of time. 88

4.3 Comparison of single and multi-pollutant morbidity models for synthetic and discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON 89

4.4 Comparison of single and multi-pollutant mortality models for discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON 90

4.5 Comparison of single and multi-pollutant morbidity models for discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON 90

B.1 Annual mortality (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year 121

B.2 Annual mortality (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for three lags on the pollutant data (o days; 1 day; and 2 days) 122

B.3 Annual morbidity (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for three lags on the pollutant data (o days; 1 day; and 2 days) 122

C.1 Comparison of single and multi pollutant morbidity models for synthetic and discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC 123

C.2 Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC 124

C.3 Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC 124

C.4 Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Edmonton, AB 125

C.5 Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Edmonton, AB 125

C.6 Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Ottawa, ON 126

C.7 Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Ottawa, ON 126

List of Tables

2.1 Count of negative hourly raw pollutant measurements by census division, sorted by decreasing counts from 1980 – 2019. The table was filtered to show CDs with more than 100 negative values. 11

2.2 Count of negative pollutant concentrations at each step by census division, sorted by location (east to west). Note that these counts are prior to any correction-for-negatives being applied. 14

2.3 Percentage of non-missing daily ozone concentrations for CD 5917, Capital, BC. 20

2.4 Percentage of non-missing daily mean nitrogen dioxide concentrations for CD 5917, Capital BC 23

2.5 Percentage of non-missing daily mean sulphur dioxide concentrations for CD 5917, Capital BC 24

2.6 Percentage of non-missing daily mean $PM_{2.5}$ concentrations for CD 5917, Capital BC 25

2.7 Percentage of non-missing daily mean PM_{10} concentrations for CD 5917, Capital BC 26

2.8 Summary of calculated compounds and components. 35

2.9 List of NAPS stations and corresponding CD used. 36

3.1 Model parameters used for one pollutant component models. 64

3.2 All-year (x 10^{-3}) risk estimates for $PM_{2.5}$, in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) with a SLP smooth function of time using 12 degrees of freedom per year and varying pollutant and temperature lags. 72

3.3 Average pollutant proportions across all CD's in the study by season. Table is sorted by highest all year proportion. 75

3.4 All-year combined scaled model estimates ($\times 10^{-3}$) for census division 3520, Toronto, with a Discrete Prolate Spheroidal Sequence with 6 degrees of freedom per year for the smooth function of time. Components are sorted with $PM_{2.5}$ first followed by components compromising the highest proportion of $PM_{2.5}$ mass. Lower and Upper bounds are for 95% confidence intervals on the coefficient estimates. 79

A.1 Count of negative pollutant concentrations by pollutant and census divisions. Missing values indicate a count of zero. 117

A.2 Mapping from census division to NAPS stations 120

1. *Introduction*

Air pollution is a significant public health issue with evidenced adverse human health effects due to both short and long term exposure [30, 35, 11, 31, 82, 37]. Epidemiological studies have shown evidence of cardiovascular, respiratory, and general physiological effects, with links to other chronic conditions such as atherosclerosis, cognitive function, and diabetes [62, 12, 40, 46, 87, 31, 36, 56]. This indicates that air pollution is a major public health concern with far-reaching effects on human health.

The World Health Organization has set global air pollution standards for pollutant concentrations with the aim of improving air quality and reducing the health burden [56]. In addition to these standards, many countries have established regional guidelines to reflect the short and long term health effects of pollutant exposure. In Canada, the Canadian Ambient Air Quality Standards (CAAQS) sets air quality guidelines for maximum yearly or daily pollutant concentrations, with thresholds currently being decreased every 5 years.

The National Air Pollution Surveillance (NAPS) Program was established in Canada to collect and monitor air pollution. This program enables evaluation of long-term air pollution trends and assessment of health, environmental and economic impacts. Accordingly, it has helped inform public health guidelines and establish regulatory standards. It has also allowed initiatives, such as the Canadian Ambient Air Quality Standards (CAAQS), the Canadian Environmental Sustainability Indicators (CESI), the Air Health Trend Indicator (AHTI) and the Air Health Quality Index

(AHQI).

In this work, we used hourly or daily air pollutant data collected by the NAPS Program to model adverse health effects of such pollutants. We began by cleaning and organizing the data, discussed in Chapter 2. We then looked at seasonal, regional and temporal differences in air pollution concentrations across Canada. In Chapter 3, we used a previously developed one-pollutant model [91, 16], which estimates the health risk associated with exposure to a single pollutant, and applied this model to the integrated data organized and cleaned in the previous chapter. Then, in Chapter 4, we present a bivariate model developed using a thin plate spline to estimate the health effects of two air pollutants simultaneously, while accounting for their correlations. Finally, in Chapter 5, we conclude with a discussion of the implications of this work, and future areas of possible research.

2. *Monitoring Ambient Air Pollution in Canada*

Scripts created for the work in this chapter to download, organize and prepare pollutant data, as well as the data itself, are publicly available [47, 48].

2.1 Air Pollution Data

The National Air Pollution Surveillance (NAPS) Program is a national system of air sampling sites managed by Environment and Climate Change Canada [70]. The monitoring sites are equipped with particulate detection air monitors, which obtain samples of the ambient air. Samples are collected by drawing air through an inlet, and using either filters to capture particles, or directly distributing the sample to an analyzer. Of the pollutants measured, particulate matter of maximum diameter 2.5 micrometers ($PM_{2.5}$), has been found to be highly significant for health effects studies [54]. For context, the diameter of $PM_{2.5}$ particles are approximately 20 times smaller than the diameter of a strand of hair [101].

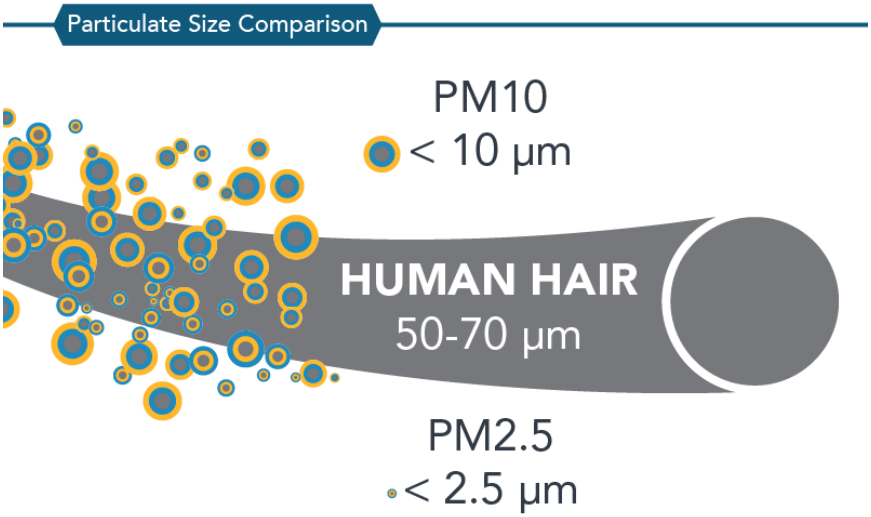


Figure 2.1: Particulate matter size comparison [18].

The NAPS program was established in 1969 and data is available for public access, published on a yearly basis [70]. The NAPS database is accessible through the Environment and Climate Change Canada open data portal [70]. As of 2021, the network consists of 277 active stations, spanning across 130 census divisions (CDs). Data is organized by pollutant and/or station as comma separated values (CSV) files via the indicated data portal.

There are two primary types of monitoring equipment used by the NAPS program; continuous (averages measured over an hour) and integrated (averages measured over 24 hours). The type is distinguished based on the equipment and collection time frames used. As such, monitoring sites vary in equipment, and contain one or more sampling units. NAPS provides metadata about each station, including: its location; population density (urban, rural); elevation; time zone; which stations they may have combined with; and collection start/end dates. Stations can move locations (combine with an existing or a new station) and begin or end monitoring at any point. Typically,

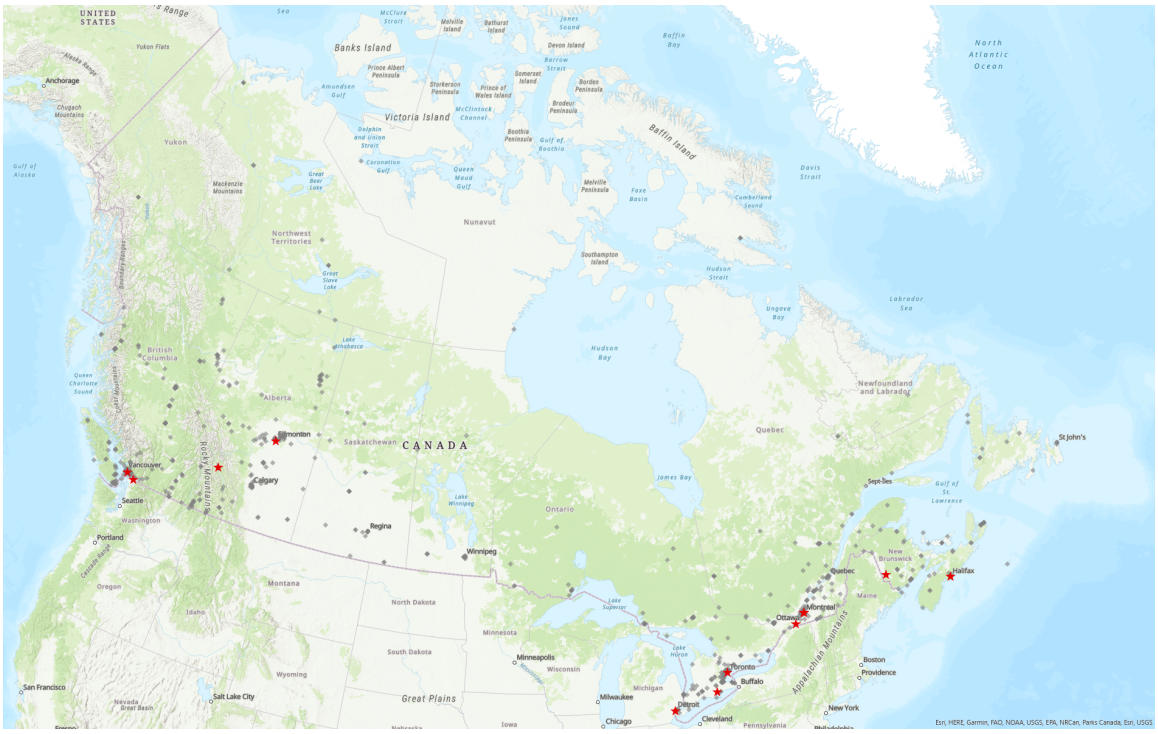


Figure 2.2: Locations of NAPS stations. Stations represented by red stars are locations that are highlighted in subsequent analyses, while locations in grey are not analyzed in this work.

when stations combine, it is due to an existing station closing, and another station at a new location beginning collection.

2.2 Continuous Data

Continuous data is measured at most monitoring sites using continuous monitoring equipment, generally gas and particulate monitors. However, the exact type of monitor is not identified in the raw data. The monitors use both analyzers and filters to directly measure hourly concentrations (in $\mu g \cdot m^{-3}$, ppb or ppm) for eight pollutants;

- carbon monoxide (CO)
- nitrogen dioxide (NO_2)

- nitric oxide (NO)
- nitrogen oxides (NO_X)
- ozone (O_3)
- sulphur dioxide (SO_2)
- particulate matter less than or equal to 2.5 micrometres ($PM_{2.5}$)
- particulate matter less than or equal to 10 micrometres (PM_{10})

Measurements require minimal human involvement as the continuous monitors can autonomously collect, log and distribute data. Quality assurance, in accordance with the NAPS guidelines [71], is done on the raw data to ensure it meets predefined standards. Data is then uploaded to the Environment and Climate Change Canada open data portal on a yearly basis by pollutant. That is, the measured hourly concentrations for each pollutant, across all stations, in a given year are contained in one file. The yearly pollutant files contain one row for all hourly measurements within a single day, for each NAPS station.

Data from 1980 – 2019 was downloaded from the open data portal and organized into a database using `R`, a statistical computing software language. The continuous data was well-structured and consistent across the time period. The raw data contained both missing and negative pollutant concentrations, with missing values being indicated by -999.

The continuous data was used to update a non-public `R` package and database, called `AHItools`, created by Dr. Wesley Burr. The aim was to expand the number of CD's analyzed from 24 (used in the previous version of `AHItools`) to 53, a significant expansion. The additional CDs were selected based on the population and availability of air pollution data. The `AHItools` database was also expanded temporally to include data up to 2019.

2.2.1 Cleaning of Continuous Data

As previously stated, the continuous data was used to update the existing `AHItools` database, which consists of daily concentrations by the CD level. Accordingly, the raw continuous data requires cleaning and transformation, with the following main objectives:

- interpolate missing data per station; this is done both at the hourly and daily level;
- aggregate multiple stations to CD-level daily concentrations.

We will begin by discussing checks done on the raw data.

2.2.1.1 Step 1: Checking the Raw Data

Checks were done when parsing each yearly pollutant file. This was to ensure the file structure and contents were valid and as expected. The checks included:

- cross-verification of latitude and longitude of stations with the station metadata;
- ensuring all measurements within each yearly pollutant file were exclusively for that pollutant and year;
- ensuring no duplicate measurements for each date and station;
- checking that the number of station observations was equal to the number of days in the year,

In addition to checks on the files, negative pollutant concentrations, which are not physically possible, were replaced with missing values. A detailed discussion on the presence of negative values throughout each cleaning step is expanded upon in Section 2.2.2.

After parsing the raw data and executing checks, summary statistics, including the daily minimum, maximum and mean were calculated. The summaries were plotted and checked for unusual data points, which signal possible issues in the data.

2.2.1.2 Step 2: Interpolate Missing Hourly Data

Interpolation of the hourly data was performed on gaps of ten hours or less. To facilitate this, another non-public `R` package, `CATNAPS` (created by David Riegert, Dr. Aaron Springford and Dr. Wesley Burr) was used. `CATNAPS` stands for: Checking and Tuning National Air Pollution Surveillance (NAPS) Data. It uses `SQLite`, a flatfile database engine embedded within `R` to initialize, access and store the NAPS data.

Minimal modifications were made to the `CATNAPS` package, but included adding the ability to read pollutant CSV files (previously required fixed width files), updates of the pollutant metadata to include additional pollutants and updates to the station metadata to include additional CDs. In accordance with the objective of aggregating the station-level NAPS data to CD level, the package requires a metadata file mapping NAPS station (by NAPS ID) to a Canadian CD. All stations not mapped to a CD are disregarded in the subsequent processes. Additionally, the package required initialization of the range of possible dates (start date and end date) and a list of pollutants.

The process begins by creating an `SQL` database and initializing the aforementioned metadata, that is: the mapping from NAPS station to CD; date range; and pollutants. Then, the database is populated by the pollutant data from Step 1. The files are required to be organized into directories by pollutant, and within each pollutant directory, the yearly CSV files should be present. When populating the database, the NAPS metadata (i.e., station name, location, status, ...) and pollutant metadata (name, code, units, ...) are associated with the corresponding station and pollutant

concentrations.

Then, `CATNAPS` data screening is used to compute statistics and flag measurements to be replaced with missing values. The following screening steps are implemented and values failing the conditions are flagged for interpolation:

- Ten or more zeroes in a row
- Baseline shift; shifts in daily minima concentrations
- Truncated daily maxima values (saturation); more daily maxima concentrations than expected
- More than 3 observations with the same daily maxima concentration in a row
- Outlier zeroes; zeroes whose concentration is more than 20 units away from the daily median concentration

After screening, the interpolation of univariate time series pollutant data was performed on the flagged (or missing values) with gaps less than 10 hours. Any negative values due to the interpolation were replaced by zeroes, due to the physicality constraint.

2.2.1.3 Step 3: Aggregate Hourly Data for Each Census Division

The `CATNAPS` hourly interpolated measurements at the NAPS station-level (Step 2) are aggregated to daily metrics by CD. The metrics included a 24-hour daily mean (calculated for all pollutants) and a daily maximum 8 hour mean (calculated for ozone).

2.2.1.4 Step 4: Interpolate Missing Daily Data

A daily univariate interpolation, using the `tsinterp` R package developed by Dr. Wesley Burr [13], for all missing values (between the first and last non-missing value) was

applied to the aggregated `CATNAPS` gap-filled metrics. A significance clip of 0.95 and time difference of 1 day (86400 seconds) was used.

Some negative values were induced by the interpolation procedure, and these were replaced by zeroes where possible (as they are predicted to have below-threshold values), or by missing where not (gaps which are too long to effectively fill). At the end of this step, the final database is produced, consisting of daily pollutant metrics by CD.

2.2.2 Data Issues

The following section details issues identified during the cleaning and interpolation procedures.

2.2.2.1 Negative Pollutant Concentrations

Throughout the raw data and interpolation process, negative pollutant concentrations were present. Table 2.1 shows the counts of negative hourly pollutant measurements by CD from the raw data. The table was filtered to show all CD's with more than 100 negative values. In total, there were 72 CDs and pollutant combinations with at least 1 negative value. Of those, there were 41 unique CDs.

Census Division (ID)	No. NAPS stations	Pollutant	No. negative hourly measurements
Haldimand-Norfolk, ON (3528)	2	SO_2	7607
Lambton, ON (3538)	5	SO_2	7145
Essex, ON (3537)	6	SO_2	6460
Greater Sudbury, ON (3553)	6	SO_2	3409
Hamilton, ON (3525)	9	SO_2	3182
Ottawa, ON (3506)	4	SO_2	2888
Hamilton, ON (3525)	9	$PM_{2.5}$	783
Toronto, ON (3520)	23	SO_2	614
Toronto, ON (3520)	23	$PM_{2.5}$	207
Division No. 11, MB (4611)	7	NO_2	177
Haldimand-Norfolk, ON (3528)	2	NO_2	175
York, ON (3519)	2	$PM_{2.5}$	173
Peterborough, ON (3515)	2	$PM_{2.5}$	142

Table 2.1: Count of negative hourly raw pollutant measurements by census division, sorted by decreasing counts from 1980 – 2019. The table was filtered to show CDs with more than 100 negative values.

Negative measurements were observed for all pollutants (NO_2, O_3, PM_{10}, $PM_{2.5}$ and SO_2), but were predominantly observed for SO_2, contributing to 70% of the negative values. Additionally, the majority of negatives were measured in Ontario (ON), seen in Figure 2.3. This may be a consequence of the data collection in Ontario, as NAPS relies upon collection and cooperation from provincial and regional government networks.

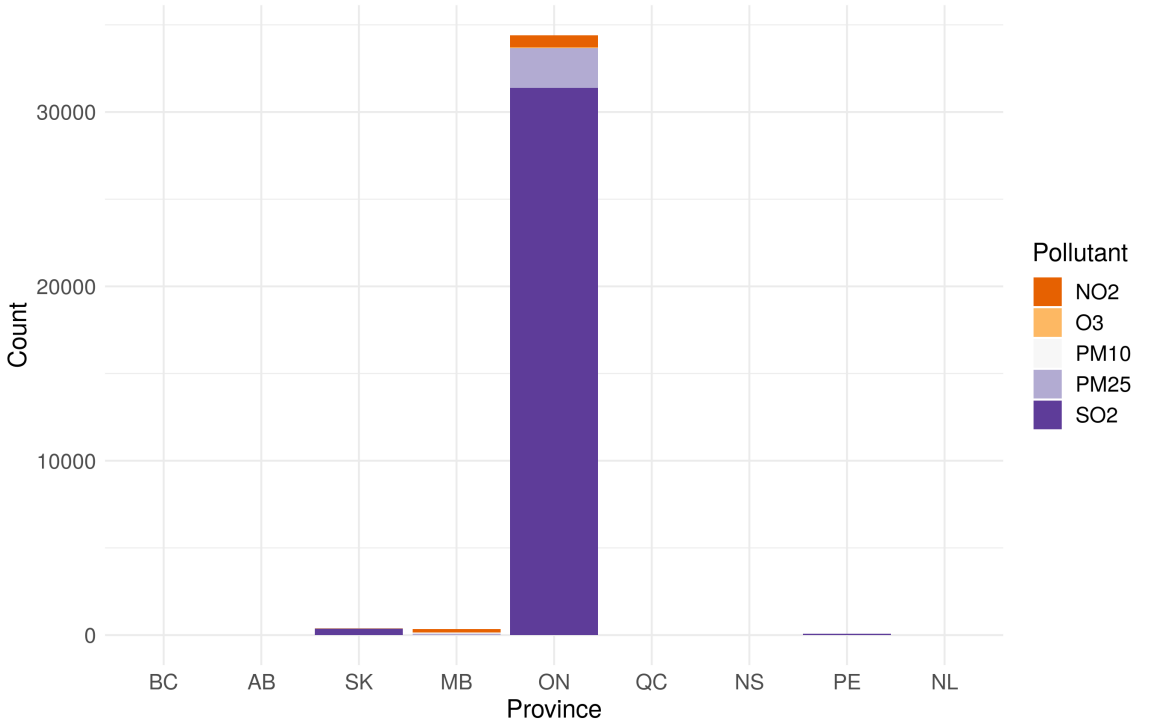


Figure 2.3: Count of negative raw continuous pollutant concentrations by province, organized from west to east.

To mitigate this issue (as negatives are physically impossible), negative raw pollutant measurements were replaced with missing values. To reiterate, after the removal of negative pollutant measurements, the count of negative values for all CDs (Table 2.1) was zero.

This removal of measured negative concentrations did not prevent negative values from arising in the interpolation steps. Table 2.2 shows the count of negative concentrations at each interpolation step. Negative values that occurred after each interpolation step were replaced with zeroes. There are two cases from which negative pollutant concentrations are produced: from the `CATNAPS` hourly interpolation step (step 2) and from the daily univariate interpolation step (step 4).

Figure 2.4 shows the count of negative daily pollutant concentrations by province,

after the final daily univariate interpolation procedure. Recall that all the negative pollutants were replaced by missing values/zeros before each subsequent step. Thus, before beginning the daily interpolation, there were no negative data points. We can see that SO_2 still comprises the majority of negative values, although the total amount of such negative values has been significantly reduced. However, it is no longer just Ontario which contains the majority of negatives. The presence of negatives for SO_2 can likely be explained by its low baseline concentration across all provinces. When the baseline is around 0, it is more likely that negative values will be estimated, particularly if there is a slight downward tend in concentrations. We will explore this further using data from the Lambton, ON CD.

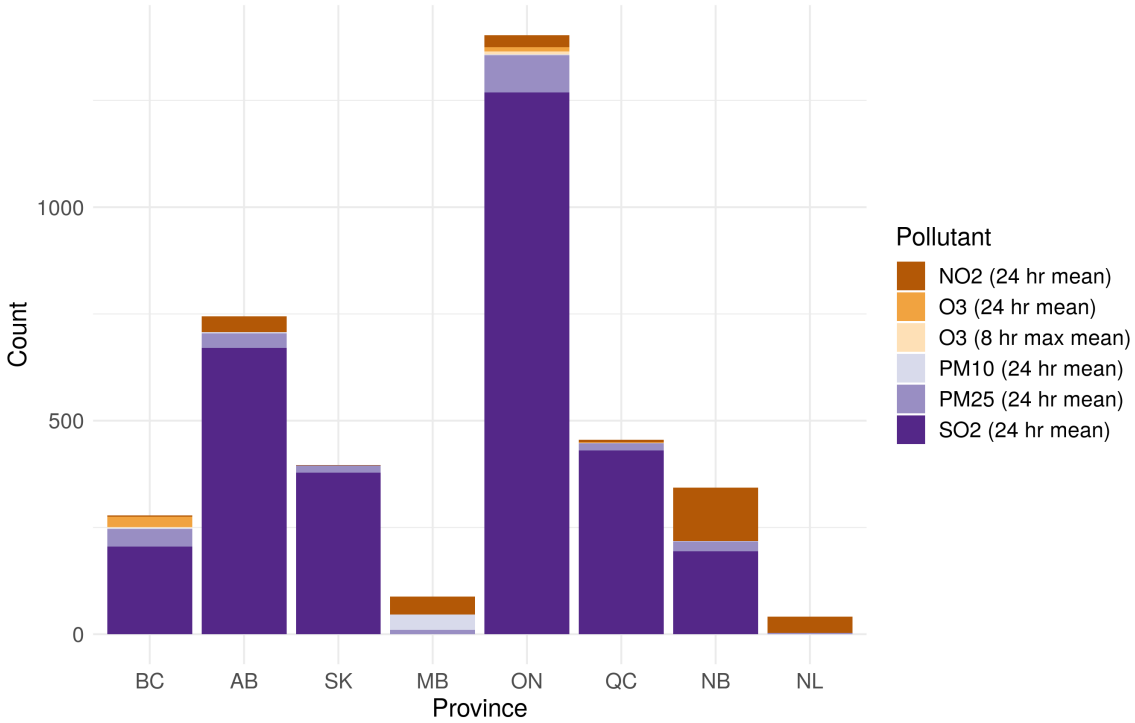


Figure 2.4: Count of negative daily interpolated continuous pollutant metric concentrations by province, organized from west to east.

Census Division (ID)	No. NAPS stations	Date range	Pollutant	No. negative hourly measurements	No. negative CATNAPS NAPS station hourly concentrations	No. negative CATNAPS CD daily concentrations	No. negative daily interpolated concentrations
Halifax, NS (1209)	10	1980:2019	NO_2	5	57	1	0
Saint John, NB (1301)	9	1996-11-20:2004-05-30	PM_{10}	0	47	1	1
Saint John, NB (1301)	9	1980-01-27:2019	SO_2	0	395	1	194
Longueuil, QC (2458)	4	1982-01-03:2019	SO_2	1	13	0	391
Haldimand-Norfolk, ON (3528)	2	1980:2019	NO_2	175	12	0	8
Haldimand-Norfolk, ON (3528)	2	2001-01-01:2019	SO_2	7607	42	1	82
Essex, ON (3537)	6	1980:2019	O_3	21	114	0	7
Essex, ON (3537)	6	1980:2019	SO_2	6460	403	2	0
Essex, ON (3537)	6	1980:2019	O_3	21	114	0	7
Lambton, ON (3538)	5	1980:2019	SO_2	7145	782	13	0
Greater Sudbury, ON (3553)	6	1980:2019	SO_2	3409	366	7	0
Division No. 7, MB (4607)	3	1993-12-10:2018-12-18	NO_2	17	30	0	36
Division No. 7, MB (4607)	3	1997-10-08:2018-12-31	PM_{10}	13	102	1	19
Division No. 11, MB (4611)	7	1993-06-01:2018-12-31	PM_{10}	54	16	0	17
Division No. 11, MB (4611)	7	1980:2018-12-31	NO_2	177	69	0	6
Division No. 16, AB (4816)	15	1986-07-03:2019	NO_2	0	462	1	2

Table 2.2: Count of negative pollutant concentrations at each step by census division, sorted by location (east to west). Note that these counts are prior to any correction-for-negatives being applied.

A complete table with counts of negative values for all CDs are included in Appendix A.

Example: SO_2 Concentrations in Lambton, ON (CD 3538)

We will highlight the SO_2 concentrations in Lambton, ON. This census division is situated on the south-east bank of Lake Huron, bordering the United States and close to Detroit, Michigan. There are five NAPS stations within the region, three of which we have SO_2 data for.

In this exploration, we will look at the number of negative concentrations, the number of missing values and the baseline pollutant concentration. This CD was chosen as it has a relatively large amount of negative concentrations due to `CATNAPS` interpolation. Figure 2.5 shows the hourly `CATNAPS` interpolated SO_2 concentration for each NAPS station. Negative values are highlighted in maroon. In total, there are 782 negative values amongst the stations. Figure 2.6 shows the corresponding daily average concentration, with 19 negative data points highlighted.

From these analyses, it was observed that stations with a significant portion of missing values and with a pollutant concentration that is closer to 0 has a greater chance of producing negative concentrations due to interpolation. Recall that the `CATNAPS` interpolation is at the hourly station level, which are aggregated to daily CD level metrics. Since multiple stations are located in one CD, a single negative hourly concentration at a NAPS station doesn't have a large influence on bringing down the daily mean when aggregating. However, if the majority of stations and/or hourly measurements are negative, the daily CD concentration will be much smaller. That is to say, for a daily CD concentration to be negative, there must be a heavy influence of negative hourly concentrations at each station.

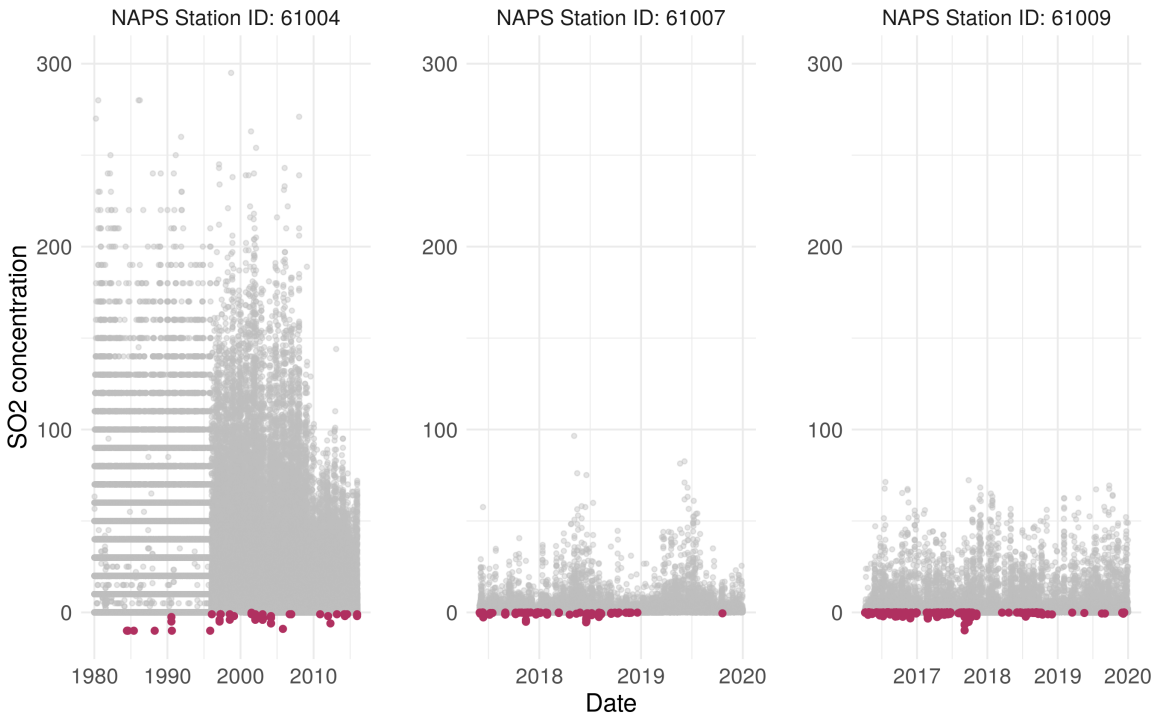


Figure 2.5: Hourly CATNAPS interpolated SO_2 concentration for NAPS stations in Lambton, ON from 1990-2019. Highlighted points (in maroon) are negative values.

Figure 2.6 aggregated the hourly station-level data from Figure 2.5 to a 24-hour mean measurement. The majority of daily negative pollutants (in maroon) occur in 2016 and 2017. Figure 2.7 shows the daily data for a 6-month period from June 2016 to December 2016.

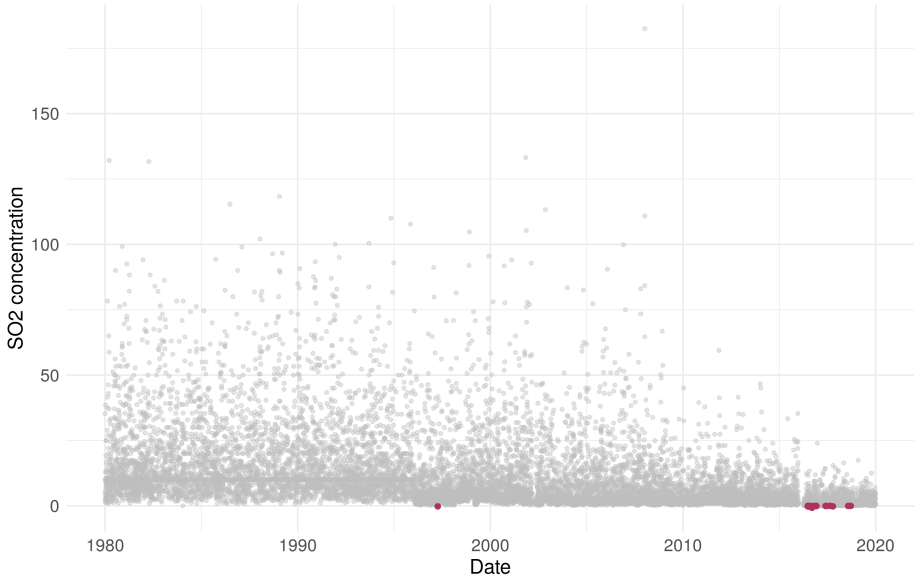


Figure 2.6: Daily average SO_2 concentration for Lambton, ON after CATNAPS interpolation from 1990-2019. Highlighted points (in maroon) are negative values.

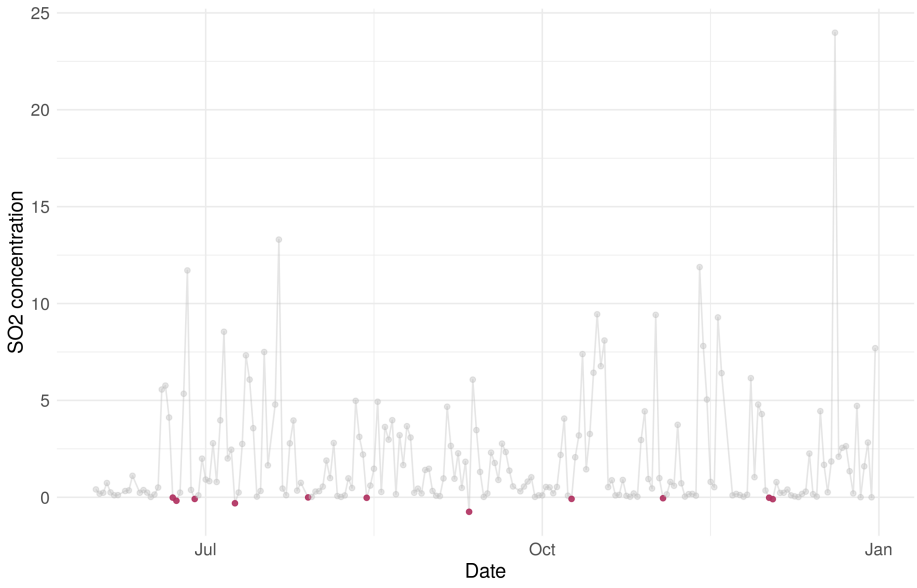


Figure 2.7: Daily average SO_2 concentration for Lambton, ON after CATNAPS interpolation, from June 2016 to December 2016. Highlighted points (in maroon) are negative values.

The negative values occurred near concentrations with a baseline around 0. Notice the decreasing values on the days before the gap, and non-zero after: this is functionally a "valley" shape, so the interpolation's best trend guess is that the slope continued down to a minimum, then climbed back upwards. Obviously nonsense, but again: interpolation is not magic, it's just mathematical and statistical guessing.

2.2.2.2 Interpolation Between Large Gaps

Additional issues could arise when interpolating gaps larger than 6 months. As previous reports on the interpolation methods have indicated, anything more than $4-6$ months is simply impossible to interpolate without secondary data sources, and even then, the values should not be taken to be all that accurate. There's no structure to follow through the gaps!

This phenomena was observed for four CDs across two pollutants: SO_2 (CDs: 3529 - Brant ON; 3548 - Nipissing ON; and 4611 - Division No. 11 MB) and PM_{10} (2466 - Montreal QC). Figure 2.8 shows this interpolation issue for CD 2466, Montreal QC. In this case, we clearly should simply not attempt to interpolate these missing values: this is the better part of a decade missing, and should just be left out of the data.

Thus, in all of these cases where the algorithm inadvertently interpolated something that really should have been left out, we have removed these imputed points, and added a logical check to prevent interpolations for longer than 180 days total time, for the daily series.

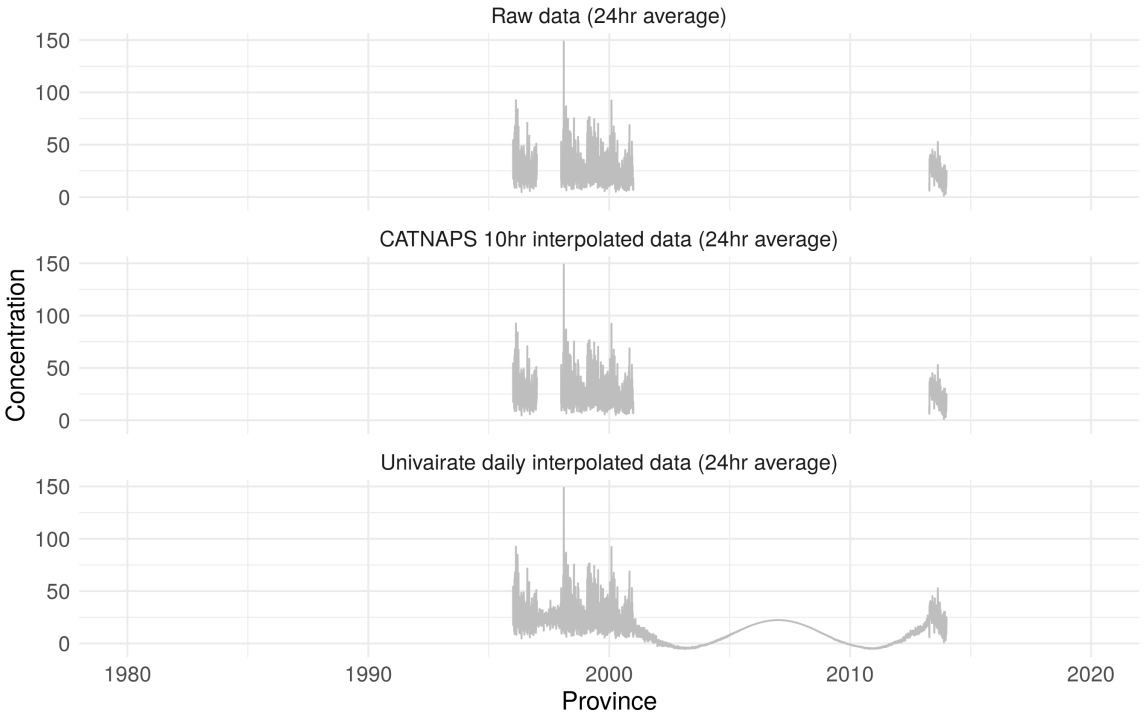


Figure 2.8: Example of error when interpolating large gaps of data using plots of mean 24-hour PM_{10} concentrations for CD 2466, Montreal QC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

2.2.3 Analysis of Continuous Data

For the analyses, we will show data summaries for Capital, a census division in BC. The Capital, BC census division is composed of 13 NAPS stations. It was selected to showcase as it displays the effect of both the `CATNAPS` and univariate interpolation on the number of missing data points. A series of six plots and corresponding table of percentage of missing data points will be shown for each pollutant. We will begin by examining ozone, O_3.

2.2.3.1 O3 Data

The 8-hour maximum mean and 24-hour mean ozone concentrations are shown in Figures 2.9 and 2.10, respectively. Each plot is divided into three segments to show the pollutant concentration at every interpolation step. In the top plot, the mean daily O_3 concentrations for all NAPS stations within Capital, BC is presented. The middle plot, `CATNAPS` 10-hour interpolated data, shows the daily mean concentration after applying the `CATNAPS` flagging and interpolation. The bottom plot shows the daily interpolated O_3 concentration. Gaps present in the `CATNAPS` plot were interpolated to fill missing data points between the first and last measurement. Table 2.3 summarizes the percent of non-missing data points at each step.

Ozone exhibits seasonal cycles, with maxima in the winter/spring and minima in the summer/fall.

Metric	Available Years	% Non-Missing Raw data	% Non-Missing CATNAPS 10hr Interpolated Data	% Non-Missing Daily Interpolated Data
O_3 24 hr mean	1980-2019	88.87	87.85	93.21
O_3 8 hr max	1980-2019	88.63	87.85	93.22

Table 2.3: Percentage of non-missing daily ozone concentrations for CD 5917, Capital, BC.

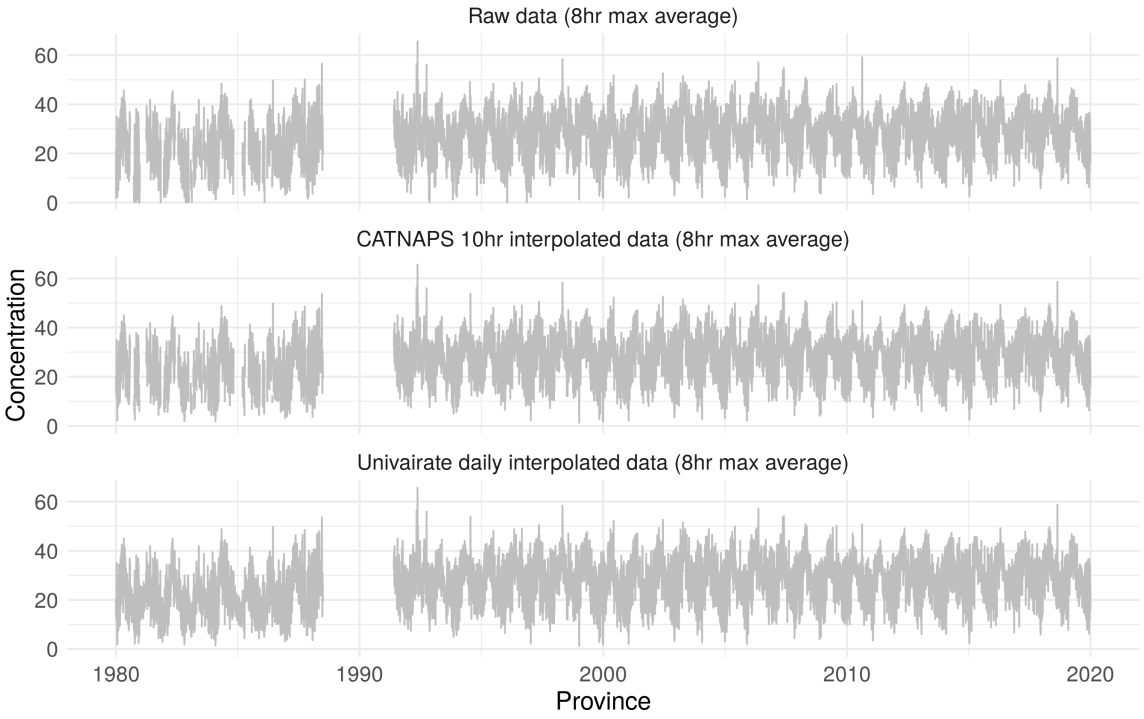


Figure 2.9: Plots of mean 8-hour maximum O_3 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

From Table 2.3, it can be seen that the percentage of non-missing data points slightly decreased between the raw data and the `CATNAPS` interpolation. That is, the `CATNAPS` flagging and interpolation introduced additional missing values in the data. If we look at the counts of negative values (Appendix A), there were no negative daily `CATNAPS` concentrations. Thus, the increase in missing data points is a consequence of: (1) negative hourly concentrations being interpolated (there were 173 negative hourly interpolated concentrations); and (2) concentrations being flagged by `CATNAPS`, but not being interpolated.

The univariate daily interpolation decreased the number of missing data points, as expected. The interpolation can be seen in Figures 2.10 and 2.9, particularly from 1980 – 1987. There is a large gap in data from 1998 – 1991, which was not

interpolated, as it is greater than 6 months. Three negative data points were incurred in the interpolation, and replaced by zeros.

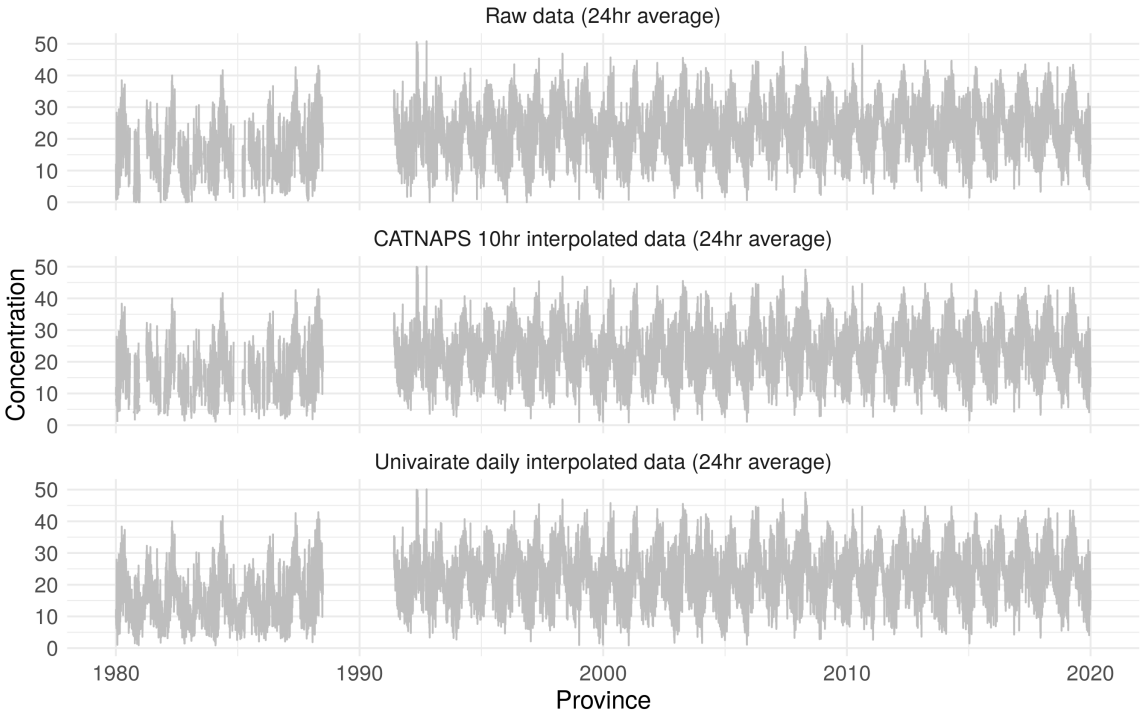


Figure 2.10: Plots of mean 24-hour O_3 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

2.2.3.2 NO2 Data

In the NO_2 data, there are a large amount of gaps that are larger than 6 months apart. These gaps can be seen throughout all the panes in Figure 2.11.

The concentrations are highly variable and don't exhibit seasonal trends.

Available Years	% Non-Missing Raw data	% Non-Missing CATNAPS 10hr Interpolated Data	% Non-Missing Daily Interpolated Data
1981-2019	76.66	76.32	80.61

Table 2.4: Percentage of non-missing daily mean nitrogen dioxide concentrations for CD 5917, Capital BC

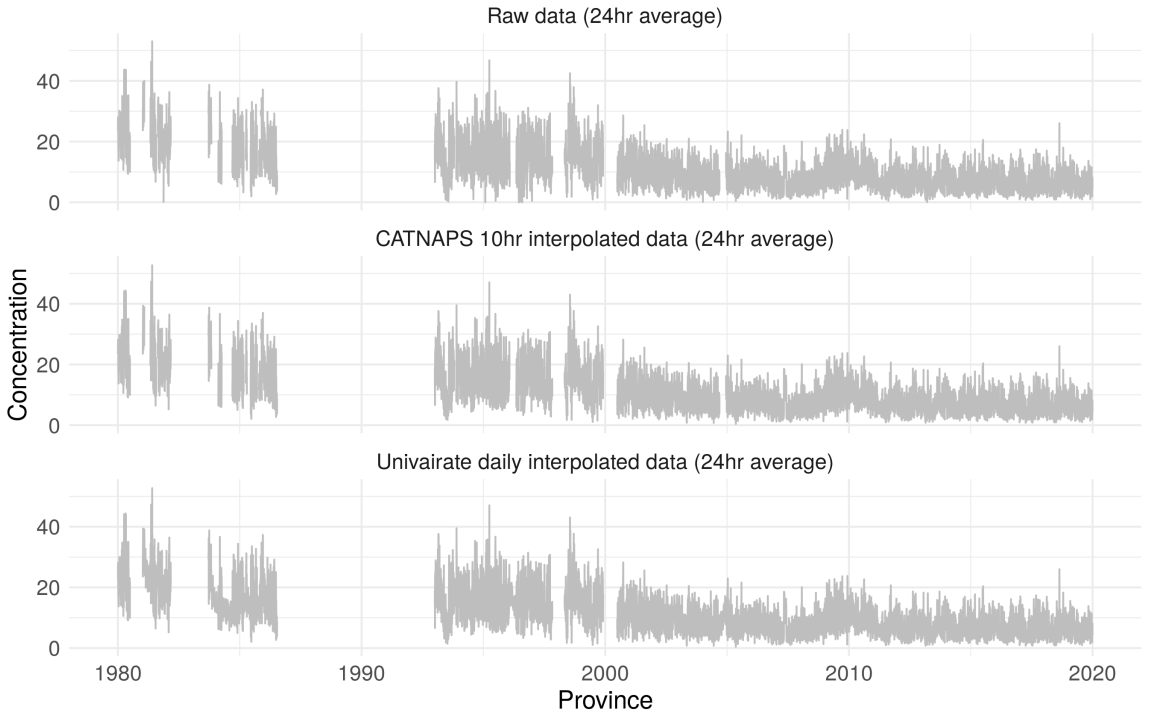


Figure 2.11: Plots of mean 24-hour NO_2 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

2.2.3.3 SO2 Data

Available Years	% Non-Missing Raw data	% Non-Missing CATNAPS 10hr Interpolated Data	% Non-Missing Daily Interpolated Data
1981-2019	87.5	70.96	95.14

Table 2.5: Percentage of non-missing daily mean sulphur dioxide concentrations for CD 5917, Capital BC

In the `CATNAPS` flagging and interpolation, 17% of the data points were removed. This is particularly evident around 1995, visualized in the middle pane of Figure 2.12. In the third pane, the points were interpolated and substantially reduced the amount of small gaps.

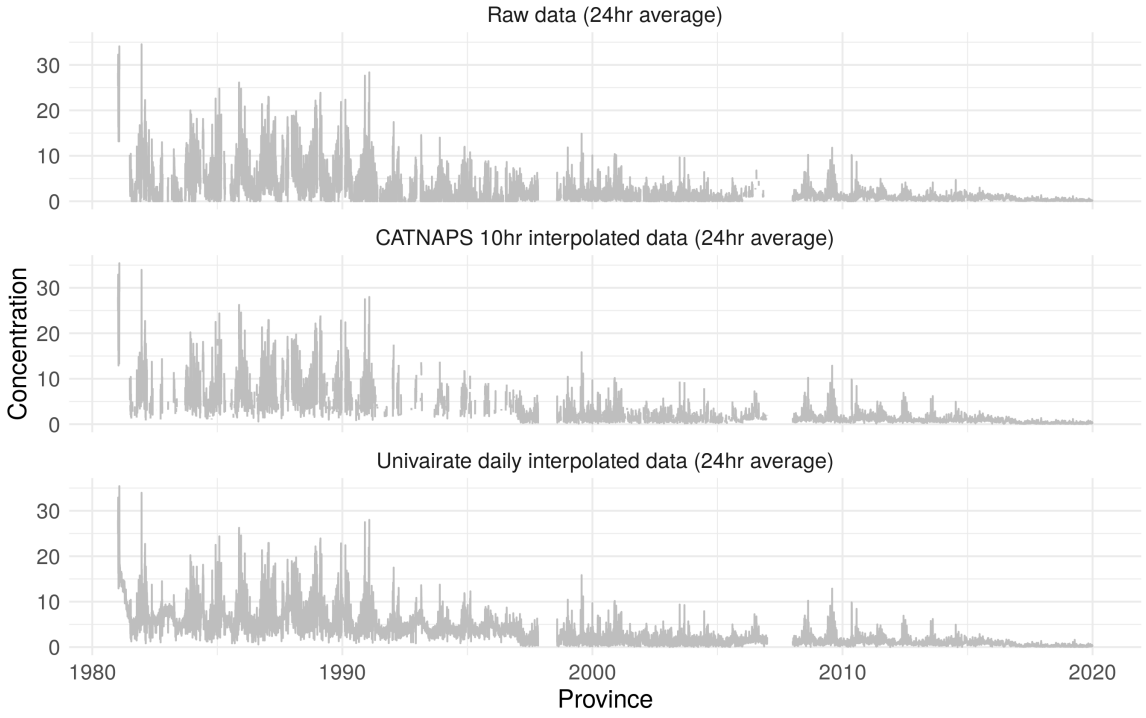


Figure 2.12: Plots of mean 24-hour SO_2 concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

SO_2 concentrations show seasonal variation. Reflecting on previous discussion of negative values in the data, which predominately occur for SO_2, we can see how the "wiggliness" with decreases and increases in concentration would generate negative concentrations in the interpolation steps.

2.2.3.4 PM2.5 Data

$PM_{2.5}$ shows seasonal (annual) cycle, but is otherwise fairly regular - high variance in late winter and early spring, lower variable in summer and autumn. It also shows a few days of very high $PM_{2.5}$ concentrations, due to forest fires throughout British Columbia in 2018.

Available Years	% Non-Missing Raw data	% Non-Missing CATNAPS 10hr Interpolated Data	% Non-Missing Daily Interpolated Data
1998/05/02 – 2019	98.53	98.53	99.99

Table 2.6: Percentage of non-missing daily mean $PM_{2.5}$ concentrations for CD 5917, Capital BC

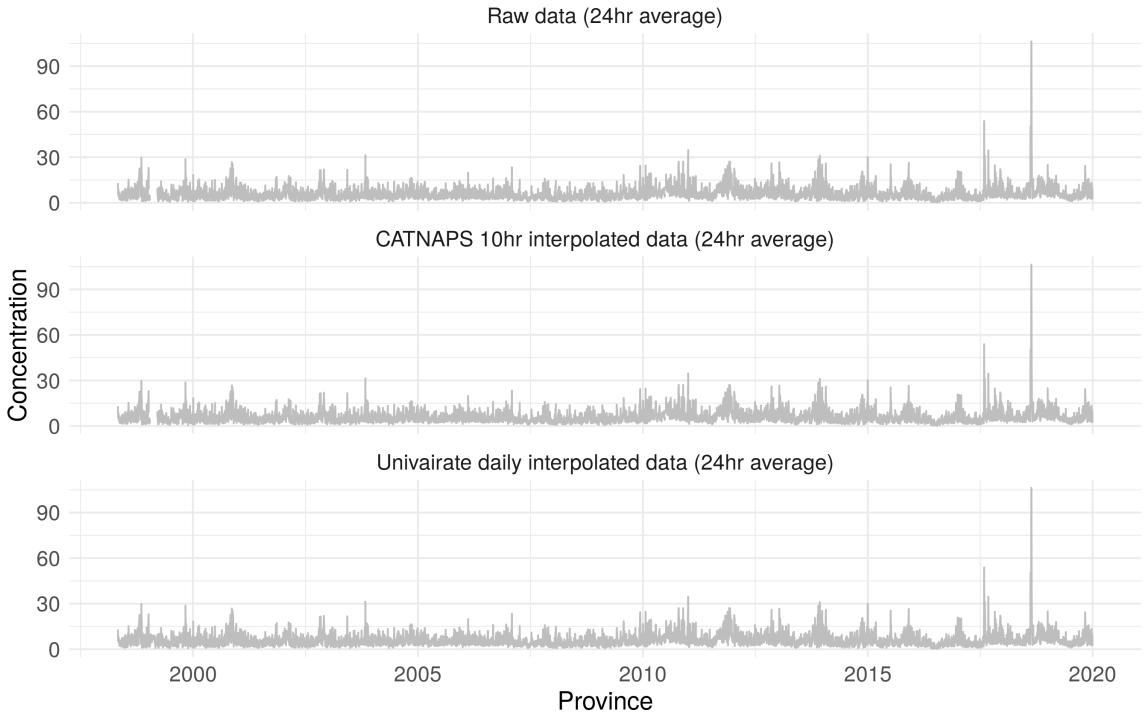


Figure 2.13: Plots of mean 24-hour $PM_{2.5}$ concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

The number of missing data points were consistent through the interpolation steps.

2.2.3.5 PM10 Data

The PM_{10} data is irregular, with some seasonal variation. The percentage of missing data points is consistent throughout the three steps.

Available Years	% Non-Missing Raw data	% Non-Missing CATNAPS 10hr Interpolated Data	% Non-Missing Daily Interpolated Data
1998/05/02 – 2019	84.33	84.33	84.96

Table 2.7: Percentage of non-missing daily mean PM_{10} concentrations for CD 5917, Capital BC

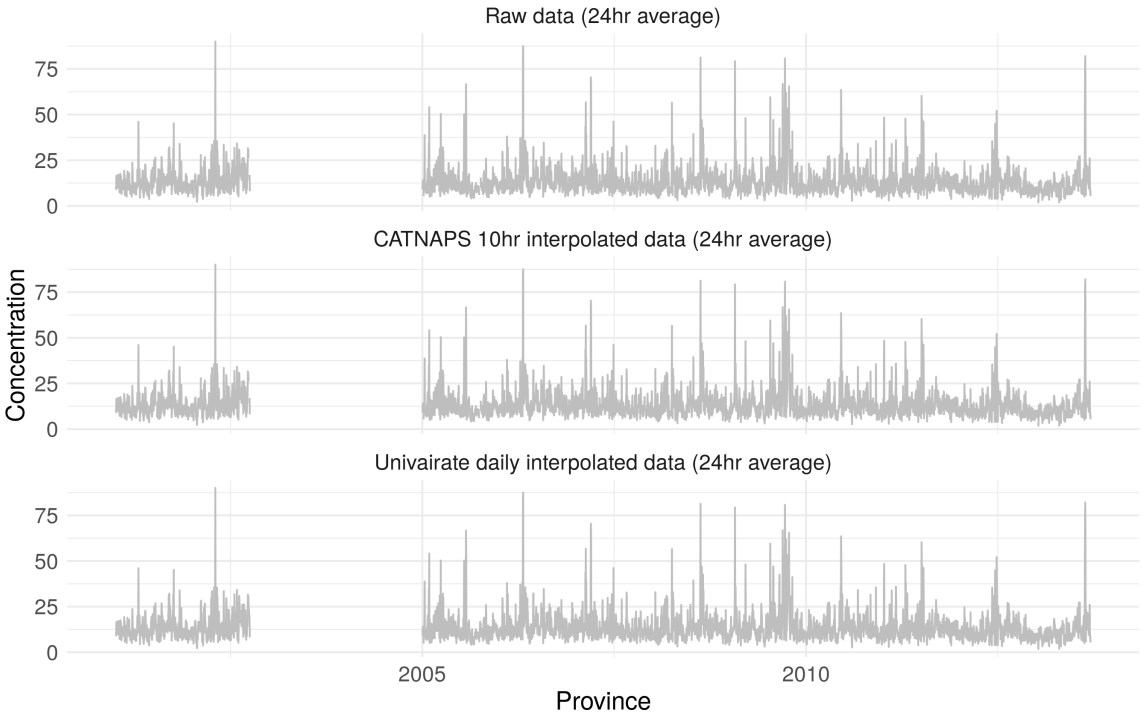


Figure 2.14: Plots of mean 24-hour PM_{10} concentrations for CD 5917, Capital BC, for the raw daily pollutant concentrations, CATNAPS 10 hour interpolated concentrations and daily interpolated concentrations.

2.3 Integrated Data

The collection and laboratory analysis of integrated data is more complex than the continuous data. As such, the data is collected by a smaller subset of sites. In this type of sampling, cartridges with filters and denuders constructed of varying media are used to collect and retain distinct compounds. The cartridges are placed in the sampler for 24 hours and are transported to a laboratory in Ottawa for analysis.

Each monitoring site varies in age and equipment, and contain one or more integrated sampling units: dichotomous samplers; speciation samples; and manual samplers. All samplers collect data over a 24 hour period at specified frequencies: once every 3, 6 or 12 days.

The total particulate matter mass (at various levels, e.g., 2.5 microns or 10 microns) from a given day is obtained by weighing the filter before and after collection [71]. Then the samples undergo chemical analyses to decompose into its components. The following components are obtained:

- Particulate Matter (PM) mass
- Metals (Ag, Al, Si, S, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Zn, Se, Br, Rb, Sr, Cd, Sn, Cu, Cs, Ba, and Pb)
- Ions (F^-, Cl^-, NO_2^-, NO_3^-, SO_4^{-2}, PO_4^{-3}, acetate, formate, propionate, methanosulfonate, oxalate, Na^+, NH_4^+, K^+, Mg^{+2}, Mn^{+2}, Ca^{+2}, Sr^{+2} and Ba^{+2})
- Ammonia, nitric acid, sulphur dioxide
- Organic carbon, elemental carbon

Each of the compounds could be measured and analyzed using several different methods. For example, metals were analyzed by an Energy Dispersive X-ray Fluorescence (ED-XRF) for total elemental mass. Acid digestion and inductively-coupled plasma mass spectrometry (ICP-MS) was also used to analyze the metal samples for near total and for water-soluble masses. This makes it essential that each observation be associated with the correct sampling machine, filter/cartridge media and analysis method. In conjunction with the total mass, a minimum detection limit (MDL) is reported for each compound observation in the raw data.

The organization of integrated data was more involved as file structure and compound naming changed over time. Additionally, the type of sampling unit, cartridge and analysis method used for each observation needed to be obtained. A description of the data wrangling and cleaning follows.

2.3.1 Cleaning of Integrated Data

As stated previously, the integrated data required numerous steps to read and organize the files such that it could be stored in a properly formatted and well organized structure. All work was done in `R`. The process used all data from 2003 to 2019. As of May 2022, data from 2020 and 2021 has not been uploaded to NAPS database, due to issues arising from the COVID-19 pandemic.

2.3.1.1 Step 1: Retrieving the Data

The process began by retrieving yearly files from Environment and Climate Change Canada's NAPS database. The files on the NAPS database were organized into sub-directories based on either sampling unit or filter size (PM_{10}, $PM_{2.5}$). A diagram of the file organization is displayed in Figure 2.15. At this point, the approach diverges for files before and after 2010 due to a major change in the file structure on the part of Environment and Climate Change Canada.

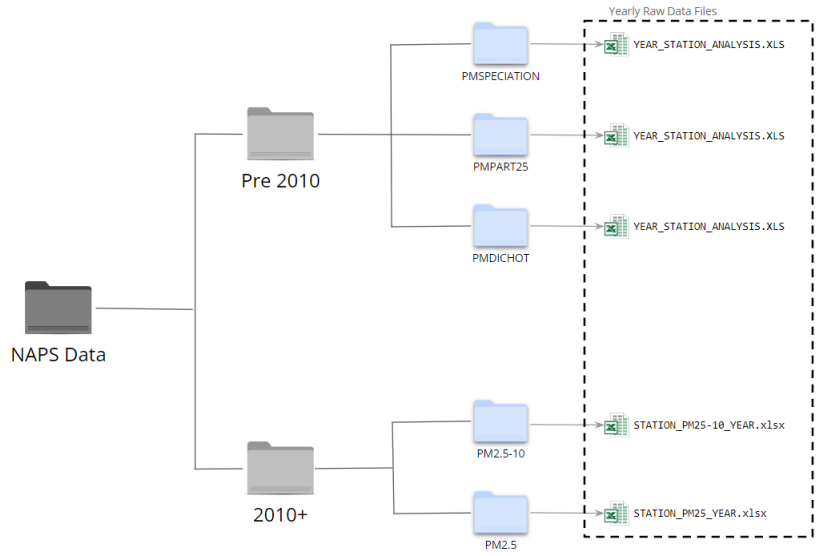


Figure 2.15: Schematic detailing the organization and file naming of NAPS raw data

Essentially, the pre-2010 data was split into numerous files, where each file contained data for a specific sampling unit and analysis method. In the 2010+ data, all samples and analysis methods were contained in one file. The retrieval of data and its organization into sensible structures set the framework for the remaining steps.

Pre-2010 Data

The pre-2010 data was organized by sampling unit (directories included: `PMDICHOT`, `PMSEPCTATION` and `PMPART25`). Within each directory, yearly station files for each analysis method contained data for both the fine and coarse filters. For example, the `PMSPECIATION` directory was used for the speciation sampler data. Within, files were named using the scheme; `YEAR_STATION_ANALYSIS_METHOD`. The following figure shows a snapshot of the file contents for station 60427 in Toronto in 2009 for Ion Chromatography (IC) analyses.

IC Results from Teflon/Nylon Filter Pack (ug/m3) - TORONTO COLLEGE & ROSS NAPS Station: 60427

For More Details on Data Contained in this File and For Contact/Acknowledgement Information Please see File SPEC_METADATA.XLS in Web-Information Folder

Date	Cartridge	Media	Speciation	Fluoride	F-MDL	Acetate	Acet-MDL	Formate	Form-MDL	Propionat	Prop-MDL	MSA	MSA-MDL	Chloride	Cl-MDL	Nitrite	NO2-MDL
1/1/09	C	T	5.935754	0	0.002087	0	0.010435	0.025421	0.010435	0	0.010435	0	0.010435	0.01847	0.010435	0	0.010435
1/1/09	C	N												0.107634	0.022261	0.026702	0.005565
1/4/09	C	T	7.680492	0	0.002087	0.074715	0.010435	0.04435	0.010435	0	0.010435	0	0.010435	0.124853	0.010435	0	0.010435
1/4/09	C	N												0.149087	0.022261	0.031387	0.005565
1/7/09	C	T	8.72966	0	0.002087	0	0.010436	0.027277	0.010436	0	0.010436	0	0.010436	0.042451	0.010436	0	0.010436
1/7/09	C	N												0.078937	0.022264	0.025264	0.005566
1/10/09	C	T	6.003071	0	0.002087	0	0.010433	0.032256	0.010433	0	0.010433	0	0.010433	0.236314	0.010433	0	0.010433
1/10/09	C	N												0.058245	0.022256	0.026607	0.005564
1/13/09	C	T	9.285764	0	0.002087	0	0.010434	0.032808	0.010434	0	0.010434	0	0.010434	0.041518	0.010434	0	0.010434
1/13/09	C	N												0.051514	0.022259	0.026326	0.005565
1/16/09	C	T	10.68063	0	0.002087	0	0.010434	0.031102	0.010434	0	0.010434	0	0.010434	0.156967	0.010434	0	0.010434
1/16/09	C	N												0.060094	0.022259	0.030844	0.005565
1/19/09	C	T	29.67463	0	0.002087	0.013514	0.010434	0.077418	0.010434	0	0.010434	0	0.010434	0.197467	0.010434	0	0.010434
1/19/09	C	N												0.066383	0.022259	0.04321	0.005565
1/22/09	C	T	30.17392	0	0.002088	0	0.010438	0.060638	0.010438	0	0.010438	0	0.010438	0.181223	0.010438	0	0.010438
1/22/09	C	N												0.08889	0.022269	0.041244	0.005567
1/23/09	FB	T		0	0.002083	0.058519	0.010417	0.035765	0.010417	0	0.010417	0	0.010417	0.009663	0.010417	0	0.010417
1/23/09	FB	N															

Figure 2.16: Snapshot of contents of speciation file for station 60427, Toronto ON, `2009_S60427_IC.XLS`

Within this file, and all other pre-2010 files, the cartridge column is used to identify the type of molecule size. *C* denotes a coarse filter (corresponding to PM 10), *F* denotes a fine filter (PM 2.5) and *FB* denotes a field blank. The field blank sample is used at least once a month in place of the routine sample for quality assurance. The

media column distinguishes the material of filter used, where a T refers to a Teflon filter and N is a Nylon filter. To reiterate, all observations within this file are obtained from a speciation sampling unit in Toronto for 2009 using IC analysis method. This file system made the association between each observation and corresponding sampling unit, filter/cartridge type and analysis method simple.

Post 2010 Data

The 2010+ data contained directories organized by particulate matter size (directories: `PM 2.5` and `PM 2.5-10`). Within each directory, one file per year and station contained all data (all analyses methods and all sampling units). Multiple spreadsheets within a file were used to organize the data, with each analysis methods contained in one sheet. Sheets included:

- Station info: site information, samplers used, collection frequency and analysis techniques used
- Metadata information: description of sampler cartridges, material and observation type
- Analysis method specific sheets: sheet names included PM 2.5, ED-XRF, IC-PMS (near total and water-soluble), Ions-Spec, Volatile Nitrate, OCEC, OC, Biomass burning markers, ammonia and acidic gaseous PM precursors

A snapshot of the $PM_{2.5}$ sheet for Station 60427 in 2010 is shown in Figure 2.17.

Sampler			S-1	S-1	S-1	S-1	S-1	S-1	S-1	S-1	S-2	S-2	S-2	S-2	S-2
Cartridge			N/A	N/A	N/A						B	B	B		
Medium			TF	TF	TF						TF	TF	TF		
Observation Type			FP	FP	FP						FP	FP	FP		
Analytical Instrument			Microbalance	Microbalance							Microbalance	Microbalance	Microbalance		
Sample Preparation			Temperature and humidity equilibration								Temperature and humidity equilibration				
CAS #															
Units	yyyy/mm/dd		µg/m3	µg/m3		kPa	°C	HH:MM	HH:MM	m3	µg/m3	µg/m3		kPa	°C
NAPS Site ID	**Sampling Date**	**Sample Type**	**PM2.5**	**PM2.5-MDL**	**PM2.5-Vflag**	**Pres.**	**Temp.**	**Start Time**	**End Time**	**Actual Volume**	**PM2.5**	**PM2.5-MDL**	**PM2.5-Vflag**	**Pres.**	**Temp.**
060427	2010/01/02	R	3.565	0.463	H1	101.725	-10.300	00:00	24:00	21.600	-999.000	0.694	M1	-999.000	-999.000
060427	2010/01/05	R	0.509	0.463	H1	101.325	-4.600	00:00	24:00	21.600	-999.000	0.694	M1	-999.000	-999.000
060427	2010/01/08	R	2.269	0.463	H1	101.592	-6.000	00:00	24:00	21.600	-999.000	0.694	M1	-999.000	-999.000
060427	2010/01/11	R	19.259	0.463	H1	101.992	-2.000	00:00	24:00	21.600	20.806	0.698	H1	100.392	-3.200
060427	2010/01/14	R	29.722	0.463	H1	101.858	3.500	00:00	24:00	21.600	32.954	0.698	H1	100.258	1.900
060427	2010/01/17	R	21.111	0.463	H1	101.325	4.300	00:00	24:00	21.600	21.231	0.698	H1	99.725	2.900
060427	2010/01/20	R	4.491	0.463	H1	101.858	-0.500	00:00	24:00	21.600	9.707	0.698	H1	100.258	-1.600
060427	2010/01/23	R	6.898	0.463	H1	102.258	1.700	00:00	24:00	21.600	9.567	0.698	H1	100.658	0.400
060427	2010/01/26	R	5.972	0.463	H1	100.258	2.900	00:00	24:00	21.600	11.313	0.698	H1	98.659	1.700
060427	2010/01/29	R	2.731	0.463	H1	102.925	-10.100	00:00	24:00	21.600	5.935	0.698	H1	101.458	-11.000
060427	2010/02/01	R	-999.000	-999.000	M1	-999.000	-999.000	-999.000	-999.000	-999.000	9.915	0.698	H1	100.792	-3.700

Figure 2.17: Snapshot of contents of speciation file for station 60427, Toronto ON, `S60427_PM25_2010.xlsx`

The first 6 rows correspond to information about the sampling unit. Acronyms (S-1, S-2) must be associated with the corresponding sampler indicated in the 'station info' spreadsheet. For example, for Station 60427 in 2010, sampler S-1 is a Dichotomous Sequential Air Sampler (2025-D Thermo) and sampler S-2 is a Partisol 2300 Speciation sampler. The sampler types were found by parsing the 'station info' sheet and looking for the sampler code (i.e. S-1, S-2) followed by the description. If the description could not be found, manual intervention was needed to associate the sampler code to the correct unit. This association was imperative as samples obtained from different types of machines should not be merged.

2.3.1.2 Step 2: Addressing Inconsistencies in Compound Naming

Over time, the naming of compounds changed. For example, silicon could be called Silicon, Si or Silicon (Si). These changes throughout the files needed to be tracked to generate one consistent naming scheme. A file called a 'header dictionary' was created to contain all possible compound names, so they could be combined.

The header dictionary was created by reading all files and creating a list of all unique column names. Names containing 'MDL' or 'flag' were removed from the list and compounds were manually associated with one another. This facilitated a consistent name for each compound in the data. In hindsight, using string matching or pattern matching may have been a much easier approach than manual intervention,

although there were so many patterns, the implemented solution was the most time efficient.

2.3.1.3 Step 3: Creating the Database

The header dictionary in conjunction with the sampling unit, analysis method and filter/cartridge media were used to rename columns in the data. The naming structure was:

`sampling_unit.analysis_method.filter_cartridge_type.compound`

After addressing inconsistencies in step 2, all data could be combined into an organized structure. Two separate lists were created: one for PM less than 2.5 microns and one for PM less than 10 microns. Both lists used the same structure; a list of stations with all component time series data.

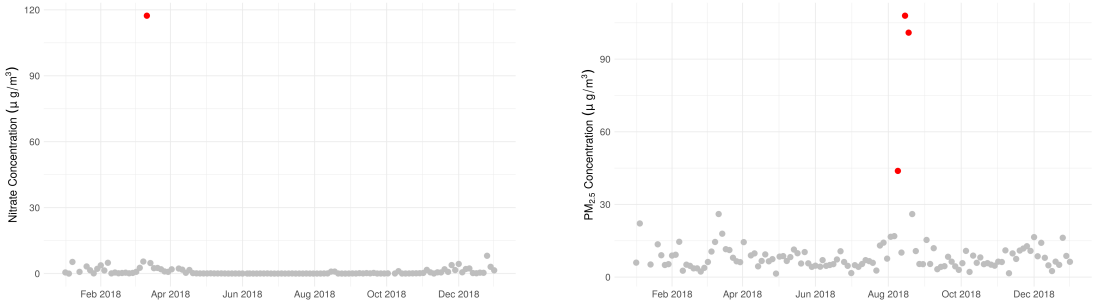
2.3.1.4 Step 4: Data Verification

The data was screened for errors or inconsistencies. Recall that each measurement had a corresponding minimum detection limit (MDL). All concentrations which fell below their MDL were replaced by half the MDL. Abnormal values were identified by preforming summary statistics and checking data points which fell 2 standard deviations outside of the mean concentration.

The verification of data was imperative, so as to check for errors in the raw data that may have incurred due to cleaning steps or within the raw data itself. Abnormal values in the raw data may be an indication of errors, whether observational or due to instruments. An example of this type of error is provided in Figure 2.18. In Figure 2.18a, an abnormal nitrate concentration was measured in March 2018 at the Edmonton, Alberta station. A concentration of 151 $\mu g/m^{-3}$ was measured which is well beyond typical concentration at that site (mean value of 1.01 $\mu g/m^{-3}$). The time

series of nitrate was reviewed to determine if such a high value is atypical. It was deemed to be an erroneous measurement and replaced by a missing value.

In Figure 2.18b, several abnormal $PM_{2.5}$ values were observed in August 2018. The steady increase of values, peaking after a few days, and not a one-off high value, indicates potential environmental causes. After a search on the internet, it was determined that these values were due to nearby forest fires. Several new outlets depicted pictures of an orange smoggy sky and headlines read "Lovely morning in the apocalypse" [6]. Those high values were unusual, but not erroneous.



(a) (b) .

Figure 2.18: Example of an abnormal measured value (in red) for pollutant concentrations measured in Division No. 11 (Edmonton), AB for 2018.
(a)Nitrate Concentrations.
(b) $PM_{2.5}$ Concentrations.

2.3.1.5 Step 5: Compound Calculations

For mass reconstruction, chemical components were calculated as in the previous work by Dabek-Zlotorzynska *et al.* [25]. The following components were calculated from the speciation data: ammonium nitrate (ANO_3); ammonium sulphates (ASO_4); organic matter (OM); elemental carbon (EC); crustal matter (SOIL); sodium chloride (NaCl); particle-bound water (PBW); and reconstructed mass (RCM, the sum of all calculated components). Trace element oxides were not calculated, as several of the

elements were not collected by the speciation sampler. Details of the compounds and their calculation are provided in Table 2.8.

Table 2.8: Summary of calculated compounds and components.

Compound name	Symbol	Component
ammonium nitrate	ANO_3	$1.29NO_3^-$
ammonium sulphates	ASO_4	SO_4^{2-} + NH_4^+ - $0.29NO_3^-$
organic matter	OM	kOC
elemental carbon	EC	EC
crustal matter	SOIL	3.48Si + 1.63Ca + 2.42Fe + 1.41K + 1.94Ti
sodium chloride	NaCl	Na + Cl
particle-bound water	PBW	$0.32(SO_4^{2-} + NH_4^+)$
reconstructed mass	RCM	ANO_3 + ASO_4 + OM + EC + SOIL + NaCl + PBW

Correction factor, k, calculated from the SANDWICH method [25]

2.3.2 Analysis of Integrated Data

In total, eleven NAPS stations were selected for analysis, based on availability of the organic matter correction factor, k [25]. For the remainder of the work, stations will be referred to by census division name, with multiple NAPS stations being merged (as there was no overlap between stations at a common geographic location - the stations were physically relocated).

Figure 2.19: Map of Canada with location of NAPS speciation stations indicated in red

Table 2.9: List of NAPS stations and corresponding CD used.

Census Division (ID)	NAPS ID	NAPS Station Name	NAPS Station City
Columbia-Shuswap, BC (5939)	103202	Golden Hospital Helipad	Golden
Greater Vancouver, BC (5915)	100119	Burnaby South	Metro Van - Burnaby
Fraser Valley, BC (5909)	101004	Abbotsford Airport	Metro Van - Abbotsford
	101005	ABBOTSFORD AIRPORT-2	Metro Van - Abbotsford
Division No. 11, AB (4811)	90132	Edmonton Mcintyre	Edmonton
Essex, ON (3537)	60211	Windsor West	Windsor
Haldimand-Norfolk, ON (3528)	62601	Experimental Farm Simcoe	Simcoe
Toronto, ON (3520)	60427	Gage Institute	Toronto
	60439	Roadside – Wallberg (Uoft)	Toronto
	60445	TORONTO DOWNTOWN	Toronto
Le Haut-Saint-Laurent, QC (2469)	54401	Saint-Anicet	Saint-Anicet
Montréal, QC (2466)	50104	Ontario	Montreal
	50134	Montréal-Molson Or Motreal-Saint-Joseph	Montreal
York, NB (1310)	40801	Canterbury	Dow Settlement
Halifax, NS (1209)	30113	Johnston Building	Halifax
	30118	Vogue Building	Halifax

Of the selected sites, Greater Vancouver, British Columbia, was the only station with observations for the entire period of interest, 2003–2019. A timeline of the available data for the selected stations is shown in Figure 2.20.

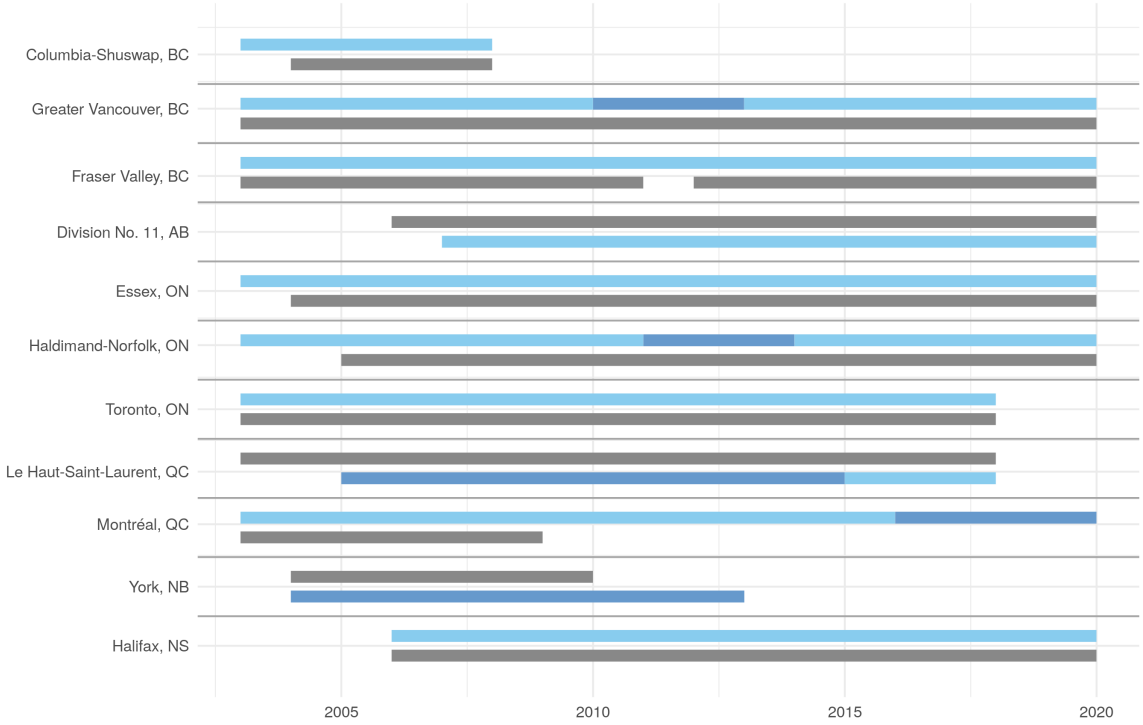


Figure 2.20: Timeline of data availability at NAPS stations for speciation (grey), dichotomous (light blue) and manual (dark blue) air samplers from 2003–2019.

It is of interest to examine the seasonal variation in particulate matter concentrations, and the implications for component concentration. We begin with the monthly average levels of $PM_{2.5}$ and organic carbon.

2.3.2.1 PM2.5 Mass Concentrations

Organizing the available sites from west (Greater Vancouver) to east (Halifax), we plot the monthly average $PM_{2.5}$ mass concentration in μgm^{-3} in Figure 2.21, from 2010–2019. The whiskers in this plot represent the 2^{nd} (lower) and 98^{th} (upper) percentiles of the available data. Median $PM_{2.5}$ concentrations ranged from 5 to 9 $\mu g\ m^{-3}$. Six concentrations above 50 $\mu g\ m^{-3}$ were observed, all in Division No. 11, Alberta (commonly known as the city of Edmonton). They can largely be attributed

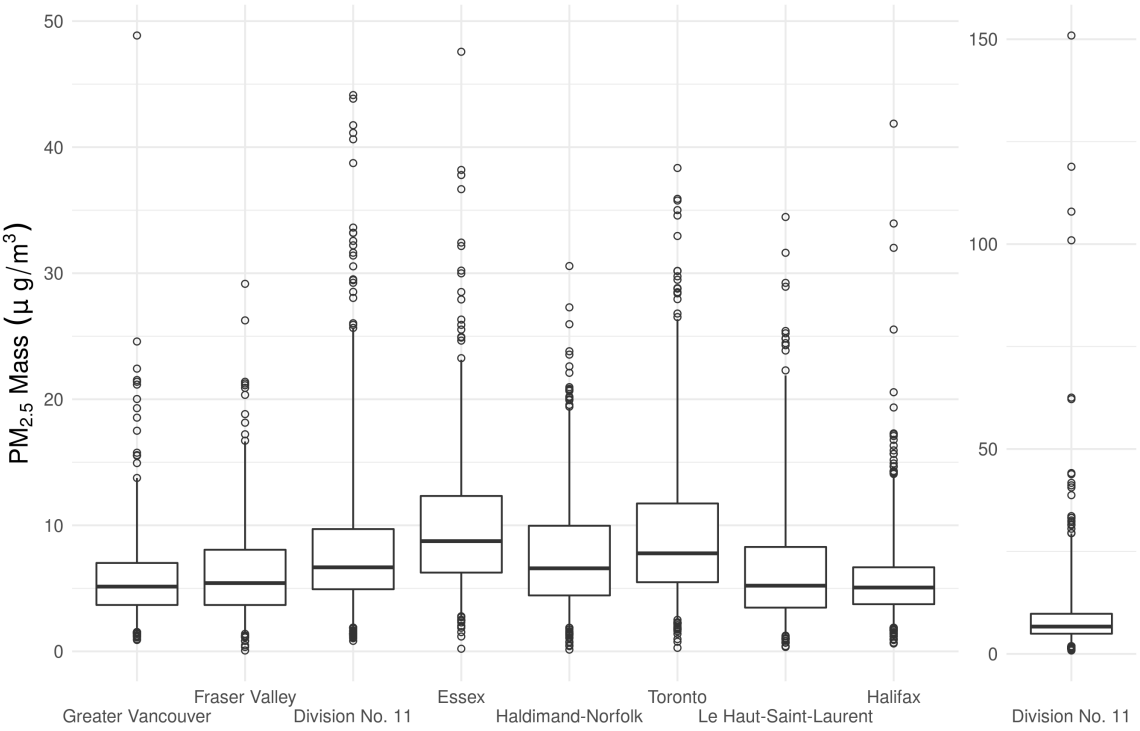


Figure 2.21: Measured $PM_{2.5}$ monthly average mass in μg m^{-3}, by location, organized west to east from 2010–2019

to months with large-scale wildfires (especially 2015 and 2018). Similarly, the outlier for Greater Vancouver (48 μg m^{-3}) was also due to a fire on Burnaby Mountain in July 2015.

Figure 2.22 takes the same data as Figure 2.21, but broken down by month. All stations exhibit seasonal variation, with maximums observed in the North American summer months (July/August). Essex, Toronto and Le Haut Saint-Laurent all exhibit peaks in early winter, which may be attributable to select years with dry or limited snow pack, where road grit is easily stirred up and aerosolized. Additionally, high PM levels may be explained by meteorological effects, such as wind and temperature inversions, which act to trap particulate matter at the surface and prevent its dispersion into the atmosphere [103, 102, 108]. Temperature inversions occur at a higher frequency in the winter [72].

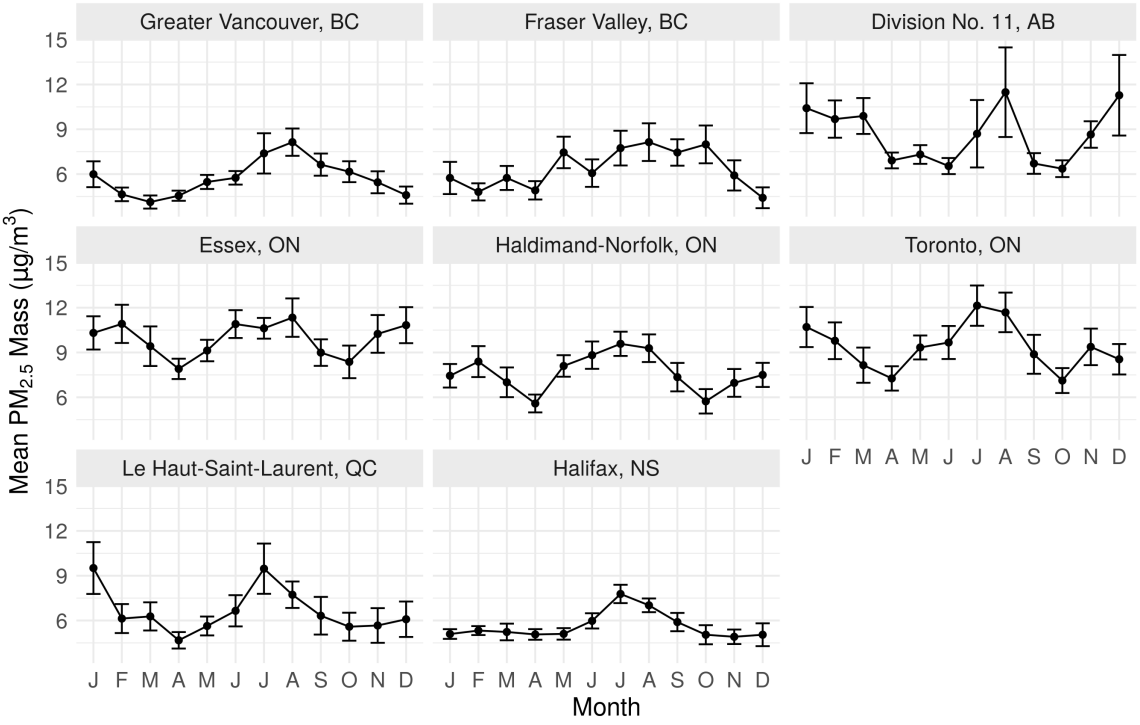


Figure 2.22: Monthly mean $PM_{2.5}$ concentration in μg m^{-3}, with the error bars representing 90% percentiles.

The highest average monthly $PM_{2.5}$ levels were observed in the urbanized locations of Division No. 11 and Toronto, as well as Essex (a mid-sized city with a heavily industrial and commercial freight presence, down-wind of Detroit, Michigan). Much of the excessive variability in Division No. 11 (Edmonton) can be explained by months with wildfires, as these act as outliers and expand the variation tremendously.

Figure 2.23 further breaks down the data from the previous figures and displays the data collected in January. The measured PM mass for each day is displayed, and outliers (whose concentration is more than three standard deviations from the mean) are highlighted in red. There can be more than one measurement per day, as the figure displays all January data from 2010 – 2019.

There is significant variation in Division No. 11, where the outliers range from 41-62 μg m^{-3}. Additionally, two outliers were identified in Greater Vancouver and one

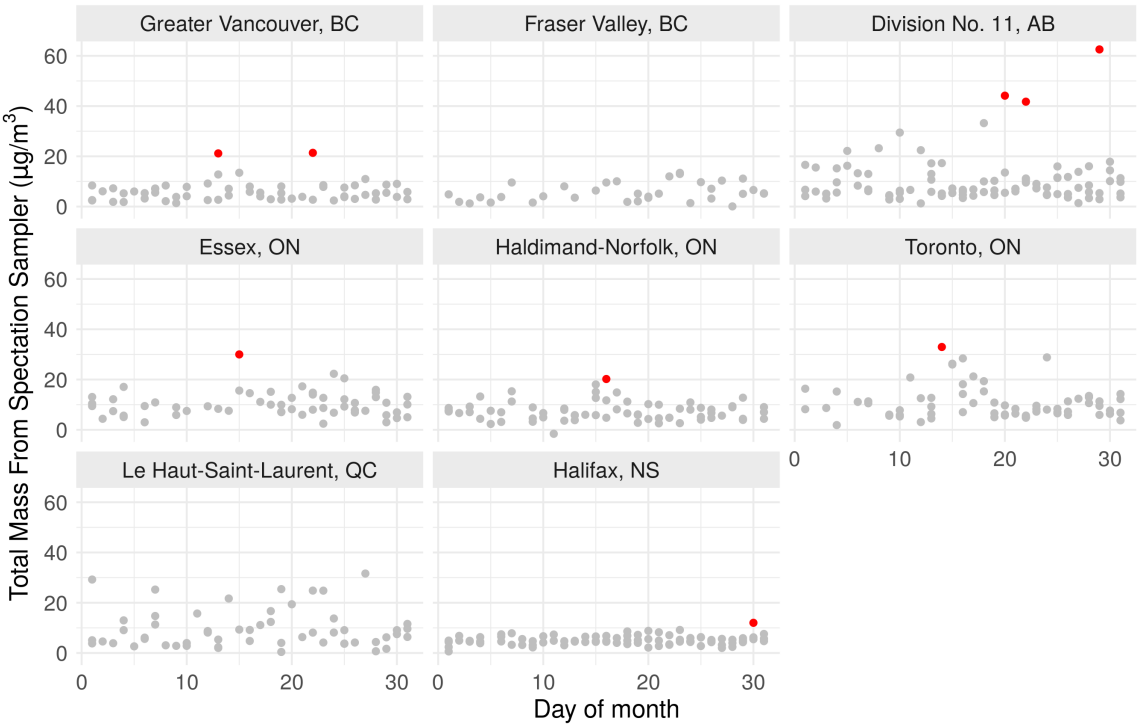


Figure 2.23: Daily $PM_{2.5}$ concentration in μg m^{-3}, for January with outliers ($> 3\sigma$) indicated in red

in each of Halifax, Haldimand-Norfolk, Toronto and Essex. The presence of outliers can explain variability observed in Figure 2.22 for January.

The mean monthly $PM_{2.5}$ observations for 2010–2019 are on average lower than those from 2003–2009, with a mean monthly concentration difference of -1.6 μg m^{-3}. The largest differences occurred in Essex and Haldimand-Norfolk where they recorded a mean monthly average $PM_{2.5}$ difference of -3μg m^{-3} and -2.9 μg m^{-3} for all months, respectively.

One important component of PM mass is organic carbon (OC). In Figure 2.24, we examine mean organic carbon concentrations across the sites. Mean organic carbon is measured using a quartz filter (cartridge A), while an active organic carbon blank is simultaneously measured using a Teflon filter (cartridge B). The mean passive travel and field blank OC concentrations ranged from 0.49 to 0.64 μg m^{-3}, and did not show

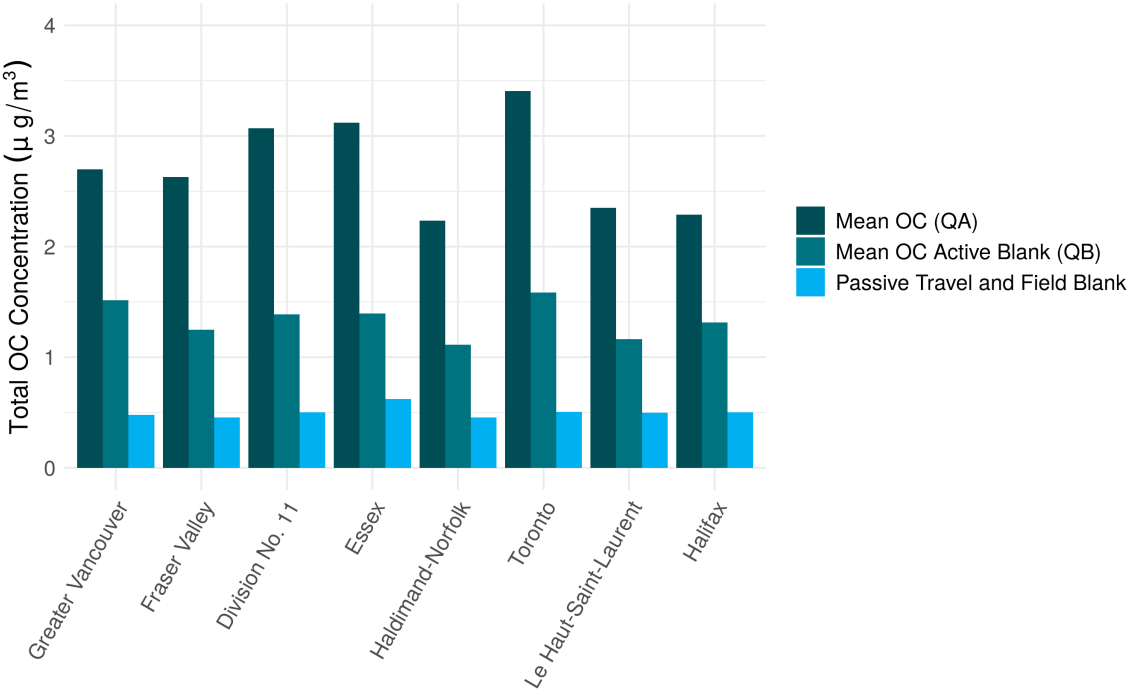


Figure 2.24: Mean organic carbon concentration, in μg m^{-3}, by cartridge and site

any obvious relationship to mean OC. Mean active blank values ranged from 1.1 to 1.6 μg m^{-3}. Lower active blank values were observed in rural sites (Haldimand-Norfolk, Le Haut Saint-Laurent) as compared to urban sites (Toronto, Barnaby, Essex). There were no obvious trends in mean organic carbon levels based on location.

When comparing the measured organic carbon levels, the passive travel and field blank was 14-22% of the mean OC concentration. The mean active blank was 45-57% of the mean OC concentration. Additionally, when comparing the mean OC concentrations from 2003–2009 to 2010–2019, the mean concentration for measured parameters decreased by 0.3-0.9 times, across the sites.

2.3.2.2 Reconstructed PM2.5 Mass

The reconstructed mass (RCM), calculated as described in Section 2.3.1.5, is presented in Figure 2.25 for samples collected during warm (April-September) and cold

(October-March) seasons. On average, there was a small discrepancy between the RCM and the $PM_{2.5}$ concentration. The mean difference between RCM and measured $PM_{2.5}$ was 0.96 μg m^{-3}.

The largest contributor to the reconstructed mass was the combination of ASO_4 and PBW. As shown in Figure 2.25, the components had the highest concentration in Ontario and Quebec. Of the eastern sites, the average contribution to the RCM was 25-45% in the summer and 30-39% in the winter. For the remaining sites, ASO_4 and PBW accounted for 21-25% of mass in the warm season and 16-23% in the cold season.

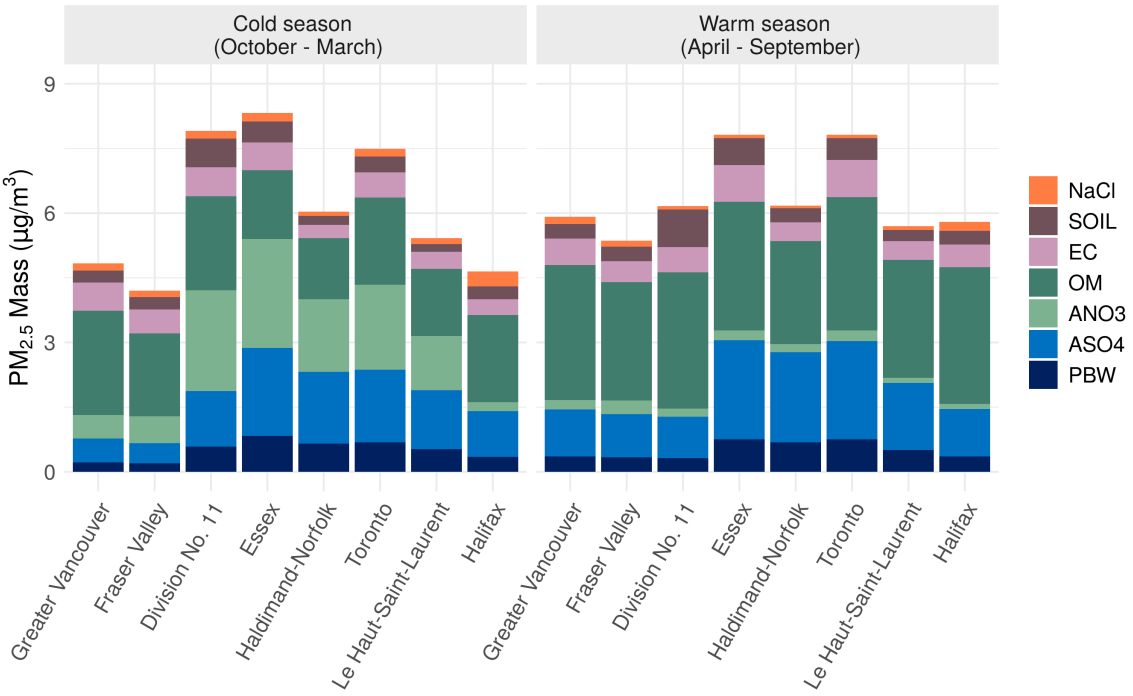


Figure 2.25: Reconstructed $PM_{2.5}$ mass from 2010–2019 by mean compound for the cold season (October - March, left) and the warm season (right, April - September).

The next largest component contributor was organic matter (OM), which accounted for 38-55% of the mass in the warm season and 19-50% in the cold season.

The largest contributions of OM occurred at coastal and residential sites (Greater Vancouver, Fraser Valley and Halifax). A decrease of OM contribution in the winter may be explained by the OM correction factor, k, used to calculate the component [25, 34]. On average, the correction factor is higher in the summer and spring season, while the monthly OC concentration showed seasonal patterns with peaks in the winter months for all sites.

The contribution of ANO_3 was greatest in the cold season for Alberta, Ontario and Quebec where the average percent composition ranged from 23-30% and 2-3% in the summer. In British Columbia, the ANO_3 percent composition ranged from 3-5% in the summer to 11-15% in the winter, while in Halifax, the composition remained at 3-4% year round. There was up to an eleven-fold increase in ANO_3 during the winter months compared to summer. This may be explained by the low temperature in the winter, where the formation of ANO_3 is chemically favoured [90].

NaCl levels were between 1-2% at all non-coastal sites. At coastal sites, Halifax NaCl concentration comprised 9% of the total RCM in the cold season and 4% in the warm season, while Greater Vancouver and Fraser Valley observed an average percent composition of 3%. The wintertime concentration was higher than the summertime at all sites. Higher levels on coastal sites are expected due to the influence of saltwater.

The contribution of SOIL was slightly higher in the warmer season where it ranged from 5-14% compared to 4-8% in the colder season. The cities with the largest contribution were Division No. 11, Toronto and Essex. It is suspected that this is due to the soil composition, coarseness and moisture in the regions.

Figure 2.26 shows the reconstructed PM mass for the years 2003–2009. On average, secondary ASO_4 and associated PBW accounted for 31-46% of total $PM_{2.5}$ mass at the eastern sites during the warm season, and 20-34% in the cold season. For the western sites, ASO_4 and associated PBW accounted for 9-20% of mass in the warm season and 7-11% in the cold season. The ASO_4 and PBW levels in the cold season

increased by 1.1-2.4 times from 2003-2009 to 2010-2019. There were no obvious trends in the summer season.

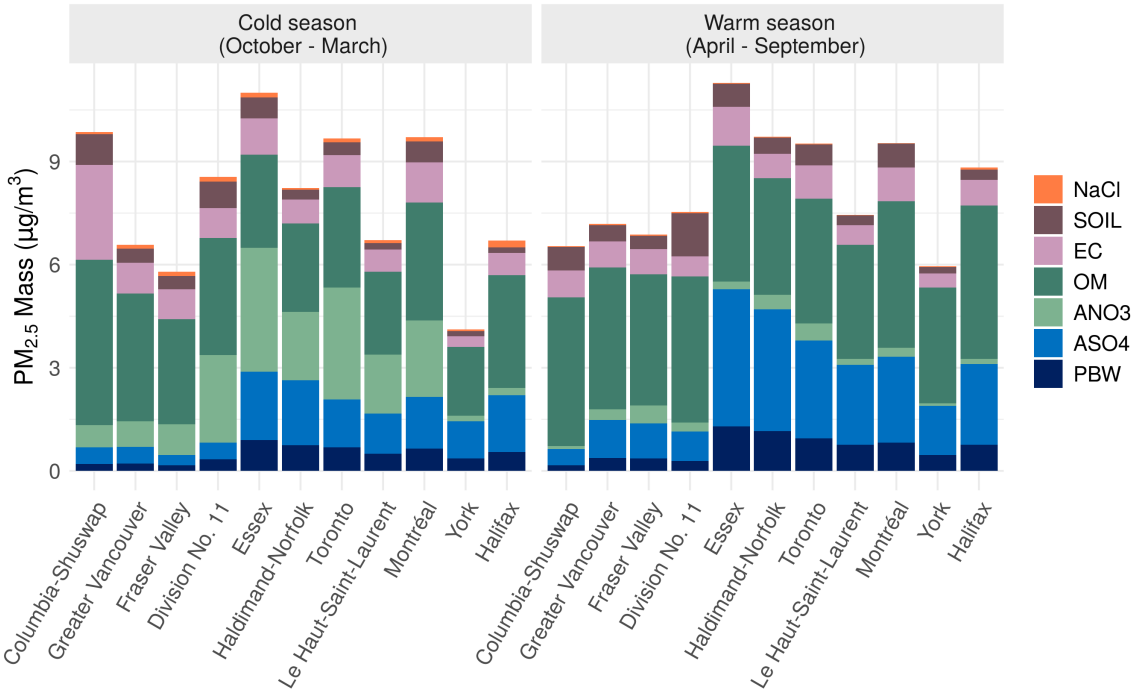


Figure 2.26: Reconstructed $PM_{2.5}$ mass from 2003 – 2009 by mean compound for the cold season (October - March, left) and the warm season (right, April - September).

There were no obvious patterns in the difference of ANO_3 levels from 2003–2009 to 2010–2019 as the levels were relatively consistent when comparing the average RCM percentage.

When comparing the average percent of RCM composition for NaCl from 2003–2009 to 2010–2019, the levels decreased. Specifically, NaCl decreased by 1.2-4.2 times in the cold season and 1.1-5.3 times in the warm season.

Figure 2.27 shows the reconstructed mass for ten days in the warm and cold seasons with the largest measured $PM_{2.5}$ concentrations. The mean total concentration for the highest 10 days was largest for Division No. 11 (Edmonton) and sites in Ontario. Mean

total mass concentration ranged from 18 μg m^{-3} (Halifax) to 59 μg m^{-3} (Division No. 11) for the 10 highest days in the summer and ranged from 16 μg m^{-3} (Fraser Valley) to 51 μg m^{-3} (Division No. 11) for the 10 highest days in the winter.

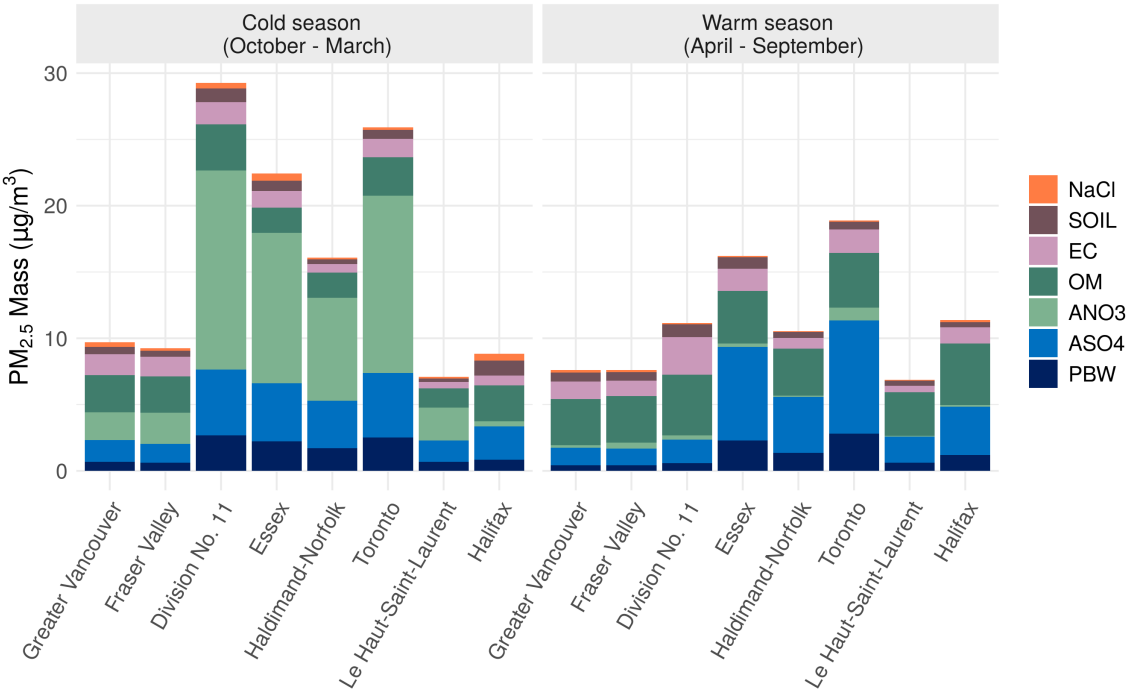


Figure 2.27: Reconstructed $PM_{2.5}$ mass by mean compound for the 10 highest mass concentration days for the cold season (October - March, left) and the warm season (right, April - September) from 2010–2019.

In the 10 highest days of the warm season in Division No. 11 and the eastern sites, the mass was primarily composed of PBW and ASO_4 (42 - 63 %). In the warm season at the western sites, organic matter encompassed the largest proportion of the mass (44 - 46 %). Similar findings were obtained for the 10 highest days in the cold season, where, PM was primarily composed of ANO_3 in Division No. 11, Ontario and Quebec (49 - 52 %) and OM in the west (28 - 30%).

2.3.2.3 Monthly Variations in the Major PM2.5 Components and Vapour Phase Species

Ammonium sulphate and ammonium nitrate: The median ammonium sulphate and ammonium nitrate concentrations by site and month is displayed in Figure 2.28. The western sites (Greater Vancouver and Fraser Valley), show some seasonal variation of ammonia sulphate with a summertime median maximum of 1.1 μg m^{-3} in August and minimum of 0.2 μg m^{-3} in the colder months (November/December/January). The remaining sites, particularly those in Ontario, have a higher median ASO_4 concentration, on average. There is no obvious seasonal pattern in the eastern sites. The eastern sites had median monthly ASO_4 values ranging from 0.3-1.8 μg m^{-3}. The average median ammonium sulphate level in Ontario was 1.2 μg m^{-3}, compared to 0.6 μg m^{-3} for the remaining provinces.

By contrast, ammonium nitrate displayed a seasonal pattern for all sites with peaks in the colder months (October - March). Most sites had peak concentrations in January or December. Peak levels ranged from 0.6-3.4 μg m^{-3}, with sites in Ontario exhibiting the highest concentrations (1.8-3.4 μg m^{-3}). Similar observations were obtained from the median ammonium sulphate and ammonium nitrate concentration from 2003–2009.

Carbonaceous compounds: The median elemental carbon (EC) and organic matter (OM) concentrations by site and month are displayed in Figure 2.29. Organic matter showed seasonal patterns for all sites, with maximums in the range of 2.9 - 3.5 μg m^{-3} occurring in July/August. Minimum OM values were observed in October-January and ranged from 1.1 to 2 μg m^{-3}.

Figure 2.29 shows similar trends for elemental carbon and organic carbon. Some seasonal patterns can be seen in Ontario and Halifax, where peaks occur in the summer and minima in the cooler months. The lowest concentrations of EC were

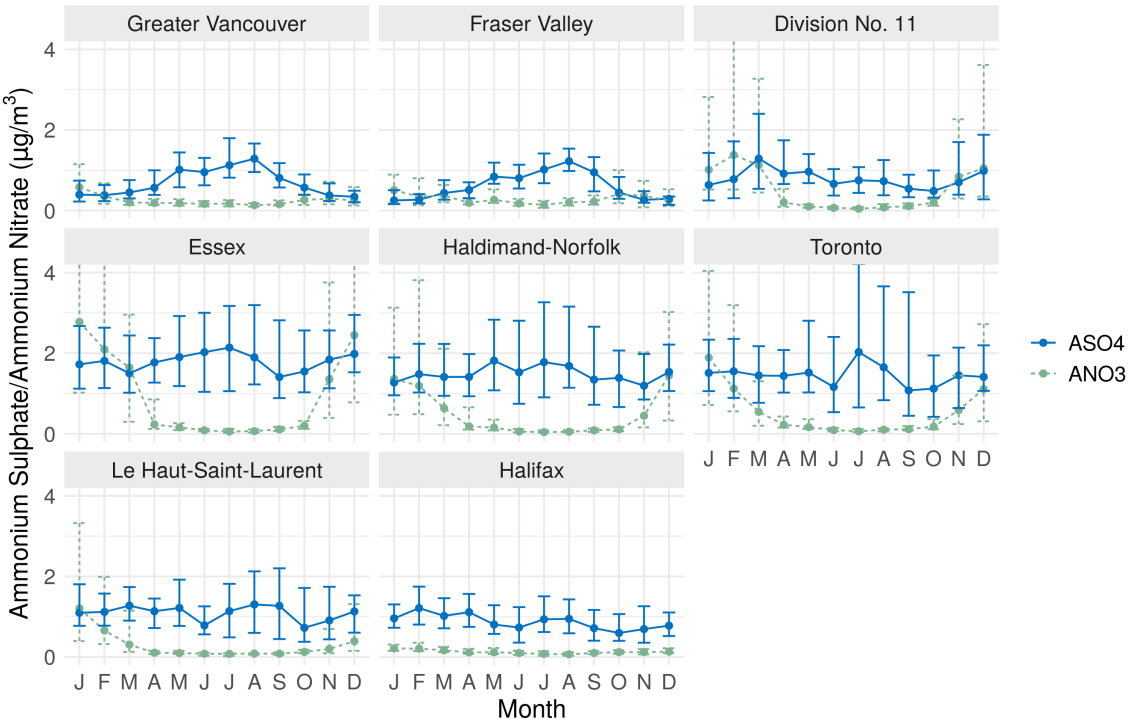


Figure 2.28: Ammonium sulphate and ammonium nitrate concentrations, in μg m^{-3}, by site and month (median and interquartile range). Note that the scale for Fraser Valley is an order of magnitude larger than the others.

observed in Saint-Ancient and Halifax where values ranged from 0.1 - 0.4 μg m^{-3}. The largest EC concentrations were observed in Toronto, Fraser Valley and Essex (0.8 - 1.0 μg m^{-3}). Median organic matter values ranged from 0.8 - 4.5 μg m^{-3}, with the largest concentrations observed in Fraser Valley and Essex. Large variation observed in Greater Vancouver in August is attributed to forest fires.

In 2003-2009, the monthly pattern for OM and EC was similar. This was also observed in 2010-2019.

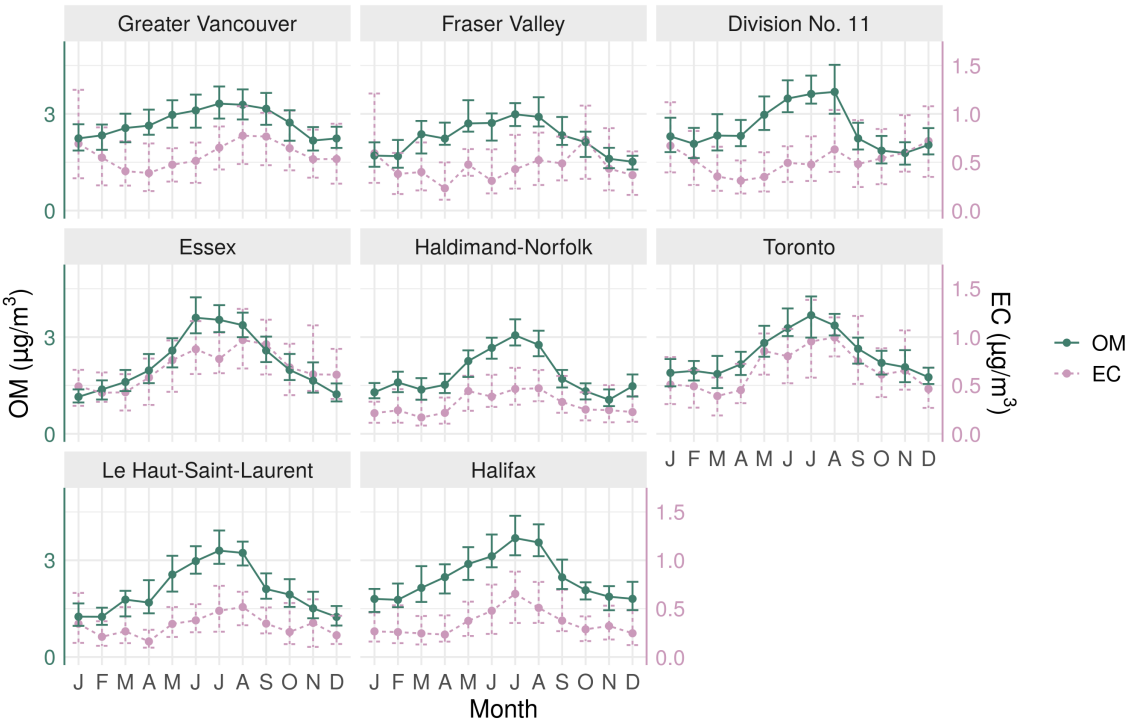


Figure 2.29: Elemental carbon (EC) and organic matter (OM) concentration, in μg m^{-3}, by site and month (median and interquartile range)

SOIL and sodium chloride: Figure 2.30 displays the median SOIL and sodium chloride concentrations by site and month. Sites in Alberta and Ontario showed some seasonal variation of soil concentrations, with higher median levels in the warm months (April-July). Median SOIL concentrations in Ontario and Alberta ranged from 0.1-0.7 μg m^{-3} and ranged from 0.1-0.4 μg m^{-3} at the remaining sites. The average median SOIL concentration for all sites was 0.3 μg m^{-3}.

The largest NaCl concentration was observed in Halifax (0.1-0.4 μg m^{-3}), as expected due to its geographic location on a coast. For all eastern sites, the lowest NaCl concentrations were observed in the summer (June/July/August). In the Fraser Valley and Greater Vancouver, the lowest concentrations were observed in November. From 2003–2009, the mean NaCl level for all observations was 0.32 μg m^{-3}. This decreased in 2010–2019, where the average concentration for all sites was 0.15 μg

m^{-3}.

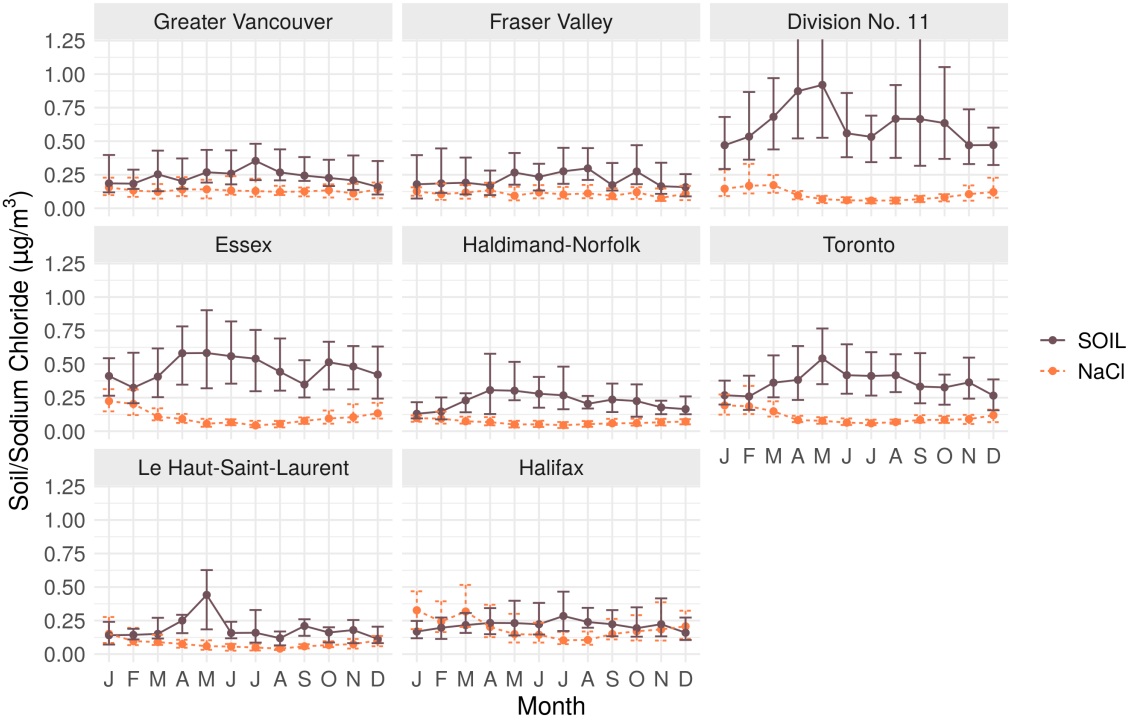


Figure 2.30: Soil and sodium chloride concentration, in μg m^{-3}, by site and month (median and interquartile range)

Gas-phase species: The median ammonia mixing ratio (in parts per million) is shown in Figure 2.31. The concentration for Fraser Valley is using a scale around ten times larger than the other sites in. Data for Montreal is missing. There are seasonal patters in the ammonia concentration for the eastern sites, where peaks occur in the warmer months (May/June/August) and ranged from 1.3 - 3.4 ppb. Minimum concentrations for the eastern sites occurred in January and February and was in the range of 0.4-1.2 ppb. The seasonal trends and range of median concentrations observed in 2010–2019 was similar to those from 2003–2009.

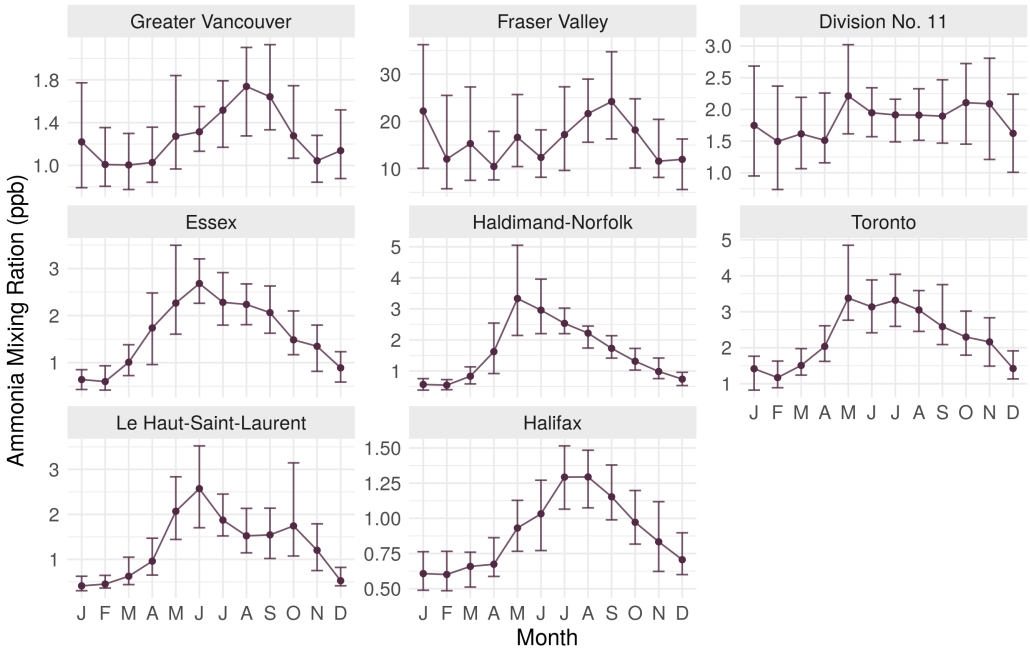


Figure 2.31: Ammonia mixing ratio, in ppb, by site and month (median and interquartile range)

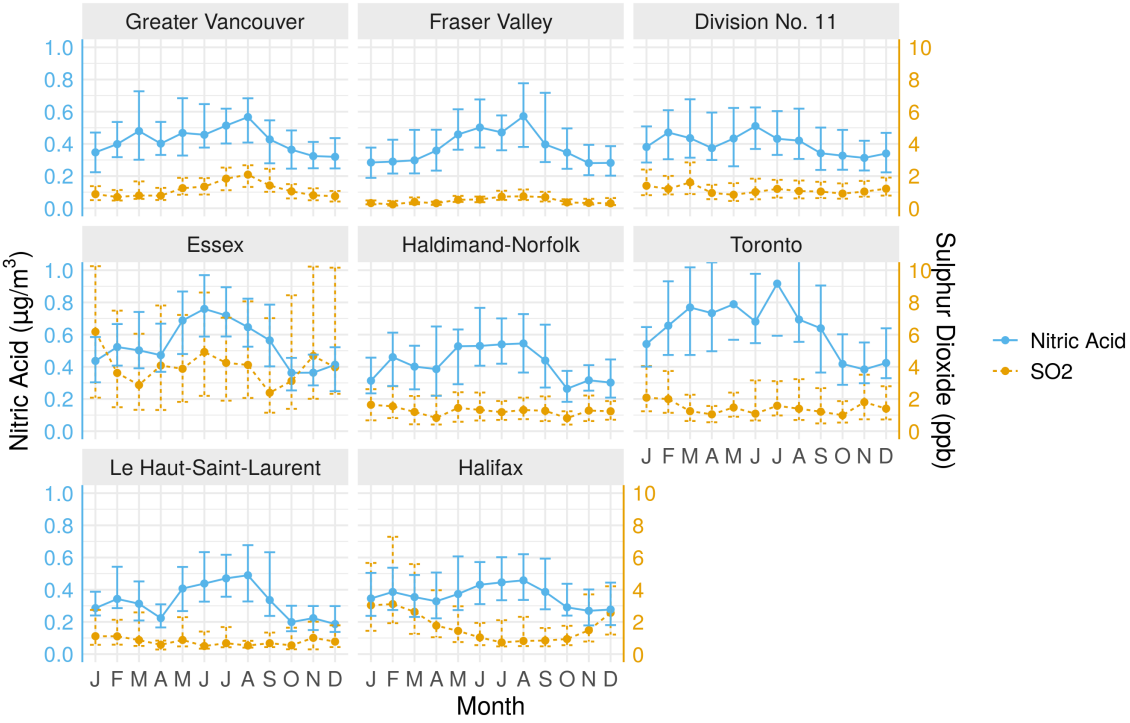


Figure 2.32: Sulphur dioxide mixing ratio, in ppb, and nitric acid concentration, in μg m^{-3}, by site and month (median and interquartile range).

Figure 2.32 displays the median sulphur dioxide mixing ratio and nitric acid concentrations by site and month. Most sites observed higher median nitric acid concentrations in the warm months (May-August). Seasonal patterns were only apparent in Essex, Toronto and Saint-Ancient, where the concentration peaked in May (0.6-0.8 μg m^{-3}) and minimums were observed in October/November (0.2-0.4 μg m^{-3}).

The highest median sulphur dioxide concentration was observed in Essex, where it ranged from 1.9-4.6 ppb. The remaining sites ranged from concentrations of 0.4-2.4 ppb. When comparing the monthly median SO_2 value from 2003-2009 to 2010-2019, on average, the concentrations decreased in Halifax, Essex and Haldimand-Norfolk and increased at all other sites.

2.3.2.4 Discussion and Conclusions

Ammonium sulphate and nitrate were the largest contributors to the reconstructed $PM_{2.5}$ mass, particularly in the winter where the concentration of ASO_4 increased at all sites. Division No. 11 (Edmonton) experienced the largest seasonal difference with a 15-fold increase in the percentage of reconstructed mass during the winter (October to March).

The formation of ammonium nitrate is controlled by an equilibrium equation with ammonia and nitric acid as precursors. Hence, it is limited by the availability of precursors, the ambient temperature and the relative humidity [89]. The observed increase in winter time ammonium nitrate concentrations at all sites is a consequence of the lower temperatures, where the formation of ammonium nitrate is favoured [81]. Additionally, the increase corresponds to increases in ammonium and nitric acid at the sites.

Ammonium sulphate's formation is dependent on the concentration and abundance of ammonia moles (which must be two or more times greater than sulphuric

acid) [81]. Consequently, peaks in particulate ammonium sulphate roughly corresponds with peaks in ammonia.

Ammonia concentrations in Fraser Valley were one order of magnitude larger than all other sites. The elevated concentrations are primarily a consequence of poultry broiler farms but nitrogen fertilizers and wastewater treatment also play a role [69, 57, 38].

Contributors to particulate ammonium nitrate and sulphate include agriculture (nitrogen fertilizer and livestock), coal-burning power plants and vehicle emissions [25, 104, 83, 7]. High density Canadian agricultural operations located in southeast Ontario and southern Quebec [20] may explain increased levels of ammonia, ammonium sulphates and ammonium nitrates in nearby sites (Windsor, Haldimand-Norfolk, Toronto and Saint-Ancient). Additionally, Toronto and Essex are urban areas with heavy traffic. Notably, Essex is situated on the busiest land boarder crossing between Canada and the United States (Detroit-Windsor). Particulate matter from Detroit and border traffic, plus being down-wind of the Ohio River Valley and the heavy industry located there, all influence Essex's air.

Emissions from the combustion of fossil fuels, in particular coal, contributes to particulate ammonia sulphate and nitrate [23]. Although coal-burning power plants are being phased out, they are currently used in Alberta (near Division No. 11), Detroit (near Essex), New Brunswick and Nova Scotia [68].

Carbonaceous compounds (elemental carbon and organic matter) comprises a significant portion of the reconstructed mass (RCM) and its contribution increased in the winter months (October–March) compared to the summer months (April–September). There were no apparent differences in concentration for urban and rural sites, but this finding may be a consequence of the limited number of rural sites in the study. Tai P.K. Amos *et al.* showed a positive correlation between temperature and OC/EC [98]. These findings are consistent with our observations.

Particulate soil is a consequence of melting snowpack, agriculture, soil composition, moisture and coarseness [1]. The observed soil trends for most sites, with soil concentrations rising in March/April and peaking around May, correlates with the release of pollutants during spring snow melt. The largest soil concentrations were observed in urban locations; Division No. 11 (Edmonton), Essex and Toronto. The lowest particulate soil concentrations were observed in Greater Vancouver, Fraser Valley and Halifax.

Non-coastal urban sites (Division No. 11, Essex, Toronto) showed seasonal variation of sodium chloride levels, with peak levels occurring in January, a consequence of winter road salting. High NaCl levels were observed at coastal sites (Halifax, Greater Vancouver and Fraser Valley), due to ocean salinity. NaCl had the smallest contribution to the reconstructed mass.

Sulphur dioxide is primarily emitted with fossil fuel combustion and industrial processes. It is also produced from marine phytoplankton (via the oxidation of dimethyl sulfide), by ships and biomass burning [64, 60]. The highest concentrations of SO_2 were observed in Essex, largely due to the impact of emissions from Michigan.

Nitric acid is emitted from industry (manufacturing, printing), farms, wastewater and vehicle emissions [39]. The highest concentrations were observed in Essex and Toronto as they are urban locations with heavy traffic and manufacturing.

In summary, urban locations (Toronto, Essex) had the highest concentrations of particulate matter and its constituents due to high population density, heavy traffic and manufacturing facilities. Across Canada, at rural and agricultural sites, farming (livestock, poultry and fertilizer use) had a significant influence on ammonia and nitric acid emissions which in turn influenced the production of ammonium sulphate and nitrate. Emissions from the United States, particularly in Michigan and Ohio, had a significant impact on Canadian air quality in Essex.

When comparing the RCM from 2003–2009 to 2010–2019, the mean concentration of all constituents has decreased. All sites observed a reduction in mean $PM_{2.5}$ concentration except for the summertime levels of Greater Vancouver, Fraser Valley and Division No. 11 (Edmonton). It is suspected that the increase in summertime particulate matter is due to biomass burning. The ammonia mixing ratio was the only component studied whose mean concentration increased in 2010–2019 compared to 2003–2009, across all sites and seasons.

2.4 Temperature Data

The temperature data was obtained from the Meteorological Service of Canada climate database. The climate database consists of daily temperature observations from climate stations across Canada. The data for each climate station is organized by monthly CSV files. The database also contains station metadata with the location and name of each sampled location.

The mapping between weather station and census division was done based on the station location, and then mean daily temperatures by census division were calculated. The temperature has been known as a strong confounder of the association between air pollution and health outcomes [21, 4]. This requires daily temperature data in the model presented in the next chapters.

3. *Modelling Adverse Health Effects of Air Pollution*

Research to assess the impact of air pollution on human health has been a major topic in epidemiological studies, seeking to understand how factors such as long-term or short-term exposure, type of pollutant, day of the week or weather affect health outcomes. With this information, we can investigate the burden of pollution in order to improve public health and inform air pollution related policy. Additionally, although this is not a topic of this study, air pollution has far-reaching impacts such as: ecosystems, food production, climate change and economic development.

3.1 Literature Review

Early epidemiological studies (1950-1990s)

In 1961, the World Health Organization published an article identifying air pollution as a significant problem on human health [44]. It was a broad report detailing sources of air pollution and identifying historical environmental episodes, primarily smog, which resulted in adverse health effects. This type of observational health effects (i.e., an environmental phenomenon occurred and resulted in excess morbidity or mortality in the population) was common at this point in time (1994, 1979) [2, 58].

In subsequent decades, modernization and the development of continuous monitoring equipment drove the implementation of initiatives, including the creation of the NAPS program in 1969, collecting hourly data at 15 sites across Canada. Throughout the 1980's, there was an increasing number of studies looking at daily mortality and air pollution concentrations which relied upon multiple regression analysis [2, 66, 59, 55, 74].

Initiatives also facilitated the enactment of legislation, such as the Canadian Environmental Protection Act (1988) and Canada-United States Air Quality Agreement (1991) [19]. Significant advancements in data collection, technology and statistical approaches were made. Consequently, epidemiological studies with greater statistical power could be performed.

In 1997, Kelsall *et al.* [52] developed a Poisson regression model to estimate the risk of mortality in Philadelphia while controlling for time, season and weather. The pollutants of interest were total suspended particle and ozone. Similar analyses were done in Lyon, France (1996) [111], London, England (1996) [3] and Milan, Italy (1996) [100], among others. The advantage of this model was that it produced relative risks which were easily interpretable and comparable across regions.

By 1990, Generalized Linear Models [67] and Generalized Additive Models [42, 43], had begun to be used in epidemiological studies [88].

Epidemiological studies in the 2000's

In the early 2000s significant contributions in Europe and the US included two studies: Air Pollution and Health: A European Approach (APHEA 1 and 2) and the National Morbidity and Mortality Study (NMMAPS) [51, 5, 79, 80]. Later, a project with Europe, US and Canada was created, called Air Pollution and Health: a European and North American Approach (APHENA), to combine and standardize the

existing analyses [50]. By this point, the association between particulate matter exposure and cardiovascular or cardiopulmonary health effects was widely acknowledged [82].

The first of this work in early 2000s was in the U.S. from Samet, Dominici, Curriero, Coursac and Zeger [86] which assessed risk estimates of five pollutants across 20 American cities. A generalized additive model (GAM) for city-level estimates and a hierarchical Bayesian model was used to obtain an overall estimate. Confounding was controlled by including a non-parametric smooth function of temperature and dew point, and a day of week term. The analysis was later expanded to include 90 cities [27].

Using data from NMMAPS project, Dominici, McDermott, Zeger, and Samet analyzed the effects of GAM convergence parameters [28]. Bell, Samet and Dominici summarized the history and evolution of air pollution studies [9]. They also discussed a distributed lag model, in which terms in the model are lagged by a specified number of days. Continuing analyses with the NMMAPS project, Peng, Dominici and Louis [78] compared different approaches to control for seasonal and long-term trends in mortality, by comparing the implication of different smooth functions of time and degrees of freedom.

Liu *et al.* used a multiple logistic regression to model the association between ambient air pollutants and adverse birth outcomes in Vancouver, Canada [61]. They found low concentrations of sulfur dioxide, nitrogen dioxide, carbon monoxide, and ozone to be associated with adverse birth outcomes. Later, a national study using data across Canada found a significant association between nitrogen dioxide with small for gestational age and birth weight [95].

Generalized Linear Models were applied to a Canadian city, Hamilton, ON to investigate the effect of socioeconomic characteristics on a risk estimate model [49]. The city was divided into 5 sub-regions and increased mortality was associated with

areas of lower socioeconomic characteristics.

Another Canadian study by Shin *et al.* examined the short-term association between ozone and circulatory mortality across 24 urban regions using generalized additive Poisson models [92]. They found significant associations in the warm season and by sex, with females having a higher risk than males.

PM component studies

More recently, researchers have begun to look at the components of particulate matter to identify which are the most influential. Bell, Dominici, Ebisu, Zegerand Samet analyzed 52 components of $PM_{2.5}$ [8]. They identified seven components which compromised the majority of $PM_{2.5}$: NH_4^+, EC, Organic carbon matter (OCM) , NO_3^-, Si, Na^+, and SO_4^{2-}. They also identified six components which were correlated with $PM_{2.5}$: NH_4^+, SO_4^{2-}, OCM, NO_3^-, Br, and EC.

Peng *et al.* [77] looked at the association between $PM_{2.5}$ constituents and emergency room visits for cardiovascular and respiratory outcomes in people over 65. They found a positive association between elemental carbon (EC) with cardiovascular admissions and an association between organic carbon matter (OCM) with respiratory admissions.

Additionally, Burnett et al. investigated the association between particulate matter components and mortality across eight Canadian cities [97]. They found that sulfate, iron, nickel, and zinc were most strongly associated with mortality. The availability of PM constituent data was limited, and the authors indicated that there were additional significant components not examined in the study. In particular, they suggested that elemental and organic carbon should be measured as they are the largest constituents of $PM_{2.5}$.

In a meta-analysis, Yang *et al.* [109] reviewed 42 papers published after 2008

which identified associations between $PM_{2.5}$ components with morbidity and mortality. Associations between all adverse health effects and black carbon, organic carbon, potassium, sulfate, NH_4^+, zinc and silicon were observed. Associations occurred between cardiovascular effects and sulfate, NH_4^+, zinc and silicon, nickel, vanadium, sodium and iron. Respiratory associations were identified for organic carbon, nitrate, sulfate and vanadium. Identifying which components are most harmful to human health can inform air pollution policies and the understanding of biological processes.

Environmental epidemiology is increasingly being used to study under-represented regions and community-level metrics. Such studies have incorporated satellite measurements and chemical transport models to estimate pollutant concentrations in unmeasured regions [26] and the inclusion of regional or individual metrics such as; socioeconomic status, sex, race, smoking and obesity rates, education, median household income and city greenness [53, 26, 94, 85].

Mortality, seasonality and lagging confounders

Curriero *et al.* [24] investigated the relationship between temperature and mortality in 11 American cities. They found that the greatest mortality occurred in the winter months, however there was a stronger association between mortality and cold temperatures in southern regions, while northern regions showed a stronger association in warm temperatures. In addition to these findings, they showed that temperature lags of 0, 1, 2 or 3 days were the most strongly associated with mortality. Understanding this relationship is valuable for the construction of our models.

Another study in the Neatherlands looked at the seasonal variation of deaths with the aim of assessing why there is an excess of deaths in the winter [63]. Two periods of increased mortality were observed in late January and March. These periods corresponded to the lowest yearly temperatures and influenza season.

More recently, Burr *et al.* [14] supported the use of a lag-0 non-parametric temperature covariates when a smooth function of time is already incorporated in the health effects model.

Burr *et al.* also developed a novel lag-aggregation technique, called synthetic lag [15]. In this approach, components are transformed to the frequency-domain and the cross-spectrum is used to capture the behaviour of individual periodic components so that they can be lagged by targeted amounts in order to phase-align the components and response. The series is then transformed back to the time-domain to form one pollutant series, that has been synthetically lagged.

3.2 Generalized Additive Model

Generalized Linear Model (GLM)

Hastie and Tibshirani set the groundwork for current air pollution models with the development of Generalized Additive Models (GAMs) [42, 43, 88]. GAMs are an extension of Generalized Linear Models (GLMs) [67] in which the response variable is modelled by a linear function of covariates. Formally this is written as [67]:

$$g(\mu) = \mathbf{X}\boldsymbol{\beta} \tag{3.1}$$

$$= \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \tag{3.2}$$

where μ is the expected value of response variable, Y;

$$\mu \equiv E(Y) \tag{3.3}$$

and g is the link function (describing how the expected response is related to the explanatory variables), $\mathbf{X}_i$ are explanatory (or predictor) variable(s) and $\boldsymbol{\beta}$ are regression coefficients, estimated by the model. The estimation of $\boldsymbol{\beta}$ allows the prediction of the response.

The power of GLM's come from the flexibility of the distribution of the response

variable: it does not need to be normally distributed as is required in most linear regression models. In health models, we are generally modelling daily counts of morbidity or mortality, naturally this is the over-dispersed Poisson distribution:

$$Y \sim Poisson(\mu) \tag{3.4}$$

In this case of a Poisson distribution, the typical link function is the log link function;

$$g(\cdot) = log(\cdot) \tag{3.5}$$

To obtain an estimate of $\boldsymbol{\beta}$, we seek to maximize the likelihood function. Due to the generalization of the response variable, this requires solving the maximum log likelihood equations iteratively, a process called iteratively re-weighted least-squares [45]. This can be done in `R` using the `glm()` function, a part of the `stats` package [84].

Generalized Additive Model (GAM)

Hastie and Tribanishae recognized that they could replace the linear predictor in GLMs with an additive smoother. That is, the response variable can be modelled by smooth functions of the explanatory variables. Formally, this is written as;

$$g(\mu) = \beta_0 + \beta_1 Z + \sum_{i=1}^{k} s_i(X_i) \tag{3.6}$$

$$= \beta_0 + \beta_1 Z + s_1(X_1) + \cdots + s_k(X_k) \tag{3.7}$$

where s_i are smooth non-parametric functions, which can be linear or nonlinear. One caveat with this model is the distribution of response variable is required to be from the exponential family (this includes; normal, exponential, chi-squared, Poisson, ...).

There are many advantages to using GAMs, including;

- Ability to capture nonlinear covariate effects.
- Easy interpretation of the model outputs. Due to additivity, the effect and

interpretation of covariate coefficients does not depend on the values of other coefficients.

- Can control the smoothing parameter and degrees of freedom in the predictor functions (i.e., can set smoothing parameters for s_i). This changes the 'wiggliness' and number of knots in the curve.

- A choice in the smoothers used, which for example can include: local regression (loess); smoothing splines; and regression splines.

In `R`, GAMs can be estimated using the `mgcv` package [105, 106] or the `gam` package, among others.

3.3 PM Component Models

A single pollutant GAM for short-term exposure, as presented above, originally developed by Dominici *et al.* [29], and expanded [78, 16] was used. The GAM estimates the relative mortality associated with the pollutant, after accounting for confounders. This estimate can be called a health risk, and is estimated annually or for all years combined. The logarithms of the daily human mortality or morbidity (health effect) are modelled by the additive combination of the pollutant of interest, a smooth function of time (either natural cubic regression splines or Discrete Prolate Spheroidal Sequences with 6 or 12 degrees of freedom per year [16]), daily average temperature and the day of week. This can be formally written as:

$$log(\mu) = \beta_0 + \beta_1 \times \text{pollutant} + s_1(\text{time}, \text{df} = 6 \text{ or } 12) + \text{temperature} + \text{DOW} \quad (3.8)$$

where μ is the mean response, β_0 and β_1 are linear coefficients estimated by the GAM, s_1 is a smooth function of time, temperature is same-day average temperature, DOW is the day-of-week and *pollutant* is the pollutant concentration of interest, possibly

lagged. As discussed in [14], a smooth function is not applied to the temperature term. The estimated value of β_1 is the pollutant specific risk estimate. That is, it is the log-relative risk of an adverse health effect due to a one unit increase in pollutant concentration, accounting for confounders. The risk estimate is a measure of the association between a pollutant exposure and health effect, it does not imply causation.

Historically, Dominici *et al.* [29] used a natural spline with 6 df as the choice of the smooth function of time. More recently, Burr *et al.* [14] demonstrated the 6 df does not remove all of the desired long time scale effects and advocated the use of 12 or 14 df per year. Burr also showed that Discrete Prolate Spheroidal Sequences are much better at filtering the long time scale variation than the natural spline.

The health data, mortality and morbidity counts due to cardiovascular and pulmonary causes for all age groups, were extracted based on ICD-10 codes [107]. The mortality data runs from 1984–2015, inclusive, while the morbidity data spans from 1996–2019.

To run these models in `R`, the `AHItools` package was used. The package contains a series of function to run GAMs based on user specification of parameters, outlined in Table 3.1. The function has parameters as follows;

```
compute_models(cutoffs, seasons, resp.terms, AP.terms, temp.lags,
               estimate.block, time.dfs, selectCDs, yrRange, fNameOut,
               subfolder, local_db_path, gam.control)
```

Model parameter	Description	Example Value
cutoffs	% of allowable missing data	0.5, 0.7
seasons	seasonality cutoff	JanDec (all months), AprSep, OctMar
response terms	health effect response variables	mortality and morbidity counts
AP.terms	air pollution variables	sulphate, nitrate, zinc, silicon, iron, nickel, vanadium, potassium, OC (corrected), OM, EC, TC (corrected), $PM_{2.5}$
temp.lags	temperature lag	daily mean temperature lagged 0, 1, 2 or 3 days
estimate.block	estimation block	Annual, AllUni (all years)
time.dfs	degrees-of-freedom per year for the smooth function of time	6, 12
selectCDs	selected CD	3520 (Toronto)
yrRange	first and last years of data	2010, 2019
fNameOut	file names	
subfolder	subfolder name	
local_db_path	desired file directory path	
gam.control	list of gam parameters	default values are; max_iter = 100, bf.max_iter = 20 epsilon = 1e-09, bf.epsilon = 1e-09

Table 3.1: Model parameters used for one pollutant component models.

As part of this thesis work, the `AHItools` package was extensively modified to facilitate parallel computation. Risk estimates for all model combinations from a CD could be estimated simultaneously, under the constraints of the CPU cores. This drastically reduced the computation time to run the models. For each model combination, two time smothers: natural cubic spline and slp (Discrete Prolate Spheroidal (Slepian) Sequence regression smoothers, from the `AHIsmooth` and `slp` packages by Dr. Wesley Burr) were used such that each combination had two model fits calculated.

3.3.1 Lagging variables

Lags are used to account for the delayed effects of confounders at short timescales. Generally, lags are selected based on which produce the largest risk estimate. It has lead to several guidelines, such as the maximum association for ozone, O_3, is observed at a lag of 1 day for respiratory mortality [93].

3.3.1.1 Discrete Lags

Discrete lags are when a variable is lagged by a specific number of days. They can be applied to the pollutant data itself [22] as well as to the temperature [24]. Typically, temperature lags of 0, 1, 2 or 3 days are used as they are the most strongly associated with mortality [24]. Pollutant lags of 0, 1 or 2 days are most commonly used.

The discrete lags manifest in the GAM equation as follows:

$$\begin{aligned} log(\mu) =& \beta_0 + \beta_1 \times \text{pollutant}_{\text{lag } t_1} + s_1(\text{time}, \text{df} = 6 \text{ or } 12) + \\ & \text{temperature}_{\text{lag } t_2} + \text{DOW} \end{aligned} \tag{3.9}$$

where $\text{pollutant}_{\text{lag } t_1}$ is the pollutant data with a lag of t_1 days and $\text{temperature}_{\text{lag } t_2}$ is the temperature data with a lag of t_2 days.

3.3.1.2 Synthetic Lags

As described in Section 3.1, Burr *et al.* developed a novel lag-aggregation technique, called synthetic lag [15]. The method applies principles of spectrum estimation to lag individual components by targeted amounts. Code to apply this method is part of the `AHItools` package.

The lag manifests in the model as:

$$\begin{aligned} log(\mu) =& \beta_0 + \beta_1 \times \text{pollutant}_{\text{synth lag}} + s_1(\text{time}, \text{df} = 6 \text{ or } 12) + \\ & \text{temperature}_{\text{lag } t_2} + \text{DOW} \end{aligned} \tag{3.10}$$

where, $\text{pollutant}_{\text{synth lag}}$ is the pollutant series with synthetic lag applied. The discrete lag works by applying a lag of an integer number of days (e.g. 1, 2 days), whereas the synthetic lag can manifest as a non-integer lag (e.g. 1.2, 2.8 days). Thus, the synthetic lag can capture missed association and has been shown to, in general, produce a more positive result [15].

3.4 Particulate Matter Risk Estimates

For this study, association refers to risk estimates obtained from Generalized Additive Models, modeling mean log health counts against additive combinations of daily pollutants and confounding variables, equation 3.8. When risk estimates are discussed as being "similar", we mean similar trends over time: peaks and troughs in the same years with the similar relative magnitudes. To look at these trends between different components, scaled risk estimates were used: the risk estimates were scaled by the average concentration of that pollutant component (with all concentrations measured in μg/m^3). This is to reflect the relative proportion and relative impact of each compound, as the components additively form the particulate mass.

In Section 3.4.1, we will examine changing parameters in the GAM for risk estimates in Toronto, ON. In Section 3.4.2, we will explore the risk estimates for constituents of PM across the eight CDs previously selected in Section 2.3.2.

3.4.1 Exploring GAM Parameters for PM2.5 Risk Estimates

We will explore associations based on the choice of the smooth function of time, seasonality, lagging and morbidity/mortality. $PM_{2.5}$ and ozone, O_3, concentrations from the updated `AHItools` package is used, with the ozone risk estimates presented in Appendix B. The analyses were performed for all 52 CDs from 1984 – 2019, but Toronto, ON was selected for demonstration. The cutoff for proportion of allowable missing data of was 50% (i.e., `cutoffs = 0.5` in the `compute_models()` function) and the default `gam.control` parameters were used.

To begin, we will look at the risk estimates for Toronto. Toronto, Ontario is a large city in the center-east of Canada, with a population of 2.1-2.9 million across the time span of available data. Toronto sits in the middle of an extended metropolitan

area ("The Greater Toronto Area") with population near 10 million, covering over 7,000 square kilometers of land.

Smooth function of time

We will first consider the choice of smooth function of time and degrees of freedom with seasonality. We modelled the logarithms of the daily mortality counts (both pulmonary and circulatory) by the additive combination of the average daily $PM_{2.5}$ concentration, the day of the week, the average daily temperature (lag 0) and a smooth function of time. That is:

```
Health outcome  ~  PM 2.5 (lag0) + Day Of Week (DOW) +
                                Temperature (lag0) + s(time, df = 6,12)
```

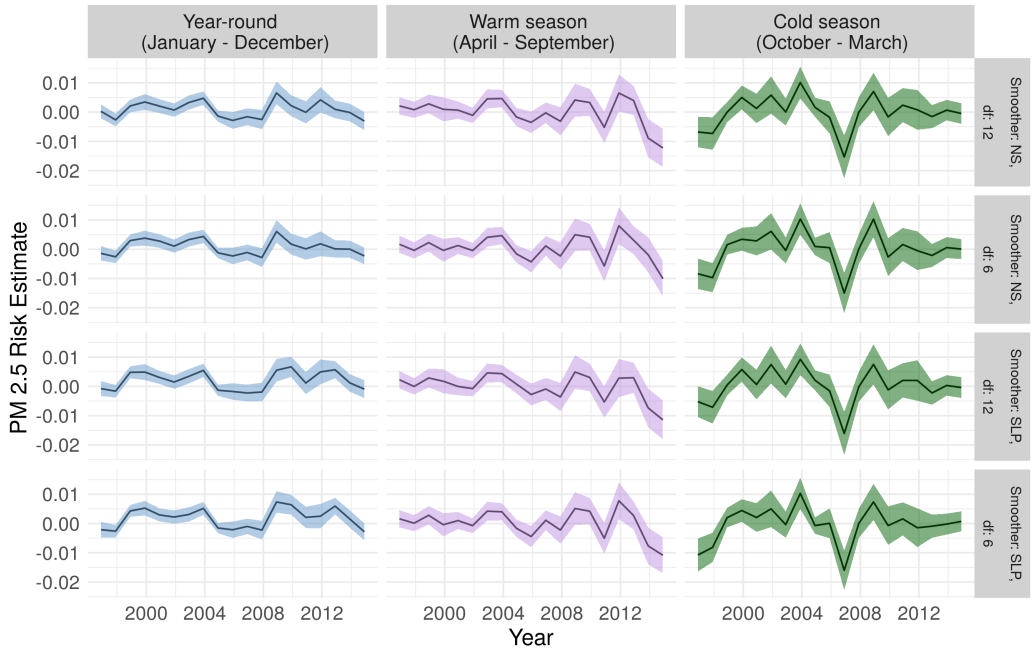


Figure 3.1: Annual **mortality** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year. No temperature lag is used.

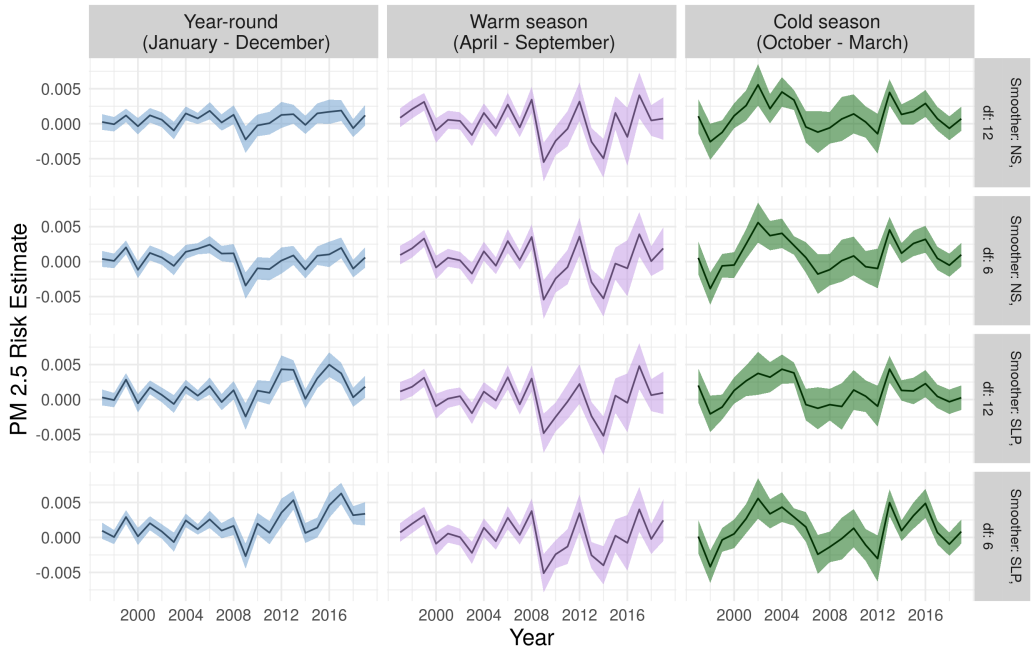


Figure 3.2: Annual **morbidity** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year. No temperature lag is used.

There is considerable variation in the risk estimates between seasons (columns in Figure 3.1), but not amongst the smooth functions of time and degrees of freedom (rows in Figure 3.1). For the remainder of the analyses, a SLP with 6 df/year will be used.

In the Canadian cold season, there is a large fluctuation in the risk estimates across the years, with a considerable decrease in 2007. The all year and warm season risk estimates are more alike, with peaks and troughs occurring around the same time, with similar magnitudes. The all year risk estimates are more consistent than the seasonal breakdowns. This seasonality illustrates that weather has an effect on the association between mortality and $PM_{2.5}$ concentrations.

Lagging pollutant concentrations

The relationship does change with different lags applied to the PM data. Figure 3.3 shows seasonal differences for discrete pollutant lags (0, 1 and 2 days), and synthetic lag, using a Slepian smooth function of time. The following is being modelled:

```
Health outcome  ~  PM 2.5 (lagX) + Day Of Week (DOW) +
                                Temperature (lag0) + SLP(time, df = 12)
```

$PM_{2.5}$ risk estimates for the synthetic lag are not available before 1998 and after 2014 (inclusive) as there isn't enough air pollution data to surpass the cutoff of 50%.

The first row of Figure 3.3 is showing the same data as the third row of Figure 3.1; the $PM_{2.5}$ concentration with no lag.

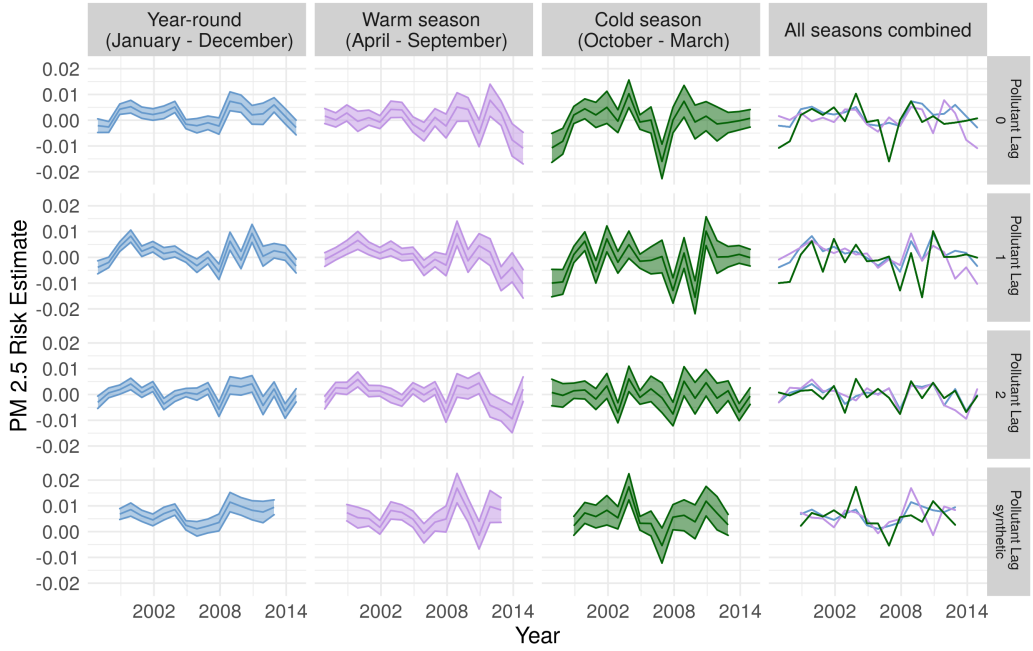


Figure 3.3: Annual **mortality** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for four lags on the pollutant data (0 days; 1 day; 2 days; and synthetic)

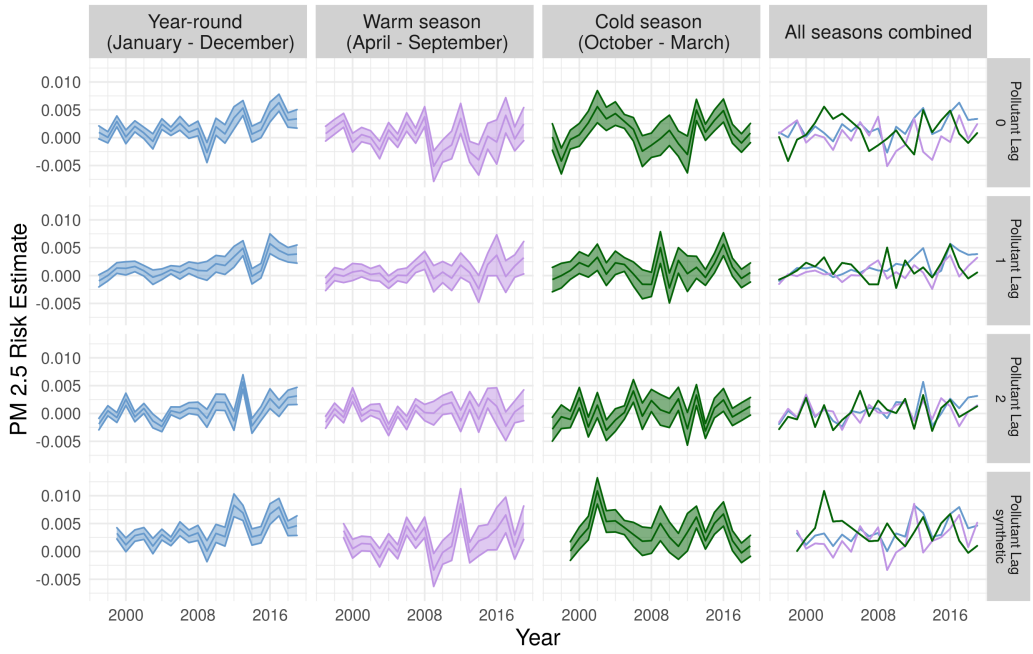


Figure 3.4: Annual **morbidity** (both cardiovascular and pulmonary) risk estimates for $PM_{2.5}$ in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for four lags on the pollutant data (0 days; 1 day; 2 days; and synthetic)

When a discrete lag is applied to the pollutant concentration, either a lag of 1 or 2 days, the variation in the risk estimates increases and the warm and all year estimates are no longer as aligned. The largest fluctuations in the risk estimate occurs when the pollutant is lagged by 1 day. The synthetic lag risk estimates have a larger magnitude and are not as variable.

Health effect cause

We can also look at the daily morbidity and mortality counts for cardiovascular and pulmonary issues, individually. The pulmonary estimates are much more erratic and have a greater magnitude than the cardiovascular and the combined estimates, shown in Figure 3.5. This indicates that deaths due to pulmonary causes are more associated with $PM_{2.5}$ than due to cardiovascular and both causes combined. Another

interesting note is that the estimates for cardiovascular and both causes (cardiovascular and pulmonary) have similar trends; that is, peaks and troughs in the same years with similar values, which are all positive.

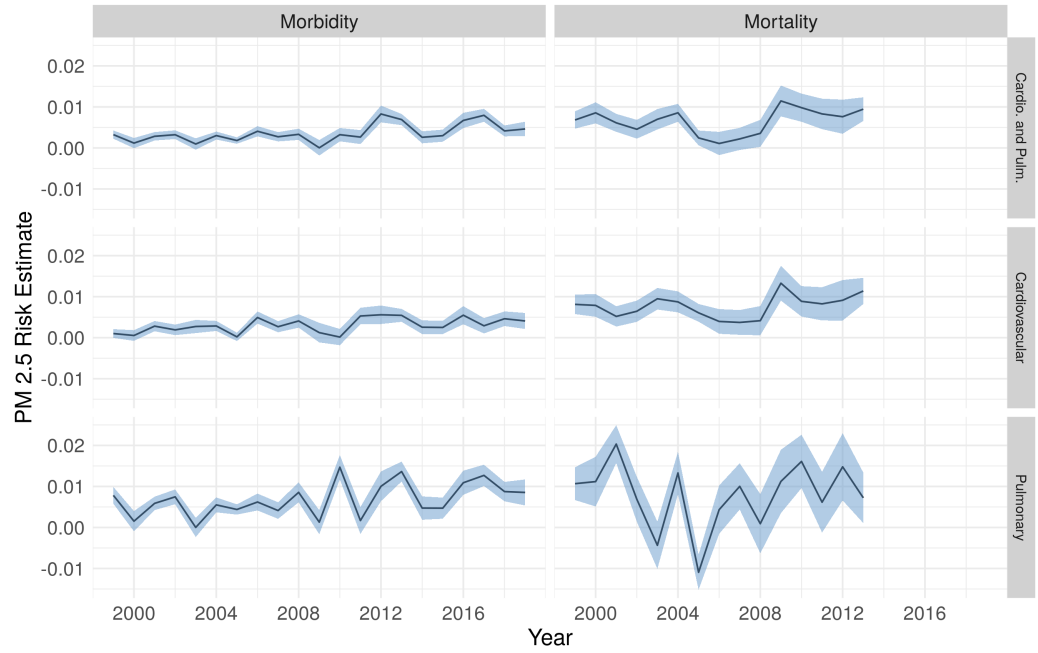


Figure 3.5: Annual health risk estimates, by cause, for $PM_{2.5}$ in Toronto, ON, for three temperature lags (0 days; 1 day; and 2 days) for all year synthetically lagged pollutant concentrations.

All years combined risk estimate

The all-year mortality risk estimates (that is, one estimate for all years of data) is in Table 3.2. When looking down a column (pollutant lags) in Table 3.2 by season, the risk estimates are decreasing, particularly for a temperature lag of 1 day. There is variation due to seasonality and due to the type of pollutant lag. For the synthetic lag, a slight increase in estimates is observed as the temperature lags increase (across a row). No trends are observed for the discrete pollutant lags.

Season	Pollutant Lag	Mortality	Morbidity
	synthetic	5.21 ± 0.7	2.07±0.28
All year	0 days	0.67 ± 0.6	0.57±0.25
(January - December)	1 day	0.68 ± 0.57	0.45±0.24
	2 days	0.48 ± 0.53	0.07±0.23
	synthetic	4.93 ± 0.96	1.73±0.39
Warm season	0 days	0.28 ± 0.83	0.47±0.35
(April - September)	1 day	0.83 ± 0.72	0.24±0.31
	2 days	0.69 ± 0.66	0.12±0.28
	synthetic	5.7 ± 1.26	3.01±0.45
Cold season	0 days	-0.11 ± 1.1	1.07±0.44
(October - March)	1 day	-0.34 ± 1.04	1.01±0.42
	2 days	0.12 ± 0.97	0.09±0.39

Table 3.2: All-year (x 10^{-3}) risk estimates for $PM_{2.5}$, in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) with a SLP smooth function of time using 12 degrees of freedom per year and varying pollutant and temperature lags.

Table 3.2 shows the all-year mortality and morbidity risk estimates. When looking down a column (pollutant lags) in Table 3.2, the risk estimates are decreasing. Hence, a synthetic lag produces the largest risk estimate, across all seasons.

Discussion

In conclusion, in this study, the choice of the smooth function of time was used to remove the long time scale variation - for this reason, a SLP with 6 d.f. was used in analyses. It was observed that seasonality had an effect on the association between mortality and $PM_{2.5}$ concentrations, with the most erratic estimates being calculated in the cold season. The largest risk estimates occur with a synthetic lag. There were

significant differences between the morbidity and mortality risk estimates, where both estimates were highly erratic. Pulmonary deaths and morbidity are associated with $PM_{2.5}$ concentrations.

3.4.2 PM Component Risk Estimates

Single pollutant models were used to identify the association between component pollutants and adverse health effects over time at a single location. Thirteen components of $PM_{2.5}$ were selected including: Sulphate, Nitrate, Zinc, Silicon, Iron, Nickel, Vanadium, Potassium, Organic Carbon (OC, corrected), Organic Matter (OM), Elemental Carbon (EC), Total Carbon (TC, corrected); and models were also fit for $PM_{2.5}$ mass (as in Section 3.4.1). The analyses were performed for eight CD's from 2003 – 2019 (a reduced number due to available data). For this work, the integrated data, described in Section 2.3 was used.

3.4.2.1 PM Constituents

Before diving into the PM component risk estimates, we will first look at the constituents that make up the particulate matter mass. This will help us identify which of the selected components compose the bulk of the PM mass, and will assist in understanding subsequent analyses.

Breakdown of PM by Proportion

The breakdown of $PM_{2.5}$ into its components by CD is shown in Figure 3.6. Components with a very small proportion (such as Nickel, Vanadium, Zinc, Potassium, Iron and Silicon) cannot be seen in Figure 3.6. The carbonaceous compounds (OM, TC, OC and EC) compose a significant portion of the total $PM_{2.5}$ mass. Sulphate and nitrate compose the next largest portion.

There are seasonal and geographical differences in the composition. The presence of sulphate is larger in eastern Canada compared to western Canada. Nitrate concentrations are significantly larger in the cold season (October – March) than the warm season (April – September). These findings are consistent with those observed when looking at the reconstructed mass, Section 2.3.

The average percentages, by site and season, are provided in Table 3.3. Organic matter, total carbon (corrected), organic carbon (corrected), sulphate, elemental carbon and nitrate compromise approximately 98.5 % of the mass.

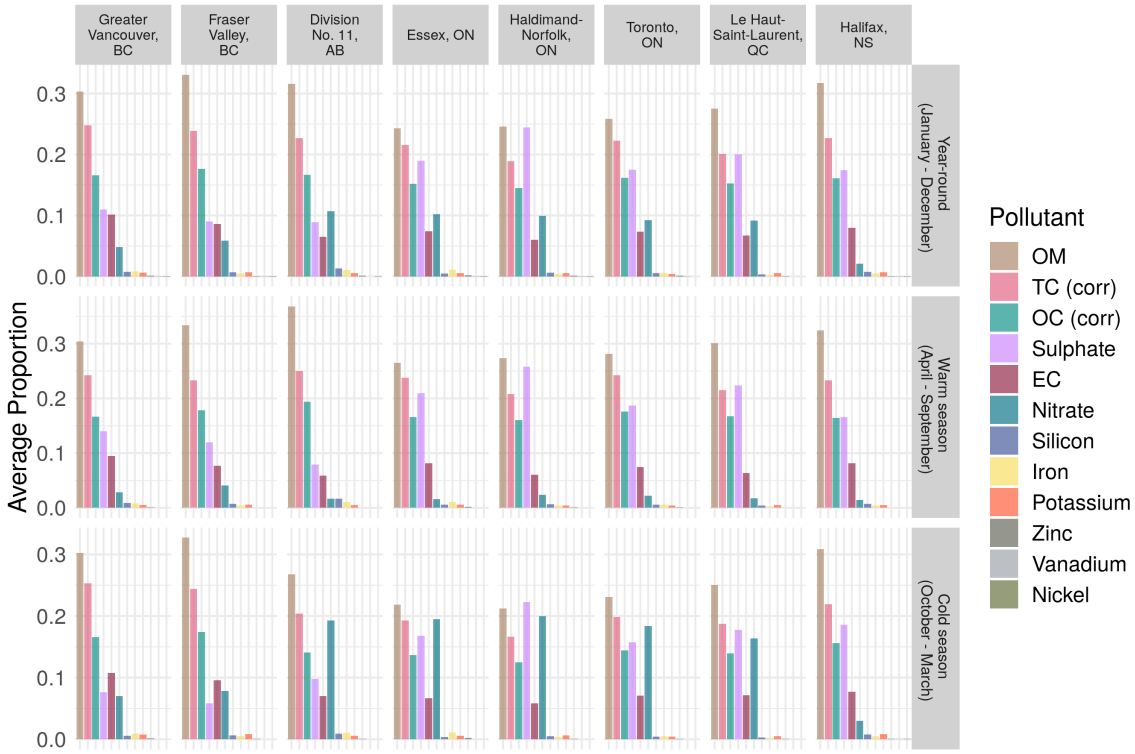


Figure 3.6: Proportional composition of components in $PM_{2.5}$ mass by season and census division.

Pollutant	All year proportion (%)	Warm season proportion (%)	Cold season proportion (%)
OM	28.6	26.5	30.6
TC (corr)	22.1	20.8	23.3
OC (corr)	16.0	14.8	17.1
Sulphate	15.9	14.3	17.3
Nitrate	7.8	13.9	2.2
EC	7.6	7.7	7.4
Silicon	0.7	0.6	0.8
Iron	0.6	0.6	0.6
Potassium	0.6	0.6	0.5
Zinc	0.1	0.1	0.1
Nickel	0.03	0.03	0.03
Vanadium	0.02	0.02	0.02

Table 3.3: Average pollutant proportions across all CD's in the study by season. Table is sorted by highest all year proportion.

Component Correlations

Figure 3.7 shows correlation coefficients between $PM_{2.5}$ and its components, for each CD by season (i.e., each set of seasonal data, for a specific location, and a specific component, is considered against $PM_{2.5}$ for the same). Of these, high correlations were observed between $PM_{2.5}$ and: sulphate; OC (corrected); OM; EC; and TC (corrected). For all CD's, a higher correlation coefficient between $PM_{2.5}$ and both nitrate and zinc was observed for the warm season (April – September) compared to both the cold season (October – March) and all months. In Greater-Vancouver, BC, a higher correlation was observed between $PM_{2.5}$ and all components in the warm season (April – September) compared to the cold season (October – March) and all months (January – December).

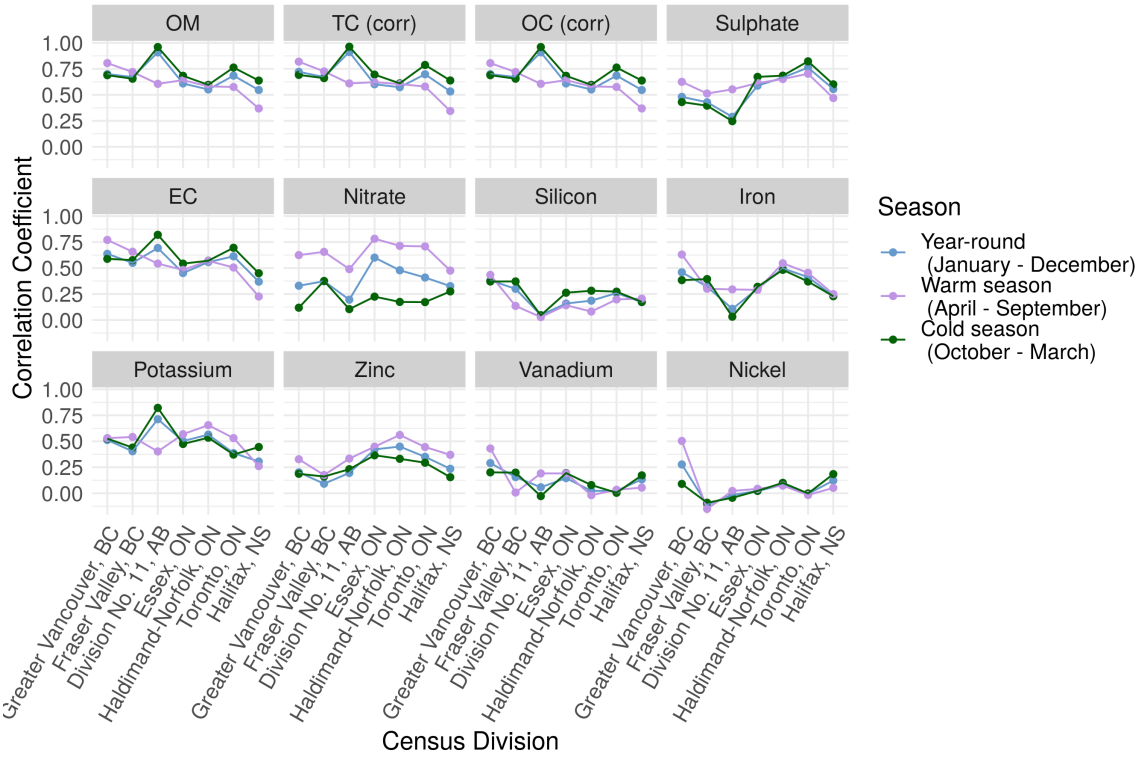


Figure 3.7: Correlation coefficient between $PM_{2.5}$ and its components, by season, for CD's organized from east to west. Note that the lines are visual aids, and do not signify series.

3.4.2.2 Risk Estimates for Toronto - CD 3520

To begin, we examine the largest city in Canada (Toronto, ON), followed by an examination of all relevant CD's in Section 3.4.2.3. The scaled risk estimates for daily cardio-pulmonary morbidity and mortality across all age groups for Toronto are shown in Figures 3.8 and 3.9. The risk estimates were scaled by the average concentration of that pollutant component (with all concentrations measured in $\mu g/m^3$). This is to reflect the relative proportion and relative impact of each compound, as the components additively form the particulate mass.

In these figures, broken down by component, with 2.5 micrometer mass as the first entry, there are a few erratic estimates across time. To give context to these annual outliers, we also estimated all-year risk estimates (i.e., one risk estimate for all years

of data), shown in Table 3.4. These all-year estimates reflect the findings in the literature: that, on the whole, the constituent components of particulate matter are themselves also quite associated with human health effects, in statistically significant ways.

The most prevalent association was observed for $PM_{2.5}$ and carbonaceous compounds (EC, OC, OM and TC), nitrate and sulphate. This is especially true when looking at mortality estimates compared to morbidity estimates. When looking at the mortality estimates (Figure 3.8) all components generally track the $PM_{2.5}$ estimates, i.e., the trends follow a similar pattern with minima in 2012 and 2014. However, when looking at the annual morbidity estimates (Figure 3.9) EC, OC (corrected), OM and TC (corrected) all exhibit common trends which *differ* from the $PM_{2.5}$ trend. All compounds appear to have a similar magnitude risk estimates and standard deviations, with the exceptions of nitrate and sulphate.

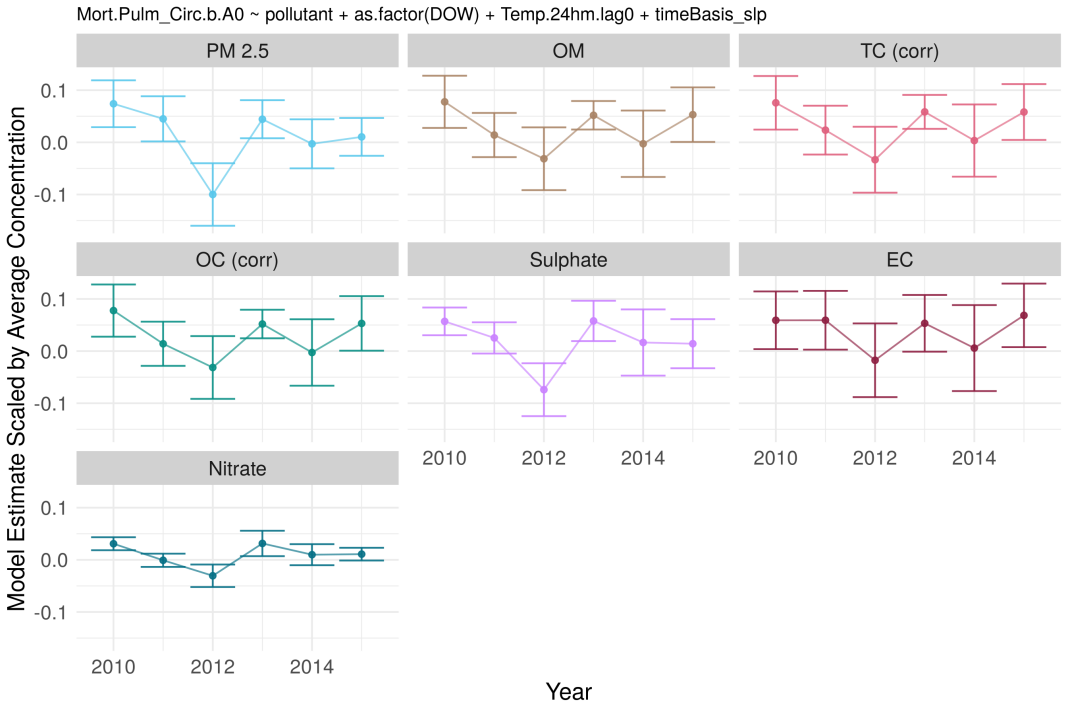


Figure 3.8: Annual scaled model estimates for census division 3520, Toronto, for response variable pulmonary and cardiac **mortality**.

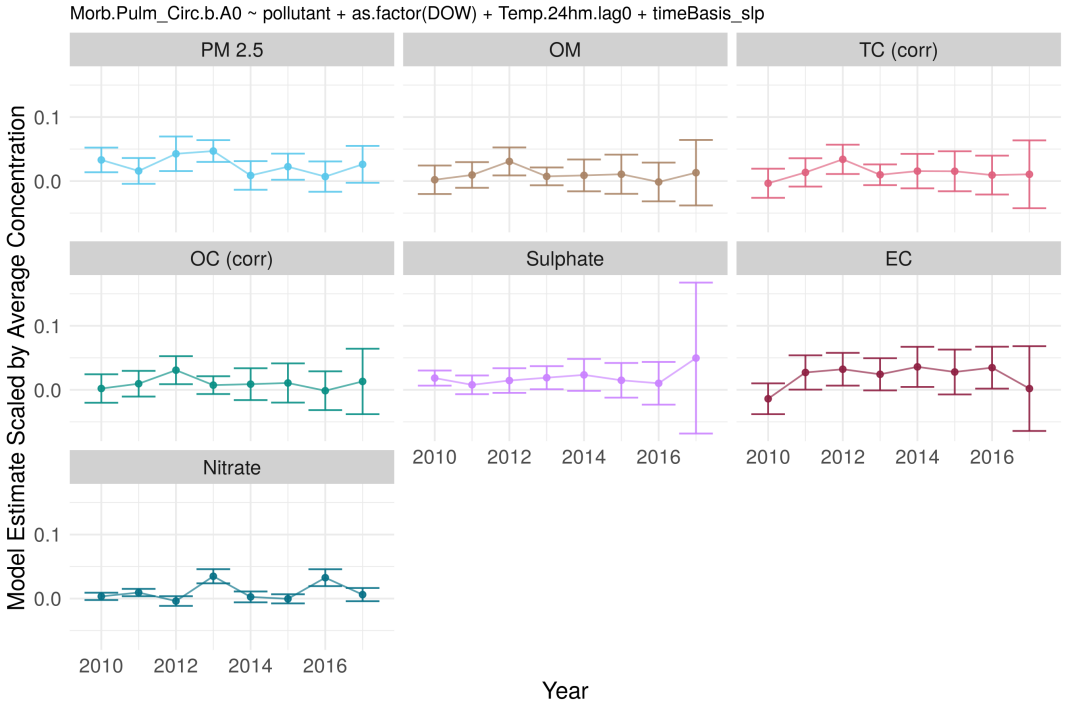


Figure 3.9: Annual scaled model estimates for census division 3520, Toronto, for response variable pulmonary and cardiac **morbidity**.

In Figure 3.9, the large standard error in 2017 is a consequence of component collection stopping halfway through the year.

Health Parameter	Component	Lower CI Bound	Scaled Estimate	Upper CI Bound
	$PM_{2.5}$	1.0	17.0	33.0
Mortality	OM	14.0	31.0	47.0
risk estimate	TC (corr)	16.0	34.0	52.0
	OC (corr)	14.0	31.0	47.0
	Sulphate	4.0	18.0	32.0
	EC	17.0	39.0	62.0
	Nitrate	-2.0	5.0	11.0
	$PM_{2.5}$	4.0	11.0	18.0
Morbidity	OM	1.0	8.0	14.0
risk estimate	TC (corr)	2.0	9.0	17.0
	OC (corr)	1.0	8.0	14.0
	Sulphate	1.0	6.0	12.0
	EC	6.0	15.0	24.0
	Nitrate	-1.0	2.0	4.0

Table 3.4: All-year combined scaled model estimates ($\times 10^{-3}$) for census division 3520, Toronto, with a Discrete Prolate Spheroidal Sequence with 6 degrees of freedom per year for the smooth function of time. Components are sorted with $PM_{2.5}$ first followed by components compromising the highest proportion of $PM_{2.5}$ mass. Lower and Upper bounds are for 95% confidence intervals on the coefficient estimates.

3.4.2.3 Risk Estimates by Census Divisions

The mortality and morbidity risk estimates for Halifax, Toronto, Division No. 11 (Edmonton), the Fraser Valley and Greater Vancouver are shown in Figures 3.10 – 3.11. Haldimand-Norfolk and Essex are not displayed due to their limited availability of data.

The greatest similarity in annual risk estimate trends appears in the three larger urban areas: Vancouver (population $\sim$ 2 million); Division No. 11 (Edmonton, $\sim$

1 million); and Toronto (as mentioned above, $\sim$ 3 million). Both the Fraser Valley and Halifax are smaller areas, with the Fraser Valley being sparsely populated in particular. In general, the larger urban areas show quite tight clustering of annual risks, especially for all months and the cold season (running October-March in Canada). The warm season is more variable, especially in Vancouver, which may indicate differing climate patterns across the different regions.

The morbidity risk estimates (Figure 3.11) showed similar findings, but with greater variability. In the warm season (April–September) the nitrate mortality and morbidity risk estimates are more erratic, particularly in Greater Vancouver, Division No. 11 and Halifax. In fact, nitrate and sulphate showed the least common trending with $PM_{2.5}$ of all observed components in Figures 3.10 and 3.11

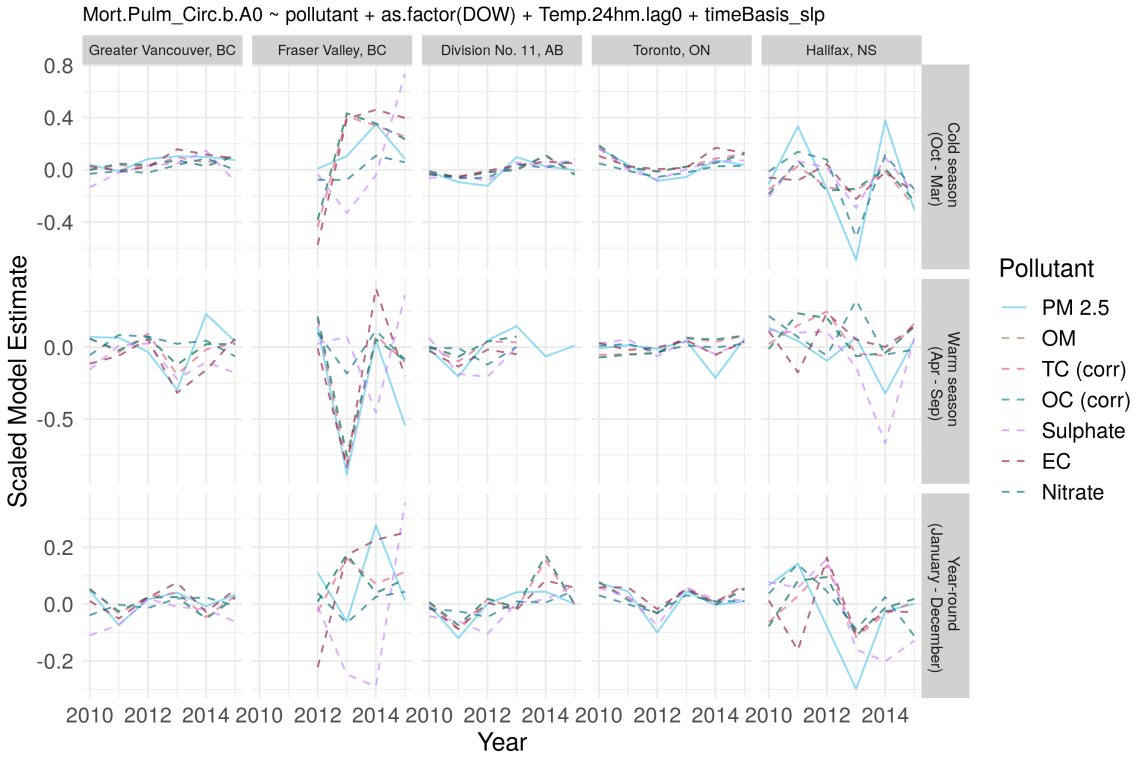


Figure 3.10: Annual scaled mortality risk estimates for $PM_{2.5}$ and components, across five census divisions, for three seasonal breakdowns (all months; cold season; and warm season).

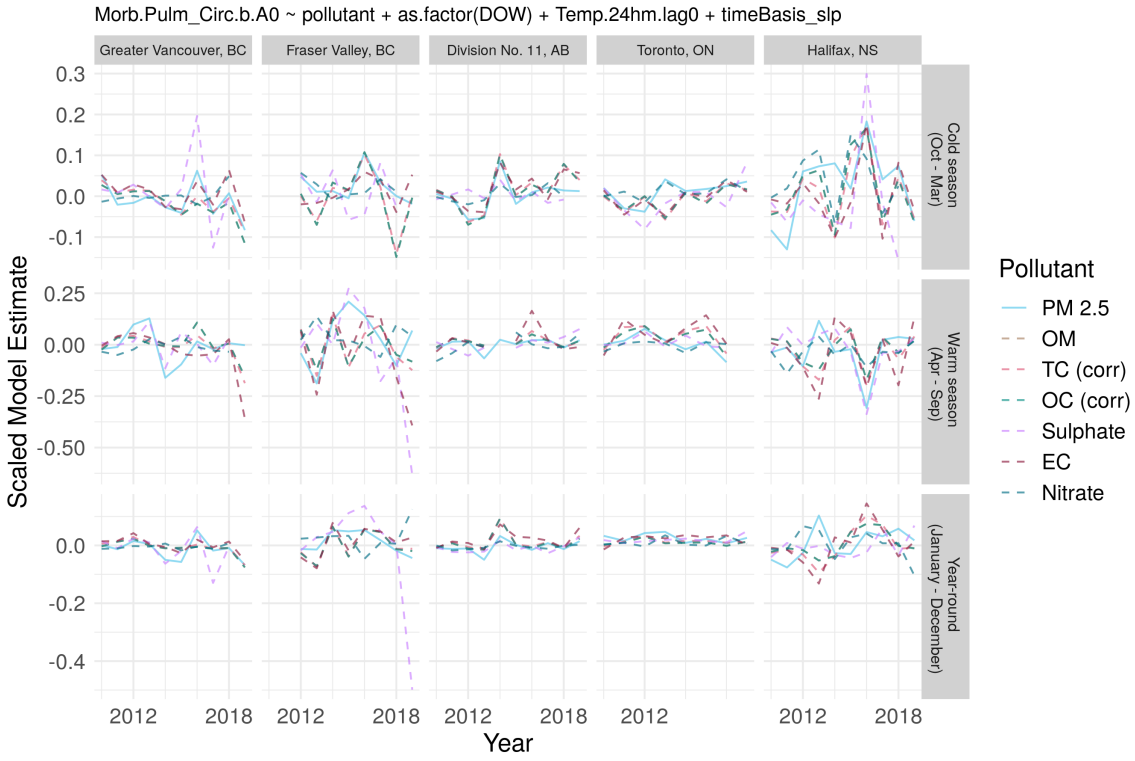


Figure 3.11: Annual scaled morbidity risk estimate for $PM_{2.5}$ and components, across five census divisions, for three seasonal breakdowns (all months; cold season; and warm season).

3.5 Discussion and Conclusions

As with many of these acute risk models, for short(er) time scales the estimates are quite noisy; we see this in the significant differences in the risk estimates for both mortality and morbidity by season. What is most interesting about these sets of models is comparing the trend and behaviour of particulate mass ($PM_{2.5}$) to carbonaceous compounds (EC, OC, OM and TC), particularly for mortality. These findings coincide with the investigation of the proportional composition of $PM_{2.5}$ above, where carbonaceous compounds, sulphate and nitrate made up the bulk of $PM_{2.5}$. The breakdown proportionally being high for these compounds does not necessarily imply

that the proportionality is stationary with time, but the associations being so similar in the risk estimates implies that at least as far as the association between the compounds and human health is concerned, these compounds can be thought of as representative of the whole.

We also see in these results (Figures 3.10, 3.11) some wildly erratic selected years. When working with annual estimates, for multi-day steps in the observations, there really are not very many observations to work with. This, compounded by the known volatility of these risk models to limited time spans, limited observations, low concentration pollutants, and smaller populations, results in limitations on the stability of these estimates. Previous work on development of the Air Health Trend Indicator (e.g., [91]) suggests that a small number of periodic components in a time series can dominate the correlation/association from a regression model, and that their phases are most important for determining this behaviour. However, estimating such periodic components in such limited data availability regimes is quite challenging.

Overall, from this limited Canadian data set we see that particulate matter of 2.5 micrometers or less is well represented temporally by carbonaceous compounds. These compounds then demonstrate associations that are very similar to aggregate particulate matter. These associations suggest that it may be possible to use both sources of information to develop multiple source approaches, using information on overall mass to inform compound observations, and vice versa. This has some philosophical connections to the work being done by remote sensing and ground-truth observation matching: use what observations are available to impute or inform the missing observations of the opposing instrument.

Particulate matter is a highly relevant air pollutant, with a wealth of evidence pointing to a broad range of adverse health effects. However, much of particulate matter pollution is not controllable, or not anthropogenic, and thus we must determine what portion and contribution in the health risk due to air pollution is from

actual anthropogenic (and hence, controllable) sources, to better inform policymakers who seek to balance improvements to the environment and overall societal costs.

4. *Modelling Adverse Health Effects of Two Pollutants*

In this chapter, we explore the development of a multiple pollutant method that includes multiple, possibly correlated, pollutants (in this case, particulate matter with a diameter less than 2.5 μm ($PM_{2.5}$) and ozone (O_3)) with the use of a response surface, and thin plate splines. A portion of this chapter has been accepted for publication in the proceedings of the International Conference on Statistics: Theory and Applications (2022), and is in press.

4.1 Literature Review

As previously discussed, the evidence of association due to both short (acute) and long-term exposures of ambient air pollution on cardiovascular and respiratory health impacts has been examined and found to be significant [73]. More recently, studies examining human health effects associated with multiple pollutants simultaneously has been investigated, as a more realistic framework for actual exposure. Several recent attempts (and successes) include [10, 75, 41], where the two project teams explored models, mostly in Bayesian frameworks, for assessing the statistical effect of multiple air pollution constituents and unknown numbers of major sources.

In 2010, Mauderly *et al.* presented key issues in multi-pollutant models including;

interactions between pollutants, relationship between exposure and response, individual exposure, pollutant data availability and the accuracy of health data [65]. They reasoned that improvements in models and data could make multi-pollutant modelling more feasible in the coming decade. Since then, Bobb *et al.* [10] used a kernel function to incorporate a mixture of air pollution constituents in a Bayesian regression model. A hierarchical approach with prior specification about the pollutants was used for correlated pollutants. Other approaches involved investigating the characteristics of air pollution mixtures that effects health. Under a grant from the Health Effects Institute, Park *et al.* used a hierarchical model to incorporate source-specific health effects [76]. This approach provided improved exposure estimates and allowed for predictions of pollutant sources at sites that were not monitored.

Thin plate splines have been applied to the modelling of spatial surfaces in epidemiological studies. Szpire *et al.* used a two-stage analysis and a thin plate spline to model pollutant exposure across the region of study [96]. This spatial exposure surface was then used to estimate health effects. Additionally, Yanosky *et al.* used a thin plate spline and generalized additive mixed model to predict monthly average PM concentrations across the United States, at high resolution [110]. The bivairate thin plate spline was used to model spatial variability over monthly time periods. The model accounted for covariates including; average wind speed, temperature, total precipitation, and air stagnation. Ettinger *et al.* used a bivariate thin plate spline to model ozone concentrations and found that estimates required less computational time and were more accurate [33].

4.2 Multiple Pollutant Model

We will begin by discussing our motivation for the bivariate model by exploring the correlation between pollutants, namely $PM_{2.5}$ and O_3. The $PM_{2.5}$, O_3 and temperatures for Toronto, ON is shown in Figure 4.1. There are apparent seasonal ozone and temperature patterns, where in the warmer season (May – October) ozone concentrations are higher. Patterns can also be seen for the $PM_{2.5}$ concentrations. An increase of concentrations in warmer seasons is a consequence of atmospheric chemical formation, where O_3 forms from nitrogen oxides via the addition of solar radiation, a process that largely occurs above 17°C [73]. These patterns, and hence correlation, between pollutants compel our exploration into the development of a multi-pollutant model that controls for confounding. They are seen and vary across all CD's in the study. Note that a naive Pearson correlation between the two gives mild correlation only, but careful examination of the time series structure makes it clear that there are deeper relationships – these pollutants cannot be considered independent.

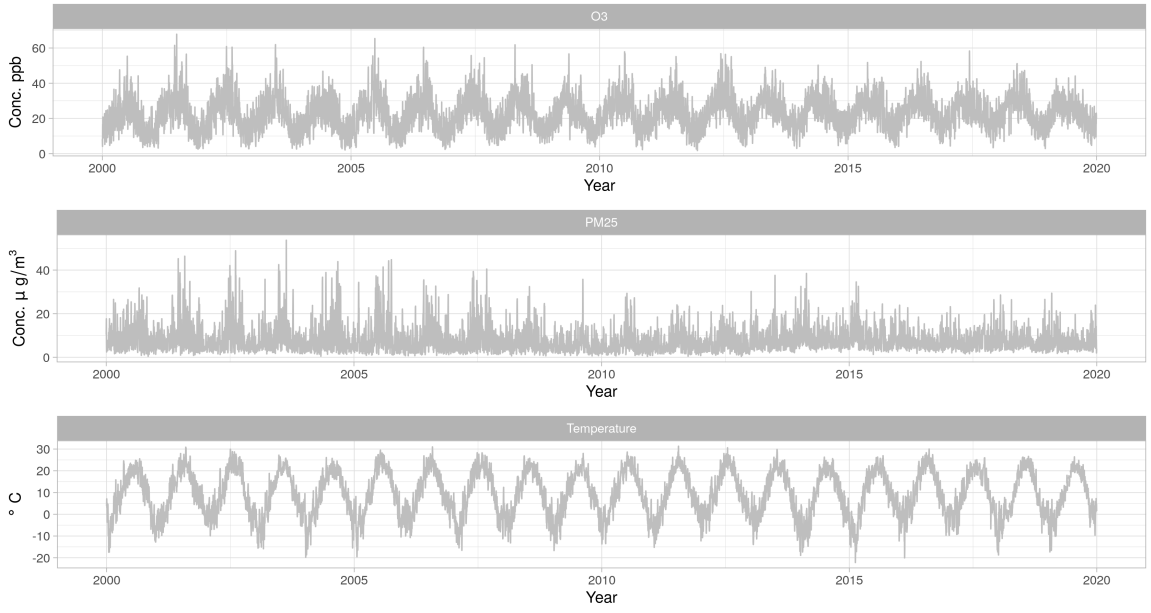


Figure 4.1: Daily ozone (top), $PM_{2.5}$ (middle) and temperature (bottom) measurements for Toronto, ON census division.

4.2.1 Brief Introduction to Thin Plate Splines (TPS)

Thin plate splines [32] are a smooth function of one or more predictor variables, such that the response surface of the spline models the combined effect of multiple predictors. They do not require prior knowledge about the relationship between predictors or specification of knot locations. This flexibility motivates its use in modelling confounding pollutants.

4.2.2 The TPS Bivariate Model

To incorporate multiple pollutants, the single pollutant model, from Equation 3.8, is adjusted to incorporate a TPS and is written as:

$$\begin{aligned} log(\mu) =& \beta_0 + s_{TPS}(\text{pollutant}_1, \text{pollutant}_2) + \\ & s_1(\text{time}, \text{df} = 6 \text{ or } 12) + \text{temperature} + \text{DOW} \end{aligned} \tag{4.1}$$

where s_{TPS} is the thin plate spline, and the smooth functions are the same as Equation 3.8. In this approach, the location on the response surface (after accounting for the temperature and a smooth function of time) that corresponds to the average of a particular pollutant concentration, and this location (average) plus one unit (e.g., ppb) can be determined. In traditional models, the result of interest for these models (e.g., a β_1) is geometrically interpreted as a slope, so we replicate this in the context of the bivariate surface as one of three quantities of interest: the slope of each pollutant at the location of the bivariate average, in the direction of a unit increase in that pollutant; and the slope of the pollutants jointly in the direction of their joint unit increase (taken to be equal contribution, so an increase of 1 unit in a bisecting angle to the two pollutant axes). This approach has the advantage that the interpretation stays the same as the univariate models, allowing easy comparison of the magnitudes and temporal trends.

The thin plate spline surface of the residual mortality is visualized in Figure 4.2.

Grids in the centre plot are used to show the cross-sections (planes corresponding to the average $PM_{2.5}$ and O_3 concentrations). In Figure 4.2, left, the residual mortality TPS is plotted against the O_3 concentration while the $PM_{2.5}$ concentration is held at its mean value, 9.12 μg m^{-3}, while the right shows the TPS residual response surface versus $PM_{2.5}$ concentration while the O_3 concentration is held at its mean value of 23.49 ppb.

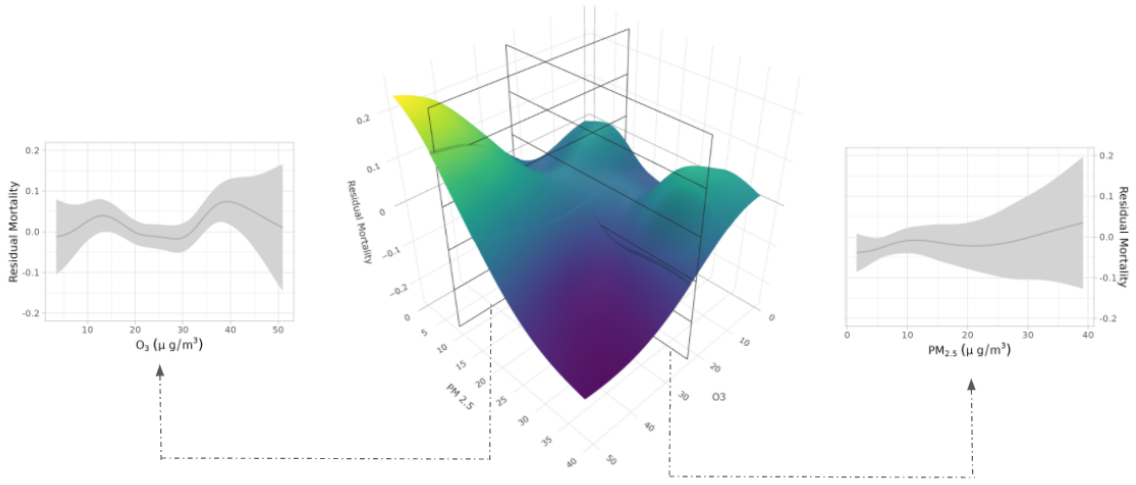


Figure 4.2: Thin Plate Spline residual mortality and cross-sectional plots at the average pollutant concentrations for Toronto, ON throughout 2014. Each of the single pollutant plots shows the surface and the prediction interval along that pollutant's axis, with the other pollutant fixed at its arithmetic overall average value. The response visualized here is the residual log mortality after accounting for temperature, DOW and the smooth function of time.

When looking at the ozone slice of the thin plate spline surface (Figure 4.2, left), the relationship forms an 'M' shaped curve. This is a peculiar finding, as it suggests exposures to both low and high concentrations of ozone (when exposed to the average $PM_{2.5}$ concentration) results in a similar mortality risk. Risk estimates were expected to increase with larger concentrations. This unanticipated relationship of the O_3 slice was observed across multiple years. It suggests that the model is not appropriate and that a constraint on the thin plate spline surface should be investigated so that at

higher pollutant concentrations a larger residual health effect is observed.

4.3 Comparison of Single and Two Pollutant Models

The cardiovascular and respiratory mortality and morbidity estimates from the thin plate spline (TPS) approach were compared to the single pollutant models (Figures 4.3 – 4.5) for Toronto, ON from 1997 to 2019. Plots for Ottawa, Edmonton and Vancouver CDs are included in Appendix C. For these estimates, the smooth function of time was a DPSS (Slepian) spline with 12 degrees of freedom per year. In general, large increases and decreases in $PM_{2.5}$ and O_3 are captured in the TPS estimates. The TPS approach has significantly larger prediction error than either of the two individual models, or their additive combination. This is not entirely surprising, as the correlation structure of the two pollutants is known to vary over time due to component-level correlations between the respective elements of the series.

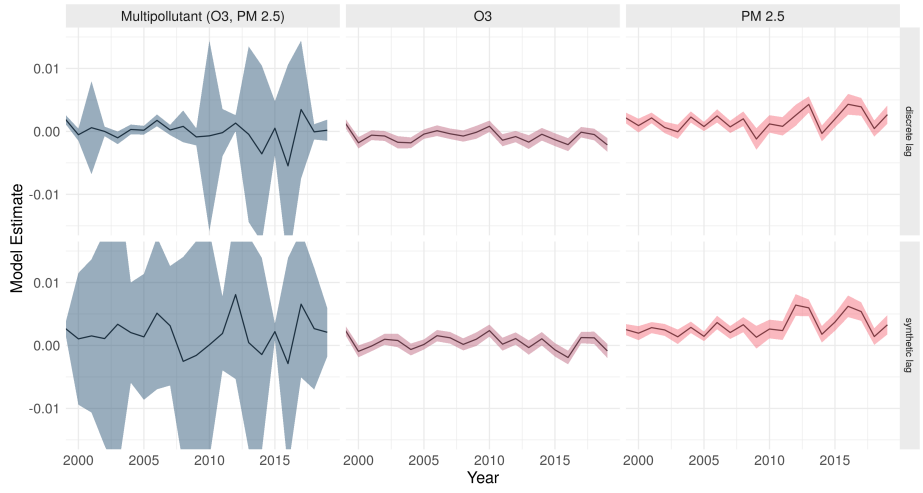


Figure 4.3: Comparison of single and multi-pollutant morbidity models for synthetic and discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON

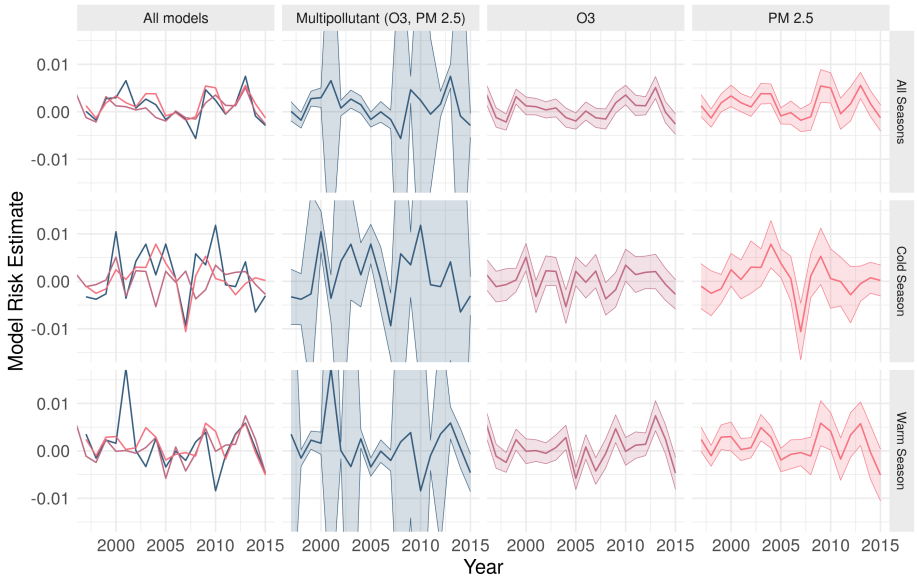


Figure 4.4: Comparison of single and multi-pollutant mortality models for discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON

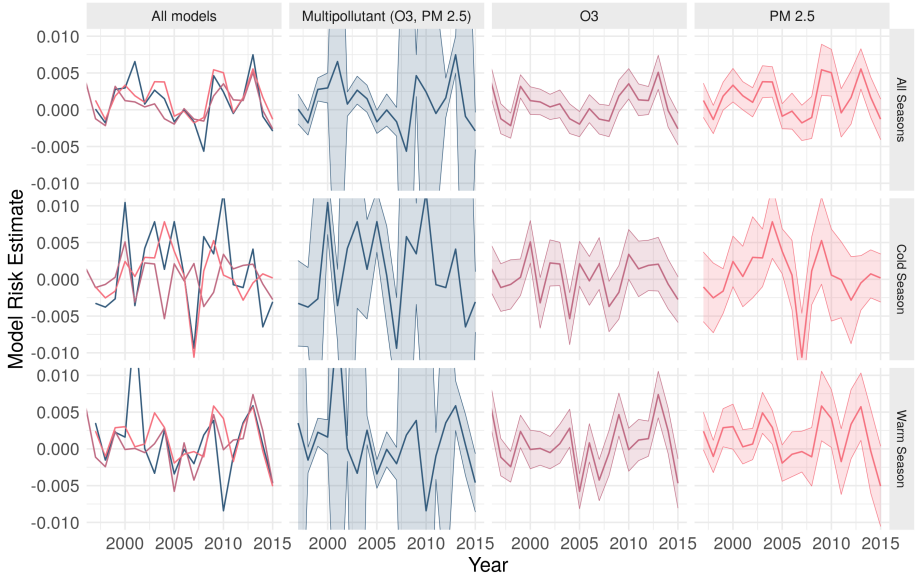


Figure 4.5: Comparison of single and multi-pollutant morbidity models for discrete lag, using an SLP with 12 df for the smooth function of time in Toronto, ON

4.4 Discussion

The mean association, jointly, between the two pollutants is quite similar to the average of two univariate models, but is not identical, demonstrating that modeling both simultaneously does have some value. The use of thin-plate splines is quite flexible, and could allow for further work to more carefully model the correlation and association structure of the pollutants, rather than treating them as simple bivariate contributors. In particular, the extreme swings in confidence for the estimates of the TPS-based model are likely a sign of extreme values and contamination in the estimation of the surface, so robust alternatives which may help control this variation need to be investigated.

The extraction of univariate 'slices' from the bivariate surface is also a technique that demonstrates promise, as these extracted slices are comparable in form to non-linear modeling outputs for the impact of air pollution on human health. In conclusion, this new method shows promise, and allows a new viewpoint on a modeling problem which is of high interest to a broad global community.

5. *Conclusions*

In this thesis, we explored and analyzed Canadian air pollutant data (Chapter 2) for both continuous (hourly) and integrated (daily) data. We began by discussing how the raw pollutant data was cleaned and unified to produce a well-structured and organized database. Scripts used to generate the database, from downloading the pollutant files for select years to addressing file inconsistencies, is publicly available. Additionally, the database themselves for both integrated and continuous data from 1980 – 2019 are available. Data preparation and management accounted for more than half of the work in this thesis. Thus, the accessibility of this data can remove such a burden in future studies. This air pollution data can not only be applied to health studies, but could be used for crop monitoring, investigating climate change and many other domains [17, 99].

We explored seasonal, regional and temporal differences in air pollution concentrations as well as the cause of such differences. Urban locations (Toronto, Essex) had the highest concentrations of particulate matter and its constituents due to high population density, heavy traffic and manufacturing facilities. Overall, we saw a reduction in pollutant concentrations since 2000, except select sites which observed increased summertime levels due to biomass burning. We identified that carbonaceous compounds (elemental carbon and organic matter) comprise a significant portion of the reconstructed mass (RCM) with is proportional contribution increasing in the winter months (October–March).

We used a single pollutant Generalized Additive Model to estimate the short-term association between component pollutants and adverse health effects over time at a single location, while accounting for confounders (Chapter 3). We explored associations based on the choice of the smooth function of time, seasonality, lagging and morbidity/mortality variable. We observed that seasonality had an effect on the association between mortality and $PM_{2.5}$ concentrations, with the most erratic estimates being calculated in the cold season. There were significant differences between the morbidity and mortality risk estimates, where both estimates were highly erratic. There were several limitations in this work, including;

- study locations were selected based on population and are not representative of regions with small populations;
- pollutant exposure measured at NAPS sites was generalized as the population exposure for the census division;
- limited data availability, particularly for PM components which were measured once every 3, 6 or 12 days;
- census division estimates do not reflect individual behaviour and exposure.

In this work we showed that carbons, sulphate and nitrate are the components of $PM_{2.5}$ which may be responsible for the majority of the acute human health effect association as they are representative of the PM risk estimates. We identified that the PM risk estimate can be used to inform the risk estimates of its largest components and vice versa.

We developed a bivariate model (Chapter 4) to model multiple correlated pollutants (in this case, $PM_{2.5}$ and O_3) using response surface and thin plate spline bases.

This method showed promise, and allows a new viewpoint on a modeling problem which is of high interest to a broad global community.

5.1 Future Work

The spatial coverage of this work was limited by the availability of data and census divisions were selected based on their population. Combining *in situ* monitoring data with other sources such as satellite measurements could grossly expand the availability of data.

Following the work of [53, 26, 94, 85], our health effects models could be used to study community-level metrics, with the addition of such variables in the models. Examples of regional or individual metrics includes; regional socioeconomic status, sex, race, smoking and obesity rates, education, median household income and city greenness. Of course, this is quite idealistic and is highly dependent on the availability of such data.

A large source of future work is improvements in the bivariate TPS model. We saw extreme swings and large standard errors in confidence estimates for the estimates of the TPS-based model, which is likely a sign of extreme values and contamination in the estimation of the surface, as the total number of inputs is quite limited given the total area of the surface. As this portion of the work was highly speculative, we leave the solution of this issue to a future research project.

To summarize, key recommendations for future work are as follows;

- incorporate additional source(s) of air pollution data to increase the spatial scope and estimate missing data;
- extend the spatial scope of the study;
- inclusion of additional confounders to the one pollutant model;

- improvements to the bivariate model;
  - investigate the inclusion of parameters to require the residual mortality of the thin plate spline surface to increase as pollutant concentrations increase;
  - apply the model to all years combined (i.e. one model for all available years of data).

Bibliography

[1] Silvia Aimar, Mariano Mendez, Roger Funk, and Daniel Buschiazzo. Soil properties related to potential particulate matter emissions (PM10) of sandy soils. *Aeolian Research*, 3, 01 2012.

[2] H Ross Anderson. Air pollution and mortality: A history. *Atmospheric Environment*, 43(1):142–152, 2009.

[3] H Ross Anderson, Antonio Ponce de Leon, J Martin Bland, Jonathan S Bower, and David P Strachan. Air pollution and daily mortality in London: 1987-92. *British Medical Journal*, 312(7032):665–669, 1996.

[4] Susan C Anenberg, Shannon Haines, Elizabeth Wang, Nicholas Nassikas, and Patrick L Kinney. Synergistic health effects of air pollution, temperature, and pollen exposure: a systematic review of epidemiological evidence. *Environmental Health*, 19(1):1–19, 2020.

[5] Richard W Atkinson, H Anderson, H Ross, Jordi Sunyer, JON Ayres, Michela Baccini, Judith M Vonk, Azzedine Boumghar, Francesco Forastiere, Bertil Forsberg, Giota Touloumi, et al. Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project. *American Journal of Respiratory and Critical Care Medicine*, 164(10):1860–1866, 2001.

[6] Karen Bartko. Lovely morning in the apocalypse: Edmonton wakes up to orange, smoky sky. `https://globalnews.ca/news/4388625/edmonton-apocalypse-sky-wildfire-smoke/`, Aug 2018.

[7] William Battye, Viney P Aneja, and Paul A Roelle. Evaluation and improvement of ammonia emissions inventories. *Atmospheric Environment*, 37(27):3873–3883, 2003.

[8] Michelle L Bell, Francesca Dominici, Keita Ebisu, Scott L Zeger, and Jonathan M Samet. Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. *Environmental Health Perspectives*, 115(7):989–995, 2007.

[9] Michelle L Bell, Jonathan M Samet, and Francesca Dominici. Time-series studies of particulate matter. *Annu. Rev. Public Health*, 25:247–280, 2004.

[10] Jennifer F Bobb, Linda Valeri, Birgit Claus Henn, David C Christiani, Robert O Wright, Maitreyi Mazumdar, John J Godleski, and Brent A Coull. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*, 16(3):493–508, 2015.

[11] Robert D Brook, Barry Franklin, Wayne Cascio, Yuling Hong, George Howard, Michael Lipsett, Russell Luepker, Murray Mittleman, Jonathan Samet, Sidney C Smith Jr, et al. Air pollution and cardiovascular disease: a statement for healthcare professionals from the expert panel on population and prevention science of the American heart association. *Circulation*, 109(21):2655–2671, 2004.

[12] Robert D Brook and Sanjay Rajagopalan. Particulate matter air pollution and atherosclerosis. *Current Atherosclerosis Reports*, 12(5):291–300, 2010.

[13] Wesley S. Burr. tsinterp: A time series interpolation package for R. `https://github.com/wesleyburr/tsinterp`, 2022.

[14] Wesley S Burr and Hwashin H Shin. Accounting for temperature when modeling population health risk due to air pollution. In *Interdisciplinary Topics in Applied Mathematics, Modeling and Computational Science*, pages 105–112. Springer, 2015.

[15] Wesley S Burr, Hwashin H Shin, and Glen Takahara. Synthetically lagged models. *Statistics & Probability Letters*, 144:37–43, 2019.

[16] Wesley S Burr, Glen Takahara, and Hwashin H Shin. Bias correction in estimation of public health risk attributable to short-term air pollution exposure. *Environmetrics*, 26(4):298–311, 2015.

[17] Andrzej Bytnerowicz, Kenji Omasa, and Elena Paoletti. Integrated effects of air pollution and climate change on forests: A northern hemisphere perspective. *Environmental Pollution*, 147(3):438–445, 2007.

[18] California Air Resources Board. Inhalable Particulate Matter and Health (PM2.5 and PM10). `https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health`, 2022. [Online; accessed September 20, 2022].

[19] Environment Canada and Climate Change. Timeline: Major milestones of environment and climate change Canada. `https://www.canada.ca/en/environment-climate-change/campaigns/50-years-environmental-action/eccc-timeline.html`, Dec 2021.

[20] Statistics Canada. Farm and farm operator data. `https://www150.statcan.gc.ca/n1/pub/95-640-x/2016001/article/14816-eng.htm`, July 2020.

[21] Kai Chen, Kathrin Wolf, Regina Hampel, Massimo Stafoggia, Susanne Breitner, Josef Cyrys, Evangelia Samoli, Zorana Jovanovic Andersen, Getahun Bero-Bedada, Tom Bellander, et al. Does temperature-confounding control influence the modifying effect of air temperature in ozone–mortality associations? *Environmental Epidemiology*, 2(1):e008, 2018.

[22] Chiogna, Monica Carlo Gaetan, and Carlo Gaetan. Dynamic generalized linear models with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(4):453–468, 2002.

[23] National Research Council. *Review of the U.S. Department of Energy Office of*

Fossil Energy's Research Plan for Fine Particulates. The National Academies Press, Washington, DC, 1999.

[24] Frank C Curriero, Karlyn S Heiner, Jonathan M Samet, Scott L Zeger, Lisa Strug, and Jonathan A Patz. Temperature and mortality in 11 cities of the eastern United States. *American Journal of Epidemiology*, 155(1):80–87, 2002.

[25] Ewa Dabek-Zlotorzynska, Tom F. Dann, P. Kalyani Martinelango, Valbona Celo, Jeffrey R. Brook, David Mathieu, Luyi Ding, and Claire C. Austin. Canadian national air pollution surveillance (NAPS) PM2.5 speciation program: Methodology and PM2.5 chemical composition for the years 2003–2008. *Atmospheric Environment*, 45(3):673–686, 2011.

[26] Qian Di, Yan Wang, Antonella Zanobetti, Yun Wang, Petros Koutrakis, Christine Choirat, Francesca Dominici, and Joel D Schwartz. Air pollution and mortality in the medicare population. *New England Journal of Medicine*, 376(26):2513–2522, 2017.

[27] Francesca Dominici, Aidan McDermott, Michael Daniels, Scott L. Zeger, and Jonathan M. Samet. Revised analyses of the national morbidity, mortality, and air pollution study: Mortality among residents of 90 cities. *Journal of Toxicology and Environmental Health, Part A*, 68(13-14):1071–1092, 2005. PMID: 16024489.

[28] Francesca Dominici, Aidan McDermott, Scott L Zeger, and Jonathan M Samet. On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156(3):193–203, 2002.

[29] Francesca Dominici, Jonathan M. Samet, and Scott L. Zeger. Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 163(3):263–302, 2000.

[30] Ken Donaldson, Nicholas Mills, William MacNee, Simon Robinson, and David

Newby. Role of inflammation in cardiopulmonary health effects of PM. *Toxicology and Applied Pharmacology*, 207(2):483–488, 2005.

[31] Sara D Dubowsky, Helen Suh, Joel Schwartz, Brent A Coull, and Diane R Gold. Diabetes, obesity, and hypertension may enhance associations between air pollution and markers of systemic inflammation. *Environmental Health Perspectives*, 114(7):992–998, 2006.

[32] Jean Duchon. Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, pages 85–100. Springer, 1977.

[33] Bree Ettinger, Serge Guillas, and Ming-Jun Lai. Bivariate splines for ozone concentration forecasting. *Environmetrics*, 23(4):317–328, 2012.

[34] Neil H Frank. Retained nitrate, hydrated sulfates, and carbonaceous mass in federal reference method fine particulate matter for six eastern US cities. *Journal of the Air & Waste Management Association*, 56(4):500–511, 2006.

[35] Barry A Franklin, Robert Brook, and C Arden Pope III. Air pollution and cardiovascular disease. *Current Problems in Cardiology*, 40(5):207–238, 2015.

[36] Nannan Gao, Chunhou Li, Jiadong Ji, Yanli Yang, Shaoting Wang, Xinlun Tian, and Kai-Feng Xu. Short-term effects of ambient air pollution on chronic obstructive pulmonary disease admissions in Beijing, China (2013–2017). *International Journal of Chronic Obstructive Pulmonary Disease*, 14:297, 2019.

[37] Andrew J Ghio and Yuh-Chin T Huang. Exposure to concentrated ambient particles (CAPs): a review. *Inhalation Toxicology*, 16(1):53–59, 2004.

[38] Éric Giroux, Helmut Roth, Dazhong Yin, and Weimin Jiang. Modeling and processing ammonia emissions for particulate matter studies in the lower Fraser Valley. In *Proceedings of the EI11 - Emission Inventories - Partnering for the Future - Conference*, pages 1–11, 2011.

[39] Australian Government. Nitric acid. `http://www.npi.gov.au/resource/`

nitric-acid, July 2020.

[40] Sandie Ha, Hui Hu, Dikea Roussos-Ross, Kan Haidong, Jeffrey Roth, and Xiaohui Xu. The effects of air pollution on adverse birth outcomes. *Environmental Research*, 134:198–204, 2014.

[41] Tavoos Hassan Bhat, Guo Jiawen, and Hooman Farzaneh. Air pollution health risk assessment (AP-HRA), principles and applications. *International Journal of Environmental Research and Public Health*, 18(4):1935, 2021.

[42] Trevor J. Hastie and Robert J. Tibshirani. Generalized additive models. *Statistical Science*, 1(3):297–310, 1986.

[43] Trevor J. Hastie and Robert J. Tibshirani. *Generalized additive models*. Chapman & Hall, 1990.

[44] Harry Heimann. Effects of air pollution on human health. *Air Pollution, World Health Organisation Monograph Series*, (46):159–220, 1961.

[45] Paul W Holland and Roy E Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods*, 6(9):813–827, 1977.

[46] Mohd Faiz Ibrahim, Rozita Hod, Azmawati Mohammed Nawi, and Mazrura Sahani. Association between ambient air pollution and childhood respiratory diseases in low-and middle-income Asian countries: A systematic review. *Atmospheric Environment*, 256:118422, 2021.

[47] Shannon Jarvis. shannonjarvis9/naps-continuous-data: Naps continuous data. https://doi.org/10.5281/zenodo.6901267, July 2022.

[48] Shannon Jarvis. shannonjarvis9/NAPS-integrated-data: NAPS Integrated Data. https://doi.org/10.5281/zenodo.7097402, September 2022.

[49] Michael Jerrett, RT Burnett, J Brook, P Kanaroglou, C Giovis, N Finkelstein, and B Hutchison. Do socioeconomic characteristics modify the short term association between air pollution and mortality? evidence from a zonal time series in

hamilton, canada. *Journal of Epidemiology & Community Health*, 58(1):31–40, 2004.

[50] Klea Katsouyanni, Jonathan M Samet, H Ross Anderson, Richard Atkinson, Alain Le Tertre, Sylvia Medina, Evangelia Samoli, Giota Touloumi, Richard T Burnett, Daniel Krewski, et al. Air pollution and health: a European and North American approach (APHENA). *Research Report (Health Effects Institute)*, (142):5–90, 2009.

[51] Klea Katsouyanni, Joel Schwartz, C Spix, Giota Touloumi, D Zmirou, A Zanobetti, B Wojtyniak, JM Vonk, A Tobias, A Pönkä, et al. Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S12–S18, 1996.

[52] JE Kelsall, Jonathan M Samet, SL Zeger, and J Xu. Air pollution and mortality in Philadelphia, 1974–1988. *American Journal of Epidemiology*, 146(9):750–762, 1997.

[53] Marianthi-Anna Kioumourtzoglou, Joel Schwartz, Peter James, Francesca Dominici, and Antonella Zanobetti. PM2.5 and mortality in 207 US cities: modification by temperature and city characteristics. *Epidemiology (Cambridge, Mass.)*, 27(2):221, 2016.

[54] Itai Kloog, Bill Ridgway, Petros Koutrakis, Brent A Coull, and Joel D Schwartz. Long-and short-term exposure to PM2.5 and mortality: using novel exposure models. *Epidemiology (Cambridge, Mass.)*, 24(4):555, 2013.

[55] Edward L. Korn and Alice S. Whittemore. Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, 35(4):795–802, 1979.

[56] Michal Krzyzanowski and Aaron Cohen. Update of WHO air quality guidelines. *Air Quality, Atmosphere & Health*, 1(1):7–13, 2008.

[57] Anthony Lau, Shabtai Bittman, and Derek Hunt. Development of ammonia

emission factors for the land application of poultry manure in the lower Fraser Valley of British Columbia. *Canadian Biosystems Engineering*, 50, 01 2008.

[58] Lester B. Lave and Eugene P. Seskin. Air pollution and human health. *Science*, 169(3947):723–733, 1970.

[59] Lester B Lave and Eugene P Seskin. An analysis of the association between US mortality and air pollution. *Journal of the American Statistical Association*, 68(342):284–290, 1973.

[60] Chulkyu Lee, Randall V Martin, Aaron van Donkelaar, Hanlim Lee, Russell R Dickerson, Jennifer C Hains, Nickolay Krotkov, Andreas Richter, Konstantine Vinnikov, and James J Schwab. SO2 emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations. *Journal of Geophysical Research: Atmospheres*, 116(D6), 2011.

[61] Shiliang Liu, Daniel Krewski, Yuanli Shi, Yue Chen, and Richard T Burnett. Association between gaseous ambient air pollutants and adverse pregnancy outcomes in Vancouver, Canada. *Environmental health perspectives*, 111(14):1773–1778, 2003.

[62] Jackson G Lu. Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*, 32:52–65, 2020.

[63] JP Mackenbach, AE Kunst, and CW Looman. Seasonal variation in mortality in the Netherlands. *Journal of Epidemiology & Community Health*, 46(3):261–265, 1992.

[64] Artur Mardyukov and Peter R Schreiner. Atmospherically relevant radicals derived from the oxidation of dimethyl sulfide. *Accounts of Chemical Research*, 51(2):475–483, 2018.

[65] Joe L Mauderly, Richard T Burnett, Margarita Castillejos, Halûk Özkaynak, Jonathan M Samet, David M Stieb, Sverre Vedal, and Ronald E Wyzga. Is the air pollution health research community prepared to support a multipollutant

air quality management framework? *Inhalation toxicology*, 22(sup1):1–19, 2010.

[66] Sati Mazumdar, Herbert Schimmel, and Ian T. T. Higgins. Relation of daily mortality to air pollution: An analysis of 14 London winters, 1958/59–1971/72. *Archives of Environmental Health: An International Journal*, 37(4):213–220, 1982. PMID: 7114901.

[67] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[68] Government of Canada. Final report by the task force on just transition for Canadian coal power workers and communities: section 6, July 2020.

[69] Government of Canada. The Georgia Basin-Puget sound airshed characterization report 2014: chapter 10. `https://www.canada.ca/en/environment-climate-change/services/air-pollution/publications/georgia-basin-puget-sound-report-2014/chapter-10.htm`, July 2020.

[70] Government of Canada Open Data Portal. National Air Pollution Surveillance Program. `https://data-donnees.ec.gc.ca/data/air/monitor/national-air-pollution-surveillance-naps-program/?lang=en`, July 2020.

[71] Canadian Council of Minister of the Environment. Ambient air monitoring and quality assurance/quality control guidelines. `https://ccme.ca/en/res/ambientairmonitoringandqa-qcguidelines_ensecure.pdf`.

[72] K Frans G Olofson, Patrik U Andersson, Mattias Hallquist, Evert Ljungström, Lin Tang, Deliang Chen, and Jan BC Pettersson. Urban aerosol evolution and particle formation during wintertime temperature inversions. *Atmospheric Environment*, 43(2):340–346, 2009.

[73] World Health Organization et al. Review of evidence on health aspects of air pollution: REVIHAAP project: technical report. Technical report, World

Health Organization. Regional Office for Europe, 2021.

[74] Bart Ostro. Urban air pollution and morbidity: A retrospective approach. *Urban studies (Edinburgh, Scotland)*, 20(3):343–351, 1983.

[75] E.S. Park, E. Symanski, D. Han, and C. Spiegelman. Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants. Technical Report 183-II, Health Effects Institute, 2015. Part II of III.

[76] Eun Sug Park, Elaine Symanski, Daikwon Han, and Clifford Spiegelman. Part 2. development of enhanced statistical methods for assessing health effects associated with an unknown number of major sources of multiple air pollutants. *Research report (Health Effects Institute)*, (183 Pt 1-2):51–113, 2015.

[77] Roger D Peng, Michelle L Bell, Alison S Geyh, Aidan McDermott, Scott L Zeger, Jonathan M Samet, and Francesca Dominici. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environmental Health Perspectives*, 117(6):957–963, 2009.

[78] Roger D Peng, Francesca Dominici, and Thomas A Louis. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(2):179–203, 2006.

[79] Roger D Peng and Leah J Welty. The NMMAPS data package. *R News*, 4(2):10–14, 2004.

[80] Roger D Peng, Leah J Welty, and Aidan McDermott. The national morbidity, mortality, and air pollution study database in R, 2004.

[81] Marc Pitchford, Richard Poirot, Bret Schichtel, and William Maim. Characterization of the winter midwestern particulate nitrate bulge. *Journal of the Air & Waste Management Association (1995)*, 59:1061–9, 09 2009.

[82] C Arden Pope III and Douglas W Dockery. Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management*

Association, 56(6):709–742, 2006.

[83] A. Pozzer, A. P. Tsimpidi, V. A. Karydis, A. de Meij, and J. Lelieveld. Impact of agricultural emission reductions on fine-particulate matter and public health. *Atmospheric Chemistry and Physics*, 17(20):12813–12826, 2017.

[84] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[85] Jongeun Rhee, Francesca Dominici, Antonella Zanobetti, Joel Schwartz, Yun Wang, Qian Di, John Balmes, and David C Christiani. Impact of long-term exposures to ambient PM2.5 and ozone on ARDS risk for older adults in the United States. *Chest*, 156(1):71–79, 2019.

[86] Jonathan M Samet, Francesca Dominici, Frank C Curriero, Ivan Coursac, and Scott L Zeger. Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *New England Journal of Medicine*, 343(24):1742–1749, 2000.

[87] Tamara Schikowski and Hicran Altuğ. The role of air pollution in cognitive impairment and decline. *Neurochemistry International*, 136:104708, 2020.

[88] Joel Schwartz. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics*, 22(4):471–487, 1994.

[89] John H. Seinfeld and Spyros N. Pandis. *10.4 Thermodynamics of atmospheric aerosol systems*, chapter 10.4, pages 423–442. John Wiley and Sons, 2016.

[90] Viral Shah, Lyatt Jaeglé, Joel A. Thornton, Felipe D. Lopez-Hilfiker, Ben H. Lee, Jason C. Schroder, Pedro Campuzano-Jost, Jose L. Jimenez, Hongyu Guo, Amy P. Sullivan, Rodney J. Weber, Jaime R. Green, Marc N. Fiddler, Solomon Bililign, Teresa L. Campos, Meghan Stell, Andrew J. Weinheimer, Denise D. Montzka, and Steven S. Brown. Chemical feedbacks weaken the wintertime response of particulate sulfate and nitrate to emissions reductions over the eastern

United States. *Proceedings of the National Academy of Sciences*, 115(32):8110–8115, 2018.

[91] Hwashin H Shin, Wesley S Burr, Dave Stieb, Lani Haque, Harun Kalayci, Branka Jovic, and Marc Smith-Doiron. Air Health Trend Indicator: association between short-term exposure to ground ozone and circulatory hospitalizations in Canada for 17 years, 1996–2012. *International Journal of Environmental Research and Public Health*, 15(8):1566, 2018.

[92] Hwashin Hyun Shin, Rajendra Prasad Parajuli, Aubrey Maquiling, and Marc Smith-Doiron. Temporal trends in associations between ozone and circulatory mortality in age and sex in Canada during 1984–2012. *Science of The Total Environment*, 724:137944, 2020.

[93] Angelo Roldão Soares and Carla Silva. Review of ground-level ozone impact in respiratory health deterioration for the past two decades. *Atmosphere*, 13(3):434, 2022.

[94] Ji-Young Son, M Benjamin Sabath, Kevin J Lane, Marie Lynn Miranda, Francesca Dominici, Qian Di, Joel Schwartz, and Michelle L Bell. Long-term exposure to PM2.5 and mortality for the older population: Effect modification by residential greenness. *Epidemiology*, 32(4):477–486, 2021.

[95] David M Stieb, Li Chen, Perry Hystad, Bernardo S Beckerman, Michael Jerrett, Michael Tjepkema, Daniel L Crouse, D Walter Omariba, Paul A Peters, Aaron van Donkelaar, et al. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999–2008. *Environmental Research*, 148:513–526, 2016.

[96] Adam A. Szpiro and Christopher J. Paciorek. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, 24(8):501–517, 2013.

[97] R T. Burnett, J Brook, T Dann, C Delocla, O Philips, S Cakmak, R Vincent,

M S. Goldberg, and D Krewski. Association between particulate-and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. *Inhalation toxicology*, 12(sup4):15–39, 2000.

[98] Amos PK Tai, Loretta J Mickley, and Daniel J Jacob. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmospheric Environment*, 44(32):3976–3984, 2010.

[99] Daniel Tong, Rohit Mathur, Kenneth Schere, Daiwen Kang, and Shaocai Yu. The use of air quality forecasts to assess impacts of air pollution on crops: Methodology and case study. *Atmospheric Environment*, 41(38):8772–8784, 2007.

[100] Maria Angela Vigotti, G Rossi, L Bisanti, A Zanobetti, and J Schwartz. Short term effects of urban air pollution on respiratory health in Milan, Italy, 1980-89. *Journal of Epidemiology & Community Health*, 50(Suppl 1):s71–s75, 1996.

[101] Thomais Vlachogianni, Konstantinos Fiotakis, Spyridon Loridas, Stamatis Perdicaris, and Athanasios Valavanidis. Potential toxicity and safety evaluation of nanomaterials for the respiratory system and lung cancer. *Lung Cancer (Auckland)*, 4:71–82, 2013.

[102] Julie Wallace, Denis Corr, and Pavlos Kanaroglou. Topographic and spatial impacts of temperature inversions on air quality using mobile air pollution surveys. *Science of the Total Environment*, 408(21):5086–5096, 2010.

[103] Julie Wallace and Pavlos Kanaroglou. The effect of temperature inversions on ground-level nitrogen dioxide (NO2) and fine particulate matter (PM2.5) using temperature profiles from the atmospheric infrared sounder (AIRS). *Science of the Total Environment*, 407(18):5085–5095, 2009.

[104] Shanshan Wang, Jialiang Nan, Chanzhen Shi, Qingyan Fu, Song Gao, Dongfang Wang, Huxiong Cui, Alfonso Saiz-Lopez, and Bin Zhou. Atmospheric ammonia

and its impacts on regional air quality over the megacity of Shanghai, China. *Scientific Reports*, 5(1):1–13, 2015.

[105] S. N. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467):673–686, 2004.

[106] S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.

[107] World Health Organization. International Statistical Classification of Diseases and Related Health Problems (ICD). `https://www.who.int/standards/classifications/classification-of-diseases`, 2022. [Online; accessed September 20, 2022].

[108] Tingting Xu, Yu Song, Mingxu Liu, Xuhui Cai, Hongsheng Zhang, Jianping Guo, and Tong Zhu. Temperature inversions in severe polluted days derived from radiosonde data in north China from 2011 to 2016. *Science of the Total Environment*, 647:1011–1020, 2019.

[109] Yang Yang, Zengliang Ruan, Xiaojie Wang, Yin Yang, Tonya G Mason, Hualiang Lin, and Linwei Tian. Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis. *Environmental Pollution (1987)*, 247:874–882, 2019.

[110] Jeff D Yanosky, Christopher J Paciorek, Francine Laden, Jaime E Hart, Robin C Puett, Duanping Liao, and Helen H Suh. Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. *Environmental Health*, 13(1):1–15, 2014.

[111] D Zmirou, T Barumandzadeh, F Balducci, P Ritter, G Laham, and JP Ghilardi. Short term effects of air pollution on mortality in the city of Lyon, France, 1985-90. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S30–S35, 1996.

A. *Data Summaries*

Negative pollutant by census division

The count of negative integrated pollutant concentrations by CD are in the following table. Blank values represent a count of zero.

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Division No. 1, NL (1001)	3	1990-01-18:2019-12-31	NO_2				11
Division No. 1, NL (1001)	3	1989-04-01:2019-12-31	O_3		6		
Division No. 1, NL (1001)	3	1980-01-01:2019-12-31	SO_2		61		
Division No. 1, NL (1001)	3	1998-01-01:2019-12-31	$PM_{2.5}$		160		
Division No. 5, NL (1005)	3	2009-12-31:2019-12-31	SO_2		1		1
Division No. 5, NL (1005)	3	2001-07-01:2019-12-31	NO_2		3		22
Division No. 5, NL (1005)	3	2001-07-07:2019-12-31	$PM_{2.5}$	2	46		1
Halifax, NS (1209)	10	1980-01-01:2019-12-31	O_3		2		
Halifax, NS (1209)	10	2001-06-19:2019-12-31	$PM_{2.5}$		70		1
Halifax, NS (1209)	10	1980-01-02:2019-12-31	SO_2		383	1	
Halifax, NS (1209)	10	1980-01-01:2019-12-31	NO_2	5	57	1	
Saint John, NB (1301)	9	1980-01-01:2019-12-31	O_3		1		
Saint John, NB (1301)	9	1980-01-27:2019-12-31	SO_2		395	1	162
Saint John, NB (1301)	9	1980-08-08:2019-12-31	NO_2		21		11
Saint John, NB (1301)	9	1996-11-21:2019-12-31	$PM_{2.5}$		261		12
Saint John, NB (1301)	9	1996-11-20:2004-05-30	PM_{10}		47	1	1
Westmorland, NB (1307)	1	1998-09-01:2019-12-31	NO_2		1		49
Westmorland, NB (1307)	1	1998-07-17:2019-12-31	O_3		4		
Westmorland, NB (1307)	1	1999-11-06:2019-12-31	$PM_{2.5}$		50		11
York, NB (1310)	3	1999-04-24:2019-12-31	$PM_{2.5}$		121		11
York, NB (1310)	3	1994-08-01:2019-12-31	O_3		3		
York, NB (1310)	3	1999-04-17:2019-12-31	NO_2		3		65
Quebec, QC (2423)	9	1980-01-01:2019-12-31	NO_2		16		
Quebec, QC (2423)	9	1980-01-01:2019-12-31	SO_2		119		
Quebec, QC (2423)	9	1998-04-08:2019-12-31	$PM_{2.5}$		21		14
Quebec, QC (2423)	9	1980-01-01:2019-12-31	O_3		239		1
Le Haut-Richelieu, QC (2456)	1	1999-08-12:2019-12-31	NO_2		4		6
Le Haut-Richelieu, QC (2456)	1	1999-11-27:2019-12-31	$PM_{2.5}$		49	1	4
Le Haut-Richelieu, QC (2456)	1	1999-08-12:2019-12-31	O_3		20		
Longueuil, QC (2458)	4	2003-01-25:2019-12-31	$PM_{2.5}$		14		
Longueuil, QC (2458)	4	1982-01-03:2019-12-31	NO_2		12		
Longueuil, QC (2458)	4	1982-01-03:2019-12-31	SO_2	1	13		390
Longueuil, QC (2458)	4	1982-01-03:2019-12-31	O_3		127		1
Laval, QC (2465)	3	1980-01-02:1992-12-31	SO_2		1		39
Laval, QC (2465)	3	2003-10-09:2019-12-31	$PM_{2.5}$		7		

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Laval, QC (2465)	3	1980-01-01:2019-12-31	NO_2		5		
Laval, QC (2465)	3	1980-01-01:2019-12-31	O_3		61		
Montreal, QC (2466)	25	1980-01-01:2019-12-31	O_3		27		
Montreal, QC (2466)	25	1980-01-01:2019-12-31	NO_2		26		
Montreal, QC (2466)	25	1980-01-01:2019-12-31	SO_2	1	30		
Montreal, QC (2466)	25	1997-09-01:2019-12-31	$PM_{2.5}$		78		1
Gatineau, QC (2481)	3	1980-01-01:2019-12-31	NO_2		8		
Gatineau, QC (2481)	3	2003-04-16:2019-12-31	$PM_{2.5}$		7		
Gatineau, QC (2481)	3	1980-01-02:2019-12-31	SO_2		17		
Gatineau, QC (2481)	3	1992-01-01:2019-12-31	O_3		10		
Stormont, Dundas and Glengarry, ON (3501)	2	1980-01-01:2019-12-31	O_3		5		
Stormont, Dundas and Glengarry, ON (3501)	2	2003-04-17:2019-12-31	$PM_{2.5}$	82	11		
Stormont, Dundas and Glengarry, ON (3501)	2	1980-01-01:2019-12-31	NO_2		19		
Stormont, Dundas and Glengarry, ON (3501)	2	1980-01-01:2002-10-09	SO_2		41		89
Ottawa, ON (3506)	4	1980-01-01:2019-12-31	O_3		68		
Ottawa, ON (3506)	4	1980-01-01:2019-12-31	NO_2		22		
Ottawa, ON (3506)	4	1998-01-01:2019-12-31	$PM_{2.5}$		14		2
Ottawa, ON (3506)	4	1980-01-01:2019-12-31	SO_2	2888	77		
Peterborough, ON (3515)	2	1993-01-06:2019-12-31	O_3		2		
Peterborough, ON (3515)	2	1998-12-02:2001-02-26	PM_{10}		1		
Peterborough, ON (3515)	2	1980-01-01:2003-11-17	SO_2				29
Peterborough, ON (3515)	2	1999-01-21:2019-12-31	NO_2	6			8
Peterborough, ON (3515)	2	2001-02-28:2019-12-31	$PM_{2.5}$	142	5		3
Durham, ON (3518)	3	1980-01-01:2017-12-31	O_3	1	37		
Durham, ON (3518)	3	1980-01-01:2017-12-31	NO_2		2		
Durham, ON (3518)	3	1980-01-01:2005-12-31	SO_2				49
Durham, ON (3518)	3	1997-04-19:2017-12-31	$PM_{2.5}$	19	12		
York, ON (3519)	2	2001-06-19:2019-12-31	$PM_{2.5}$	173	13		
York, ON (3519)	2	1980-01-01:2019-12-31	O_3		6		
York, ON (3519)	2	2001-06-09:2006-12-31	SO_2				13
York, ON (3519)	2	1983-06-09:2019-12-31	NO_2	2			
Toronto, ON (3520)	28	1980-01-01:2019-12-31	SO_2	622	56		
Toronto, ON (3520)	28	1997-01-01:2019-12-31	$PM_{2.5}$	231	72		
Toronto, ON (3520)	28	1980-01-01:2019-12-31	NO_2	4	9		
Toronto, ON (3520)	28	1980-01-01:2019-12-31	O_3	6	246		

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Peel, ON (3521)	5	1980-01-01:2019-12-31	O_3	4	40		
Peel, ON (3521)	5	1980-01-01:2019-12-31	NO_2		1		
Peel, ON (3521)	5	1980-01-01:2016-05-31	SO_2	22			194
Peel, ON (3521)	5	1997-01-01:2019-12-31	$PM_{2.5}$	68	17		
Halton, ON (3524)	5	1980-01-01:2019-12-31	NO_2	54	27		
Halton, ON (3524)	5	1980-01-01:2019-12-31	O_3	2	19		
Halton, ON (3524)	5	1980-01-01:2006-12-31	SO_2		7		36
Halton, ON (3524)	5	1999-05-11:2001-04-23	PM_{10}		1		
Halton, ON (3524)	5	2001-04-26:2019-12-31	$PM_{2.5}$	74	15		
Hamilton, ON (3525)	10	1980-01-01:2019-12-31	SO_2	3231	375		
Hamilton, ON (3525)	10	1980-01-01:2019-12-31	NO_2		13		
Hamilton, ON (3525)	10	1980-01-01:2019-12-31	O_3		55		
Hamilton, ON (3525)	10	1998-01-01:2019-12-31	$PM_{2.5}$	783	43		
Niagara, ON (3526)	4	1980-01-01:2019-12-31	O_3	1	34		
Niagara, ON (3526)	4	1980-01-01:2019-12-31	NO_2		12		
Niagara, ON (3526)	4	1998-01-01:2019-12-31	$PM_{2.5}$	18	59		
Niagara, ON (3526)	4	1980-01-01:2006-12-31	SO_2		11		38
Haldimand-Norfolk, ON (3528)	2	1980-01-01:2019-12-31	NO_2	175	12		8
Haldimand-Norfolk, ON (3528)	2	1998-01-01:2019-12-31	$PM_{2.5}$	31	14		2
Haldimand-Norfolk, ON (3528)	2	1980-01-01:2019-12-31	O_3		6		
Brant, ON (3529)	2	2004-01-02:2019-12-31	NO_2	5			4
Brant, ON (3529)	2	2004-01-01:2019-12-31	$PM_{2.5}$	1	8		
Brant, ON (3529)	2	1980-01-01:2006-12-31	SO_2				1
Waterloo, ON (3530)	3	1980-01-01:2019-12-31	O_3	24	19		
Waterloo, ON (3530)	3	1980-01-01:2019-12-31	NO_2	3			
Waterloo, ON (3530)	3	1998-01-01:2019-12-31	$PM_{2.5}$	34	100		7
Waterloo, ON (3530)	3	1980-01-01:2006-12-31	SO_2				46
Essex, ON (3537)	6	1980-01-01:2019-12-31	O_3	21	114		7
Essex, ON (3537)	6	1980-01-01:2019-12-31	SO_2	6460	403	2	
Essex, ON (3537)	6	1980-01-01:2019-12-31	O_3	21	114		7
Essex, ON (3537)	6	1980-01-01:2019-12-31	NO_2	2	1		
Essex, ON (3537)	6	1999-03-18:2019-12-31	$PM_{2.5}$	49	21		5
Lambton, ON (3538)	5	1980-01-01:2019-12-31	O_3		24		
Lambton, ON (3538)	5	1980-01-01:2019-12-31	NO_2	15	10		
Lambton, ON (3538)	5	1980-01-01:2019-12-31	SO_2	7145	782	13	

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Lambton, ON (3538)	5	2000-02-17:2019-12-31	$PM_{2.5}$	36	4		5
Middlesex, ON (3539)	6	1980-01-22:2010-12-31	SO_2		16		112
Middlesex, ON (3539)	6	1980-01-01:2019-12-31	NO_2	6	4		
Middlesex, ON (3539)	6	2001-01-01:2019-12-31	$PM_{2.5}$	22	9		
Middlesex, ON (3539)	6	1980-01-01:2019-12-31	O_3	3	40		
Simcoe, ON (3543)	3	2001-12-01:2019-12-31	NO_2		2		
Simcoe, ON (3543)	3	1981-06-27:2019-12-31	O_3		8		
Simcoe, ON (3543)	3	2001-12-01:2019-12-31	$PM_{2.5}$	45	33		
Simcoe, ON (3543)	3	2001-12-01:2006-12-31	SO_2				8
Nipissing, ON (3548)	1	1999-09-10:2019-12-31	$PM_{2.5}$	37	9		
Nipissing, ON (3548)	1	1988-01-01:2019-12-31	NO_2		3		
Greater Sudbury / Grand Sudbury, ON (3553)	6	1980-01-01:2019-12-31	NO_2		6		
Greater Sudbury / Grand Sudbury, ON (3553)	6	2004-06-28:2019-12-31	$PM_{2.5}$	66	11		20
Greater Sudbury / Grand Sudbury, ON (3553)	6	1980-01-01:2019-12-31	SO_2	3409	366	7	
Greater Sudbury / Grand Sudbury, ON (3553)	6	1980-01-01:2019-12-31	O_3		5		
Algoma, ON (3557)	7	1996-01-01:2001-12-31	PM_{10}		7		2
Algoma, ON (3557)	7	2000-01-01:2019-12-31	$PM_{2.5}$		13		
Algoma, ON (3557)	7	1983-02-23:2019-12-31	NO_2		10		7
Algoma, ON (3557)	7	1983-02-23:2019-12-31	O_3		8		1
Algoma, ON (3557)	7	1980-01-01:2019-12-31	SO_2		21		583
Thunder Bay, ON (3558)	5	2001-06-22:2019-12-31	$PM_{2.5}$		9		
Thunder Bay, ON (3558)	5	1980-01-01:2003-12-31	SO_2				9
Thunder Bay, ON (3558)	5	1981-01-01:2019-12-31	O_3		5		2
Division No. 7, MB (4607)	3	1993-12-10:2018-12-18	NO_2	17	30		44
Division No. 7, MB (4607)	3	2001-06-20:2018-12-31	$PM_{2.5}$		25		3
Division No. 7, MB (4607)	3	1987-01-02:2018-12-31	O_3	1	45		
Division No. 11, MB (4611)	7	1993-06-01:2018-12-31	PM_{10}	54	16		17
Division No. 11, MB (4611)	7	1980-01-01:2018-12-31	NO_2	177	69		
Division No. 11, MB (4611)	7	1980-01-01:2018-12-31	O_3		191		
Division No. 11, MB (4611)	7	1980-01-01:2018-12-31	SO_2		38		
Division No. 11, MB (4611)	7	1997-09-06:2018-12-31	$PM_{2.5}$	81	14		1
Division No. 6, SK (4706)	6	1980-01-01:2019-12-31	O_3	5	2		
Division No. 6, SK (4706)	6	1999-01-01:2017-12-31	PM_{10}		1		
Division No. 6, SK (4706)	6	1980-04-15:2019-12-31	NO_2		3		
Division No. 6, SK (4706)	6	1980-02-08:2019-12-31	SO_2		11		262

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Division No. 6, SK (4706)	6	2001-01-01:2019-12-31	$PM_{2.5}$		26		1
Division No. 11, SK (4711)	3	1984-01-01:2019-12-31	NO_2				1
Division No. 11, SK (4711)	3	2003-07-11:2019-12-31	$PM_{2.5}$		42		3
Division No. 11, SK (4711)	3	1980-02-20:2019-12-31	SO_2		2		103
Division No. 11, SK (4711)	3	1984-01-01:2019-12-31	O_3	1	15		
Division No. 2, AB (4802)	1	2004-01-02:2019-12-31	NO_2		4		4
Division No. 2, AB (4802)	1	2004-01-02:2019-12-31	O_3		5		
Division No. 2, AB (4802)	1	2003-10-02:2019-12-30	$PM_{2.5}$		41		
Division No. 2, AB (4802)	1	2004-01-02:2019-12-31	SO_2				65
Division No. 6, AB (4806)	13	1980-01-02:2019-12-31	NO_2		12		
Division No. 6, AB (4806)	13	1980-01-02:2019-12-31	O_3		176		
Division No. 6, AB (4806)	13	1996-01-02:2012-04-16	PM_{10}		3		2
Division No. 6, AB (4806)	13	1997-11-02:2019-12-31	$PM_{2.5}$		54		
Division No. 6, AB (4806)	13	1980-01-02:2019-12-31	SO_2		46		388
Division No. 8, AB (4808)	5	2000-01-02:2019-12-31	NO_2		1		1
Division No. 8, AB (4808)	5	2000-01-02:2019-12-31	O_3	1	87		
Division No. 8, AB (4808)	5	2001-01-03:2019-12-31	$PM_{2.5}$		26		1
Division No. 8, AB (4808)	5	2001-01-02:2019-12-31	SO_2		3		55
Division No. 10, AB (4810)	5	1994-01-02:2019-12-31	NO_2		16		16
Division No. 10, AB (4810)	5	1993-04-02:2019-12-31	O_3		15		
Division No. 10, AB (4810)	5	2004-01-02:2014-12-31	PM_{10}		5		1
Division No. 10, AB (4810)	5	2004-01-02:2019-12-31	$PM_{2.5}$		19		1
Division No. 10, AB (4810)	5	2004-01-02:2019-12-31	SO_2		48		31
Division No. 11, AB (4811)	27	1980-01-02:2019-12-31	NO_2		231		
Division No. 11, AB (4811)	27	1980-01-02:2019-12-31	O_3		226		
Division No. 11, AB (4811)	27	1994-01-02:2014-12-31	PM_{10}		29		
Division No. 11, AB (4811)	27	1998-04-18:2019-12-31	$PM_{2.5}$	1	247		3
Division No. 11, AB (4811)	27	1980-01-02:2019-12-31	SO_2		555		36
Division No. 16, AB (4816)	15	1986-07-03:2019-12-31	NO_2		462	1	2
Division No. 16, AB (4816)	15	1986-07-02:2019-12-31	O_3		243		
Division No. 16, AB (4816)	15	1998-01-02:2019-12-31	$PM_{2.5}$		309		
Division No. 16, AB (4816)	15	1991-01-02:2019-12-31	SO_2		786		60
Division No. 19, AB (4819)	4	1998-01-02:2019-12-31	NO_2		3		15
Division No. 19, AB (4819)	4	1998-01-02:2019-12-31	O_3		29		
Division No. 19, AB (4819)	4	2004-02-02:2019-12-30	$PM_{2.5}$		113		1

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
Division No. 19, AB (4819)	4	2001-01-02:2019-12-31	SO_2		43		44
Fraser Valley, BC (5909)	10	1986-01-02:2019-12-31	NO_2		10		
Fraser Valley, BC (5909)	10	1982-08-02:2019-12-31	O_3		857		1
Fraser Valley, BC (5909)	10	1994-07-20:2019-12-31	PM_{10}		20		
Fraser Valley, BC (5909)	10	1995-06-02:2019-12-31	$PM_{2.5}$		23		3
Fraser Valley, BC (5909)	10	1998-09-19:2019-12-31	SO_2		13		9
Greater Vancouver, BC (5915)	34	1980-01-02:2019-12-31	NO_2		90		
Greater Vancouver, BC (5915)	34	1980-01-02:2019-12-31	O_3		2265		
Greater Vancouver, BC (5915)	34	1994-01-02:2019-12-31	PM_{10}		15		
Greater Vancouver, BC (5915)	34	1999-01-15:2019-12-31	$PM_{2.5}$	1	77		
Greater Vancouver, BC (5915)	34	1980-01-02:2019-12-31	SO_2		328		1
Capital, BC (5917)	13	1980-01-02:2019-12-31	NO_2		59		
Capital, BC (5917)	13	1980-01-02:2019-12-31	O_3		173		3
Capital, BC (5917)	13	2001-01-04:2013-09-16	PM_{10}		1		
Capital, BC (5917)	13	1998-05-02:2019-12-31	$PM_{2.5}$		60		1
Capital, BC (5917)	13	1981-01-10:2019-12-31	SO_2		117		55
Squamish-Lillooet, BC (5931)	3	2003-07-05:2019-12-31	NO_2		6		
Squamish-Lillooet, BC (5931)	3	1995-07-06:2019-12-31	O_3		158		1
Squamish-Lillooet, BC (5931)	3	1998-01-03:2011-01-28	PM_{10}		51		1
Squamish-Lillooet, BC (5931)	3	2005-01-02:2019-12-31	$PM_{2.5}$		71		1
Squamish-Lillooet, BC (5931)	3	1998-01-02:2019-12-31	SO_2		17		18
Thompson-Nicola, BC (5933)	2	1998-01-03:2019-12-31	NO_2		33		1
Thompson-Nicola, BC (5933)	2	1998-01-03:2019-12-31	O_3		67		2
Thompson-Nicola, BC (5933)	2	1995-01-02:2008-12-31	PM_{10}		9		1
Thompson-Nicola, BC (5933)	2	1998-01-02:2019-12-30	$PM_{2.5}$		19		
Thompson-Nicola, BC (5933)	2	1980-02-23:2019-12-31	SO_2		27		21
Central Okanagan, BC (5935)	1	1998-01-03:2019-06-03	NO_2		16		
Central Okanagan, BC (5935)	1	1987-07-02:2019-06-03	O_3		53		7
Central Okanagan, BC (5935)	1	1987-07-02:2019-06-03	O_3		53		1
Central Okanagan, BC (5935)	1	1995-01-02:2019-06-03	PM_{10}		4		
Central Okanagan, BC (5935)	1	1998-01-03:2019-06-12	$PM_{2.5}$		14		3
Central Okanagan, BC (5935)	1	2001-01-03:2019-03-10	SO_2		1		4
North Okanagan, BC (5937)	1	2002-10-16:2019-12-31	NO_2		7		
North Okanagan, BC (5937)	1	2002-10-16:2019-12-31	O_3		35		10
North Okanagan, BC (5937)	1	2002-10-18:2019-11-21	PM_{10}		4		1

Census Division	No. NAPS stations	Date range	Pollutant	No. negative raw values	No. negative CATNAPS hourly by NAPS station	No. negative CATNAPS hourly by CD	No. negative daily
North Okanagan, BC (5937)	1	2002-10-16:2019-12-31	$PM_{2.5}$		17		
North Okanagan, BC (5937)	1	2003-12-18:2013-10-29	SO_2				3
Cariboo, BC (5941)	7	1998-01-03:2019-12-31	NO_2		63		
Cariboo, BC (5941)	7	1992-04-16:2019-12-31	O_3		140		
Cariboo, BC (5941)	7	1995-01-02:2019-12-31	PM_{10}		70		
Cariboo, BC (5941)	7	2000-03-11:2019-12-31	$PM_{2.5}$		385		1
Cariboo, BC (5941)	7	2006-07-27:2019-12-31	SO_2		35		9
Fraser-Fort George, BC (5953)	11	1998-01-03:2019-12-31	NO_2		7		2
Fraser-Fort George, BC (5953)	11	1995-04-29:2019-12-31	O_3		121		1
Fraser-Fort George, BC (5953)	11	1995-01-02:2019-12-31	PM_{10}		53		
Fraser-Fort George, BC (5953)	11	1998-01-02:2019-12-31	$PM_{2.5}$	2	166		7
Fraser-Fort George, BC (5953)	11	2001-01-03:2019-12-31	SO_2		541		39

Table A.1: Count of negative pollutant concentrations by pollutant and census divisions. Missing values indicate a count of zero.

Mapping from CD to NAPS Station

The table on the next page shows the mapping between census division (name and unique identifier) to naps station, by station ID.

Census Division	NAPS Station ID(s)
Division No. 1, NL (1001)	10102, 10401, 10101
Division No. 5, NL (1005)	10201, 10301, 10602
Halifax, NS (1209)	30113, 30115, 30116, 30117, 30118, 30120, 30601, 31001, 30101, 30114
Saint John, NB (1301)	40202, 40203, 40204, 40206, 40207, 40209, 40501, 40208, 40201
Westmorland, NB (1307)	40302
York, NB (1310)	40103, 40104, 40801
Quebec, QC (2423)	50304, 50306, 50307, 50308, 50309, 50310, 50311, 50303, 50305
Le Haut-Richelieu, QC (2456)	55301
Longueuil, QC (2458)	50108, 50119, 50121, 50122
Laval, QC (2465)	50111, 50112, 50113
Montreal, QC (2466)	50102, 50103, 50104, 50107, 50109, 50110, 50115, 50116, 50120, 50123, 50126, 50128, 50129,
	50130, 50134, 50135, 50136, 50138, 50133, 50101, 50105, 50106, 50117, 50118, 50131
Gatineau, QC (2481)	50203, 50204, 50201
Stormont, Dundas and Glengarry, ON (3501)	61201, 65701
Ottawa, ON (3506)	60101, 60104, 60105, 60106
Peterborough, ON (3515)	61103, 61104
Durham, ON (3518)	61701, 61702
York, ON (3519)	63201, 65101
Toronto, ON (3520)	60401, 60402, 60403, 60410, 60413, 60414, 60417, 60418, 60419, 60420, 60421, 60422,
	60423, 60424, 60425, 60429, 60430, 60433, 60435, 60440, 60412, 60409, 60411
Peel, ON (3521)	60415, 60428, 60432, 60434, 60450
Halton, ON (3524)	61601, 61602, 61603, 63001
Hamilton, ON (3525)	60501, 60512, 60513, 60514, 60515, 60505, 60511, 60518, 60519
Niagara, ON (3526)	61301, 61302, 62901, 62902
Haldimand-Norfolk, ON (3528)	62601, 62701
Brant, ON (3529)	61402, 61401
Waterloo, ON (3530)	61501, 61502, 62801
Essex, ON (3537)	60204, 60211, 65601, 60203, 60201, 60212
Lambton, ON (3538)	61004, 61005, 61009, 61007, 61001
Middlesex, ON (3539)	60901, 60903, 60904, 62401, 63601, 64301
Simcoe, ON (3543)	63101, 64401, 65001
Nipissing, ON (3548)	62001
Greater Sudbury / Grand Sudbury, ON (3553)	60602, 60607, 60609, 60610, 63501, 60606
Algoma, ON (3557)	60706, 60707, 60709, 64101, 60704, 60705, 60708
Thunder Bay, ON (3558)	60806, 60807, 60809, 63401, 63901
Division No. 7, MB (4607)	70203, 70201, 70202

Census Division	NAPS Station ID(s)
Division No. 11, MB (4611)	70118, 70119, 70101, 70102, 70103, 70104, 70120
Division No. 6, SK (4706)	80109, 80110, 80111, 80801, 80901, 80108
Division No. 11, SK (4711)	80209, 80211, 80202
Division No. 2, AB (4802)	90502
Division No. 6, AB (4806)	90218, 90222, 90227, 90228, 90229, 90230, 90250, 92701, 93301, 93401, 94101, 90219, 90220
Division No. 8, AB (4808)	90302, 90303, 90304, 93501, 93701
Division No. 10, AB (4810)	90606, 90609, 90901, 91101, 92201
Division No. 11, AB (4811)	90120, 90121, 90122, 90130, 90133, 90134, 90135, 90136, 90601, 90607, 90608, 91301, 91401, 92601, 93101, 90602, 90603, 90605, 92301, 93801, 94201, 94202, 90114, 90131, 90604, 90132, 92801
Division No. 16, AB (4816)	90701, 90702, 90801, 90806, 90807, 90808, 91801, 94601, 90703, 90805, 92101, 92102, 90802, 90803, 90804
Division No. 19, AB (4819)	91501, 92001, 93001, 94001
Fraser Valley, BC (5909)	100143, 101001, 101002, 101003, 101004, 101005, 101101, 101401, 103602, 105401
Greater Vancouver, BC (5915)	100103, 100106, 100108, 100109, 100110, 100111, 100112, 100118, 100119, 100120, 100121, 100122, 100123, 100124, 100125, 100126, 100127, 100128, 100129, 100130, 100131, 100132, 100134, 100135, 100140, 100141, 101201, 101202, 101301, 101501, 100105, 100136, 100137, 100138
Capital, BC (5917)	100302, 100303, 100304, 100307, 100308, 100314, 100316, 102001, 100301, 100317, 100318, 100313, 100312
Squamish-Lillooet, BC (5931)	101601, 101603, 105001
Thompson-Nicola, BC (5933)	100401, 100402
Central Okanagan, BC (5935)	100701
North Okanagan, BC (5937)	104003
Cariboo, BC (5941)	101701, 102701, 102706, 101702, 101703, 101704, 102702
Fraser-Fort George, BC (5953)	100202, 100201, 100205, 100210, 100211, 100212, 100213, 100214, 106101, 100203, 100209

Table A.2: Mapping from census division to NAPS stations

B. *Ozone Risk Estimates*

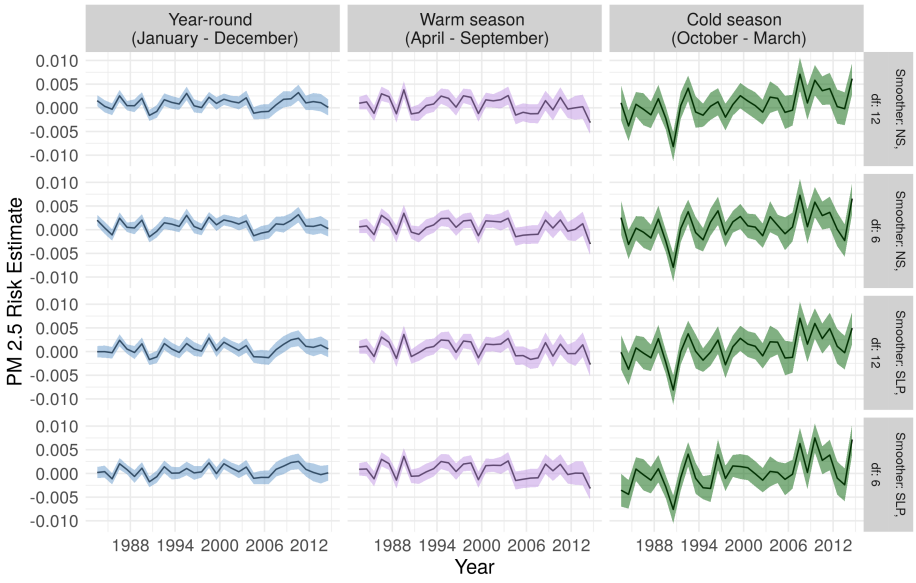


Figure B.1: Annual mortality (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for varying smooth functions of time (natural spline (NS); and Discrete Prolate Spheroidal (Slepian) Sequence (SLP)), with 6 and 12 degrees of freedom (df) per year

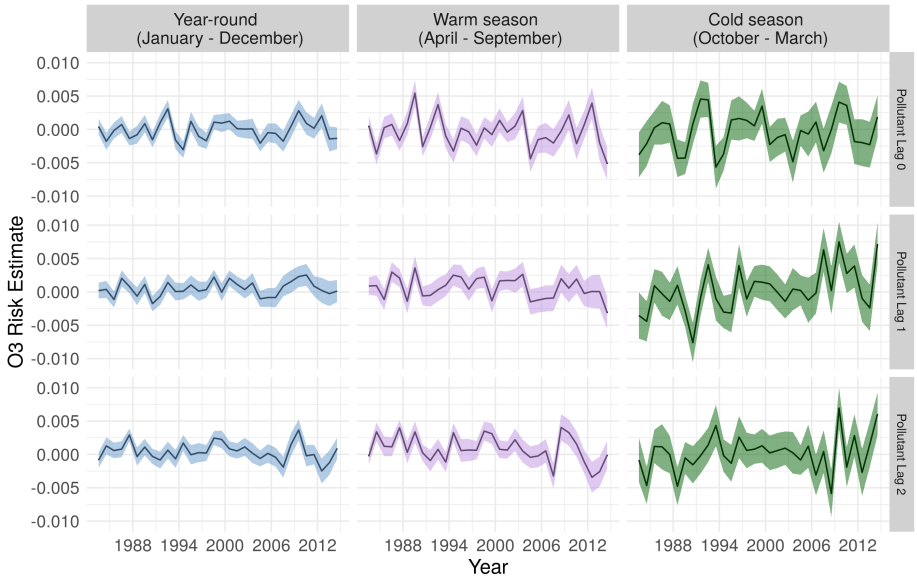


Figure B.2: Annual mortality (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for three lags on the pollutant data (o days; 1 day; and 2 days)

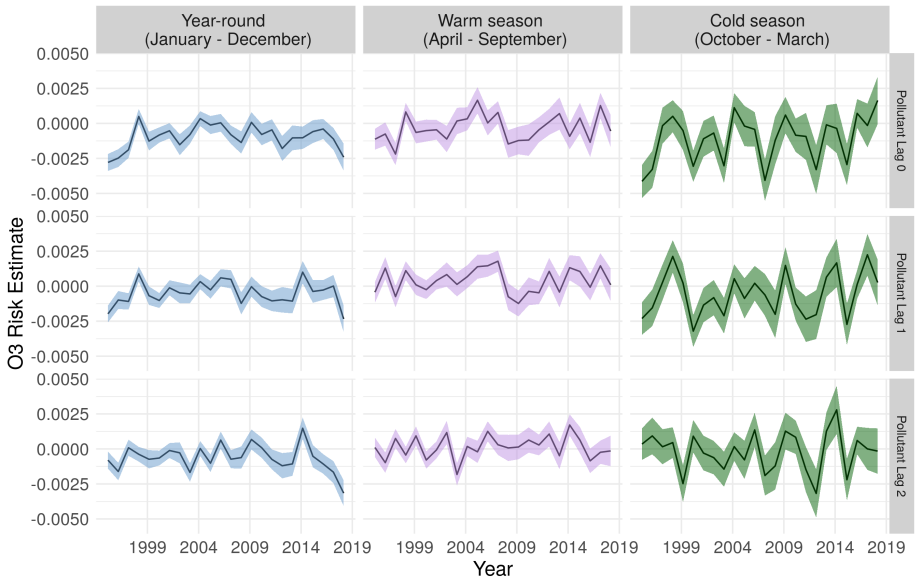


Figure B.3: Annual morbidity (both cardiovascular and pulmonary) risk estimates for O_3 in Toronto, ON, for three seasonal breakdowns (all year; warm season; and cold season) and for three lags on the pollutant data (o days; 1 day; and 2 days)

C. *Bivariate Risk Estimates*

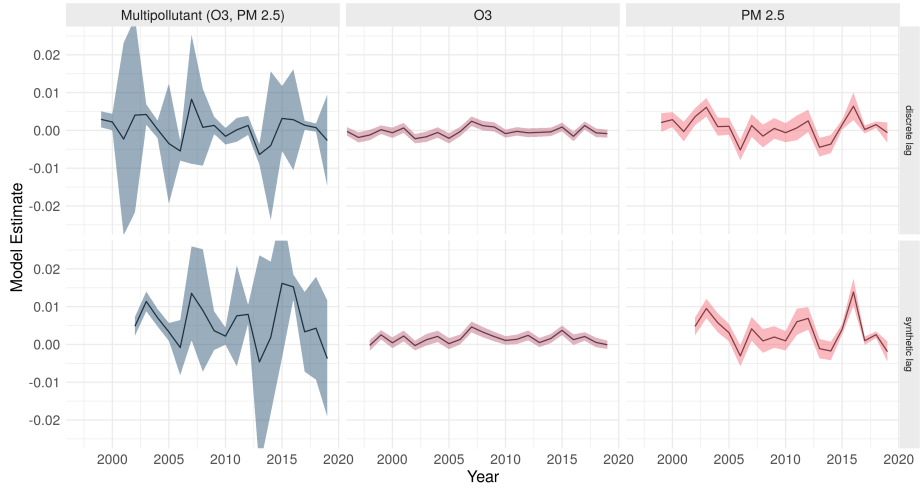


Figure C.1: Comparison of single and multi pollutant morbidity models for synthetic and discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC

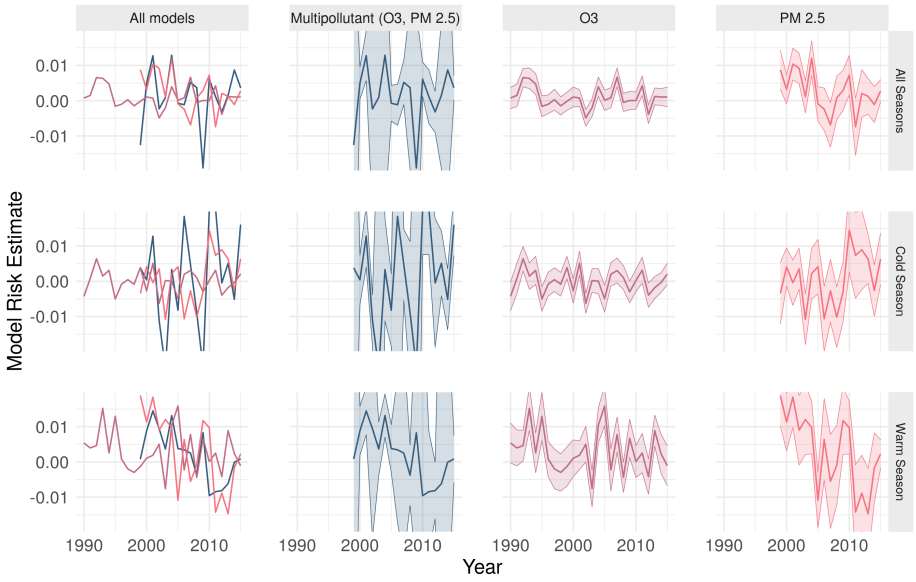


Figure C.2: Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC

Figure C.3: Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Greater Vancouver, BC

Figure C.4: Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Edmonton, AB

Figure C.5: Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Edmonton, AB

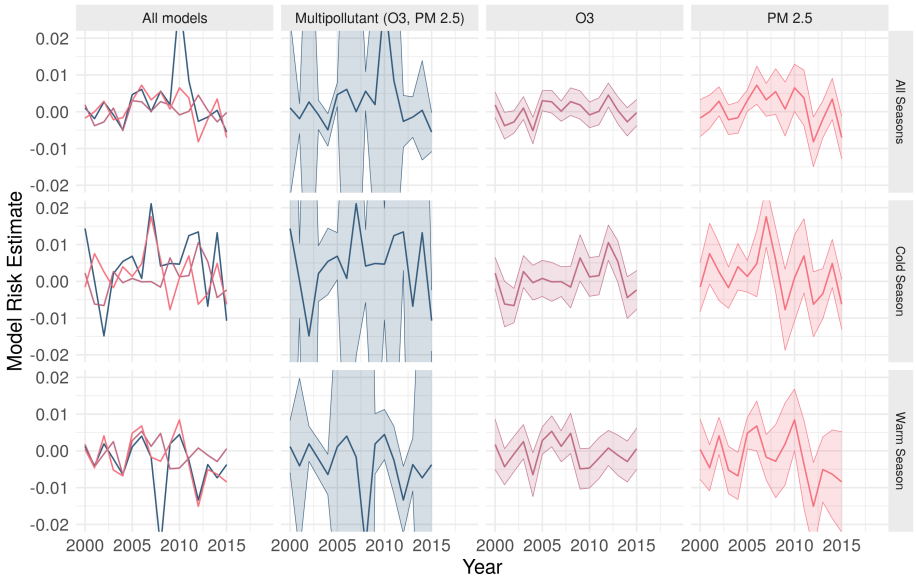


Figure C.6: Comparison of single and multi pollutant mortality models for discrete lag, using a SLP with 12 df for the smooth function of time in Ottawa, ON

Figure C.7: Comparison of single and multi pollutant morbidity models for discrete lag, using a SLP with 12 df for the smooth function of time in Ottawa, ON