tps://theplanetd.com/tofino-and-ucluelet-british-columbia-canada/

Trend analysis of background levels of ground-level ozone in Pacific Northwest

Preparation for renewal of the Regional Ground-Level Ozone Strategy

Prepared by: Carlina Kim, UBC Sustainability Scholar, 2020

Prepared for: John Lindner, Air Quality Planner, Metro Vancouver

August 2020

Disclaimer

This report was produced as part of the UBC Sustainability Scholars Program, a partnership between the University of British Columbia and various local governments and organizations in support of providing graduate students with opportunities to do applied research on projects that advance sustainability across the region.

This project was conducted under the mentorship of Metro Vancouver. The opinions and recommendations in this report and any errors are those of the author and do not necessarily reflect the views of Metro Vancouver or the University of British Columbia.

Acknowledgements

The author would like to thank the following individuals for their contribution, feedback, and support throughout this project.

Kyle Howe – Air Quality Planner, Metro Vancouver
John Lindner – Air Quality Planner, Metro Vancouver
Derek Jennejohn – Lead Senior Engineer, Metro Vancouver
Karen Taylor – Program Manager, University of British Columbia

Additional thank you to the following individuals for their guidance on the data and analysis.

Bruce Ainslie – Environment and Climate Change Canada
Keith Jones – Environment and Climate Change Canada
Ian McKendry – UBC Geography
Marina Richter – Fraser Valley Regional District
Jerome Robles – BC Ministry of Environment and Climate Change Strategy
Annie Seagram – BC Ministry of Environment and Climate Change Strategy
Corrine Schiller – Environment and Climate Change Canada

Cover photo of Ucluelet from: https://theplanetd.com/tofino-and-ucluelet-british-columbia-canada/

Contents

List of Figures

List of Tables

Executive Summary

This data analysis report was written in support of the planned renewal of the *Regional Ground-Level Ozone Strategy* (RGLOS). Specifically, the project investigated the background levels of ground-level ozone from the decommissioned Marine Boundary Layer Station (MBLS) in Ucluelet, British Columbia. One main focus was to find the underlying trend of background ozone over the seven years of operation of the MBLS, after removing for seasonality and other contributing variables.

We also explored the impact of precursor emissions on MBLS background ozone levels, classified geographic source regions of ozone levels on a given day based on back-trajectories, as well as compared MBLS to ozone trend levels at other Pacific Northwest monitoring stations. The other stations are in Whistler, BC as well as Trinidad Head, Mt. Bachelor, and Cheeka Peak, all in the United States. We also compared MBLS pollutant levels to air quality ambient objectives for ozone and nitrogen dioxide.

Using the statistical software R, we used various air quality packages including `openair`, `forecast`, and `deweather` to conduct time series analysis, correlation between pollutants, and air quality modelling. We applied several statistical methods to extract the trend levels for all stations including STL decomposition, TheilSen from `openair` and Boosted Regression Tree.

Main findings for MBLS include:

- Overall trend is decreasing at -0.4 ppb/year according to TheilSen statistical package.
- Ozone levels peak in the spring and minimum levels occur in the summer.
- Ozone levels are lowest during mornings and peak in the late afternoons.
- Minimal correlation of ozone with the other pollutants including particulate matter, sulphur dioxide, carbon monoxide, carbon dioxide, nitrogen oxides and nitrogen dioxide.
- Ozone trends found using STL decomposition and Boosted Regression Tree are similar with increasing annual trend observed until late 2013, where it starts decreasing.
- Ozone and nitrogen dioxide levels are below Metro Vancouver and provincial ambient air quality objectives, with ozone 8-hr metrics to be decreasing over the years.
- For 10m and 100m back-trajectory source regions, higher ozone levels are likely transported from Asia, California and Hawaii. Lowest levels came from Aleutian region for both measurements.

Main findings in comparing MBLS to other stations include:

- Cheeka Peak could potentially be used as a proxy ozone background station for MBLS. Additional analysis is needed to confirm its applicability, including statistical significance such as the WilCox method and comparisons against other pollutants.
- Ozone trends found using STL decomposition show that most station's trends do not line up with MBLS.
- Cheeka Peak and MBLS ozone levels have a seasonal alignment when plotted side by side in monthly ozone distributions.
- TheilSen from `openair` show Cheeka Peak to have a weak downward ozone trend and Mt Bachelor to have a weak upward trend. Trinidad Head has evidence of a strong upward trend with a statistical significance of $p > 0.001$.

Moving forward, we recommend additional evaluation of ozone levels at Cheeka Peak, to confirm its applicability as a proxy background ozone station in place of MBLS. We also recommend using the same time series analysis tools for air quality stations in the Lower Fraser Valley to see what the ozone trends are within the LFV region. We also suggest exploring other statistical methods for trend analysis, such as the Kolmogorov-Zurbenko method and plots such as wind rose/ pollution rose plots, which were not conducted due to limited project time.

Further, we suggest investigating the reasoning behind the observed trends in our region such as researching wildfires occurring near the region during the time periods. Lastly, we recommend applying more complex data science and machine learning mechanisms to the air quality data to gain better understanding of pollutant levels and influences from meteorological parameters.

1 Introduction

The RGLOS, published in 2014, provides a "foundation for policies and strategic directions to control ozone precursors in the Canadian Lower Fraser Valley." The Canadian Lower Fraser Valley (LFV) airshed includes Metro Vancouver and the western portion of the Fraser Valley Regional District (FVRD). RGLOS was adopted by Metro Vancouver, FVRD, BC Ministry of Environment and Climate Change Strategy (BCENV), Environment and Climate Change Canada (ECCC), and the Port of Vancouver. Metro Vancouver is responsible for managing air quality within the Metro Vancouver region by "continuing to identify air contaminants in the region, identifying priorities and pursuing effective actions to reduce pollutants," as stated in their 2019-22 Board Strategic Plan.

To support a planned renewal of the RGLOS, additional research is needed to understand how background levels of ground-level ozone impact the measured ozone levels in the LFV airshed. Specifically, this data analysis project investigated pollutant datasets from the former decommissioned Marine Boundary Layer Station (MBLS) in Ucluelet, British Columbia.

The main objectives of the project are:

- To analyze and summarize trends of background levels of ground-level ozone from MBLS, Ucluelet, British Columbia;
- To explore the potential impact of precursor emissions on background ozone levels measured at MBLS;
- To classify background ozone levels at MBLS based on their likely geographic source;
- To compare MBLS pollutant levels to air quality ambient objectives for ozone and nitrogen dioxide; and
- To compare MBLS data with current trends of ozone from other background stations across the Pacific Northwest.

2 Background

2.1 What is Ground-level Ozone?

Ground-level ozone is a secondary pollutant produced near the surface of the earth at the troposphere level and can cause environmental, health, and economic impacts. It is a secondary pollutant because it is formed when nitrogen oxides (NO_x) and volatile organic compounds (VOC), the "precursor" chemicals, react in the presence of sunlight as shown in

Figure 1 (Metro Vancouver, 2014). Sunlight is an important part in the production of ground-level ozone and generally, the highest levels in the Canadian LFV are typically observed during the late afternoons of hot, sunny summer days (Chen, 2019). In the Canadian LFV, there has been a trend of non-peak ozone levels rising over time in both western and eastern portions of our region which is of concern due to health impacts from ozone exposure, especially towards immunocompromised and vulnerable communities (Metro Vancouver, 2014).

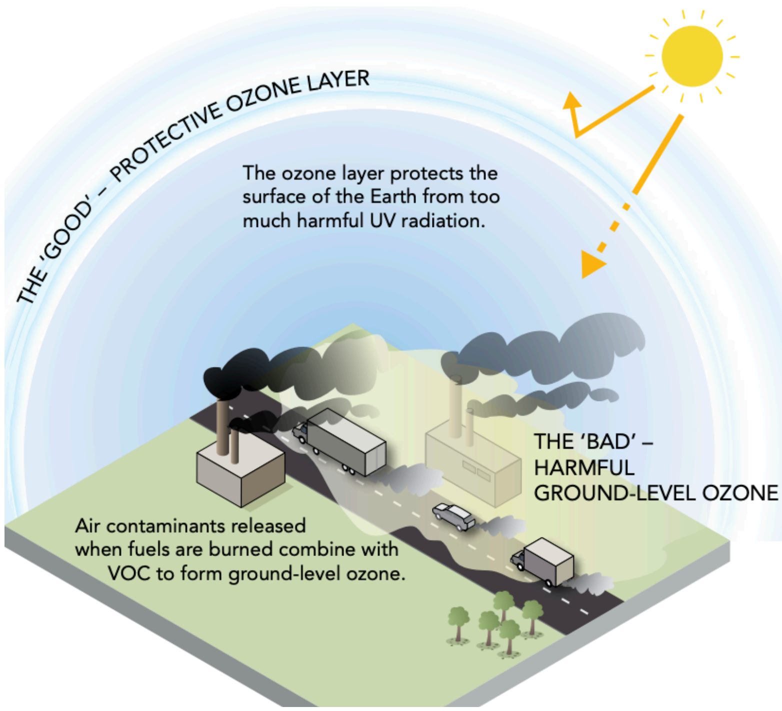


Figure 1. Ground-level ozone versus the ozone layer. Source *Caring for the Air 2012* (Metro Vancouver, 2014).

An additional concern is the potential impact of medium and long range transport of pollutants from human activities outside the Pacific Northwest, including from Asia, on background ozone levels in the Canadian LFV (Metro Vancouver, 2014). For these health and environmental reasons, Metro Vancouver and its RGLOS partners are interested in the background levels of ozone and how it could affect the ground-level ozone levels in the Canadian LFV.

2.2 Marine Boundary Layer Station

"Background" ozone is defined in the Environmental Protection Agency (US EPA) as the "concentrations that would occur in the absence of human emissions and free of contamination from local, regional, and continental sources" (McKendry *et al*, 2014). It is the level of ozone present even without local contributing emissions. The Marine Boundary Layer Station (MBLS) was installed at Amphitrite Point in Ucluelet, British Columbia because previous research indicated that the location was minimally influenced by anthropogenic emissions in North America (McKendry *et al*, 2014). Therefore, this station was suitable for observing our region's background ozone levels.

The station collected data for the period from June 2010 to July 2017, inclusively. The station was operated in partnership by the BCENV, ECCC, and Metro Vancouver. Pollutants measured include ozone (O_3), nitrogen dioxide (NO_2), nitrogen oxide (NO_x), sulphur dioxide (SO_2), carbon dioxide (CO_2), carbon monoxide (CO), and fine particulate matter ($PM_{2.5}$). Ozone was measured in parts per billion (ppb) and hourly averages were provided. ECCC has preprocessed the MBLS data, corrected for outliers, and handled for other errors prior to the analysis described in this report. No processing was done to account for the impacts of wildfires on ozone levels in 2017, given limited project time.

2.3 Comparing other Pacific Northwest Background Stations

In July 2017, the MBLS was decommissioned, ending data collection. Thus, Metro Vancouver and its RGLOS partners are interested whether other stations in the Pacific Northwest region could potentially be used as proxy background stations for the Canadian LFV to support future policy decision-making, particularly around ground-level ozone. The trends of background ozone levels at these stations were calculated and compared with MBLS ozone levels to observe any similarities between them.

The stations used for comparison and were chosen through guidance and recommendations made by ozone experts, including Dr. Bruce Ainslie (ECCC), Dr. Corinne Schiller (ECCC), Dr. Ian McKendry (UBC) and Metro Vancouver staff. The comparison stations were:

1. Trinidad Head Observatory in California, USA
2. Mt. Bachelor Observatory in Oregon, USA
3. Whistler station in Whistler, BC
4. Cheeka Peak station in Washington, USA

The locations of these stations in relation to MBLS are shown in Figure 2. Trinidad Head data was downloaded from the National Oceanic and Atmospheric Administration Earth System Research Laboratories[1], Mt Bachelor data from the University of Washington research portal[2], Cheeka Peak data from the US EPA database[3], and Whistler data was provided by the BCENV[4]. For all stations, hourly data was downloaded except for Cheeka Peak, where only daily averages were available.

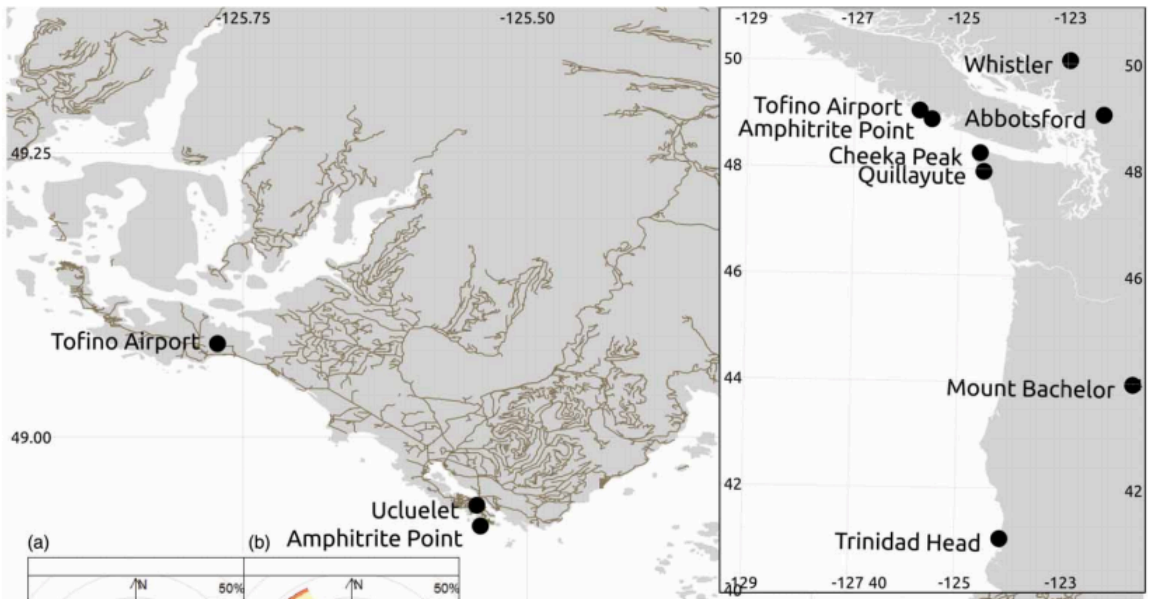


Figure 2 Map showing the location of Ucluelet in relation to other Pacific Northwest air quality monitoring stations. Map on the left shows the location of Amphitrite Point, where the Marine Boundary Layer Station was located, relative to the other communities in the region. Map on the right shows the other four comparison stations in relation to the Ucluelet station. Comparison stations are Whistler, Cheeka Peak, Mount Bachelor, and Trinidad Head (Abbotsford and Quillayute were not used). Map provided by McKendry 2014.

Air quality monitoring stations within the Canadian LFV (specifically 'YVR Airport' and 'Tsawwassen') were initially considered for analysis as potential proxies for a background station, but were determined to be too close to local emission sources, though they could be evaluated in the future.

[1] NOAA Earth System Research Laboratories
[2] University of Washington research portal
[3] US EPA Map of Air Quality Monitors
[4] BC Government air data portal

2.4 Ambient Air Quality Objectives

To support continuous improvement of air quality, Metro Vancouver and other jurisdictions adopts ambient air quality objectives. Metro Vancouver's ambient air quality objectives are largely aligned with provincial objectives and federal Canadian Ambient Air Quality Standards (CAAQS) (Metro Vancouver, 2019). These health-based objectives are used to provide context on current or historical air quality trends and help guide air management decisions (Metro Vancouver, 2020). The MBLS data was compared against the ozone and nitrogen dioxide objectives to determine whether the measured levels achieve or exceed the objectives in use in the Canadian LFV.

Table 1 shows the current Metro Vancouver air quality objectives for ozone and nitrogen dioxide, updated in January 2020.

Table 1: Metro Vancouver's Ambient Air Quality Objectives. Updated in January 2020 (Metro Vancouver, 2020)

Air Contaminant	**Averaging period**	**Ambient Air Quality** $\mu g/m^3$	**Objective** parts per billion
Nitrogen dioxide (NO_2)	1-hour	113	60
	Annual	32	17
Ozone (O_3)	1-hour	161	82
	8-hour	122	62

This project specifically assessed ozone and nitrogen dioxide given the project's focus on ozone formation. Under the ambient objectives, ozone levels are calculated using the annual 4th highest daily maximum 8-hour average concentration, averaged over three consecutive years. For nitrogen dioxide, the levels are calculated using the annual 98th percentile of the daily maximum 1-hour concentration, averaged over three consecutive years. The annual nitrogen dioxide levels, were calculated as the average of the valid 1-hour concentrations, over the year. (Metro Vancouver, 2020)

Note that there are currently no ambient air quality objectives for VOC.

2.5 Back-trajectories

Back-trajectories are defined as the past paths of infinitesimally small particles of air as they move through time and space using interpolated measured meteorological fields such as wind speed and direction (Akhtar *et al*, 2003). We can use these back-trajectories to trace back where the measured MBLS ozone concentrations were likely transported from on any given day. This was done using cluster analysis where MBLS's back-trajectories are grouped together by similar air mass origins (Carslaw, D. 2015) to create cluster region labels for any given day. The clustered data was provided by ECCC and we post-processed the ozone levels according to these cluster regions to gain more insight on the measured pollutant's likely source.

3 Analysis of Marine Boundary Layer Station

The following methodologies were all developed using R, an open source scripting language, commonly used in scientific analysis. The software provides many air quality packages to support statistical analysis and modelling. The R software can be downloaded at https://www.r-project.org/.

3.1 Exploratory Data Analysis

Several exploratory data analyses were conducted to gain a better understanding of the MBLS dataset. Background ozone hourly data was collected from 2010-06-01 1:00 to 2017-07-17 11:00 in PDT/PST time[5].

Intra-annual and seasonal variations of the MBLS background ozone concentrations were explored using the `openair` R package, an open source tool for analyzing air pollution data. The results for the intra-annual and seasonal variations included time of day, monthly, weekday, daily variations and were calculated using `timeVariation` from `openair` (Figure 3).

The bold red line in the plots show the mean levels and the shaded red shows the 95% confidence interval in the mean calculated through bootstrap re-sampling, a random sampling with replacement that assigns measures of accuracy to sample estimates (Carslaw, D. 2015).

[5] To account for daylight savings in R, all hourly data were forced to UTC time zone but kept in their PDT time (-7 UTC). All stations in Pacific Northwest region were in PDT/PST time.

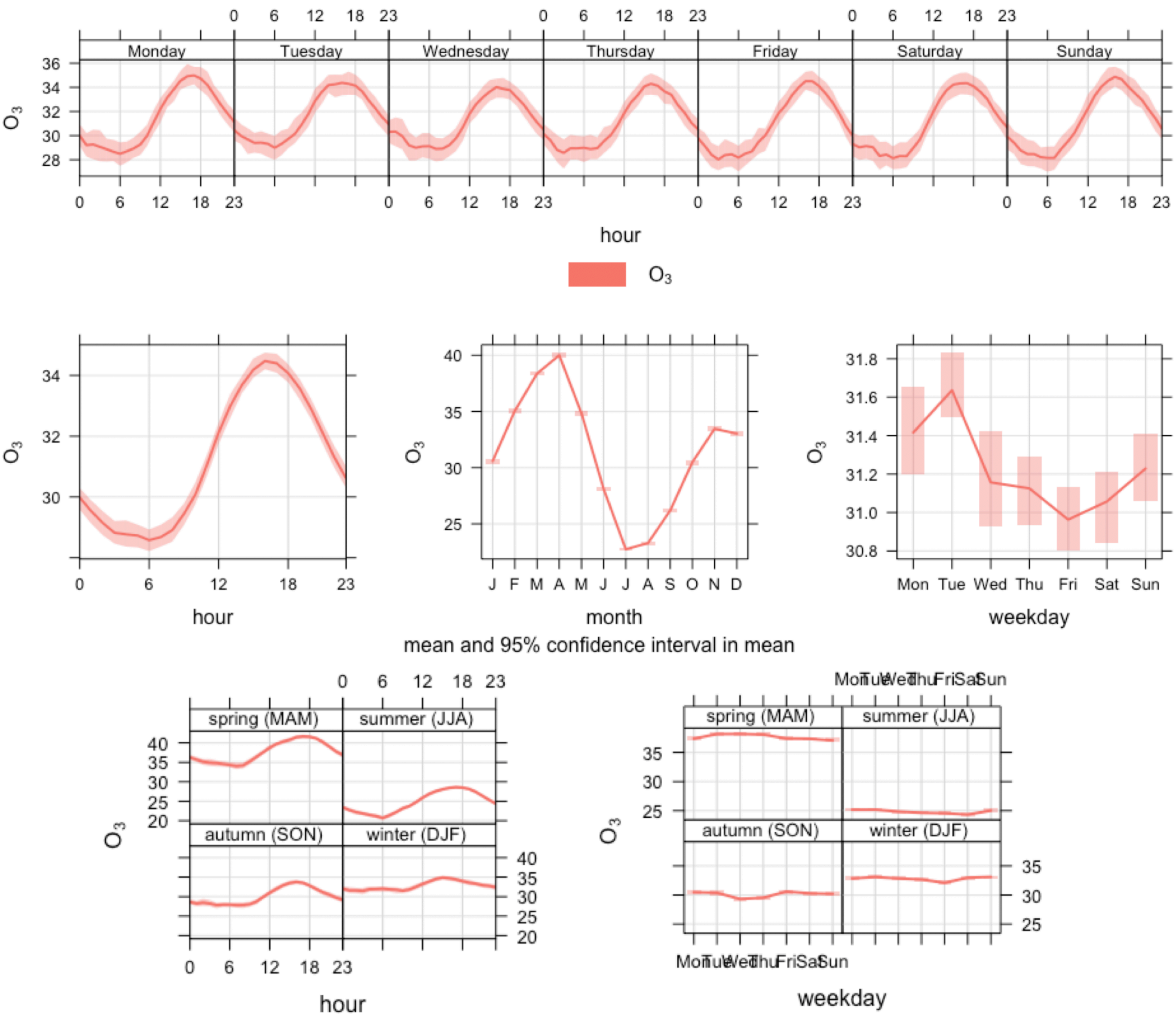


Figure 3: Results of timeVariation from `openair` package for ground-level ozone levels (ppb) from the Marine Boundary Layer Station. First 4 plots show combined hour or day-day of week plot, mean hour of day variation (using Pacific Daylight Time), monthly plot and day of week variation. Last row's 2 plots show the hourly and day of week variations by season. Bold red line show the mean levels and shaded red show the 95% confidence intervals.

Some of the key findings for background ozone levels from MBLS in Figure 3 are:

- The monthly graph (center) shows that the spring season has the highest monthly mean ozone concentrations with peak levels in April, followed by a steep decline until the lowest monthly mean ozone level observed in July.
- The bottom 2 graphs also show that summer seasons observe the lowest ozone levels while the spring seasons observe the highest levels of ozone.

- In the mean hour of day variation plot (center-left), we observe a drop in ozone levels in the morning (6am) and peaks around the afternoon (3pm).
- In the day of week variation plot (center-right), the highest ozone levels are observed to be on Tuesdays. The differences between the weekdays are minimal and confidence intervals overlap suggesting the finding may not be significant.

3.2 MBLS Pollutants Analysis

We also used `timeVariation` to compare the other pollutants measured at MBLS with ozone. Pollutants can have different ranges from each other, so values were normalized by dividing the concentration of the pollutant by its mean value. We compared ozone levels to measured SO_2, NO_x, NO_2, CO, and $PM_{2.5}$.[6] The results of the normalized pollutants are shown in Figure 4 below for averaged hourly, monthly, and weekday.

As before, the bold line in the plots show the 95% confidence interval in the mean and the shaded red shows the uncertainty intervals calculated through bootstrap re-sampling, a random sampling with replacement that assigns measures of accuracy to sample estimates (Carslaw, D. 2015).

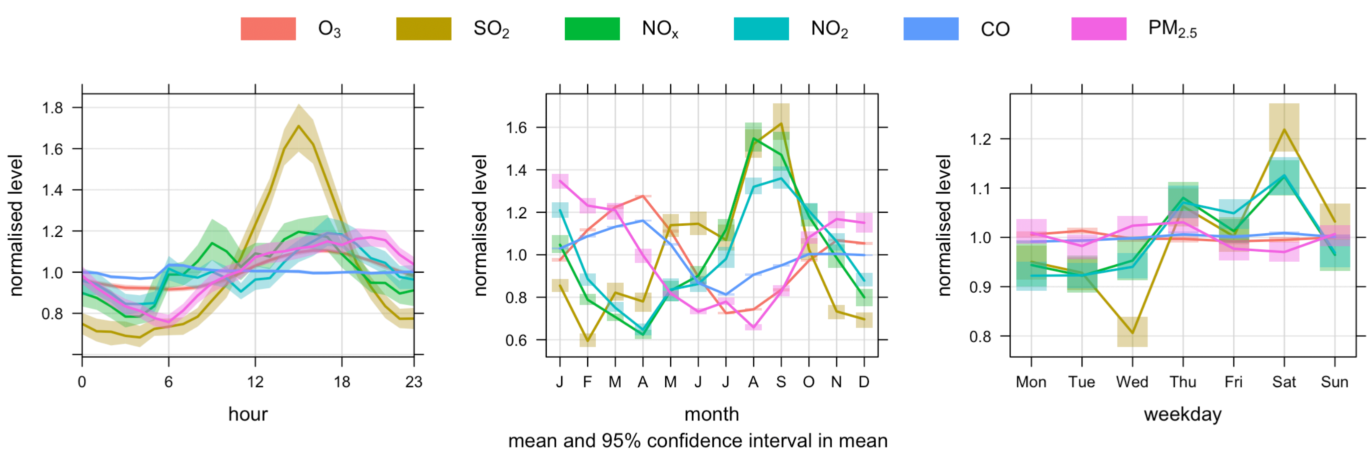


Figure 4: timeVariation function in `openair` package results of pollutants from Marine Boundary Layer Station. Plots show a combined mean hour of day variation, monthly plot and day of week variation on normalized values of pollutants. Bold lines show the mean levels and shaded areas show the 95% confidence interval in the mean.

[6] CO_2 and NO were removed for the analysis in Figure 6. CO_2 was observed to be constant in their normalized state while normalized NO had fluctuating behaviour that made the graph hard to interpret.

Some of the key findings from the normalized values of pollutants graphs in Figure 4 are:

- In the mean hour of day and weekday plots, $PM_{2.5}$ and CO have similar patterns to ozone, where they all remain relatively constant over time.
- In the monthly plot, SO_2, NO_x and NO_2 are observed to have opposite patterns to O_3, where an increase in concentration for these pollutants are in the spring and summer months while ozone is decreasing during these months.
- Higher SO_2 values are likely associated with marine vessels operating near MBLS.
- In the day of week variation plot, SO_2 has the most fluctuating pattern with lowest levels on Wednesdays and highest levels on Saturdays. Ozone is relatively constant in comparison.

Since air pollutants have non-normal distributions, most pollutants are skewed. As most measurements are low with periods of extreme highs, normalization does not typically apply. Given time constraints we were unable to conduct further comparison analysis, but we suggest using non-parametric methods such as a Q-Q plot, a modified scatter plot that uses percentiles, as a more appropriate step for the next analysis.

Another way to observe relationships between pollutants is to calculate the correlation coefficient values for pollutant pairs. This is found using the Pearson correlation coefficients (r) to characterize the strength of correlation between the pollutants. A correlation matrix was created using daily averages (Figure 5) for all the pairs of pollutants collected from MBLS[7]. The coefficients are represented as whole numbers, with 100 being r = 1 (strongest positive correlation) and r = 0 meaning no correlation. Positive correlations indicate that the correlated pollutant falls when ozone falls, while negative correlations indicate that when ozone falls the correlated pollutant rises (McKendry *et al*, 2014).

[7] See in Appendix A for the correlation matrixes separated by season.

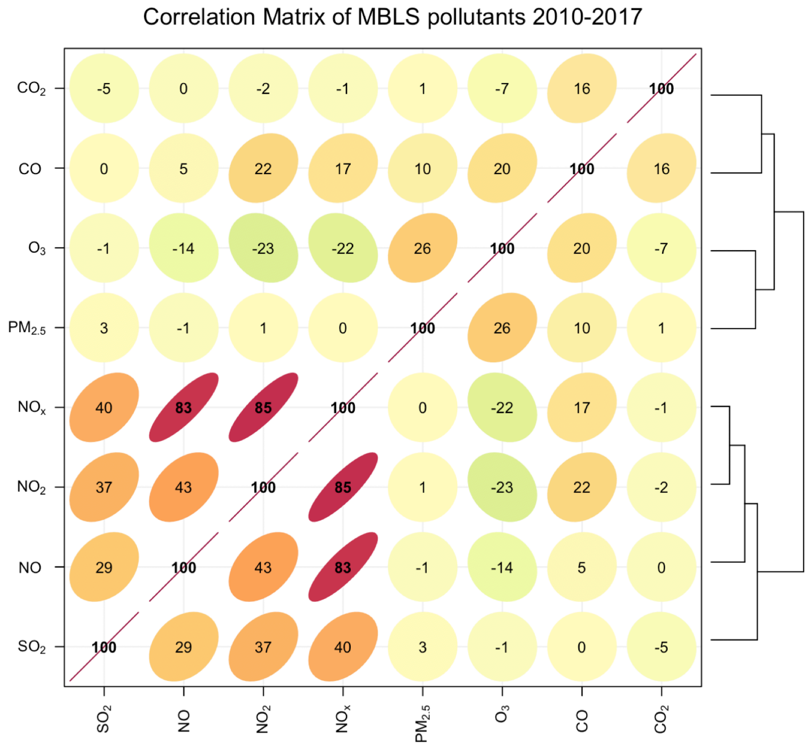


Figure 5: Correlation matrix of MBLS pollutants using corPlot from `openair` package. Ellipses are visual representations of scatter plot to show the relationship, 0 correlation shows a perfect circle shape. The values represent the correlation coefficient in whole numbers, with 100 being r = 1. A dendogram is displayed on the right to visualize how groups of variables are related to one another.

Results show there is no clear strong correlation between ozone and the other pollutants, with the strongest being $PM_{2.5}$ with a weak positive correlation (r = 0.26). Additionally, we observe a weak negative correlation for O_3 with NO_x and NO_2 with r= -0.22 and -0.23 respectively. This makes sense as ozone is a secondary pollutant that forms when NO_x reacts with other precursor emissions, so as NO_x decreases we will generally observe an increase in ozone levels. NO_x is also observed to have high positive correlations with NO_2 and NO with r = 0.83 and 0.85 respectively.

This analysis shows no strong evidence of precursor emissions having an impact on background ozone levels measured at MBLS, based on daily averages. Additional analysis could explore these relationships further.

3.3 Time Series Analysis: STL Decomposition

In time series, there are various components that describe the observed data including trends, seasonality, and repeating cycles. As we are observing data which fluctuate between seasons and have shown cyclic patterns, we need to remove the seasonality and other parameters to observe the underlying background ozone trends over the measured years.

The decomposition methods from the `forecast` package in R splits the time series into three components: the trend-cycle, seasonal, and the remainder (containing anything else like random noise from the time series analysis) (Hyndman, R.J., & Athanasopoulos, G., 2018). For an additive decomposition, we can write the equation as such:

$$y_t = S_t + T_t + R_t$$

where y_t is the data, S_t is the seasonal component, T_t is the trend-cycle component, and R_t is the remainder component, with all at period t.

It should be noted that with air quality data, trends are challenging to isolate given that monitored readings can vary significantly from year to year. In some cases, these variations are due to changing pollutant emissions, but often changing weather patterns can drive large changes in measured pollutant levels. Confirming the presence of a 'trend' in air quality generally requires decades of data. While 'trend-cycle' is the common terminology for decomposition of time series, the 7 years of available MBLS data is insufficient to confirm a true trend in air quality for the Ucluelet area. For clarity, we will describe the 'trend-cycle' as the 'decomposition trend'.

In our decomposition method, we used the "Seasonal and Trend decomposition with Loess" or STL from the `forecast` package. STL is a versatile and robust method for decomposing time series as it uses Loess to decompose. Loess is a locally weighted smoothing method that combines linear least-squares and nonlinear regression (Glen, 2013) to create a smooth line through the time series plot. Note that although STL captures the varying effects in seasonality, we set the s.window = 'periodic' as a first order assumption of constant seasonal pattern over time. Additional analysis could be conducted using a non-periodic window.

Hourly data was aggregated into daily average ozone concentrations and then further aggregated to monthly mean for better interpretability when using decomposition plots. Figure 6 shows the monthly averaged data, the decomposition trend, seasonal, and remainder components for ozone from MBLS over measured years. For STL decomposition, all time zones were converted to UTC for consistency.

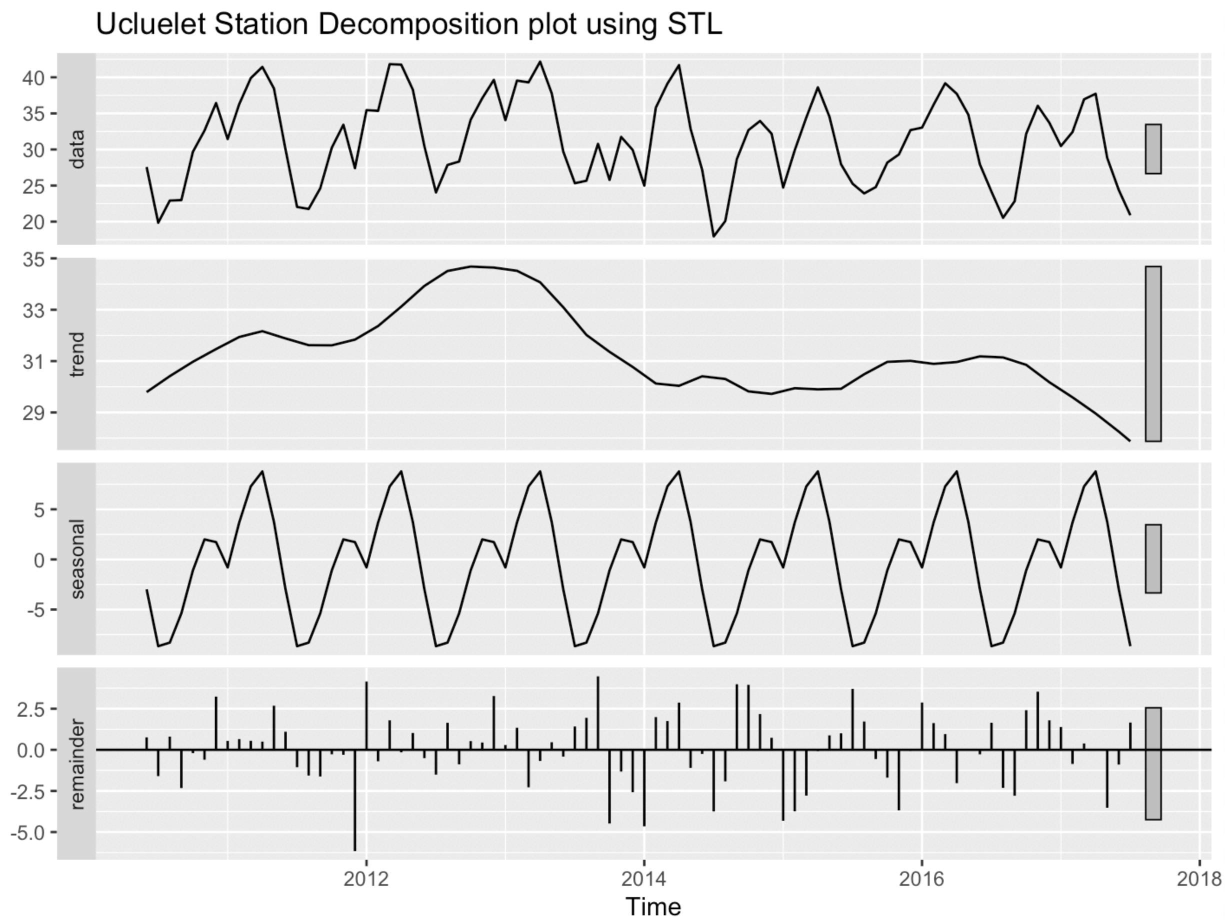


Figure 6: STL additive decomposition plot of MBLS average monthly ozone time series (top plot titled "data") using `forecast` in R showing decomposition trend, seasonal, and remainder components.

As mentioned earlier, we set the seasonality window as "periodic" to assume constant seasonal pattern over time and calculated as averages of de-trended values. Note that this is not always the case for air quality data, but was implemented for simplicity and due to project time constraints. For the ozone decomposition trend in MBLS, we see an increase in background ozone levels from 2010 to 2013 and a slower decrease over time up to 2017.

3.4 Time Series Analysis: Boosted Regression Tree

Another time series analysis method was by using the `deweather` package, developed for the purpose of removing the influence of meteorology from air quality time series data (Carslaw, 2017). The package uses a boosted regression tree (BRT) approach for processing air quality data. BRT combines statistical methods of 1) regression trees (models that relate a response to their predictors by recursive binary splits) and 2) boosting (adaptive method for combining many simple models for improved predictive performance) (Elith, 2008). This can be useful when observing whether the change in concentration is due to emissions or meteorological factors such as temperature, humidity and wind speed. Hourly data was used for the BRT analysis.

Using BRT also allows us to explore the partial dependencies. Partial dependencies show the relationship between the pollutant of interest and the covariates (meteorological factors) used in the model while holding the value of other covariates at their mean level (Carslaw, 2017). We first use a simpler approach of obtaining the background ozone trend of MBLS from 2010-2017 using the BRT approach without any meteorological parameters as covariates (Figure 7).

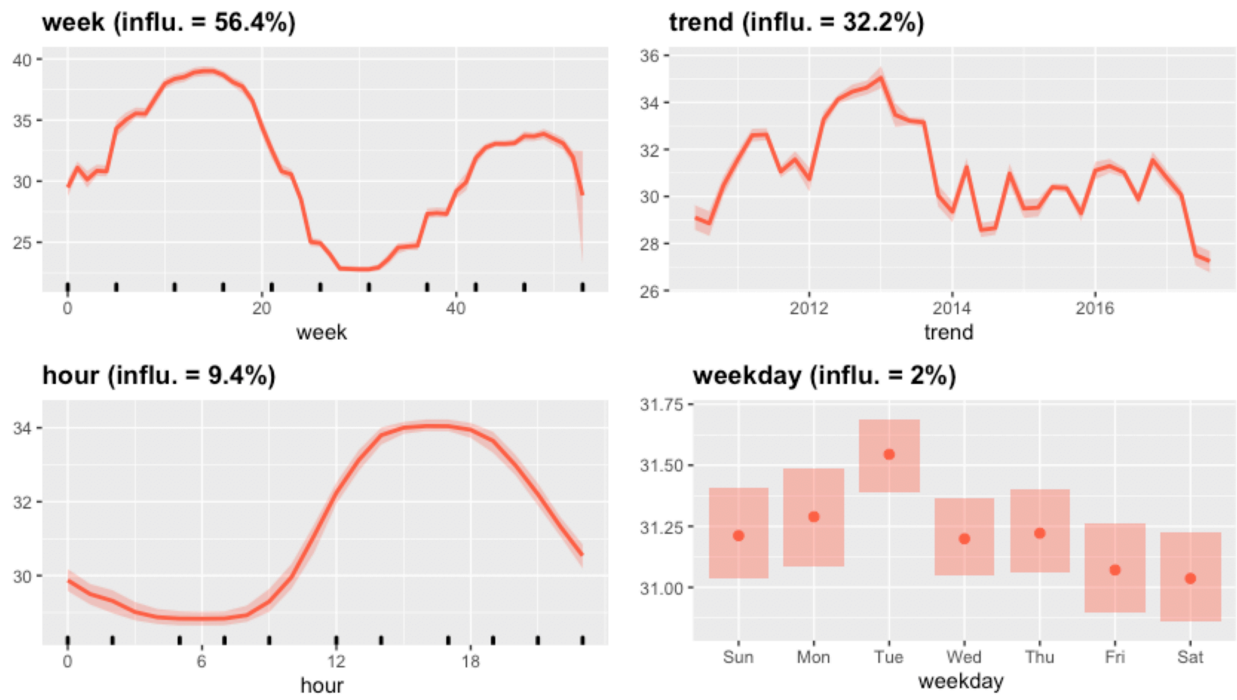


Figure 7: Results using buildMod in `deweather` for 2011-2017 MBLS ozone data. Plot shows the partial dependencies of ozone levels while holding the value of other covariates at their mean. Covariates used here were week, trend, hour, and weekday. Bold red line is mean level and shaded red is 95% confidence intervals.

The MBLS background ozone trend observed using BRT aligns with the monthly and annual trends obtained using STL decomposition (Figure 6). The hour and weekday plots also align with Figure 3's mean hour of day and day of the week variation plots. The influence percentage represents the relative importance and levels of the parameters.

Further, the BC government[8] has provided meta-data from 2011-2013 for MBLS with air temperature, wind speed, wind direction, and relative humidity (RH) data. We used these meteorological parameters, along with the ozone levels from 2011-2013 from MBLS to develop a model that shows a more accurate representation of the trend and to remove any influence of weather variables. Figure 8 shows the results of a BRT model with trend, wind speed, wind direction, hour, weekday, air temperature, RH, and week as covariates.

[8] BC Government data can be found here. Ucluelet MBLS code is E282169.

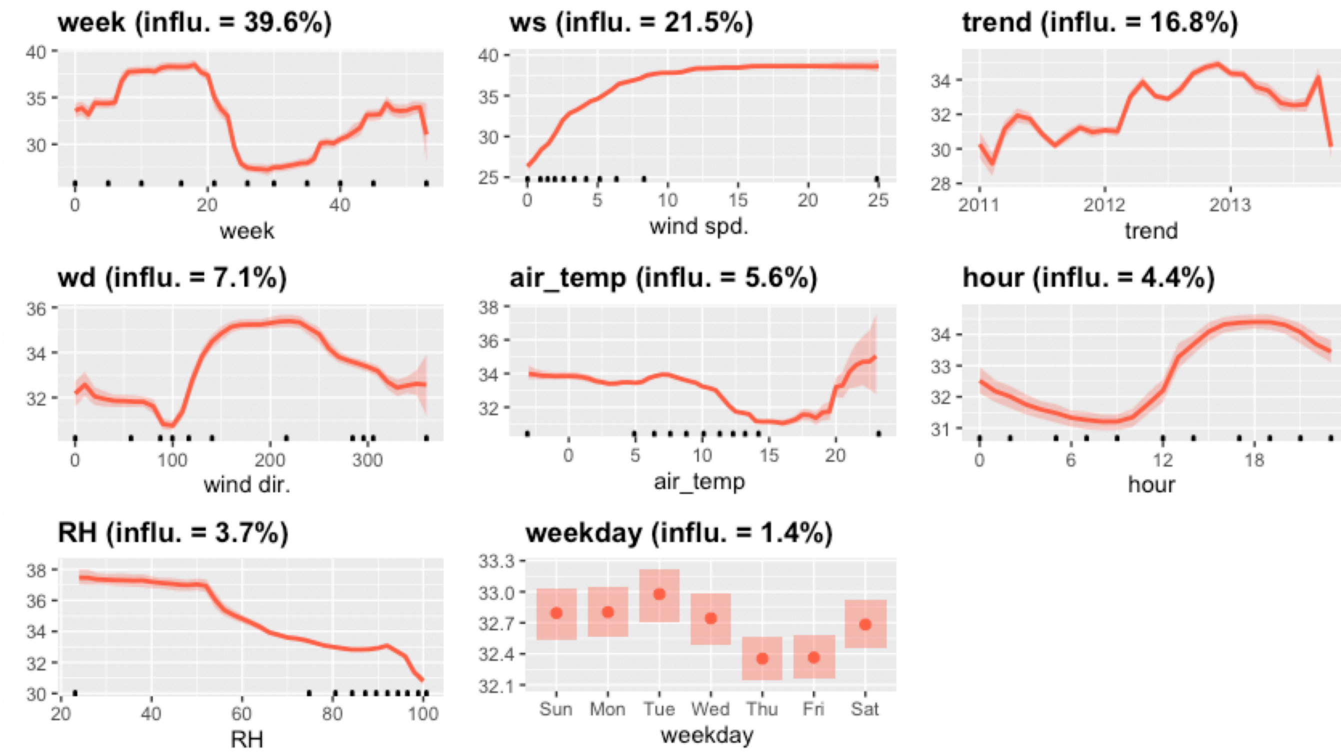


Figure 8: Results using buildMod in `deweather` for 2011-2013 MBLS ozone data. Plot shows the partial dependencies of ozone levels while holding the value of other covariates at their mean. Covariates used here were week, wind speed, trend, wind direction, air temperature, hour, RH, and weekday.

The 2011-2013 MBLS ozone BRT model shows similar patterns for week, hour and weekday as Figure 7. The trend is similar to the one found in Figure 7 for 2011-2013. The influence of the trend has decreased compared to Figure 7's trend influence which may suggest the importance of meteorological parameters on the ozone trend levels. Given project time constraints and limited meta-data, we were unable to run the model on the entire 2010-2017 datasets to explore that potential any further.

We also suggest using these meteorological parameters to create PollutionRose or WindRose plots from `openair`. These plots show how the concentrations of pollutants vary by wind direction or wind speed and time period.

STL decomposition and Boosted Regression Trees were the main two time-series trend analysis techniques carried out for MBLS for their background ozone levels.

Another method that was attempted, but not run due to limited project time and complexity of implementation, was the Kolmogorov-Zurbenko Filter or KZ method recommended by Dr. Bruce Ainslie. The KZ filter is a low-pass filter produced through repeated iterations of a moving average (Wise, 2015) which can be represented as:

$$A_t = e_t + S_t + W_t$$

where A(t) is the original time series, e(t) is the long-term trend component, S(t) is seasonal variation, and W(t) is the short-term component such as weather and short-term fluctuations in precursor emissions (Wise, 2015). This could be implemented in the future to confirm the trend and seasonality results extrapolated for MBLS.

3.5 Back-trajectories calculated to source regions

ECCC clustered the back-trajectories of MBLS data into 6 different source regions to identify where the measured pollutant concentrations likely originated from for any day of the year through 2010-2017. These source regions were calculated from 10m and 100m releases for 121 hour back-trajectories. Then, ECCC clustered the results into specific source regions to simplify the understanding of the potential sources of background ozone and their contribution to measured levels of ozone in the Canadian LFV and other nearby areas.

The source regions are Siberian, Asian, Aleutian, Hawaiian, Californian, and N. American. Figure 9 shows the location of trajectories of these source regions with respect to Ucluelet, BC on a global map.

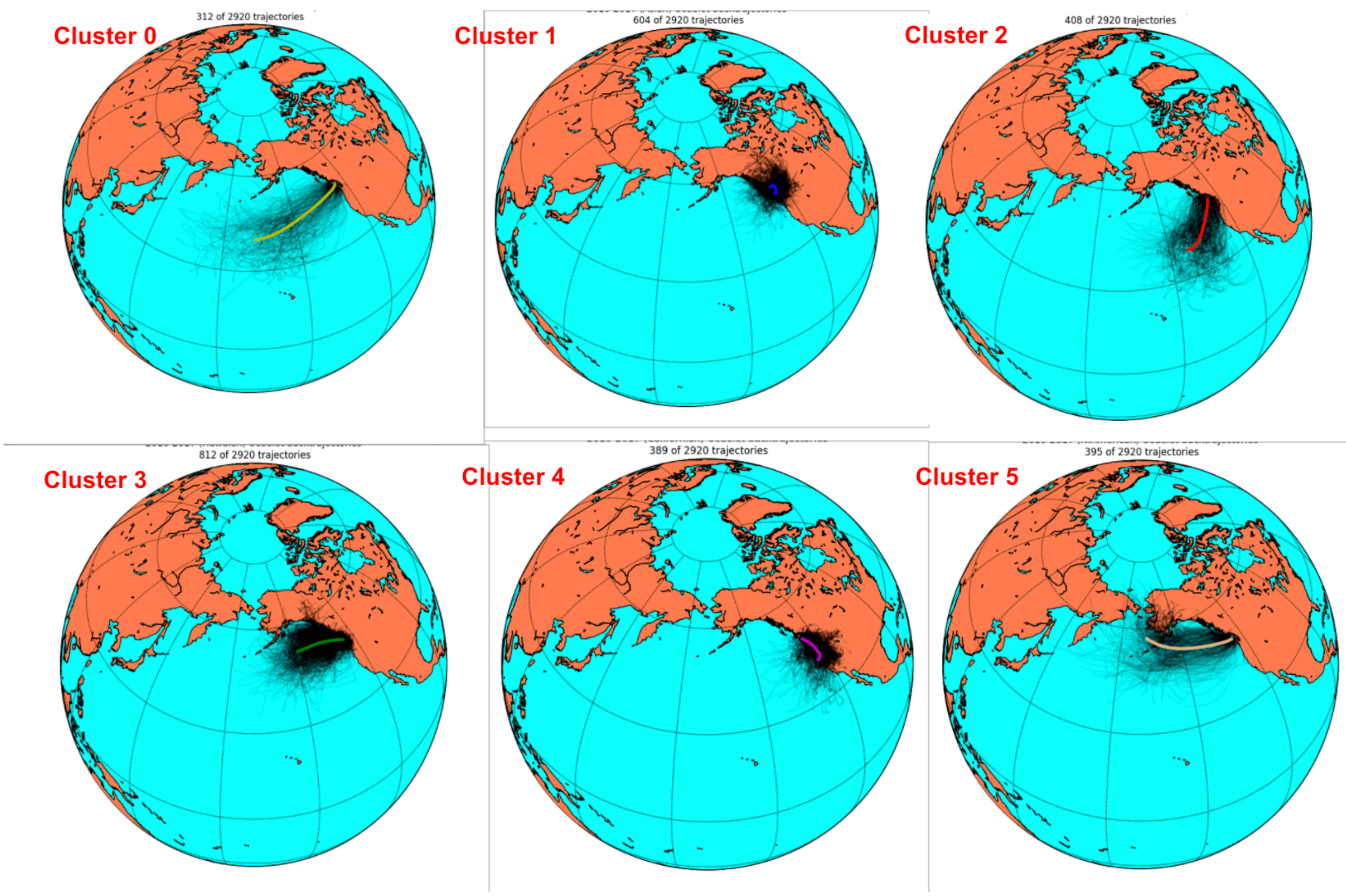


Figure 9: Back- trajectory paths of source regions using 10m release from 2010-2017 MBLS data provided by Environment and Climate Change Canada. Locations include starting from top left: Asian, N. American, Hawaian, Aleutian, Californian, and Siberian.

The source regions are combined with the ozone data from MBLS to determine the distribution of ozone levels from each region.

Each daily mean background ozone level for MBLS was assigned to a cluster region from the back trajectories using data provided by ECCC. These ozone levels for each release measurement were plotted using boxplots and violin plots for each cluster region as shown in Figure 10 and 11. Violin plots allow a clear visualization on the distribution of ozone levels from each region, with the shape representing the density estimate of the data. The more data points in a specific range, the larger the violin is for that range (Holtz, 2018).

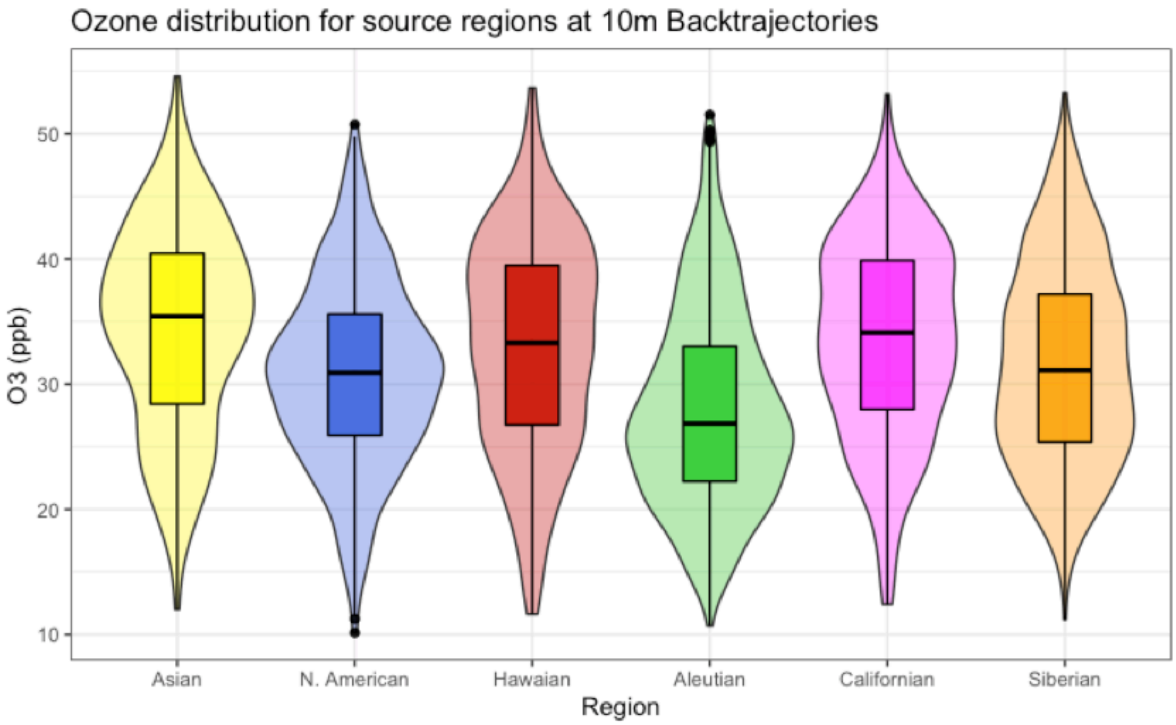


Figure 10: MBLS ozone distribution for 6 source regions clustered by back-trajectory calculations at 10m release. Boxplot shows median and standard deviation. The violin ridges show the density distribution for each region.

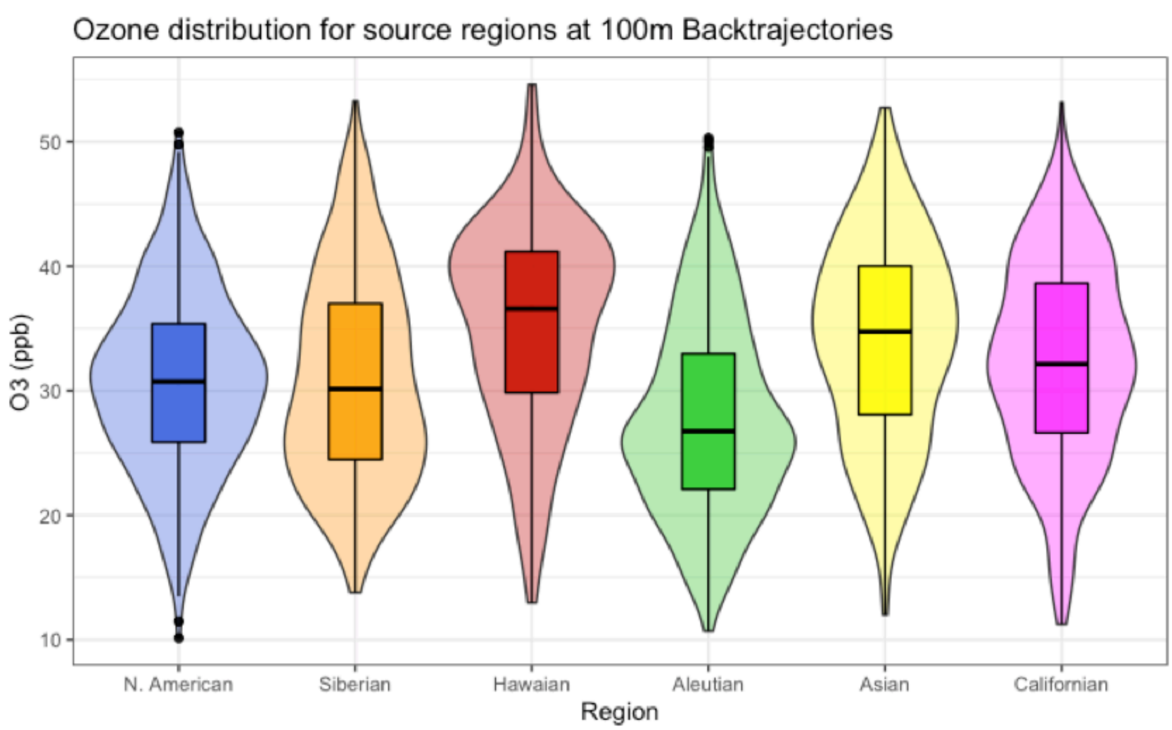


Figure 11: MBLS ozone distribution for 6 source regions clustered by back-trajectory calculations at 100m release. Boxplot shows median and standard deviation. The violin ridges show the density distribution for each region.

The results are shown from in order from Cluster 0 – 5. The colors for the regions in the two figures are the same for comparison.

The ozone distribution of the source regions shows:

- The highest ozone levels for 10m release (Figure 10) were transported to MBLS from the Asian region, followed by California and Hawaii;
- For 100m release (Figure 11), highest ozone level transported to MBLS from Hawaii region, followed by Asia and California;
- The lowest ozone levels were transported to MBLS from the Aleutian region for both 10m and 100m releases.

3.6 Ambient Air Quality Objectives

To calculate the ambient air quality objectives for pollutants, published CAAQS methods were followed to assess data completeness criteria (i.e., is there enough data for the dataset to be considered representative for that season or year) and specific calculation steps. We calculated values for O_3 and NO_2 concentrations from MBLS using code provided by Kyle Howe, Metro Vancouver.

Metro Vancouver's Ambient Air Quality Objective (Table 1 above) for O_3 is 62 ppb (8-hour) and NO_2 is 60 ppb (1-hour). Achievement of the ozone objective is based on the annual 4th highest daily maximum 8-hour average concentration (Metro Vancouver, 2020). Note these values are averaged over 3 years. Table 2 shows the results calculated for MBLS ozone.

Table 2: Marine Boundary Layer Station's ozone measurements compared against Metro Vancouver 8-hour ozone Ambient Air Quality Objective. Averaged over 3 consecutive years on annual 4th highest daily maximum 8 hour-average concentration.

Year	Ozone (O_3) parts per billion (ppb)
2013	53.3
2014	53.9
2015	53.2
2016	52.1
2017[9]	51.25

[9] 2017 objective was based on a 2-year average due to partial data for 2017 (up to July) resulting in the year's invalidity to use for calculation. CAAQS methodology allows using the valid years if 2 of the 3 years average values

The objectives were achieved for ozone levels in MBLS, with values all being below the objective level. We also see a decrease in the 8-hr metrics over the years, which aligns with the decreasing trend value we found using TheilSen (Figure 16) This is expected due to a lack of significant local NO_x and VOC emission sources.

Due to insufficient project time, we did not compare the measured NO_2 levels against the objectives. However, given that the NO_2 levels for MBLS ranged between -0.03 and 23.66 ppb for 2010-2017, we expect the calculated values to be below the objective level. This range is far lower than the NO_2 objective of 60 ppb for its daily maximum 1 hour-average. We also do not expect the NO_2 levels from MBLS to exceed the annual objective of 17 ppb given these low results. Additionally, previous findings at MBLS has found NO_2 mean to be 0.775 ppb for 2010- 2011 (Schiller, et al. 2011) which suggests that nitrogen dioxide levels at MBLS are below the objective.

There is also an R package created to facilitate the calculation of air quality metrics according to the CAAQS. You can find the package here https://github.com/bcgov/rcaaqs which may be used in the next objectives calculations.

pass data completeness and validity requirements (i.e., the non-valid year is ignored). See Appendix A for the data completeness table.

4 Comparison of MBLS to other background stations

We explored the monthly distributions of ozone levels in MBLS and created a similar graph as Professor Ian McKendry produced on his initial investigation on the measured MBLS data (2014). Professor McKendry's original graph is shown in Figure 12 below.

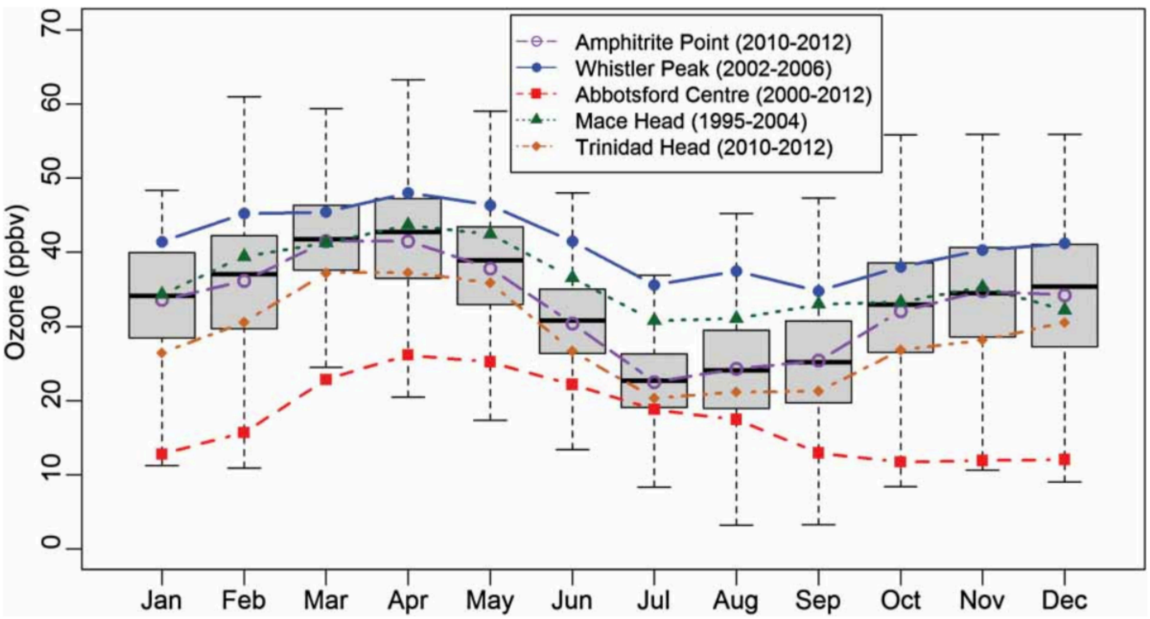


Figure 12: Mean monthly averaged O3 ozone levels from Amphitrite Point Marine Boundary Layer Station, Whistler Peak, Abbotsford Centre, Mace Head in Ireland, and Trinidad Head. The boxplot is for Amphitrite Point Marine Boundary Layer Station and shows median values with box constrained by first and third quartiles. The bottom whisker is the 1st quartile –1.5 × the inter-quartile range, and the top whisker is the 3rd quartile +1.5 × inter-quartile range (McKendry *et al,* 2014).

We calculated the distribution of the hourly background ozone data from MBLS by month using boxplots along with the average monthly ozone trends from the different stations (Figure 13).

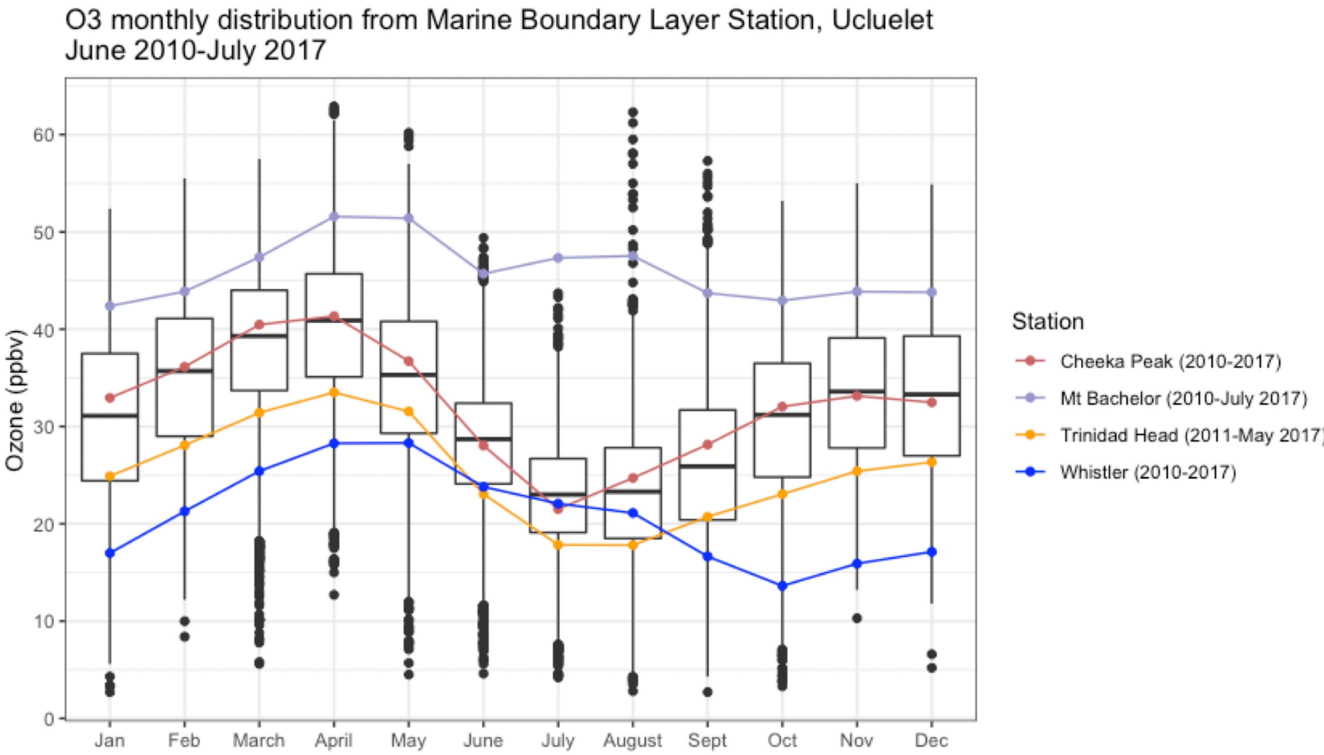


Figure 13: Boxplot of Marine Boundary Layer Station ozone monthly distribution and monthly averaged ozone levels from Cheeka Peak, Mt Bachelor, Trinidad Head, and Whistler. The boxplot is for Marine Boundary Layer Station and shows median values with box constrained by first and third quantiles. The bottom whistler is 1st quantile – 1.5 x inter-quantile range, and the top whisker is the 3rd quartile + 1.5 x inter-quartile range; outliers shown as black dots.

In our boxplot, the monthly ozone distributions of MBLS and monthly averaged ozone trend for Trinidad Head (yellow line) align well when comparing to Professor Ian McKendry's boxplot (Figure 12). Our results for Whistler (blue line) had much lower ozone levels in comparison. It was determined that this Whistler data was from Meadow Park station instead of Whistler Peak High Elevation Research site, which was used for Figure 12. We suggest looking into the Whistler Peak data as a next step in future comparison.

Findings from the boxplot include:

- Cheeka Peak's monthly mean ozone levels follow a similar pattern to MBLS's monthly median values as the points overlay close to the boxplot's median lines.
- Mt. Bachelor has the highest levels of ozone, possibly due to its high elevation.

Due to evidence of Cheeka Peak showing similar ozone trends to MBLS, we created a boxplot of the monthly ozone distributions of the two stations for a more detailed comparison (Figure 14).

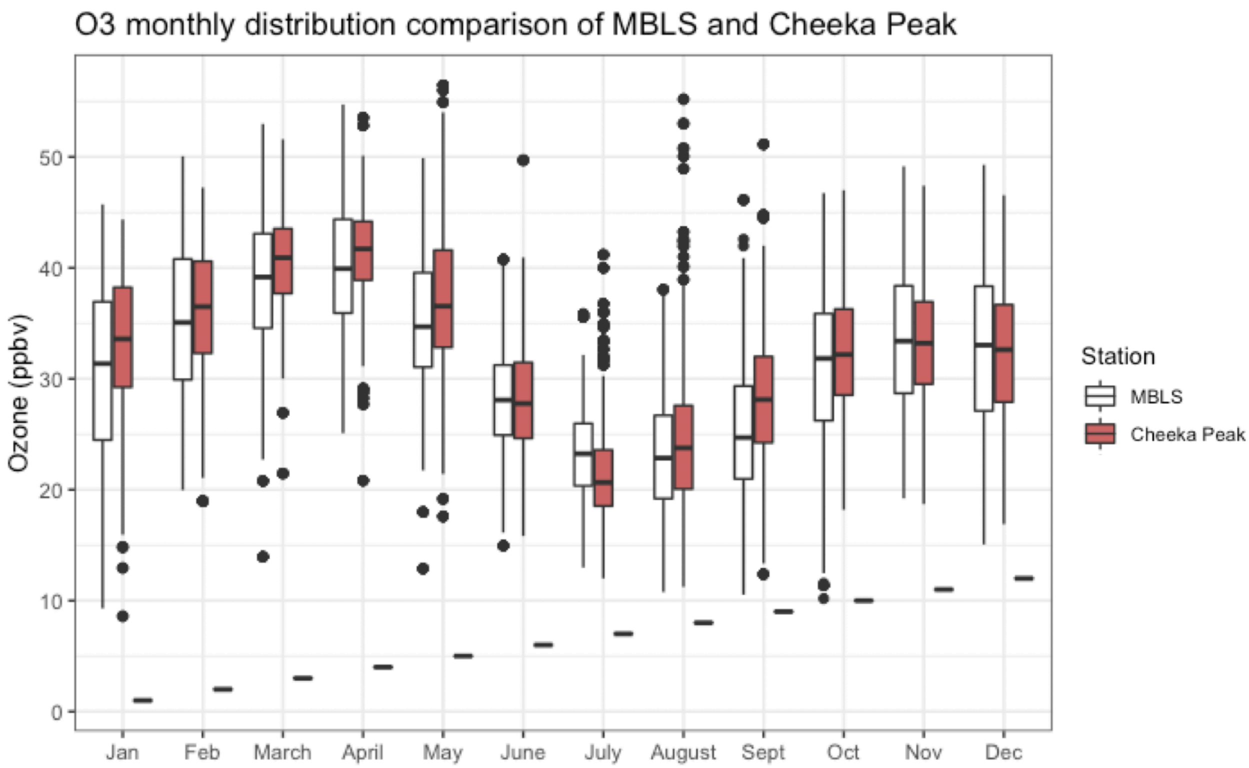


Figure 14: Boxplot comparing monthly ozone distribution of Marine Boundary Layer Station and Cheeka Peak. The boxplots show median values with box constrained by first and third quantiles. The bottom whistler is 1st quantile – 1.5 x inter-quantile range, and the top whisker is the 3rd quartile + 1.5 x inter-quantile range; outliers are shown as black dots.

Cheeka Peak and MBLS show strong alignment in their monthly ozone distributions for 2010-2017, suggesting Cheeka Peak could potentially be used as a proxy background station for MBLS. To conduct a statistical comparison, we suggest using the Wilcox distribution. This function determines the uniformity between two datasets in their distribution levels.

To ensure that the comparisons of the other station's ozone trends to MBLS were accurate, we took several statistical approaches to analyze and compare the time series trends.

4.1 Trend comparison: STL Decomposition

The first approach was using the same method as the MBLS time series analysis by creating STL decomposition plots for all other stations to visually compare the de-seasonalized trends.

Repeating the same method, hourly data was aggregated into daily average ozone concentrations and then further aggregated to monthly mean for better interpretability when using decomposition plots. Cheeka Peak had daily average data available (not hourly), so this ozone dataset was aggregated to monthly mean for STL decomposition. For STL decomposition, all time zones were converted to UTC for consistency.

The STL decomposition across all stations is shown in Figure 15. Each plot shows the monthly averaged data, trend-cycle component, seasonal component and the remainder component.

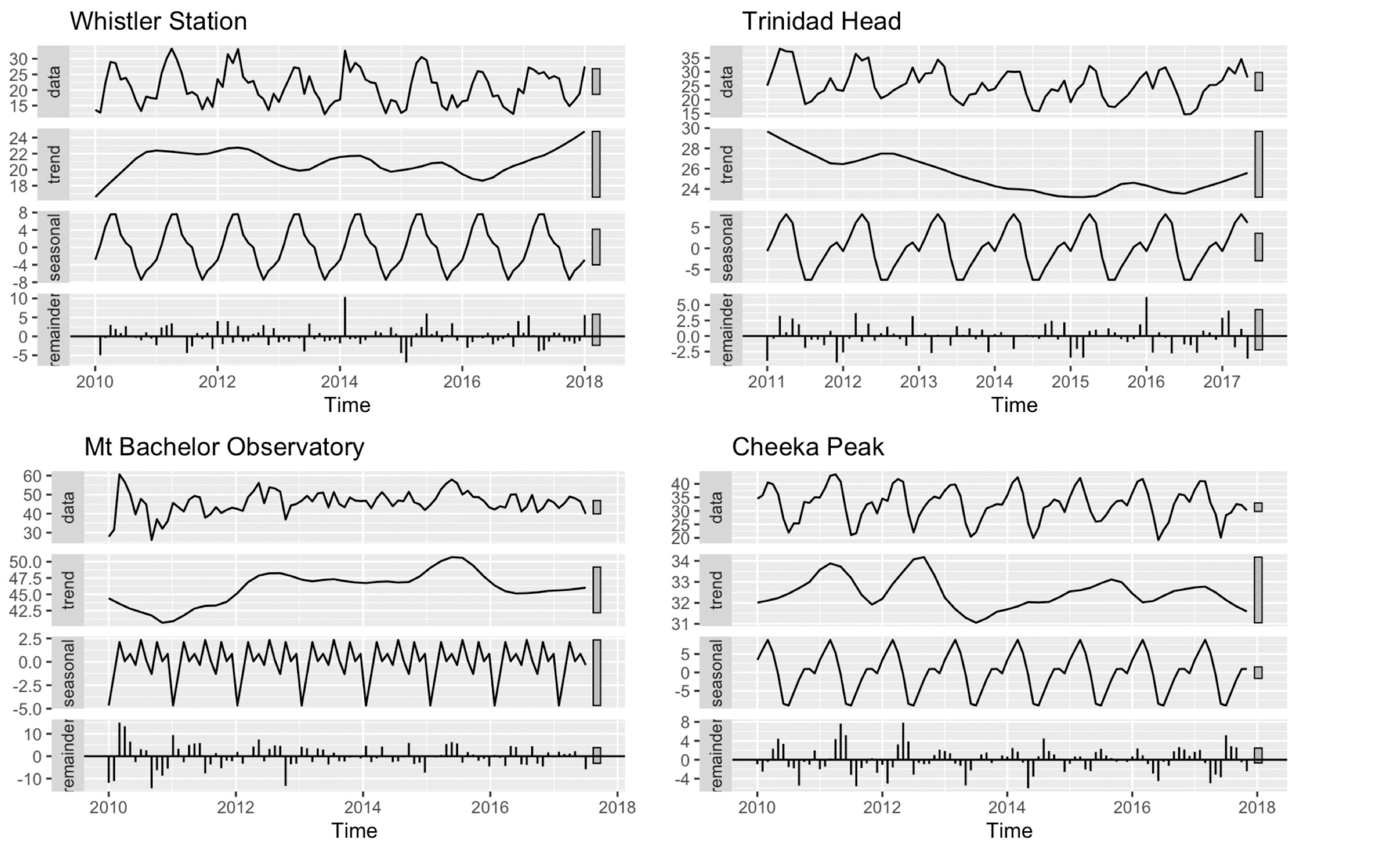


Figure 15: STL decomposition plots for Whistler Station, Trinidad Head Observatory, Mt. Bachelor Observatory, and Cheeka Peak Station based on ozone monthly time series using `forecast` in R.

The stations all have notable trend patterns after removing seasonal influences:

- Trinidad Head has an overall decreasing ozone trend, with a slight increase in ozone levels starting around 2016.
- Mt Bachelor has an increasing trend from 2010 to 2012, a constant trend until we see a decrease in 2016, and levels off until 2017.
- Cheeka Peak has a fluctuating trend from 2010 to 2012, followed by a slow increasing trend until end of 2017.
- The stations do not line up with MBLS decomposition trend.

Whistler's results were based on the incorrect station, Meadow Park station, instead of Whistler Peak High Elevation Research site, which is usually used to compare background stations.

Note that Mt. Bachelor's seasonal component is not in line with the rest of the stations. One reason could be that the station's seasonality is not constant over time where we would need to set t.window to a specific value. Another reason could be that this time series is a multiplicative decomposition instead of additive like the other station's time series. The equation for multiplicative decomposition is:

$$y_t = S_t * T_t * R_t$$

where y_t is the data, S_t is the seasonal component, T_t is the trend-cycle component, and R_t is the remainder component, with all at period t.

The multiplicative decomposition plot for Mt. Bachelor can be found in Appendix A. The plot shows a significant drop at the start of every year, which is not generally observed for the other stations. Note that STL decomposition does not handle multiplicative methods well. Thus, the classical decomposition method was implemented which uses moving averages instead of loess.

4.2 Trend comparison: TheilSen

We also used TheilSen, another time series analysis tool from `openair`, to determine trends in pollutant concentrations over time. The basic understanding of TheilSen is that given a set of x, y pairs, with n pairs, the slopes between all pairs of points are first calculated. The Theil-Sen estimate of the slope is the median of all these slopes. An advantage of using Theil-Sen is that it is resistant to outliers and yield accurate confidence intervals even with non-normal data (Carslaw, D. 2015). It also provides an estimate of a p-value (testing for statistical significance) for the slope through bootstrap resampling.

TheilSen handles hourly data, which was used for all stations except for Cheeka Peak which only had daily averages available, to aggregate and plot monthly mean concentrations. The results using TheilSen for MBLS ozone data is shown in Figure 16, after seasonality was removed.

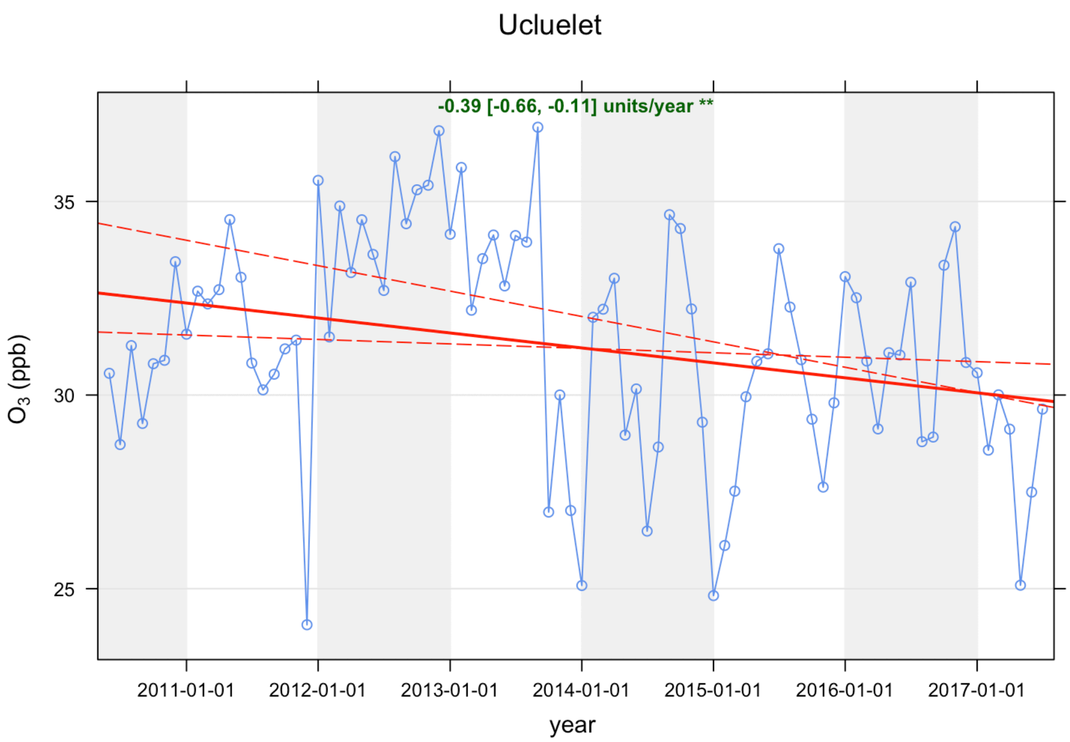


Figure 16: Trends in ozone at Marine Boundary Layer Station using TheilSen from `openair`. Plot shows deseasonalized monthly mean concentrations of O_3. Solid red line shows trend estimate and dashed red shows 95% confidence intervals for trend based on resampling method.

The plot shows the deseasonalized monthly mean concentrations of O_3 from MBLS. The deseasoning is done using STL (seasonal trend decomposition using loess). The solid red line shows the trend estimate and the dashed red lines show the 95% confidence intervals based on runs of simulations of bootstrapping resampling to estimate the uncertainty of the slope (Carslaw, D 2015).

The overall trend value is shown at the top of the graph, with the overall trend being -0.39 ppb per year and the 95% confidence intervals in the slope ranging from -0.66 to -0.11 ppb/year.

The symbols shown next to each trend estimate relate to how statistically significant the trend estimate is, with p-values represented as: $p < 0.001$ = ∗∗∗, $p < 0.01$ = ∗∗, $p < 0.05$ = ∗ and $p < 0.1$ = +. The ** symbol in Figure 15 shows that the ozone trend for MBLS is significant with $p < 0.01$, meaning there is strong evidence that the concentrations are in fact decreasing over the years of 2010-2017 for MBLS.

The next figure below (Figure 17) shows TheilSen plots applied for all the other stations, using the same methodology as above.

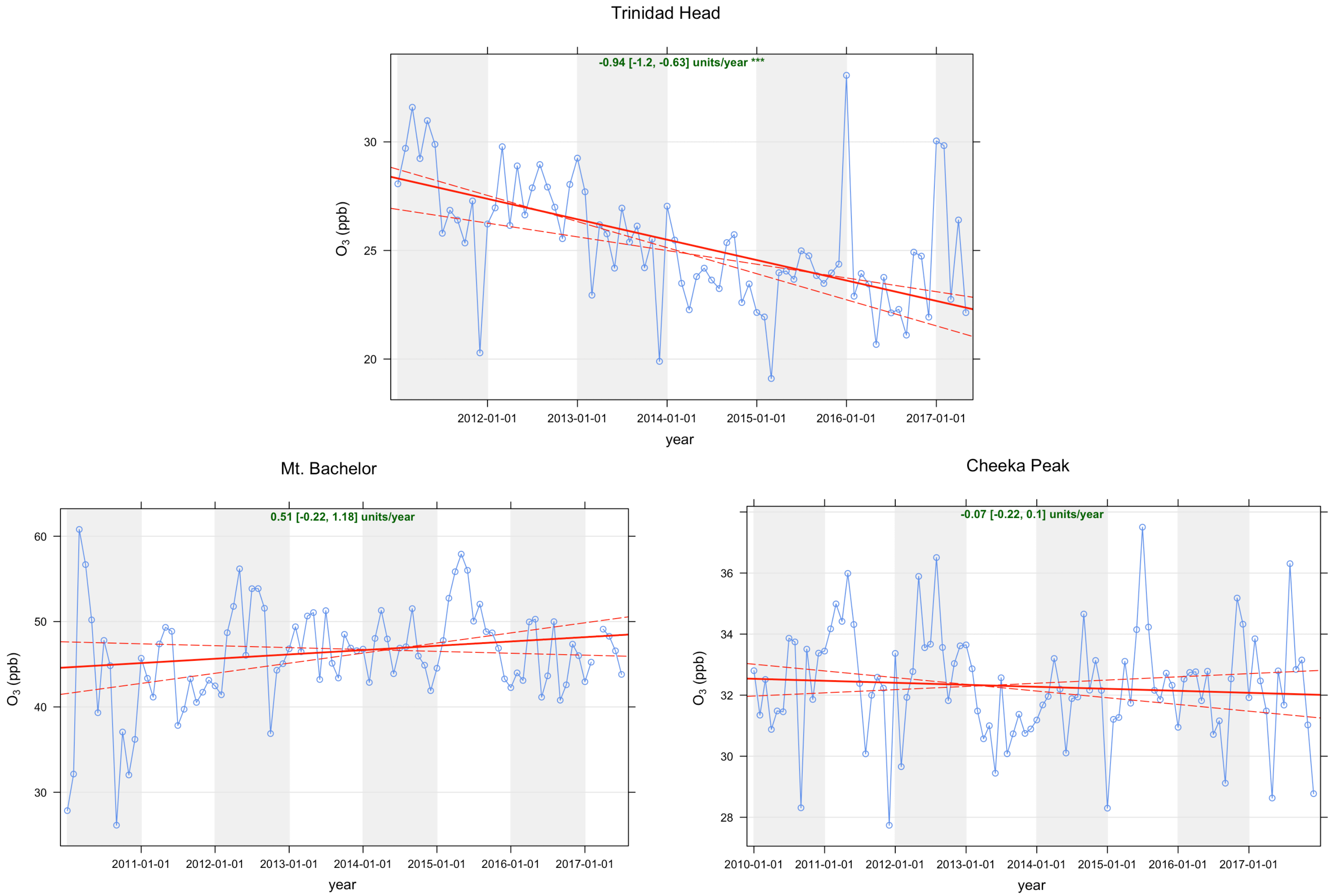


Figure 17: Trends in ozone at Trinidad Head, Mt Bachelor, and Cheeka Peak using TheilSen from `openair`. Plot shows deseasonalized monthly mean concentrations of ozone. Solid red line shows trend estimate and dashed red shows 95% confidence intervals for trend based on resampling method.

The plots show the deseasonalized monthly mean concentrations of ozone for the 4 different stations. The findings for these stations using TheilSen are as follows:

- Trinidad Head observes a significant decreasing trend of -0.94ppb/year with 95% confidence intervals ranging from -1.2 to -0.63 ppb/year. The *** shows that this trend finding is statistically significant with $p < 0.001$ and there is strong evidence for this decreasing trend for Trinidad Head.
- Mt. Bachelor has an upward trend of 0.51 ppb/year with 95% confidence intervals ranging from -0.22 to 1.18 ppb/year. The trend estimate is not statistically significant.
- Cheeka Peak has a relatively constant trend over time with -0.07 ppb/year with 95% confidence intervals ranging from -0.22 to 0.1 ppb/year. The trend is not statistically significant.

Previous research findings show that trends calculated from monitors located in western North America do not show statistically significant trends (either positive or negative) (Ainslie, 2018). These findings of trends levels align with what we observed using STL decomposition with the forecast package. With these findings using TheilSen and STL decomposition, the other stations do not pose similar trend patterns to MBLS in Ucluelet to become a proxy station for the station.

Whistler was not displayed in report due to incorrect station.

5 Summary + Conclusion

5.1 MBLS

Our findings show the Marine Boundary Layer Station in 2010-2017 to have ozone peaks in the spring and minimum levels in the summer. In time of day, ozone levels are lowest during the mornings (around 6am PDT) and peak in the late afternoons (around 3pm PDT).

We also found minimal correlation for ozone with the other pollutants including particulate matter, sulphur dioxide, carbon monoxide, carbon dioxide, nitrogen oxides and nitrogen dioxide. The best correlation found was between nitrogen oxides and nitrogen dioxide with a high negative correlation of $r = -0.85$, which is to be expected.

In the STL decomposition, we found that MBLS had an increasing annual trend until late 2013 where it started decreasing. From the TheilSen method, we found the overall ozone trend to be decreasing at -0.4 ppb/ year. This is interesting to note as we are seeing a background increase in ozone levels in the Canadian LFV. Findings from STL decomposition and Boosted Regression Trees are similar suggesting that the trends and seasonality we're observing are valid from the two different methods

The ozone and NO_2 levels are below CAAQS and the other objectives, which is expected given that they are background stations. We also observe the 8-hr metrics to be decreasing over the years which align with the decreasing trend value found for MBLS using TheilSen.

For the back-trajectories source regions, higher ozone levels are likely transported from Asian, Hawaiian, and Californian source regions. The source region that contributed the lowest levels of ozone in both 10m and 100m releases was Aleutian. Further analysis is recommended to explore the reasons behind these findings.

5.2 Comparison to other stations

The boxplot created to replicate Professor Ian McKendry's work (Figure 12), shows a good correlation between the boxplot results. There is a seasonal alignment for MBLS and Cheeka Peak.

In the STL decomposition we find that most stations' trends do not line up with MBLS, expect for Cheeka Peak, suggesting that this station in Washington, USA could potentially be used as a proxy background station for MBLS. TheilSen results show Trinidad Head, Whistler and Cheeka Peak have overall downward trends in their ozone levels, with the strongest downward trend from Trinidad Head. Mt. Bachelor has an upward trend in their ozone levels for the given years.

With these findings, the RGLOS steering committee will be able to strategize guidelines to broad policies, regulations or standards for ground-level ozone in the Canadian LFV.

Recommendations + Next Steps

The primary next step to support the ongoing RGLOS renewal is to confirm whether Cheeka Peak could be used as a proxy background station. Additional analysis is likely needed to confirm this initial result, including checks for statistical significance (including the WilCox text), comparisons to time series of other pollutants (e.g., CO, others), and testing Cheeka Peak with boosted regression. If it can be determined that Cheeka Peak could be used as a proxy background station, Metro Vancouver and its partners should explore the potential risks associated with that approach, given its reliance on other governments, and identify the key objectives for having a background station.

Additional next steps could include evaluation of the LFV stations (and perhaps the provincial Saturna station) using the same time series analysis tools to observe what ozone trends are visible for these stations. As mentioned earlier, stations at YVR and Tsawwassen are likely too close to local emissions to be used as background stations and so a suggestion is to restrict to only off-shore levels for these stations in the ongoing analysis. We also suggest exploring RH trigger point and boundary vs free troposphere to identify specific stations that can be used as background stations under specific conditions. Additional filtering of the MBLS data by wind speed (low speeds indicate stagnant air) could provide other results.

To carry out the report's analysis further, we suggest applying the boosted regression tree method to the other background stations to see whether the trends are the same as the ones found using STL decomposition. MBLS can be further modelled using boosted regression tree for 2010-2017 by integrating the other meteorological parameters, including the back-trajectories and NO, to determine any influences of these parameters to the observed ozone concentrations. We also suggest exploring the implementation of the KZ method to compare to the other statistical methods.

We also want to further understand why we are observing these trends in our region and what could be driving them. One of the suggestions during the data analysis was to look into the wildfires occurring in the Pacific Northwest during those increasing years and see whether they contributed to the increase in background ozone levels.

Lastly, using machine learning and other advanced data science techniques could serve good use to Metro Vancouver and its air quality work. For example, the forecast package allows use of algorithms to predict and forecast future trends using data. This can help prepare policies for expected levels of background ozone or other pollutant levels and bring greater insight from data.

References

Ainslie, B., et. Al (2018). The spatiotemporal response of summertime tropospheric ozone to changes in local precursor emissions in the lower fraser valley, British Columbia. *Atmospheric- Ocean Journal. 303-321.* Doi: https://doi-org.ezproxy.library.ubc.ca/10.1080/07055900.2018.1517721

Akhtar, F., Whitesides, B., Yan, B. (2003, November). Back trajectory techniques in air pollution. *Georgia Institute of Technology of Earth and Atmospheric Sciences.* Retrieved from: http://apollo.eas.gatech.edu/EAS6792/2003/presentations/back_trajectory.ppt

Carslaw, D. (2015, Jan). *The openair manual. Open-source tools for analysis air pollution data.* King's College London. Available from http://www.openair-project.org/PDF/OpenAir_Manual.pdf

Carslaw, D. (2017). Deweather: an R package to remove meteorological variation from air quality data. Available from: https://github.com/davidcarslaw/deweather

Chen, W. (2019). *State of the science on Ozone. Updating the Regional Ground-Level Ozone Strategy.* Available from https://sustain.ubc.ca/sites/default/files/2019-36_State%20of%20Science%20on%20Ozone_Chen.pdf

Glen, S. (2013, Oct) "Lowess Smoothing in Statistics: What is it?" From StatisticsHowTo.com: Elementary Statistics for the rest of us! Available from: https://www.statisticshowto.com/lowess-smoothing

Holtz, Y. (2018). Violin Chart. *The R graph Gallery.* Available from: https://www.r-graph-gallery.com/violin.html

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on June 2020.

McKendry, I., et al. (2014, May). *Low Ozone Episodes at Amphitrite Point Marine Boundary Layer Observatory*, British Columbia, Canada, Atmosphere-Ocean, 52:3, 271-280, DOI: 10.1080/07055900.2014.910164

Metro Vancouver (2020, January). *Metro Vancouver ambient air quality objectives.* Air Quality Publications. Available from http://www.metrovancouver.org/services/air-quality/AirQualityPublications/CurrentAmbientAirQualityObjectives.pdf

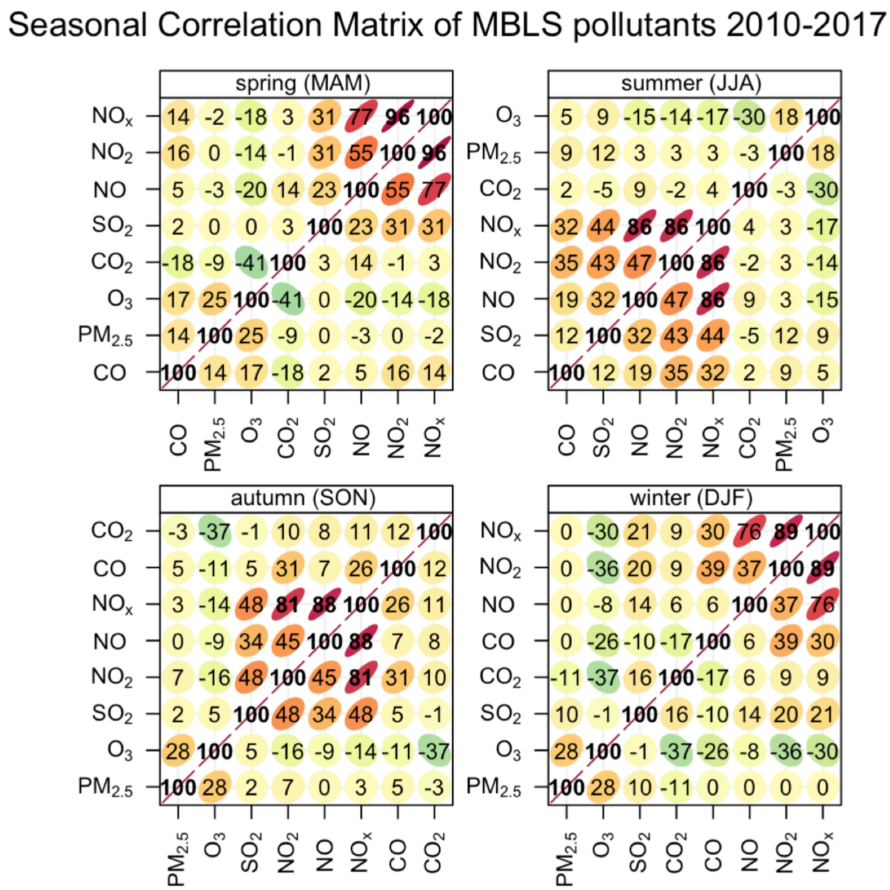
Metro Vancouver. (2019, May). Proposed Changes to Metro Vancouver's Ambient Air Quality Objectives for Nitrogen Dioxide, Ground-Level Ozone and Carbon Monoxide. Retrieved from: http://www.metrovancouver.org/services/airquality/AirQualityPublications/MV_Consultation_PaperProposedChangesMV_Ambient_Air_Quality_Objectives_May2019.pdf

Metro Vancouver, et al. (2014, April). *Regional Ground-Level Ozone Strategy. For the Canadian Lower Fraser Valley Region*. Available from http://www.metrovancouver.org/services/air-quality/AirQualityPublications/RGLOS2014.pdf

Schiller, C., Wingarzan R., et al (2011, December). Trace gas and particulate observations at the Marine Boundary Layer Station near Ucluelet on Vancouver Island. *Environment Canada.* Available from: https://www.researchgate.net/publication/258457444_Trace_Gas_and_Particulate_Observations_at_the_Marine_Boundary_Layer_Site_near_Ucluelet_on_Vancouver_Island

Wise, E. (2005). Extending the Kolmogorov-Zurbenko Filter: Application to ozone particulate matter, and meteorological trends. *Journal of the Air & Waste Management Association. Pubmed.* DOI: 10.1080/10473289.2005.10464718

Appendix A

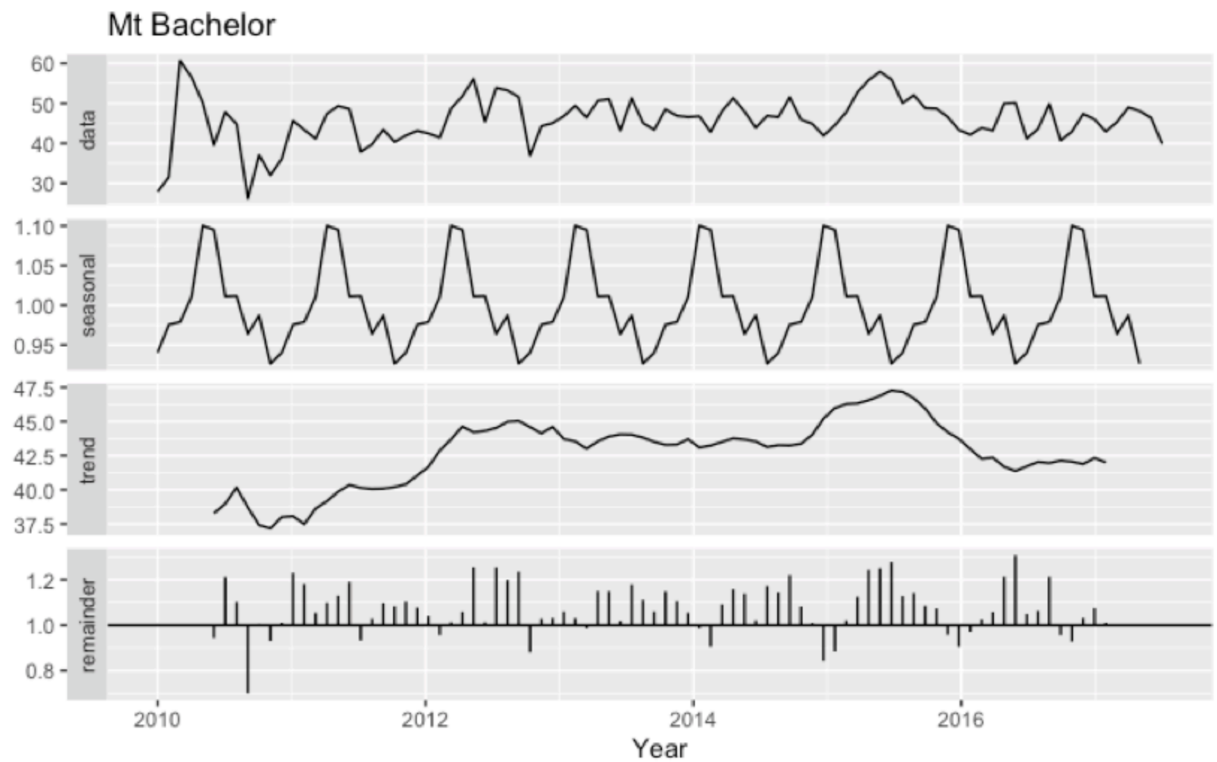


Seasonal correlation matrix using `corPlot` from `openair` package.

	year	O3_4th_raw	valid_year	valid_q2	valid_q3	valid_q2_q3	annual_avg	perc_q2	perc_q3
	<int>	<dbl>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	2010	45.6	NA	NO	YES	NO	NA	24.17582	88.04348
2	2011	52.1	NA	YES	YES	YES	NA	85.71429	95.65217
3	2012	54.0	NA	YES	YES	YES	NA	98.90110	91.30435
4	2013	53.9	NA	YES	YES	YES	NA	83.51648	91.30435
5	2014	53.8	NA	YES	YES	YES	NA	82.41758	69.56522
6	2015	51.9	NA	YES	YES	YES	NA	100.00000	77.17391
7	2016	50.6	NA	YES	YES	YES	NA	100.00000	93.47826
8	2017	48.2	NA	YES	NO	NO	NA	100.00000	18.47826

8 rows | 1-10 of 10 columns

CAAQS data completeness table for Marine Boundary Layer Station for ozone. Data table shows the O_3 4^{th} highest raw values for every year using 8 hour rolling average. Years with valid_q2_q3 having NO is excluded in calculation as there is not enough data in the year to satisfy data completeness criteria. For the 3 year average, if 2/3 of the years had all valid q2 and q3, then it could be used. For example, for 2017, we can use 2015 and 2016 values and average it to obtain the objective value for 2017 CAAQS.



Multiplicative decomposition using classical decomposition plot for Mt Bachelor. STL does not handle multiplicative decompositions well. Classical decomposition uses moving averages instead of loess used in STL.