

7-19-2021

A Ground-Based Assessment Framework for Validating Diesel Particulate Emission Models and Applicability in Portland, OR

Kirsten Marie Sarle
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Environmental Sciences Commons](#), and the [Environmental Studies Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Sarle, Kirsten Marie, "A Ground-Based Assessment Framework for Validating Diesel Particulate Emission Models and Applicability in Portland, OR" (2021). *Dissertations and Theses*. Paper 5733.
<https://doi.org/10.15760/etd.7604>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

A Ground-Based Assessment Framework for Validating Diesel Particulate
Emission Models and Applicability in Portland, OR.

by

Kirsten Marie Sarle

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
In
Environmental Science and Management

Thesis Committee:
Linda George, Chair
Kelly Gleason
Meenakshi Rao

Portland State University
2021

Abstract

Exposure to diesel emissions causes a range of health effects throughout the body, impairing; respiratory, cardiovascular, central nervous, renal, and cognitive systems. Diesel particulate matter (DPM) in Portland, Oregon is prevalent due to the layout of highly trafficked roadways, rail lines, and marine ports exposing a dense population to high levels of exhaust pollution. These high concentrations of ambient diesel emissions disproportionately impact minority and low-income populations.

Ground-based monitoring and modeling are two ways to assess ambient DPM. However, there are uncertainties in modeled DPM due to knowledge gaps in emissions inventories as well as lack of model validation against ground-based measurements. We propose a framework for efficient assessment of localized diesel emission sources, and model validation.

Sources of diesel identified as having the largest uncertainty in previous modeling studies were assessed for activity data and emissions were sampled for each main source type. We monitored for a range of traffic related air pollutants such as Black carbon and Nitrogen Oxides in two communities. These measurements will enable us to assess dispersion models, and better characterize DPM sources that are impacting the health of these communities. Fuzzy cluster analysis's applicability in air quality is shown through several studies but not yet for diesel identification. Fuzzy Cluster analysis was investigated as a potential tool for simplified source characterization. We demonstrate its practical use and discuss the opportunities and challenges of interpreting fuzzy clustering output.

In summary we present a suite of tools, accessible to most municipalities in the US, that can be used to fill in knowledge gaps or validate models to help communities to better understand and plan to mitigate their health risk from exposure to DPM.

Acknowledgments

Starting a graduate degree was not something I would have considered without the encouragement of my advisor Linda George. She has guided and supported me throughout this time of growth.

I would like to thank Meenakshi Rao and Kelly Gleason for taking on the extra work of being a part of my thesis committee. They have been patient, provided valuable advice, and have been flexible during tough times.

Thank you to everyone in our diesel collaboration group. Especially, Julianne Fry for continual support and enthusiastic collaboration. As well as all the students that have put in hard work on our project under Julianne Frye with Reed College. It was a pleasure working with everyone at Oregon DEQ who played a major role in collaboration, and community monitoring.

Thank you to my peers in the Sustainable Atmospheres Research Lab: Philip Orlando, Modi Raduma, Lyndsey Boyle, T'sai Cookey, and Andrew Rodgers, for your assistance and comradery. As well as Research for Undergraduate students that helped during the summers of 2018 and 2019 Aneesha Gharpurey and Ashley Kahl-Placek.

Thank you to all my dearest friends that have been a network of strength and growth. Lastly, to my family, for cultivating curiosity in the environment, and the belief in myself to follow my passions.

Table of Contents

A Ground-Based Assessment Framework for Validating Diesel Particulate Emission Models and Applicability in Portland, OR.	
Abstract	i
List of Figures	vi
List of Tables and Equations.....	viii
Chapter 1. Introduction.....	1
Chapter 2. Model Source Inputs Improvement	6
2.1 Model improvements Introduction.....	6
2.2 Rail Pilot Study	8
2.3 Rail Mobile Monitoring	10
2.4 Marine	13
2.5 Distribution Centers	15
2.6 Construction	17
2.7 Summary of source sampling.....	19
Chapter 3. Community Monitoring	20
3.1 Community Monitoring Introduction.....	20
3.2 Cully Monitoring.....	22
3.2.1 Cully Sampling Methods	22
3.2.2 Cully Results.....	23
3.2.3 PAH Analysis at Cully	29
3.3 Jade Monitoring.....	32
3.3.1 Sampling Methods.....	32
3.3.2 Jade Monitoring Results	33
3.4 SE Lafayette Comparison.....	39
Chapter 4. Model validations.....	42
Chapter 5. Fuzzy C-means Cluster Analysis	45
5.1 Cluster Rationale	45
5.2 Cully Clustering Results.....	47
5.3 Cluster Results for Jade.....	54
5.6 Summary of Fuzzy Clustering.....	60
Chapter 6. Conclusion	61
6.1 Summary of Project.....	61
6.2 Use as a Framework	64

References	66
Appendix A. Instrumentation.....	70
Portable Measurement System (PMS)	70
Aethalometer BC Source Apportionment	71
Appendix B. Fuzzy C-Means Methods.....	73
Cluster validation methods.....	73
Cluster Analysis Workflow.....	73

List of Figures

Figure 1-1. PATS dispersion output for DPM aggregated to census tracts (left). High spatial resolution NO ₂ map from predictive model and ground-based measurements (right).	5
Figure 2-1. Example plume capture for calculation for a BC emission factor using equation 1.....	7
Figure 2-2. Average BC emission factor by rail company, and total average emission factor = 0.9 +/- 0.23 g BC/ kg diesel total n = 23. UP = 13, Amtrak n= 8, BNSF = 2	12
Figure 2-3. Number of vessel calls to the Port of Portland annually from 2010 to 2020.	14
Figure 3-1. Vulnerability map for Portland OR, with scaled diesel vulnerability score at the census tract level.	21
Figure 3-2. Windrose of Cully Monitoring during diesel pollution sampling.....	25
Figure 3-3. Polar Plots of main pollutants at the Cully monitoring site.	26
Figure 3-4. fmass over air quality permitted facilities (squares) around the Cully monitoring site.	27
Figure 3-5. fmass and bbmass at Cully compared by hour and day of the week (top), hour (bottom left), month (bottom middle), and day of week (bottom right).....	28
Figure 3-6. fmass and bbmass compared by month and by hour averages.	29
Figure 3-7. PAH diagnostic ratio from DEQ PUF sampler.....	30
Figure 3-8. Comparison between p-PAH monitor and PUF 1-in-6 day sampler.....	31
Figure 3-9. Jade site wind rose for sampling duration.....	34
Figure 3-10. Polar plots of main pollutants at the Jade monitoring site.	35
Figure 3-11. fmass overlaying a map of air quality permitted facilities in Jade.	36
Figure 3-12. fmass and bbmass at Jade compared by hour and day of the week (top), hour (bottom left), month (bottom middle), and day of week (bottom right).....	37
Figure 3-13. Traffic volume from on ramp from Powell to I-205 from 2019 to the end of monitoring at Jade.....	38
Figure 3-14. SEL fmass and bbmass by hour (left), and by month (right).....	40
Figure 4-1. Comparison of measurements PATS and C-PORT modelled DPM.....	43
Figure 5-1. Sample partitioning among various groups (left), and best number of partitions using Calinski criterion (right) for Cully.	48

Figure 5-2. Cluster plot of Cully fuzzy clustering.	49
Figure 5-3. Silhouette width plot for 4 partitions for Cully	49
Figure 5-4. Polar plot of four cluster fuzzy memberships in Cully.	51
Figure 5-5. Cully's Time Variation plot of fuzzy cluster memberships equaling 1. Hour and day of week (top), average hour contribution (bottom-left), month (bottom-middle), day-of-week (bottom-right)	52
Figure 5-6. Cully's correlation plot based on Pearson correlation coefficient between fuzzy cluster memberships and model variables.	53
Figure 5-7. sample partitioning among various groups (left), and best number of partitions using Calinski criterion (right) for Jade.	54
Figure 5-8. Silhouette width plot for 4 partitions for Jade.	55
Figure 5-9. Cluster plot of Jade fuzzy clustering.	55
Figure 5-10. Polar plot of four cluster fuzzy memberships in Jade.	58
Figure 5-11. Jade's time Variation plot of fuzzy cluster memberships equaling 1. Hour and day of week (top), average hour contribution (bottom-left), month (bottom-middle), day-of-week (bottom-right)	58
Figure 5-12. Jade's correlation plot based on Pearson correlation coefficient between fuzzy cluster memberships and model variables.	59
Figure 6.2-1. Flow chart of proposed framework steps	64
Figure A-1. Magee Scientific AE33 internal calculation of fmass and bbmass.	72

List of Tables and Equations

Table 1. Instrumentation setup for Portable Pollution Monitoring.....	8
Table 2. Instrument used at the Brooklyn Railyard Pilot study	9
Table 3. Summary statistics for Cully site measurements	24
Table 4. Descriptive Statistics for Jade site monitoring	34
Table 5. Aethalometer averages for SE Lafayette Monitoring site.	40
Table 6. Comparison of ground-based measurements, PATS, and CMAS C-Port models.	43
Table 7. Instrumentation setup for Portable Pollution Monitoring.....	71
Equation 1.....	6

Chapter 1. Introduction

Our modern society is reliant on diesel engines for construction, goods movement, and transportation (Miller & Jin, 2019). Yet, exposure to emissions from these diesel engines has been shown to cause a spectrum of health issues (Lutes, 2007). Diesel exhaust, which is made up of a mixture of mostly nitrogen, water, carbon dioxide (CO₂), particulate matter (PM), gases, polycyclic aromatic hydrocarbons (PAHs), and metals (Wang, Wang, Bai, Wang, & Zhao, 2019), causes oxidative stress, tissue damage, and inflammation, leading to a range of effects throughout the body including respiratory, cardiovascular, central nervous, renal, and cognitive systems (Reis et al., 2018). Diesel emissions are also a climatic concern. Black carbon (BC), which is major particulate component of diesel exhaust, (Reşitoğlu, Altinişik, & Keskin, 2015) is also the second highest pollutant in global warming potential second to CO₂ (Bond et al., 2013).

Diesel emissions pose a wicked problem: diesel engines form the backbone of our transportation, and construction infrastructure, but their emissions our damaging human and environmental health. Yet, implementing a solution is not easy as there is little incentive for owners to bear the cost of switching to newer cleaner engines or retrofitted emissions control technology when the health and environmental benefits accrue largely to others.

Diesel emissions in Portland are concerning because Multnomah County was ranked in the top five percent in the nation for diesel exposure (US EPA, 2014, NATA).

The Portland Air Toxics Solutions (PATS) found minority and low-income communities in the Portland metropolitan region are disproportionately impacted by air toxics. One aim of this thesis is to serve Environmental Justice (EJ) communities of Portland with better knowledge of sources elevating their exposure. How does a community know what concentrations of DPM it might be exposed to? The two general methods of finding ambient concentrations of DPM are dispersion models and receptor monitoring. Receptor monitoring is the breakdown of pollution contribution to various sources from monitored data. Dispersion models use emissions inventories that include rate of operation, and rates of pollution from exhausts. These terms are activity rate (AR) for rate of pollutant and emissions factor (EF) for rate of pollution.

The PATS study used CALPUFF (Oregon Department of Environmental Quality, 2012), an atmospheric dispersion model with inputs of activity rates (AR) and emissions factors (EF), which is the amount of pollutant per amount of fuel burned. Inputs are spatially and temporally used in conjunction with meteorology and topography to create a map of pollution across a study area. The source emissions were estimated in PATS using data inputs that were mostly from national fleet averages, and fleet turnover (Oregon Department of Environmental Quality, 2012). These assumptions particularly for the Portland Metro area have been shown in recent statewide emission inventory update that Oregon and Portland specifically vary from what was expected in PATS (Eastern Research Group, 2020).

Originally, the model was performed in 2007 as the Portland Air Toxics Assessment (PATA), and it was performed again with updated information in 2011 to

PATS as a predictive model for the year of 2017 (Oregon Department of Environmental Quality, 2012). From the PATs report, DPM primarily comes from cars, trucks, marine, construction, and rail sources. The sources above were the focus of this study excluding cars since detailed information was already available for cars through emissions testing with the state Department of Motor Vehicles.

DPM is not a single pollutant, and thus there is no simple way of measuring it. DPM is made up of a soot-particle core, volatile organic compounds (VOC) and Polycyclic Aromatic Hydrocarbons (PAH), with gases absorbed onto the surface and absorbed into the particles (Downing, 2015). The components of diesel are not unique to diesel, and each pollutant is emitted by other combustion sources in various amounts and ratios. There is no direct way to measure DPM as predicted by models. There is a need to ground-truth results from dispersion models, to be able to confirm accurate input data, and find impactful sources.

DEQ, in partnership with Portland State University (PSU), Reed College, and Neighbors for Clean Air, received funding to improve the understanding of diesel emissions exposure in Portland OR, and how environmental justice communities are disproportionately impacted. To improve the PATS predictions of DPM, data was collected on emissions factors (EF) from each primary source present in the metropolitan area. Activity rate and site specific information is important in improving models for locally relevant exposure estimates (Isakov & Venkatram, 2006). Therefore, source activity assessments were performed to understand the spatial and temporal variability of how diesel emissions are being emitted in the Portland Metro region.

After source data inputs were assessed, community monitoring took place to ground-truth models in critical neighborhoods. Current ground-based methods for finding diesel by field sampling use filter-based methods and, apply multivariate statistical methods such as the EPA Positive Matrix Factorization (PMF) and Principal Components Analysis (PCA) (Hopke, 2016). The filter-based method of sampling takes in a minimum of 24-hour air flow onto a filter paper that is then analyzed in a laboratory setting for a matrix of analytes. EPA PMF typically needs at least a few hundred filter sample days per site to properly converge with reliable results (US-EPA, 2014). Due to these requirements, the cost of laboratory analysis, labor costs, and siting, add up to be expensive and time intensive.

Monitoring methods that take extensive time for collected potentially can potentially miss changes in diesel exposures. Diesel sources are transient in nature, such as construction sites operating in different areas, and moving locations when projects are complete. Construction diesel emissions are also largely associated with the phase of large soil movement called the earthworks phase, and later phases emit considerably less emissions (Heidari & Marr, 2015). Diesel emissions from on-road and non-road engines are dictated by regulation, through increasingly stringent performance-based emission standards. Emission standards need to be met at time of engine manufacturing and need to meet a tier set through the US EPA (Coglianese & Nash, 2017). Current methods for meeting standards are through Exhaust Gas Recirculation (EGR), Diesel Particulate Filters (DPF) and Selective Catalytic Reduction (SCR) (Reşitoğlu et al., 2015). As fleets turnover to newer tiered engines, pollutant contributions by source will change. Marine

and rail emissions are dictated through EPA emissions tiers, market demands for various goods, contractual agreements with suppliers, and through agreements between Ports, rail and, government agencies (Oregon Department of Transportation, 2014). The disparity between the changing nature of diesel, the time and cost of traditional ground-based monitoring show a need for a new method of DPM determination. Based on the above knowledge gaps, there is a need for low-cost and less time-intensive methods of measuring DPM, a method of inferring local diesel sources from measurements, and accessible information to communities of disproportionate impact from such measurements. The updated knowledge of pollution sources is needed to make direct and effective solutions. Lastly, the result of this study is a framework that can be used in future diesel emissions ground-based assessments.

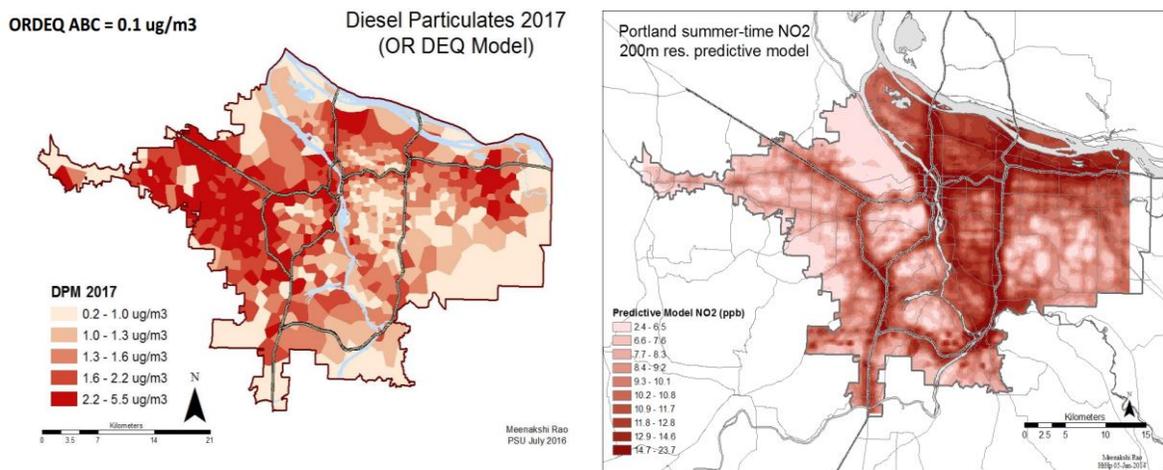


Figure 1-1. PATS dispersion output for DPM aggregated to census tracts (left). High spatial resolution NO₂ map from predictive model and ground-based measurements (right).

Chapter 2. Model Source Inputs Improvement

2.1 Model improvements Introduction

The two main types of data needed for source emission estimation are activity rate (AR) and emission factor (EF). AR is traffic volume, speed, and temporal data to average activity at a location. Emission factors are a mass of pollutant emitted per amount of fuel burned. For example, a common emission factor is g BC/ kg diesel fuel, which can also be in power usage as well, often seen in fleet standards regulations. The calculation adapted from Tang et. Al, 2015 of EF for BC is shown in Equation. 1.

Equation 1.

$$EF_{BC} = \frac{\int_{t_1}^{t_2} [BC(t) - BC(t_1)] dt}{\int_{t_1}^{t_2} [CO_2(t) - CO_2(t_1)] dt} * W_C$$

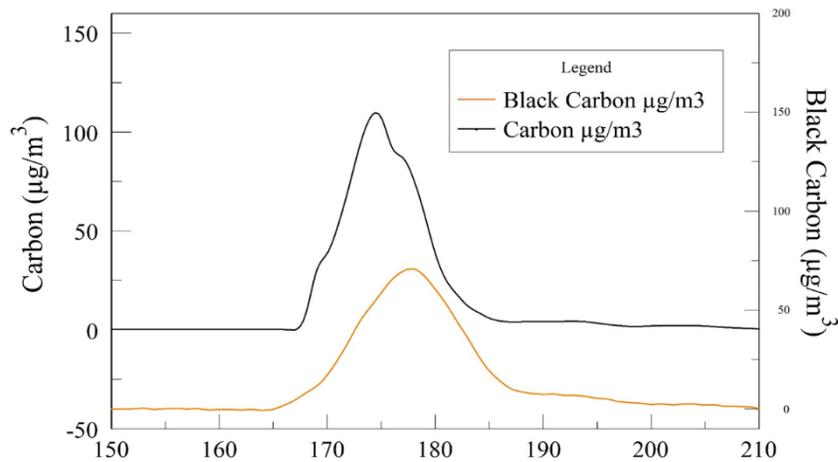


Figure 2-1. Example plume capture for calculation for a BC emission factor using equation 1.

Where the denominator is the integration of CO₂ from the starting and ending inflection points of the rise in emissions measured, minus the background. An example plume from this study is shown in Figure 2-1. The background for each plume is found by averaging five minutes before and after a plume. Then, divide the pollutant, in this case BC integrated from start to finish. Multiply the ratio by a carbon weight fraction for a specific fuel. Diesel’s carbon weight fraction is 0.87 (Johnson et al., 2013a). Johnson et al., also explain that CO₂ is converted to just carbon based on the assumption that all other gaseous carbon species are in trace amounts. This is done through a mass balance equation using the molecular weight ratio of carbon and oxygen.

Rail was the first source assessed. A pilot data collection for rail took place during the month of August 2018. After the pilot study, a portable measurement system (PMS)

was designed for mobile monitoring and capture of emissions further described in Appendix A. Full data collection for the main sources took place from March 2019-October 2019.

Table 1. Instrumentation setup for Portable Pollution Monitoring

<i>Instrument</i>	<i>Measurement</i>	<i>Units</i>	<i>Time base</i>
Magee Scientific AE33	Black carbon	ng/m³ or µg/m³	1 sec
Li-COR 850	CO₂	µ mol/mole (ppm)	1 sec
EcoChem PAS2000	p-PAH	ng/m³	6 sec
TSI DustTrak 8533	PM_{2.5} & PM₁₀	µg	1 sec
MicroAeth MA300	Black carbon & Lat & Long	ng/m³	1 sec

2.2 Rail Pilot Study

Rail emissions come from rail-line movement, and railyards. Railyards pose a different difficulty in modelling emissions due to the complexity of activity that happens onsite. Railyard emissions were investigated to capture plumes and characterize emissions from a Portland specific yard. a large comprehensive study of yard emissions and activities has not yet been performed in Portland. Methods similar to (Johnson et al., 2013b), were attempted at the Brooklyn Railyard. A monitoring trailer was set up at a garage door manufacturing facility with a shared fence to the Brooklyn railyard. Table 2 goes over the measurements made at the Brooklyn railyard. In addition, the instruments

listed in Table 2, a video camera watching the rail activity recorded for 24 hours a day during sampling for the month of August 2018.

Table 2. Instrument used at the Brooklyn Railyard Pilot study

<i>Instrument</i>	<i>Measurement</i>	<i>Unit</i>	<i>Time base</i>
Teledyne T200	NOx	Ppb	1 minute
Teledyne	SO2	Ppb	1 minute
thermo scientific Ozone Analyzer	Ozone	Ppb	1 sec
TSI NanoScanSMPS 3910	particle size distribution of 10nm to 420nm	Nm	1 minute
TSI Optical Particle Sizer 3330	particle size distribution of 0.3 to 10 µm	Nm	1 minute
Magee Scientific AE33	Black carbon	ng/m3	1 sec
EcoChem PAS2000	p-PAH	ng/m3	1 minute

Data was continuously monitored at the lowest time resolution obtainable, and a plume was detected if there was a rise in CO₂ of at least 25 ppm above background concentrations with a locomotive occurrence within a period of a minute or two via video footage.

Despite having setup with what seemed like an ideal sampling location and design, plumes that connected with locomotives passing were not able to be distinguished. A potential hypothesis for this is that the exhaust plumes were higher than

the sampling inlets. Another potential error could be that the truck loading station was nearby to the west of the rail transect, and with many trucks moving and idling nearby could have made the CO₂ plume indistinguishable from the rail passing by.

2.3 Rail Mobile Monitoring

Due to the rail pilot study instrumentation being unable to detect plumes next to the locomotive stacks, a new sampling method was designed. Based on trucking tunnel studies (Dallmann et al., 2013, Ban-Weiss et al., 2009) and other rail plume capture studies (Tang et al., 2015), being in a location above a passing locomotive exhaust was ideal. Pedestrian bridges that were far enough away from major traffic sources, and where locomotives passed underneath, representing typical inner-city rail speeds and load was key. Multiple locations were found that fit the criteria via Google maps and Geographic Information System (GIS) rail routes. Ease of access and transportation to each sampling location was also important. Sampling took place on the East Bank Esplanade on the waterfront in Downtown, but road traffic emissions interfered heavily with plume detection. Pier Park in North Portland was attempted but infrequent trains, difficult terrain, and distance made sampling here less than ideal. Thus, most samples took place at Brooklyn railyard crossing at the northern end of the yard and the Spring Water Corridor for ease of bike access, little to no traffic interference, and frequent train traffic under a variety of operating conditions.

It was found that plume captures were most successful under low wind conditions and thus in the early morning hours. Each sampling day began with zero calibrating the

DustTrak and p-Trak, allowing the aethalometer roughly five minutes to warm up, while the CO₂ and p-PAH data logging was set up. Then, the aethalometer and p-PAH monitors' connected sampling lines routed to a PM_{2.5} impactor were hung up on the side of bridge fence above where exhaust stacks would pass underneath. The CO₂ Teflon tubing was duct-taped to the outside of the impactor with the inlet as close to the particle inlet as possible to ensure the same air mass was sampled.

Once a train was close, the researcher would walk off the bridge far enough away as to mitigate personal exposure, take a photo or video with a timestamp, and track the speed using a radar gun. Time, train identification numbers, speed, rail owner, and other notes were recorded after the passing.

Emission factors were calculated for BC and some p-PAH based off Equation 1 using R programming (R Core team, 2020) for data pre-processing. P*S*i-plot software was used to find the area under the curve for both CO₂ and BC. Results of BC EF's by company type is shown in Figure 2-2. Amtrak had the highest emissions factor on average, but across all owners there are high standard deviations. A higher sample size is suggested to increase certainty of averages. Even though there was a wide range of EF's and difference between company, the total average emission factor was on par with Tang et al. (2015), of 0.87 +/- 0.66 g BC/ kg fuel burned.

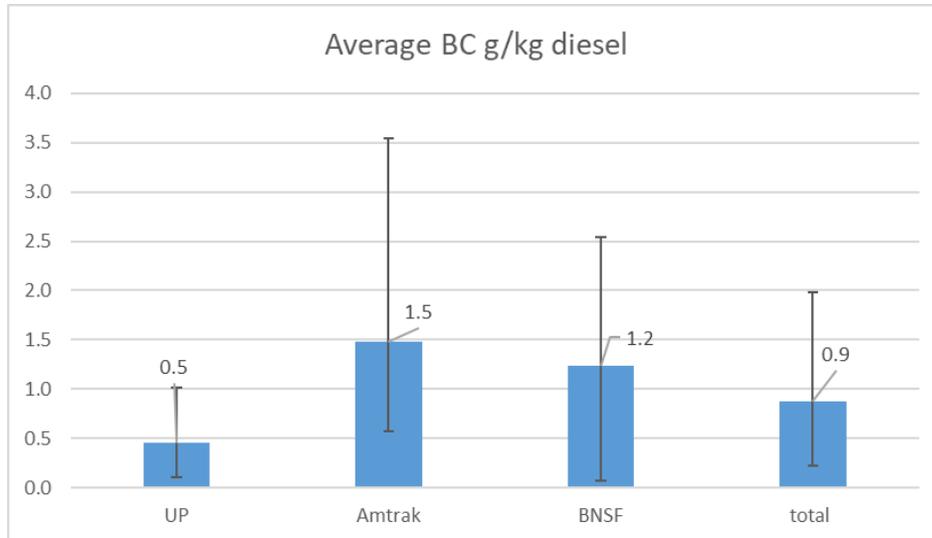


Figure 2-2. Average BC emission factor by rail company, and total average emission factor = 0.9 +/- 0.23 g BC/ kg diesel total n = 23. UP = 13, Amtrak n= 8, BNSF = 2

To check rail AR, a schedule or previous detailed rail crossings would be needed. Rail schedules are not released to the public due to the risk to national security. A way to find activity rates based off locally relevant data is from the state department of transportation. In this instance for safety and public concerns, Oregon Department of Transportation keeps a record of average annual daily train crossings. Based on rail rights and ownership, rail lines can be assigned an activity rate and emission factor, from how many crossings occur on average in a day.

Obtaining the emissions factors and comparing to the national average, along with rail line specific ARs are useful for future modelers. By checking predicted versus actual emission rates, it can be determined how to best perform future model runs.

2.4 Marine

Sampling for marine emissions was very similar to rail mobile monitoring, without the ability to be above the exhaust stack. Sampling took place at Kelly Point Park at the most northern tip of Portland, where the Columbia River and the Willamette converge. Due to the difference in stack height to inlet, and dilution by the time a plume reached the inlets, detecting marine emissions using the above method was not feasible. A BC monitor (Magee Scientific aethalometer, AE33) and a CO₂ (Li-COR 850) were setup at a DEQ site on Sauvie Island near the Columbia River. Similar to plume capture attempts on Kelly Point, plumes were too diluted for detection and Island traffic often interfered with measurements.

In addition to the above difficulties, we were informed by the Port of Portland that there was a large decrease in vessel traffic due to market forces and loss of contracts. This is confirmed through Port of Portland's report of vessel data years 2010 through 2020 shown in Figure 2-3. Since 2010 there has been a decline of roughly 180 vessels docking at Port of Portland per year, from 575 in 2010 to 394 per year in 2020. The reduced activity made it unfavorable for measuring emission factors, because of time restraints for measurements and low likelihood of plume capture. For this reason, activity rates were focused on for the remaining portion of marine vessel source characterization.

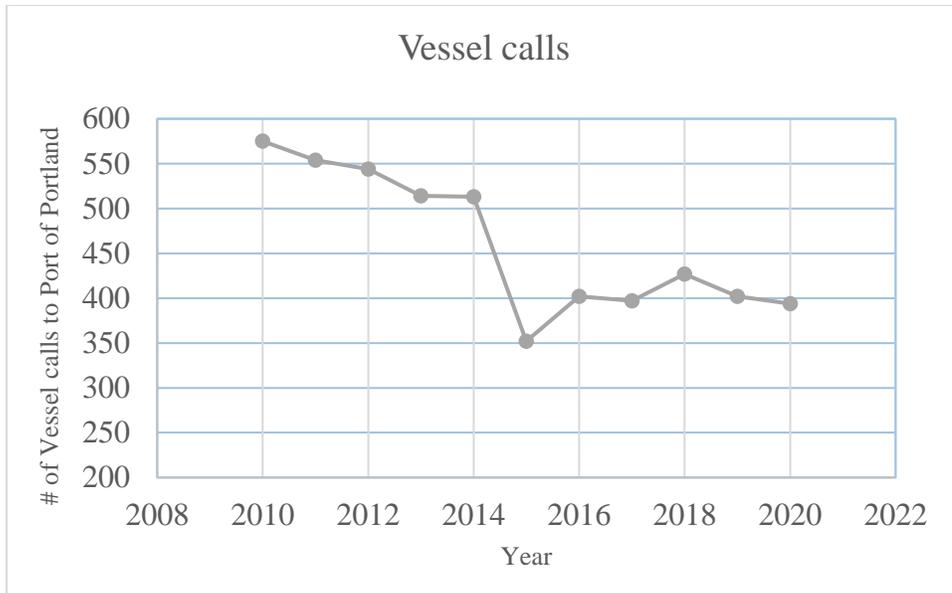


Figure 2-3. Number of vessel calls to the Port of Portland annually from 2010 to 2020.

Evidenced by Port of Portland’s decline in vessel traffic, marine activity can vary greatly due to market forces and port dynamics. Emissions estimates could be quickly outdated if based on long term port activity assumptions. A way to ground truth what is in a model for activity assessment is to compare input data to vessel Automatic Identification System (AIS) data. AIS is sent from every ship in the world that is over 300 gross tonnage, with location, time, speed, type, direction, length, beam, and draft (Brown & Aldridge, 2019). The data can be as detailed as every minute. For dispersion models this level of data is likely not needed. However, factors such as speed and type and engine usage could be useful in correcting assumptions made in models previously (Brown & Albridge, 2019). Data is free and publicly available for the United States at Marinecadastre.gov.

2.5 Distribution Centers

Trucking included in PATS as On-Road mobile was identified as an area that needed assessment. When looking at the evidence for trucking, normal operations were well known. Emission factors are well studied, and activity is tracked by ODOT and used currently by DEQ. An area within the trucking industry that is less investigated is the impact of distribution centers on local diesel emissions exposure by attracting more diesel truck activity.

To investigate the impact that distribution centers have above normal traffic and trucks, sampling of NO₂ for fresh emissions was performed around three distribution centers of varying sizes. Ogawa[®] passive NO₂ samplers were set out for at least 10-12 days during summer months and analyzed in the Sustainable Atmospheres Research (STAR) Lab by undergraduate students Ashley Placek and Tiffany Mosher. NO₂ is makes up a large portion of diesel emissions and is considered the second-best proxy for diesel emissions (___). Results showed that surrounding distribution centers even near highways had spots of higher NO₂ concentrations compared to residential neighborhoods. In Figure 2, there were high levels of NO₂ on the opposite side of a distribution center, and lower levels of NO₂ on the freeway side of the Albertsons distribution center (in blue). Distribution centers thus are a source that could use further investigation as a potential source for the additional activity.

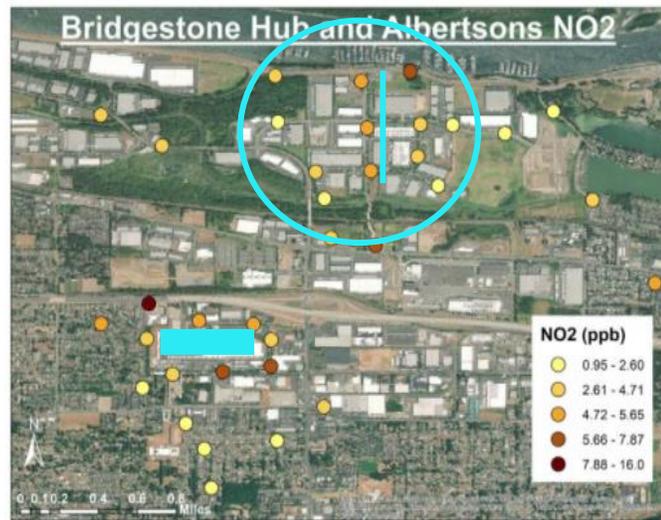


Figure 2-4. NO₂ passive sampler results surrounding distribution centers. WinCo distribution in Woodburn, OR. Albertsons (bottom right), and Bridgestone (top right) in Portland, OR. Blue squares represent location of distribution centers. Figure 1a shows a trucking entrance on the right blue square. Figure 1b, shows a distribution center hub, identified with a circle and main traffic corridor in a line and Albertsons distribution center in the bottom square.

2.6 Construction

Construction EFs are difficult to sample because of site liability and access. Therefore, a similar sampling design to rail mobile monitoring was created to assess relative contribution from construction. The bike stroller was used just as a stroller for walking around sites and inside contained the sampling instruments. On the stroller handle was the inlet and the cover of the stroller was placed over the front for ambiguity. Walking transects took place in blocks looping directly around the perimeter of the construction site and then an outer loop. Each sampling day these loops were repeated from before construction started at 7am until roughly noon when winds picked up and caused too much variability to reliably catch construction emissions. Sampling on weekends was also performed in the mornings as well to show the difference between weekend versus weekday rush hour impact.

The results are shown in Figure 2-. Figure 2-a., shows weekend sampling. Figure 2-b, shows before construction starts, and Figure 2-c, shows while active construction is happening at a single construction site in downtown operating diesel equipment. It is seen in Figure 2-c, that emissions are higher directly surrounding the site. This is not due to general rush hour traffic or road traffic, because the surrounding area was consistently at least $2 \mu\text{g}/\text{m}^3$ lower than the construction loop.

Construction emissions were not able to be quantified with EFs in this study, however, it was shown that construction adds to the pollution load in a city above normal condition. When a site is active in highly polluting stages of construction such as the

earthworks phase, then this could cause hotspots of pollution within a city. As well as a city that is going through growth, with many construction sites, could cause a regional background concentration to increase. In the updated emissions inventory of Oregon non-road diesel, construction made up 18.2 percent of total diesel emissions for the state, with the there being a larger proportion of construction in urban areas versus rural. This means that Portland’s construction emissions are hard to spatially predict but may cause a region-wide risk to increased diesel particles.

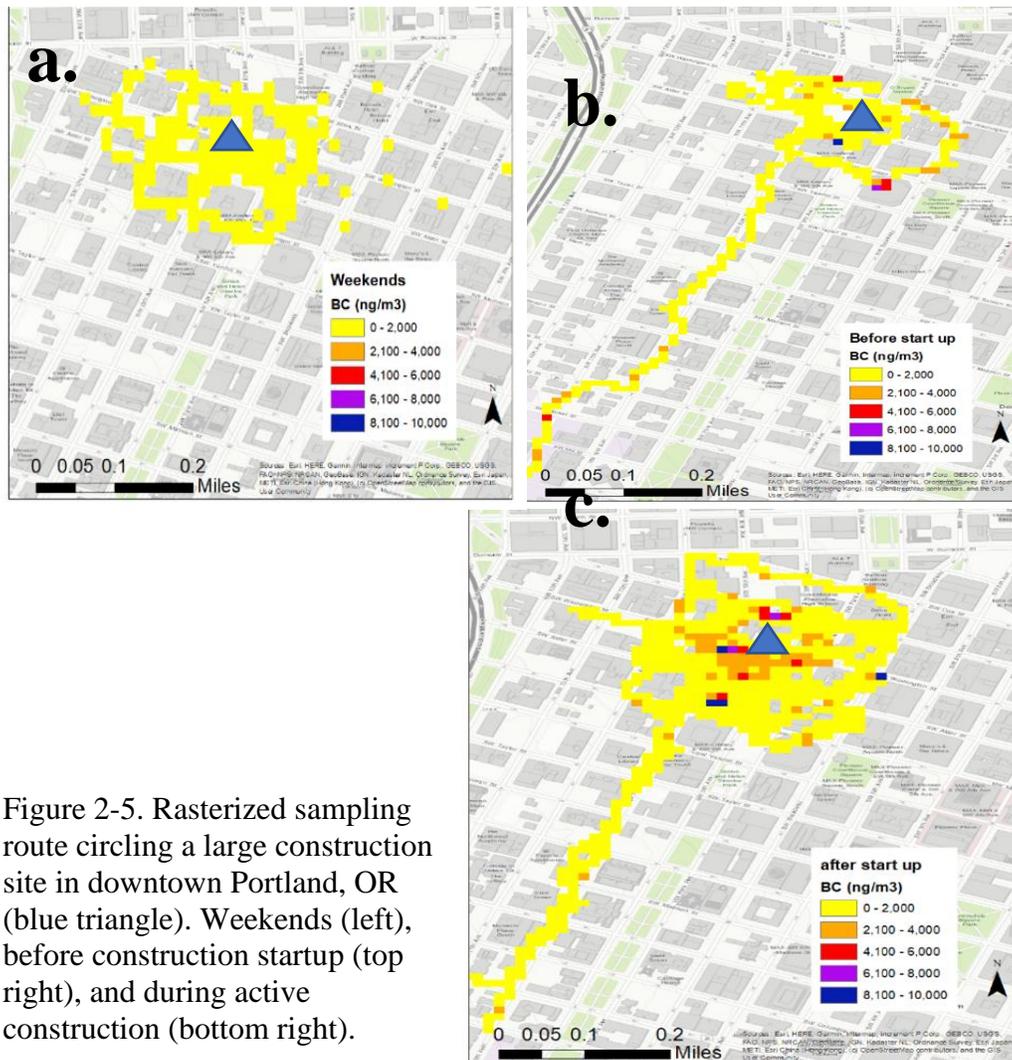


Figure 2-5. Rasterized sampling route circling a large construction site in downtown Portland, OR (blue triangle). Weekends (left), before construction startup (top right), and during active construction (bottom right).

2.7 Summary of source sampling

All the primary sources sampled for were assessed with locally relevant emissions measurements and data. Rail emissions estimates are improved through collection of a sample of EFs of locomotives in operation through Portland. ODOT rail crossing data was collected and used to create an up-to-date activity assessment, that can be compared to inputs for dispersion models. Marine emissions were not successfully sampled. However, marine AIS data is a detailed up-to-date data source for AR. The AIS data can be used to develop a better future understanding of marine emissions through comparison between previous model data inputs. Emission proxies can be created using ship type, and engine usage within the modelling domain.

Each activity rate data source obtained was free, publicly available, and available in most other cities outside of the Portland region. For these reasons, the above methods can be implemented as part of a framework for assessing dispersion model accuracy in other regions that would like to assess dispersion models, or diesel exposure.

Source sampling investigations provide a better understanding of activity rates, what certain fleet makeups are, and spatial distribution of source activity. This supplies knowledge to the upcoming community sampling analysis, in which major sources can be compared with monitored DPM proxies.

Chapter 3. Community Monitoring

3.1 Community Monitoring Introduction

Community monitoring took place to find sources of diesel pollution that were most impacting the health of EJ communities in Portland. It was decided that two neighborhoods would be monitored over different seasons to capture a range of seasonal variability. It would be reasonable to assess at maximum two neighborhoods based on resources. Each neighborhood was monitored for 4-6 months using a similar suite of instruments and set up as the rail pilot study. The neighborhoods were chosen using a scaled vulnerability and exposure map produced by Vivek Shandas and Yasuyo Makido of the PSU Sustainable Urban Places Research (SUPR) lab. The exposure assessment was based off a high spatial resolution study of NO₂ in Portland using ground-based sampling and LUR modelling (Rao, et al., 2014), and the PATS estimate of DPM. The exposures were scaled with socioeconomic metrics that have been linked to individuals more vulnerable to exposure, including poverty, non-white races, elderly living alone, percent under 18 years of age, and population density. The resulting two neighborhoods that had the highest vulnerability score were the Cully and Jade districts. Both neighborhoods have a high proportion of people of color and low-income housing. Jade particularly is home to a large proportion of Pacific-Islander and Asian heritage.

Diesel : NO2 : Socioeconomic : Population = 1 : 1 : 1 : 1

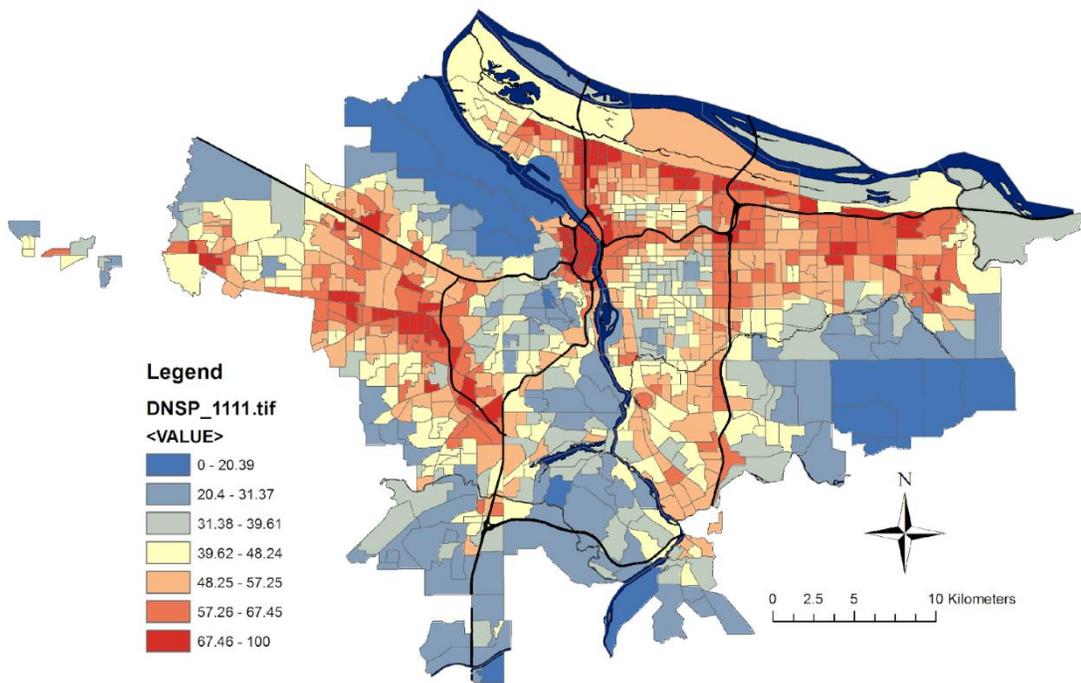


Figure 3-1. Vulnerability map for Portland OR, with scaled diesel vulnerability score at the census tract level.

The Cully site was monitored first since DEQ already had an air toxics monitoring site set up in Cully. Adding to the suite of sampling would provide a robustness and attention to analysis. The Jade site was contracted with the Bureau of Environmental Services at the Brooklyn Pump station after Cully monitoring was complete.

3.2 Cully Monitoring

The Cully monitoring site is located at Helensview School (45.562234, -122.575693) in North Portland about 780 meters from the interstate 205 ramp and 1.3 miles or 2 km Southeast of the airport. It is located on the south side of Killingworth Ave. and Columbia Blvd., both of which are heavily trafficked during rush hours, and by HDD traffic. Parallel to Columbia Blvd. are train tracks that carry mostly Union Pacific locomotives. North of the airport is the Columbia River with the marine diesel activity coming from the port of Vancouver and pleasure craft and local tugboats. The site is located on a relatively flat and open field next to a community garden.

3.2.1 Cully Sampling Methods

In June of 2018, the Oregon DEQ set up a long-term monitoring station in the Cully neighborhood at Helensview School. The site consisted of a trailer containing a BC monitor and nephelometer for PM mass. Toxics were sampled every six days using high and low volume samplers. Metals, PAH's, carbonyls, and VOCs were analyzed at the DEQ's Hillsboro chemistry department. Previous results and findings are published periodically in the station report (2018 Oregon Air Toxics Monitoring Summary, 2018). In mid-October of 2019, the PSU STAR lab added gas analyzers; NO_x/NO₂ (Teledyne T200), Ozone (Dasibi 1008), particle analyzers; p-PAH (EcoChem PAS2000), and Optical Particle sizer (OPS) (TSI 3330). Before the equipment was set up, each instrument was calibrated accordingly. The Nanoscan Mobility Particle Sizer (TSI 3910) was added to this site, but due to pump failure, it was removed early in the campaign.

Each instrument was setup for one minute sampling, except the ozone monitor was set up for 30 second sampling via analogue to digital converter. The OPS was set up for scheduled monitoring for the maximum memory it could hold of 10 days. Within the 10-day window the instruments were checked, data was downloaded, and the OPS data was cleared and setup for monitoring again. Failure to start sampling occurred on multiple occasions, causing blocks of data for the OPS to be missing. Continuous sampling took place from June 2018 to March 2020 when PSU STAR lab equipment was removed. DEQ is continuing monitoring of toxics, including BC and PM measurements.

3.2.2 Cully Results

General descriptive statistics including mean, range, and median are shown for Cully at one hour averaging intervals in Table 3. Source apportionment of BC to proportion of fossil fuel associated BC (ffmass) and biomass wood burning BC (bbmass) masses is explained in Appendix A. The BC and specifically the ffbmass are 7 and 5 times higher respectively than the Oregon ambient benchmark concentration (ABC) for DPM. NO_x and NO₂ are still within attainment based on the National Ambient Air Quality Standards, but the maximum seen for all pollutants agree that there are events of high concentrations that could cause acute exposure health risks (Functions et al., 1998). Such as the maximum of ffbmass and BC being 4.7 µg/m³, which is 47 times the cancer benchmark risk. Most of the wind during this time came from the East and South Directions as seen in Figure 3-2, from the direction of I-205 and the surrounding neighborhood, implying that on-road traffic is a major source for this site.

Table 3. Summary statistics for Cully site measurements

	<i>WD</i>	<i>WS</i>	<i>BC</i>	<i>ffmass</i>	<i>bbmass</i>	<i>NO2</i>	<i>NOx</i>	<i>O3</i>
<i>Units</i>	<i>DEGREES</i>	<i>MPH</i>	$\mu\text{G}/\text{M}^3$	$\mu\text{G}/\text{M}^3$	$\mu\text{G}/\text{M}^3$	<i>PPB</i>	<i>PPB</i>	<i>PPB</i>
<i>Min</i>	<i>0.5</i>	<i>0.6</i>	<i>0.0</i>	<i>0.0</i>	<i>0.0</i>	<i>0.4</i>	<i>0.9</i>	<i>0.1</i>
<i>Median</i>	<i>137.8</i>	<i>5.0</i>	<i>0.4</i>	<i>0.3</i>	<i>0.1</i>	<i>10.0</i>	<i>13.1</i>	<i>16.7</i>
<i>Mean</i>	<i>161.3</i>	<i>5.7</i>	<i>0.7</i>	<i>0.5</i>	<i>0.3</i>	<i>10.7</i>	<i>18.0</i>	<i>16.9</i>
<i>Max</i>	<i>359.4</i>	<i>19.2</i>	<i>6.7</i>	<i>4.7</i>	<i>2.5</i>	<i>41.7</i>	<i>153.6</i>	<i>43.4</i>

Despite the wind mostly coming from the East and Southeast, when the wind does come from the North and West there are higher concentrations of pollutants seen in Figure 3-3. When the winds shift in the summer months pollutants would be blowing high concentrations into the surrounding neighborhood more often. The ffmass in Figure 3-3 lines up in better agreement with NO_x and NO_2 Polar Plots than the bbmass in Figure 3-3. The agreement of diesel proxy measurements over bbmass shows that there is traffic activity evidenced in Cully.

Fossil fuel mass polar plot can be compared with the Oregon DEQ Facility Profiler Lite map (ODEQ, 2020) of air quality permits. By looking at all the sources together compared to where high concentration loads are coming from,

source characterization is more obtainable. In Figure 3-4 the, the ffmass polar plot from Figure 3-3 is laid over the map of permitted facilities. It is seen that there is some directional correlation between northern air masses and the freeway. There are less distinct hot spots of ffmass concentration as seen in NO_x and NO_2 in Figure 3-3. There are many permitted facilities in the area for NO_x , and these hot spots in the polar plot suggest point sources. In Figure 3-4 there is a lack of hotspots, and this suggests a non-point source or more specifically a mobile source that is spatially dispersed.

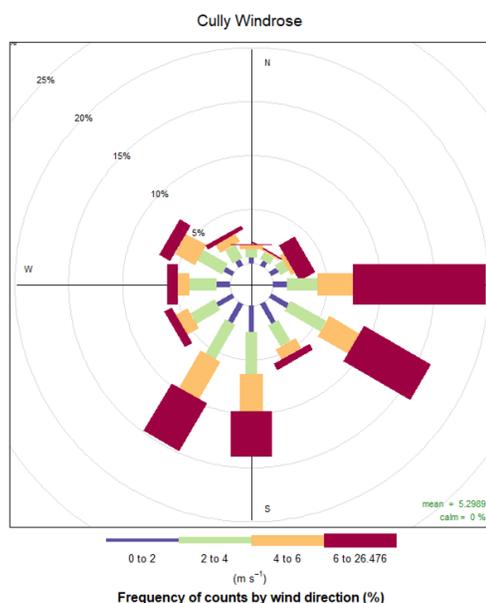


Figure 3-2. Windrose of Cully Monitoring during diesel pollution sampling.

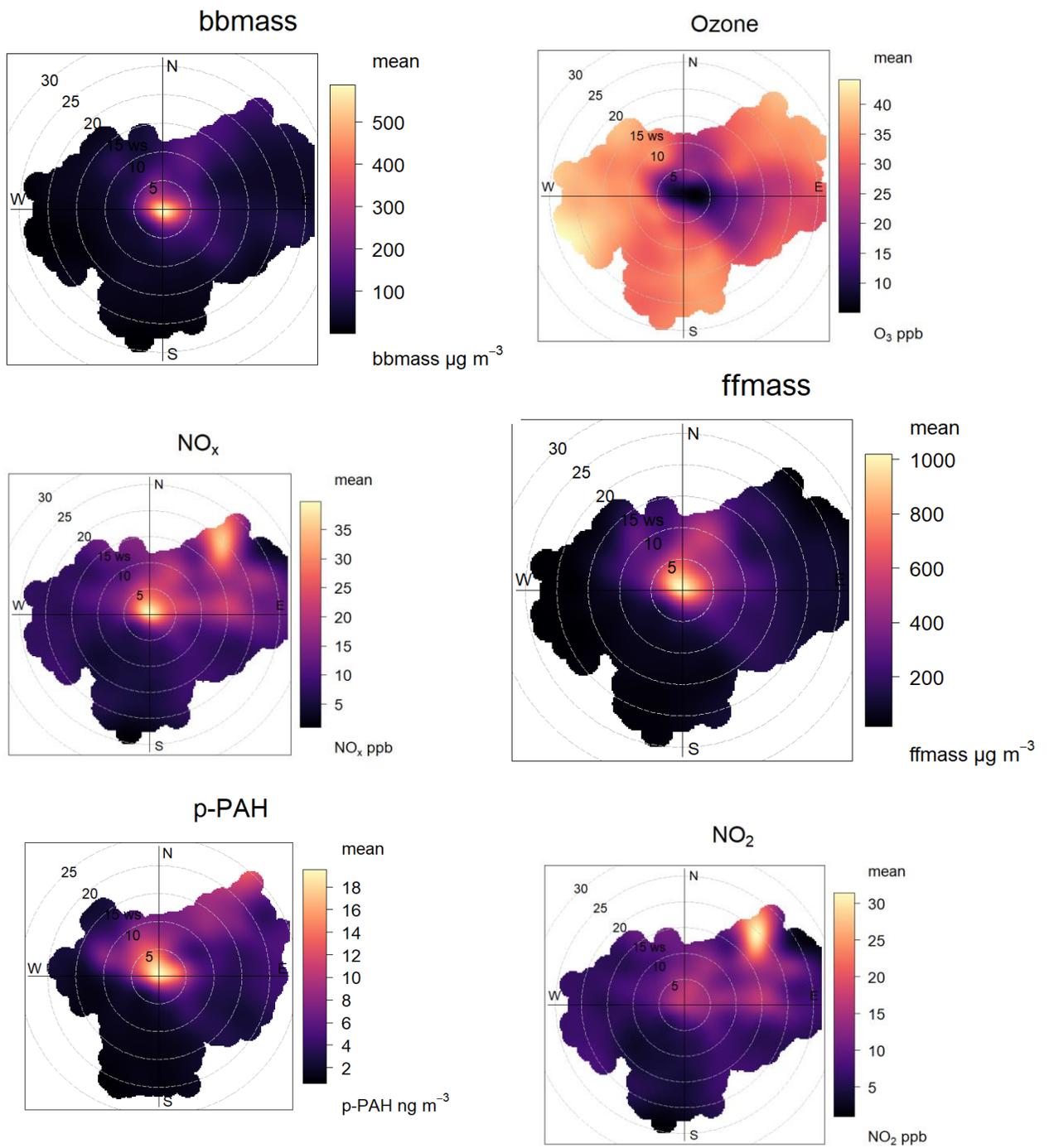


Figure 3-3. Polar Plots of main pollutants at the Cully monitoring site.

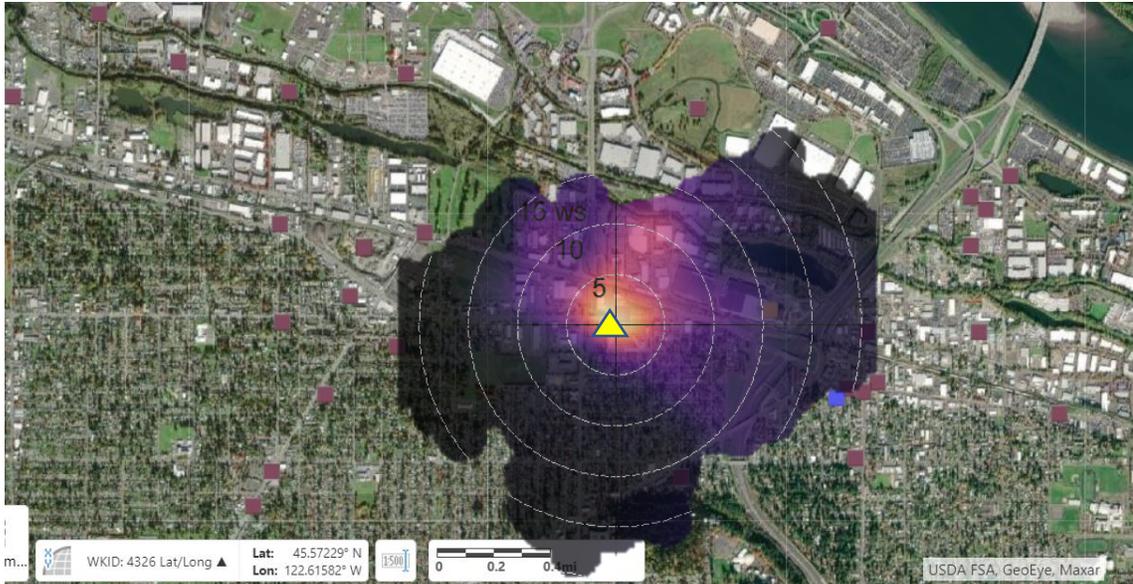


Figure 3-4. fmass over air quality permitted facilities (squares) around the Cully monitoring site.

In Figure 3-5 you can see that bbmass is lower than fmass through all metrics. There are times at night when fossil fuel activity is down and wood burning is up that causes the two to be nearly equal, but bbmass does not make up the major source of exposure. In Figure 8 the difference between fmass and bbmass is greater in the summer but less in the winter. However, in total fmass is much higher than bbmass contribution. Fmass reduces on the weekends, while bbmass does not change. The above means that traffic related activities are likely the culprit of health risks to the community.

Seasonally, Figure 3-5 and Figure 3-6 it appears there is less pollution than in summer months. However, this is more likely due to higher atmospheric boundary layer in the summer causing more mixing and dilution of pollutants. Therefore, activity and sources likely remain constant, but the exposure is slightly less due to dilution.

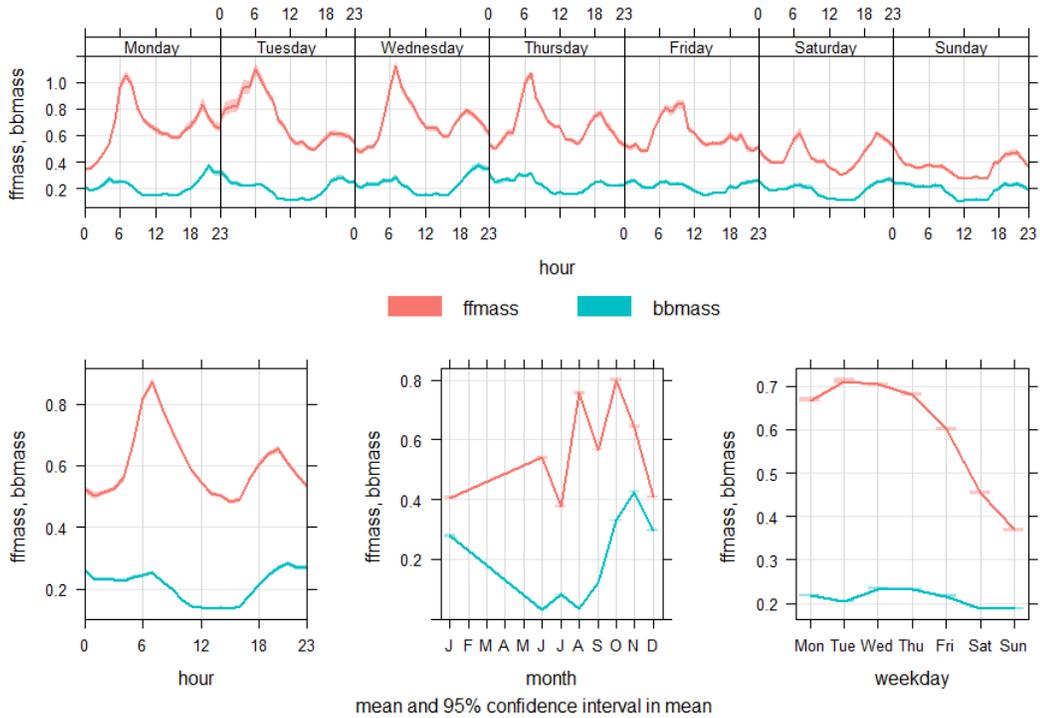


Figure 3-5. fmass and bbmass at Cully compared by hour and day of the week (top), hour (bottom left), month (bottom middle), and day of week (bottom right)

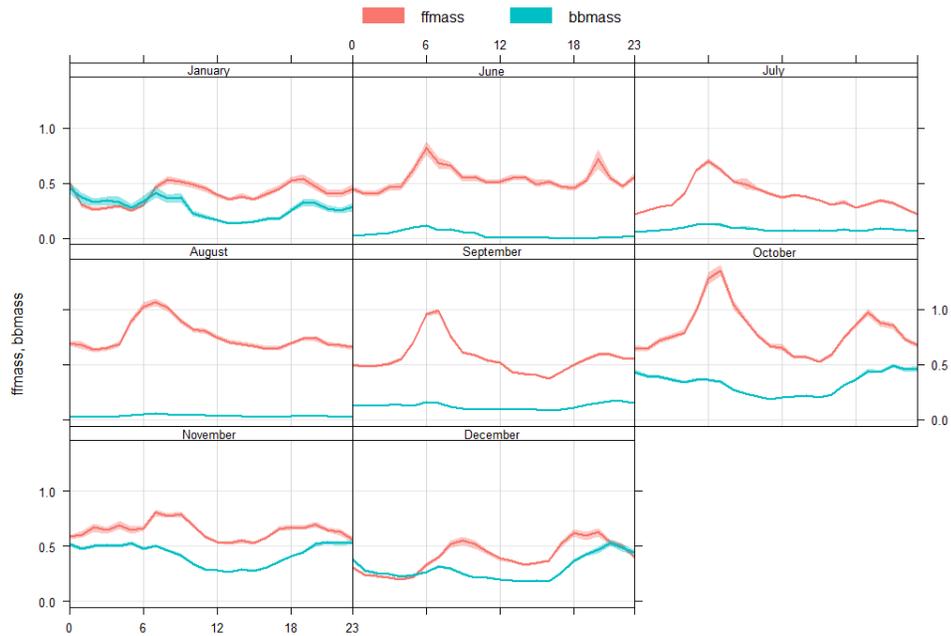


Figure 3-6. fmass and bbmass compared by month and by hour averages.

3.2.3 PAH Analysis at Cully

PAHs are known to be carcinogenic and is a major part of the danger of diesel emissions exposure (Lewtas, 2007). Lewtas discusses how PAHs are composed of hundreds of different semi-volatile compounds that have multiple naphthalene rings. The instrument EcoChem PAS2000 used for measuring PAHs for this study does real-time photo-electric ionization of particles (EcoChem Analytics, 2000). Particles alone should not have a charge, but the PAH attached to the particle will produce an electron if excited by a photon. Thus, the measurements from this instrument are only from particle bound PAHs. This is a different measurement technology compared to more traditional ways of

collecting PAH data through full speciation on a polyurethane foam (PUF) (Ravindra, Sokhi, & Van Grieken, 2008).

Data of speciated PAHs and a real-time PAH monitor were compared. First using the speciated PAHs from DEQ, diagnostic ratios were explored. Diagnostic ratios are ratios of a pair of PAHs in this case plotted against a ratio pair of another set of PAHs. PAHs composition is dependent of engine and fuel parameters, and can be different sources will emit PAHs with different characteristics (de Souza & Corrêa, 2016). The placement of points on the graph shows by what source that sample/day was dominated. Figure 3-7 shows that the results from Cully all fall within the biomass combustion section. Based on Ravindra, et. al. (2008), measurements of PAHs using PUF samplers are often hard to perform correctly, and much of the common methods in use lose the more volatile compounds during sampling. This could be seen in Figure 3-7 if there are

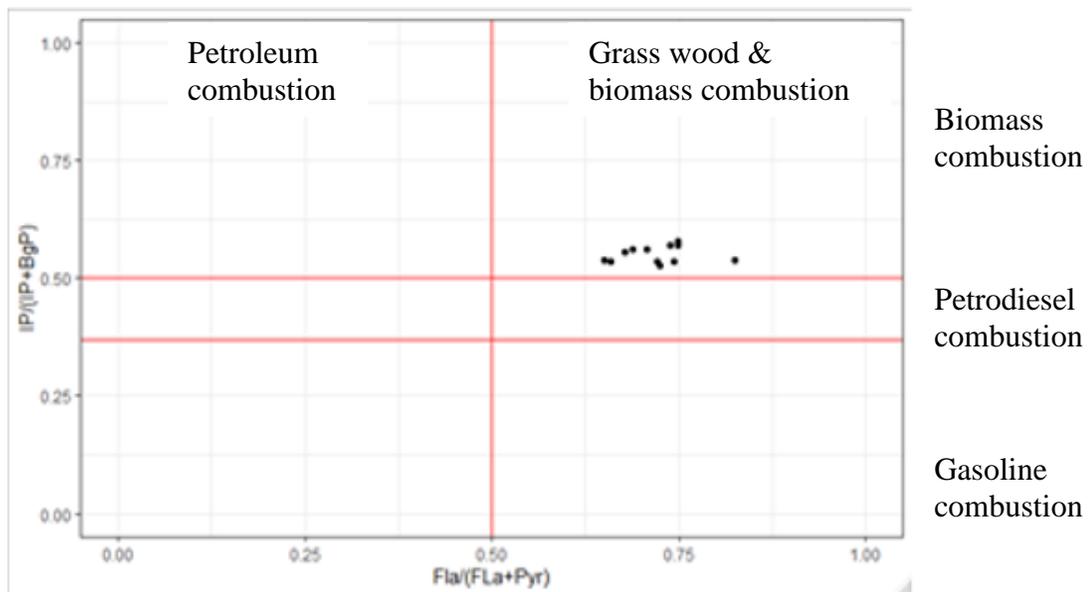


Figure 3-7. PAH diagnostic ratio from DEQ PUF sampler.

critical PAHs lost from the sample in warmer months because diesel is dominated by lower molecular weight PAHs.

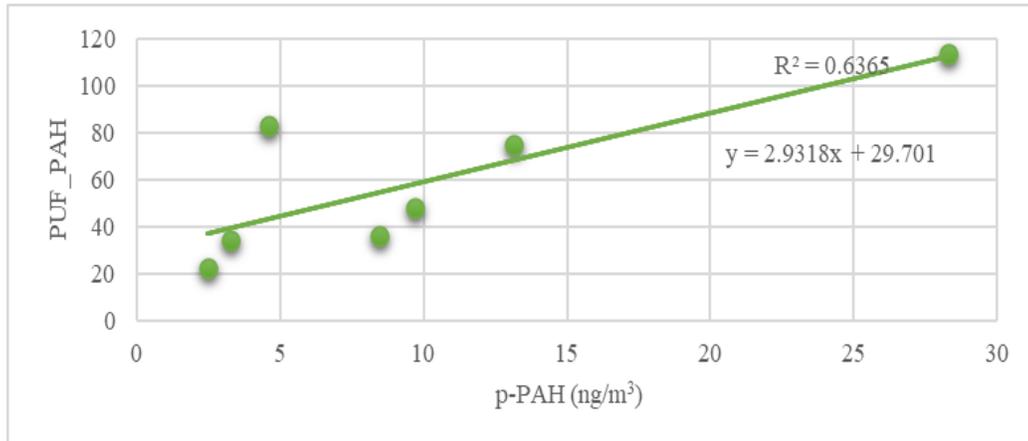


Figure 3-8. Comparison between p-PAH monitor and PUF 1-in-6 day sampler.

Next, the total PAHs from the PUF samplers and the average daily p-PAH were compared in Figure 3-8. There is some agreement with an R-squared values of 0.64 but not strong between the two and a larger sample size is needed to truly compare. Since these could only be compared on the one in six days that the PUF sampler was run, there was a small sample size during the Cully monitoring. The p-PAH monitor only detects PAHs with three or more rings, making Naphthalene undetected by this monitor. Naphthalene was taken out of the total for the PUF sampler results and compared to the p-PAH but the R^2 value reduced. More samples would show the variability across different scenarios.

3.3 Jade Monitoring

3.3.1 Sampling Methods

The Jade district site (45.500758, -122.571661) is located west of Interstate 205, South of Division Street and North of Powell Boulevard within a residential neighborhood. Specifically, the site is located at a wastewater pump station with a pond and trees on the east and to the west is a gravel road. This road could cause spikes of high counts for coarse particle dust. Traffic from arterials and the interstate here are the main obvious sources impacting the sites. Topologically, Jade is located between two Buttes, which could cause wind shadows. Kelly Butte is located East by ~ 1000 m, and Mount Tabor is Northwest by 1.13 miles.

In conjunction with Bureau of Environmental Services, DEQ was able to place a trailer on the Brooklyn pump station site and secure power from the pump station. The trailer had an AC unit, a computer, a met tower for wind speed and direction, and an Aethalometer measuring BC (Magee Scientific AE33) with a PM_{2.5} cyclone impactor all provided and maintained by DEQ. The NO_x, OPS, O₃ and SO₂ monitors were brought from Cully to the lab for maintenance checks and calibrations. In addition to the above monitors to replace PM mass estimation, a TSI DustTrak was added to the sampling suite. The Jade site was fully set up on March 23rd, 2020. This was a day after the Portland stay-at home order due to COVID-19 had been issued. Monitoring at Jade continued until the Wildfires and smoke events that started on September 6th, 2020.

Site visits were similar in manner to Cully. Most of the instruments were connected to the DEQ data logging system Envidas. Instruments that were not included the DustTrak and OPS, which needed periodic scheduling, zeroing, and data downloading for memory storage. Thus, site visits were between once to every two weeks at minimum. The Jade site did not include p-PAH so that the instrument could remain longer at Cully for future comparisons of PAH monitoring methods. Jade did not include any filter based or toxics measurements.

3.3.2 Jade Monitoring Results

Summary statistics are shown in Table 4. The mean of BC and fmass are 0.3 and 0.2 $\mu\text{g}/\text{m}^3$, respectively. This is double the DPM ambient benchmark set by Oregon. NOx and NO2 are 5 and 6 ppb, respectively. These values are not particularly high, especially compared to Cully, and other monitoring sites. There are high maximum values but for BC these are not as high as seen in Cully as well. The median and mean values of pollutants are relatively close meaning that the concentrations during sampling remained somewhat stable, despite the presence of outliers.

Table 4. Descriptive Statistics for Jade site monitoring

	<i>WD</i>	<i>WS</i>	<i>BC</i>	<i>ffmass</i>	<i>bbmass</i>	<i>NO2</i>	<i>NOx</i>	<i>O3</i>
<i>Units</i>	degrees	mph	μg/m3	μg/m3	μg/m3	ppb	ppb	ppb
<i>min</i>	0	0.1	0	0	0	0	0.9	0.5
<i>median</i>	191	0.9	0.24	0.17	0.06	4.1	4.6	21.6
<i>mean</i>	167	1	0.34	0.23	0.11	5.1	6	22.4
<i>max</i>	360	4.2	4.35	2.81	3.27	24.1	64.87	92.8

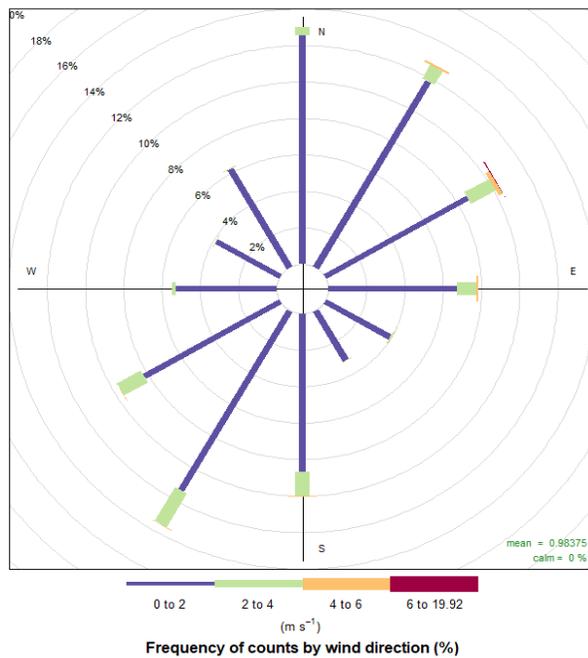


Figure 3-9. Jade site wind rose for sampling duration.

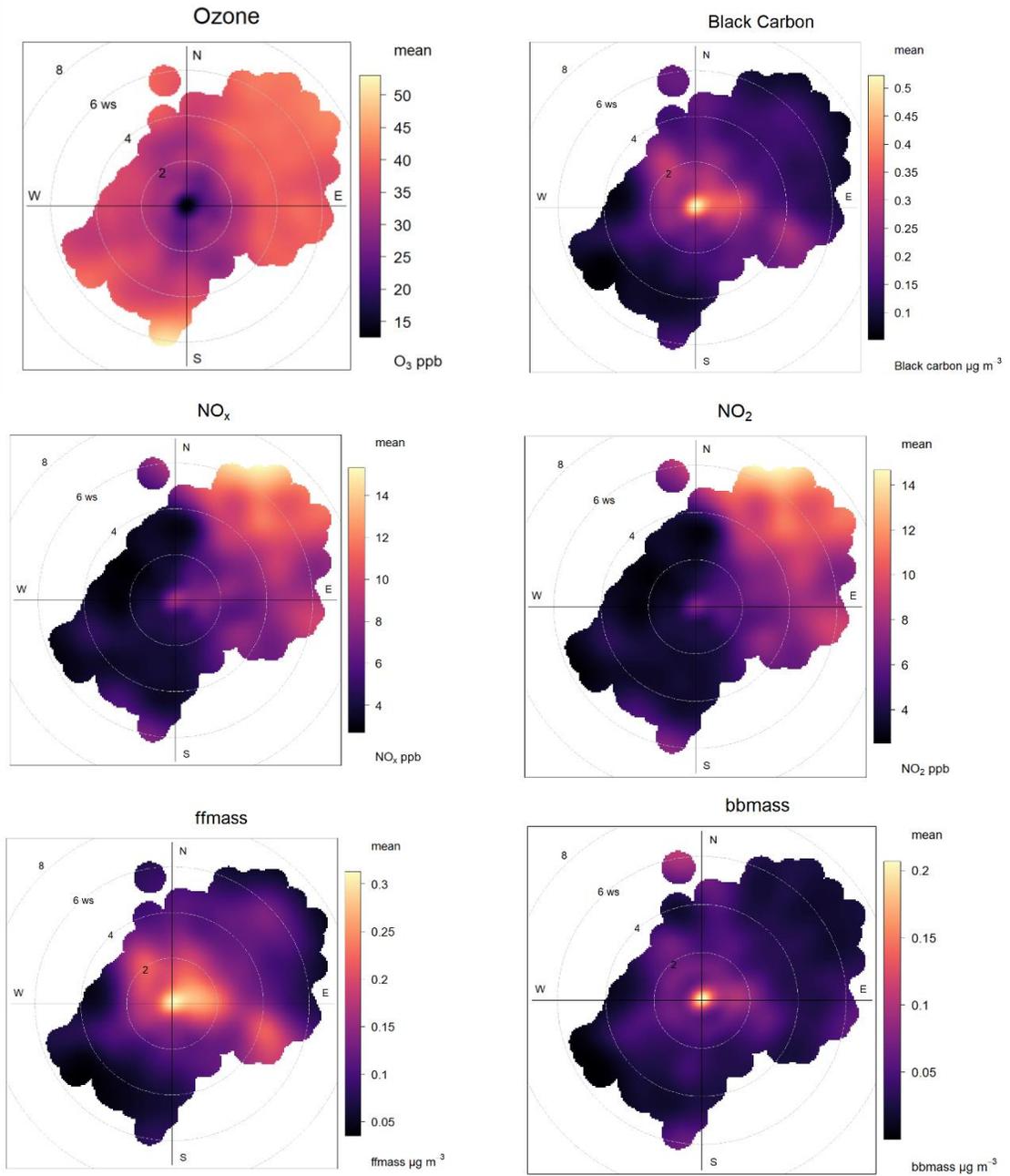


Figure 3-10. Polar plots of main pollutants at the Jade monitoring site.

The fmass polar Plot from Figure 3-10 of fmass is shown in Figure 3-11 over a map of the Jade district. The hotspots for higher pollution concentration appear over the highway and directly over the site itself. The higher contribution close in and around the site is indicative of lower wind speeds. Generally, with polar plots alone it would be hard to definitively point to a source because the plot is wind speed and direction, and there is no way of determining distance. However, it is reasonable to assume that there are not major diesel sources on Kelly Butte, the butte causes a wind shadow which would slow the transport of pollutants from behind. The highway therefore is claimed to be a not surprising source for fossil fuel BC. The only permitted sources within figure _ are gas stations and this would not be a major contributor over highway traffic emissions.

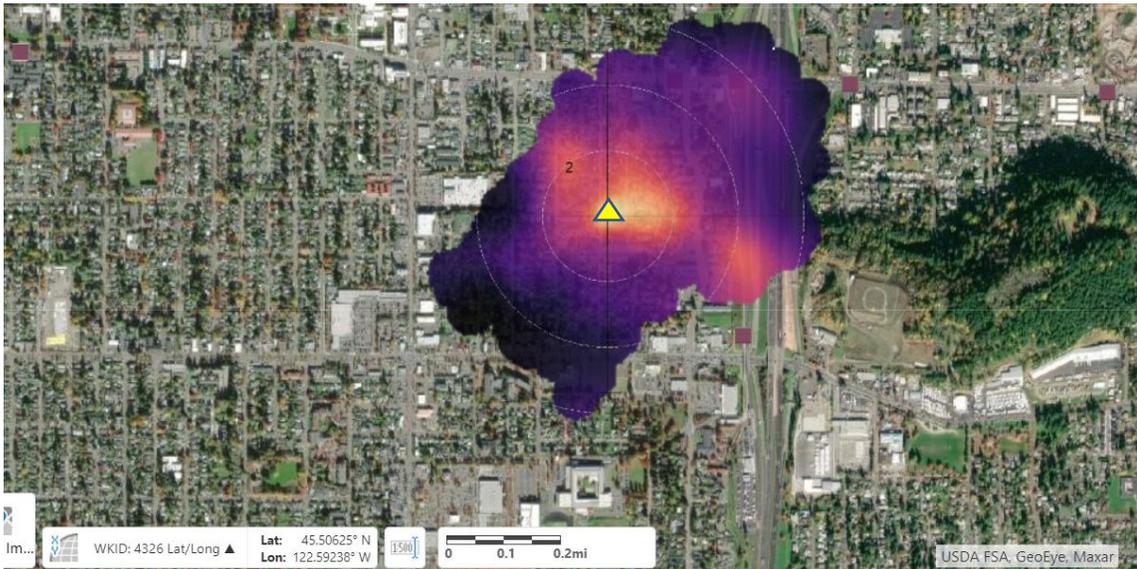


Figure 3-11. fmass overlaying a map of air quality permitted facilities in Jade.

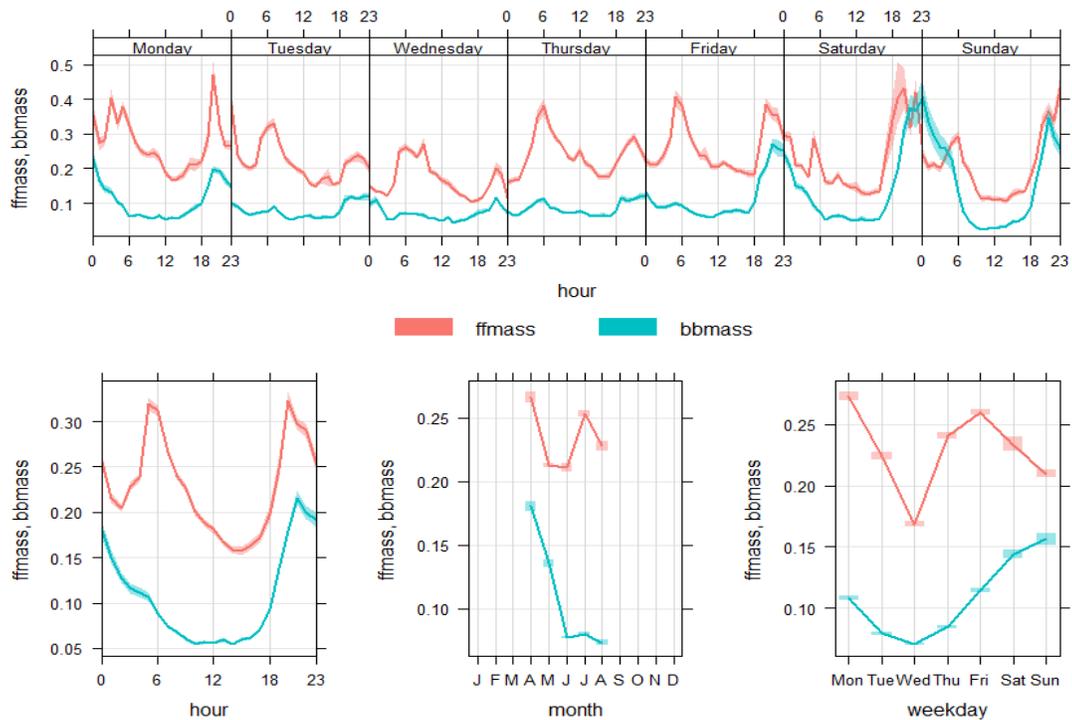


Figure 3-12. fmass and bbmass at Jade compared by hour and day of the week (top), hour (bottom left), month (bottom middle), and day of week (bottom right)
Z

Figure 3-12 shows bbmass and fmass as they vary through hour of the day for all days (top) then hour of the day averaged for all seven days (bottom left), average by month (bottom- middle), and by average for each weekday (bottom- right). Uncertainty is shown by the width of the bars and lines. There is relatively high certainty with these two variables for each time component, and fmass is higher in each way. Expected common patterns are observed in each time variable. In the hourly averaged plot (bottom-left) fmass shows evidence of impact of both morning and evening rush hours. In the same plot bbmass shows a general diurnal pattern that lacks the morning rush hour peak. The

particles the linger from the day stick around in the evening, and potentially more from wood fired stoves due to stagnation, and when atmospheric boundary layer raises, the bbmass lowers. Wood stoves are evidenced in the monthly averaged plot (bottom-middle) because of the large drop in the warmer months. The top plot is useful to check the hourly averaged plot to make sure there are not abnormal patterns or odd days. On Saturday night there is a higher amount of bbmass, which could also be attributed to wood stoves or bonfires.

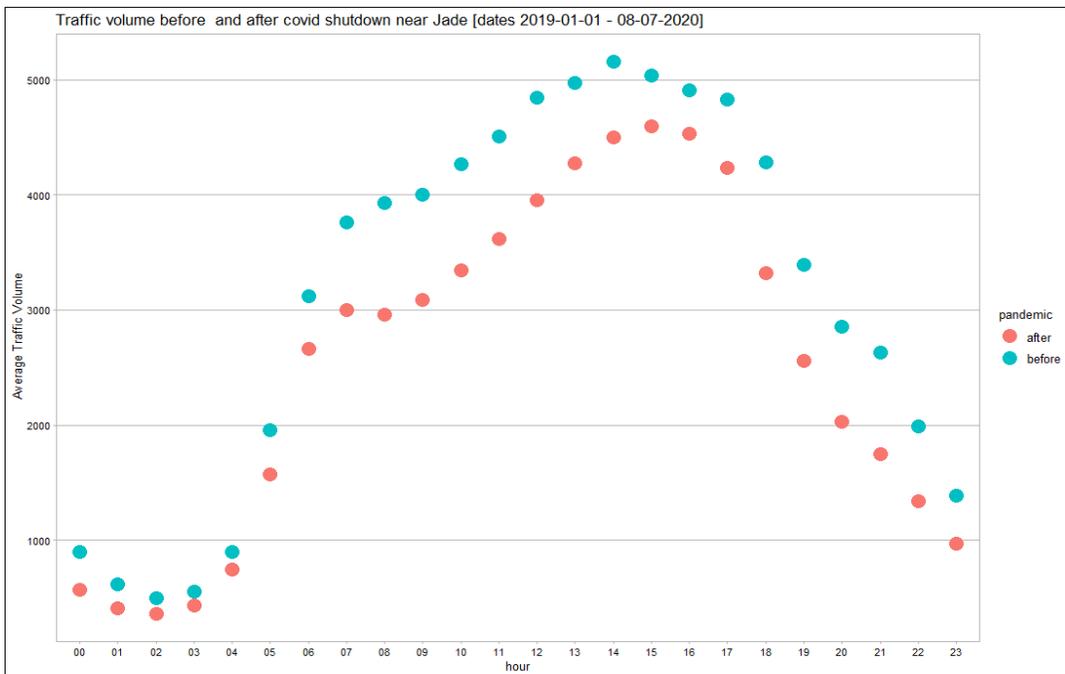


Figure 3-13. Traffic volume from on ramp from Powell to I-205 from 2019 to the end of monitoring at Jade.

Figure 3-12, shows how the ffbmass and BC in general is less predictable compared to other sites. It was investigated as to whether this was an effect of COVID-19 shutdown. Traffic volume via Portal (Portal, 2021) from Powell to I-205 on ramp near

the Jade site was pulled from January 1st 2019 to August 7th, 2020, when reliable data from Jade ended. The difference of average traffic volume by hour from before March 22nd, 2020, to the end of sampling compared to the year prior is shown in Figure 3-13. There is a small decline in traffic volume, however, not enough to substantially explain the difference in concentration and variability between monitoring sites. The Jade monitoring site is nestled in a residential community with two buttes surrounding it. The topography likely adds variability to the wind patterns in the community, making it harder to detect common patterns. Also, concentrations are likely less than in Cully due to the lack of many sources and industrial activity. Even though $ffmass$ is not expected from industrial plants and manufacturing, the draw of diesel trucks, and trains in the area do contribute to this. By comparing Jade and Cully's $ffmass$, Cully is likely more impacted by diesel PM compared to Jade. Jade is only connected to on-road traffic, and this is not distinguishable from diesel specifically. However, since Jade is only traffic and is lower than Cully, it can be considered that the addition of exposure in Cully is due to diesel emissions.

3.4 SE Lafayette Comparison

DEQ monitors pollutants at an inner SE Portland location that is representative of urban residential background concentration pollution. This site is on SE Lafayette Street, a block south of Powell Blvd. By comparing urban background concentrations to the Cully and Jade site, excess diesel pollution burden can be examined. In Table 5 the average concentrations of BC, $ffmass$ and $bbmass$ from an Aethalometer (Magee Scientific, AE33) are shown. BC was elevated in comparison to Jade and less than Cully.

The fmass and bbmass were 0.35 $\mu\text{g}/\text{m}^3$ and 0.28 $\mu\text{g}/\text{m}^3$, respectively. The percent contribution from both BC measurements are closer compared to Cully and Jade. The hourly graph of fmass and bbmass in Figure 3-14 show there is a morning rush hour peak for fmass and bbmass lacks the same rush hour pattern. The difference between biomass burning (wood smoke) widens during the day when the atmospheric boundary layer rises, then narrows as the evening hours approach with presumably residential wood-smoke rises with cooler nighttime temperature and fireplace usage. The residential wood-smoke is also seen in the right graph in Figure 3-14, when the bbmass exceeds or matches fmass, while the difference is more pronounced in the warmer months.

Table 5. Aethalometer averages for SE Lafayette Monitoring site.

<i>measurement</i>	<i>Average $\mu\text{g}/\text{m}^3$</i>
BC	0.64
ffmass	0.35
bbmass	0.28

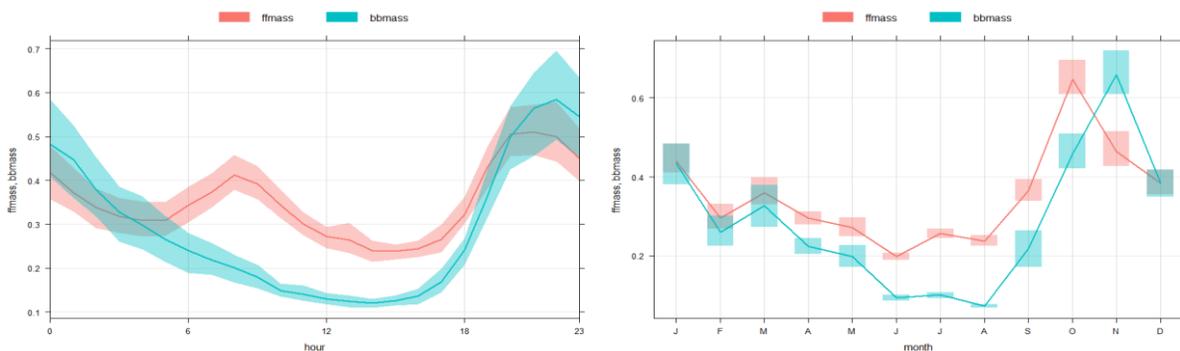


Figure 3-14. SEL fmass and bbmass by hour (left), and by month (right).

SEL had a pollution load of total BC that is between Cully and Jade sites. There was a larger proportion on bbmass meaning that woodsmoke in this site and has a more dominant impact than the other sites identified with diesel vulnerability. Even though bbmass had a higher relative contribution at SEL, fmass still outweighed total contribution of BC with a higher overall average. This shows that even in residential background stations, that traffic sources dominate BC exposure.

Chapter 4. Model validations

The output for the PATS model has the nearest two receptors concentration at 14 times the Oregon DPM ambient benchmark concentration of $1.4 \mu\text{g}/\text{m}^3$ at Cully. At the Jade site, the concentration is $0.4 \mu\text{g}/\text{m}^3$ for the PATS model. However, measured concentrations of BC were 0.5 and $0.2 \mu\text{g}/\text{m}^3$ at Cully and Jade, respectively. That is a difference of 0.9 and 0.2 at Cully and Jade respectively of the PATs model compared to ground-based monitoring. Further, the source category that were most influential the receptors in the PATS model receptors was construction making up $0.37 \mu\text{g}/\text{m}^3$ at Cully and $0.2 \mu\text{g}/\text{m}^3$ at Jade. This proportion of construction that is included in PATS subtracted from the total makes the values much closer to the ground-based quantities.

The Community Monitoring and Analysis System (CMAS) C-Port was created by the University of North Carolina Chapel Hill and supported by the United States EPA. CMAS is an online tool that does quick easily performed dispersion models for research grade analysis of fossil fuel emissions (Hill, 2016). It includes much of the same inputs as PATS with traffic volume, rail, marine, and point sources. After running a model for wintertime average conditions (Cully sampling season), with neutral atmospheric conditions, and not modifying any of the default inputs, the modelled concentrations for

PATS, CMAS, and observed are shown in

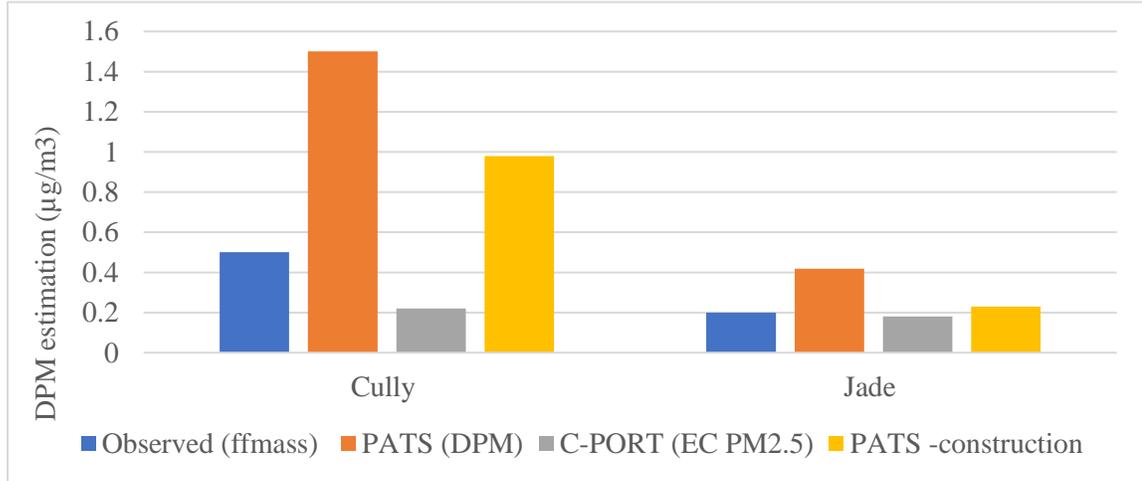


Figure 4-1. At Cully, the fmass concentration was $0.5 \mu\text{g}/\text{m}^3$, which is a $1.0\mu\text{g}/\text{m}^3$ less than the PATS predicted. The C-PORT concentration for this area was roughly $0.3 \mu\text{g}/\text{m}^3$ lower than the observed concentrations. For warm weather and neutral conditions at the Jade site the CMAS modelled concentration was $0.18 \mu\text{g}/\text{m}^3$ and the measured fmass was $0.2 \mu\text{g}/\text{m}^3$, the PATS model predicted this area to have $0.42 \mu\text{g}/\text{m}^3$. The PATS was much higher than the CMAS and observed concentrations. However, Jade was sampled only during the peak of the 2020 Pandemic shutdown and thus should have somewhat lower concentrations to normal.

Table 6. Comparison of ground-based measurements, PATS, and CMAS C-Port models.

	<i>Observed (ffmass $\mu\text{g}/\text{m}^3$)</i>	<i>PATS (DPM $\mu\text{g}/\text{m}^3$)</i>	<i>C-PORT (EC PM2.5 $\mu\text{g}/\text{m}^3$)</i>	<i>PATS - construction</i>
<i>Cully</i>	0.5	1.5	0.22	0.98
<i>Jade</i>	0.2	0.42	0.18	0.23

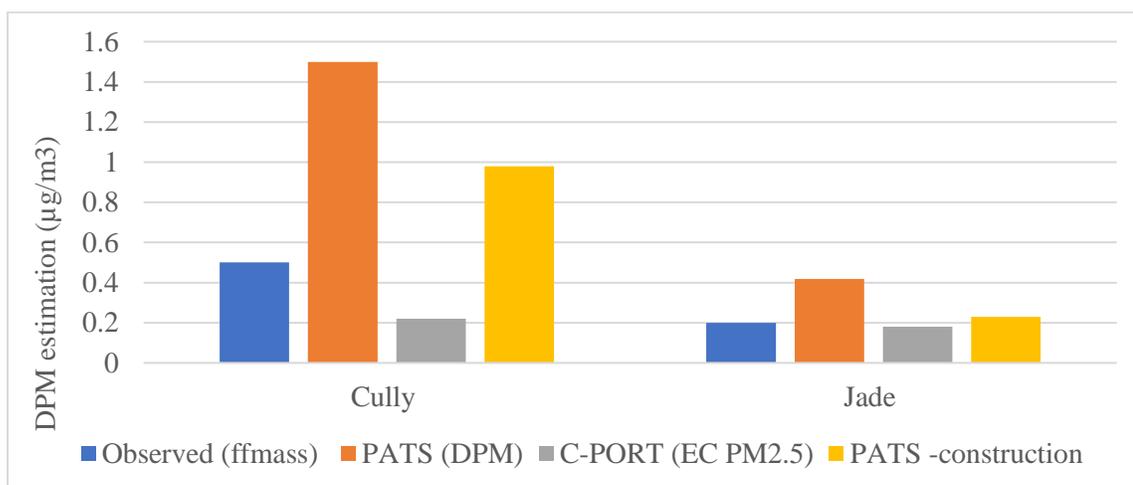


Figure 4-1. Comparison of measurements PATS and C-PORT modelled DPM

The main source affecting both the Cully and the Jade sites in the PATS model is the expected construction emissions. The contribution from construction sector were expected to be 0.4 and 0.2 $\mu\text{g}/\text{m}^3$ for Cully and Jade, respectively. If these concentrations were deducted, then the concentrations match more closely to observed values. Jade's concentration is less than the expected post-construction deduction. C-PORT expects that 99 percent of Jade's emission are from traffic, so this could be due to the slight decline from traffic. However, the modelled and observed results are very close and it performed well at estimating exposure. Most sources are relatively the same between PATS and C-PORT, except construction is not included in C-PORT. It should be noted that construction was not paused or reduced due to COVID-19 as this was a business sector that could remain working outside.

Chapter 5. Fuzzy C-means Cluster Analysis

5.1 Cluster Rationale

Source characterization was attempted at each neighborhood that would lead to a more detailed understanding of community exposure. The traditional method of PMF analysis requires uncertainty values for each sample (US-EPA, 2014) and is designed for long-term sampling. Other methods could have been used, but many of such methods rely on back trajectories in conjunction with emission inventories (Salimi et al., 2014). Back trajectories using Hysplit were explored, however, for this application it was found that the minimum height Hysplit uses would be too high for an inter-urban and high-resolution analysis sought for in this study. Hysplit is meant more for a regional or city wide at best resolution. Therefore, there is not much improvement of knowledge over general meteorological measurements included in the analysis already.

Using the in-situ measurements, a greater level of detail is obtained through higher temporal resolution. Through the greater level of detail, it was expected that patterns would arise that would differentiate diesel from gasoline sources. Cluster analysis was chosen as the optimal choice for the task. There are few studies that use cluster analysis in air pollution apportionment,(Thrane & Gunderson, 1986) (Thrane, 1987)(Salimi et al., 2014), and none to date for diesel emissions, or using the same pollutants. Testing the ability of cluster analysis for high time-resolution sampling furthers our understanding of potential for low-cost source apportionment, and limitations of cluster analysis.

Cluster analysis is simply a multivariate analysis technique like Principal Components Analysis by using a distance matrix, and by finding the closest points to each other. Further, clusters are found by minimizing the distance between points. Where the center of each cluster is, is a characteristic or pattern that is descriptive of all the points that belong to that cluster. Since air masses are not from one source only, a fuzzy cluster analysis was used instead to capture the mixture of contribution more accurately. In the popular k-means clustering and other agglomerative clustering techniques, a sample or data point can only belong to a single cluster. Fuzzifying a cluster analysis shows each data points strength of relationship to the cluster center from zero to one. Complete membership or absolute center of the cluster would be one and not belonging at all to the cluster would be zero (Dalmaijer, Nord, & Astle, 2020). By using a fuzzy analysis instead of a regular k-means, each data sample of air mass can have a score for how much it relates to each cluster pattern. Then those scores can be shown through other variables such as wind patterns, time variations or how each cluster correlates to individual variables.

Hierarchical techniques were not used because atmospheric sampling dynamics make for fast switching sources, and rapid changes in the data. Therefore, hierarchical clustering would not perform well, and the other option was agglomerative clustering. This technique places highly dimensional data into a distance matrix. Then, based on this distance matrix, points close to each other are placed in a group.

The measure of distance chosen was the Pearson Correlation (Kassambara & Mundt, 2019). This measure finds the most linearly similar points. Other potential and

popularly used distance measures are the Euclidean, Spearman, Manhattan, and more. Pearson was chosen as the best fit for this analysis based on the data not fitting assumptions of the Euclidean distance for normality..

The analysis starts with selecting the variables of interest, that would be useful in finding diesel sources. Thus, variables associated with fossil fuel combustion were used as analysis inputs. Variables used in each mode are shown in the correlation result plot in figure Figure 5-6 for Cully and Figure 5-12 for Jade. The measurements included in the analysis were bbmass, ffbmass, ozone, SO₂ and PAH when available, Particle size distribution was measured with an Optical Particle Sizer (OPS), which measures particle sizes 0.3-10 µm distributed into 17 bins.

5.2 Cully Clustering Results

Cully clustering used OPS size bins 1-3 (0.3 -0.5 µm), ffbmass, bbmass, SO₂, Ozone, PAH, NO₂ and NO_x. The Calinski method in Figure 5-1 shows that even though two is marked in the red dot as the most optimal, but this is the default that is shown when no real clustering is found. Therefore, four clusters were chosen based off the orange (next best).

A reasonable cluster structure was found, based off silhouette width plot in Figure 5-3 shows the average width to be 0.57. The sample sizes are also shown in Figure 5-3, and there are very unequal sample sizes between groups, with cluster three being the largest. The resulting cluster plot is shown in Figure 5-2 where the main two dimensions of variability are shown on the axes. It can also be observed in Figure 5-2 that there are not major outliers in the data but little separation between clusters.

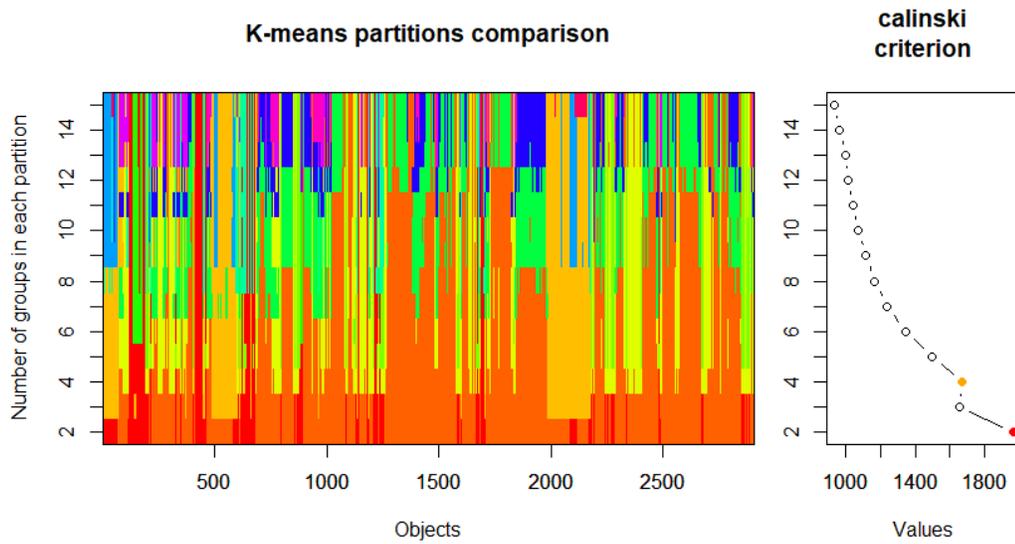


Figure 5-1. Sample partitioning among various groups (left), and best number of partitions using Calinski criterion (right) for Cully.

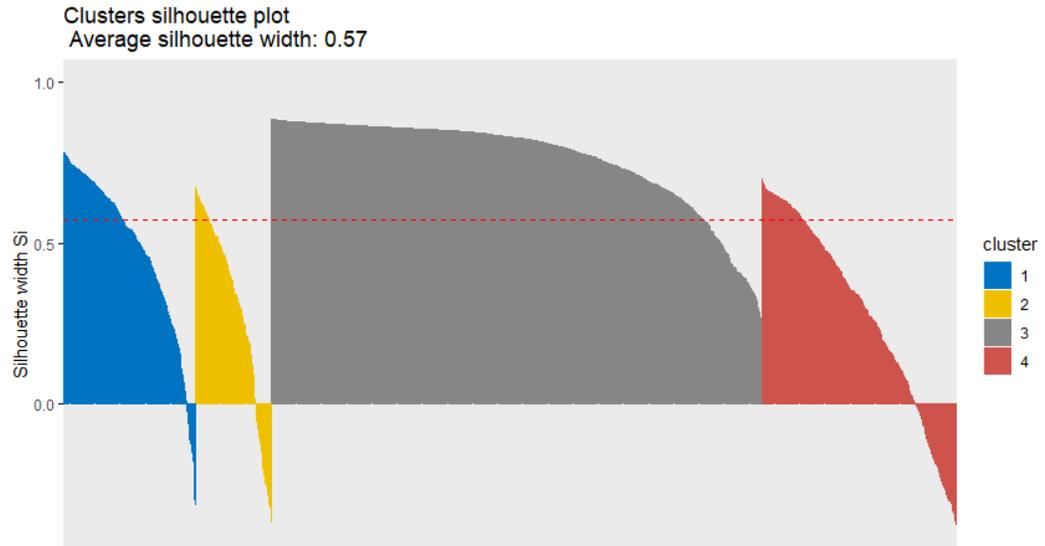


Figure 5-3. Silhouette width plot for 4 partitions for Cully

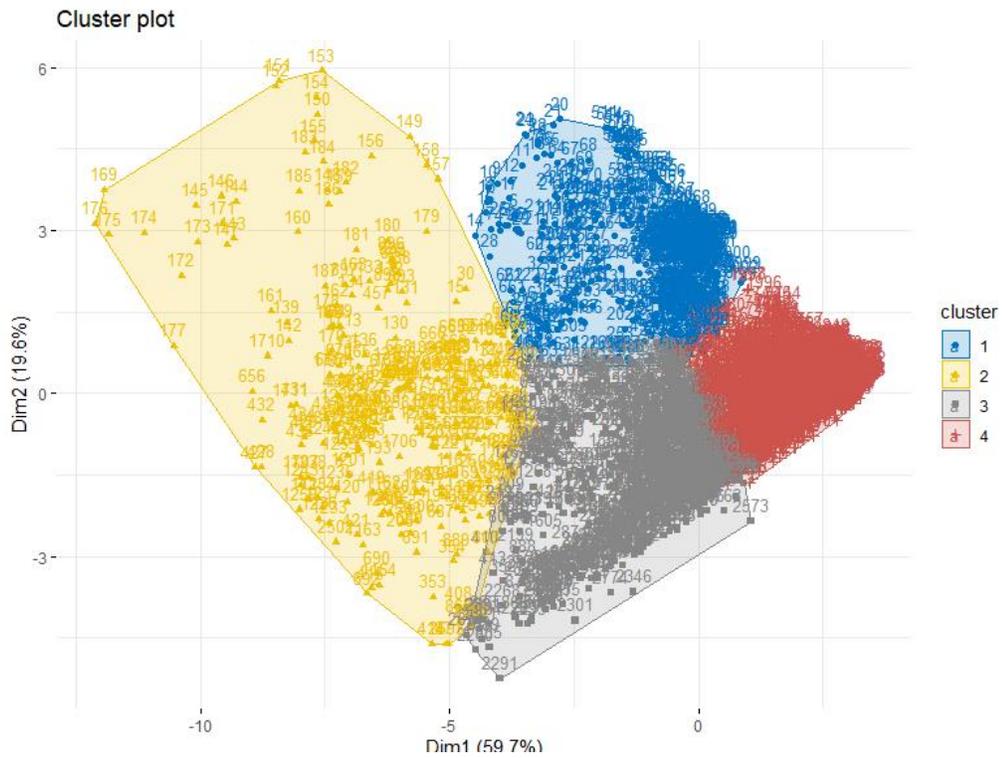


Figure 5-2. Cluster plot of Cully fuzzy clustering.

Cluster one has a distinct hot spot in Figure 5-4 to the east. It lacks a time-of-day variation in Figure 5-5. On Monday and Sunday, it does have higher values, this could be due to a large contribution or extreme value leading this value.

Cluster 2 has the smallest sample size, and a weak silhouette width of 0.29. In Figure 5-4 the cluster seems to appear from North easterly winds and low mixing states. In Figure 5-5 there is no large variations in time of day, or day of week. Cluster 2 does correlate well with bbmass, PAH, fmass, BC, and NO_x. It appears that it correlates the best with BC with 80 Pearson correlation coefficients, this would explain why its correlated with both bbmass and fmass.

Cluster 3 has the most samples and highest silhouette width of 0.76 and thus has a strong structure. In Figure 5-4, cluster 3 comes from well mixed air seen by the low membership in the bullseye of the chart, meaning that the air is moving when cluster three is present. When it is present it is coming from Southeast winds and North westerly winds roughly between five and ten mph. These directions are potentially around Killingsworth and I-205. Cluster 3 is not highly correlated with any particular pollutant as the other clusters are. However, it is most correlated with NO₂, which is indicative of fresh combustion emissions. In Figure 5-5 this cluster is elevated during working hours, and with the NO₂ correlation, and potential traffic directionality, it could be identified as a traffic source.

Cluster 4 also has a very low silhouette width of 0.28, and a medium sample size. In Figure 5-4 there is a striking directionality with this source. It has a strong presence

with winds from the Southwest. In Figure 5-6 it is also strongly correlated with Ozone and somewhat wind direction. For these reasons it would be hypothesized that cluster four is a solar radiation and Ozone dominated cluster. In the average by hour shows that there is a morning peak of cluster four. This peak could be due to the reduction of O₃ in the ambient atmosphere due to the reaction with NO and NO₂, and thus the reduction during morning rush hour.

In conclusion cluster 3 is the strongest cluster and can be potentially interpreted as fresh traffic emissions. To confirm this there a correlation to traffic volume would be ideal. Cluster 1 is the next best cluster and could potentially just be represented by particle count. Cluster 2 and 4 are the weakest and should be interpreted with limited

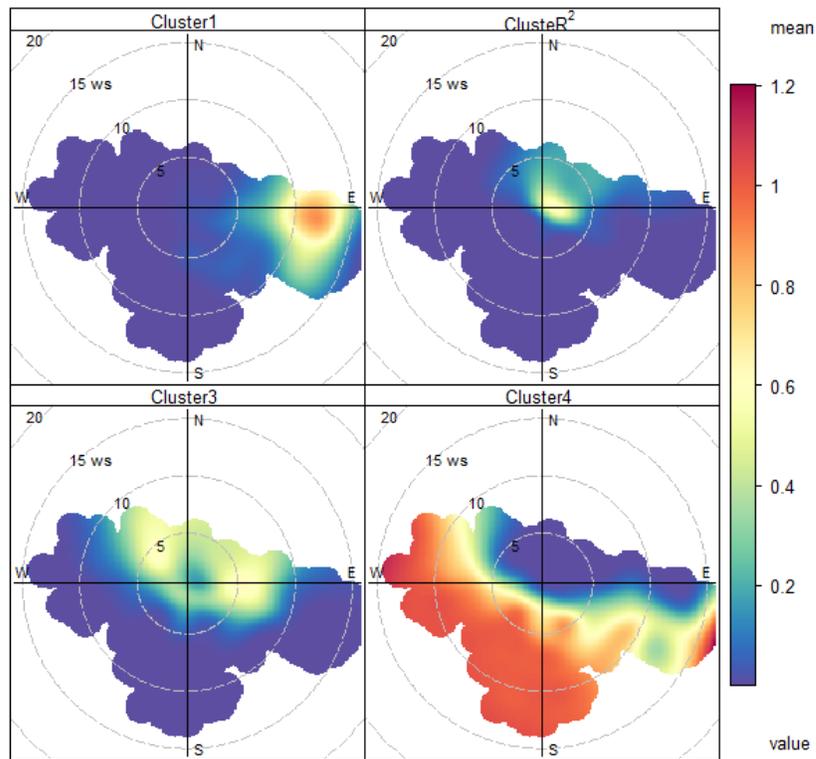


Figure 5-4. Polar plot of four cluster fuzzy memberships in Cully.

confidence. Overall, there was a reasonable structure in the clustering that was influenced by the strength of Cluster 3.

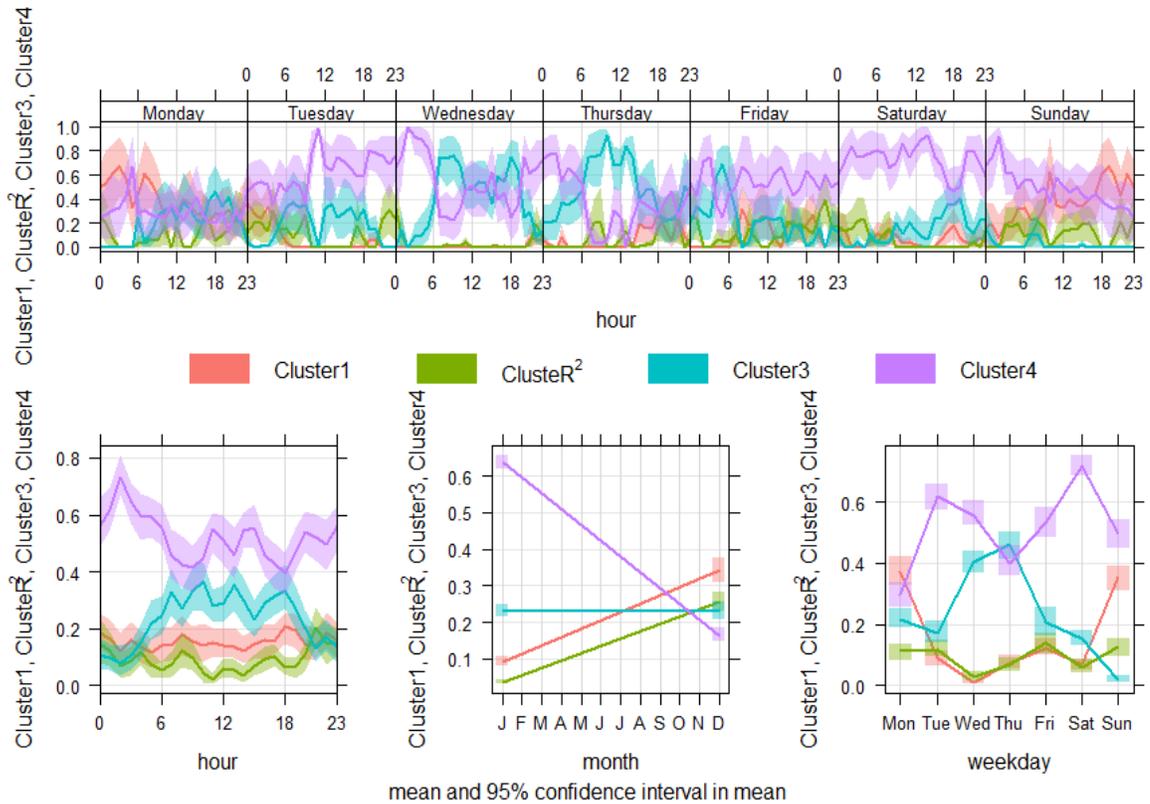


Figure 5-5. Cully's Time Variation plot of fuzzy cluster memberships equaling 1. Hour and day of week (top), average hour contribution (bottom-left), month (bottom-middle), day-of-week (bottom-right)

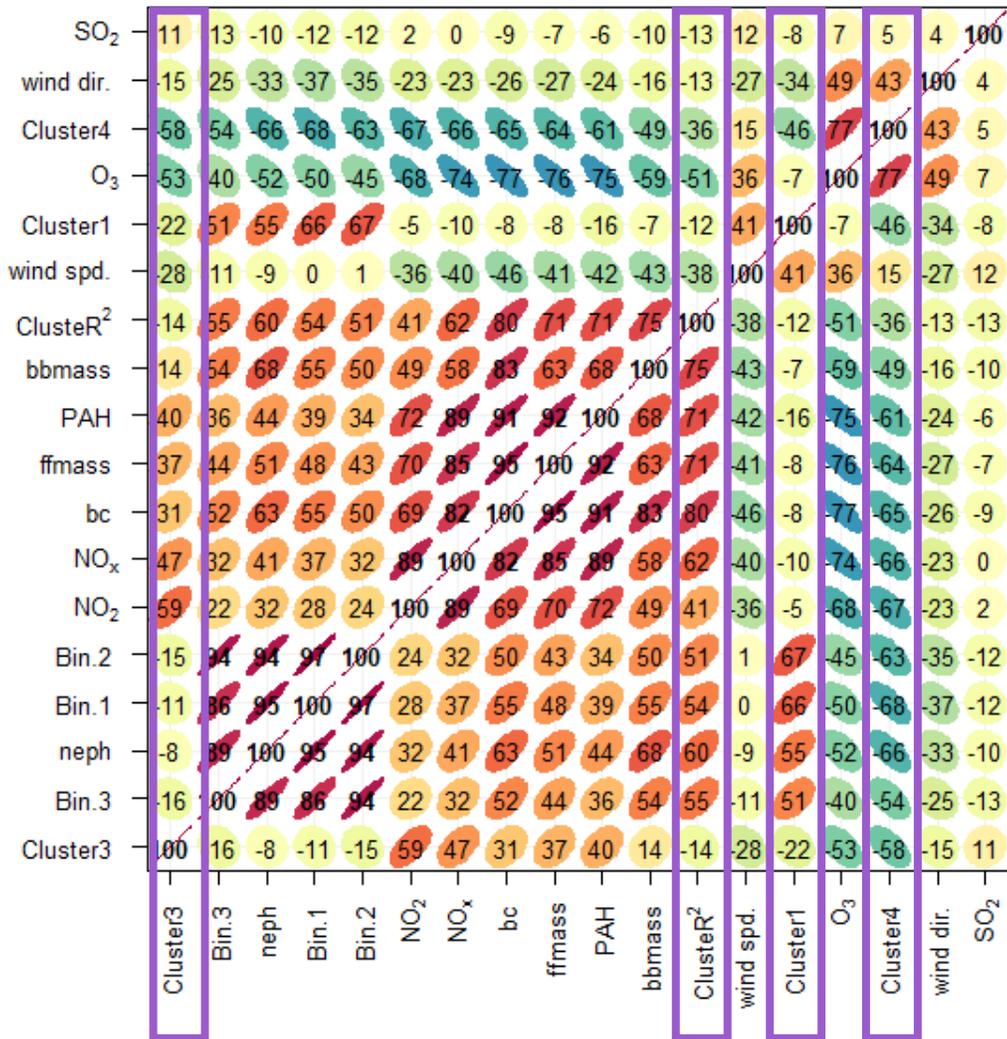


Figure 5-6. Cully's correlation plot based on Pearson correlation coefficient between fuzzy cluster memberships and model variables.

5.3 Cluster Results for Jade

Jade clustering was best performed with variables NO_x , NO_2 , bbmass , ffmass , and particle size bins 1-3 ($0.3 - 0.5 \mu\text{m}$) the same as Cully. The data was also averaged to 15 minutes to match the Cully analysis and reduce run times. The Calinski criterion showed 2 as optimal but as described in the Cully clustering, the next best number was chosen again.

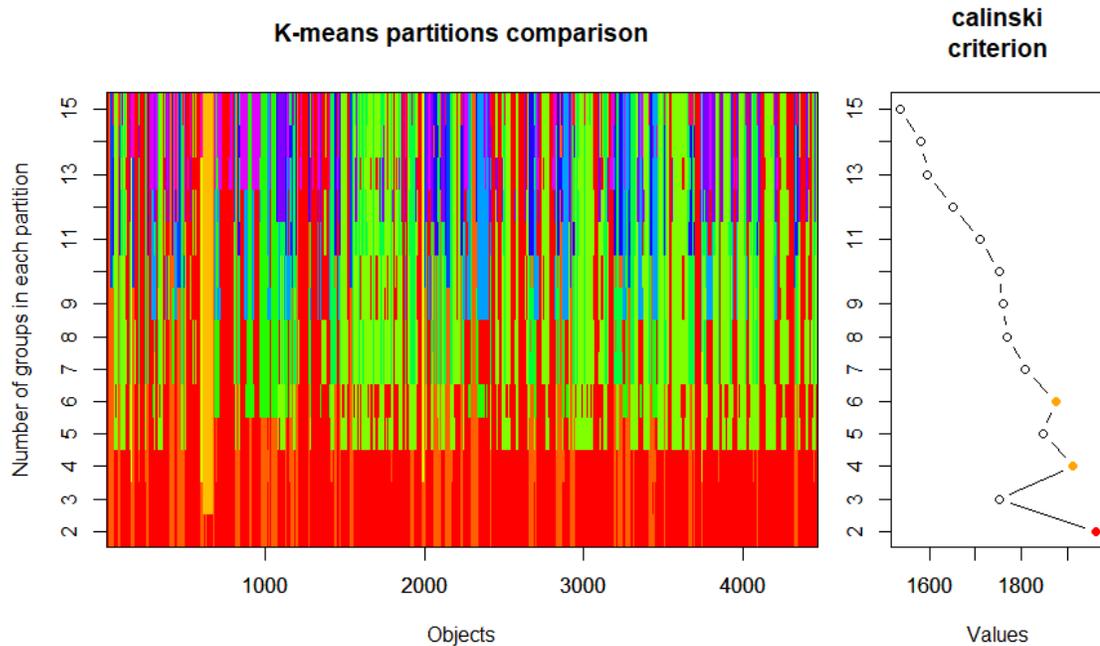


Figure 5-7. sample partitioning among various groups (left), and best number of partitions using Calinski criterion (right) for Jade.

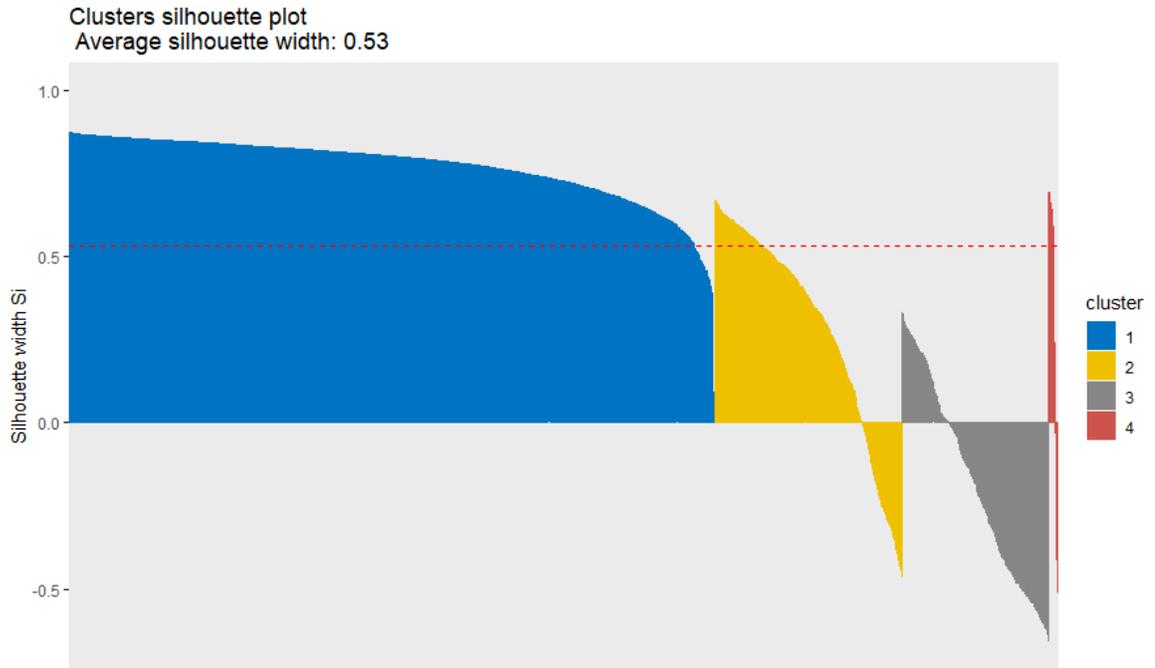


Figure 5-9. Silhouette width plot for 4 partitions for Jade.

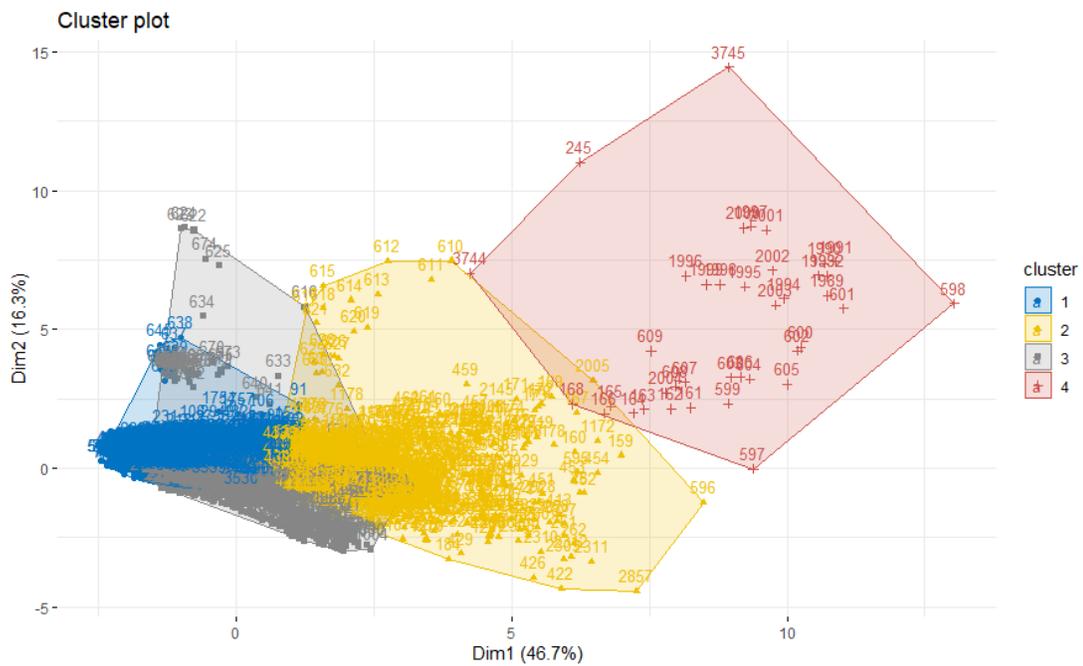


Figure 5-8. Cluster plot of Jade fuzzy clustering.

The average Silhouette width for Jade is 0.53, which, is slightly less than Cully and barely making it into the category of having a reasonable structure. The sample sizes per cluster are very different with cluster one having the largest group by far and the strongest width of 0.77. Clusters 2, 3 and 4 have silhouette widths of 0.28, -0.18, and 0.33, respectively. There does not appear to be strong outliers in Figure 5-8, but there is strong overlap between cluster 3 and cluster 1. The overlap is the reason for the negative silhouette score for cluster 3.

In Figure 5-8 Cluster 1 is seen to be very squished together at the left side of the cluster plot, while the other clusters spread off from there. In Figure 5-10 cluster 1 comes from the south and from the northwest and a smaller hotspot to the East. There is little variation in time patterns seen in Figure 5-11, however, this as to be expected has the most contribution of membership seen in all the time variations. This cluster is not correlated with other variables in Figure 5-12. Cluster 1 could be a background variable that just describes where the general wind and there are no extreme pollutants standing out.

Cluster 2 in yellow in Figure 5-9 and 26 is the next largest sample size. Cluster 2 is shown in Figure 5-8 to be present during low mixing and stagnation, with small amounts coming from 45 and roughly 120 degrees. This cluster is correlated most with fmass NO_x and NO₂, while being negatively correlated with Ozone. If this were a diesel or traffic source it would be expected that there would be a general traffic pattern but

there is only a morning peak at 6 am and then a complete drop for the afternoon hours that would be expected to be higher.

Cluster 3 in grey for Figure 5-9 and 26 has a negative silhouette score and thus the negative samples could have been assigned to the wrong cluster. With this in mind, there are still interesting patterns that could be looked at. Starting with Figure 5-12 this cluster is correlated with Ozone and wind speed. This is similar to what was seen in Cully except for with wind direction. In Figure 5-10 there is a hot spot in 5 mph NE of the site that could mean when winds are high, its able to transport Ozone. As well as in Figure 5-11 there is a peak after noon. All the post analysis graphs for cluster 3 point towards a solar radiation and Ozone pattern.

Cluster 4 has the smallest sample size and contains the most spread-out points in Figure 5-8. Due to the small sample size Figure 5-10 and Figure 5-11 have little contribution or variability. In Figure 5-11 there is the faintest rise in concentration during nighttime hours, and a dip from March to April. Figure 5-12 shows cluster 4 is correlated with bbmass and the particle size bins. All the observations indicate that cluster 4 could be wood smoke. As the time frame for Jade took place during spring and summer wood smoke from household fireplaces likely reduced and accounts for the level of membership for this cluster. Comparing this cluster to a wintertime analysis could be useful to see the impact wood smoke has on the community.

In conclusion, cluster 1 appears to be a background concentration. Cluster 2 could be a traffic source, but without more evidence. Cluster 3 is likely based on solar

radiation and ozone but is weak. The overlap between solar radiation and ozone to a background source however does make sense, as these two in theory have regional and dispersed natures, unlike nearby combustion sources. Lastly cluster 4 is likely wood smoke.

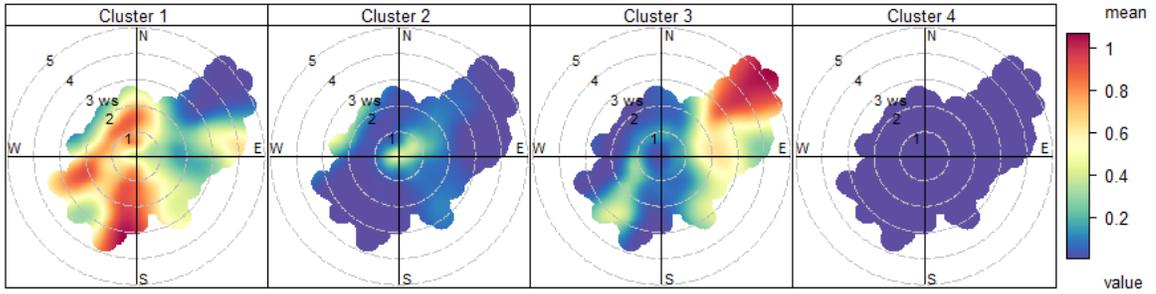


Figure 5-10. Polar plot of four cluster fuzzy memberships in Jade.

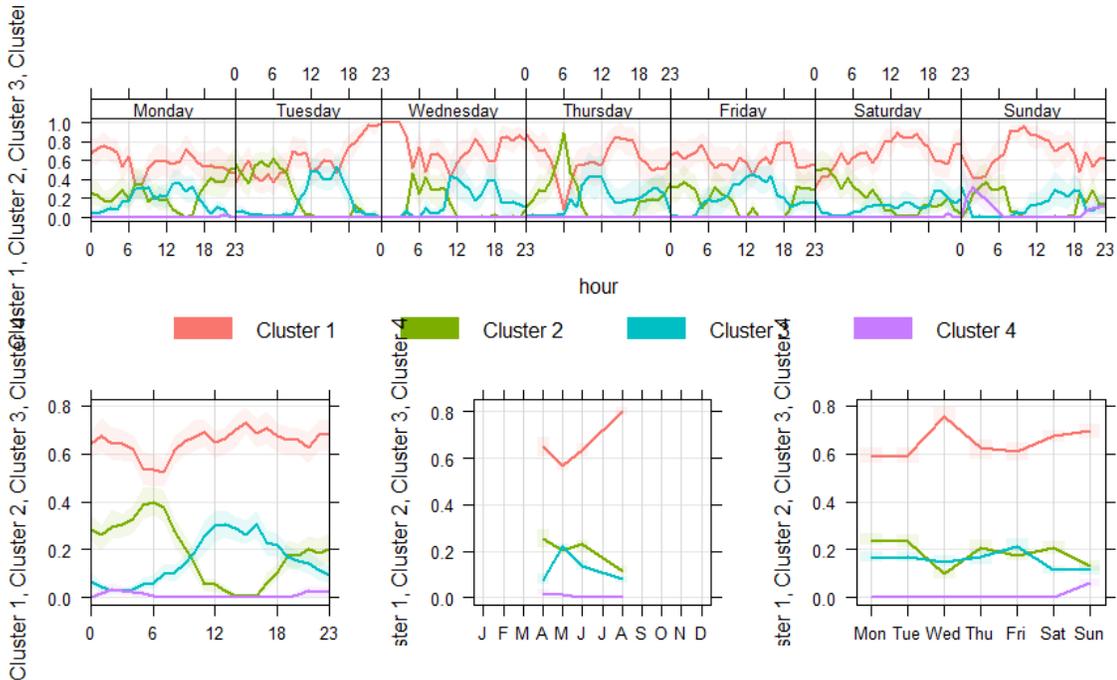


Figure 5-11. Jade's time Variation plot of fuzzy cluster memberships equaling 1. Hour and day of week (top), average hour contribution (bottom-left), month (bottom-middle), day-of-week (bottom-right)

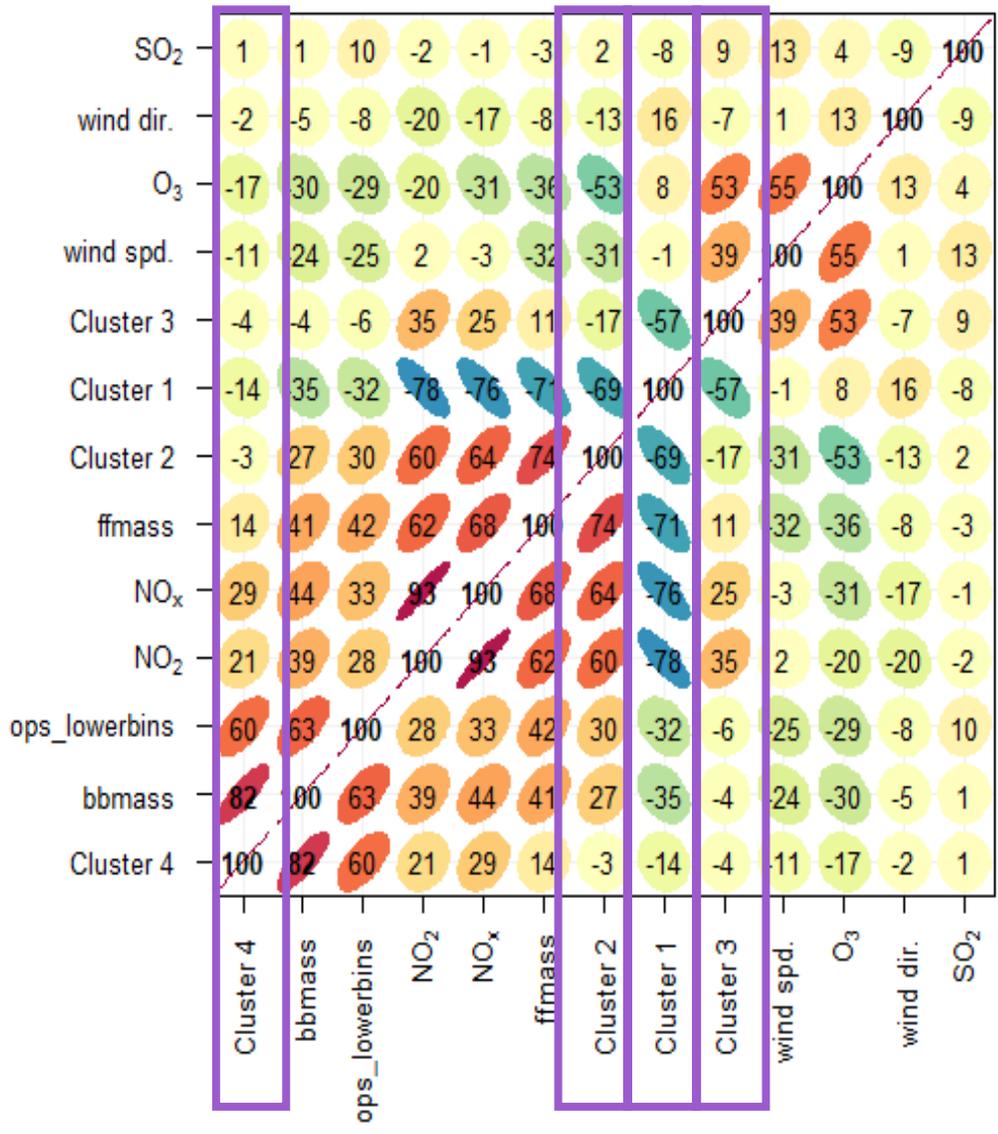


Figure 5-12. Jade's correlation plot based on Pearson correlation coefficient between fuzzy cluster memberships and model variables.

5.6 *Summary of Fuzzy Clustering*

On average reasonable structures were found in the clusters for Cully and Jade, with some clusters having more strength than others. The ability to interpret clusters however is limited. The OPS measured at a range that has little overlap in diesel and gasoline particle size mode ranges. If used, the TSI SMPS would have been able to pick up the differences between gasoline and diesel (Rivas et al., 2020).

Cluster analysis is an exploratory tool. One should be wary of performing a full source apportionment with cluster analysis. However, patterns could arise from stronger clusters and then the analyzer could look through clusters for patterns such as was shown above. Fuzzy c-means analysis is a promising tool for future works that aim to look at high time-resolution multivariate data.

Other ways in which this analysis could be improved in the future would be to ensure more complete data, to create a larger sample size. Imputations were an option but not performed based on the idea that the missingness differed across variables. If instruments were not recording and data was imputed during that time frame and there was a spike in concentration for related pollutants, the strength of the analysis would be reduced.

Chapter 6. Conclusion

6.1 Summary of Project

Field investigations were performed for each of the four sources with the highest uncertainty from the PATS model, namely rail, marine, distribution centers and construction.

- (1) Rail: Estimations of DPM emissions from rail were improved by obtaining locally relevant emission factors and activity assessments.
- (2) Marine vessel activity was improved with a high-resolution activity data.
- (3) Distribution centers were identified as a viable source for modelling and proves to be a section of urban emissions for further research.
- (4) Construction has been shown through community monitoring and direct measurements that it does in fact cause local impacts, but for chronic exposure modelling methods have a need improved inputs.

Modeled ambient concentrations of DPM in the PATS model were considerably higher than at each of the ground-based monitoring sites. The main contribution of DPM for each site was construction, and with this source deducted from the totals, the ambient concentrations matched more reasonably to observation, although not perfectly.

Therefore, we concluded that construction may be the cause of the most model error for the following reasons. Construction emissions contributions were much higher than all the other sources. It is not feasible for there to be extreme improvement upon the other

sources because the other sources (i.e., traffic, rail, marine, NRMs) have better inventories, emission factors, and activity data representations. As well as the differences seen in the normal traffic contribution and minimum in COVID-19 are not large enough to conclude the disparity between modelled versus observed traffic emissions. Considering that monitoring has taken place before and during the pandemic shutdown shows that the model does not perform well enough to predict DPM for a variety of scenarios.

Potential reasons that the expected emissions are so much higher is that models assume engines to be running 24 hours a day, as well as the same emissions all throughout the construction site lifetime. In real site operations there are more limited working hours, and emissions largely occur during the earthworks phase (Heidari & Marr, 2015). Despite construction not showing to be a major contributing factor based on long-term sampling, the source is not negligible to adverse health effects. Construction near residents has been shown to cause elevated concentrations as seen in figure 3 from the construction sampling performed. For short term acute exposure, we need to be weary of emissions. As well as for a global standpoint of emissions for climate warming impacts (Kassambara & Mundt, 2019).

Ground-based monitoring has been able to prove useful in finding sources with disparities in models. Using locally relevant data inputs and using measurement proxies is a useful way of ground-truthing models for DPM. For future modeling, these methods can be used individually to assess each source for potential improvements. Such checks include emission factors, and activity rate assessments, and ground-based monitoring to

compare rough values of fmass and total DPM model results. Models and especially predictive models need to be compared to measured concentrations, continuing work in comparisons and thus model improvements will continue to grow our understanding of urban air pollution exposure and highlight gaps in knowledge.

It is recommended for future dispersion modelers to create clear documentation of data inputs, to ensure future comparisons. Clearer data inputs could also allow for more citizen participation in the future. When a community knows what is going into a model, there is more opportunity for involvement and citizen data collection. Which when communities are involved with creating and solving problems, sustainable and long-lasting solutions are more likely to be reached, that benefit the community, and ideally avoiding unintentional consequences (Wolch, Byrne, & Newell, 2014).

6.2 Use as a Framework

The above methods can be used as a framework for assessing diesel emissions for other regions and cities in the United States or if similar resources are available. Collecting data around hard to measure sources can show relative impact from a source in an area, such as how construction was monitored. Collecting a sample of emission factors from a subset of fleet can help to assess whether input emission factors are way off from the fleet that is in-use in a region. As well as using high resolution AIS data for tracking ship movements is a great tool, especially for port cities. Then to ground-truth models and give EJ communities an extra layer of knowledge, monitoring using in-situ measurements that are descriptive of diesel emissions. Each of these data sources are free and available to the public, making this framework accessible.

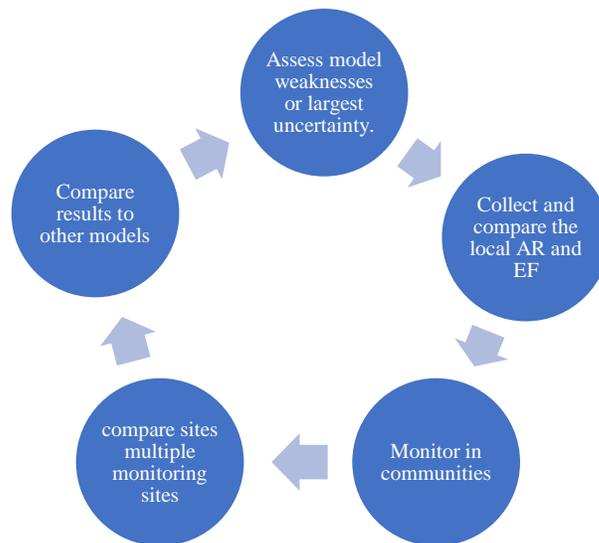


Figure 6.2-1. Flow chart of proposed framework steps

Figure 6.2-1 shows the framework as a flow chart. First, atmospheric dispersion modelers should assess what sources included the most uncertainty with the largest potential for exposure based on previous models. Then, data collection through sources identified in this study should be collected for AR and targeted source emissions testing if possible. Then, monitor ffmass and bbmass in multiple communities that are suspected to have a disproportionate vulnerability to diesel exposure. Compare patterns in each monitoring site, expected sources and other easy to run models.

The above framework for assessing diesel emissions is useful to speed up the process of research and study design development, thus making diesel model ground-truthing more accessible and efficient. This is an iterative process, and results will feed into future models that will need similar validations.

References

- Ban-Weiss, G. A., Lunden, M. M., Kirchstetter, T. W., & Harley, R. A. (2009). Measurement of black carbon and particle number emission factors from individual heavy-duty trucks. *Environmental Science and Technology*, 43(5), 1419–1424. <https://doi.org/10.1021/es8021039>
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., Deangelo, B. J., ... Zender, C. S. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research Atmospheres*, 118(11), 5380–5552. <https://doi.org/10.1002/jgrd.50171>
- Brown, I. N., & Aldridge, M. F. (2019). Power models and average ship parameter effects on marine emissions inventories. *Journal of the Air and Waste Management Association*, 69(6), 752–763. <https://doi.org/10.1080/10962247.2019.1580229>
- Coglianesi, C., & Nash, J. (2017). The Law of the Test: Performance-Based Regulation and Diesel Emissions Control. *Yale Journal on Regulation*, 34(1), 2.
- Dallmann, T. R., Kirchstetter, T. W., Demartini, S. J., & Harley, R. A. (2013). Quantifying on-road emissions from gasoline-powered motor vehicles: Accounting for the presence of medium- and heavy-duty diesel trucks. *Environmental Science and Technology*, 47(23), 13873–13881. <https://doi.org/10.1021/es402875u>
- Dalmajer, E. S., Nord, C. L., & Astle, D. E. (2020). Statistical power for cluster analysis. *ArXiv*.
- De Souza, C. V., & Corrêa, S. M. (2016). Polycyclic aromatic hydrocarbons in diesel emission, diesel fuel and lubricant oil. *Fuel*, 185, 925–931. <https://doi.org/10.1016/j.fuel.2016.08.054>
- Downing, K. (2015). *The Concerns about Diesel Engine Exhaust Contact: (February)*, 1–8.
- EcoChem Analytics. (1998). *User's Guide for PAS2000, Edition 2.0*.
- Functions, E., Subcommittee, R. T., Toxicants, R., Isbn, C., Pdf, T., Press, N. A., ... Press, N. A. (1998). Assessment of Exposure-Response Functions for Rocket-Emission Toxicants. In *Assessment of Exposure-Response Functions for Rocket-Emission Toxicants*. <https://doi.org/10.17226/6205>
- Group, E. R. (2020). *Oregon Nonroad Diesel Equipment Survey and Emissions Inventory, Final Report*.

- Hedmer, M., Wierzbicka, A., Li, H., Albin, M., Tinnerberg, H., & Broberg, K. (2017). Diesel Exhaust Exposure Assessment Among Tunnel Construction Workers- Correlations Between Nitrogen Dioxide, Respirable Elemental Carbon, and Particle Number. *Annals of Work Exposures and Health*, 61(5), 539–553. <https://doi.org/10.1093/annweh/wxx024>
- Heidari, B., & Marr, L. C. (2015). Real-time emissions from construction equipment compared with model predictions. *Journal of the Air and Waste Management Association*, 65(2), 115–125. <https://doi.org/10.1080/10962247.2014.978485>
- Hill, I. for the E. U. of N. C. at C. (2016). *User's Guide for C-PORT Version 3.0 C-PORT: Community modeling system for near-PORT*.
- Hopke, P. K. (2016). Review of receptor modeling methods for source apportionment. *Journal of the Air and Waste Management Association*, 66(3), 237–259. <https://doi.org/10.1080/10962247.2016.1140693>
- Isakov, V., & Venkatram, A. (2006). Resolving neighborhood scale in air toxics modeling: A case study in Wilmington, CA. *Journal of the Air and Waste Management Association*, 56(5), 559–568. <https://doi.org/10.1080/10473289.2006.10464473>
- Johnson, G. R., Jayaratne, E. R., Lau, J., Thomas, V., Juwono, A. M., Kitchen, B., & Morawska, L. (2013a). Remote measurement of diesel locomotive emission factors and particle size distributions. *Atmospheric Environment*, 81, 148–157. <https://doi.org/10.1016/j.atmosenv.2013.09.019>
- Johnson, G. R., Jayaratne, E. R., Lau, J., Thomas, V., Juwono, A. M., Kitchen, B., & Morawska, L. (2013b). Remote measurement of diesel locomotive emission factors and particle size distributions. *Atmospheric Environment*, 81, 148–157. <https://doi.org/10.1016/j.atmosenv.2013.09.019>
- Kassambara, A., & Mundt, F. (2019). *Package 'factoextra.'* Retrieved from <http://www.sthda.com/english/rpkgs/factoextra>
- Lewtas, J. (2007). Air pollution combustion emissions: Characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects. *Mutation Research - Reviews in Mutation Research*, 636(1–3), 95–133. <https://doi.org/10.1016/j.mrrev.2007.08.003>
- Magee Scientific Inc. (2016). *Aethalometer® Model AE33 User Manual, March 2016, ver 1.54.* (March).
- Miller, J., & Jin, L. (2019). *Global progress toward soot-free diesel vehicles in 2019.* 1–27. Retrieved from <https://theicct.org/publications/global-progress-toward-soot-free-diesel-vehicles-2019#:~:text=A 75%25 reduction in global,2020 to 2025 time frame.>

- Oksanen, J., F. Guillaume Blanchet, R. K., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., ... M. Henry H. Stevens, H. W. (2019). Package 'vegan.' *R Package Version 3.4.0*.
- Oregon Department of Environmental Quality. (2012). *Portland Air Toxics Solutions Committee Report and Recommendations*. (April), 1–12. Retrieved from <http://www.deq.state.or.us/aq/planning/patsReport.htm>
- Oregon Department. of Environmental Quality DEQ Facility Profiler-Lite Permitted Sites
Sitermitted Sites
- Oregon Department of Environmental Quality. (2018). *Oregon Air Toxics Monitoring Summary*.
- Oregon Department of Transportation. (2014). *Oregon State Rail Plan, An Element of the Oregon Transportation Plan*.
- Rao, M., George, L. A., Rosenstiel, T. N., Shandas, V., & Dinno, A. (2014). Assessing the relationship among urban trees, nitrogen dioxide, and respiratory health. *Environmental Pollution*, 194, 96–104. <https://doi.org/10.1016/j.envpol.2014.07.011>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ravindra, K., Sokhi, R., & Van Grieken, R. (2008). Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmospheric Environment*, Vol. 42, pp. 2895–2921. <https://doi.org/10.1016/j.atmosenv.2007.12.010>
- Reşitoğlu, I. A., Altinişik, K., & Keskin, A. (2015). The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems. *Clean Technologies and Environmental Policy*, 17(1), 15–27. <https://doi.org/10.1007/s10098-014-0793-9>
- Rivas, I., Beddows, D. C. S., Amato, F., Green, D. C., Järvi, L., Hueglin, C., ... Kelly, F. J. (2020). Source apportionment of particle number size distribution in urban background and traffic stations in four European cities. *Environment International*, 135(November 2019), 105345. <https://doi.org/10.1016/j.envint.2019.105345>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rousseeuw, P., Struyf, A., Hubert, M., Studer, M., & Roudier, P. (2021). Package 'cluster' *R topics documented* :

- Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., ... Morawska, L. (2014). Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment. *Atmospheric Chemistry and Physics*, *14*(21), 11883–11892. <https://doi.org/10.5194/acp-14-11883-2014>
- Sandradewi, J., Prévôt, A. S. H., Szidat, S., Perron, N., Alfarra, M. R., Lanz, V. A., ... Baltensperger, U. R. S. (2008). Using aerosol light absorption measurements for the quantitative determination of wood burning and traffic emission contribution to particulate matter. *Environmental Science and Technology*, *42*(9), 3316–3323. <https://doi.org/10.1021/es702253m>
- Thrane, K. E. (1987). Ambient air concentrations of polycyclic aromatic hydrocarbons, fluoride, suspended particles and particulate carbon in areas near aluminum production plants. *Atmospheric Environment (1967)*, *21*(3), 617–628. [https://doi.org/10.1016/0004-6981\(87\)90044-8](https://doi.org/10.1016/0004-6981(87)90044-8)
- Thrane, K. E., & Gunderson, R. W. (1986). Source allocation of organic air pollutants by application of fuzzy c-varieties pattern recognition. *Analytica Chimica Acta*, *191*(C), 309–317. [https://doi.org/10.1016/S0003-2670\(00\)86317-5](https://doi.org/10.1016/S0003-2670(00)86317-5)
- US EPA (Environmental Protection Agency). (2014). *EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and*. 136. Retrieved from https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf
- US EPA (Environmental Protection Agency). (2014). NATA: National Air Toxics Assessment Results. Retrieved March 19, 2021, from <https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results>
- Wang, X., Wang, Y., Bai, Y., Wang, P., & Zhao, Y. (2019). An overview of physical and chemical features of diesel exhaust particles. *Journal of the Energy Institute*, *92*(6), 1864–1888. <https://doi.org/10.1016/j.joei.2018.11.006>
- Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities “just green enough.” *Landscape and Urban Planning*, *125*, 234–244. <https://doi.org/10.1016/j.landurbplan.2014.01.017>
- Tang, N. W., Apte, J. S., Martien, P. T., & Kirchstetter, T. W. (2015). Measurement of black carbon emissions from in-use diesel-electric passenger locomotives in California. *Atmospheric Environment*, *115*, 295–303. <https://doi.org/10.1016/j.atmosenv.2015.05.001>

Appendix A. Instrumentation

Portable Measurement System (PMS)

Emissions measurements were made using a portable measurement system (PMS) designed by the PSU STAR lab. The system consisted of a bike carrier/stroller (Burley D'Lite) with a wood board inside as a base on top of foam cushioning. Screw hooks were used around the instruments to secure with bungee cords and extra foam was placed between instruments if needed for shock absorption. Particulate analyzers used non-static tubing. The BC (Magee Scientific AE33) and the particle bound PAH (p-PAH) (EcoChem PAS2000) analyzers were connected using a T fitting from an omnidirectional inlet with a PM_{2.5} impactor. The Li-COR 850 CO₂ was the only gas analyzed and used Teflon tubing and a particle filter for the inlet. The system was powered by a Yeti1400 Goalzero Lithium battery. A summary of all the instruments used in the PMS is outlined in table 1.

Table 7. Instrumentation setup for Portable Pollution Monitoring

INSTRUMENT	MEASUREMENT	UNITS	TIME BASE
<i>MAGEE SCIENTIFIC AE33</i>	<i>BLACK CARBON</i>	<i>ng/m³</i>	<i>1 SEC</i>
<i>LI-COR 850</i>	<i>CO₂</i>	<i>μ MOL/MOLE (PPM)</i>	<i>1 SEC</i>
<i>ECOCEM PAS2000</i>	<i>P-PAH</i>	<i>ng/m³</i>	<i>6 SEC</i>
<i>TSI DUSTTRAK 8533</i>	<i>PM 2.5 & PM10</i>	<i>μg</i>	<i>1 SEC</i>
<i>MICROAETH MA300</i>	<i>BLACK CARBON</i>	<i>ng/m³</i>	<i>1 SEC</i>
	<i>LAT & LONG</i>		

Aethalometer BC Source Apportionment

For the community monitoring, additional measurements from the aethalometer were used. The instrument has seven wavelengths that beam onto a filter paper loaded with the sample air and then a separate spot with clean air. Wavelengths measure between 370 and 950 nm. The total BC is considered the light absorption at 880 nm. Aside from BC there is also brown carbon, which, is from biomass burning instead of fossil fuel sources which is absorbed best at 950 nm (Sandradewi et al., 2008, Magee Scientific Inc, 2016). Brown carbon is absorbed at the 450 nm while fossil fuel is absorbed at 950 nm.

The aethalometer uses the Sandradewi et al., (2008) method, with the below equation from the AE33 manual (Magee Scientific Inc, 2016) to calculate internally biomass burning percent of total BC. The portion of BC that is not BB is fossil fuel percent. If both proportions are multiplied by the total BC concentration, then a mass

concentration is obtained for fossil fuel mass (ffmass) and biomass burning mass (bbmass).

$$\frac{b_{abs}(470\text{ nm})_{ff}}{b_{abs}(950\text{ nm})_{ff}} = \left(\frac{470}{950}\right)^{-\alpha_{ff}} \quad ,$$

$$\frac{b_{abs}(470\text{ nm})_{bb}}{b_{abs}(950\text{ nm})_{bb}} = \left(\frac{470}{950}\right)^{-\alpha_{bb}} \quad ,$$

$$b_{abs}(470\text{ nm}) = b_{abs}(470\text{ nm})_{ff} + b_{abs}(470\text{ nm})_{bb} \quad ,$$

$$b_{abs}(950\text{ nm}) = b_{abs}(950\text{ nm})_{ff} + b_{abs}(950\text{ nm})_{bb} \quad ,$$

$$BC_{bb} = BB * BC \quad ,$$

$$BC_{ff} = (1 - BB) * BC \quad \quad BB(\%) = \frac{b_{abs}(950\text{ nm})_{bb}}{b_{abs}(950\text{ nm})}$$

Figure A-1. Magee Scientific AE33 internal calculation of ffmass and bbmass.

Appendix B. Fuzzy C-Means Methods

Cluster validation methods

The most traditional way of finding an optimal number of clusters is with the Calinski method, but this only works if there are equal sample sizes within the cluster (Kassambara & Mundt, 2019). Therefore, the next best option is the simple structure index (SSI). However, through trial and error, manually finding the best model outputs agreed with the Calinski method.

Results are assessed based off the tightness of their clusters and against maximizing the distance between clusters (P. J. Rousseeuw, 1987). The silhouette width is most often used as a metric for assess cluster strength for individual clusters and total average. The silhouette width ranges from -1 to 1. It is considered to have a strong structure if greater than 0.71, reasonable if between 0.51 and 0.7, weak between 0.26 and 0.5, and no substantial structure less than 0.25 (P. J. Rousseeuw, 1987).

Cluster Analysis Workflow

The workflow for performing cluster analysis in R programming went as follows. Variables of interest were chosen if time averaging was used then this would be next. Outliers are removed by removing anything outside of 1.5 times the interquartile range method. Then rows with missingness were removed. Next, the data is scaled to a z-score

and centered to ensure comparability across differing units. The optimal number of clusters are found next using the `CascadeKM` function from the `vegan` package (Oksanen et al., 2019). The `fanny` function in the package from the `Cluster` package was used for its simplicity (P. Rousseeuw, Struyf, Hubert, Studer, & Roudier, 2021). A member exponent is the factor applied to the algorithm to make it fuzzy and 1.5 most typically used. This number ranges from 1 (completely crisp) to 2 (full fuzziness). It was found that using a slightly crispy cluster analysis provided the best results with a membership exponent of 1.2. Then, the number of clusters and data frame are specified in the `fanny` function.

After the `fanny` function is performed and output stored, silhouette widths are analyzed using a summary function and graphically using the `factoextra` package (Kassambara & Mundt, 2019). Then, all post-analysis visualizations were performed from `openair` (Kassambara & Mundt, 2019).