Excitation Emission Matrix Fluorescence Spectroscopy for analysis of Reactive Oxygen Species in Combustion Particulate Matter

Ishrat Singh

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Reading Committee:

Igor Novosselov, Chair

Christopher D. Simpson

John C. Kramlich

Program Authorized to Offer Degree:

Mechanical Engineering

University of Washington

Abstract

Excitation Emission Matrix Fluorescence Spectroscopy for analysis of Reactive Oxygen Species in Combustion Particulate Matter

Ishrat Singh

Chair of the Supervisory Committee:
Professor Igor Novosselov
Department of Mechanical Engineering

The presence of particulate matter (PM) in the environment can lead to adverse health impacts, including respiratory, cardiovascular, neurological, and lung cancer. Air pollution has been estimated to cause 4.9 million deaths, and 94% were caused by PM. Sources of combustion-generated PM range include wildfires, residential wood burning, traffic emissions, etc. While the epidemiological link between PM exposure and adverse health effects is clear, there is a lack of information regarding source-specific differences in PM toxicity. Thus, there is a clear need to quantify PM presence in the environment and identify its sources and toxicity. Reactive oxygen species (ROS) have been proposed as one surrogate metric for the toxicity of PM. Excitation emission matrix (EEM) spectroscopy has been well documented as a low-cost, reliable method for analyzing the organic fraction of PM and can be used in source apportionment. In this study, we

investigate the correlation between EEM signature and ROS measurements in PM. PM collected from laboratory flame cookstove (natural and forced draft), and wildfire smoke (collected indoor and outdoor) are analyzed by EEM and the dithiothreitol (DTT) assay for ROS. While total integrated fluorescence of EEM does not provide clear patterns of association with ROS, with the implementation of principal component analysis (PCA) and regression of the EEM spectra, we show that specific EEM patterns can be correlated with the ROS measurement. EEM-PCA may be a useful alternative to evaluate the ROS level in combustion-generated aerosols.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Professor Igor Novosselov for providing me with the opportunity, funds, and mentoring me during the course of this thesis. It would not have been possible without his endless support, both as a mentor and a friend.

I would also like to thank the members of my thesis committee, Professor Chris Simpson and Professor John Kramlich for their time and evaluation of my work.

This would not have been possible without Jiayang He, who has helped, taught, motivated, and guided me towards the final results and has been with me from the start. Thank you for all your time and support.

Lastly, I would thank my parents, my sister, and my friends who have always been there for me and pushed me throughout my MS program. For that, I am eternally grateful.

Chapter 1. INTRODUCTION

Environmental and occupational exposure to particulate matter (PM) is linked to adverse health impacts, including respiratory disease, cardiovascular diseases, neurological diseases, and lung cancer. Air pollution was estimated to be responsible for 4.9 million deaths and 147 million disability-adjusted life years in 2017 [1]. The PM arises from a variety of sources, including agricultural processes, traffic, industrial sources, forest fires, and other natural and anthropogenic sources [2]. Combustion-generated PM accounts for around 45-60% of urban $PM_{2.5}$ (particles with a size less than 2.5 μm) and 40-58% of urban PM_{10} (particles with a size less than 10 μm) [3]. $PM_{2.5}$ has been associated with adverse health impacts, as these particles can penetrate deep within the respiratory system and, in some cases, translocate intact into the bloodstream. Growing evidence from in vitro studies, animal models, and human studies shows that PM exposure can induce oxidative stress, offering one potential mechanism of PM toxicity [4-14]. Several studies have suggested that quinones present in PM can act as catalysts to produce Reactive Oxygen Species (ROS) directly and may be critical compounds in the PM-based oxidative stress [15-18]. ROS are any oxygen-containing molecules that have one or more unpaired electrons, making them highly reactive, and include species like hydrogen peroxide (H_2O_2), superoxide radical (O_2^-), and hydroxyl radical ($^{\bullet}OH$) [18]. Although there is a strong epidemiological link between PM exposure and adverse health effects, there is only limited evidence of differential toxicity of source-specific PM. Several authors commented that the evidence of source-specific PM's health impact is insufficient to warrant regulation of specific types of PM air pollution more than others [19-22].

Quantitative measurement of the capacity of PM to generate ROS has been proposed as a surrogate metric for the PM toxicity [23]. Recent epidemiological studies demonstrated that the

ability of particles to consume antioxidants and/or generate ROS, referred to as oxidative potential (OP), can be more strongly associated with the reduction in respiratory function compared to PM mass, suggesting that OP of PM is a health-relevant metric for air quality [24-26]. The dithiothreitol (DTT) assay is an acellular technique that has been used to quantify particle OP by monitoring the consumption of DTT, a surrogate for biological antioxidants, in the presence of particle components under physiologically relevant conditions [27, 28].

Wildfires have become more frequent and intense in the Western United States. Summer wildfire seasons are 40 to 80 days longer on average than 30 years ago [29]. The intensified wildfires will release more smoke into the atmosphere [30], traveling significant distances [31]. Fine particulate matter ($PM_{2.5}$), a major pollutant in smoke from wildfires, can travel deep into the respiratory tract [32]. The combustion-generated aerosols consist of elemental and organic carbon fractions, which may be more toxic than other PM2.5 sources and have long-lasting impacts on the health [33-36]. Complex flow structures associated with large-scale flames and low flame temperature in biomass burning lead to low carbonization of organic carbon, thus -- high levels of potentially carcinogenic polycyclic aromatic compounds [2, 37-39]. Exposure to $PM_{2.5}$, particularly combustion-generated aerosols, has been linked to adverse respiratory and cardiovascular health effects, including ischemic heart disease, stroke, cardiovascular mortality, and exacerbations of asthma and chronic obstructive pulmonary disease [40-44]. More recently, wildfire smoke exposures have been linked to higher severity and mortality of SARS-CoV-2 [45-48]. A series of large wildfires impacted air quality in western regions of the United States in 2020. The episode measured in this study (2020 Washington Labor Day fires) in September 2020. The fires burned over 41,000 acres of the forest [49].

Recent advancements in low-cost particulate matter (PM) sensors led to their extensive use in various applications, such as air quality (AQ) monitoring in indoor [50-53] and outdoor [54-58] environments, large-scale deployments in [59-62] including wildfire smoke monitoring [63-66]. Typical optical PM sensors rely on elastic light scattering providing size-resolved PM concentrations in the 0.3 – 10.0 μm range and requiring calibration against reference instruments, e.g., [67]. The scattering light intensity depends on particle size, morphology, complex index of refraction (CRI), and sensor geometry. [68, 69] CRI sensitivity can be addressed by optimizing the design to simultaneously measure scattered light at multiple angles or by employing dual-wavelength techniques. [70, 71] However, these solutions do not address the analysis of aerosol chemical composition. Some recent approaches suggested using EEM spectroscopy for sensor design, which can potentially analyze PM composition in a compact form factor [72].

Fluorescence spectroscopy provides information on the composition of PM by measuring fluorescent emissions from the many chemicals present in combustion-derived PM, such as PAHs [73]. EEM has been applied to the analysis of environmental water samples as well as atmospheric PM samples [74-80]. EEM provides complex spectral information consisting of thousands of wavelength-dependent fluorescence intensities. Various approaches have been used to interpret EEM spectra [81, 82]. For example, EEM and principal component regression (PCR) can be used to classify the PAH content in PM [83]. PCR has also been applied to EEM data for analysis of ternary mixtures of organic acids [84] and the PAH content of laboratory PM samples [2]. PCR is a two-step process in which predictors are transformed into a lower-dimensional representation by principal component analysis (PCA), and linear regression is performed on a chosen number of the principal components (PCs) [85]. The number of PCs used is much smaller than the number of original predictors and is selected to capture nearly all the variance in the data, giving the best

possible model fit and avoiding overfitting. This work investigates the correlation between EEM signatures and ROS measurements for biomass combustion aerosols. PM samples collected from two cookstove designs and wildfire smoke from the 2020 Washington Labor Day wildfire event were analyzed by EEM and the DTT assay for ROS.

Chapter 2. METHODS

2.1 Sample Collection and Preparation

We collected 18 samples for four types of biomass combustion sources: (a) 6 for cookstove smoke with a natural draft, (b) 6 for cookstove smoke with a forced draft design, (c) 3 wildfire samples collected outdoors, and (d) 3 wildfire samples collected indoors. These samples were later defined in three sub-categories; cookstove (natural-draft), cookstove (forced-draft), and wildfire smoke samples. Cookstove samples were generated by burning 1½ by ¾ inch Douglas fir sticks cut from dimensional lumber in side-feed, natural-draft [86], or forced-draft cookstoves [87]. The cookstoves were located inside a fume hood; the sampling devices were placed in a sealed chamber connected to the exhaust duct of the fume hood. The natural-draft and forced-draft cookstove samples were collected as replicates of each other using the same conditions respectively. The variation between them was mainly due to the variation in flow rates of air pumps used in the study. Wildfire smoke samples were collected during the 2020 wildfire season in Seattle, WA. The Harvard impactors collected triplicates of samples at indoor (NanoES building, University of Washington) and outdoor (in the location near UW campus) locations over a 3-day period. The indoor location did not have an active HEPA filtration system [15]. The two types of wildfire samples were considered as one sub-category since the wildfires are large-scale events so the sample composition does not change over short locations.

PM was collected using 1.0 µm pore PTFE membrane filters (Fluoropore® membrane filters, # FALP03700, 37 mm, MilliporeSigma, Burlington, MA, USA) housed in the Harvard impactor (Cat. # HP2518, BGI, Butler, NJ, USA) sampling cassettes placed in the sealed chamber. The sampling flow rate was set to 1.8 Lpm using portable sampling pumps (AirChek XR5000

pump, SKC Inc., Eighty-Four, PA, USA). Flowrates were verified using a flow calibrator (Gilian Gilibrator PN# 800268, Sensidyne, St. Petersburg, FL, USA).

Following collection, filters are removed from the samplers and placed in a chamber with 37% (SD = 4%) relative humidity for 24 hours. The filters were then weighed using a micro-balance with 0.5 µg resolution (Mettler-Toledo UMT-2, Greifensee, Switzerland). Initial weights of the filters were recorded in the same manner, and the difference in weights was used to calculate the mass of $PM_{2.5}$ in each experiment. Weight change for the filters was divided by the air volume sampled to calculate the PM mass concentration in the original air samples. Filters were then submerged in 7 mL methanol (Sigma Aldrich, HPLC grade) and sonicated in an ultrasonic water bath for 60 minutes. Following sonication, the filters were removed from the extracts and discarded. We prepared a filtered extract and an unfiltered extract for each sample as follows: For the filtered extract, 3.5 m of the extract was pushed through a 0.2 µm syringe filter (VWR Cat. #28145-491) into a 4 mL glass vial to remove insoluble particles. The remaining portion of the solution was transferred directly into another 4 mL glass vial. PM extracts were stored in the vials until analysis (see Figure 2.1).

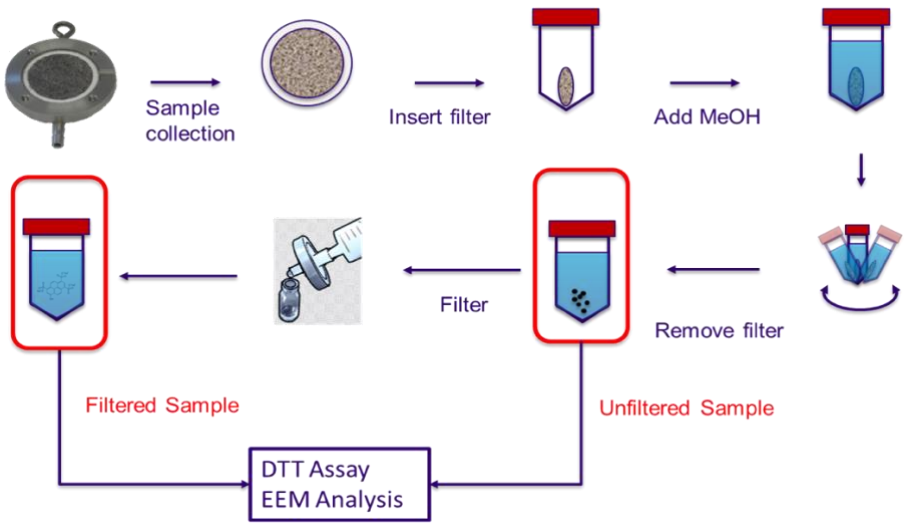

Figure 2.1. Sample preparation for the EEM analysis and DTT assay.

2.2 EXCITATION EMISSION MATRIX (EEM) ANALYSIS

EEM data are 2-dimensional matrices where the x-axis is the excitation wavelength and the y-axis is the emission wavelength with fluorescence intensity data for each excitation-emission wavelength, as shown in Figure 2.2. For EEM spectroscopy, ~ 3mL of the filtered and unfiltered PM extracts were transferred to a 1 cm x 1 cm quartz cuvette (Item # CV10Q3500FS, Thorlabs Inc., Newton, New Jersey). The extracts were analyzed using a fluorescence spectrofluorometer with an extended-UV 150W xenon-arc lamp (Aqualog-880-C, HORIBA Instruments Inc. Edison, NJ, USA). Samples were exposed to the excitation wavelengths from 200 to 600 nm with increments of 2 nm. Emission spectra were recorded from 246 to 826 nm with 5nm increments. The instrument's raw fluorescent signal is corrected for detector response and lamp intensity [88] and normalized to Raman Units [89]. Raman unit is a method for normalizing the raw fluorescent spectra of each sample collected from EEM [89]. Daily solvent blanks are recorded and used for blank subtraction to minimize the effect of Rayleigh and Raman scattering. To further reduce the effects of Rayleigh scatter values within 10 nm of the first and second-order Rayleigh scattering bands were excised, followed by replacing the excised values using a 2-dimensional interpolation [90]. The processed EEMs were then passed through a Gaussian filter (sigma=2), and negative values were removed numerically in Python 3.6.0 to smooth the data (see Figure 2.3) [81].

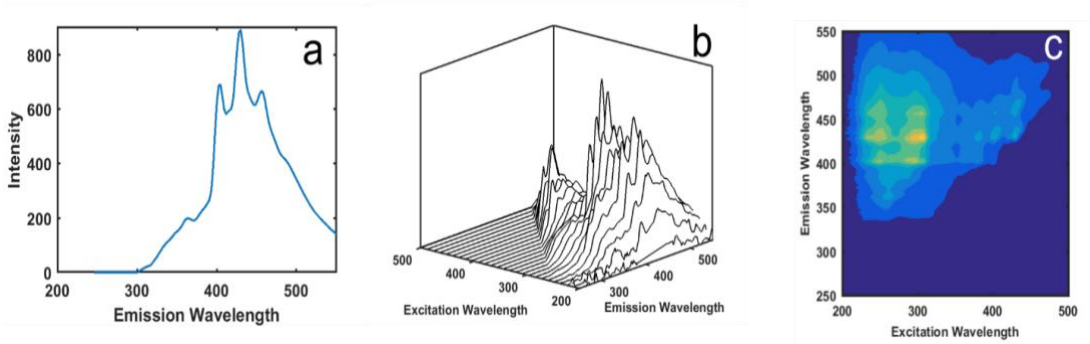

Figure 2.2. EEM spectroscopy, (a) Individual fluorescence emission spectra are collected, (b) the excitation wavelength is incremented, and additional spectra are collected, and (c) The EEM fingerprint is represented using a filled contour plot.

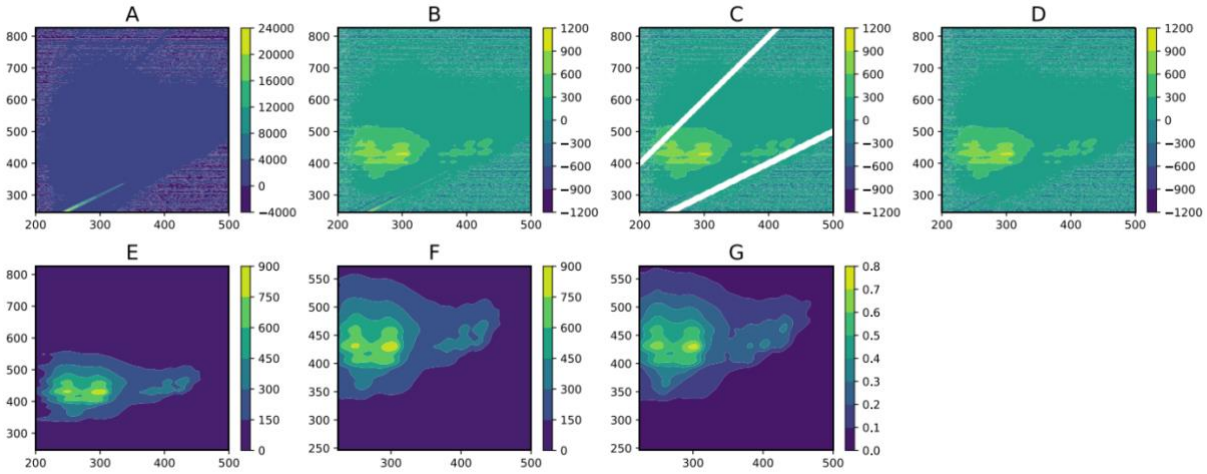

Figure 2.3. EEM of wood smoke extract showing data processing steps: Starting with (a) the raw data, (b) the solvent blank is subtracted, (c) then values are excised for scattering removal, (d) the excised values are interpolated, (e) negative values are replaced with zero, and 2D-Gaussian smoothing is applied, (f) the EEM is cropped and (g) normalized to Raman units.

2.3 Reactive Oxygen Species Analysis

PM samples collected on PTFE filters are also analyzed for DTT activity. The DTT assay measures the presence or formation of ROS via the formation of the DTT-disulfide [5]. Unreacted DTT is detected colorimetrically after reaction with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB),

producing 5-mercapto-2-nitrobenzoic acid. The rate of disappearance of DTT is proportional to the oxidant activity. The assay uses 1,8-phenanthraquinone as a positive control and methanol extracts of blank filters as a negative control.

We adapted the procedure of Li et al. [5] for use with a microplate reader to improve sample throughput, sensitivity, and precision. Particle-containing and particle-free extracts were prepared for DTT analysis as described above. The samples were well mixed, and 150 µL aliquots, including solvent and suspended particles, were removed for analysis. The remaining portion of each extract was concentrated to 2 mL under a flow of nitrogen at 50°C in a Turbovap evaporative concentrator, vortexed, and a second 150 µL aliquot was removed for possible analysis if the unconcentrated extract showed low activity. The filtered extracts were reduced to dryness and reconstituted in 150 µL of methanol. After vortexing, water and phosphate buffer were also added to the samples. To measure DTT reactivity, the extracts were then incubated in the presence of DTT in a 96-well plate. At designated time points (0, 10, 20, 30, 40, and 50 minutes) aliquots of the reaction mixture were withdrawn and added to microplate wells containing tris HCl in 20 mM EDTA and DTNB (5,5'-dithiobis-(2-nitrobenzoic acid)). Absorption at 412 nm was recorded. The rate of DTT consumption is calculated from a plot of absorbance vs. time, and this value is corrected for atmospheric oxidation of DTT calculated from a blank filter extract time series run with each microplate.

2.4 Principal Component Regression Analysis

The PCR analysis was done in three different analyses. As we move deep into each analysis in the results section, we would discuss the reasons behind each analysis. The overview of the three analyses is given below:

1. Using all 18 samples and training 816 models, that can be possible by different combinations of splitting 18 samples into 15:3 ratio. Then we selected the models which have R^2 (test > 0.8 and used the average of coefficients and intercepts of these models to get a general model to predict the DTT value. This cross-validation technique removes the bias that may arise since some samples would not be in the training set if taken individually.
2. Training the models using the sub-categories of samples; cookstove (natural-draft), cookstove (forced-draft), and wildfire samples. Since the number of samples in the training set for PCR is reduced to 6, we do not use the cross-validation technique for this step.
3. Using 12 cookstove samples (natural and forced-draft) and training 220 models that can be possible by different combinations of splitting 12 samples into a 9:3 ratio. Then we selected the models which have R^2 (test) > 0.8 to find a superior correlation.

PCR was performed as a two-step process: (1) transforming the EEM data onto its principal components (PCs) and (2) fitting regression models on the transformed EEM data and DTT consumption rates from the ROS analysis. EEM spectra were randomly split into training spectra and test spectra to train the models. We choose 3 samples every time for the test set. The reason is (i) a minimum number of samples in the test set to test the model is 3 and (ii) we have a small dataset so we need a maximum number of samples in the training set too. The EEM data contains ~ 201,000 fluorescent values (predictors) in each data matrix. Linear regression could be performed using all values in the EEM independently but would result in overfitting. Instead, linear regression is performed on the PC representation of the data. The number of PCs used is smaller

than the number of original predictors. It is selected to capture nearly all the variance in the data, give the best possible model fit, and avoid overfitting.

We did three different analyses with a different number of samples in the training set. Here we describe the overview of how PCR was conducted for all three analyses. In the first step of PCR, each EEM 2D matrix (1000 x 201) was unfolded into a 1D row vector (1 x 20100). The EEM row vectors of the training samples (n) were stacked to create a training data matrix C (n x 201000). The data matrix C was reduced with principal component analysis (PCA) using the sci-kit-learn library in Python 3.6.0 [91]. The number of PCs (d), required to represent the data in the data matrix C was varied from 1 to n. The data matrix C can be represented as the matrix product of s and v and a residual matrix E, shown in (2.1).

$$C = sv + E \tag{2.1}$$

where s (n x d) is the scores matrix and v (d x 201000) is the loadings matrix having d PC vectors. The scores matrix s is the reduced form of data matrix C; the operation reduces the number of data points required to represent each EEM from 20100 to d. The value of d is chosen such that the number of PCs that account for the maximum variance in the EEM dataset is minimized.

For the second step of PCR, the scores matrix s (n x d) is used to fit linear models for estimating concentrations of ROS. We used the ordinary least squares method in the sci-kit-learn library in Python 3.6.0 to perform multiple linear regression (MLR) to fit the linear model between the EEM scores matrix and the DTT consumption rates from ROS analysis (i) for unfiltered data and the the the (ii) filtered data, for each analysis. The mean squared error (MSE), representing the error in correlation for each model, was calculated using the "leave one out cross-validation" method (LOOCV) [91].

The models for estimating ROS concentrations trained on the smoke samples were tested using the three random samples from the 18 samples. The EEM row vectors of the test samples were stacked to create a test data matrix C_{test} (3 x 201000), where 3 is the total number of test samples. The test data matrix C_{test} was reduced along the PCs obtained for training samples to obtain a test scores matrix (3 x d). The coefficient of determination (R^2) for both models is reported.

Chapter 3. RESULTS AND DISCUSSION

3.1 EEM ANALYSIS

PM extracts from natural-draft cookstove smoke, forced-draft cookstove smoke, and wildfire smoke show unique EEM spectra (Figure 3.1). Natural-draft cookstove smoke has spectra consisting of four peaks in the region λ_{ex} = 225 – 275 nm, λ_{em} = 400 – 475 nm. Some fluorescence can also be observed in the region λ_{ex} = 350 – 450 nm, λ_{em} = 400 – 475 nm. Forced-draft cookstove smoke has distinctly different spectra consisting of two peaks at λ_{ex} 350 – 450 nm, ~ λ_{em} 350 nm. The indoor and outdoor wildfire smoke spectra are similar, with a single primary peak at λ_{em} = 350 – 400 nm and some fluorescence surrounding the peak. Natural-draft cookstove smoke shows fluorescence over the broadest region and generally at higher emission wavelengths than forced-draft cookstove smoke and wildfire smoke. The spectra from the different sources have overlapping regions, so there can be some difficulty in identifying individual sources from mixed samples.

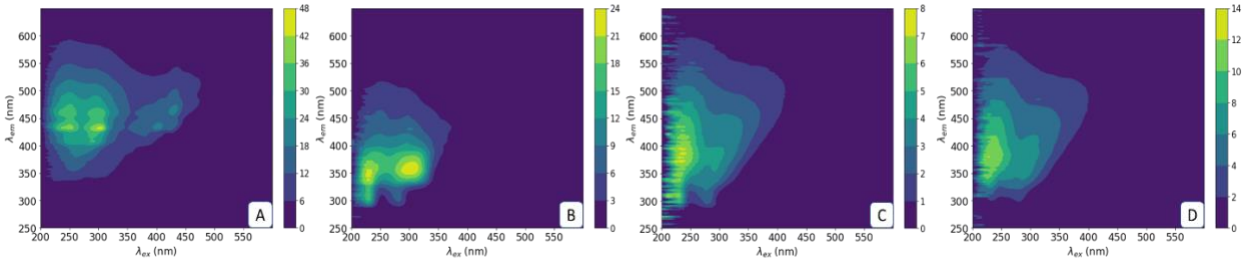

Figure 3.1. Representative EEM spectra for (A) natural draft cookstove, (B) forced draft cookstove, and 2020 Washington wildfire (C) indoor and (D) outdoor samples. The EEM spectra are from the extract in methanol. The spectra were normalized to Raman units (R.U.)

We also compared the spectra from filtered and unfiltered samples to see how insoluble material affects EEM spectra. Figure 3.2 shows the EEM spectra of a filtered natural-draft cookstove smoke sample and the spectrum of an unfiltered sample. The spectra have primary peaks

at identical locations. The peak fluorescence intensity of the unfiltered extract is three times higher than the filtered extract from the same sample. No notable difference in fluorescence contour can be observed between these two spectra.

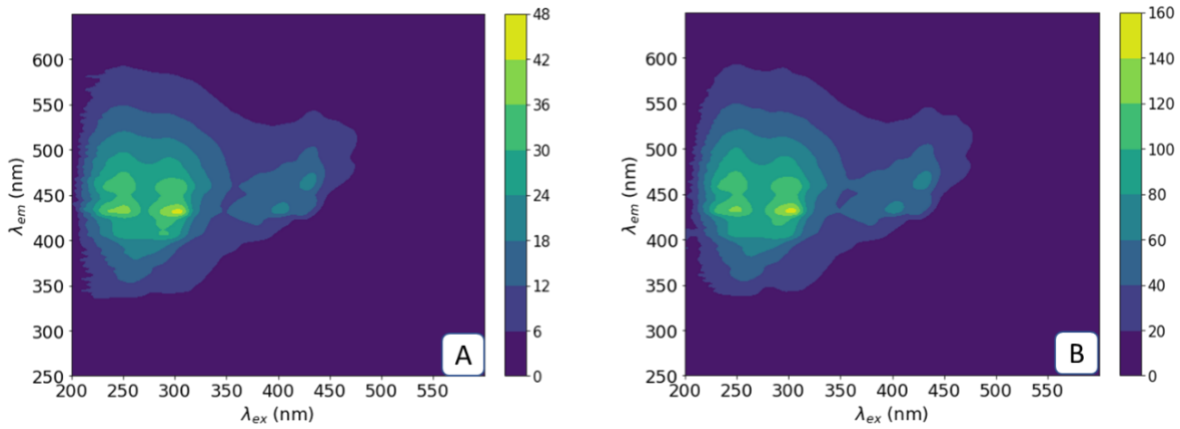

Figure 3.2. EEM spectra of natural-draft cookstove smoke from (a) a filtered extract, and (b) an unfiltered extract. The EEM spectra are from the extract in methanol. The spectra were normalized to Raman units (R.U.)

3.2 ROS Analysis

We measured the oxidative potential of both filtered samples and unfiltered samples using the DTT assay. Sample mass and DTT consumption rate for the smoke samples are summarized in Table 3.1. The highest DTT consumption rate per mass of collected PM was found in the natural draft cookstove samples. Visually the PM samples collected on filters were darker in color, indicating a higher fraction of elemental carbon. DTT consumption rate for the unfiltered natural-draft cookstove smoke samples is higher than the filtered samples. This indicates that for these samples, ROS is associated with MeOH extract and the PM-bound compounds (not extracted by MeOH).

Table 3.1. Summary of sample mass and DTT consumption rate for the smoke samples

Sample Name	Sample Mass (mg)	DTT Consumption (ng/min/ug)	
		Unfiltered	Filtered
Natural-draft Cookstove Smoke	0.11	1.8	2.12
Natural-draft Cookstove Smoke	0.135	1.64	1.73
Natural-draft Cookstove Smoke	0.151	2.01	1.21
Natural-draft Cookstove Smoke	0.118	2.8	1.88
Natural-draft Cookstove Smoke	0.146	2.1	1.78
Natural-draft Cookstove Smoke	0.157	0.94	0.9
Wildfire Smoke	0.325	4.76	5.04
Wildfire Smoke	0.234	4.75	5.83
Wildfire Smoke	0.297	6.38	5.25
Wildfire Smoke	0.371	4.88	4.75
Wildfire Smoke	0.392	4.65	4.32
Wildfire Smoke	0.425	4.74	4.4
Forced-draft Cookstove Smoke	0.114	7.84	5.58
Forced-draft Cookstove Smoke	0.122	7.56	5.87
Forced-draft Cookstove Smoke	0.132	6.35	4.71
Forced-draft Cookstove Smoke	0.132	7.36	5.4
Forced-draft Cookstove Smoke	0.126	6.85	5.58
Forced-draft Cookstove Smoke	0.125	8.42	5.43

Table 3.2. The average DTT consumption rate for the samples sub-categories

Sample Name	DTT Consumption (ng/min/ug)	
	Unfiltered	Filtered
Natural-draft Cookstove Smoke	1.88	1.60
Wildfire Smoke	5.03	4.93
Forced-draft Cookstove Smoke	7.40	5.43

The DTT consumption rate per gram of PM in the forced-draft cookstove is significantly higher than the natural-draft cookstove and is similar to the DTT consumption rate for the wildfire samples as seen in Table 3.2. We have previously shown via GCMS analysis that the forced-draft

cookstove sample has a significantly lower concentration of PAH species than the natural-draft cookstove sample [73]. Interestingly the DTT consumption is similar in the filtered and unfiltered extracts for the wildfire samples and the natural-draft cookstove PM samples, suggesting that DTT consumption in these samples is dominated by chemicals that are soluble in the methanol solvent, possibly organic carbon. In contrast, there is a greater difference in DTT consumption between the filtered vs. non-filtered extract of the forced air cookstove – suggesting that a greater proportion of the DTT reactivity remains associated with methanol insoluble chemicals in the forced draft cookstove.

3.3 Total Integrated Fluorescence and ROS

To investigate the correlation between EEM spectra and ROS, we first analyzed the relation between the mass normalized integrated fluorescent intensity of the EEM and the DTT consumption rate for each sample. Integrated fluorescence was calculated by adding all the intensity values to a 2D EEM matrix. Figure 3.3 shows the integrated fluorescent intensity measurement per sample mass vs. DTT consumption rate for the filtered and unfiltered smoke samples. Samples from the same group cluster with the natural-draft cookstove smoke samples have higher integrated fluorescence intensity than others. We observed no strong correlation between integrated fluorescence intensity and DTT consumption rate for the filtered and unfiltered smoke samples. Total fluorescence is greater in the unfiltered samples compared to the filtered extracts for all three particle sources. These data show that less than half of the total fluorescence is extractable into methanol in contrast to DTT reactivity.

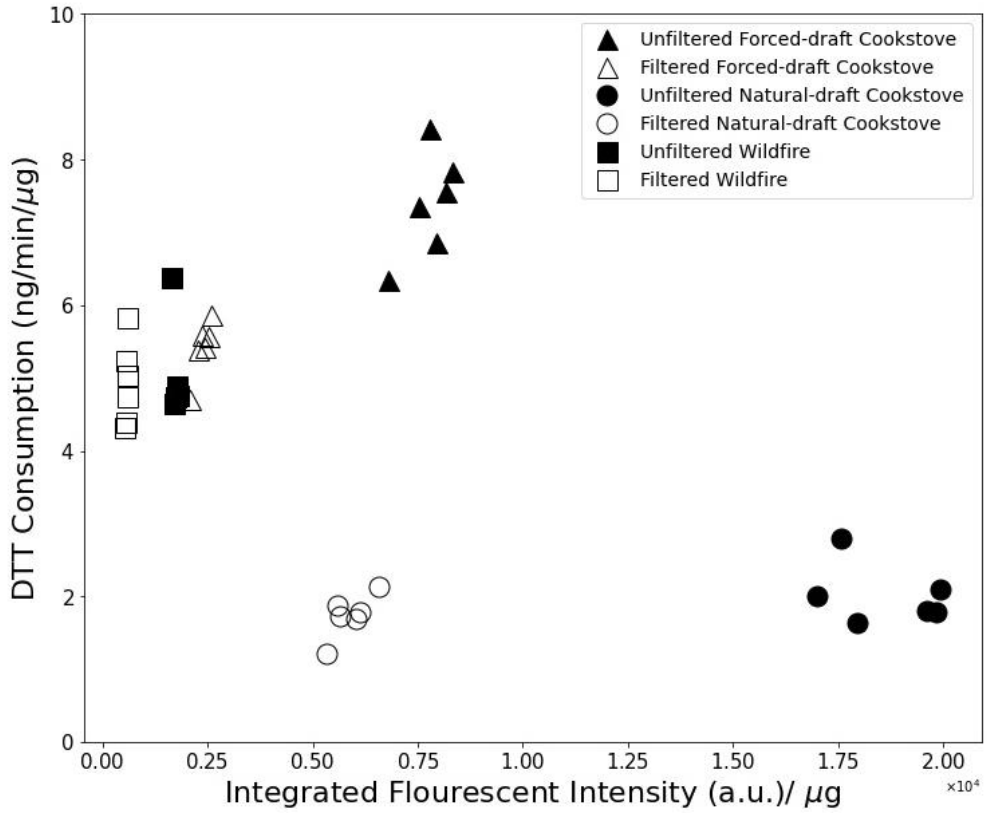

Figure 3.3. Mass normalized Integrated Fluorescent Intensity v/s DTT consumption rate for filtered and unfiltered samples. Filtered samples are shown in solid symbols

3.4 PCR ANALYSIS

The total integrated intensity plot shows a qualitative relationship between the sample sub-categories but not a clear correlation of the DTT with integrated intensity; hence, we turn to PCR analysis to find an underlying correlation. The PCR analysis was performed on the dataset to evaluate the correlation between the mass-normalized 2D EEM spectra and DTT consumption rates. We took three different approaches for the PCR analysis.

3.4.1 *First Analysis - PCR Analysis of 18 Samples*

PCA was performed on the 15 training spectra to generate the PCs. The first three PCs for the unfiltered and filtered smoke samples are shown in Figure 3.4. The first three PCs account for

over 98% of the variance in the dataset showing that although the original data contain over twenty-thousand fluorescent values, the spectra can be effectively represented in a lower-dimensional space.

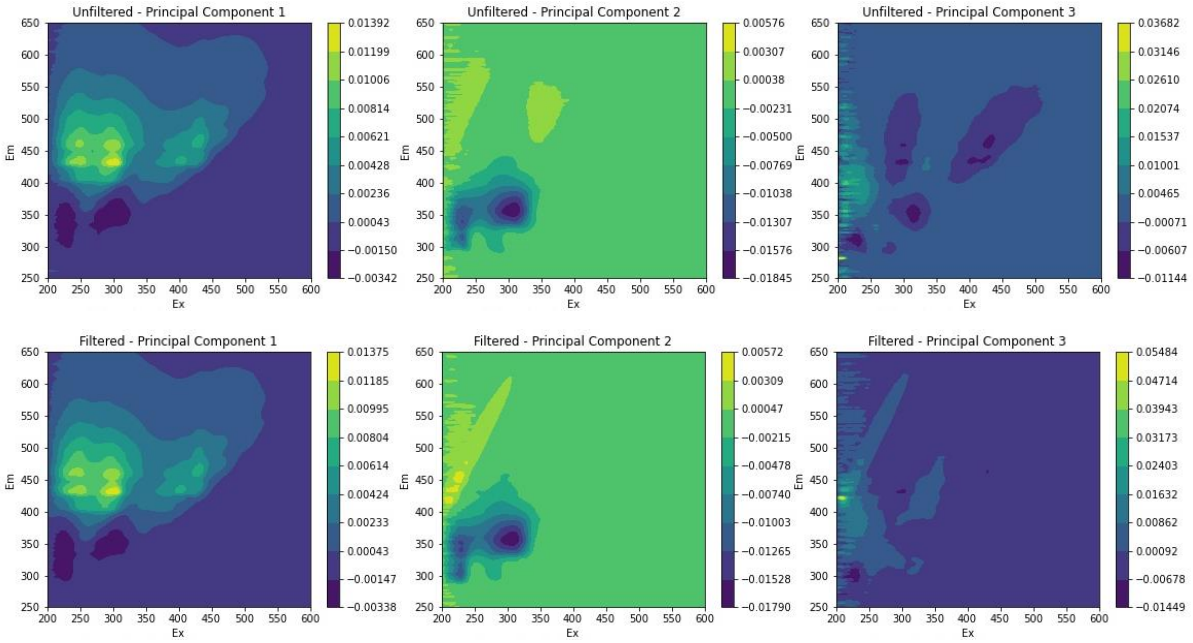

Figure 3.4. The first three principal components of the 15 training EEM spectra for filtered (top) and unfiltered (bottom) smoke samples. These three PCs shown here account for 99% of the variance in the dataset

Figure 3.5 plots the R^2 value using each of the 15 principal components of filtered and unfiltered samples calculated from PCR. It shows that we can effectively reduce the higher dimensional data to lower dimensions using PCA while retaining most of the information from the data. Based on the variance in the EEM data for $d = 1$ - 15 PCs and the R^2 values, $d = 3$ for unfiltered and filtered samples were chosen for fitting the linear models for DTT consumption rate. We ran a total of 816 different cases for all possible models that could be formed from the 18 samples in data; Figure 3.6 shows the R^2 (test) for all the 816 models.

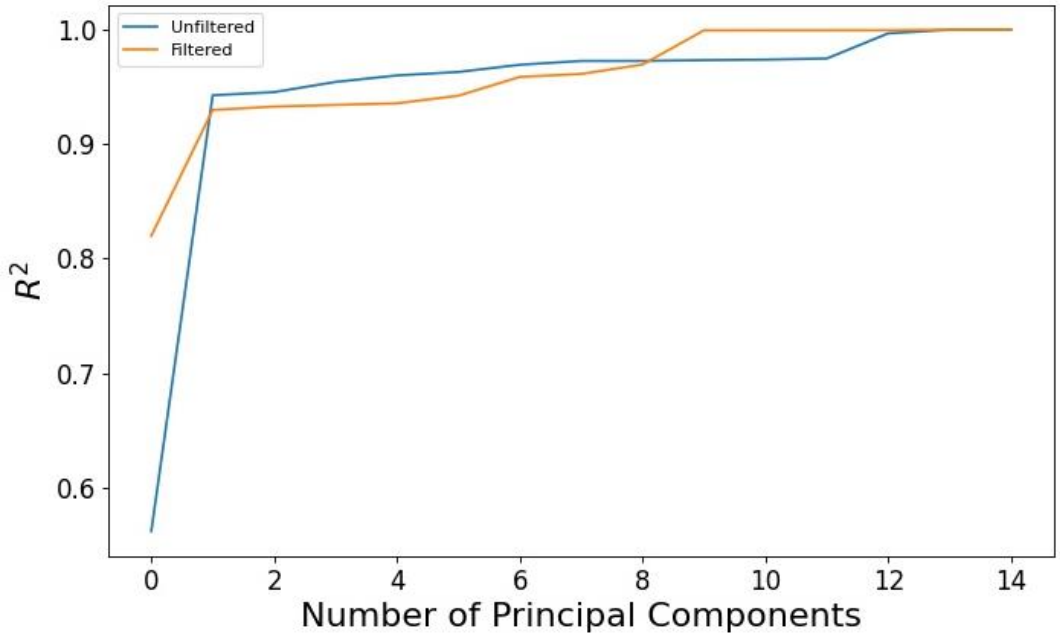

Figure 3.5. PCR R^2 vs. number of PCs used in PCR for DTT consumption rate. As the number of PCs is increased from one to fifteen, the fit to the training data continually improves. The fit to the filtered sample data is always better than the fit to the unfiltered sample data

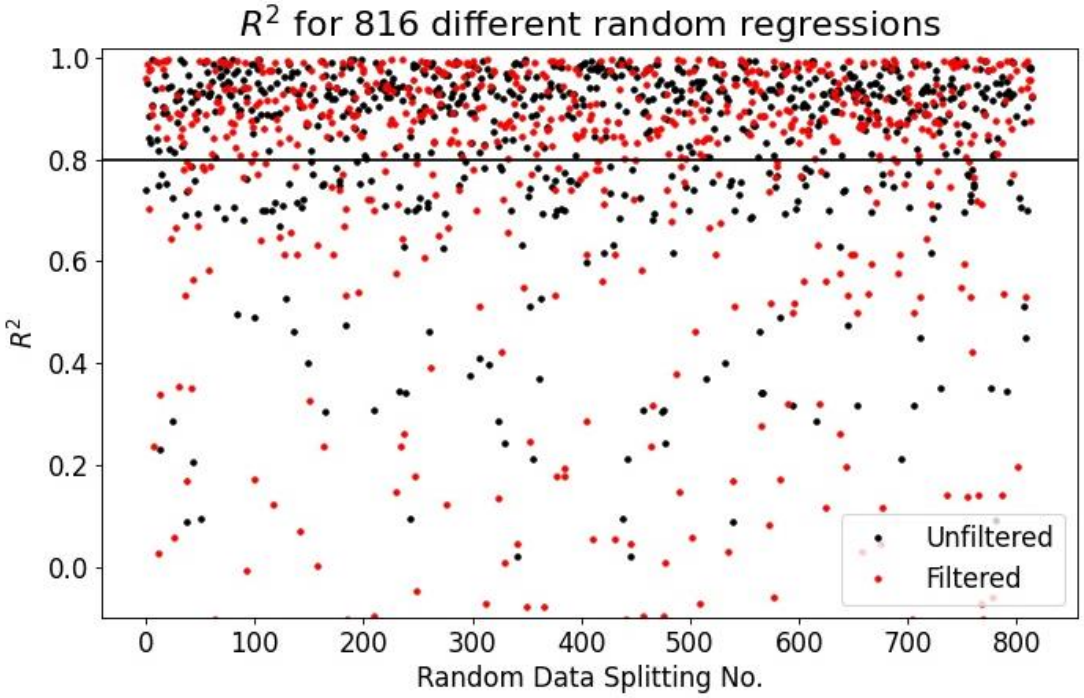

Figure 3.6. PCR R^2 for all 816 models

Out of 816 models, we found that over 66% of the unfiltered data models had an R^2 (test) > 0.8. And filtered data had around 59% of models with R^2 (test) > 0.8, as seen in Figure 3.7. This shows that for the models trained using PCR analysis on this dataset, we can predict the ROS activity with R^2 (test) > 0.8 over 60%. Part of the reason that around 30% of models did not give good results, is because two-thirds of data is from cookstove sources (natural and forced-draft). For the models where the test samples are not significantly different from training data, the R^2 values would be worse since the model would be trained on different samples than test samples.

Next, we calculated the coefficients of three PCs and the intercept value for all the models with R^2 (test) > 0.8 to get the average of the coefficients and intercepts to generate a model to predict the DTT in the samples. We plot the coefficients of PCs and intercepts for unfiltered samples in Figure 3.8.

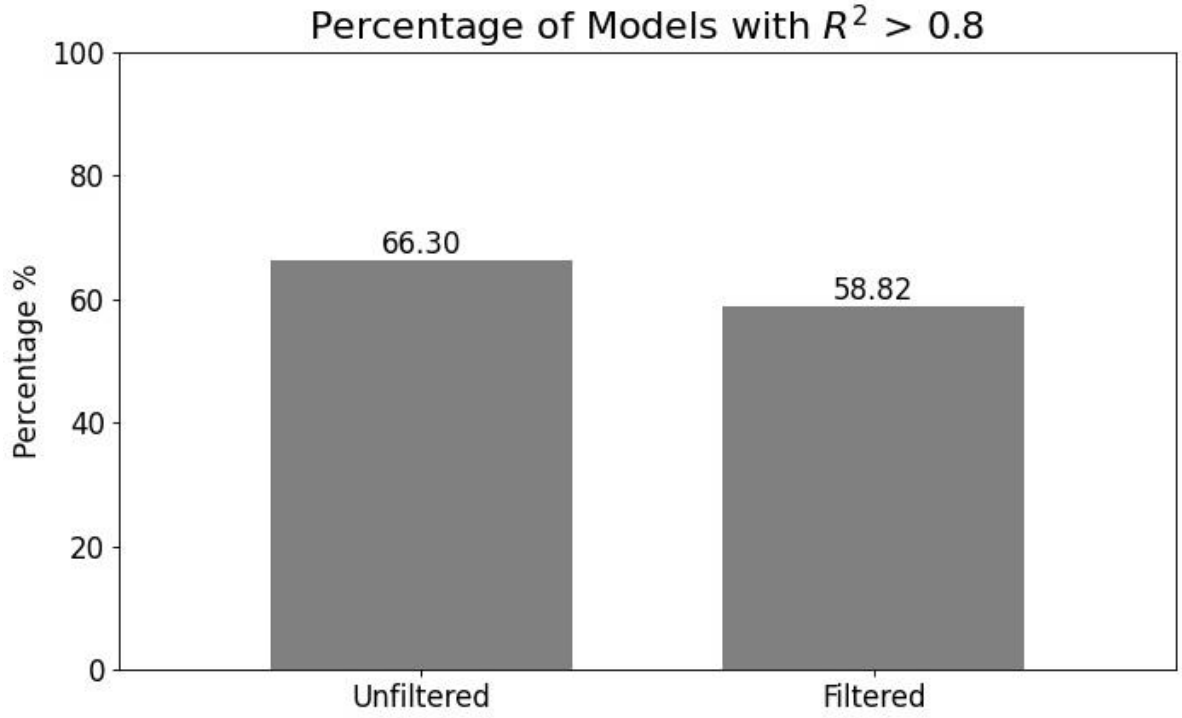

Figure 3.7. Percentage of PCR models with R^2 (test) > 0.8 vs. sample type

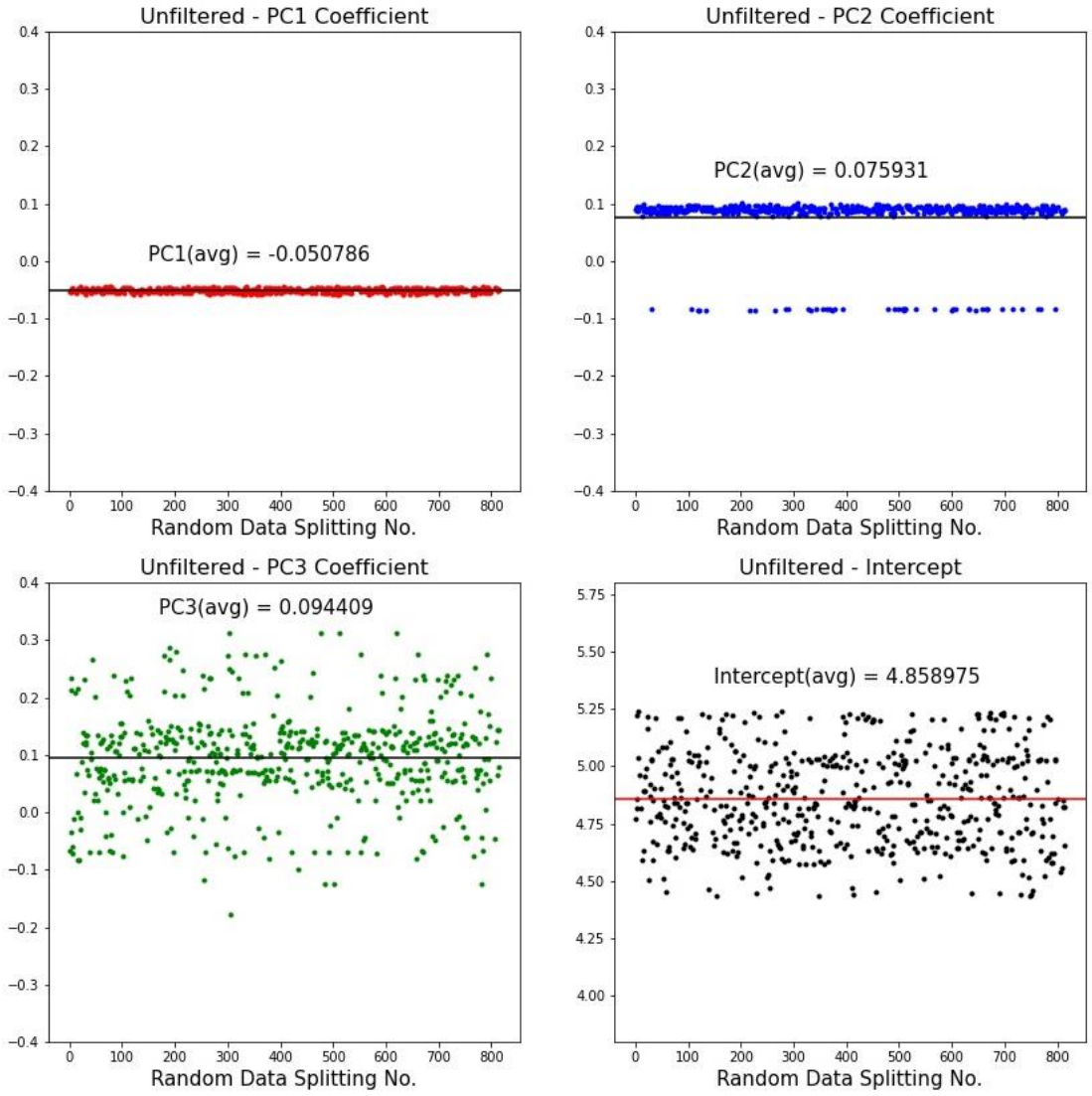

Figure 3.8. Coefficients of PCs and intercepts for all the models with R^2 (test) > 0.8 for unfiltered samples.

We plot the coefficients of PCs and intercepts for filtered samples in Figure 3.9.

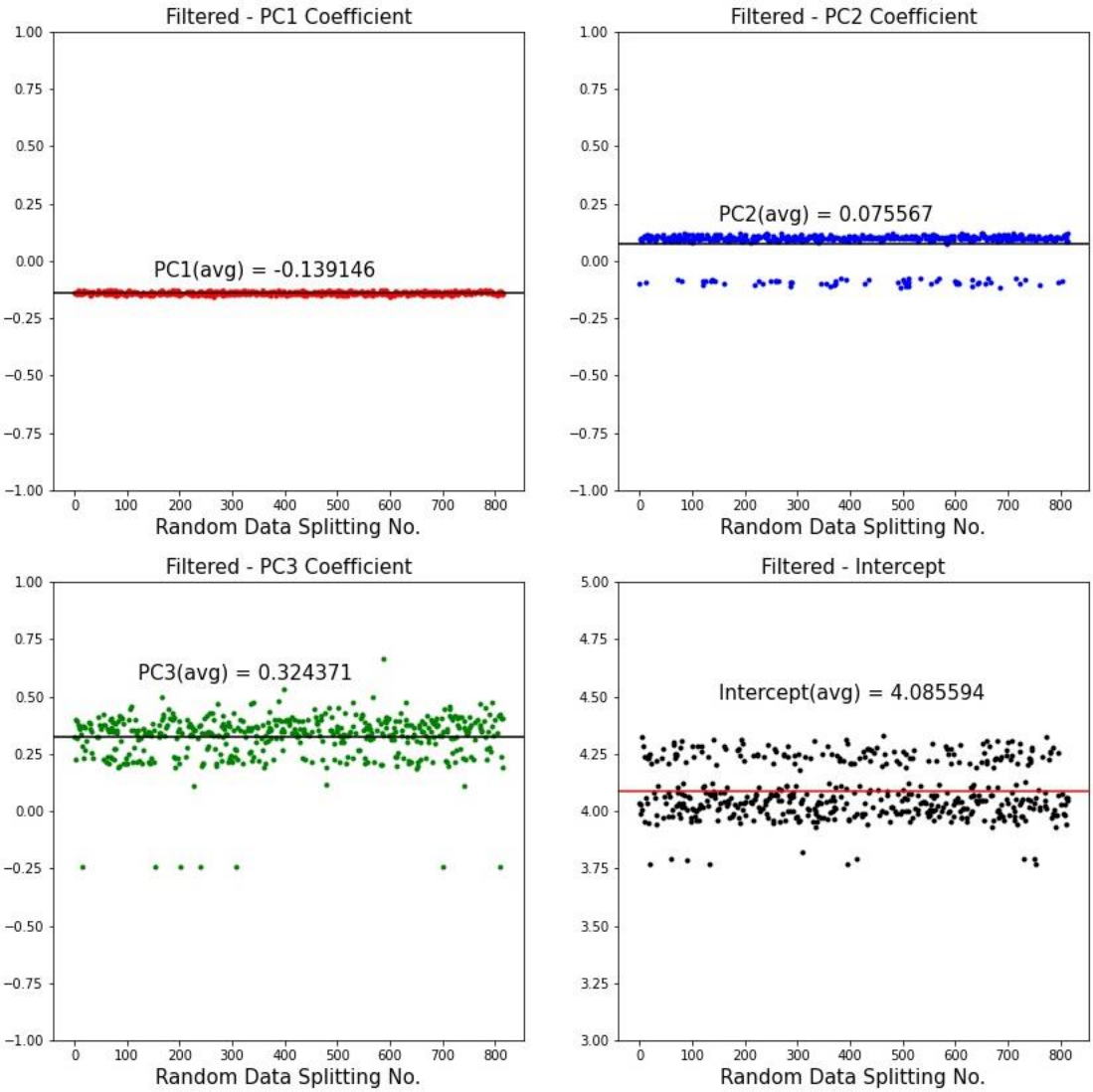

Figure 3.9. Coefficients of PCs and intercepts for all the models with R^2 (test) > 0.8 for filtered samples.

Using the average values of coefficients of PCs and intercept, we generated a model to predict the DTT reactivity of the samples. The averaged models generated using unfiltered (3.1) and filtered (3.2) samples are shown in the equations below:

$$\mathrm{DTT}_{(\mathrm{pred})}^{Unfiltered} = -0.050 \times \mathrm{PC}_1 + 0.076 \times \mathrm{PC}_2 + 0.094 \times \mathrm{PC}_3 + 4.86 \quad (3.1)$$

$$\mathrm{DTT}_{(\mathrm{pred})}^{Filtered} = -0.139 \times \mathrm{PC}_1 + 0.076 \times \mathrm{PC}_2 + 0.324 \times \mathrm{PC}_3 + 4.08 \quad (3.2)$$

Figure 3.10 shows the plot where we use the above model to predict the DTT of the 18 unfiltered and filtered samples.

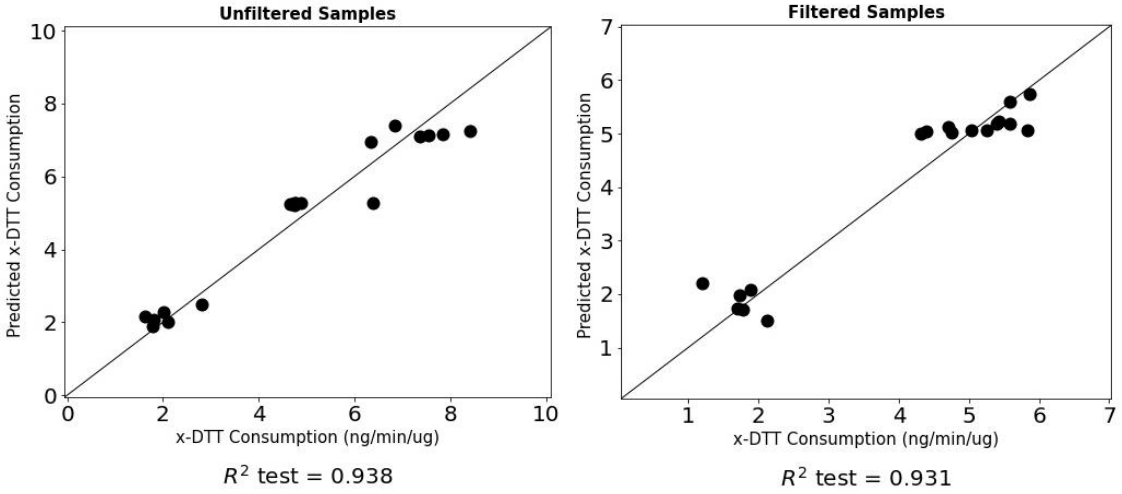

Figure 3.10. Parity plots showing results of PCR with an averaged model for (a) unfiltered and (b) filtered samples.

As we can see from the plots above, these models can predict the ROS activity in a combustion PM sample very accurately, giving us a platform to build on with more data to make an even better model.

Although we have a good correlation using the averaged model, there is a huge intercept in both unfiltered and filtered models. In the ideal case, there should not be an intercept as zero fluorescence means the blank sample, so there can be no ROS activity. The presence of the intercept led us to believe that the different sample sub-categories would have different models and cannot be combined into one general group. Hence, we move to the second analysis.

3.4.2 *Second Analysis - PCR Analysis of Subcategories of Samples*

We generated models for each sub-category. For training the model, we used all six samples in each sub-category. Further, we also test the accuracy of the model of one sub-category to predict the other two categories.

3.4.2.1 Natural-draft Cookstove Models:

Using PCA and linear regression (PCR), we get the equations for both unfiltered (3.3) and filtered (3.4) natural draft cookstove models, as shown below.

$$\mathrm{DTT}_{(\mathrm{pred})}^{Unfiltered} = 0.018 \times \mathrm{PC}_1 + 0.199 \times \mathrm{PC}_2 - 0.187 \times \mathrm{PC}_3 + 2.024 \quad (3.3)$$

$$\mathrm{DTT}_{(\mathrm{pred})}^{Filtered} = 0.105 \times \mathrm{PC}_1 - 0.019 \times \mathrm{PC}_2 - 0.378 \times \mathrm{PC}_3 + 1.73 \quad (3.4)$$

Figure 3.11 plots predicted vs. actual DTT consumption for unfiltered and filtered natural-draft cookstove samples using the above models. The unfiltered model has a better R^2 (test) = 0.934, while the filtered model has R^2 (test) = 0.805.

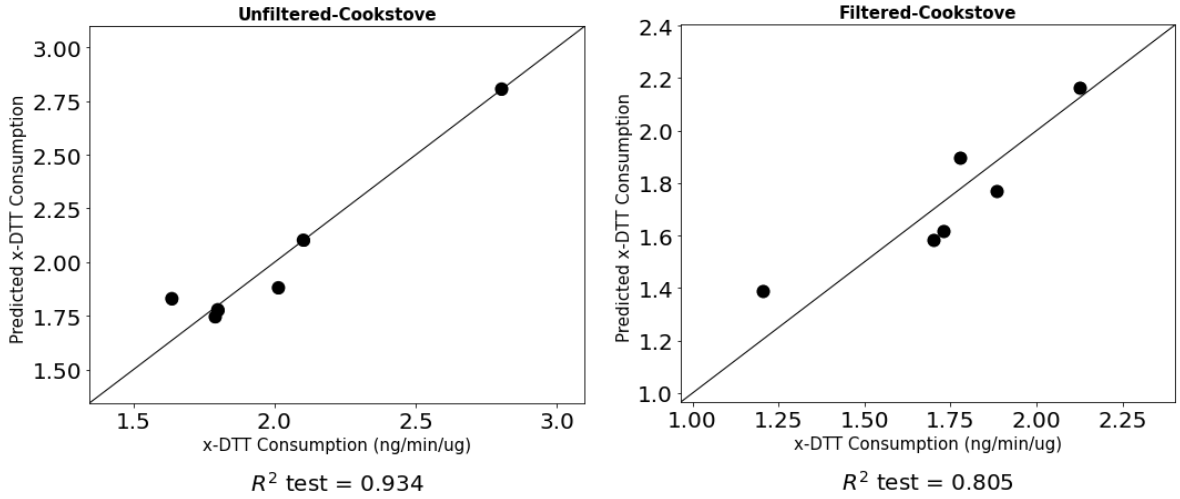

Figure 3.11. Parity plots showing results of PCR with the model for (a) unfiltered and (b) filtered natural-draft cookstove samples.

This model is then used to predict the ROS activity of other types of samples. Figure 3.12 shows how the natural-draft cookstove model predicts the ROS activity in wildfire samples using EEM and PCR for unfiltered and filtered samples. Black dots are for the natural-draft cookstove, and red squares are for the wildfire samples. Here, the model under-predicts the ROS activity for wildfire samples.

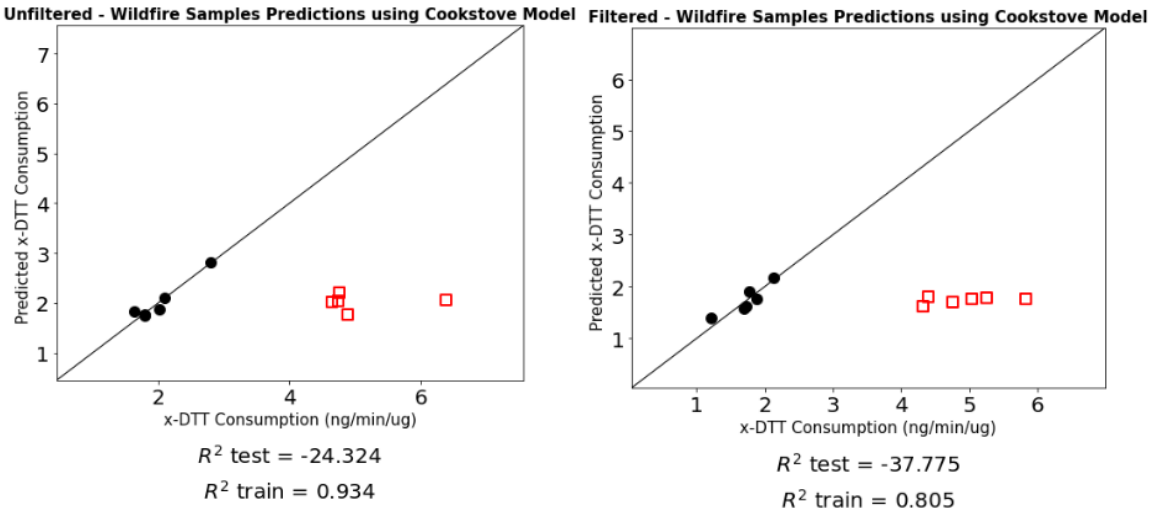

Figure 3.12. Parity plots showing results of DTT prediction of wildfire samples using the natural-draft cookstove model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (natural-draft cookstove), and red squares plot the test data (wildfire samples)

Similarly, this model is also used to predict the forced-draft cookstove samples, and plots are shown in Figure 3.13. Similar to wildfire samples, the natural-draft cookstove model under-predicted the ROS activity of forced-draft samples.

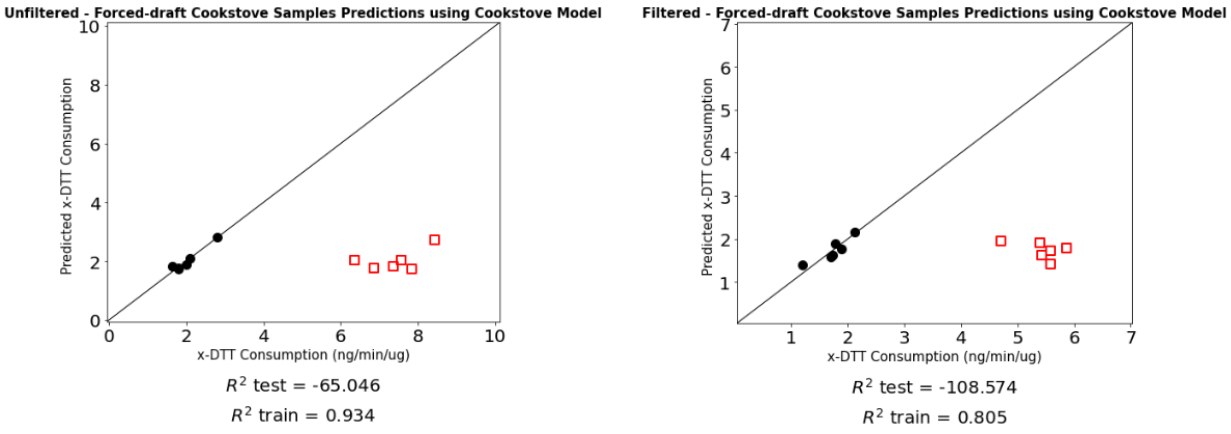

Figure 3.13. Parity plots showing results of DTT prediction of forced-draft samples using natural draft cookstove model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (natural-draft cookstove) and red squares plot the test data (forced-draft samples)

3.4.2.2 Wildfire and Forced-draft Cookstove Models:

Similar results were seen in wildfire and forced-draft cookstove models, where the models were able to predict the samples they were trained on but not the samples of different sub-categories. More information is added to APPENDIX A. One exception was the unfiltered wildfire model that did not correlate with R^2 (test) = 0.174. This was due to the noise in unfiltered wildfire samples.

This analysis provides us with a better picture of the models for different samples that are very different from each other. We also see that the models are not good at predicting the ROS activity of other sub-category samples. We could not perform the cross-validation method here because as stated earlier, the minimum number of required samples for the test set is 3 and that would make the training set so small that it would not provide any meaningful results.

From the results from the second analysis, we decided to do another analysis using just the cookstove samples (both natural and forced-draft cookstove). One, because we wanted to remove the wildfire samples with the noise, and also the cookstove samples were collected in a laboratory setting so that data is much better than wildfire samples.

3.4.3 *Third Analysis - PCR Analysis of Cookstove Samples (Natural and Forced-draft)*

PCA was performed on the nine training spectra to generate the PCs. The first three PCs for the filtered and unfiltered smoke samples are shown in Figure 3.14. The first three PCs account for over 78% of the variance in the dataset.

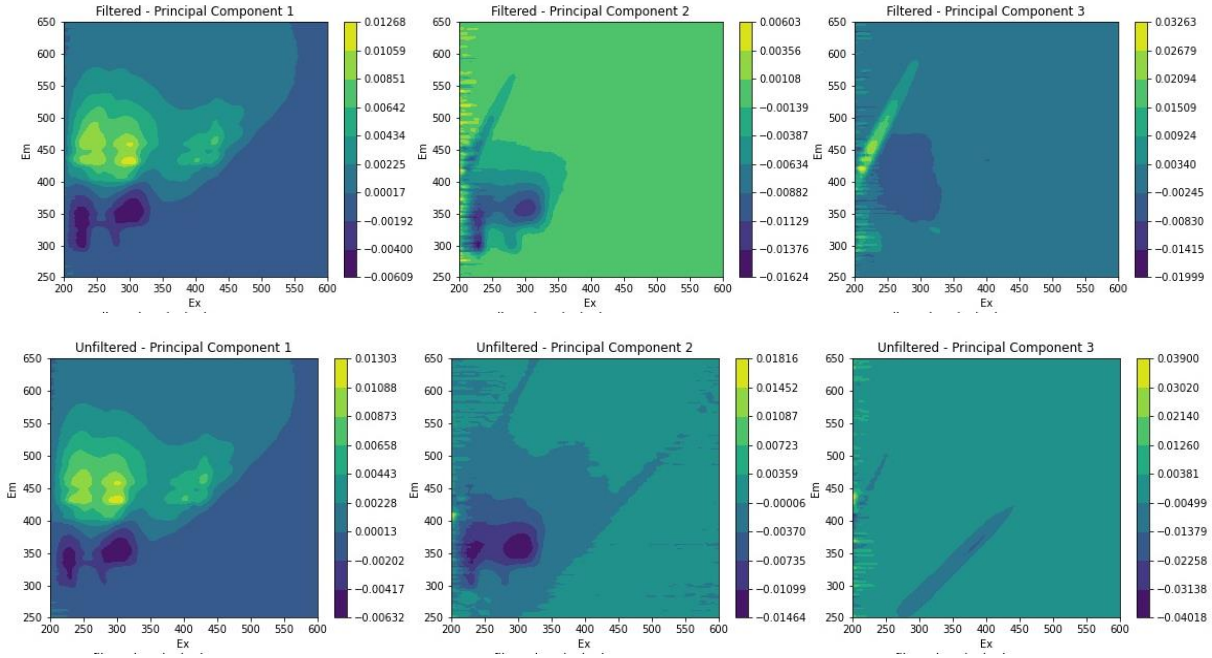

Figure 3.14. The first three principal components of the nine training EEM spectra for filtered (top) and unfiltered (bottom) smoke samples. These three PCs shown here account for 78% of the variance in the dataset.

Figure 3.15 plots the R^2 value using each of the 9 principal components of the filtered and unfiltered samples calculated from PCR. We ran a total of 220 cases for all possible models that could be formed from the 12 samples in the data.

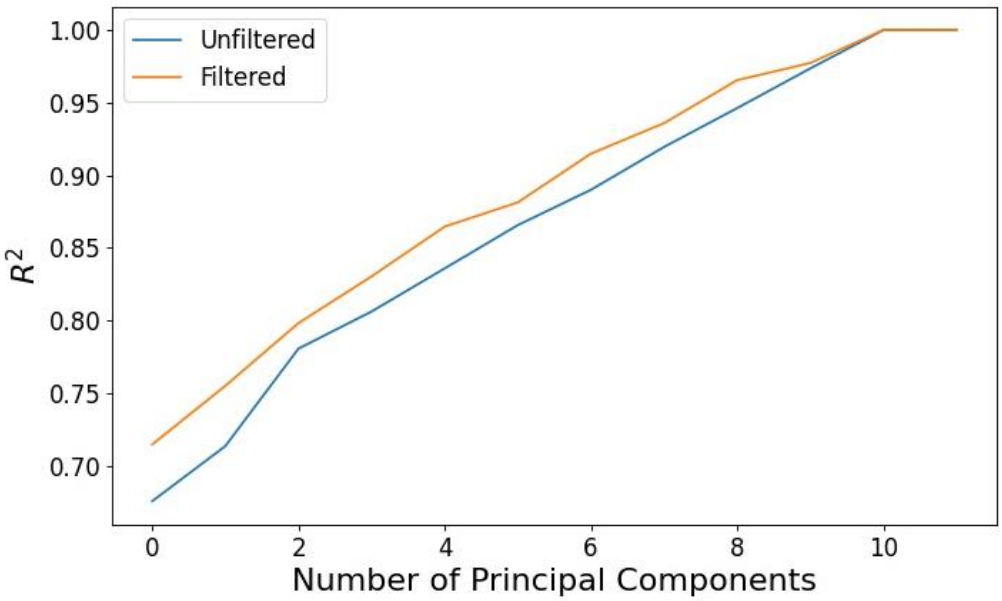

Figure 3.15. PCR R^2 vs. number of PCs for all cookstove samples

Figure 3.16 shows R^2 (test) for all the 220 models. Out of 220 models, we found that around 83% of the unfiltered data models had an R^2 (test) value greater than 0.8.

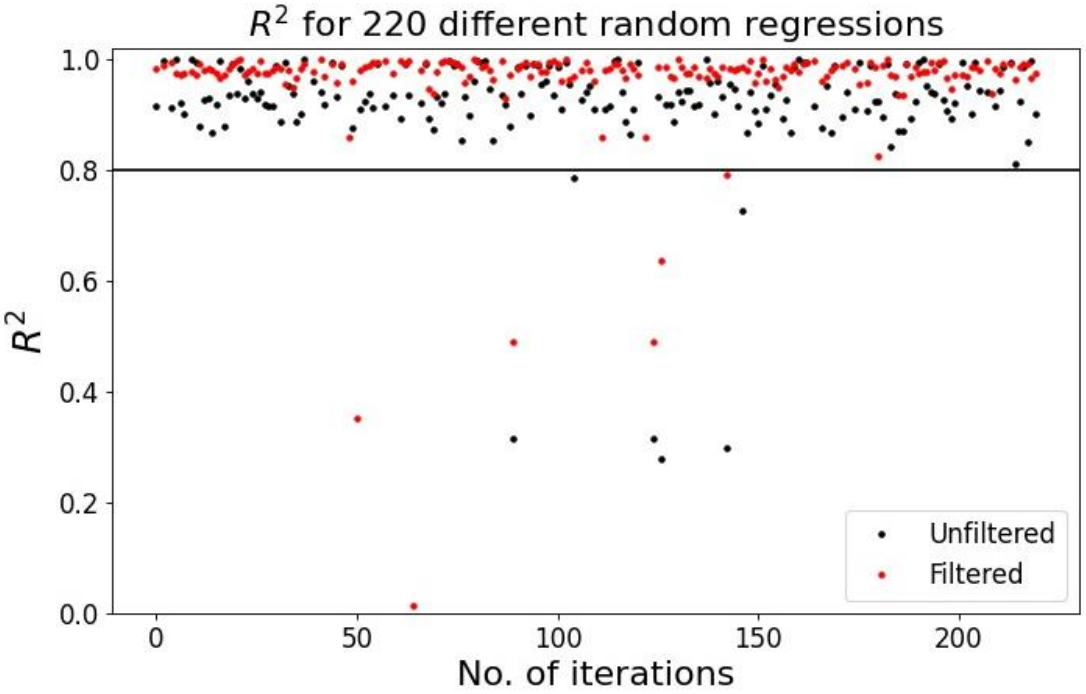

Figure 3.16. R^2 vs. model iterations for all cookstove samples

And filtered data had around 85% of models with an R^2 (test) > 0.8, as seen in Figure 3.17. This shows the reliability of this model for predicting ROS using EEM data. Further, we also calculated the coefficients of three PCs and the intercept value for all the models with R^2 (test) > 0.8.

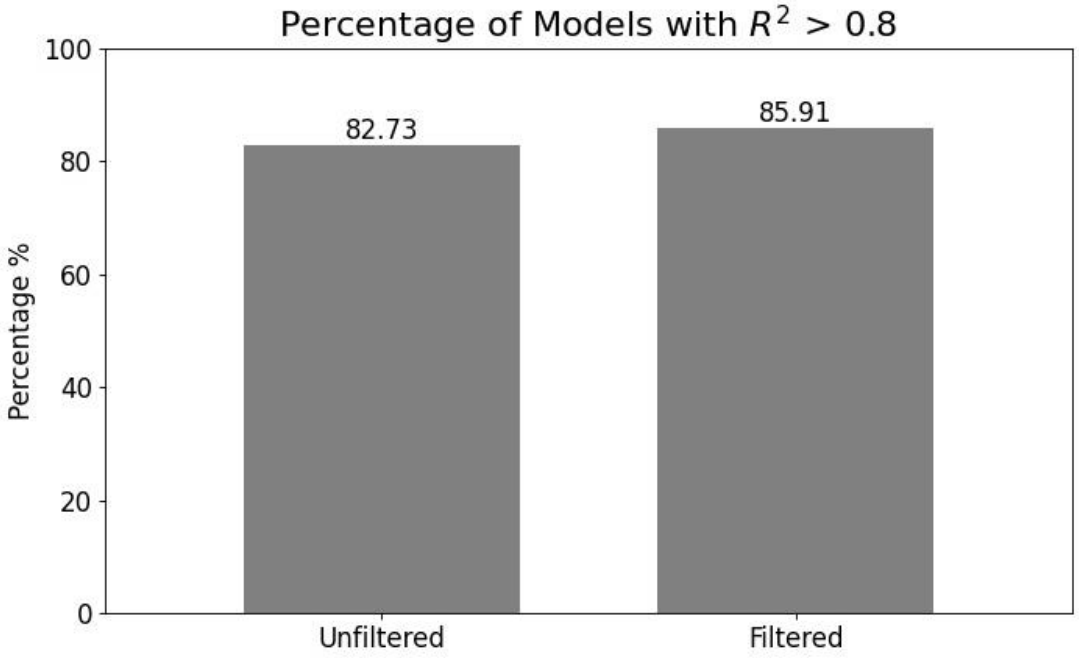

Figure 3.17. Percentage of PCR models with R^2 (test) > 0.8 vs. sample type

We plot the coefficients of PCs and intercepts for unfiltered samples in Figure 3.18. We see the segregation of the coefficients since the coefficients of PCs of sub-category models of natural and forced-draft cookstoves have opposite signs.

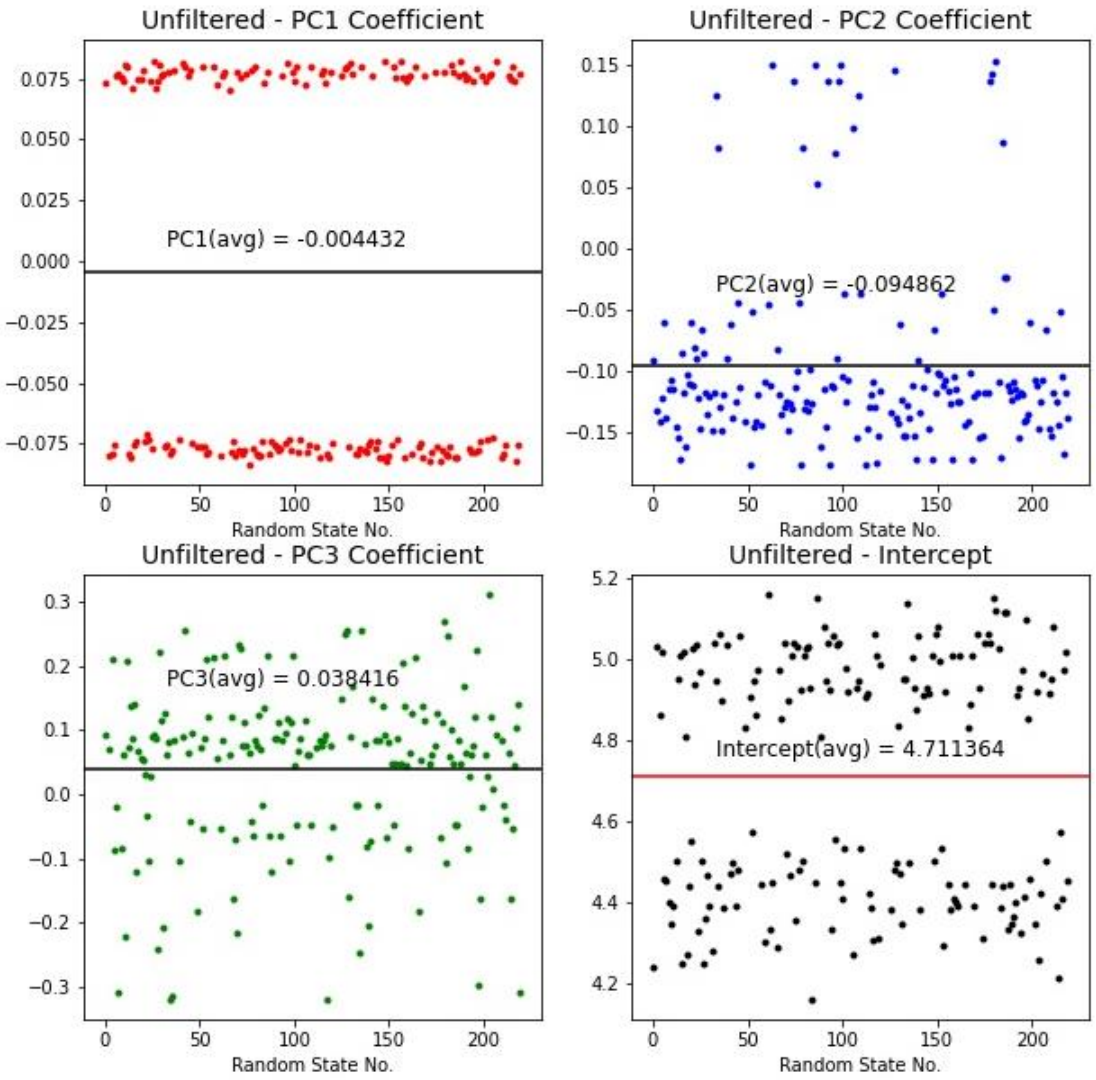

Figure 3.18. Coefficients of PCs and intercepts for all the models with R^2 (test) > 0.8 for unfiltered samples

We plot the coefficients of PCs and intercepts for filtered samples in Figure 3.19.

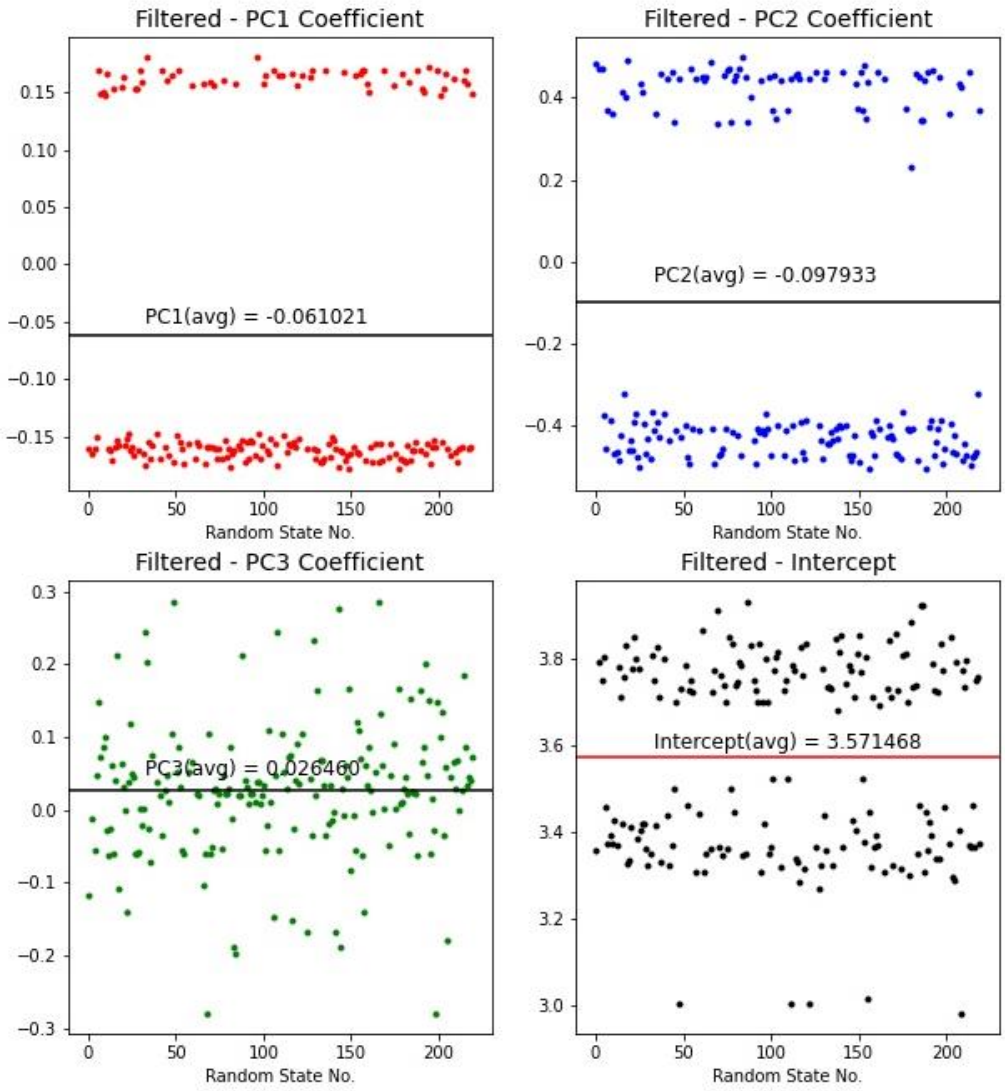

Figure 3.19. Coefficients of PCs and intercepts for all the models with R^2 (test) > 0.8 for filtered samples

The averaged model could not be produced as it was done in the first analysis because the coefficients of some PCs in both filtered and unfiltered models were evenly separated around the zero value and the average came out to be close to zero. This segregation was because, in linear regression, the combined model of two sub-categories would be a linear combination of the models of individual sample categories. And when we look at the individual models of natural-draft and forced-draft cookstove samples, the coefficient of PC_1 for unfiltered models of both natural and

forced-draft samples had opposite numerical signs, one is a positive one is negative. Similarly, it was the case for coefficients of PC_1 and PC_2 for filtered models of natural and forced-draft samples. Hence, we see two bands of points in the plots.

Chapter 4. CONCLUSIONS

This study demonstrated the correlation between EEM signatures and ROS measurements using PCR for biomass combustion aerosol, albeit it depends on the source of the combustion aerosol. PM collected from the cookstove and wildfire smoke were analyzed by EEM and the DTT assay. The unfiltered samples had higher ROS signals for natural-draft cookstoves than filtered samples, which can be attributed to polycyclic aromatic compounds (PACs) bounded to the particles. For the forced draft cookstove, both unfiltered and filtered samples had similar ROS signals. We saw that the filtered and unfiltered samples of each type of smoke had relatively similar DTT consumption, but the total fluorescence did not show the same pattern. Further, total fluorescence shows the grouping of the sample types but does not provide quantitative correlations.

So, we used EEM-PCR analysis to show that EEM data can be used to predict ROS content in PM samples using multivariate features of EEM. We ran all the possible models from the 18 samples dataset and found that the unfiltered dataset gave a model with R^2 (test) > 0.8 for 67% (541) of a total of 816 models. The filtered dataset also produced a good 58% (480) of models with R^2 (test) > 0.8, suggesting that this method can provide reliable predictions of ROS activity. Further using the coefficients of PCs for these models (with R^2 (test) > 0.8), we also generated the model using the average of the coefficients of PCs and intercept and got the resulting DTT prediction with the accuracy of R^2 (test) = 0.938 for unfiltered data and R^2 (test) = 0.931 for filtered data.

Applying PCR to sub-categories shows that different combustion-generated PM samples have different attributes and generate different predictive models. Based on this information, one can quantitatively correlate DTT consumption (ROS activity), for known PM sources. This point is further reaffirmed when we use all the cookstove samples (natural and forced-draft) and use

PCR, it provides many reliable results, with over 83% of unfiltered models and 85% of filtered models giving R^2 (test) = 0.8.

The filtered models had similar results to unfiltered models in terms of the correlation coefficient, R^2, which was also the case in total fluorescence analysis where we saw ROS did not change on filtration. Further, the filtered models had smaller intercepts compared to the unfiltered models of the same training sets. This inference could be useful to decide on using the filtered samples for EEM-ROS analysis going forward.

Although we see a correlation between the EEM and ROS in these samples, there are some limitations to this analysis. The models are generated on a small dataset and these models are not generalizable yet. The models we got also had a huge intercept in most cases which does not make reasonable sense in the physical world but that would change with more samples to train the models. This study used linear regression for all the analysis, for a reason that it is the simplest and most efficient method in supervised machine learning. But there are other methods that could improve the results. Another side of this is ROS analysis, where we used a DTT assay for the measurement of ROS activity. Understanding the underlying chemical process and training the model based on that rather than using a linear model could provide significant improvements to these results.

In conclusion, these results do show that the EEM with PCR method can provide a good correlation with the DTT measurements for different sources of combustion PM. This method could be used as an alternative method to direct measurement of DTT reactivity to evaluate the ROS levels in combustion-generated aerosols, provided we have a trained model with a greater number of samples.

Chapter 5. FUTURE WORK

This study demonstrates a framework for using EEM-PCR for ROS analysis of combustion-generated PM. Moving forward, several steps would help establish this method for ROS assessment.

1. Collecting more samples from different combustion sources would be the first step. This would help remove the intercept in the model as having a more comprehensive range of data would aid model training over a broader range of values.
2. Currently, we use linear regression, which is a good starting point; however, other methods could be potentially better for supervised machine learning, such as polynomial regression or a random forest tree method.
3. Understanding the underlying principles of ROS analysis can give a better insight into the model relating PAH fluorescence and DTT consumption. Baseline experiments with neat PAH compounds and their mixture are required.
4. Chemical analysis (GS/MS or LC/MS) for combustion-generated PM samples would be beneficial to aid the correlation.

BIBLIOGRAPHY

[1] I. f. H. M. a. E. (IHME). (2018). *GBD Compare Data Visualization, Seattle, WA*. Available: http://vizhub.healthdata.org/

[2] G. Mahamuni, J. Rutherford, J. Davis, E. Molnar, J. D. Posner, E. Seto, *et al.*, "Excitation–Emission Matrix Spectroscopy for Analysis of Chemical Composition of Combustion Generated Particulate Matter," *Environmental Science & Technology*, vol. 54, pp. 8198-8209, 2020/07/07 2020.

[3] F. Karagulian, C. A. Belis, C. F. C. Dora, A. M. Prüss-Ustün, S. Bonjour, H. Adair-Rohani, *et al.*, "Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level," *Atmospheric Environment*, vol. 120, pp. 475-483, 2015/11/01/ 2015.

[4] K. Donaldson, V. Stone, A. Seaton, and W. MacNee, "Ambient particle inhalation and the cardiovascular system: potential mechanisms," *Environmental Health Perspectives*, vol. 109, pp. 523-527, 2001/08/01 2001.

[5] N. Li, C. Sioutas, A. Cho, D. Schmitz, C. Misra, J. Sempf, *et al.*, "Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage," *Environ Health Perspect*, vol. 111, pp. 455-60, Apr 2003.

[6] A. Baulig, M. Garlatti, V. Bonvallot, A. Marchand, R. Barouki, F. Marano, *et al.*, "Involvement of reactive oxygen species in the metabolic pathways triggered by diesel exhaust particles in human airway epithelial cells," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 285, pp. L671-L679, 2003/09/01 2003.

[7] M. Strak, A. H. Janssen Nicole, J. Godri Krystal, I. Gosens, S. Mudway Ian, R. Cassee Flemming, *et al.*, "Respiratory Health Effects of Airborne Particulate Matter: The Role of Particle Size, Composition, and Oxidative Potential—The RAPTES Project," *Environmental Health Perspectives*, vol. 120, pp. 1183-1189, 2012/08/01 2012.

[8] B. Dellinger, W. A. Pryor, R. Cueto, G. L. Squadrito, V. Hegde, and W. A. Deutsch, "Role of Free Radicals in the Toxicity of Airborne Fine Particulate Matter," *Chemical Research in Toxicology*, vol. 14, pp. 1371-1377, 2001/10/01 2001.

[9] H. Autrup, B. Daneshvar, L. O. Dragsted, M. Gamborg, M. Hansen, S. Loft, *et al.*, "Biomarkers for exposure to ambient air pollution--comparison of carcinogen-DNA adduct levels with other exposure markers and markers for oxidative stress," *Environmental Health Perspectives*, vol. 107, pp. 233-238, 1999/03/01 1999.

[10] M. Sørensen, B. Daneshvar, M. Hansen, O. Dragsted Lars, O. Hertel, L. Knudsen, *et al.*, "Personal PM2.5 exposure and markers of oxidative stress in blood," *Environmental Health Perspectives*, vol. 111, pp. 161-166, 2003/02/01 2003.

[11] A. K. Prahalad, J. Inmon, L. A. Dailey, M. C. Madden, A. J. Ghio, and J. E. Gallagher, "Air Pollution Particles Mediated Oxidative DNA Base Damage in a Cell Free System and in Human Airway Epithelial Cells in Relation to Particulate Metal Content and Bioreactivity," *Chemical Research in Toxicology*, vol. 14, pp. 879-887, 2001/07/01 2001.

[12] J. C. Ball, A. M. Straccia, W. C. Young, and A. E. Aust, "The Formation of Reactive Oxygen Species Catalyzed by Neutral, Aqueous Extracts of NIST Ambient Particulate Matter and Diesel Engine Particles," *Journal of the Air & Waste Management Association*, vol. 50, pp. 1897-1903, 2000/11/01 2000.

[13] U. P. Kodavanti, M. C. Schladweiler, A. D. Ledbetter, R. Hauser, D. C. Christiani, J. McGee, *et al.*, "TEMPORAL ASSOCIATION BETWEEN PULMONARY AND SYSTEMIC EFFECTS OF PARTICULATE MATTER IN HEALTHY AND CARDIOVASCULAR COMPROMISED RATS," *Journal of Toxicology and Environmental Health, Part A*, vol. 65, pp. 1545-1569, 2002/10/12 2002.

[14] A. Saffari, N. Daher, M. M. Shafer, J. J. Schauer, and C. Sioutas, "Global Perspective on the Oxidative Potential of Airborne Particulate Matter: A Synthesis of Research Findings," *Environmental Science & Technology*, vol. 48, pp. 7576-7583, 2014/07/01 2014.

[15] D. Gao, S. Ripley, S. Weichenthal, and K. J. Godri Pollitt, "Ambient particulate matter oxidative potential: Chemical determinants, associated health effects, and strategies for risk management," *Free Radic Biol Med*, vol. 151, pp. 7-25, May 1 2020.

[16] T. Xia, P. Korge, N. Weiss James, N. Li, M. I. Venkatesen, C. Sioutas, *et al.*, "Quinones and Aromatic Chemical Compounds in Particulate Matter Induce Mitochondrial Dysfunction: Implications for Ultrafine Particle Toxicity," *Environmental Health Perspectives*, vol. 112, pp. 1347-1358, 2004/10/01 2004.

[17] N. Li, M. I. Venkatesan, A. Miguel, R. Kaplan, C. Gujuluva, J. Alam, *et al.*, "Induction of Heme Oxygenase-1 Expression in Macrophages by Diesel Exhaust Particle Chemicals and Quinones via the Antioxidant-Responsive Element," *The Journal of Immunology*, vol. 165, p. 3393, 2000.

[18] J. T. Bates, T. Fang, V. Verma, L. Zeng, R. J. Weber, P. E. Tolbert, *et al.*, "Review of Acellular Assays of Ambient Particulate Matter Oxidative Potential: Methods and Relationships with Composition, Sources, and Health Effects," *Environmental Science & Technology*, vol. 53, pp. 4003-4019, 2019/04/16 2019.

[19] L. W. Stanek, J. D. Sacks, S. J. Dutton, and J.-J. B. Dubois, "Attributing health effects to apportioned components and sources of particulate matter: An evaluation of collective results," *Atmospheric Environment*, vol. 45, pp. 5655-5663, 2011/10/01/ 2011.

[20] J. J. West, A. Cohen, F. Dentener, B. Brunekreef, T. Zhu, B. Armstrong, *et al.*, "'What We Breathe Impacts Our Health: Improving Understanding of the Link between Air Pollution and Health'," *Environmental Science & Technology*, vol. 50, pp. 4895-4904, 2016/05/17 2016.

[21] K. Adams, D. S. Greenbaum, R. Shaikh, A. M. van Erp, and A. G. Russell, "Particulate matter components, sources, and health: Systematic approaches to testing effects," *Journal of the Air & Waste Management Association*, vol. 65, pp. 544-558, 2015/05/04 2015.

[22] N. J. Hime, G. B. Marks, and C. T. Cowie, "A Comparison of the Health Effects of Ambient Particulate Matter Air Pollution from Five Emission Sources," *International Journal of Environmental Research and Public Health*, vol. 15, p. 1206, 2018.

[23] N. Li, M. I. Venkatesan, A. Miguel, R. Kaplan, C. Gujuluva, J. Alam, *et al.*, "Induction of heme oxygenase-1 expression in macrophages by diesel exhaust particle chemicals and quinones via the antioxidant-responsive element," *J Immunol*, vol. 165, pp. 3393-401, Sep 15 2000.

[24] J. T. Bates, R. J. Weber, J. Abrams, V. Verma, T. Fang, M. Klein, *et al.*, "Reactive Oxygen Species Generation Linked to Sources of Atmospheric Particulate Matter and Cardiorespiratory Effects," *Environmental Science & Technology*, vol. 49, pp. 13605-13612, 2015/11/17 2015.

[25] A. Yang, N. A. H. Janssen, B. Brunekreef, F. R. Cassee, G. Hoek, and U. Gehring, "Children's respiratory health and oxidative potential of PM_{2.5}: the PIAMA birth cohort study," *Occupational and Environmental Medicine*, vol. 73, pp. 154-160, 2016.

[26] J. Y. Abrams, R. J. Weber, M. Klein, S. E. Samat, H. H. Chang, M. J. Strickland, *et al.*, "Associations between Ambient Fine Particulate Oxidative Potential and Cardiorespiratory Emergency Department Visits," *Environmental Health Perspectives*, vol. 125, p. 107008, 2017.

[27] Y. Kumagai, S. Koide, K. Taguchi, A. Endo, Y. Nakai, T. Yoshikawa, *et al.*, "Oxidation of Proximal Protein Sulfhydryls by Phenanthraquinone, a Component of Diesel Exhaust Particles," *Chemical Research in Toxicology*, vol. 15, pp. 483-489, 2002/04/01 2002.

[28] A. K. Cho, E. Di Stefano, Y. You, C. E. Rodriguez, D. A. Schmitz, Y. Kumagai, *et al.*, "Determination of Four Quinones in Diesel Exhaust Particles, SRM 1649a, and Atmospheric PM2.5 Special Issue of Aerosol Science and Technology on Findings from the Fine Particulate Matter Supersites Program," *Aerosol Science and Technology*, vol. 38, pp. 68-81, 2004/12/01 2004.

[29] W. M. Jolly, M. A. Cochrane, P. H. Freeborn, Z. A. Holden, T. J. Brown, G. J. Williamson, *et al.*, "Climate-induced variations in global wildfire danger from 1979 to 2013," *Nature Communications*, vol. 6, p. 7537, 2015-11-01 2015.

[30] X. Yue, L. J. Mickley, J. A. Logan, and J. O. Kaplan, "Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century," *Atmospheric Environment*, vol. 77, pp. 767-780, 2013-10-01 2013.

[31] S. Tiwari, U. Dumka, A. Gautam, D. Kaskaoutis, A. Srivastava, D. Bisht, *et al.*, "Assessment of PM2. 5 and PM10 over Guwahati in Brahmaputra River Valley: Temporal evolution, source apportionment and meteorological dependence," *Atmospheric Pollution Research*, vol. 8, pp. 13-28, 2017.

[32] M. Burke, A. Driscoll, S. Heft-Neal, J. Xue, J. Burney, and M. Wara, "The changing risk and burden of wildfire in the United States," *Proceedings of the National Academy of Sciences*, vol. 118, p. e2011048118, 2021-01-12 2021.

[33] V. Samburova, B. Zielinska, and A. Khlystov, "Do 16 polycyclic aromatic hydrocarbons represent PAH air toxicity?," *Toxics*, vol. 5, p. 17, 2017.

[34] R. Aguilera, T. Corringham, A. Gershunov, and T. Benmarhnia, "Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from Southern California," *Nature Communications*, vol. 12, Mar 5 2021.

[35] S. Magzamen, R. W. Gan, J. Y. Liu, K. O'Dell, B. Ford, K. Berg, *et al.*, "Differential Cardiopulmonary Health Impacts of Local and Long-Range Transport of Wildfire Smoke," *Geohealth*, vol. 5, Mar 2021.

[36] E. L. Landguth, Z. A. Holden, J. Graham, B. Stark, E. B. Mokhtari, E. Kaleczyc, *et al.*, "The delayed effect of wildfire season particulate matter on subsequent influenza season in a mountain west region of the USA," *Environment International*, vol. 139, Jun 2020.

[37] C. P. West, A. P. S. Hettiyadura, A. Darmody, G. Mahamuni, J. Davis, I. Novosselov, *et al.*, "Molecular Composition and the Optical Properties of Brown Carbon Generated by the Ethane Flame," *ACS Earth and Space Chemistry*, vol. 4, pp. 1090-1103, 2020/07/16 2020.

[38] J. Davis, E. Molnar, and I. Novosselov, "Nanostructure transition of young soot aggregates to mature soot aggregates in diluted diffusion flames," *Carbon*, vol. 159, pp. 255-265, 2020/04/15/ 2020.

[39] J. Davis, K. Tiwari, and I. Novosselov, "Soot morphology and nanostructure in complex flame flow patterns via secondary particle surface growth," *Fuel*, vol. 245, pp. 447-457, 2019/06/01/ 2019.

[40] S. Deflorio-Barker, J. Crooks, J. Reyes, and A. G. Rappold, "Cardiopulmonary Effects of Fine Particulate Matter Exposure among Older Adults, during Wildfire and Non-Wildfire Periods, in the United States 2008–2010," *Environmental Health Perspectives*, vol. 127, p. 037006, 2019-03-01 2019.

[41] Y. Liu, E. Austin, J. Xiang, T. Gould, T. Larson, and E. Seto, "Health Impact Assessment of the 2020 Washington State Wildfire Smoke Episode: Excess Health Burden Attributable to Increased $PM_{2.5}$ Exposures and Potential Exposure Reductions," *GeoHealth*, vol. 5, 2021-05-01 2021.

[42] M. Dennekamp and M. J. Abramson, "The effects of bushfire smoke on respiratory health," *Respirology*, vol. 16, pp. 198-209, 2011-02-01 2011.

[43] C. E. Reid, M. Brauer, F. H. Johnston, M. Jerrett, J. R. Balmes, and C. T. Elliott, "Critical review of health impacts of wildfire smoke exposure," *Environmental health perspectives*, vol. 124, pp. 1334-1343, 2016.

[44] C. J. Matz, M. Egyed, G. Xi, J. Racine, R. Pavlovic, R. Rittmaster, *et al.*, "Health impact analysis of PM2. 5 from wildfire smoke in Canada (2013–2015, 2017–2018)," *Science of the Total Environment*, vol. 725, p. 138506, 2020.

[45] S. A. Meo, A. A. Abukhalaf, A. A. Alomar, O. M. Alessa, W. Sami, and D. C. Klonoff, "Effect of environmental pollutants PM-2.5, carbon monoxide, and ozone on the incidence and mortality of SARS-COV-2 infection in ten wildfire affected counties in California," *Science of the Total Environment*, vol. 757, p. 143948, 2021.

[46] J. Cortes-Ramirez, R. Michael, L. Knibbs, H. Bambrick, M. Haswell, and D. Wraith, "The association of wildfire air pollution with COVID-19 incidence in New South Wales, Australia," *Science of The Total Environment*, p. 151158, 2021.

[47] D. Kiser, G. Elhanan, W. J. Metcalf, B. Schnieder, and J. J. Grzymski, "SARS-CoV-2 test positivity rate in Reno, Nevada: association with PM2. 5 during the 2020 wildfire smoke events in the western United States," *Journal of exposure science & environmental epidemiology*, vol. 31, pp. 797-803, 2021.

[48] K. M. Navarro, K. A. Clark, D. J. Hardt, C. E. Reid, P. W. Lahm, J. W. Domitrovich, *et al.*, "Wildland firefighter exposure to smoke and COVID-19: A new risk on the fire line," *Science of The Total Environment*, vol. 760, p. 144296, 2021.

[49] N. I. F. Center. (2020). *National Large Incident Year-to-Date Report*. Available: https://gacc.nifc.gov/index.php

[50] S. Makhsous, J. M. Segovia, J. He, D. Chan, L. Lee, I. V. Novosselov, *et al.*, "Methodology for Addressing Infectious Aerosol Persistence in Real-Time Using Sensor Network," *Sensors*, vol. 21, p. 3928, 2021.

[51] S. Hegde, K. T. Min, J. Moore, P. Lundrigan, N. Patwari, S. Collingwood, *et al.*, "Indoor household particulate matter measurements using a network of low-cost sensors," *Aerosol and Air Quality Research*, vol. 20, pp. 381-394, 2020.

[52] J. Li, H. Li, Y. Ma, Y. Wang, A. A. Abokifa, C. Lu, *et al.*, "Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network," *Building and Environment*, vol. 127, pp. 138-147, 2018.

[53] P. Kumar, A. N. Skouloudis, M. Bell, M. Viana, M. C. Carotta, G. Biskos, *et al.*, "Real-time sensors for indoor air monitoring and challenges ahead in deploying them to urban buildings," *Science of the Total Environment*, vol. 560, pp. 150-159, 2016.

[54] E. Seto, E. Austin, I. Novosselov, and M. G. Yost, "Use of low-cost particle monitors to calibrate traffic-related air pollutant models in urban areas," in *International Environmental Modelling and Software Society*, 2014.

[55] X. Liu, R. Jayaratne, P. Thai, T. Kuhn, I. Zing, B. Christensen, *et al.*, "Low-cost sensors as an alternative for long-term air quality monitoring," *Environmental research*, vol. 185, p. 109438, 2020.

[56] T. Kuhn, R. Jayaratne, P. K. Thai, B. Christensen, X. Liu, M. Dunbabin, *et al.*, "Air quality during and after the Commonwealth Games 2018 in Australia: Multiple benefits of monitoring," *Journal of Aerosol Science,* vol. 152, p. 105707, 2021.

[57] W. Jiao, G. Hagler, R. Williams, R. Sharpe, R. Brown, D. Garver, *et al.*, "Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States," *Atmospheric Measurement Techniques,* vol. 9, pp. 5281-5292, 2016.

[58] M. Gao, J. Cao, and E. Seto, "A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China," *Environmental Pollution,* vol. 199, pp. 56-65, 4// 2015.

[59] P. Kumar, L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. Di Sabatino, *et al.*, "The rise of low-cost sensing for managing air pollution in cities," *Environment international,* vol. 75, pp. 199-205, 2015.

[60] J. Li, H. Zhang, C.-Y. Chao, C.-H. Chien, C.-Y. Wu, C. H. Luo, *et al.*, "Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2. 5," *Atmospheric Environment,* vol. 223, p. 117293, 2020.

[61] C.-Y. Chao, H. Zhang, M. Hammer, Y. Zhan, D. Kenney, R. V. Martin, *et al.*, "Integrating Fixed Monitoring Systems with Low-Cost Sensors to Create High-Resolution Air Quality Maps for the Northern China Plain Region," *ACS Earth and Space Chemistry,* vol. 5, pp. 3022-3035, 2021.

[62] X. Qiao, Q. Zhang, D. Wang, J. Hao, and J. Jiang, "Improving data reliability: A quality control practice for low-cost PM2. 5 sensor network," *Science of The Total Environment,* vol. 779, p. 146381, 2021.

[63] O. Stampfer, E. Austin, T. Ganuelas, T. Fiander, E. Seto, and C. J. Karr, "Use of low-cost PM monitors and a multi-wavelength aethalometer to characterize PM2. 5 in the Yakama Nation reservation," *Atmospheric Environment,* vol. 224, p. 117292, 2020.

[64] J. He, C.-H. Huang, N. Yuan, E. Austin, E. Seto, and I. Novosselov, "Network of Low-cost Air Quality Sensor for Monitoring Indoor, Outdoor, and Personal PM2. 5 Exposure in Seattle during the 2020 Wildfire Season," *arXiv preprint arXiv:2203.14140,* 2022.

[65] N. W. May, C. Dixon, and D. A. Jaffe, "Impact of Wildfire Smoke Events on Indoor Air Quality and Evaluation of a Low-cost Filtration Method," *Aerosol and Air Quality Research,* vol. 21, pp. 210046-210046, 2021.

[66] J. Xiang, C.-H. Huang, J. Shirai, Y. Liu, N. Carmona, C. Zuidema, *et al.*, "Field measurements of PM2. 5 infiltration factor and portable air cleaner effectiveness during wildfire episodes in US residences," *Science of The Total Environment,* vol. 773, p. 145642, 2021.

[67] E. Austin, I. Novosselov, E. Seto, and M. G. Yost, "Laboratory Evaluation of the Shinyei PPD42NS Low-Cost Particulate Matter Sensor," *PloS one,* vol. 10, p. e0137789, 2015.

[68] T. Njalsson and I. Novosselov, "Design and optimization of a compact low-cost optical particle sizer," *Journal of Aerosol Science,* vol. 119, pp. 1-12, 2018.

[69] C.-H. Huang, J. He, E. Austin, E. Seto, and I. Novosselov, "Assessing the value of complex refractive index and particle density for calibration of low-cost particle matter sensor for size-resolved particle count and PM2. 5 measurements," *Plos one,* vol. 16, p. e0259745, 2021.

[70] J.-B. Renard, F. Dulac, G. Berthet, T. Lurton, D. Vignelles, F. Jégou, *et al.*, "LOAC: a small aerosol optical counter/sizer for ground-based and balloon measurements of the size distribution and nature of atmospheric particles–Part 2: First results from balloon and unmanned aerial vehicle flights," *Atmospheric Measurement Techniques,* vol. 9, pp. 3673-3686, 2016.

[71] A. Nagy, W. Szymanski, P. Gal, A. Golczewski, and A. Czitrovszky, "Numerical and experimental study of the performance of the dual wavelength optical particle spectrometer (DWOPS)," *Journal of aerosol science,* vol. 38, pp. 467-478, 2007.

[72] G. Mahamuni, J. He, J. Rutherford, B. Ockerman, A. Majumdar, E. Seto, *et al.*, "Solid-phase excitation-emission matrix spectroscopy for chemical analysis of combustion aerosols," *PloS one,* vol. 16, p. e0251664, 2021.

[73] G. Mahamuni, "Excitation Emission Matrix Fluorescence Spectroscopy based Sensing of Combustion Generated Particulate Matter," Ph.D., Mechanical Engineering, University of Washington, 2020.

[74] R. Aryal, B.-K. Lee, S. Beecham, J. Kandasamy, N. Aryal, and K. Parajuli, "Characterisation of Road Dust Organic Matter as a Function of Particle Size: A PARAFAC Approach," *Water, Air, & Soil Pollution,* vol. 226, p. 24, 2015/02/06 2015.

[75] Q. Chen, Y. Miyazaki, K. Kawamura, K. Matsumoto, S. Coburn, R. Volkamer, *et al.*, "Characterization of Chromophoric Water-Soluble Organic Matter in Urban, Forest, and Marine Aerosols by HR-ToF-AMS Analysis and Excitation–Emission Matrix Spectroscopy," *Environmental Science & Technology,* vol. 50, pp. 10351-10360, 2016/10/04 2016.

[76] S. Elcoroaristizabal, A. de Juan, J. A. García, I. Elorduy, N. Durana, and L. Alonso, "Chemometric determination of PAHs in aerosol samples by fluorescence spectroscopy and second-order data analysis algorithms," *Journal of Chemometrics,* vol. 28, pp. 260-271, 2014/04/01 2014.

[77] J. T. V. Matos, S. M. S. C. Freire, R. M. B. O. Duarte, and A. C. Duarte, "Natural organic matter in urban aerosols: Comparison between water and alkaline soluble components using excitation–emission matrix fluorescence spectroscopy and multiway data analysis," *Atmospheric Environment,* vol. 102, pp. 1-10, 2015/02/01/ 2015.

[78] N. Mladenov, L. Alados-Arboledas, F. J. Olmo, H. Lyamani, A. Delgado, A. Molina, *et al.*, "Applications of optical spectroscopy and stable isotope analyses to organic aerosol source discrimination in an urban area," *Atmospheric Environment,* vol. 45, pp. 1960-1969, 2011/04/01/ 2011.

[79] N. Mladenov, J. Lípez-Ramos, D. M. McKnight, and I. Rechea, "Alpine lake optical properties as sentinels of dust deposition and global change," *Limnology and Oceanography,* vol. 54, pp. 2386-2400, 2009/11/01 2009.

[80] H. Nakajima, K. Okada, Y. Kuroki, Y. Nakama, D. Handa, T. Arakaki, *et al.*, "Photochemical formation of peroxides and fluorescence characteristics of the water-soluble fraction of bulk aerosols collected in Okinawa, Japan," *Atmospheric Environment,* vol. 42, pp. 3046-3058, 2008/04/01/ 2008.

[81] J. W. Rutherford, N. Dawson-Elli, A. M. Manicone, G. V. Korshin, I. V. Novosselov, E. Seto, *et al.*, "Excitation emission matrix fluorescence spectroscopy for combustion generated particulate matter source identification," *Atmospheric Environment,* vol. 220, p. 117065, 2020/01/01/ 2020.

[82] J. W. Rutherford, T. V. Larson, T. Gould, E. Seto, I. V. Novosselov, and J. D. Posner, "Source apportionment of environmental combustion sources using excitation emission matrix fluorescence spectroscopy and machine learning," *Atmospheric Environment,* vol. 259, p. 118501, 2021/08/15/ 2021.

[83] G. Mahamuni, "Excitation–Emission Matrix Spectroscopy for Analysis of Chemical Composition of Combustion Generated Particulate Matter," ed, 2020.

[84] A. Muñoz de la Peña, I. Durán-Merás, M. D. Moreno, F. Salinas, and M. M. Galera, "Resolution of ternary mixtures of salicylic, salicyluric and gentisic acids by partial least squares and principal component regression: Optimization of the scanning path in the excitation-emission matrices," *Fresenius' Journal of Analytical Chemistry*, vol. 351, pp. 571-576, 1995/01/01 1995.

[85] G. G. M. James, *An Introduction to Statistical Learning: With Applications in R*. New York: Springer texts in statistics, 2013.

[86] B. Sullivan, G. Allawatt, A. Emery, P. Means, J. Kramlich, and J. Posner, "Time-Resolved Particulate Emissions Monitoring of Cookstove Biomass Combustion Using a Tapered Element Oscillating Microbalance," *Combustion Science and Technology*, vol. 189, pp. 923-936, 2017.

[87] M. Barbour, D. Udesen, S. Bentson, A. Pundle, C. Tackman, D. Evitt, *et al.*, "Development of wood-burning rocket cookstove with forced air-injection," *Energy for Sustainable Development*, vol. 65, pp. 12-24, 2021/12/01/ 2021.

[88] A. J. W. G. V. Yves Engelborghs, *Fluorescence Spectroscopy and Microscopy*: Humana Press, 2014.

[89] K. R. Murphy, K. D. Butler, R. G. M. Spencer, C. A. Stedmon, J. R. Boehme, and G. R. Aiken, "Measurement of Dissolved Organic Matter Fluorescence in Aquatic Environments: An Interlaboratory Comparison," *Environmental Science & Technology*, vol. 44, pp. 9405-9412, 2010/12/15 2010.

[90] R. G. Zepp, W. M. Sheldon, and M. A. Moran, "Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices," *Marine Chemistry*, vol. 89, pp. 15-36, 2004/10/01/ 2004.

[91] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

Chapter 6. APPENDIX A

6.1 WILDFIRE MODELS:

Using PCR for wildfire samples gives the model equations (6.1) and (6.2) for unfiltered and filtered samples.

$$\mathrm{DTT}_{(\mathrm{pred})}^{Unfiltered} = -0.241 \times \mathrm{PC}_1 - 0.181 \times \mathrm{PC}_2 - 0.452 \times \mathrm{PC}_3 + 5.027 \quad (6.1)$$

$$\mathrm{DTT}_{(\mathrm{pred})}^{Filtered} = 2.201 \times \mathrm{PC}_1 - 0.402 \times \mathrm{PC}_2 - 0.106 \times \mathrm{PC}_3 + 4.93 \quad (6.2)$$

Figure 6.1**Error! Reference source not found.** plots predicted vs. actual DTT consumption for unfiltered and filtered wildfire models. The unfiltered model has the worse R^2 (test) = 0.174, which could be because of the noise in the unfiltered wildfire data, while the filtered model shows a good R^2 (test) = 0.977.

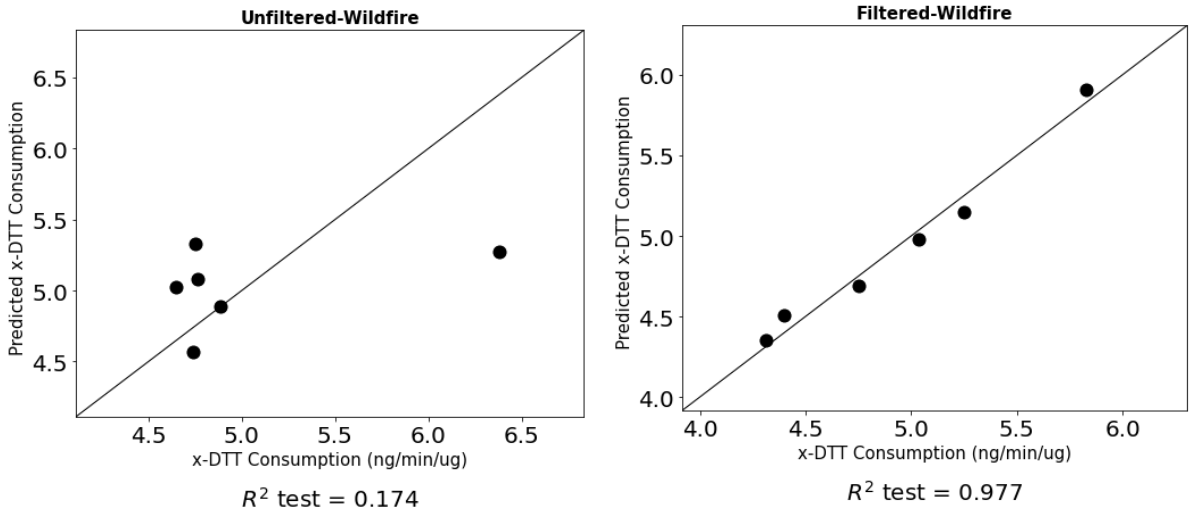

Figure 6.1. Parity plots showing results of PCR with the model for (a) unfiltered and (b) filtered wildfire samples.

Figure 6.2Figure 6.2**Error! Reference source not found.** shows how the wildfire model predicts the ROS activity in natural-draft cookstove samples using EEM and PCR for unfiltered and filtered samples. Black dots are for the wildfire, and red squares are for the natural-draft

cookstove samples. Here, we see that the model over-predicts the ROS activity for natural-draft cookstove samples, but the unfiltered wildfire model itself is not good.

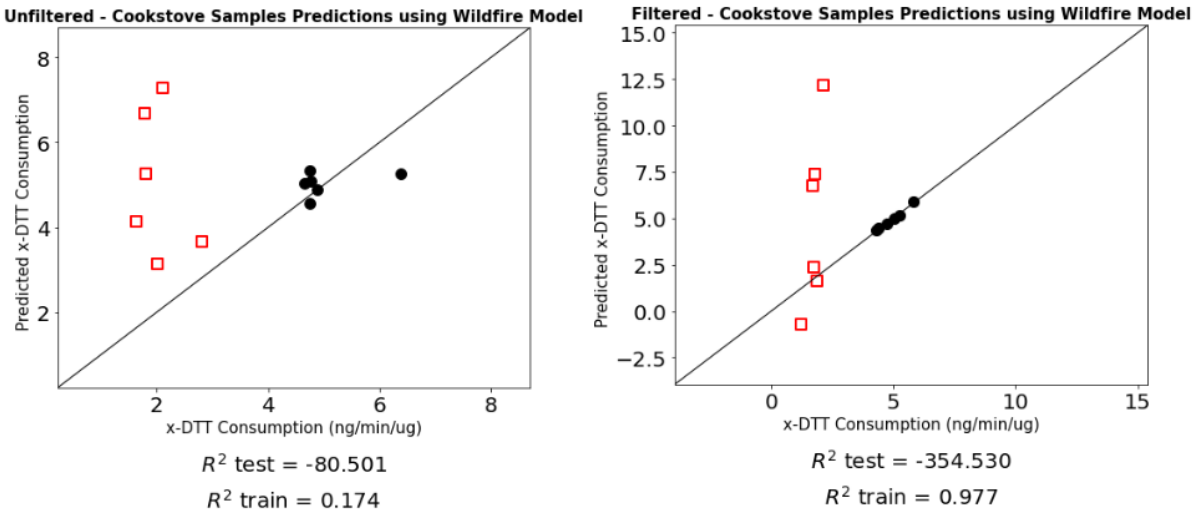

Figure 6.2. Parity plots showing results of DTT prediction of natural-draft cookstove samples using wildfire model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (wildfire samples), and red squares plot the test data (natural-draft cookstove)

Figure 6.3 shows how the wildfire model predicts the ROS activity in forced-draft cookstove samples using EEM and PCR for unfiltered and filtered samples. Black dots are for the wildfire, and red squares are for the forced-draft cookstove samples. Here, the model over-predicts the ROS activity for forced-draft cookstove samples.

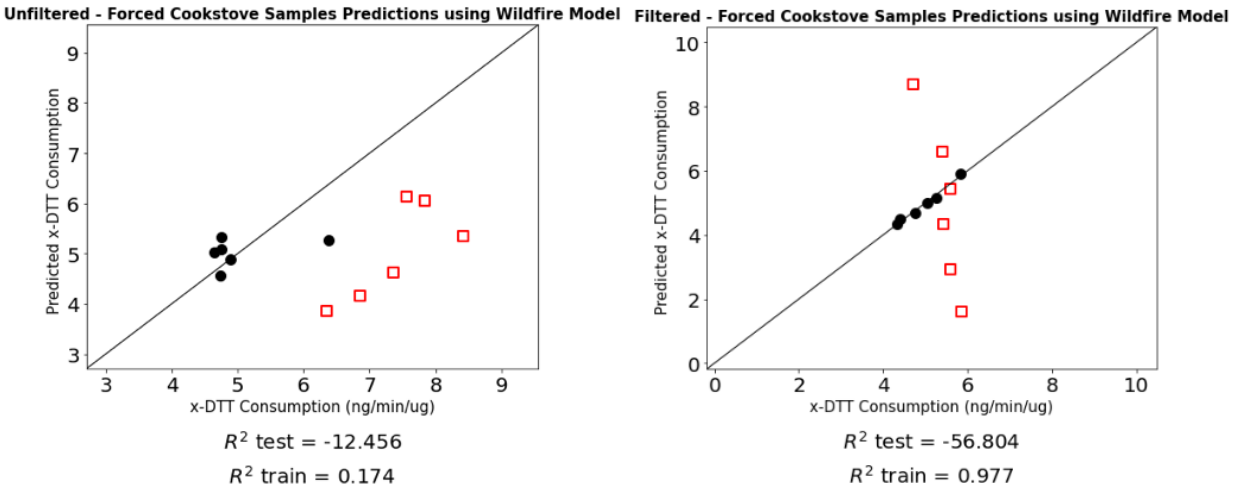

Figure 6.3. Parity plots showing results of DTT prediction of natural-draft cookstove samples using wildfire model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (wildfire samples), and red squares plot the test data (natural-draft cookstove)

6.2 FORCED-DRAFT COOKSTOVE MODELS:

PCR for forced-draft cookstove samples gives the model equations (6.3) and (6.4) for unfiltered and filtered samples.

$$\mathrm{DTT}_{(\mathrm{pred})}^{Unfiltered} = -0.155 \times \mathrm{PC}_1 + 0.162 \times \mathrm{PC}_2 - 0.329 \times \mathrm{PC}_3 + 7.39 \quad (6.3)$$

$$\mathrm{DTT}_{(\mathrm{pred})}^{Filtered} = -0.293 \times \mathrm{PC}_1 + 0.155 \times \mathrm{PC}_2 - 0.106 \times \mathrm{PC}_3 + 5.42 \quad (6.4)$$

Figure 6.4 plots predicted vs. actual DTT consumption for unfiltered and filtered forced-draft cookstove models. The unfiltered model has a good R^2 (test) = 0.852, while the filtered model shows a good R^2 (test) = 0.849.

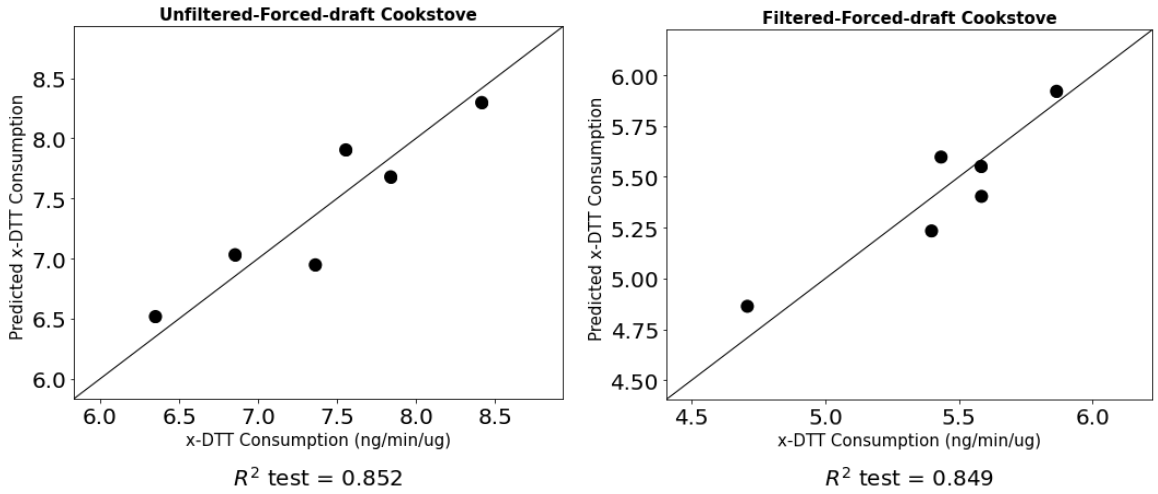

Figure 6.4. Parity plots showing results of PCR with the model for (a) unfiltered and (b) filtered forced-draft cookstove samples.

Figure 6.5 shows how the forced-draft cookstove model predicts the ROS activity in wildfire samples using EEM and PCR for unfiltered and filtered samples. Black dots are for the forced-draft cookstove, and red squares are for the wildfire samples. Here, the model over-predicts the ROS activity for wildfire samples.

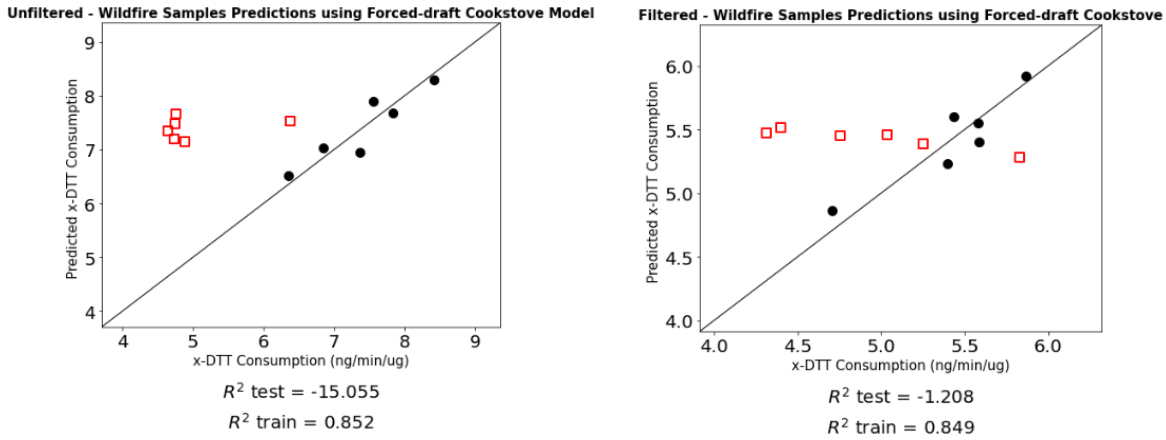

Figure 6.5. Parity plots showing results of DTT prediction of wildfire samples using forced-draft cookstove model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (forced-draft cookstove samples) and red squares plot the test data (wildfire).

Figure 6.6 shows how the forced-draft cookstove model predicts the ROS activity in natural-draft cookstove samples using EEM and PCR for unfiltered and filtered samples. Black

dots are for the forced-draft cookstove, and red squares are for the natural-draft cookstove samples. Here, we find that the model over-predicts the ROS activity for natural-draft cookstove samples.

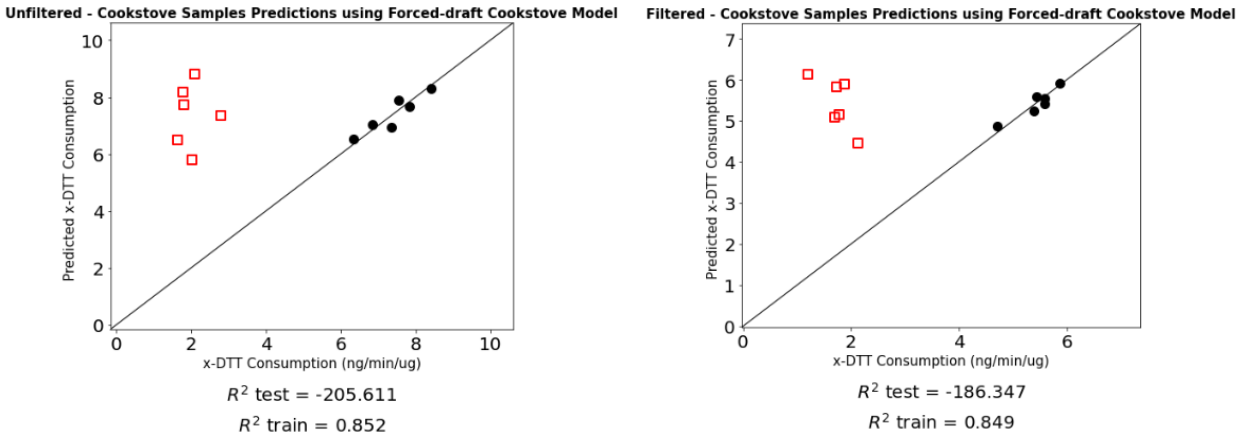

Figure 6.6. Parity plots showing results of DTT prediction of natural-draft cookstove samples using forced-draft cookstove model for (a) unfiltered and (b) filtered samples. Black dots plot the training data (forced-draft cookstove samples) and red squares plot the test data (natural-draft cookstove).