Atmospheric Environment 220 (2020) 117065

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: http://www.elsevier.com/locate/atmosenv

Excitation emission matrix fluorescence spectroscopy for combustion generated particulate matter source identification

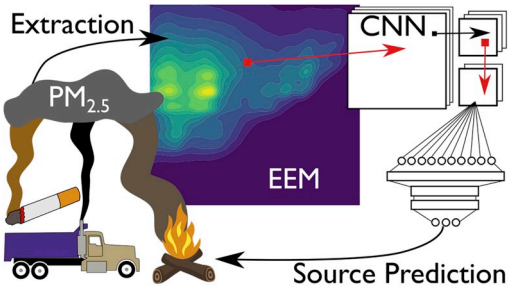
Jay W. Rutherford [a], Neal Dawson-Elli [a], Anne M. Manicone [e], Gregory V. Korshin [b], Igor V. Novosselov [b], Edmund Seto [d], Jonathan D. Posner [a,b,c,*]

[a] Department of Chemical Engineering, University of Washington, United States
[b] Department of Mechanical Engineering, University of Washington, United States
[c] Department of Family Medicine, University of Washington, United States
[d] Environmental and Occupational Health Sciences, University of Washington, United States
[e] Department of Medicine: Pulmonary, Critical Care and Sleep Medicine, University of Washington, United States

HIGHLIGHTS

- Excitation Emission Matrix (EEM) fluorescence used for source apportionment.
- Convolutional Neural Network used to interpret EEM spectra.
- Predicted the presence or absence of three sources with an overall accuracy of 89%.
- Limit of detection in air from 0.7 to 2.6 μg/m^3 for 24-h sample at 1.8 lpm.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Fluorescence
Source apportionment
Particulate matter
Diesel
Woodsmoke
Neural network

ABSTRACT

The inhalation of particulate matter (PM) is a significant health risk associated with reduced life expectancy due to increased cardio-pulmonary disease and exacerbation of respiratory diseases such as asthma and pneumonia. PM originates from natural and anthropogenic sources including combustion engines, cigarettes, agricultural burning, and forest fires. Identifying the source of PM can inform effective mitigation strategies and policies, but this is difficult to do using current techniques. Here we present a method for identifying PM source using excitation emission matrix (EEM) fluorescence spectroscopy and a machine learning algorithm. We collected combustion generated $PM_{2.5}$ from wood burning, diesel exhaust, and cigarettes using filters. Filters were weighted to determine mass concentration followed by extraction into cyclohexane and analysis by EEM fluorescence spectroscopy. Spectra obtained from each source served as training data for a convolutional neural network (CNN) used for source identification in mixed samples. This method can predict the presence or absence of the three laboratory sources with an overall accuracy of 89% when the threshold for classifying a source as present is 1.1 μg/m^3 in air over a 24-h sampling time. The limit of detection for cigarette, diesel and wood are

* Corresponding author. Department of Family Medicine, University of Washington, United States.
E-mail address: jposner@uw.edu (J.D. Posner).

https://doi.org/10.1016/j.atmosenv.2019.117065
Received 5 August 2019; Received in revised form 14 October 2019; Accepted 17 October 2019
Available online 22 October 2019

0.7, 2.6, 0.9 μg/m^3, respectively, in air assuming a 24-h sampling time at an air sampling rate of 1.8 L per minute. We applied the CNN algorithm developed using the laboratory training data to a small set of field samples and found the algorithm was effective in some cases but would require a training data set containing more samples to be more broadly applicable.

1. Introduction

According to the Global Burden of Disease study, air pollution is the world's largest environmental health risk accounting for 4.9 million deaths and 147 million disability adjusted life years in 2017 (Stanaway et al., 2018). Particulate matter (PM) originates from a wide range of natural and anthropogenic sources. Most PM is a result of emissions directly from sources such as diesel engines, agricultural burning, cooking with biomass, electrical power generation, pollen, bacteria, and soil. In addition to these primary sources, PM results from chemical processes in the atmosphere, referred to as secondary sources (Godish, 1997). In this work we focus on PM from cigarettes, diesel and woodsmoke due to their relevance to health. Cigarette smoking is estimated to cause 7.1 million deaths each year and second-hand smoke is estimated to cause 1.2 million deaths. Solid fuel combustion for cooking is a significant contributor to indoor air pollution and was estimated to be responsible for 1.6 million deaths in 2017. Finally, ambient PM was estimated to account for 2.9 million deaths (Stanaway et al., 2018; Institute for Health Metrics and Evaluation (IHME), 2018). Diesel exhaust is a contributor to transportation emissions, which are estimated to account for 25% of total ambient PM pollution in large cities (Karagulian et al., 2015). The human respiratory system is effective at removing many of the particles that enter the respiratory tract before they enter deep into the lungs. Larger particles are removed by impaction and sedimentation in the upper respiratory tract and branching airways of the lungs while smaller particles can penetrate deeper into the lungs (Tsuda et al., 2013). A size cut-off of 2.5 μm in aerodynamic equivalent diameter ($PM_{2.5}$) and smaller has been established for the purpose of regulation of PM pollution (Miller et al., 1979). The USEPA standards for maximum exposure levels are 12 μg/m^3 annual average and 35 μg/m^3 daily average (US EPA, 2018). The World Health Organization (WHO) guidelines are lower at 10 and 25 μg/m^3 , respectively (World Health Organization, 2006). Studies suggest that some sources of PM are worse for health than others. For example, black carbon, which is associated with traffic, has been shown to be worse for health than $PM_{2.5}$ alone (Janssen et al., 2011; Bell et al., 2014), but the body of evidence as a whole does not conclusively show what sources or combinations thereof are the worst for health (Stanek et al., 2011; West et al., 2016; Adams et al., 2015; Hime et al., 2018).

Asthma is known to be exacerbated by $PM_{2.5}$ exposure (Koenig, 1999). In practice, clinicians advise asthmatic patients to avoid exposure to pollution that they are sensitive to and to avoid exertion outdoors when air quality is poor (National Asthma Education and Prevention Program, 2007). A study using parental questionnaires and proximity to roadways found increased asthma risk in children with exposure to second hand smoke, but not with roadway proximity (Lewis et al., 2005). User-friendly and inexpensive tools for monitoring source specific PM exposure will enable continued and more quantitative epidemiological research in the area of source specific health impacts and may enable regulations targeting the worst sources of PM pollution (Duncan et al., 2018).

In this work, we use excitation-emission matrix (EEM) fluorescent spectroscopy and machine learning to identify the source of PM. Fluorescence spectroscopy is a sensitive analytical technique with the ability to detect fluorescence from a single molecule using sophisticated instrumentation (Moerner and Fromm, 2003). With widely available benchtop fluorimeters limits of detection are around 1 ng/mL for polycyclic aromatic hydrocarbons, a common chemical component of PM air pollution (Elcoroaristizabal et al., 2014a; Nahorniak and Booksh, 2006). For this reason, fluorescence spectroscopy is an attractive analytical technique for PM analysis due to typical sample sizes of PM being small. Although fluorescence is a very sensitive technique it is not highly specific due to many analytes having overlapping spectra. EEM spectroscopy can increase the specificity of fluorescence spectroscopy by collecting fluorescent emission spectra at multiple excitation wavelengths, giving a 2D dataset or matrix of fluorescence intensities (Johnson et al., 1977). EEM spectroscopy has been widely applied to analysis of complex environmental water samples (Andrade-Eiroa et al., 2013) as well as analysis of atmospheric PM (Aryal et al., 2015; Chen et al., 2016; Elcoroaristizabal et al., 2014b; Matos et al., 2015; Mladenov et al., 2009, 2011; Nakajima et al., 2008). Mladenov et al. suggested EEM could be useful as a source identification tool for atmospheric PM, but did not evaluate the ability of EEM alone to identify sources (Mladenov et al., 2011). Other work applying EEM to atmospheric aerosols discusses the chemical composition of various regions of fluorescence, but does not discuss using EEM as a source apportionment or identification tool (Aryal et al., 2015; Chen et al., 2016; Elcoroaristizabal et al., 2014b; Matos et al., 2015; Mladenov et al., 2009; Nakajima et al., 2008).

EEMs provide complex spectral information consisting of thousands of wavelength dependent fluorescent intensities (~20,000 datapoints for the EEMs in this work), as such, a variety of approaches have been used to interpret EEM spectra. Fluorescent regional integration considers specific regions of an EEM spectrum based on compounds of interest that display fluorescence in various regions (e.g. aromatic protein and humic acid like regions when applied to water samples) (Chen et al., 2016). This approach has the advantage of simplicity, but it is unable to distinguish overlapping spectra which was a challenge in our samples. Techniques used for interpreting EEM spectra that can handle overlapping spectra include partial least squares regression (PLS), parallel factor analysis (PARAFAC), and multivariate curve resolution (MCR). These techniques have been used successfully to identify specific chemical components of atmospheric PM, but have not been used for source apportionment (Elcoroaristizabal et al., 2014a; Aryal et al., 2015; Matos et al., 2015). In this work we used a convolutional neural network (CNN), a machine-learning technique that is well suited to handle 2D data like EEM spectra. Machine learning techniques, like a CNN, do not rely on underlying theory or assumptions to create a model, instead these techniques use data inputs and expected outputs, referred to as training data, to create a non-parametric model. The model framework is defined and then the training data are used to iteratively adjust model parameters to best fit the training data, this is the learning or training process.

A CNN is a combination of a neural network (Hastie, 2013) preceded by a feature recognition process called convolution (Goodfellow et al., 2016). Individual steps in the network are called layers. The first layers are convolutional layers consisting of filters of user-defined size which are iteratively scanned, or convolved, across the input data. The filter values are randomly initialized and adjusted to identify relevant features during the training process. CNNs excel at processing 2D data (e.g. image classification (Simonyan and Zisserman, 2014; Krizhevsky et al., 2012)) due to their ability to learn patterns or features encoded in spatial data using the convolutional layers. For example, the convolutional layers of an image classification CNN may recognize features such as noses, eyes, and ears when classifying images of animals (Zeiler and Fergus, 2013; Lee et al., 2009). The result of the convolution process, referred to as a feature map, is fed into a neural network to map the learned patterns and features to an output value or classification. In the animal classification example, a feature map showing the presence of

floppy ears would suggest the image was a dog. However, all features in the map are considered, so if there are also a trunk and tusks, the image would be identified as an elephant. In the case of EEM data the feature maps will highlight important spectral patterns.

Neural networks and CNNs have been applied to a variety of PM exposure related problems. For example, a CNN has been used to predict continuous $PM_{2.5}$ concentrations from discrete measurements (Di et al., 2016). Source apportionment has been conducted using a neural network with elemental composition as the input, which is similar to our work, but the input data was 1D elemental data so convolution layers are not used (Song and Hopke, 1996). Neural networks have been used to process EEM spectra for analysis of water samples, contaminates in olive oil, and antibiotics in urine (Carstea et al., 2010; Bieroza et al., 2009; García-Reiriz et al., 2007, 2008). In these examples, neural networks are used because of their ability to fit non-linear behavior without the need for assumptions about the underlying data and their ability to interpret information from the entire EEM, both of which are applicable to our work.

In this paper, we demonstrate the ability of EEM coupled with a CNN to identify PM from woodsmoke, cigarettes and diesel exhaust with a limit of detection of 2.6 μg/m^3 based on a 24-h sampling time. $PM_{2.5}$ samples were collected on PTFE filters using personal sampling devices. Samples were weighed to determine $PM_{2.5}$ mass concentration and then extracted in cyclohexane. The cyclohexane extracts were analyzed using EEM fluorescence. The spectra from the three sources show unique but overlapping fluorescent EEM spectra. We applied a CNN to identify the presence or absence of the three sources present in a set of EEM spectra consisting of one, two, or all three sources. We achieved an overall accuracy of 89% in identifying sources. This technique has a limit of detection well below the USEPA and WHO recommended exposure levels for PM and may be useful for personal monitoring in epidemiological studies of respiratory diseases such as asthma.

2. Materials and methods

2.1. Particulate matter sampling and extraction

We sampled $PM_{2.5}$ from cigarettes, diesel exhaust, and woodsmoke using 2.0 μm pore PTFE membrane filters (Pall Zefluor®, Pall Cat. #P5PJ037) housed in Harvard School of Public Health Personal Exposure Monitor (BGI, Butler, NJ Cat. # HP2518) sampling cassettes. Filters were operated at a flowrate of 1.8 lpm using either portable or stationary vacuum pumps (AirChek XR5000 pump, SKC Inc., Eighty Four, PA or VP0625-V1014-D2-0511, Medo USA, Roselle, IL with custom manifold of nine VFB-65-BV roto-meters, Dwyer Instrument, Michigan City, IN). Flow rates were verified using an air flow calibrator (Gilian Gilibrator PN# 800268, Sensidyne, St. Petersburg, FL).

We collected woodsmoke by burning 1 ½ by ¾ inch Douglas fir sticks cut from dimensional lumber in a prototype side-feed, natural-draft, improved cookstove. Our sampling devices were placed in a sealed chamber connected to the exhaust hood duct at the sampling point described for gravimetric sampling by Sullivan et al. (Sullivan et al., 2017) We collected diesel exhaust particulate from the exposure room in UW's controlled inhalation diesel exhaust exposure facility (Gould et al., 2008). The diesel PM is generated by a 435 cc direct injection single cylinder diesel engine (Yanmar LW Series) fueled with ultra-low-sulfur diesel. We collected cigarette smoke either by lighting cigarettes in a fume hood and allowing them to smolder or from the exposure chamber of a cigarette smoking machine (Model TE-10B, Teague Enterprises, Woodland, CA). The TE-10B produces mainstream smoke mixed with sidestream smoke from filtered 3R4F research cigarettes (Tobacco Research Institute, University of Kentucky, Lexington, KY). Two cigarettes were puffed simultaneously for 2 s for a total of 8 puffs, at a flow rate of 1.05 l/min. The smoke collected represents approximately 10% mainstream and 90% sidestream to more closely resemble secondhand smoke.

Following collection, filters are removed from the samplers and placed in a chamber with 37% (SD = 4%) relative humidity for 24 h (Allen et al., 2001). The filters were then weighed using a micro-balance with 0.5 μg resolution (Mettler-Toledo UMT-2, Greifensee, Switzerland). Initial weights of each filter are recorded in the same manner and we use the difference to calculate the mass of $PM_{2.5}$ collected.

We placed the filters into 20 mL glass vials (Cat # 89096-774 VWR, Edison, NJ), submerged the filters in cyclohexane (Uvasol® Cyclohexane for Spectroscopy, MilliporeSigma Cat. #1.02822.2500), and sonicated for 30 min (42 kHz, 2510R-MT Branson, Ultrasonic Corp., Danbury, CT). Filters were generally submerged in ~10 mL of cyclohexane to achieve an initial extract concentration of 5 μg PM/mL cyclohexane or greater. For filters with low PM loading we cut the filter into fourths to enable extraction in as little as 3 mL of cyclohexane to maintain extract concentrations at or above 5 μg/mL. Typically, the PM was not dislodged from the filter during extraction allowing for direct analysis of the extract. If significant PM was dislodged and suspended causing turbidity, the extract was filtered with a 0.2 μm PTFE syringe filter (VWR Cat. #28145-491) before analysis.

We collected a total of 37 filter samples and used the extracts from these filters to create 113 samples for EEM analysis. They consisted of 81 single source samples diluted to concentrations between 0.2 μg/mL and 10 μg/mL, 21 mixtures of the single source samples and five samples from filters with mixed PM from serial sampling of the sources. We also collected six spectra from liquid extracts of filters that were loaded into sampling devices and weighed, but no air was drawn through the filters (method blanks). For training our algorithm, we used 12 of the 113 EEM samples leaving a total of 101 samples for testing the algorithm. Table 1 summarizes the total number of each type of sample used for training and testing.

In addition to the samples collected in the laboratory, we collected twelve field samples to evaluate our method on real world samples. Eight field samples were taken in Seattle homes and in campus buildings (*background* field samples) and four were collected in areas we expected to be dominated by cigarette, diesel, or woodsmoke (*expected primary source* field samples). We collected and extracted the field samples using the same equipment and methods as the laboratory samples. EEM spectra collected from these two groups of field spectra are shown in Fig. S7.

2.2. Fluorescence EEM analysis

PM extracts were stored in 4 ml vials (Cat # 66009-876 VWR, Edison, NJ) until analysis. For EEM spectroscopy ~3 ml of PM extract was transferred to a 1 cm × 1 cm quartz cuvette (Item # CV10Q3500FS, Thorlabs Inc., Newton, New Jersey). We collected EEM data using a fluorometer with an extended-UV 150W xenon-arc lamp (Aqualog-880-C, HORIBA Instruments Inc. Edison, New Jersey). We excited samples between 200 and 500 nm at 2 nm increments with an excitation slit

Table 1
Number of unique filter samples and liquid extract samples generated from extracts, dilutions of extracts and mixtures of extracts for each category of sample. We collected a total of 113 EEM spectra. Twelve of these spectra were used to generate training data leaving 101 spectra in the test set.

Sample Type	Number of Unique Filters	Samples for EEM	Spectra used for training
Cigarette	9	26	4
Diesel	10	29	4
Woodsmoke	9	26	4
Extract Mixtures	N/A*	21	0
Multiple-Exposure	5	5	0
Method Blanks	4	6	0
Total	**37**	**113**	**12**

*Mixtures of Cigarette, Diesel, and Woodsmoke samples.

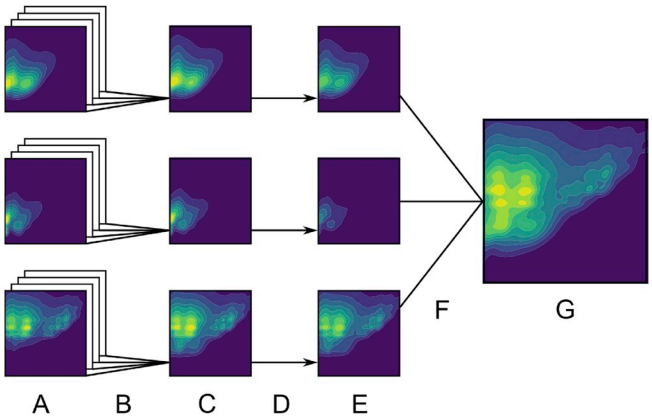


Fig. 1. Graphical representation of the loop used to generate training spectra. Twelve original spectra, four from each of cigarette, diesel, and woodsmoke, (A) are averaged using random weights (B) producing prototypical spectra for each source (C) that are scaled (D) giving single source spectra at target concentration (E) that are combined (F) to generate a training spectra (G). This loop is repeated from the beginning for each training spectra generated in order to simulate variability associated with PM sampling and EEM collection.

width of 5 nm and recorded emission spectra between 246 and 826 nm on a CCD array. The CCD array has 1000 pixels each covering 0.58 nm. We collected data using 4-pixel binning giving an effective emission slit width of 2.32 nm. We kept emission data between 246 and 572 nm and excitation data between 224 and 500 nm. Emission data above 572 nm were discarded because minimal fluorescence was observed above this wavelength and excitation wavelengths below 224 nm were removed due to low excitation lamp intensity between 200 and 224 nm yielding extensive noise in the data. The raw fluorescent signal is corrected for detector response and lamp intensity by the instrument (Engelborghs et al., 2014), and is normalized to Raman units using Raman area data collected daily from a Milli-Q water sample (Murphy et al., 2010). Daily solvent blanks are recorded and used for blank subtraction to minimize the effect of Rayleigh and Raman scatter. To further reduce the effects of Rayleigh scatter we excised values within 10 nm of the first and second order Rayleigh scattering bands followed by replacement of the values using 2-dimmensional interpolation (Zepp et al., 2004). We did not correct for the inner filter effect because we observed absorbance below 0.2 for our samples that were recorded with the Aqualog during EEM collection.

2.3. Machine learning for identification of sources present

We used a CNN to identify the presence or absence of known PM sources in the EEM spectra. The CNN was trained on 6375 training spectra generated from twelve single source spectra from cigarette, diesel, and woodsmoke PM (four from each source). We generated the training dataset using a data augmentation approach by mathematical combination of twelve original spectra, assuming fluorescence is linearly proportional to concentration (see section S3 for data supporting this assumption). The process for generating spectra is shown schematically in Fig. 1 and described in more detail in section S4. First, we created 1000 spectra for cigarette, diesel, and woodsmoke in a linearly spaced concentration range from 0 to 5 μg/mL resulting in 3000 single-source spectra. Then we created digital mixtures of the three sources in a logarithmically spaced concentration range from 0.01 to 6.3 in fifteen steps (15^3 combinations) giving 3375 training spectra consisting of mixtures. In creating mixtures by mathematically combining spectra from pure sources we assumed matrix effects of mixing to be negligible. We showed this to be a reasonable assumption by comparing digital and actual mixtures as illustrated in Fig. S5.

The CNN used in this work consists of three convolutional layers each followed by max pooling (Dumoulin and Visin, 2016), as shown in Fig. 2. All convolutions are performed using padding, so the dimensions of input and output data are the same (Dumoulin and Visin, 2016). The first convolutional layer contains twenty 5-by-5 filters equating to 11.6 nm in emission (height) and 10 nm in excitation (width) as shown by the red box (Fig. 2a). This is followed with 3-by-3 max pooling that reduces the data from 143-by-139 to 47-by-46. The second convolutional layer is ten 10-by-10 filters followed by 3-by-3 max pooling. The final convolutional layer applies ten 15-by-15 filters to the 15-by-15 feature maps. The output of the third convolutional layer is max pooled to a size of 5-by-5 and then flattened and connected to a dense neural network with three hidden layers having 512, 256, and 256 nodes in each layer, respectively. A dropout rate of 20% is used between all convolutional and fully connected layers (Srivastava et al., 2014). The exponential linear unit was used as the activation function for all convolutional and hidden layers and a linear activation function was used for the output layer (Clevert et al., 2015), the loss function was the mean-squared-error (Hastie, 2013). The results described are from a network that was trained for 80 epochs. Details of how we selected the training duration are included in the SI. The CNN was implemented in Python 3 using Keras (Chollet, 2018) and TensorFlow (Abadi et al., 2015).

Our EEMs are 2D spatial data made up of combinations of peaks and valleys, which correlate to a particular chemical or combination of chemicals that are extracted from the PM samples. These peaks and valleys vary in their intensity across emission and excitation dimensions. The convolution filters learn to fit these varying shapes to better detect peak presence as they are iteratively applied over the EEM. Subsequent convolutional layers are used to identify patterns of lower level features, for example, a second convolutional layer may look for a group of narrow peaks identified in the previous convolutional layer. The results of the convolutional layers are feature maps showing the presence or absence of features. The feature map data are fed into fully connected layers that map this information to the desired output. In our case this process assigns a predicted concentration value for each source that we use to predict the presence or absence of the sources.

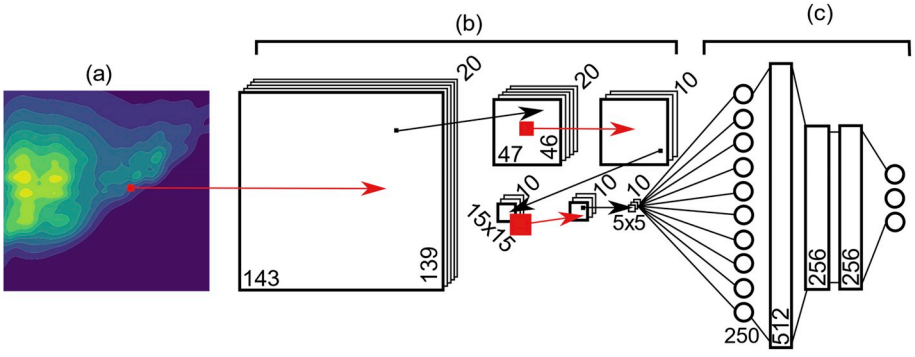


Fig. 2. CNN Network Diagram. (a) Input spectra are first convolved with twenty 5-by-5 filters. (b) Convolutional layers and max pooling layers are shown with associated data shapes. Convolutions (10×10 followed by 15×15) are shown in red and 3-by-3 max pooling is shown in black. Filters, data, and pooling sizes are shown to scale. (c) Output of the convolutional layers is flattened to a shape of 250 by 1 and fed into fully connected layers resulting in 3 output values (not to scale). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

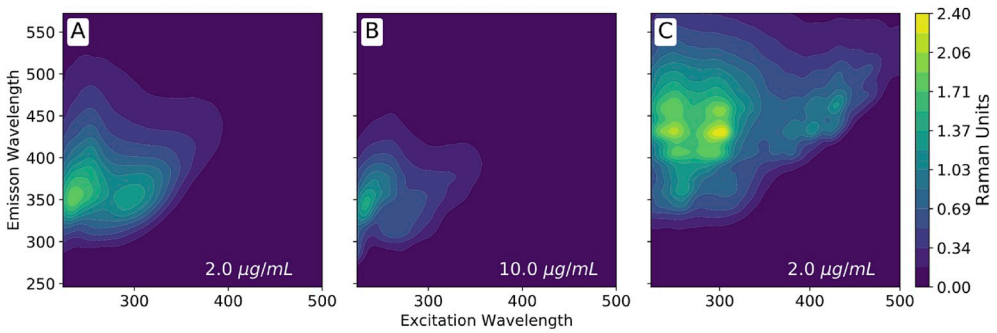


Fig. 3. Fluorescence EEM spectra of (a) cigarette smoke at an extract concentration of 2 μg/ml, (b) diesel soot at 10 μg/ml, and (c) woodsmoke at 2 μg/ml. PM samples from the three sources were extracted in cyclohexane and exhibit unique spectral fingerprints. Cigarette and woodsmoke have similar maximum fluorescent intensity on a per mass basis while diesel has a lower signal intensity.

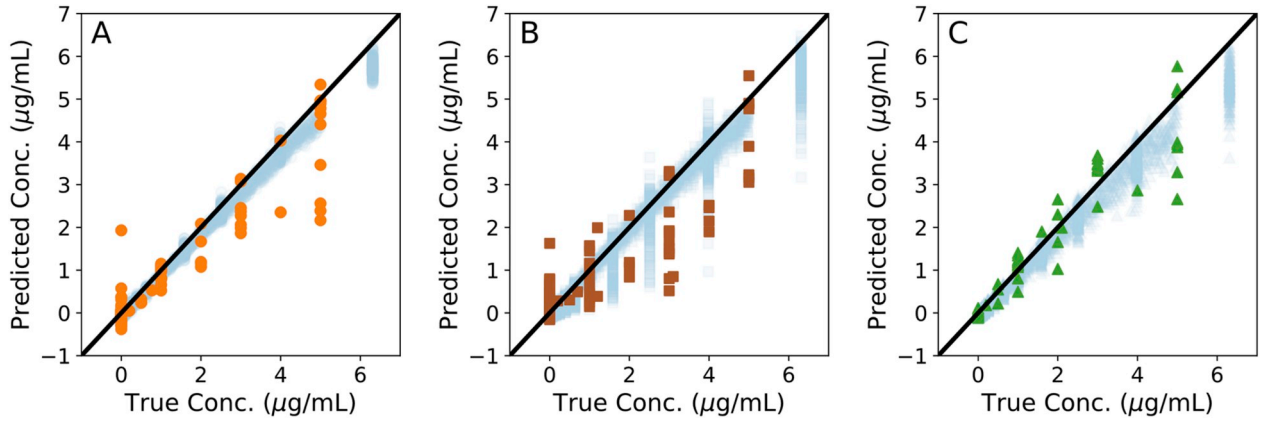


Fig. 4. Parity plots showing predicted concentration vs. true extract concentrations for (a) cigarette (R^2 training = 0.99, R^2 test = 0.86), (b) diesel (R^2 training = 0.96, R^2 test = 0.79) and (c) woodsmoke (R^2 training = 0.97, R^2 test = 0.89). The data points shown as solid colors (orange, brown and green) are from 101 test spectra, the points shown in light blue are the results for the simulated training data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3. Results and discussion

PM extracts from cigarette, woodsmoke, and diesel show unique EEM spectra, as shown in Fig. 3. We also extracted cigarette and woodsmoke in methanol and water for comparison as shown in Fig. S11. Cigarette spectra consist of two peaks at ~350 nm emission wavelength. Diesel spectra consist of single primary peak also located at ~350 nm emission with less fluorescence surrounding the peak than cigarette. Woodsmoke has spectra consisting of six peaks in the region from 400 to 475 nm emission and 225 to 275 nm excitation. Cigarette and woodsmoke have similar maximum fluorescent intensity levels per mass of PM while diesel has a lower intensity. Woodsmoke shows fluorescence over the broadest region and generally at higher emission wavelengths than cigarette and diesel. The differences in the spectra are the result of differences in chemical composition of the PM samples. Combustion generated PM contains thousands of chemical compounds, including polycyclic aromatic hydrocarbons (PAHs) (Samburova et al., 2016; Lin et al., 2018). We do not attempt to quantify these underlying differences on a compound-by-compound level, but instead use the combined effect to identify the sources of interest. PAH compounds with higher molecular weight absorb light and show fluorescence at higher wavelengths (Rieger and Müllen, 2010). Based on this approximation, woodsmoke contains compounds of higher molecular weight than cigarette and diesel; however, other factors, such as presence of functional groups (e.g. amino and hydroxyl), may modify the absorbance and fluorescence so we do not make quantitative assessments of molecular weight based on EEM alone (Rieger and Müllen, 2010; Birks, 1970). The spectra from the different sources have overlapping regions suggesting challenges in distinguishing individual sources from mixed samples.

The CNN algorithm was trained on EEM spectra with known particulate concentrations generated as described in the methods section. True concentration values are determined from PM mass measurement and extraction volume. We then provide the CNN this same training data and ask it to predict the concentration of the three sources. These results are shown in Fig. 4. The diagonal line represents perfect prediction of the samples where the CNN prediction values are equal to the values provided during training. The data points that result from the analysis of the original training data roughly follow the diagonal. The R^2 value for the fit to the training data for cigarette, diesel, and woodsmoke are 0.99, 0.97, 0.97 respectively. One reason for scatter in the training data is extracts at the same particulate matter concentration have different fluorescent signal strengths. This variation in signal strength is shown in Fig. S4 which plots fluorescent intensity vs. concentration for single source spectra.

We then predict the concentration of the 101 test spectra which are shown in the parity plots of Fig. 4. The results generally follow the diagonal trend, but there are significant under- and over-predictions. This can be attributed to the fact that total fluorescent intensity from a given source varies from sample to sample at the same concentration. The R^2 value for the fit to the test data for cigarette is 0.86, for diesel it is 0.79, and for woodsmoke it is 0.89. The lower R^2 values for the test data are due to the overlap of the signals (Fig. 4) that makes mixtures difficult to quantify and variation in fluorescent signal intensity among samples at the same concentration (Fig. S4).

We used the predicted results for samples containing only a single source to determine the limit of detection (LoD). The LoD for each source

Table 2
LoD determined by applying the CNN model to single source samples. The column reporting LoD in μg/mL is determined using PM mass measurement of filters dispersed in a volume of cyclohexane. The column reporting μg/m^3 in air is determined by converting the LoD in μg/mL to μg/m^3 assuming a 24-h sampling time at an air sampling rate of 1.8 L per minute.

Source	LoD [μg/mL cyclohexane]	LoD [μg/m^3 air]
Cigarette	0.6	0.7
Diesel	2.2	2.6
Woodsmoke	0.8	0.9

was evaluated according to the Clinical and Laboratory Standards Institute method as described in the supporting information (Tholen, 2004; Borysiak et al., 2016). The measured LoD for each source is provided in Table 2. Here we provide the detection limit in mass of particulate matter per volume of extraction liquid as well as a calculated particulate matter concentration per volume of sampled air, assuming a 24-h sampling time at 1.8 L per minute (see section S1 for calculation details). Diesel has the weakest fluorescence intensity and thus the highest LoD of 2.2 μg/mL cyclohexane or 2.6 μg/m^3 air. The LoD for each source in a 24-h sampling period is significantly lower than the WHO and USEPA 24-h mean exposure guidelines of 25 and 35 μg/m^3 respectively (Aable of Histo, 2018; World Health Organization, 2006).

The ability to identify if PM from a source is present or absent above a threshold level could be a useful tool for clinicians and asthma patients in treating asthma or for asthma research, for example. To this end, we evaluated the ability of the CNN analysis of EEM spectra to detect the presence of individual sources above a threshold of 1 μg/mL. This threshold corresponds to an average exposure of nearly 10 μg/m^3, the WHO annual average guideline, during a 3-h sampling period at 1.8 L/min. In Fig. 5, we plot the predicted concentration of each source in either a negative or positive column. Samples are considered positive if they had a true concentration (measured gravimetrically) of 1 μg/mL or greater of any of the single sources, and negative if they are below this concentration. This analysis method is based on the establishment of a cut off value for a qualitative diagnostic health test (Borysiak et al., 2016). The clinical sensitivity, specificity, and overall accuracy of the diagnostic is then determined by choosing a threshold which delineates the positive from negative results. Depending on the purpose of the diagnostic test, the threshold may be set to achieve a specific outcome. For example, in the case of screening for a deadly but treatable disease, the number of false negatives would be minimized (i.e. maximizing sensitivity) (Borysiak et al., 2016). In this work, we choose the threshold that maximizes the accuracy for each source. Fig. 5D shows a plot of source detection accuracy as a function of the calibrated threshold value used as a cut off between positive and negative detection. This plot shows that as we increase calibrated threshold value the detection accuracy for each source increases to a maximum and then decreases because as the threshold increases nearly all positive samples are classified as negative. The threshold of maximum accuracy varies with the source. The predicted concentration thresholds for maximum detection accuracy for cigarette, diesel, and woodsmoke are 0.6, 0.8, and 0.7 μg/mL, respectively. These thresholds are shown by red horizontal lines in Fig. 5A–C and when applied, we achieve an overall accuracy of 89%. The accuracies for identifying cigarette and woodsmoke were 98% and 99% respectively. Diesel was more challenging because of its low signal intensity relative to the other sources and had one false positive and seven false negatives giving an accuracy of 92%.

After setting threshold values for source classification using all 101 test spectra, we evaluated the model performance on sub-groups of the test set. The sub-group of spectra containing single sources consisted of sixty-nine spectra from sixteen filter samples ranging in concentration from 0.2 μg/mL to 10 μg/mL. Within this group we classified the samples with an overall accuracy of 91%. Cigarette and woodsmoke spectra were identified with the best results while diesel was the most often misclassified with a sensitivity of 0.91 and a specificity of 0.98. Next, we tested the algorithm on the sub-group of test spectra containing two or more sources. Twenty-one samples were generated by mixing liquid extracts together and five were from exposing an individual filter to multiple PM sources. The results of the analysis are shown in Table 3. The CNN algorithm was able to identify the sources present in mixed samples with an overall accuracy of 81%. The sensitivity and specificity for cigarette and woodsmoke was perfect; however, diesel continued to show challenges with a specificity of unity and a sensitivity of 0.75. The relatively low sensitivity of diesel is a result of the diesel spectra being weaker than and overlapping with the other sources resulting in five false negative results. Finally, we evaluated six process blank spectra and the algorithm correctly identified them all as not containing any of the sources. The results for classification of sub-groups are summarized in Table 3.

We evaluated a partial least squares (PLS) and a linear model to interpret the EEM spectra in the same manner as the CNN. The linear model achieved an overall accuracy of 68% and the PLS model had an overall accuracy of only 40%. Both the PLS and linear models performed poorly largely due to an inability to accurately predict diesel concentration. We included the results of these comparative models in section S7.

We tested the limits of our algorithm by applying it to a set of twelve

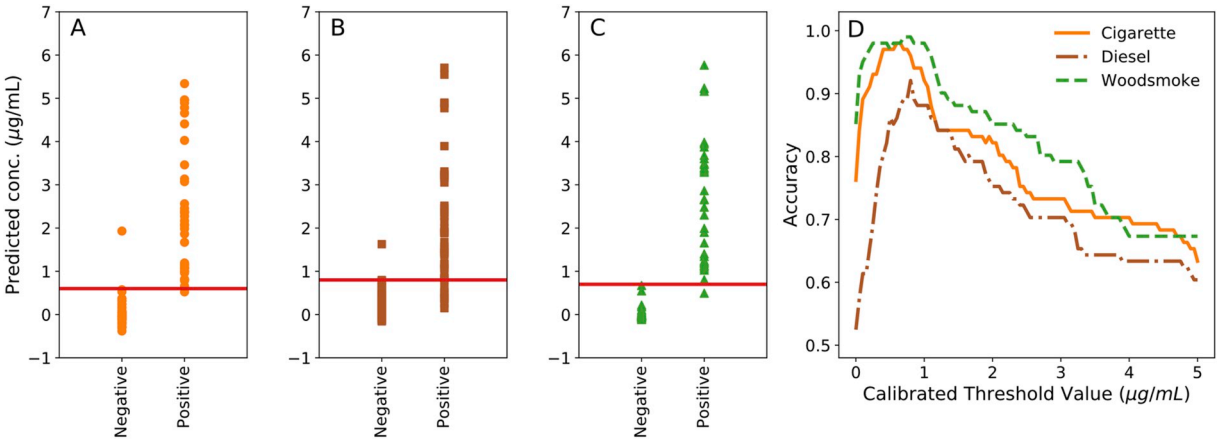


Fig. 5. Classification plots showing classification of test data for (a) cigarette, (b) diesel and (c) woodsmoke sources as present or absent. Data points above the threshold (red horizontal line) are predicted as positive for the source. The location of the threshold was chosen to give the maximum accuracy for classifying each source individually. The source detection accuracies vs. calibrated thresholds are shown in (d). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Classification results for sample sub-groups containing spectra with only one PM source and mixtures (two or three sources). The overall accuracy for the single source and mixtures groups were 91% and 81% respectively.

	Single Source			Mixtures			Process Blanks		
	Cigarette	Diesel	Wood	Cigarette	Diesel	Wood	Cigarette	Diesel	Wood
True +	16	21	17	21	16	20	0	0	0
True -	51	45	51	5	6	6	6	6	6
False +	1	1	0	0	0	0	0	0	0
False -	1	2	1	0	5	0	0	0	0
Accuracy	0.97	0.96	0.99	1.00	0.81	1.00	1.00	1.00	1.00
Sensitivity	0.94	0.91	0.94	1.00	0.75	1.00	1.00	1.00	1.00
Specificity	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00

field spectra. We had a limited number of spectra available for this test and the algorithm was optimized for the laboratory samples, therefore the results are limited in scope to understanding potential difficulties when applying this method to a larger set of field samples. EEM spectra from the field samples were mathematically normalized to an extract concentration of 10 μg/mL for ease of interpreting results: the classification threshold used to train the CNN was 1 μg/mL so a spectrum will be classified as positive for a source if that source makes up 10% or more of the normalized EEM. We expect many sources of PM to contribute to the field samples such as crustal dust and biological material. The EEMs from the eight *background* field samples looked most similar to diesel and cigarette spectra (Fig. S7A). The CNN predicted cigarette as present in three *background* samples, diesel as present in four and woodsmoke as present in two. We did not expect any of the *background* samples to contain cigarette smoke, as they came from non-smoking households and buildings, but this source was detected in three samples that had spectra of similar appearance to cigarette smoke. This suggests that some sources of PM have similar EEM spectra. Diesel may have been present as all samples were collected in urban areas of Seattle. Woodsmoke was detected in two *background* samples. The spectra where woodsmoke was detected looked most similar to diesel, but had higher fluorescent intensity than diesel at the same concentration (10 μg/mL). This illustrates that the CNN may give unexpected results when analyzing spectra that are different than spectra used in training. These results illustrate that an appropriate training set containing as many of the expected sources of PM as possible will be key for successful application to real world samples.

Woodsmoke was the expected primary source in an ambient sample taken in the UW cookstove lab and in a sample taken from an open window during a time when forest fire smoke was causing air pollution in Seattle. Woodsmoke was detected in the cookstove lab sample as expected and the EEM resembled other woodsmoke spectra (Fig. S7B). We believe this was due to small amounts of woodsmoke escaping the ventilation system during stove testing. Woodsmoke was not detected in the sample taken during the forest fire smoke episode. We believe this is due to the forest fire smoke having a different composition than the laboratory generated woodsmoke due to a combination of aging during atmospheric transport and different combustion conditions in a forest fire compared to a cookstove. Diesel exhaust was expected and detected as present in a sample taken in a mechanical room at the diesel exhaust exposure facility. We believe this was due to fugitive emissions of diesel exhaust as with the cookstove sample. Cigarette smoke was expected in a sample taken outdoors near a smoking area but only diesel was detected. In retrospect, we believe this was due to minimal amount of cigarette smoke present in the sample as the PM concentration measured by the filter was 6.3 $\mu g/m^3$, while the average concentration measured over the same time period at two nearby air monitoring sites in Seattle was 5.5 $\mu g/m^3$ showing that this outdoor sample likely consisted of a typical mixture of urban PM that would be expected to include diesel exhaust (Puget Sound Clean Air Agency, 2019).

4. Conclusions

We used a CNN model to successfully classify cigarette, diesel, and woodsmoke sources as present or absent in a series of laboratory samples. The limit of detection for our method is 0.7, 2.6, and 0.9 $\mu g/m^3$ in air for cigarette, diesel, and woodsmoke respectively. The CNN was able to identify cigarette and woodsmoke individually and in the presence of the other sources with 98% and 99% accuracy respectively, while classification of diesel was less accurate with an accuracy of 92%, sensitivity of 0.84 and specificity of 0.98. The overall classification accuracy for all three sources was 89%. When testing the limits of our algorithm by classifying field samples, some samples were classified as expected while in others, sources were detected as present even when they were not expected. This illustrates the need for a training data set with samples from more sources is needed for application to field samples. In future work we will further evaluate the performance of the EEM-CNN method on field samples by comparing it to source apportionment performed using an orthogonal method (e.g. X-Ray Fluorescence). Additionally, we expect atmospheric aging of PM may change EEM spectra in intensity and/or overall appearance. A controlled study of these effects using an atmospheric chamber would be valuable to further evaluate this method.

Notes

The authors declare no competing financial interest.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the National Institute of Biomedical Imaging and Bioengineering, United States (NIBIB) grant U01 EB021923. Data analysis was supported by the Data Intensive Research Enabling Clean Technology (DIRECT) NSF National Research Traineeship, United States (DGE-1633216). We thank Ben Sullivan, Garrett Allawatt, and Devin Udesen of the UW Clean Cookstoves Lab for their assistance collecting woodsmoke samples, Jim Stewart for his help collecting diesel samples, and Tim Gould for his guidance on weighing filters and sample collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2019.117065.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al., 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

Adams, K., Greenbaum, D.S., Shaikh, R., Erp, A. M. van, Russell, A.G., 2015. Particulate matter components, sources, and health: systematic approaches to testing effects. J. Air Waste Manag. Assoc. 65 (5), 544–558. https://doi.org/10.1080/10962247.2014.1001884.

Allen, R., Box, M., Liu, L.-J.S., Larson, T.V., 2001. A cost-effective weighing chamber for particulate matter filters. J. Air Waste Manag. Assoc. 51 (12), 1650–1653. https://doi.org/10.1080/10473289.2001.10464392.

Andrade-Eiroa, Á., Canle, M., Cerdá, V., 2013. Environmental applications of excitation-emission spectrofluorimetry: an in-depth review II. Appl. Spectrosc. Rev. 48 (2), 77–141. https://doi.org/10.1080/05704928.2012.692105.

Aryal, R., Lee, B.-K., Beecham, S., Kandasamy, J., Aryal, N., Parajuli, K., 2015. Characterisation of road dust organic matter as a function of particle size: a PARAFAC approach. Water Air Soil Pollut. 226 (2), 24. https://doi.org/10.1007/s11270-014-2289-y.

Bell, M.L., Ebisu, K., Leaderer, B.P., Gent, J.F., Lee, H.J., Koutrakis, P., Wang, Y., Dominici, F., Peng, R.D., 2014. Associations of $PM_{2.5}$ constituents and sources with hospital admissions: analysis of four counties in Connecticut and Massachusetts (USA) for persons $\geq$ 65 Years of Age. Environ. Health Perspect. 122 (2), 138–144. https://doi.org/10.1289/ehp.1306656.

Bieroza, M., Baker, A., Bridgeman, J., 2009. Exploratory analysis of excitation-emission matrix fluorescence spectra with self-organizing maps as a basis for determination of organic matter removal efficiency at water treatment works. J. Geophys. Res. Biogeosci. 114 (G4) https://doi.org/10.1029/2009JG000940.

Birks, J.B., 1970. Photophysics of Aromatic Molecules. Wiley monographs in chemical physics; Wiley-Interscience, London, New York.

Borysiak, M.D., Thompson, M.J., Posner, J.D., 2016. Translating diagnostic assays from the laboratory to the clinic: analytical and clinical Metrics for device development and evaluation. Lab Chip 16 (8), 1293–1313. https://doi.org/10.1039/c6lc00015k.

Carstea, E.M., Baker, A., Bieroza, M., Reynolds, D., 2010. Continuous fluorescence excitation–emission matrix monitoring of river organic matter. Water Res. 44 (18), 5356–5366. https://doi.org/10.1016/j.watres.2010.06.036.

Chen, Q., Miyazaki, Y., Kawamura, K., Matsumoto, K., Coburn, S., Volkamer, R., Iwamoto, Y., Kagami, S., Deng, Y., Ogawa, S., et al., 2016. Characterization of chromophoric water-soluble organic matter in urban, forest, and marine aerosols by HR-ToF-AMS analysis and excitation–emission matrix spectroscopy. Environ. Sci. Technol. 50 (19), 10351–10360. https://doi.org/10.1021/acs.est.6b01643.

Chollet, F.. Keras. https://keras.io (accessed Sep 21, 2018).

Clevert, D.-A., Unterthiner, T., Hochreiter, S., 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *ArXiv151107289 Cs*.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50 (9), 4712–4721. https://doi.org/10.1021/acs.est.5b06121.

Dumoulin, V., Visin, F., 2016. A guide to convolution arithmetic for deep learning, pp. 12–18. ArXiv160307285 Cs Stat.

Duncan, G.E., Seto, E., Avery, A.R., Oie, M., Carvlin, G., Austin, E., Shirai, J.H., He, J., Ockerman, B., Novosselov, I., 2018. Usability of a personal air pollution monitor: design-feedback iterative cycle study. JMIR MHealth UHealth 6 (12), e12023. https://doi.org/10.2196/12023.

Elcoroaristizabal, S., de Juan, A., García, J.A., Durana, N., Alonso, L., 2014. Comparison of second-order multivariate methods for screening and determination of PAHs by total fluorescence spectroscopy. Chemometr. Intell. Lab. Syst. 132, 63–74. https://doi.org/10.1016/j.chemolab.2014.01.005.

Elcoroaristizabal, S., Juan, A. de, García, J.A., Elorduy, I., Durana, N., Alonso, L., 2014. Chemometric determination of PAHs in aerosol samples by fluorescence spectroscopy and second-order data analysis algorithms. J. Chemom. 28 (4), 260–271. https://doi.org/10.1002/cem.2604.

Engelborghs, Y., Visser, A.J.W.G., 2014. Fluorescence spectroscopy and microscopy: methods and protocols. In: Clifton, N.J. (Ed.), Methods in Molecular Biology. Humana Press, New York, p. 1076.

García-Reiriz, A., Damiani, P.C., Olivieri, A.C., 2007. Analysis of amoxicillin in human urine by photo-activated generation of fluorescence excitation–emission matrices and artificial neural networks combined with residual bilinearization. Anal. Chim. Acta 588 (2), 192–199. https://doi.org/10.1016/j.aca.2007.02.020.

García-Reiriz, A., Damiani, P.C., Olivieri, A.C., Cañada-Cañada, F., Muñoz de la Peña, A., 2008. Nonlinear four-way kinetic-excitation-emission fluorescence data processed by a variant of parallel factor Analysis and by a neural network model achieving the second-order advantage: malonaldehyde determination in olive oil samples. Anal. Chem. 80 (19), 7248–7256. https://doi.org/10.1021/ac8007829.

Godish, T., 1997. Air Quality, third ed. CRC/Lewis Publishers, Boca Raton, Fla.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gould, T., Larson, T., Stewart, J., Kaufman, J.D., Slater, D., Mcewen, N., 2008. A controlled inhalation diesel exhaust exposure facility with dynamic feedback control of PM concentration. Inhal. Toxicol. 20 (1), 49–52. https://doi.org/10.1080/08958370701758478.

Hastie, T., 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, [corrected at 7th printing].; Springer Series in Statistics. Springer, New York, NY, USA.

Hime, N., Marks, G., Cowie, C., Hime, N.J., Marks, G.B., Cowie, C.T., 2018. A comparison of the health effects of ambient particulate matter air pollution from five emission sources. Int. J. Environ. Res. Public Health 15 (6), 1206. https://doi.org/10.3390/ijerph15061206.

Institute for Health Metrics and Evaluation (IHME), 2018. GBD Compare Data Visualization. IHME, University of Washington, Seattle, WA. http://vizhub.healthdata.org/gbd-compare (accessed Oct 9, 2019).

Janssen, N.A.H., Hoek, G., Simic-Lawson, M., Fischer, P., Van Bree, L., Ten Brink, H., Keuken, M., Atkinson, R.W., Anderson, H.R., Brunekreef, B., et al., 2011. Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM10 and PM2.5. Environ. Health Perspect. 119 (12), 1691–1699. https://doi.org/10.1289/ehp.1003369.

Johnson, D.W., Callis, J.B., Christian, G.D., 1977. Rapid scanning fluorescence spectroscopy. Anal. Chem. 49 (8), 747A–757A. https://doi.org/10.1021/ac50016a769.

Karagulian, F., Belis, C.A., Dora, C.F.C., Prüss-Ustün, A.M., Bonjour, S., Adair-Rohani, H., Amann, M., 2015. Contributions to cities' ambient particulate matter (PM): a systematic review of local source contributions at global level. Atmos. Environ. 120, 475–483. https://doi.org/10.1016/j.atmosenv.2015.08.087.

Koenig, J.Q., 1999. Air pollution and asthma. J. Allergy Clin. Immunol. 104 (4), 717–722. https://doi.org/10.1016/S0091-6749(99)70280-0.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems -, vol. 1. Curran Associates Inc., USA, pp. 1097–1105. NIPS'12.

Lee, H., Grosse, R., Ranganath, R., Ng, A.Y., 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning; ICML '09. ACM, New York, NY, USA, pp. 609–616.

Lewis, S.A., Antoniak, M., Venn, A.J., Davies, L., Goodwin, A., Salfield, N., Britton, J., Fogarty, A.W., 2005. Secondhand smoke, dietary fruit intake, road traffic exposures, and the prevalence of asthma: a cross-sectional study in young children. Am. J. Epidemiol. 161 (5), 406–411. https://doi.org/10.1093/aje/kwi059.

Lin, P., Fleming, L.T., Nizkorodov, S.A., Laskin, J., Laskin, A., 2018. Comprehensive molecular characterization of atmospheric Brown carbon by high resolution mass spectrometry with electrospray and atmospheric pressure photoionization. Anal. Chem. 90 (21), 12493–12502. https://doi.org/10.1021/acs.analchem.8b02177.

Matos, J.T.V., Freire, S.M.S.C., Duarte, R.M.B.O., Duarte, A.C., 2015. Natural organic matter in urban aerosols: comparison between water and alkaline soluble components using excitation–emission matrix fluorescence spectroscopy and multiway data analysis. Atmos. Environ. 102, 1–10. https://doi.org/10.1016/j.atmosenv.2014.11.042.

Miller, F.J., Gardner, D.E., Graham, J.A., Lee, R.E., Wilson, W.E., Bachmann, J.D., 1979. Size considerations for establishing a standard for inhalable particles. J. Air Pollut. Control Assoc. 29 (6), 610–615. https://doi.org/10.1080/00022470.1979.10470831.

Mladenov, N., Lípez-Ramos, J., McKnight, D.M., Rechea, I., 2009. Alpine lake optical properties as sentinels of dust deposition and global change. Limnol. Oceanogr. 54 (6part2), 2386–2400. https://doi.org/10.4319/lo.2009.54.6_part_2.2386.

Mladenov, N., Alados-Arboledas, L., Olmo, F.J., Lyamani, H., Delgado, A., Molina, A., Reche, I., 2011. Applications of optical spectroscopy and stable isotope analyses to organic aerosol source discrimination in an urban area. Atmos. Environ. 45 (11), 1960–1969. https://doi.org/10.1016/j.atmosenv.2011.01.029.

Moerner, W.E., Fromm, D.P., 2003. Methods of single-molecule fluorescence spectroscopy and microscopy. Rev. Sci. Instrum. 74 (8), 3597–3619. https://doi.org/10.1063/1.1589587.

Murphy, K.R., Butler, K.D., Spencer, R.G.M., Stedmon, C.A., Boehme, J.R., Aiken, G.R., 2010. Measurement of dissolved organic matter fluorescence in aquatic environments: an interlaboratory comparison. Environ. Sci. Technol. 44 (24), 9405–9412. https://doi.org/10.1021/es102362t.

Nahorniak, M.L., Booksh, K.S., 2006. Excitation-emission matrix fluorescence spectroscopy in conjunction with multiway analysis for PAH detection in complex matrices. Analyst 131 (12), 1308–1315. https://doi.org/10.1039/B609875D.

Nakajima, H., Okada, K., Kuroki, Y., Nakama, Y., Handa, D., Arakaki, T., Tanahara, A., 2008. Photochemical formation of peroxides and fluorescence characteristics of the water-soluble fraction of bulk aerosols collected in okinawa, Japan. Atmos. Environ. 42 (13), 3046–3058. https://doi.org/10.1016/j.atmosenv.2007.12.045.

National Asthma Education and Prevention Program, 2007. Expert panel report 3 (EPR-3): guidelines for the diagnosis and management of asthma–summary report 2007. J. Allergy Clin. Immunol. 120 (5), S94–S138. https://doi.org/10.1016/j.jaci.2007.09.029.

Puget Sound clean air agency - air graphing tool. https://secure.pscleanair.org/airgraphing (accessed Jan 22, 2019).

Rieger, R., Müllen, K., 2010. Forever young: polycyclic aromatic hydrocarbons as model cases for structural and optical studies. J. Phys. Org. Chem. 23 (4), 315–325. https://doi.org/10.1002/poc.1644.

Samburova, V., Connolly, J., Gyawali, M., Yatavelli, R.L.N., Watts, A.C., Chakrabarty, R. K., Zielinska, B., Moosmüller, H., Khlystov, A., 2016. Polycyclic aromatic hydrocarbons in biomass-burning emissions and their contribution to light absorption and aerosol toxicity. Sci. Total Environ. 568, 391–401. https://doi.org/10.1016/j.scitotenv.2016.06.026.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs.

Song, X.-H., Hopke, P.K., 1996. Solving the chemical mass balance problem using an artificial neural network. Environ. Sci. Technol. 30 (2), 531–535. https://doi.org/10.1021/es950281o.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Stanaway, J.D., Afshin, A., Gakidou, E., Lim, S.S., Abate, D., Abate, K.H., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., et al., 2018. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. The Lancet 392 (10159), 1923–1994. https://doi.org/10.1016/S0140-6736(18)32225-6.

Stanek, L.W., Sacks, J.D., Dutton, S.J., Dubois, J.-J.B., 2011. Attributing health effects to apportioned components and sources of particulate matter: an evaluation of collective results. Atmos. Environ. 45 (32), 5655–5663. https://doi.org/10.1016/j.atmosenv.2011.07.023.

Sullivan, B., Allawatt, G., Emery, A., Means, P., Kramlich, J., Posner, J., 2017. Time-resolved particulate emissions monitoring of cookstove biomass combustion using a tapered element oscillating microbalance. Combust. Sci. Technol. 189 (6), 923–936. https://doi.org/10.1080/00102202.2016.1253564.

Tholen, D.W., 2004. Protocols for Determination of Limits of Detection and Limits of Quantitation: Approved Guideline. NCCLS, Wayne, Pa.

Tsuda, A., Henry, F.S., Butler, J.P., 2013. Particle transport and deposition: basic physics of particle kinetics. Comp. Physiol. 3 (4), 1437–1471. https://doi.org/10.1002/cphy.c100085.

US EPA, OAR. Table of Historical Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS). https://www.epa.gov/pm-pollution/table-historical-particulate-matter-pm-national-ambient-air-quality-standards-naaqs (accessed May 3, 2018).

West, J.J., Cohen, A., Dentener, F., Brunekreef, B., Zhu, T., Armstrong, B., Bell, M.L., Brauer, M., Carmichael, G., Costa, D.L., et al., 2016. What we breathe impacts our health: improving understanding of the link between air pollution and health. Environ. Sci. Technol. 50 (10), 4895–4904. https://doi.org/10.1021/acs.est.5b03827.

World Health Organization, 2006. Regional Office for Europe. *Air Quality Guidelines: Global Update 2005 : Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide.* World Health Organization Europe, Copenhagen, Denmark.

Zeiler, M.D., Fergus, R., 2013. Visualizing and Understanding Convolutional Networks. https://arxiv.org/abs/1311.2901.

Zepp, R.G., Sheldon, W.M., Moran, M.A., 2004. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. Mar. Chem. 89 (1), 15–36. https://doi.org/10.1016/j.marchem.2004.02.006.